How can I fit a random intercept or mixed effects model with heteroskedastic errors in Stata?

Authored by

stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I fit a random intercept or mixed effects model with heteroskedastic errors in Stata?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=164928

In order to fit a random intercept or mixed effects model with heteroskedastic errors in Stata, one must utilize the "xtmixed" command. This command allows for the inclusion of random effects and heteroskedastic errors in the model, which accounts for the varying levels of variability in the data. Additionally, the "robust" option can be added to the command in order to adjust for heteroskedasticity. By using the "xtmixed" command and specifying the appropriate options, one can effectively fit a random intercept or mixed effects model with heteroskedastic errors in Stata.

How can I fit a random intercept or mixed effects model with heteroskedastic errors in Stata? | Stata FAQ

It is common to fit a model where a variable (or variables) has an effect on the expected mean. You may also want to fit a model where a variable has an effect on the variance, that is a model with heteroskedastic errors. One example is a hierarchical model where the error variance is not constant over some categorical variable (e.g., gender). Another example is a longitudinal model where you might want to allow the errors to vary across time. This page describes how to fit such a model and demonstrates with an example.

Example 1: Heteroskedastic errors by group

It may be useful to note that a multi-level model (or a single level model) that allows different error terms between groups is similar to a multi-group structural equation model in which all parameters in the model are constrained to equality except for the error terms.

The dataset for this example includes data on 7185 students in 160 school. This dataset and model come from the HLM manual. The variable mathach is a measure of each student's math achievement, the variable female is a binary variable equal to one if the student is female and zero otherwise, and the variable id contains the school identifier.

Below we run a model that uses female to predict mathach. This model assumes equal error variances for males and females.

use https://stats.idre.ucla.edu/stat/stata/faq/hsb, clear

mixed mathach female || id:, var mle nolog

Mixed-effects ML regression Number of obs = 7185

Group variable: id Number of groups = 160

Obs per group: min = 14

avg = 44.9

max = 67

Wald chi2(1) = 62.89

Log likelihood = -23526.66 Prob > chi2 = 0.0000

--

mathach | Coef. Std. Err. z P>|z|

-------------+--

female | -1.35939 .1714111 -7.93 0.000 -1.69535 -1.02343

_cons | 13.34526 .253935 52.55 0.000 12.84756 13.84297

--

--

Random-effects Parameters | Estimate Std. Err.

-----------------------------+--

id: Identity |

var(_cons) | 8.108982 1.018274 6.339834 10.37181

-----------------------------+--

var(Residual) | 38.84481 .6555316 37.58101 40.15112

--

LR test vs. linear regression: chibar2(01) = 936.66 Prob >= chibar2 = 0.0000

In order to fit a model with heteroskedastic errors, we must modify the above model. The model shown above can be written as:

mathach_ij = b0 + b1*female_ij + u_i + e_ij (1)

This model assumes:

e_ij ~ N(0,s2)
u_i ~ N(0,t2)

Where e_ij is the level 1 errors (i.e., residuals), and u_i is the random intercept across classrooms (i.e., the level 2 variance). We want to allow the variance of e_ij (i.e., s2) to be different for male and female students. One way to think about this would be to replace the single e_ij

term, with
separate terms for males and females, in equation form:

e_ij = e(m)_ij*male + e(f)_ij*female (2)

Where male is a dummy variable equal to one if the student is male and
equal to zero otherwise, and e(m)_ij is the error term for males. The
variable female and the term e(f)_ij are the same except that they
are for females. Another way to write this is:

e_ij ~ N(0,s2_1) males
N(0,s2_2) females

Another way to think about this is to use one
group as the reference group, and estimate the difference between the variance
of the two groups. In equation form this would be:

e_ij = e(m)_ij + e(f)_ij*female (3)

Where e(m)_ij is the error term for males, and e(f)_ij is

the difference between the errors for males and females. This can also be written:

e_ij ~ N(0,s2 + s2_2*female)

This second way of looking at the error structure is part of how we will restructure our model to allow for heteroskedastic error terms.

Equation 3 says that the residual variance is a function of gender. So we can write it as r_ijk, where i is the group, j is the subject, and k is the gender.

mathach_ij = b0 + b1*female + u_i + r_ijk (4)

To model this, we add a level to our model. Now the third level will be classrooms (previously level 2), the second level will be students (previously level 1), and level 1 will be a single case within each student.

The only random effect at level 1 is gender (even the intercept is fixed). The new model can be written as:

mathach_ij = b0 + b1*female_ij + u_i + e_ij*female + r_ij0
r_ijk = r_ij0 male
r_ij1 female=e_ij+r_ij0

Where the level one error
is represented by r_ij0. Because males are the omitted category in the
random portion of the model, the variance of r_ij0 is the variance of the
errors males (i.e., female=0). This is similar to the interpretation of
the intercept in the fixed portion of the mode, where b0 represents the
intercept for males, since they are the omitted category. The difference between
the error terms for males and females is e_ij so the error term for females is
r_ij + e1_ij.

A final consideration before we can actually fit the model is

that Stata restricts the estimates of the error variances to be greater than zero. As a result, we must use the group with the smallest variance as the omitted category, since this will result in the values of the error variances for the group(s) being positive. To establish which group has the lowest variance you can examine the residuals from a model without heteroskedastic errors. In this example, examination of the residuals shows that the residual variance is smaller for females than males.

The code below creates two new variables, male, and nid. The variable nid will be the identifier for the students.

```
recode female (0=1) (1=0), gen(male)
gen nid = _n
```

The resulting data set should look something like the one shown below. Where each classroom (identified by the variable id) has multiple rows in the dataset, but each case

(i.e., each student, identified by the variable nid) has only one row.

id nid
1224 1
1224 2
1224 3
...
1288 48
1288 49
1288 50
...
1296 73
1296 74
1296 75

Once the necessary variables are created, we can run the model as shown below, which allows for a difference in the variance of the errors for males and females. It is necessary to specify the nocons option suppresses the random intercept at level 2, so that the only random

effect at level 2 is gender (i.e., male).

mixed mathach female || id: || nid: male, nocons var mle nolog

Mixed-effects ML regression Number of obs = 7185

| No. of Observations per Group

Group Variable | Groups Minimum Average Maximum

----------------+---------------------------------------

id | 160 14 44.9 67

nid | 7185 1 1.0 1

Wald chi2(1) = 63.03

Log likelihood = -23522.932 Prob > chi2 = 0.0000

--

mathach | Coef. Std. Err. z P>|z|

-------------+--

female | -1.363969 .1718025 -7.94 0.000 -1.700695 -1.027242

_cons | 13.34707 .2548619 52.37 0.000 12.84755 13.84659

Random-effects Parameters | Estimate Std. Err.

-----------------------------+--

id: Identity |

var(_cons) | 8.090527 1.016457 6.324639 10.34946

-----------------------------+--

nid: Identity |

var(male) | 3.622413 1.333432 1.76062 7.452988

-----------------------------+--

var(Residual) | 37.13827 .8657943 35.47953 38.87456

LR test vs. linear regression: chi2(2) = 944.12 Prob > chi2 = 0.0000

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

The second table in the output shows the estimated random effects. The residual variance for females is equal to var(Residuals) = 37.138,

while the variance for males is var(Residuals) + var(male) =

37.1383 + 3.622 = 40.7607. Since the 95% confidence interval for var(male)
does not include zero, we can say that the difference between the variances is
statistically significant at the $p<0.05$ level.

Traditionally, and in some other packages (e.g., HLM), the variances in models with heteroskedastic errors have been parameterized as:

s2_ij = exp(a_0 + a_1*x_ij)

Where s2_ij is the variance of the residuals, x_ij is a dichotomous variable equal to one for one group and zero for the other, and a_0 and a_1 are estimated parameters for the
variance. a_0 is the variance in the group where x_ij = 0, and a_1 is the difference in the variances between the two groups.
If you want to compare the results of your analysis in Stata to those from another
package, it is possible to translate between the two. As a first step, let's
think about what Stata is estimating a little more. One

way to write what Stata models is:

s2_ij = var(Residuals) + var(x_ij)*x_ij

Where var(Residuals) is the variance of the level 1 errors, and var(x_ij) is the random effect of the dummy variable x_ij. Because Stata models the natural log of the standard deviation of the error term, the above is visually clear, but not quite correct. The actual model is:

s2_ij = exp(2*lnsd_0) + exp(2*lnsd_1)*x_ij

Where lnsd_0 is the natural log of the standard deviation of the level 1 errors, and lnsd_1 is the natural log of the standard deviation of the level 2 random effect. We can convert the estimates Stata gives us to a_0 and a_1 using the formulas below (each formula is written twice, the first for visual clarity, the second to reflect what is

actually modeled by
Stata):

a_0 = ln(var(Residuals))
a_0 = ln(exp(2*lnsd_0))

a_1 = ln(var(Residuals)+var(x_ij)) - ln(var(Residuals))
a_1 = ln - ln

We can use the display command to calculate these values. Stata stores the ln(sd) for the
level 1 residuals as _cons, and the ln(sd) of the random effect
of male in _cons. Below we use these values to calculate a_0 and a_1.

di "a_0=" ln(exp(2 * _cons))
a_0=3.6147746

di "a_1=" ln(exp(2 * _cons)+exp(2 * _cons)) - ln(exp(2 * _cons))
a_1=.09308814

Example 2: A growth curve model

When you run a growth curve model in the structural equation modeling framework, it is not uncommon to allow the error variance to be different across time points, or at least to test whether such differences exist. A longitudinal model that allows different error variances across time points is similar to a growth model in the structural equation modeling setting, where all parameters except for the error terms across time points are constrained to equality. Below we show how to allow for different error variances across time points, and how to test whether the error variances are significantly different from each other.

The example dataset contains data on 239 subjects. The dataset includes observations at up to five time points per subject, for a total of 1079 observations. The variable id uniquely identifies subjects. The time at which an observation was taken is indicated by the variable time,

which takes on
the values 0 to 4.

Below we use mixed to fit a model where
the variables x and time
predict the variable attit. This model assumes the variance is the same
across time points.

use https://stats.idre.ucla.edu/stat/stata/faq/nys2, clear

mixed attit x time ||id:, var mle nolog

Mixed-effects ML regression Number of obs = 1079
Group variable: id Number of groups = 239

Obs per group: min = 1
avg = 4.5
max = 5

Wald chi2(2) = 324.65
Log likelihood = 140.31043 Prob > chi2 = 0.0000

--

attit | Coef. Std. Err. z P>|z|

-------------+--

x | .0241245 .0033107 7.29 0.000 .0176357 .0306134
time | .0600884 .0039557 15.19 0.000 .0523353 .0678415
_cons | .1191882 .0184893 6.45 0.000 .0829498 .1554267

Random-effects Parameters | Estimate Std. Err.
-----------------------------+--
id: Identity |
var(_cons) | .0308536 .003533 .0246511 .0386167
-----------------------------+--
var(Residual) | .0311131 .0015204 .0282714 .0342405

LR test vs. linear regression: chibar2(01) = 324.66 Prob >= chibar2 = 0.0000

The model above can be written as:

attit_it = b0 + b1*x_it + b2*time_it + u_i + e_it (4)

It is assumed that:

e_it ~ N(0,s2)
u_i ~ N(0,t2)

Where u_i is the random intercept across individuals (i.e., the level 2 variance), and e_it represents the level 1 errors (i.e., residuals). We want to allow the variance of e_it to be different at each time point. One way to think about this would be to replace the single e_ij term, with separate terms for each time point, in equation form:

e_it = e_i0*t0 + e_i1*t1 + e_i2*t2 + e_i3*t3 + e_i4*t4

Where t0 is a dummy variable equal to one if time = 0 (i.e. the first time point) and equal to zero otherwise, and e_i0 is the error for the first measurement occasion (time=0). The variables t1 to t4, are dummy variables for time points 1 to 4. The terms e_i1 to e_i4 are the error terms for time points 1 to 4.

In order to model the heteroskedastic errors, we add a third level

to our model. In this new model,
the third level will be individuals (previously level 2), the second level will be time
points (previously level 1), and level 1 will be a single case within each time point.
Since the effect of time is in the level at model 2, only random effects for
time are included at level 1. The new model can be written as:

attit_itk = b0 + b1*x_it + b2*time_it + e_i1*t1 + e_i2*t2 + e_i3*t3 + e_i4*t4 + r_ij0

Where the level one error
is represented by r_it0. Because the time=0 is the omitted category, the variance of
r_it0 is the variance of the
errors at time=0. The
random effects e_i1 to e_i4 represent the difference between the variance
of the errors and the variance of the errors at time=0 for each time
point.

A final consideration before we can actually fit the model is that Stata restricts the estimates of the error variances to be greater than zero. As a result, we must use the time point (or other category/group) with the smallest variance as the omitted category, since this will result in the values of the error variances for the other time points being positive. To establish which time point or group has the lowest variance you can examine the residuals from a model without heteroskedastic errors. In this example, the residual variance is smallest when time=0.

In order to fit the model we will need dummy variables for each of the other time points.

```
gen t1 = (time==1)
gen t2 = (time==2)
gen t3 = (time==3)
gen t4 = (time==4)
```

Now we are ready to fit our model using mixed. The nocons option
suppresses the random intercept at level 2.

mixed attit x time ||id: ||time: t1 t2 t3 t4, nocons var mle nolog

Mixed-effects ML regression Number of obs = 1079

--
| No. of Observations per Group
Group Variable | Groups Minimum Average Maximum
----------------+--
id | 239 1 4.5 5
time | 1079 1 1.0 1
--

Wald chi2(2) = 319.37
Log likelihood = 143.67787 Prob > chi2 = 0.0000

--
attit | Coef. Std. Err. z P>|z|
-------------+--
x | .0238029 .0033059 7.20 0.000 .0173235 .0302823
time | .0597567 .0039638 15.08 0.000 .0519878 .0675256

_cons | .1208 .0178047 6.78 0.000 .0859035 .1556965

--

--

Random-effects Parameters | Estimate Std. Err.

-----------------------------+--

id: Identity |

var(_cons) | .0280981 .0034551 .0220805 .0357557

-----------------------------+--

time: Independent |

var(t1) | .0027798 .0046901 .0001018 .0758888

var(t2) | .0058393 .0053178 .0009799 .0347972

var(t3) | .0125841 .0060606 .0048964 .0323421

var(t4) | .01354 .0059784 .0056988 .0321702

-----------------------------+--

var(Residual) | .0247609 .0033388 .0190103 .032251

--

LR test vs. linear regression: chi2(5) = 331.40 Prob > chi2 = 0.0000

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

Once we have run the model, we may want to test

whether the estimated
variances are different either from zero or different from each other. Below nlcom is used to
get the estimated variance at each time point.

```
nlcom (time0: exp(2 * _cons)) ///
(time1: exp(2 * _cons) + exp(2 * _cons)) ///
(time2: exp(2 * _cons) + exp(2 * _cons)) ///
(time3: exp(2 * _cons) + exp(2 * _cons)) ///
(time4: exp(2 * _cons) + exp(2 * _cons))

time0: exp(2 * _cons)
time1: exp(2 * _cons) + exp(2 * _cons)
time2: exp(2 * _cons) + exp(2 * _cons)
time3: exp(2 * _cons) + exp(2 * _cons)
time4: exp(2 * _cons) + exp(2 * _cons)

------------------------------------------------------------------------------
attit | Coef. Std. Err. z P>|z|
-------------+----------------------------------------------------------------
time0 | .0247609 .0033388 7.42 0.000 .018217 .0313048
time1 | .0275406 .0034705 7.94 0.000 .0207385 .0343427
time2 | .0306002 .0035976 8.51 0.000 .0235491 .0376513
time3 | .037345 .0044345 8.42 0.000 .0286536 .0460364
time4 | .0383009 .0044824 8.54 0.000 .0295155 .0470862
```

We can also use nlcom to estimate the difference in the variance between time points,
in this case, the difference between the variance at time = 1 and time = 4.

nlcom (t4_t1: exp(2 * _cons) - exp(2 * _cons))

t4_t1: exp(2 * _cons) - exp(2 * _cons)

----

attit | Coef. Std. Err. z P>|z|

-------------+---

t4_t1 | .0107602 .0060541 1.78 0.076 -.0011055 .0226259

References

Gutierrez, Roberto G. (2008) Tricks of the Trade: Getting the most out of
xtmixed. 2008 Fall North American Stata Users Group Meeting, San Francisco, CA.

Raudenbush, Stephen, Anthony Bryk, Yuk Fai Cheong,

and Richard Congdon
(2001) HLM 5: Hierarchical Linear and Nonlinear Modeling. Scientific
Software International. Lincolnwood, IL.