Lecture 3: Random coefficients model

Antonello Maruotti

Lecturer in Medical Statistics, S3RI and School of Mathematics
University of Southampton

Southampton, 13 June 2014

Longitudinal data - Repeated measurements

- Longitudinal data typically refer to data containing time series observations of a number of units.
- At least two dimension are involved in the analysis of longitudinal data: a cross-sectional dimension, indicated by subscript i, and a time series dimension, indicated by subscript j.
- Repeated measurements are usually time period or units within clusters
- Observations cannot be assumed independently distributed across time

Common features of longitudinal data

- The sample of individuals n is typically relatively large
- The number of time periods (or repeated measurements) J is generally short
- Increasing data availability
- Challenging methodology

More on...Features

- Increased precision of regression estimates
- Repeated observations on units allow to isolate effects of unobserved differences between individuals
- Linear mixed models contains both fixed and random effects.
- Linear mixed models are a generalization of linear regression allowing for the inclusion of random deviations (other than those associated with the error term)

Comments on the required computational burden

- There is a special relationship between the multiple observations of a particular unit
- $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \epsilon_{ij} = \beta_0 + \beta_1 x_{ij} + u_{ij}$
- The presence of α_{0i} produces a correlation among the errors of the same cross-section unit $\Rightarrow Cov(u_{ij}, u_{ij'}) \neq 0$, though $Cov(u_{ij}, u_{i'j}) = 0$
- OLS is inefficient in the random effects model, and yields incorrect standard errors

Model specification

- In linear random intercept models, the overall level of the response, conditional on covariates, could vary across units or clusters
- In random coefficients models, we also allow the marginal effect of the covariates to vary across clusters
- $Y_{ij} = (\beta_0 + \alpha_{0i}) + (\beta_1 + \alpha_{1i})x_{ij} + \epsilon_{ij}$
- This allows for the intercept and slope coefficients to vary across clusters

Model assumptions

- β_0 and β_1 fixed intercept and slope
- $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a_0})$ random intercept effect
- $\alpha_{1i} \sim \mathcal{N}(0, \sigma^2_{a_1})$ random slope over time
- $Cov(\alpha_{0i}, \alpha_{1i}) = \tau_{a_0,a_1}$
- $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ random error, independent of α_{0i}'s and α_{1i}'s

Remarks

- The fixed portion $\mathbf{X}\beta$ is analogous to the standard regression model, with β being the regression coefficient to be estimated.
- We assume α is orthogonal to ϵ
- The random coefficients are not directly estimated, but are characterized by the elements of their variance-covariance matrix (known as variance components)
- A general specification of the random coefficients variance-covariance matrix provides flexibility

Example 1: Weight gain of pigs

Repeated measures are collected on **units of analysis** over time

Consider a study consisting of weight measurements of 48 pigs in 9 successive weeks

Dataset: simplepig.dta

- `id`: pig identifier
- `weight`: weight of pig
- `week`: time of measurement

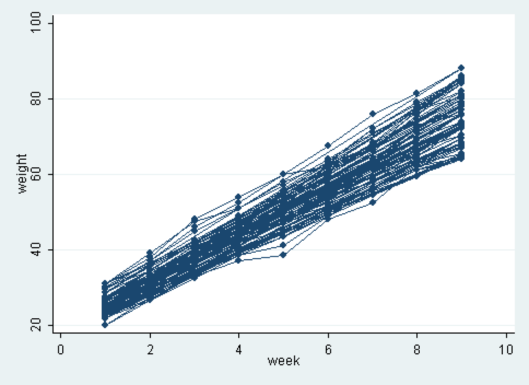


Figure: Weight trajectories for individual pigs

- Each trajectory is approximately linear in time
  $\hookrightarrow$ model an intercept β_0 and a slope β_1
- The pigs are a random sample from a larger population
  $\hookrightarrow$ pig-specific random effect α_{0i}
- $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \epsilon_{ij}, \quad i = 1, 2, \ldots, 48, \; j = 1, 2, \ldots, 9$
  - β_0 and β_1 fixed intercept and slope
  - $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a0})$ random pig effect
  - $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ random error, independent of α_{0i}'s
  - Y_{ij} weight of ith pig in week j and x_{ij} the corresponding time of measurement
- Anything else . . . ?

Random intercept

```
. xtmixed weight week || id:, ml

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log likelihood = -1014.9268
Iteration 1:   log likelihood = -1014.9268

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       432
Group variable: id                              Number of groups   =        48

                                                Obs per group: min =         9
                                                               avg =       9.0
                                                               max =         9


                                                Wald chi2(1)       =  25337.48
Log likelihood = -1014.9268                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        week |   6.209896   .0390124   159.18   0.000     6.133433    6.286359
       _cons |   19.35561   .5974056    32.40   0.000     18.18472    20.52651
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                   sd(_cons) |    3.84935   .4058114      3.130768    4.732863
-----------------------------+------------------------------------------------
                sd(Residual) |   2.093625   .0755472       1.95067    2.247056
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =   472.65 Prob >= chibar2 = 0.0000
```

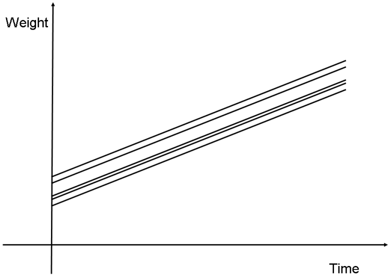


Figure: Random intercept model

Random pig effect = random intercept ⇒ can only fit parallel lines!

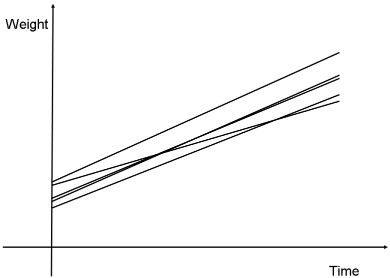


Figure: Random intercept and slope model

Random pig effect and random slope (in time) allow for individual curves

Random coefficients model

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \alpha_{1i} x_{ij} + \epsilon_{ij}, \quad i = 1, 2, \ldots, 48, \; j = 1, 2, \ldots, 9$$

- β_0 and β_1 fixed intercept and slope
- $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a0})$ random pig effect
- $\alpha_{1i} \sim \mathcal{N}(0, \sigma^2_{a1})$ random slope (for individual pigs) over time
- $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ random error, independent of α_{0i}'s and α_{1i}'s
- Y_{ij} weight of ith pig in week j and x_{ij} the corresponding time of measurement

Covariance structures

Stata provides four different covariance structures, illustrated for a model with two variance components:

- Unstructured:

$$\text{Cov}\begin{bmatrix}\alpha_{0i}\\ \alpha_{1i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a0} & \tau_{a0a1}\\ \tau_{a0a1} & \sigma^2_{a1}\end{bmatrix}$$

- Independent:

$$\text{Cov}\begin{bmatrix}\alpha_{0i}\\ \alpha_{1i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a0} & 0\\ 0 & \sigma^2_{a1}\end{bmatrix}$$

  i.e. the covariances are zero

Covariance structures (continued)

Stata provides four different covariance structures, illustrated for a model with two variance components:

- Identity:

$$\text{Cov}\begin{bmatrix}\alpha_{0i}\\ \alpha_{1i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a0} & 0\\ 0 & \sigma^2_{a0}\end{bmatrix}$$

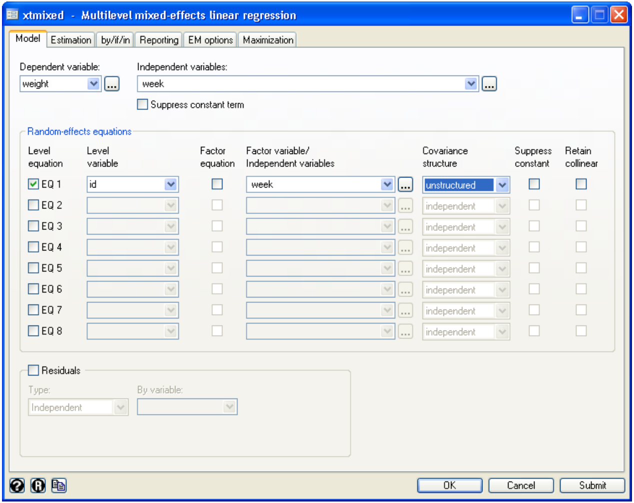
  i.e. the variances are equal and the covariances are zero

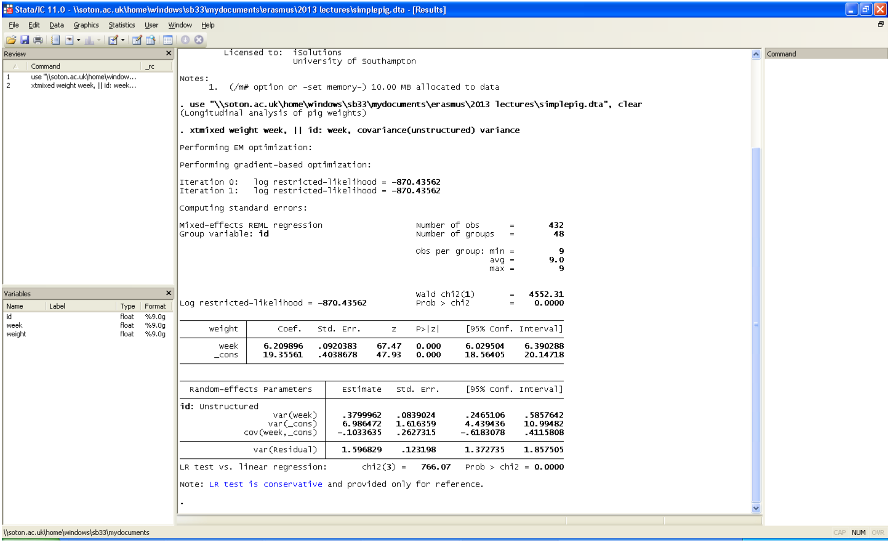
- Exchangeable:

$$\text{Cov}\begin{bmatrix}\alpha_{0i}\\ \alpha_{1i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a0} & \tau_{a0a1}\\ \tau_{a0a1} & \sigma^2_{a0}\end{bmatrix}$$

  i.e. the variances are equal and all covariances are equal, too

In Stata:


xtmixed - Multilevel mixed-effects linear regression
Model
Estimation
by/if/in
Reporting
EM options
Maximization
Dependent variable:
weight
Independent variables:
week
Suppress constant term
Random-effects equations
Level equation
Level variable
Factor equation
Factor variable/ Independent variables
Covariance structure
Suppress constant
Retain collinear
EQ 1
id
week
unstructured
EQ 2
EQ 3
EQ 4
EQ 5
EQ 6
EQ 7
EQ 8
independent
Residuals
Type:
Independent
By variable:
OK
Cancel
Submit


Stata/IC 11.0 - \\soton.ac.uk\home\windows\sb33\mydocuments\erasmus\2013 lectures\simplepig.dta - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command _rc
1 use "\\soton.ac.uk\home\window...
2 xtmixed weight week, || id: week...
Variables
Name Label Type Format
id float %9.0g
week float %9.0g
weight float %9.0g
Command
Licensed to: isolutions
University of Southampton
Notes:
1. (/m# option or -set memory-) 10.00 MB allocated to data
. use "\\soton.ac.uk\home\windows\sb33\mydocuments\erasmus\2013 lectures\simplepig.dta", clear
(Longitudinal analysis of pig weights)
. xtmixed weight week, || id: week, covariance(unstructured) variance
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0: log restricted-likelihood = -870.43562
Iteration 1: log restricted-likelihood = -870.43562
Computing standard errors:
Mixed-effects REML regression Number of obs = 432
Group variable: id Number of groups = 48
Obs per group: min = 9
avg = 9.0
max = 9
Wald chi2(1) = 4552.31
Log restricted-likelihood = -870.43562 Prob > chi2 = 0.0000
weight Coef. Std. Err. z P>|z| [95% Conf. Interval]
week 6.209896 .0920383 67.47 0.000 6.029504 6.390288
_cons 19.35561 .4038678 47.93 0.000 18.56405 20.14718
Random-effects Parameters Estimate Std. Err. [95% Conf. Interval]
id: Unstructured
var(week) .3799962 .0839024 .2465106 .5857642
var(_cons) 6.986472 1.616359 4.439436 10.99482
cov(week,_cons) -.1033635 .2627315 -.6183078 .4115808
var(Residual) 1.596829 .123198 1.372735 1.857505
LR test vs. linear regression: chi2(3) = 766.07 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
\\soton.ac.uk\home\windows\sb33\mydocuments
CAP NUM OVR

Model selection

We have fitted an `unstructured` covariance, i.e.

$$D = \text{Cov}\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} = \begin{bmatrix} \sigma^2_{a0} & \tau_{a0a1} \\ \tau_{a0a1} & \sigma^2_{a1} \end{bmatrix}.$$

The test: `LR test vs linear regression` tests
H_0: $\sigma^2_{a0} = \sigma^2_{a1} = \tau_{a0a1} = 0$ vs H_1: $\sigma^2_{a0} > 0$ or $\sigma^2_{a1} > 0$ or $\tau_{a0a1} \neq 0$

This test is conservative since H_0 is on the boundary
↪ the p-value is an upper bound for the true p-value
↪ random effects model highly significant over fixed effects model

Model selection

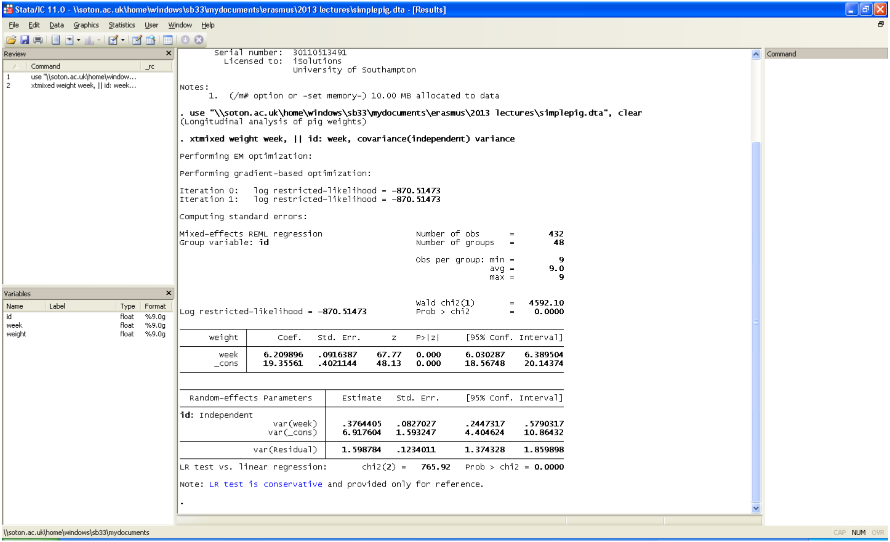
From confidence interval in `Stata` output: Covariance τ_{a0a1} does not appear to be significant (at the 5% level)

Test formally using LR test:

- Refit the model using covariance structure `independent`
  $\hookrightarrow$ log restricted-likelihood = -870.51473
- Test statistic: $2(-870.43562 - (-870.51473)) = 0.15822$
- Evaluate on χ^2 scale with 1df via
  `display chiprob(1,0.15822)`
- p-value: 0.691, so the covariance τ_{a0a1} is not needed in the model

Model selection

- For testing covariances of the random effects, the LR test is not conservative (H_0 not on the boundary)
- AIC and BIC can also be used for model selection
- LR tests with REML require identical fixed effects for both models
  ↪ No problem in this example
  ↪ Fit with ML when testing fixed effects, and then refit the final model with REML




```
     Serial number:  30110513491
       Licensed to:  iSolutions
                     University of Southampton

Notes:
      1.  (/m# option or -set memory-) 10.00 MB allocated to data

. use "\\soton.ac.uk\home\windows\sb33\mydocuments\erasmus\2013 lectures\simplepig.dta", clear
(Longitudinal analysis of pig weights)

. xtmixed weight week, || id: week, covariance(independent) variance

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -870.51473
Iteration 1:   log restricted-likelihood = -870.51473

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =       432
Group variable: id                              Number of groups   =        48

                                                Obs per group: min =         9
                                                               avg =       9.0
                                                               max =         9

                                                Wald chi2(1)       =   4592.10
Log restricted-likelihood = -870.51473          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        week |   6.209896   .0916387    67.77   0.000     6.030287    6.389504
       _cons |   19.35561   .4021144    48.13   0.000     18.56748    20.14374
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Independent              |
                   var(week) |   .3764405   .0827027      .2447317    .5790317
                  var(_cons) |   6.917604   1.593247      4.404624    10.86432
-----------------------------+------------------------------------------------
               var(Residual) |   1.598784   .1234011      1.374328    1.859898
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   765.92   Prob > chi2 = 0.0000

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

.
```

Final Model

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \alpha_{1i} x_{ij} + \epsilon_{ij} \tag{1}$$

where $\text{Cov}(\alpha_{0i}, \alpha_{1i}) = 0$

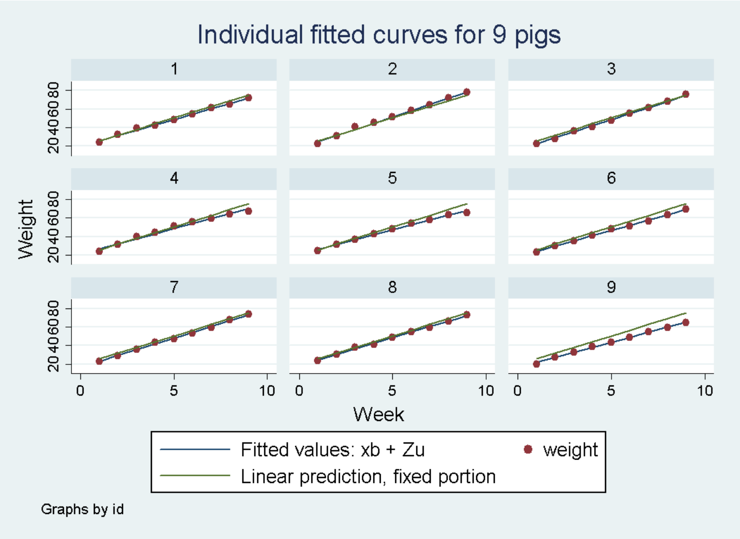
Information from Stata output

- Estimate of intercept β_0 is 19.36 with 95% CI: 18.57 - 20.14
- Estimate of slope β_1 is 6.21 with 95% CI 6.03 - 6.39
- LR test for H_0: $\sigma^2_{a0} = \sigma^2_{a1} = 0$ is rejected
  (Test statistic = 765.92, p-value (upper bound) = 0.0000)
- Estimates of variance components:
  $\sigma^2_{a0} = 6.92, \sigma^2_{a1} = 0.38, \sigma^2 = 1.599$

Predicting random effects

EBLUPs - Empirical Best Linear Unbiased Predictors

- EBLUPs are calculated as the conditional expectations of the random effects, given the data
- EBLUPs predict values for the random effects based on the observed data
- Model (in vector form) for the ith pig: $Y_i = X_i\beta + A_i\alpha_i + \epsilon_i$, where $\alpha_i \sim \mathcal{N}(0, D)$ is the vector of random effects for pig i
- EBLUPs: $\hat{\alpha}_i = E[\alpha_i | Y_i = y_i] = \hat{D}A_i^T\hat{V}_i^{-1}(y_i - X_i\hat{\beta})$ where $V_i = \text{Cov}(Y_i)$


Individual fitted curves for 9 pigs
1
2
3
4
5
6
7
8
9
Weight
20 40 60 80
Week
0 5 10
Fitted values: xb + Zu
weight
Linear prediction, fixed portion
Graphs by id

Conditional residuals

- A conditional residual is the difference between an observation and its EBLUP prediction, i.e. in vector form (for pig i):

$$\hat{\epsilon}_i = y_i - X_i \hat{\beta} - A_i \hat{\alpha}_i$$

  where $\hat{\alpha}_i$ is the vector of EBLUPs for pig i
- Useful for model diagnostics

Model diagnostics

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \alpha_{1i} x_{ij} + \epsilon_{ij}$$

Check the following assumptions

- $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, i.e. normality and constant variance
- $\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D \right)$

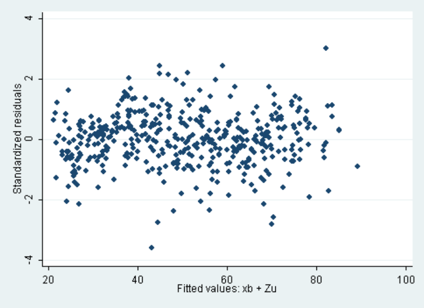
Also plot observations vs predictions to assess overall model fit

Assumption:

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, i.e. normality and constant variance

To check constant variance:

Conditional standardized residuals vs cond. fits

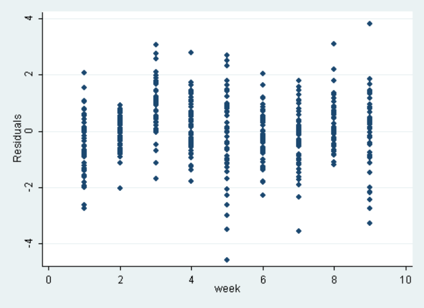


Assumption:

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, i.e. normality and constant variance

To check constant variance:

Conditional residuals vs week

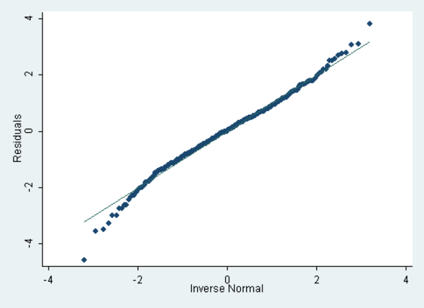


Assumption:

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, i.e. normality and constant variance

To check normality:

Normal plot of residuals



Conclusion for assumption:

$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, i.e. normality and constant variance

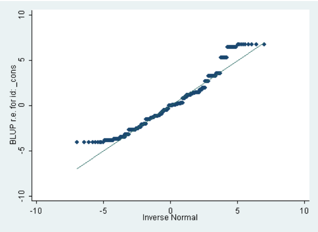
- Standardized residuals vs fits plot shows random scatter with constant amplitude, and some potential outliers
- Residuals vs week plot shows fairly similar spread for each week - somewhat higher in weeks 5 and 9, and somewhat lower in week 2
- Overall, assumption of constant variance seems acceptable
- Normal plot shows approximately a straight line except for some (potential) outliers at both ends - Normality seems acceptable

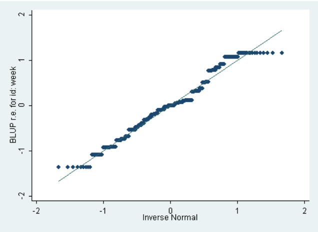
Assumption:

$$\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D \right)$$

To check marginal normality:

Normal plots of EBLUPs



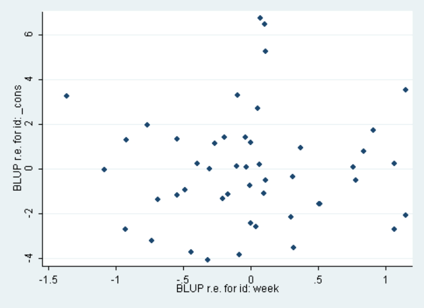


Assumption:

$$\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D \right)$$

To check joint normality:

Scatter plot of EBLUPs (intercept vs slope)

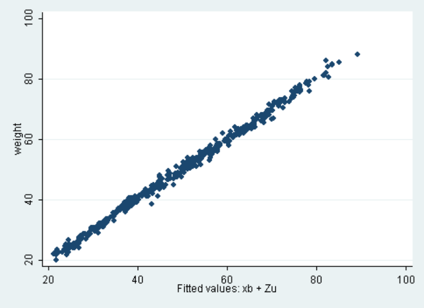


Conclusion for assumption:

$$\begin{bmatrix} \alpha_{0i} \\ \alpha_{1i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, D \right)$$

- Normal plot for random slopes shows approximately straight line, some concern about random intercepts - Marginal normality may be acceptable
- Scatter plot could come from a bivariate normal distribution
- Overall, assumption seems acceptable

Observations vs predictions



Approximately a straight line - the model seems to predict well

More about...pigs

An alternative model would be
$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_{0i} + \gamma_j + \epsilon_{ij}$

- β_0 and β_1 fixed intercept and slope
- $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a0})$ random pig effect
- $\gamma_j \sim \mathcal{N}(0, \sigma^2_{\gamma})$
- $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ random error, independent of α_{0i}'s and γ_j's
- Y_{ij} weight of ith pig in week j and x_{ij} the corresponding time of measurement

Which is the difference with (1)?

In Stata:

```
Iteration 0:   log likelihood = -1013.824
Iteration 1:   log likelihood = -1013.824

Computing standard errors:

Mixed-effects ML regression                     Number of obs      =       432
Group variable: _all                            Number of groups   =         1

                                                Obs per group: min =       432
                                                               avg =     432.0
                                                               max =       432

                                                Wald chi2(1)       =  13258.25
Log likelihood = -1013.824                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        week |   6.209896   .0539313   115.14   0.000     6.104192    6.315599
       _cons |   19.35561   .6333981    30.56   0.000     18.11418    20.59705
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
_all: Identity               |
                 sd(R.week)  |   .2915266   .1490187      .1070464    .7939334
-----------------------------+------------------------------------------------
_all: Identity               |
                   sd(R.id)  |   3.851781    .405804      3.133165    4.735216
-----------------------------+------------------------------------------------
               sd(Residual)  |      2.073   .0756007      1.929997    2.226599
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =   474.85   Prob > chi2 = 0.0000

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

Example 2: Weight gain of children

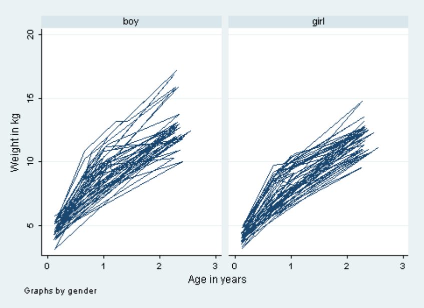
Repeated measures are collected on **units of analysis** over time, and further covariates are recorded for each unit of analysis

Weight gain of Asian children in a British community - between 1 and 5 weight measurements for each child (Rabe-Hesketh and Skrondal, 2005)

Dataset: `asian1.dta`

- `id`: child identifier
- `weight`: weight in kg
- `age`: age in years
- `gender`: boy/girl
- `girl`: dummy variable - 0 for boys, 1 for girls

Individual growth trajectories



- The trajectories are not straight lines
  ↪ include a quadratic term in the model
- Some children are consistently heavier than others
  ↪ include random coefficients

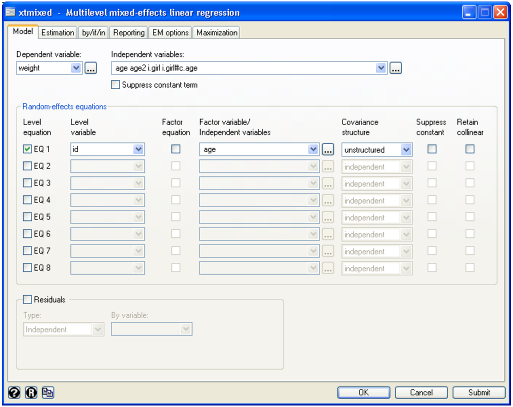
Starting model

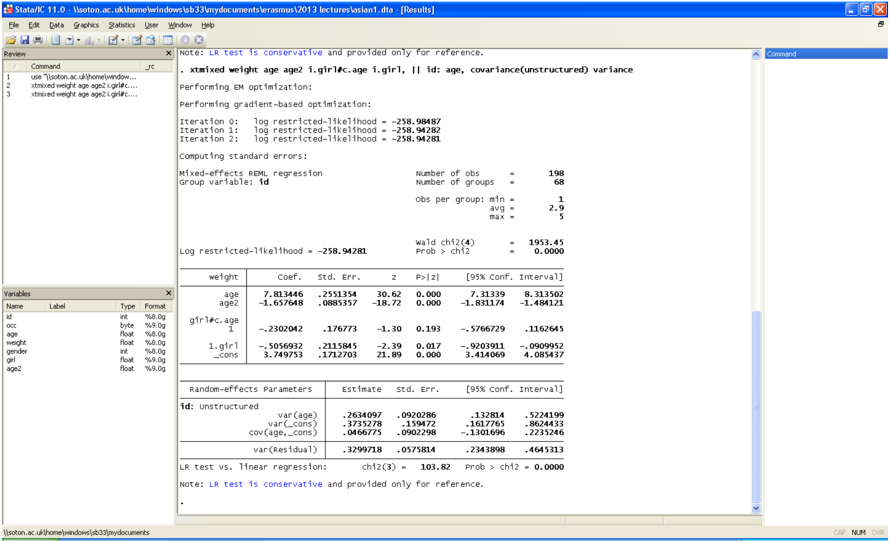
- Include systematic differences in weight for boys and girls through dummy variable girl_i (0 for a boy, 1 for a girl)
- Include interaction of girl_i with age

- $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 \text{girl}_i + \beta_4 \text{girl}_i x_{ij} + \alpha_{0i} + \alpha_{1i} x_{ij} + \epsilon_{ij}$

- β_3 is the difference between intercepts for girls and boys
- β_4 is the difference between slopes for girls and boys

In Stata


xtmixed - Multilevel mixed-effects linear regression
Model
Estimation
by/if/in
Reporting
EM options
Maximization
Dependent variable:
weight
Independent variables:
age age2 i.girl i.girl#c.age
Suppress constant term
Random-effects equations
Level equation
Level variable
Factor equation
Factor variable/ Independent variables
Covariance structure
Suppress constant
Retain collinear
EQ 1
id
age
unstructured
EQ 2
EQ 3
EQ 4
EQ 5
EQ 6
EQ 7
EQ 8
independent
Residuals
Type:
Independent
By variable:
OK
Cancel
Submit



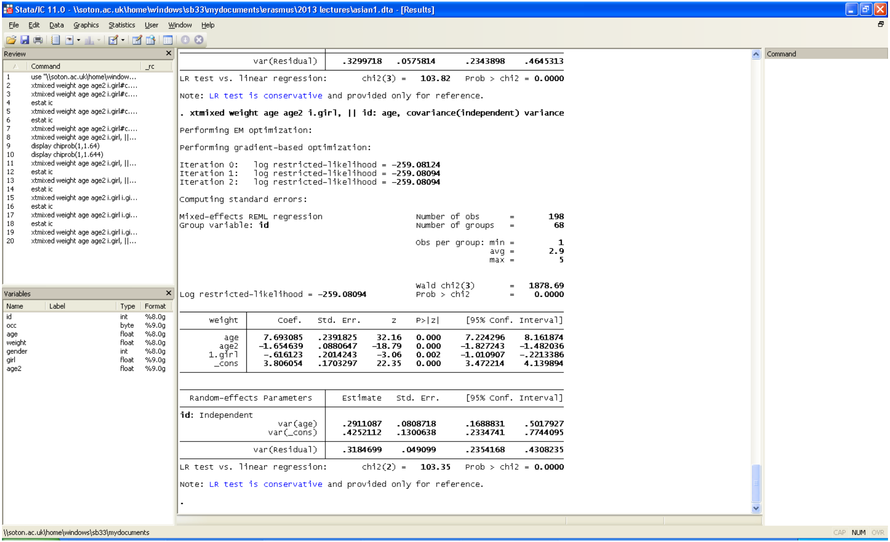
```
Note: LR test is conservative and provided only for reference.

. xtmixed weight age age2 i.girl#c.age i.girl, || id: age, covariance(unstructured) variance

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:   log restricted-likelihood = -258.98487
Iteration 1:   log restricted-likelihood = -258.94282
Iteration 2:   log restricted-likelihood = -258.94281

Computing standard errors:

Mixed-effects REML regression                   Number of obs      =       198
Group variable: id                              Number of groups   =        68

                                                Obs per group: min =         1
                                                               avg =       2.9
                                                               max =         5


                                                Wald chi2(4)       =   1953.45
Log restricted-likelihood = -258.94281          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   7.813446   .2551354    30.62   0.000      7.31339    8.313502
        age2 |  -1.657648   .0885357   -18.72   0.000    -1.831174   -1.484121
             |
  girl#c.age |
           1 |  -.2302042    .176773    -1.30   0.193    -.5766729    .1162645
             |
      1.girl |  -.5056932   .2115845    -2.39   0.017    -.9203911   -.0909952
       _cons |   3.749753   .1712703    21.89   0.000     3.414069    4.085437
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                    var(age) |   .2634097   .0920286       .132814    .5224199
                  var(_cons) |   .3735278    .159472      .1617765    .8624433
              cov(age,_cons) |   .0466775   .0902298     -.1301696    .2235246
-----------------------------+------------------------------------------------
               var(Residual) |   .3299718   .0575814      .2343898    .4645313
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   103.82   Prob > chi2 = 0.0000

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

.
```

Model selection strategy

- $\text{Cov}(\alpha_{0i}, \alpha_{1i})$ seems not significant
  $\hookrightarrow$ Refit with covariance structure `independent` and carry out LR test
  $\hookrightarrow$ Remove covariance term from the model
- In the new model `girl` $\times$ `age` interaction seems not significant
  $\hookrightarrow$
  - Refit using ML instead of REML
  - Fit model without interaction in ML and carry out LR test
  - Refit preferred model with REML

Model selection strategy

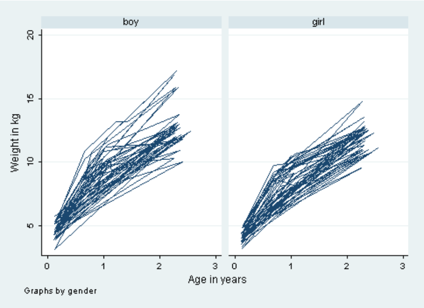
Model	LR (df)	AIC REML/ML	BIC REML/ML
"Full"	0.26 (1)	535.89/523.99	565.48/553.58
$\text{Cov}(\alpha_{0i}, \alpha_{1i}) = 0$	1.64 (1) (from ML)	534.13/522.36	560.44/548.67
$\text{Cov}(\alpha_{0i}, \alpha_{1i}) = 0$, no interaction		532.16/522.01	555.18/545.03


Stata/IC 11.0 - \\soton.ac.uk\home\windows\sb33\mydocuments\erasmus\2013 lectures\asian1.dta - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command _rc
1 use "\\soton.ac.uk\home\window...
2 xtmixed weight age age2 i.girl#c....
3 xtmixed weight age age2 i.girl#c....
4 estat ic
5 xtmixed weight age age2 i.girl#c....
6 estat ic
7 xtmixed weight age age2 i.girl#c....
8 xtmixed weight age age2 i.girl, ||...
9 display chiprob(1,1.64)
10 display chiprob(1,1.644)
11 xtmixed weight age age2 i.girl, ||...
12 estat ic
13 xtmixed weight age age2 i.girl, ||...
14 estat ic
15 xtmixed weight age age2 i.girl i.gi...
16 estat ic
17 xtmixed weight age age2 i.girl i.gi...
18 estat ic
19 xtmixed weight age age2 i.girl i.gi...
20 xtmixed weight age age2 i.girl, ||...
Variables
Name Label Type Format
id int %8.0g
occ byte %9.0g
age float %8.0g
weight float %8.0g
gender int %8.0g
girl float %9.0g
age2 float %9.0g
var(Residual) | .3299718 .0575814 .2343898 .4645313
LR test vs. linear regression: chi2(3) = 103.82 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
. xtmixed weight age age2 i.girl, || id: age, covariance(independent) variance
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0: log restricted-likelihood = -259.08124
Iteration 1: log restricted-likelihood = -259.08094
Iteration 2: log restricted-likelihood = -259.08094
Computing standard errors:
Mixed-effects REML regression Number of obs = 198
Group variable: id Number of groups = 68
Obs per group: min = 1
avg = 2.9
max = 5
Wald chi2(3) = 1878.69
Log restricted-likelihood = -259.08094 Prob > chi2 = 0.0000
weight | Coef. Std. Err. z P>|z| [95% Conf. Interval]
age | 7.693085 .2391825 32.16 0.000 7.224296 8.161874
age2 | -1.654639 .0880647 -18.79 0.000 -1.827243 -1.482036
1.girl | -.616123 .2014243 -3.06 0.002 -1.010907 -.2213386
_cons | 3.806054 .1703297 22.35 0.000 3.472214 4.139894
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]
id: Independent
var(age) | .2911087 .0808718 .1688831 .5017927
var(_cons) | .4252112 .1300638 .2334741 .7744095
var(Residual) | .3184699 .049099 .2354168 .4308235
LR test vs. linear regression: chi2(2) = 103.35 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
Command
\\soton.ac.uk\home\windows\sb33\mydocuments
CAP NUM OVR

Conclusions so far

- At a given age, girls are on average 0.616 kg lighter than boys
- Weight growth is quadratic in age
- Random effects $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a0})$ and $\alpha_{1i} \sim \mathcal{N}(0, \sigma^2_{a1})$ needed in the model
- However ...

Individual growth trajectories



- The trajectories for boys appear to be more variable around the mean model!
  ↪ The variances of the random effects may be different for boys and girls

A heteroscedastic model

Consider random effects $\alpha_{0i} \sim \mathcal{N}(0, \sigma^2_{a0})$, $\alpha_{1i} \sim \mathcal{N}(0, \sigma^2_{a1})$ if the ith child is a boy, and $\alpha_{2i} \sim \mathcal{N}(0, \sigma^2_{a2})$, $\alpha_{3i} \sim \mathcal{N}(0, \sigma^2_{a3})$ if the ith child is a girl

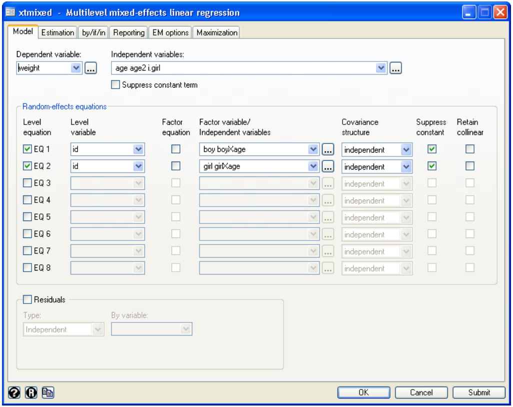
$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + \beta_3 \text{girl}_i + \alpha_{0i} \text{boy}_i + \alpha_{2i} \text{girl}_i + \alpha_{1i} \text{boy}_i x_{ij} + \alpha_{3i} \text{girl}_i x_{ij} + \epsilon_{ij}$$

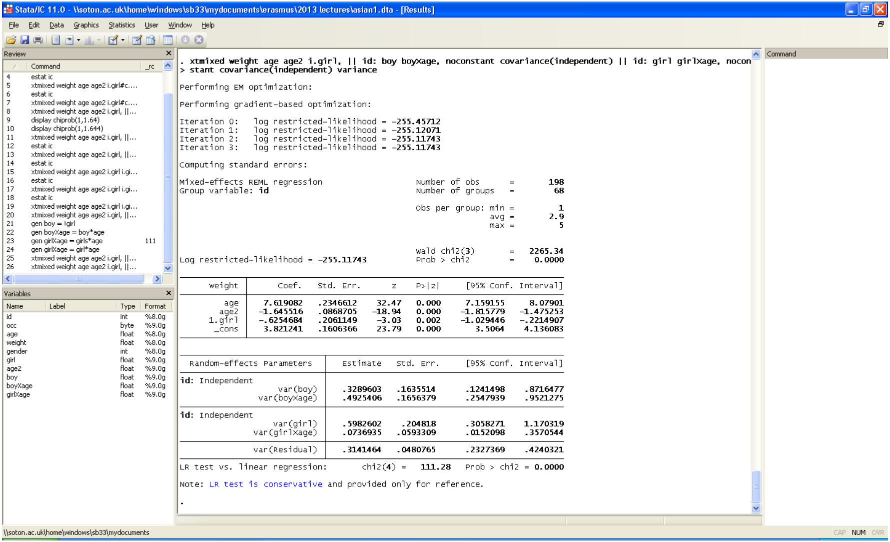
In Stata

Create the necessary dummy variables via

```
gen boy = !girl
gen boyXage = boy*age
gen girlXage = girl*age
```

In Stata


xtmixed - Multilevel mixed-effects linear regression
Model
Estimation
by/if/in
Reporting
EM options
Maximization
Dependent variable:
weight
Independent variables:
age age2 i.girl
Suppress constant term
Random-effects equations
Level equation
Level variable
Factor equation
Factor variable/ Independent variables
Covariance structure
Suppress constant
Retain collinear
EQ 1
id
boy boyXage
independent
EQ 2
id
girl girlXage
independent
EQ 3
EQ 4
EQ 5
EQ 6
EQ 7
EQ 8
Residuals
Type:
Independent
By variable:
OK
Cancel
Submit


Stata/IC 11.0 - \\soton.ac.uk\home\windows\sb33\mydocuments\erasmus\2013 lectures\asian1.dta - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command _rc
4 estat ic
5 xtmixed weight age age2 i.girl#c....
6 estat ic
7 xtmixed weight age age2 i.girl#c....
8 xtmixed weight age age2 i.girl, ||...
9 display chiprob(1,1.64)
10 display chiprob(1,1.644)
11 xtmixed weight age age2 i.girl, ||...
12 estat ic
13 xtmixed weight age age2 i.girl, ||...
14 estat ic
15 xtmixed weight age age2 i.girl i.gi...
16 estat ic
17 xtmixed weight age age2 i.girl i.gi...
18 estat ic
19 xtmixed weight age age2 i.girl i.gi...
20 xtmixed weight age age2 i.girl, ||...
21 gen boy = !girl
22 gen boyXage = boy*age
23 gen girlXage = girls*age 111
24 gen girlXage = girl*age
25 xtmixed weight age age2 i.girl, ||...
26 xtmixed weight age age2 i.girl, ||...
Variables
Name Label Type Format
id int %8.0g
occ byte %9.0g
age float %8.0g
weight float %8.0g
gender int %8.0g
girl float %9.0g
age2 float %9.0g
boy float %9.0g
boyXage float %9.0g
girlXage float %9.0g
. xtmixed weight age age2 i.girl, || id: boy boyXage, noconstant covariance(independent) || id: girl girlXage, nocon
> stant covariance(independent) variance
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0: log restricted-likelihood = -255.45712
Iteration 1: log restricted-likelihood = -255.12071
Iteration 2: log restricted-likelihood = -255.11743
Iteration 3: log restricted-likelihood = -255.11743
Computing standard errors:
Mixed-effects REML regression Number of obs = 198
Group variable: id Number of groups = 68
Obs per group: min = 1
avg = 2.9
max = 5
Wald chi2(3) = 2265.34
Log restricted-likelihood = -255.11743 Prob > chi2 = 0.0000
weight | Coef. Std. Err. z P>|z| [95% Conf. Interval]
age | 7.619082 .2346612 32.47 0.000 7.159155 8.07901
age2 | -1.645516 .0868705 -18.94 0.000 -1.815779 -1.475253
1.girl | -.6254684 .2061149 -3.03 0.002 -1.029446 -.2214907
_cons | 3.821241 .1606366 23.79 0.000 3.5064 4.136083
Random-effects Parameters | Estimate Std. Err. [95% Conf. Interval]
id: Independent |
var(boy) | .3289603 .1635514 .1241498 .8716477
var(boyXage) | .4925406 .1656379 .2547939 .9521275
id: Independent |
var(girl) | .5982602 .204818 .3058271 1.170319
var(girlXage) | .0736935 .0593309 .0152098 .3570544
var(Residual) | .3141464 .0480765 .2327369 .4240321
LR test vs. linear regression: chi2(4) = 111.28 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
Command
\\soton.ac.uk\home\windows\sb33\mydocuments
CAP NUM OVR

LR test for equal variances

H_0: $\sigma^2_{a0} = \sigma^2_{a2}$ and $\sigma^2_{a1} = \sigma^2_{a3}$ vs not equal

Test statistic: $2(-255.11743 - (-259.08094)) = 7.92702$ on 2 df

Produces a p-value of 0.0190

AIC and BIC

Model	AIC	BIC
Equal variances	532.16	555.18
Unequal variances	528.23	557.83

The heteroscedastic model is preferred by LR test and AIC, but slightly worse on BIC

References

- Rabe-Hesketh, S. and Skrondal, A. (2005). Multilevel and longitudinal modeling using Stata. Stata Press.
- Stata Help Manual
- West, B.T., Welch, K.B. and Galecki, A.T. (2007). Linear Mixed Models - A practical guide using statistical software. Chapman & Hall/CRC