mixed postestimation — Postestimation tools for mixed

### Postestimation commands

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

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>estat df</code></td>
<td>calculate and display degrees of freedom for fixed effects</td>
</tr>
<tr>
<td><code>estat group</code></td>
<td>summarize the composition of the nested groups</td>
</tr>
<tr>
<td><code>estat icc</code></td>
<td>estimate intraclass correlations</td>
</tr>
<tr>
<td><code>estat recovariance</code></td>
<td>display the estimated random-effects covariance matrices</td>
</tr>
<tr>
<td><code>estat sd</code></td>
<td>display variance components as standard deviations and correlations</td>
</tr>
<tr>
<td><code>estat wcorrelation</code></td>
<td>display model-implied within-cluster correlations and standard deviations</td>
</tr>
</tbody>
</table>

The following standard postestimation commands are also available:

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>contrast</code></td>
<td>contrasts and ANOVA-style joint tests of estimates</td>
</tr>
<tr>
<td><code>estat ic</code></td>
<td>Akaike’s and Schwarz’s Bayesian information criteria (AIC and BIC)</td>
</tr>
<tr>
<td><code>estat summarize</code></td>
<td>variance–covariance matrix of the estimators (VCE)</td>
</tr>
<tr>
<td>estimates</td>
<td>cataloging estimation results</td>
</tr>
<tr>
<td><code>hausman</code></td>
<td>Hausman’s specification test</td>
</tr>
<tr>
<td><code>lincom</code></td>
<td>point estimates, standard errors, testing, and inference for linear</td>
</tr>
<tr>
<td><code>lrtest</code></td>
<td>likelihood-ratio test</td>
</tr>
<tr>
<td><code>margins</code></td>
<td>marginal means, predictive margins, marginal effects, and average marginal</td>
</tr>
<tr>
<td><code>marginsplot</code></td>
<td>graph the results from margins (profile plots, interaction plots, etc.)</td>
</tr>
<tr>
<td><code>nlcom</code></td>
<td>point estimates, standard errors, testing, and inference for nonlinear</td>
</tr>
<tr>
<td><code>predict</code></td>
<td>predictions, residuals, influence statistics, and other diagnostic measures</td>
</tr>
<tr>
<td><code>predictnl</code></td>
<td>point estimates, standard errors, testing, and inference for generalized</td>
</tr>
<tr>
<td><code>pwcompare</code></td>
<td>pairwise comparisons of estimates</td>
</tr>
<tr>
<td><code>test</code></td>
<td>Wald tests of simple and composite linear hypotheses</td>
</tr>
<tr>
<td><code>testnl</code></td>
<td>Wald tests of nonlinear hypotheses</td>
</tr>
</tbody>
</table>

1
predict

Description for predict

predict creates a new variable containing predictions such as linear predictions, standard errors, fitted values, residuals, and standardized residuals.

Menu for predict

Statistics > Postestimation

Syntax for predict

Syntax for obtaining BLUPs of random effects and the BLUPs’ standard errors

    predict [type] newvarsspec [if] [in], reffects [reses(newvarsspec)]
        relevel(levelvar)

Syntax for obtaining scores after ML estimation

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

Syntax for obtaining other predictions

    predict [type] newvar [if] [in] [, statistic relevel(levelvar)]

newvarsspec is stub* or newvarlist.

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main</td>
<td></td>
</tr>
<tr>
<td>xb</td>
<td>linear prediction for the fixed portion of the model only; the default</td>
</tr>
<tr>
<td>stdp</td>
<td>standard error of the fixed-portion linear prediction</td>
</tr>
<tr>
<td>fitted</td>
<td>fitted values, fixed-portion linear prediction plus contributions based on predicted random effects</td>
</tr>
<tr>
<td>residuals</td>
<td>residuals, response minus fitted values</td>
</tr>
<tr>
<td>*rstandard</td>
<td>standardized residuals</td>
</tr>
</tbody>
</table>

Unstarred statistics are available both in and out of sample; type predict ... if e(sample) ... if wanted only for the estimation sample. Starred statistics are calculated only for the estimation sample, even when if e(sample) is not specified.
xb, the default, calculates the linear prediction $x\beta$ based on the estimated fixed effects (coefficients) in the model. This is equivalent to fixing all random effects in the model to their theoretical mean value of 0.

stdp calculates the standard error of the linear predictor $x\beta$.

reffects calculates best linear unbiased predictions (BLUPs) of the random effects. By default, BLUPs for all random effects in the model are calculated. However, if the relevel(levelvar) option is specified, then BLUPs for only level levelvar in the model are calculated. For example, if classes are nested within schools, then typing

```
predict b*, reffects relevel(school)
```

would produce BLUPs at the school level. You must specify $q$ new variables, where $q$ is the number of random-effects terms in the model (or level). However, it is much easier to just specify stub* and let Stata name the variables stub1, stub2, ..., stubq for you.

Rabe-Hesketh and Skrondal (2012, sec. 2.11.2) discuss the link between the empirical Bayes predictions and BLUPs and how these predictions are unbiased. They are unbiased when the groups associated with the random effects are expected to vary in repeated samples. If you expect the groups to be fixed in repeated samples, then these predictions are no longer unbiased.

reses(newvarsspec) calculates the standard errors of the BLUPs of the random effects. By default, standard errors for all BLUPs in the model are calculated. However, if the relevel(levelvar) option is specified, then standard errors for only level levelvar in the model are calculated; see the reffects option.

You must specify $q$ new variables, where $q$ is the number of random-effects terms in the model (or level). However, it is much easier to just specify stub* and let Stata name the variables stub1, stub2, ..., stubq for you. The new variables will have the same storage type as the corresponding random-effects variables.

The reffects and reses() options often generate multiple new variables at once. When this occurs, the random effects (or standard errors) contained in the generated variables correspond to the order in which the variance components are listed in the output of mixed. Still, examining the variable labels of the generated variables (with the describe command, for instance) can be useful in deciphering which variables correspond to which terms in the model.

fitted calculates fitted values, which are equal to the fixed-portion linear predictor $x\beta$ plus contributions based on predicted random effects, or in mixed-model notation, $x\beta + Zu$. By default, the fitted values take into account random effects from all levels in the model; however, if the relevel(levelvar) option is specified, then the fitted values are fit beginning with the topmost level down to and including level levelvar. For example, if classes are nested within schools, then typing

```
predict yhat_school, fitted relevel(school)
```

would produce school-level predictions. That is, the predictions would incorporate school-specific random effects but not those for each class nested within each school.

residuals calculates residuals, equal to the responses minus fitted values. By default, the fitted values take into account random effects from all levels in the model; however, if the relevel(levelvar) option is specified, then the fitted values are fit beginning at the topmost level down to and including level levelvar.

rstandard calculates standardized residuals, equal to the residuals multiplied by the inverse square root of the estimated error covariance matrix.
scores calculates the parameter-level scores, one for each parameter in the model including regression coefficients and variance components. The score for a parameter is the first derivative of the log likelihood (or log pseudolikelihood) with respect to that parameter. One score per highest-level group is calculated, and it is placed on the last record within that group. Scores are calculated in the estimation metric as stored in \texttt{e(b)}.

\texttt{scores} is not available after restricted maximum-likelihood (REML) estimation.

\texttt{relevel(levelvar)} specifies the level in the model at which predictions involving random effects are to be obtained; see the options above for the specifics. \texttt{levelvar} is the name of the model level and is either the name of the variable describing the grouping at that level or is \_all, a special designation for a group comprising all the estimation data.

## margins

### Description for margins

\texttt{margins} estimates margins of response for linear predictions.

### Menu for margins

Statistics  >  Postestimation

### Syntax for margins

\begin{verbatim}
margins \[marginlist\] \[, options\] 
margins \[marginlist\], predict(statistic ...) \[options\]
\end{verbatim}

<table>
<thead>
<tr>
<th>statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{xb}</td>
<td>linear predictor for the fixed portion of the model only; the default</td>
</tr>
<tr>
<td>\texttt{reffects}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
<tr>
<td>\texttt{scores}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
<tr>
<td>\texttt{stdp}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
<tr>
<td>\texttt{fitted}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
<tr>
<td>\texttt{residuals}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
<tr>
<td>\texttt{standard}</td>
<td>not allowed with \texttt{margins}</td>
</tr>
</tbody>
</table>

Statistics not allowed with \texttt{margins} are functions of stochastic quantities other than \texttt{e(b)}.

For the full syntax, see \texttt{[R] margins}. 
test and testparm

Description for test and testparm

test and testparm, by default, perform $\chi^2$ tests of simple and composite linear hypotheses about the parameters for the most recently fit mixed model. They also support $F$ tests with a small-sample adjustment for fixed effects.

Menu for test and testparm

Statistics > Postestimation

Syntax for test and testparm

test (spec) [(spec) ...] [, test_options small]

testparm varlist[, testparm_options small]

Options for test and testparm

- **test_options**: see [R] test options. Options df(), common, and nosvyadjust may not be specified together with small.

- **testparm_options**: see options of testparm in [R] test. Options df() and nosvyadjust may not be specified together with small.

small specifies that $F$ tests for fixed effects be carried out with the denominator degrees of freedom (DDF) obtained by the same method used in the most recently fit mixed model. If option dfmethod() is not specified in the previous mixed command, option small is not allowed. For certain methods, the DDF for some tests may not be available. See Small-sample inference for fixed effects in [ME] mixed for more details.
lincom

Description for lincom

lincom, by default, computes point estimates, standard errors, z statistics, p-values, and confidence intervals for linear combinations of coefficients after mixed. lincom also provides t statistics for linear combinations of the fixed effects, with the degrees of freedom calculated by the DF method specified in option dfmethod() of mixed.

Menu for lincom

Statistics  >  Postestimation

Syntax for lincom

lincom exp [, lincom_options small]

Options for lincom

lincom_options; see [R] lincom options. Option df() may not be specified together with small. small specifies that t statistics for linear combinations of fixed effects be displayed with the degrees of freedom obtained by the same method used in the most recently fit mixed model. If option dfmethod() is not specified in the previous mixed command, option small is not allowed. For certain methods, the degrees of freedom for some linear combinations may not be available. See Small-sample inference for fixed effects in [ME] mixed for more details.
contrast

Description for contrast

contrast, by default, performs $\chi^2$ tests of linear hypotheses and forms contrasts involving factor variables and their interactions for the most recently fit mixed model. contrast also supports tests with small-sample adjustments after mixed, dfmethod().

Menu for contrast

Statistics > Postestimation

Syntax for contrast

contrast termlist [, contrast_options small]

Options for contrast

contrast_options; see [R] contrast options. Options df() and nosvyadjust may not be specified together with small.

small specifies that tests for contrasts be carried out with the DDF obtained by the same method used in the most recently fit mixed model. If option dfmethod() is not specified in the previous mixed command, option small is not allowed. For certain methods, the DDF for some contrasts may not be available. See Small-sample inference for fixed effects in [ME] mixed for more details.
pwcompare

Description for pwcompare

pwcompare performs pairwise comparisons across the levels of factor variables from the most recently fit mixed model. pwcompare, by default, reports the comparisons as contrasts (differences) of margins along with $z$ tests or confidence intervals for the pairwise comparisons. pwcompare also supports $t$ tests with small-sample adjustments after mixed, dfmethod().

Menu for pwcompare

Statistics $>$ Postestimation

Syntax for pwcompare

pwcompare marginlist [, pwcompare_options small]

Options for pwcompare

pwcompare_options; see [R] pwcompare options. Option df() may not be specified together with small.

small specifies that $t$ tests for pairwise comparisons be carried out with the degrees of freedom obtained by the same method used in the most recently fit mixed model with the dfmethod() option. If option dfmethod() is not specified in the previous mixed command, option small is not allowed. For certain methods, the degrees of freedom for some pairwise comparisons may not be available. See Small-sample inference for fixed effects in [ME] mixed for more details.

Remarks and examples

Various predictions, statistics, and diagnostic measures are available after fitting a mixed model using mixed. For the most part, calculation centers around obtaining BLUPs of the random effects. Random effects are not estimated when the model is fit but instead need to be predicted after estimation. Calculation of intraclass correlations, estimating the dependence between responses for different levels of nesting, may also be of interest.

Example 1: Obtaining predictions of random effects and checking model fit

In example 3 of [ME] mixed, we modeled the weights of 48 pigs measured on nine successive weeks as

$$\text{weight}_{ij} = \beta_0 + \beta_1 \text{week}_{ij} + u_{0j} + u_{1j} \text{week}_{ij} + \epsilon_{ij}$$  \hspace{1cm} (1)

for $i = 1, \ldots, 9$, $j = 1, \ldots, 48$, $\epsilon_{ij} \sim N(0, \sigma^2_\epsilon)$, and $u_{0j}$ and $u_{1j}$ normally distributed with mean 0 and variance–covariance matrix

$$\Sigma = \text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \sigma^{2}_{u0} & \sigma_{01} \\ \sigma_{01} & \sigma^{2}_{u1} \end{bmatrix}$$
**Mixed Postestimation**

```plaintext
. use http://www.stata-press.com/data/r15/pig
(Longitudinal analysis of pig weights)
. mixed weight week || id: week, covariance(unstructured)
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0: log likelihood = -868.96185
Iteration 1: log likelihood = -868.96185
Computing standard errors:
Mixed-effects ML regression
Group variable: id
Number of obs = 432
Number of groups = 48
Obs per group:
min = 9
avg = 9.0
max = 9
Wald chi2(1) = 4649.17
Log likelihood = -868.96185
Prob > chi2 = 0.0000

| weight | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|--------|-----------|-------|------|----------------------|
| week   | 6.209896 | 0.0910745  | 68.18 | 0.000 | 6.031393 6.388399   |
| _cons  | 19.35561 | 0.3996387  | 48.43 | 0.000 | 18.57234 20.13889  |

Random-effects Parameters
<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>
id: Unstructured
| var(week) | .3715251   | .0812958 .2419532 .570486 |
| var(_cons)| 6.823363  | 1.566194 4.351297 10.69986 |
| cov(week,_cons) | -.0984378 | .2545767 -.5973991 .4005234 |
| var(Residual)| 1.596829 | .123198 1.372735 1.857505 |

LR test vs. linear model: chi2(3) = 764.58
Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.

Rather than see the estimated variance components listed as variance and covariances as above, we can instead see them as correlations and standard deviations using `estat sd'; see [ME] estat sd.

. estat sd

<table>
<thead>
<tr>
<th>Random-effects Parameters</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>
id: Unstructured
| sd(week)                 | .6095286  | .0666874  | .4918874 .7553052   |
| sd(_cons)                | 2.612157  | .2997895  | 2.085976 3.271064   |
| corr(week,_cons)         | -.0618257 | .1575911  | -.3557072 .243182  |
| sd(Residual)             | 1.263657  | .0487466  | 1.171638 1.362903   |

We can use `estat recovariance' to display the estimated variance components $\hat{\Sigma}$ as a correlation matrix; see [ME] estat recovariance.

. estat recovariance, correlation

Random-effects correlation matrix for level id

<table>
<thead>
<tr>
<th></th>
<th>week</th>
<th>_cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>week</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-.0618257</td>
<td>1</td>
</tr>
</tbody>
</table>
Finally, we can use `estat wcorrelation` to display the within-cluster marginal standard deviations and correlations for one of the clusters; see [ME] `estat wcorrelation`.

```
estat wcorrelation, format(%4.2g)
Standard deviations and correlations for id = 1:
Standard deviations:
         obs |   1    2    3    4    5    6    7    8    9
----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
sd        |  2.9  3.1  3.3  3.7  4.1  4.5  5.0  5.5  6.1
Correlations:
         obs |   1   2  3  4  5  6  7  8  9
----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
   1      |  1.00  0.8  0.77  0.72  0.67  0.59  0.55  0.52
   2      |   0.8  1.00  0.83  0.81  0.78  0.75  0.69  0.66
   3      |   0.77 0.83  1.00  0.86  0.85  0.83  0.79  0.77
   4      |   0.72 0.81 0.86  1.00  0.85  0.83  0.79  0.77
   5      |   0.67 0.78 0.85 0.86  1.00  0.87  0.86  0.85
   6      |   0.59 0.72 0.81 0.87 0.86  1.00  0.90  0.89
   7      |   0.55 0.69 0.79 0.86 0.86 0.90  1.00  0.93
   8      |   0.52 0.66 0.77 0.85 0.89 0.92 0.93  1.00
   9      |   0.55 0.66 0.77 0.85 0.89 0.92 0.93  0.95
```

Because within-cluster correlations can vary between clusters, `estat wcorrelation` by default displays the results for the first cluster. In this example, each cluster (pig) has the same number of observations, and the timings of measurements (week) are the same between clusters. Thus the within-cluster correlations are the same for all the clusters. In example 1 of [ME] `estat wcorrelation`, we fit a model where different clusters have different within-cluster correlations and show how to display these correlations.

We can also obtain BLUPs of the pig-level random effects ($u_{0j}$ and $u_{1j}$). We need to specify the variables to be created in the order $u_1$ $u_0$ because that is the order in which the corresponding variance components are listed in the output (week _cons). We obtain the predictions and list them for the first 10 pigs.

```
predict u1 u0, reffects
by id, sort: generate tolist = (_n==1)
list id u0 u1 if id <=10 & tolist
```

<table>
<thead>
<tr>
<th>id</th>
<th>u0</th>
<th>u1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2369444</td>
<td>-0.3957636</td>
</tr>
<tr>
<td>10</td>
<td>-1.584127</td>
<td>0.510038</td>
</tr>
<tr>
<td>19</td>
<td>-3.526551</td>
<td>0.3200372</td>
</tr>
<tr>
<td>28</td>
<td>1.964378</td>
<td>-0.7719702</td>
</tr>
<tr>
<td>37</td>
<td>1.299236</td>
<td>-0.9241479</td>
</tr>
<tr>
<td>46</td>
<td>-1.147302</td>
<td>-0.5448151</td>
</tr>
<tr>
<td>55</td>
<td>-2.590529</td>
<td>0.0394454</td>
</tr>
<tr>
<td>64</td>
<td>-1.137067</td>
<td>-0.1696566</td>
</tr>
<tr>
<td>73</td>
<td>-3.189545</td>
<td>-0.7365507</td>
</tr>
<tr>
<td>82</td>
<td>1.160324</td>
<td>0.0030772</td>
</tr>
</tbody>
</table>

If you forget how to order your variables in `predict`, or if you use `predict stub*`, remember that `predict` labels the generated variables for you to avoid confusion.
Examining (1), we see that within each pig, the successive weight measurements are modeled as simple linear regression with intercept $\beta_0 + u_{j0}$ and slope $\beta_1 + u_{j1}$. We can generate estimates of the pig-level intercepts and slopes with

```
. generate intercept = _b[_cons] + u0
. generate slope = _b[week] + u1
. list id intercept slope if id<=10 & tolist
```

Thus we can plot estimated regression lines for each of the pigs. Equivalently, we can just plot the fitted values because they are based on both the fixed and the random effects:

```
. predict fitweight, fitted
. twoway connected fitweight week if id<=10, connect(L)
```
We can also generate standardized residuals and see whether they follow a standard normal distribution, as they should in any good-fitting model:

```
predict rs, rstandard
summarize rs
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs</td>
<td>432</td>
<td>1.01e-09</td>
<td>0.8929356</td>
<td>-3.621446</td>
<td>3.000929</td>
</tr>
</tbody>
</table>

```
qnorm rs
```

Example 2: Estimating the intraclass correlation

Following Rabe-Hesketh and Skrondal (2012, chap. 2), we fit a two-level random-effects model for human peak-expiratory-flow rate. The subjects were each measured twice with the Mini-Wright peak-flow meter. It is of interest to determine how reliable the meter is as a measurement device. The intraclass correlation provides a measure of reliability. Formally, in a two-level random-effects model, the intraclass correlation corresponds to the correlation of measurements within the same individual and also to the proportion of variance explained by the individual random effect.
First, we fit the two-level model with mixed:

```
. use http://www.stata-press.com/data/r15/pefrate, clear
(Peak-expiratory-flow rate)
. mixed wm || id:
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:  log likelihood = -184.57839
Iteration 1:  log likelihood = -184.57839
Computing standard errors:
Mixed-effects ML regression  Number of obs = 34
Group variable: id  Number of groups = 17
Obs per group:
  min = 2
  avg = 2.0
  max = 2
Wald chi2(0) = .
```

|     | Coef. | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|-----|-------|-----------|-------|-------|---------------------|
| _cons | 453.9118 | 26.18617 | 17.33 | 0.000 | 402.5878 505.2357   |

```
Random-effects Parameters  Estimate  Std. Err.  [95% Conf. Interval]
id: Identity
  var(_cons) | 11458.94 | 3998.952 | 5782.176 22708.98 |
  var(Residual) | 396.441 | 135.9781 | 202.4039 776.4942 |
LR test vs. linear model: chibar2(01) = 46.27  Prob >= chibar2 = 0.0000
```

Now we use `estat icc` to estimate the intraclass correlation:

```
. estat icc
Intraclass correlation
Level      ICC  Std. Err.  [95% Conf. Interval]
  id      .9665602   .0159495   .9165853   .9870185
```

This correlation is close to 1, indicating that the Mini-Wright peak-flow meter is reliable. But as noted by Rabe-Hesketh and Skrondal (2012), the reliability is not only a characteristic of the instrument but also of the between-subject variance. Here we see that the between-subject standard deviation, \( sd(\text{cons}) \), is much larger than the within-subject standard deviation, \( sd(\text{Residual}) \).

In the presence of fixed-effects covariates, `estat icc` reports the residual intraclass correlation, the correlation between measurements conditional on the fixed-effects covariates. This is equivalent to the correlation of the model residuals.

In the presence of random-effects covariates, the intraclass correlation is no longer constant and depends on the values of the random-effects covariates. In this case, `estat icc` reports conditional intraclass correlations assuming 0 values for all random-effects covariates. For example, in a two-level model, this conditional correlation represents the correlation of the residuals for two measurements on the same subject, which both have random-effects covariates equal to 0. Similarly to the interpretation of intercept variances in random-coefficient models (Rabe-Hesketh and Skrondal 2012, chap. 4),
interpretation of this conditional intraclass correlation relies on the usefulness of the 0 baseline values of random-effects covariates. For example, mean centering of the covariates is often used to make a 0 value a useful reference.

See [ME] estat icc for more information.

Example 3: Estimating residual intraclass correlations

In example 4 of [ME] mixed, we estimated a Cobb–Douglas production function with random intercepts at the region level and at the state-within-region level:

\[ y_{jk} = X_{jk}\beta + u^{(3)}_k + u^{(2)}_{jk} + \epsilon_{jk} \]

. use http://www.stata-press.com/data/r15/productivity
(Public Capital Productivity)
. mixed gsp private emp hwy water other unemp || region: || state:
(output omitted)

We can use estat group to see how the data are broken down by state and region:

. estat group

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td>region</td>
<td>9</td>
<td>51</td>
</tr>
<tr>
<td>state</td>
<td>48</td>
<td>17</td>
</tr>
</tbody>
</table>

We are reminded that we have balanced productivity data for 17 years for each state.

We can use predict, fitted to get the fitted values

\[ \hat{y}_{jk} = X_{jk}\hat{\beta} + \hat{u}^{(3)}_k + \hat{u}^{(2)}_{jk} \]

but if we instead want fitted values at the region level, that is,

\[ \hat{y}_{jk} = X_{jk}\hat{\beta} + \hat{u}^{(3)}_k \]

we need to use the relevel() option:
. predict gsp_region, fitted relevel(region)
. list gsp gsp_region in 1/10

<table>
<thead>
<tr>
<th></th>
<th>gsp</th>
<th>gsp_region</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>10.25478</td>
<td>10.40529</td>
</tr>
<tr>
<td>2.</td>
<td>10.2879</td>
<td>10.42336</td>
</tr>
<tr>
<td>3.</td>
<td>10.35147</td>
<td>10.47343</td>
</tr>
<tr>
<td>4.</td>
<td>10.41721</td>
<td>10.52648</td>
</tr>
<tr>
<td>5.</td>
<td>10.42671</td>
<td>10.54947</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>gsp</th>
<th>gsp_region</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.</td>
<td>10.4224</td>
<td>10.53537</td>
</tr>
<tr>
<td>7.</td>
<td>10.4847</td>
<td>10.60781</td>
</tr>
<tr>
<td>8.</td>
<td>10.53111</td>
<td>10.64727</td>
</tr>
<tr>
<td>9.</td>
<td>10.59573</td>
<td>10.70503</td>
</tr>
<tr>
<td>10.</td>
<td>10.62082</td>
<td>10.72794</td>
</tr>
</tbody>
</table>

Technical note

Out-of-sample predictions are permitted after `mixed`, but if these predictions involve BLUPs of random effects, the integrity of the estimation data must be preserved. If the estimation data have changed since the mixed 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.

We can use `estat icc` to estimate residual intraclass correlations between productivity years in the same region and in the same state and region.

. estat icc
Residual intraclass correlation

<table>
<thead>
<tr>
<th>Level</th>
<th>ICC</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>region</td>
<td>.159893</td>
<td>.127627</td>
<td>.0287143 .5506202</td>
</tr>
<tr>
<td>state</td>
<td>region</td>
<td>.8516265</td>
<td>.0301733</td>
</tr>
</tbody>
</table>

`estat icc` reports two intraclass correlations for this three-level nested model. The first is the level-3 intraclass correlation at the region level, the correlation between productivity years in the same region. The second is the level-2 intraclass correlation at the state-within-region level, the correlation between productivity years in the same state and region.

Conditional on the fixed-effects covariates, we find that annual productivity is only slightly correlated within the same region, but it is highly correlated within the same state and region. We estimate that state and region random effects compose approximately 85% of the total residual variance.

Example 4: Small-sample adjusted tests for fixed effects

To illustrate the use of `test` and `testparm` with the `small` option for small-sample adjusted tests for fixed effects, we refit the dental veneer data from example 14 of [ME] `mixed` using the Satterthwaite method (option `dfmethod(satterthwaite)`) to compute the DF for fixed effects.
Mixed-effects REML regression

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>No. of Groups</th>
<th>Observations per Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Minimum</td>
<td>Average</td>
<td>Maximum</td>
<td></td>
</tr>
<tr>
<td>patient</td>
<td>12</td>
<td>2</td>
<td>9.2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>tooth</td>
<td>55</td>
<td>2</td>
<td>2.0</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

DF method: Satterthwaite

F(4, 16.49) = 1.87
Prob > F = 0.1638

Log restricted-likelihood = -420.92761

| gcf            | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|----------------|--------|-----------|-------|------|---------------------|
| followup       | 0.3009815 | 1.936863 | 0.16  | 0.879 | -3.963754 4.565717 |
| base_gcf       | -0.0183127 | 0.1433094 | -0.13 | 0.899 | -0.3065704 0.269945 |
| cda            | -0.329303 | 0.5292525 | -0.62 | 0.537 | -1.39197 0.7333636 |
| age            | -0.5773932 | 0.2139656 | -2.70 | 0.022 | -1.051598 -0.1031885 |
| _cons          | 45.73862 | 12.55497 | 3.64  | 0.001 | 19.90352 71.57372 |

Random-effects Parameters

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>patient: Unstructured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(followup)</td>
<td>41.88772</td>
<td>18.79997</td>
</tr>
<tr>
<td>var(_cons)</td>
<td>524.9851</td>
<td>253.0205</td>
</tr>
<tr>
<td>cov(followup,_cons)</td>
<td>-140.4229</td>
<td>66.57623</td>
</tr>
<tr>
<td>tooth: Identity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>var(_cons)</td>
<td>47.45738</td>
<td>16.63034</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>48.86704</td>
<td>10.50523</td>
</tr>
</tbody>
</table>

LR test vs. linear model: chi2(4) = 91.12
Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Now we can, for example, test the hypotheses that all fixed effects are zero by typing

```
. testparm *, small
```

F( 4, 16.49) = 1.87
Prob > F = 0.1638

The $F$ statistic for the overall test is 1.87, and the DDF is estimated to be 16.49. These results are different from the model test using the Kenward–Roger DDF method reported in the header of the estimation output in example 1 of [ME] estat df (the $F$ statistic is 1.47, and the model DDF is 27.96).

The results differ because the Kenward–Roger method uses an adjusted $F$-test statistic and adjusts the fixed-effects variance–covariance estimator for a small sample. Both methods, however, lead to the same conclusion of no joint significance of the fixed effects.
Without option small, the commands test and testparm report large-sample $\chi^2$ Wald tests. We can compare the small-sample and large-sample tests of the joint hypotheses that the coefficient on followup and the coefficient on age equal zero.

```
. test followup = age = 0, small
   ( 1) [gcf]followup - [gcf]age = 0
   ( 2) [gcf]followup = 0
   F(  2, 10.75) =    3.65
   Prob > F =    0.0617

. test followup = age = 0
   ( 1) [gcf]followup - [gcf]age = 0
   ( 2) [gcf]followup = 0
   chi2(  2) =    7.30
   Prob > chi2 =    0.0260
```

The DDF of the $F$ test, which is computed using the Satterthwaite method from our posted results, is 10.75. The $p$-values are very different (0.0617 versus 0.0260), and they lead to different conclusions of whether we should reject the null hypotheses at the $\alpha = 0.05$ level.

Similarly, you can use the small option with lincom to perform small-sample inference for linear combinations of fixed effects.

---

**Example 5: Small-sample adjusted contrasts**

As we did with test, after fitting a mixed model with the dfmethod() option for small-sample adjustment, we can use the small option with contrast to adjust for a small sample when estimating contrasts. Suppose we have collected data on a vigilance performance test. This experiment has been designed to test the response latency scores of two modes of signal during a four-hour monitoring period. This is a split-plot factorial design where signal is the whole-plot factor, hour is the subplot factor, and subject is the block factor. The whole-plot factor and the subplot factor are fixed; the block factor is random. Also, suppose that two measurements are missing in this dataset.

```
. use http://www.stata-press.com/data/r15/vptscores, clear
  (Vigilance Performance Test Scores With Missing Data)
. tabdisp subject hour, cellvar(score) by(signal) concise missing
```

<table>
<thead>
<tr>
<th>Subject and Monitoring period</th>
<th>Auditory</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1 2 3 4 1 2 3 4</td>
<td>5 6 7 8</td>
</tr>
<tr>
<td>1</td>
<td>3 4 7 7</td>
<td>1 2 5 10</td>
</tr>
<tr>
<td>2</td>
<td>6 5 . 8</td>
<td>2 3 6 .</td>
</tr>
<tr>
<td>3</td>
<td>3 4 7 9</td>
<td>2 4 5 9</td>
</tr>
<tr>
<td>4</td>
<td>3 3 6 8</td>
<td>2 3 6 11</td>
</tr>
</tbody>
</table>

We start by fitting a mixed model. Because the dataset is small and unbalanced, we apply the Kenward–Roger method for small-sample adjustment:
. mixed score signal#hour || subject:, reml dfmethod(kroger) nolog nogroup

Mixed-effects REML regression
Number of obs = 30
DF method: Kenward-Roger
DF: min = 16.02
    avg = 16.76
    max = 18.29
F(7, 16.08) = 43.84
Log restricted-likelihood = -32.9724
Prob > F = 0.0000

| score      | Coef. | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|------------|-------|-----------|-------|------|----------------------|

signal
  Visual    -2  .6288677  -3.18  0.005  -3.319693  -.6803071

hour
  2       .25  .5359916   0.47  0.647   -.8861371  1.386137
  3      3.108222  .5911044  5.26  0.000   1.859163  4.357281
  4      4.25  .5359916  7.93  0.000   3.113863  5.386137

signal#hour
  Visual#2  1   .7580066   1.32  0.206  -.6067405  2.606741
  Visual#3   .6417778   .7979294  0.80  0.433  -1.046666  2.330221
  Visual#4  4.044205   .7979294  5.07  0.000   2.355762  5.732649

_cons     3.75  .4446766   8.43  0.000   2.816836  4.683164

Random-effects Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
</table>

subject: Identity
  var(_cons) .2163751   .2345718   .0258477  1.811312

  var(Residual)  .574574   .2062107  .2843515  1.161011

LR test vs. linear model: chibar2(01) = 1.55
Prob >= chibar2 = 0.1069

We can test the main effects and the interaction effects by typing the contrast command. With the small option, contrast reports small-sample adjusted $F$ tests. Without the small option, contrast performs large-sample $\chi^2$ Wald tests. Below is the comparison of the small-sample and the large-sample contrasts:
From these results, we can see that the \( p \)-values for the main effect of signal and the interaction effect vary between small-sample and large-sample tests. However, both tests indicate that the hour effect and the interaction effects are significant. We can decompose the interaction effect into separate interaction contrasts for further investigation.
. contrast r.signal#ar.hour, small

From previous analysis, we already knew the overall interaction was significant. From the decomposition, we can easily see that the overall significance is driven by differences in the third and fourth hours; the change in response latency from hour three to hour four is greater for visual signals than for auditory signals.

We can also calculate the pairwise differences of the hourly marginal means by typing the `pwcompare` command. With the `small` option, `pwcompare` reports small-sample adjusted pairwise comparisons along with the degrees of freedom for each pairwise comparison.

. pwcompare hour, small

When we compare these results with the large-sample results below, we can see that the confidence interval of hour 2 versus hour 1 changes to include 0. Therefore, after adjusting for small-sample
size, we would not reject the hypothesis that the means for hour 1 and hour 2 are equivalent at the
5% significance level.

    . pwcompare hour
Pairwise comparisons of marginal linear predictions
Margins : asbalanced

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>hour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 vs 1</td>
<td>.75</td>
<td>.3790033</td>
</tr>
<tr>
<td>3 vs 1</td>
<td>3.4291111</td>
<td>.3971529</td>
</tr>
<tr>
<td>4 vs 1</td>
<td>6.272103</td>
<td>.3971529</td>
</tr>
<tr>
<td>3 vs 2</td>
<td>2.679111</td>
<td>.3971529</td>
</tr>
<tr>
<td>4 vs 2</td>
<td>5.522103</td>
<td>.3971529</td>
</tr>
<tr>
<td>4 vs 3</td>
<td>2.842991</td>
<td>.4145085</td>
</tr>
</tbody>
</table>

Stored results
pwcompare with option small stores the following in r():
Matrices
r(L_df) degrees of freedom for each margin difference
r(M_df) degrees of freedom for each margin estimate

pwcompare with options post and small stores the following in e():
Matrices
e(L_df) degrees of freedom for each margin difference
e(M_df) degrees of freedom for each margin estimate

Methods and formulas
Methods and formulas are presented under the following headings:

Prediction
Small-sample inference

Prediction
Following the notation defined throughout [ME] mixed, BLUPs of random effects $\tilde{u}$ are obtained as

$$
\tilde{u} = \tilde{G}Z'\tilde{V}^{-1} \left( y - X\hat{\beta} \right)
$$

where $\tilde{G}$ and $\tilde{V}$ are $G$ and $V = ZGZ' + \sigma^2 \epsilon R$ with maximum likelihood (ML) or REML estimates of the variance components plugged in. Standard errors for BLUPs are calculated based on the iterative technique of Bates and Pinheiro (1998, sec. 3.3) for estimating the BLUPs themselves. If estimation is done by REML, these standard errors account for uncertainty in the estimate of $\beta$, while for ML the standard errors treat $\beta$ as known. As such, standard errors of REML-based BLUPs will usually be larger.
Fitted values are given by $X\hat{\beta} + Z\tilde{u}$, residuals as $\hat{\epsilon} = y - X\hat{\beta} - Z\tilde{u}$, and standardized residuals as

$$\hat{\epsilon}_* = \hat{\sigma}_\epsilon^{-1} \hat{R}^{-1/2} \hat{\epsilon}$$

If the `relevel(levelvar)` option is specified, fitted values, residuals, and standardized residuals consider only those random-effects terms up to and including level `levelvar` in the model.

For details concerning the calculation of scores, see *Methods and formulas* in [ME] mixed.

**Small-sample inference**

For small-sample computations performed when the `small` option is used with `test`, `testparm`, `lincom`, `contrast`, or `pwcompare`, see *Denominator degrees of freedom* in *Methods and formulas* of [ME] mixed.

**References**


**Also see**

[ME] mixed — Multilevel mixed-effects linear regression

[U] 20 Estimation and postestimation commands