1 Introduction

In this practical, we will analyse longitudinal data on health functioning from a study of civil servants, the Whitehall II study. Health functioning was assessed by the SF-36, a 36 item instrument that comprises eight subscales covering physical, psychological and social functioning. These eight scales can be summarised into physical and mental health components. These are scaled using general US population norms to have mean values of 50 and low scores imply poor functioning. Physical health functioning (PCS) and mental health functioning (MCS) were measured on up to six occasions at approximately 2.5 year intervals. In this practical, PCS is the response, and there are two levels of data - person (level 2) and measurement occasion (level 1). In addition, there are three explanatory variables. The first is the person’s age, which varies from occasion to occasion and is therefore a level one variable. The other two are gender (coded 0 for males and 1 for females) and employment grade at baseline (coded 1 for high grade, 2 for intermediate grade and 3 for low grade). These vary from person to person and are thus level two variables.

In this session we will explore the following questions:

1. How does physical functioning change as people get older?
2. Does this vary from person to person?
3. Does physical functioning decline faster in people from low employment grades compared with those in high employment grades?

2 Setting up the data structure

Open the worksheet function.wsz which contains 21 variables for 8815 people as shown below in the Names display:
Column number 1 contains the person identifier. Columns 2 and 3 contain person level explanatory variables: gender (FEMALE) and employment grade (GRADE). This is followed by sets of three variables for the six measurement occasions: age, physical functioning and mental functioning at each occasion (XAGE50, XPCS, XMCS, at measurement occasion 1 etc). The variable names for measurement occasions 1 to 6 are prefixed by X, V, T, Q, M and K respectively.

*Note that the ages have been centred at age 50. In this data set, –20 represents a missing value. We can tell MLwiN that -20 is the missing value code by*

- Select the Options menu
- Select Numbers(Display precision and missing value code)
- Set the missing value code to -20
- Click the Apply button, then Done

The Names window is updated and now explicitly shows the number of missing cases for each variable.
This arrangement of the data, in which each row of a rectangular array corresponds to a different individual and contains all the data for that individual, is a natural one, but it does not reflect the hierarchical structure of measurements nested within individuals. For a multilevel analysis, the data must first be restructured so that there is one record per measurement occasion (level 1 unit). The **Split records** window (shown below), accessed via the **Data Manipulation** menu, is designed to transform an individual’s data record into separate records (or rows), one for each occasion. In the present case we shall produce six records per person, that is, 52890 records altogether. The ordering of people will be preserved, and they will become the level 2 units.

<table>
<thead>
<tr>
<th>Name</th>
<th>Cn</th>
<th>n</th>
<th>missing</th>
<th>min</th>
<th>max</th>
<th>categorical</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>8815</td>
<td>0</td>
<td>1</td>
<td>8815</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>2</td>
<td>8815</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>GRADE</td>
<td>3</td>
<td>8815</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>XAGE50</td>
<td>4</td>
<td>8815</td>
<td>0</td>
<td>14</td>
<td>14</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>XPSCS</td>
<td>5</td>
<td>8815</td>
<td>523</td>
<td>6.487695</td>
<td>71.57574</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>XMCS</td>
<td>6</td>
<td>8815</td>
<td>523</td>
<td>2.97561</td>
<td>69.64911</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>VAGE50</td>
<td>7</td>
<td>8815</td>
<td>503</td>
<td>8</td>
<td>15</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>VPCS</td>
<td>8</td>
<td>8815</td>
<td>1333</td>
<td>11.80435</td>
<td>70.37716</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>VMCS</td>
<td>9</td>
<td>8815</td>
<td>1333</td>
<td>4.231273</td>
<td>72.92052</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>TAGE50</td>
<td>10</td>
<td>8815</td>
<td>1186</td>
<td>5.296924</td>
<td>19.23751</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>TPCS</td>
<td>11</td>
<td>8815</td>
<td>2010</td>
<td>8.490927</td>
<td>69.6278</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>IMCS</td>
<td>12</td>
<td>8815</td>
<td>2010</td>
<td>2.450083</td>
<td>72.98643</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>QAGE50</td>
<td>13</td>
<td>8815</td>
<td>1658</td>
<td>-1.828125</td>
<td>21.40625</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>QPCS</td>
<td>14</td>
<td>8815</td>
<td>2350</td>
<td>9.355224</td>
<td>69.81925</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>QMCS</td>
<td>15</td>
<td>8815</td>
<td>2350</td>
<td>2.383819</td>
<td>73.94762</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>IMAGE50</td>
<td>16</td>
<td>8815</td>
<td>2017</td>
<td>0.4660045</td>
<td>24.08076</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>MPCS</td>
<td>17</td>
<td>8815</td>
<td>2346</td>
<td>9.969062</td>
<td>71.85647</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>MMCS</td>
<td>18</td>
<td>8815</td>
<td>2346</td>
<td>2.536131</td>
<td>75.21274</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>KAGE50</td>
<td>19</td>
<td>8815</td>
<td>1893</td>
<td>0.3040685</td>
<td>26.73438</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>KPCS</td>
<td>20</td>
<td>8815</td>
<td>2379</td>
<td>7.425531</td>
<td>70.09693</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>KMCS</td>
<td>21</td>
<td>8815</td>
<td>2379</td>
<td>7.746543</td>
<td>73.8034</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>c22</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>c23</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>c24</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
</tbody>
</table>
There are two types of data to consider: occasion specific data and individual (time-invariant) specific data. The former (in principle) changes from occasion to occasion, in this case, the functioning scores and the ages. The latter remain constant from occasion to occasion, in this case, the person identifiers, gender and employment grade.

First let us deal with the occasion specific data:

- Open the Split records window
- Set the Number of occasions to 6
- Set the Number of variables to 3

Doing this produces:
We need to stack the six physical functioning scores into a single column, the six mental health functioning scores into a single column and the six ages into a single column.

- In the **Stack data** grid, click on **Variable 1**
- From the drop down list that appears, select the six variables XPCS, VPCS, TPCS, QPCS, MPCS, KPCS and then click **Done**. (To make multiple selections, hold the control key down while clicking on variable names.)
- Repeat the above two steps for **Variable 2** and the six variables XMCS, VMCS, TMCS, QMCS, MMCS, KMCS
- Repeat the above two steps for **Variable 3** and the six variables XAGE50, VAGE50, TAGE50, QAGE50, MAGE50, KAGE50

Before proceeding check carefully that the **Split records** window looks like this:
Clicking on the column headings allows you to set all six occasion variables from a single pick list. The first variable on the list is assigned to occasion 1, the second to occasion 2 and so on. This works fine in our case because the variables appear on the list in the correct order. If this is not the case, you can specifically assign variables to occasions by clicking on individual cells in the grid.

- Click in turn on the three empty cells in the Stacked into row of the Stack data grid. (You may need to enlarge the window to see the whole grid.)
- From the drop-down lists that appear, select c22, c23 and c24 respectively
- Tick the Generate indicator column check box
- In the neighbouring drop down list, select c25

That deals with occasion specific data. Now we will specify the repeated data:

- In the Repeat(carried data) frame, select the three variables ID, FEMALE, GRADE as the input columns and c26,c27 and c28 as the output columns

The completed set of entries should look like this:
This will take the six physical functioning scores, each of length 8815, and stack them into a single variable in \( c_{22} \). The six mental health functioning scores will be stacked into \( c_{23} \) and the six age variables will be stacked into \( c_{24} \). Each id code will be repeated six times, and the repeated codes are stored in \( c_{26} \). Similarly, values of \text{FEMALE} \text{ and } \text{GRADE} \text{ will be repeated six times and stored in } \( c_{27} \) \text{ and } \( c_{28} \). The indicator column, which is output to \( c_{25} \), will contain occasion identifiers for the new long data set.

- Click the Split button to execute the changes
- You will be asked if you want to save the worksheet - select No

The Names window now shows the following for \( c_{22} \) through \( c_{28} \):

<table>
<thead>
<tr>
<th>Name</th>
<th>Cn</th>
<th>n</th>
<th>missing</th>
<th>min</th>
<th>max</th>
<th>categorical</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C22</td>
<td>22</td>
<td>52890</td>
<td>10932</td>
<td>6.407695</td>
<td>71.85617</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C23</td>
<td>23</td>
<td>52890</td>
<td>10932</td>
<td>2.383819</td>
<td>75.21274</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C24</td>
<td>24</td>
<td>52890</td>
<td>7257</td>
<td>-11</td>
<td>26.73438</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C25</td>
<td>25</td>
<td>52890</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>C26</td>
<td>26</td>
<td>52890</td>
<td>0</td>
<td>1</td>
<td>8815</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C27</td>
<td>27</td>
<td>52890</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C28</td>
<td>28</td>
<td>52890</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C29</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C30</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
<tr>
<td>C31</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>False</td>
<td></td>
</tr>
</tbody>
</table>
In the Names window, use Edit name to assign the names pcs, mcs, age, occasion, person, fem and occupation to c22-c28. Viewing columns 22-28 (by selecting the View or edit data from the Data Manipulation menu) will now show:

The data are now in the required form for analysis, with one row per measurement occasion. It would now be a good idea to save the worksheet, using a different name, e.g. function_long.wsz.

3 Initial data exploration

Before we start to do any modelling, we should first carry out some exploratory analysis. We will begin by looking at the mean of our outcome variable, functioning score, at each occasion.

- From the Basic Statistics menu, select Tabulate
  - Select Means as the Output Mode
  - A drop-down list labelled variate column appears. Select pcs
  - From the Columns drop-down list, select occasion
  - Click Tabulate
This produces the output:

<table>
<thead>
<tr>
<th>Variable tabulated is pcs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>MEANS</td>
</tr>
<tr>
<td>SD'S</td>
</tr>
</tbody>
</table>

Now use the Tabulate window to tabulate mean age by occasion.

<table>
<thead>
<tr>
<th>Variable tabulated is age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>MEANS</td>
</tr>
<tr>
<td>SD'S</td>
</tr>
</tbody>
</table>

The age variable has been transformed by measuring it as a deviation from age 50.

We are now almost in a position to set up a simple model, but first we must define a constant column; this is just a column of 52890 values of 1 (one for each measurement occasion).

- From the Data Manipulation menu, select Generate Vector
- Fill out the options as shown below and click Generate
- Use the Names window to assign the name cons to c29

4 A simple variance components model

We will start by examining how the total variance is partitioned into two components: between person (level 2) and between occasions within person (level
1). This variance components model is not interesting in itself but it provides a baseline with which to compare more complex models.

Set up a two-level model with pcs (physical functioning) as the outcome variable and cons as the only explanatory variable. The Equations window should appear as follows:

\[
\begin{align*}
\text{pcs}_{ij} &\sim N(\chi \mathbf{B}, \Omega) \\
\text{pcs}_{ij} &= \beta_{00} \text{cons} \\
\beta_{0ij} &= \beta_0 + u_{0j} + \varepsilon_{0ij} \\
[u_{0j}] &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} \sigma^2 \end{bmatrix} \\
[\varepsilon_{0ij}] &\sim N(0, \Omega_\varepsilon) : \Omega_\varepsilon = \begin{bmatrix} \sigma^2 \end{bmatrix}
\end{align*}
\]

Note pcs\(_{ij}\) is the physical functioning score at \(i^{\text{th}}\) measurement occasion for the \(j^{\text{th}}\) person. At convergence the estimates are:

\[
\begin{align*}
\text{pcs}_{ij} &\sim N(\chi \mathbf{B}, \Omega) \\
\text{pcs}_{ij} &= \beta_{00} \text{cons} \\
\beta_{0ij} &= 50.133(0.075) + u_{0j} + \varepsilon_{0ij} \\
[u_{0j}] &\sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 40.631(0.738) \end{bmatrix} \\
[\varepsilon_{0ij}] &\sim N(0, \Omega_\varepsilon) : \Omega_\varepsilon = \begin{bmatrix} 33.462(0.259) \end{bmatrix}
\end{align*}
\]

-2*loglikelihood(IQLS Deviance) = 282585.513(41958 of 52890 cases in use)

There is variation in physical functioning between individuals (\(\hat{\sigma}^2_{u0} = 40.6\)) and also variation between occasions within person (\(\hat{\sigma}^2_{e0} = 33.5\)). The likelihood statistic (-2 loglikelihood), found at the bottom of the Equations window, can be used as the basis for judging more elaborate models. The baseline value is 282585.
5 A linear growth curve model

A first step in modelling the between-occasion within-person, or level 1, variation is to fit a fixed linear trend. We therefore add age to our list of fixed explanatory variables in the Equations window (using Add Term). After adding age, click on More and at convergence obtain the following:

\[
\begin{align*}
\text{pcs}_{ij} &\sim N(\beta_0, \Omega) \\
\text{pcs}_{ij} &= \beta_{0j}\text{cons} + 0.244(0.005)\text{age}_{ij} \\
\beta_{0j} &= 51.671(0.080) + \nu_{0j} + \varepsilon_{0j} \\
\begin{bmatrix} \nu_{0j} \\ \varepsilon_{0j} \end{bmatrix} &\sim N(0, \Omega_\nu) : \Omega_\nu = \begin{bmatrix} 39.203(0.711) \\ \text{31.597}(0.245) \end{bmatrix} \\
-2\times\log\text{likelihood}(\text{ICLS Deviance}) &= 280335.095(\text{41958 of 52890 cases in use})
\end{align*}
\]

The estimate of the fixed parameter for age is -0.244 indicating that physical functioning declines with increasing age. Estimates of the random parameters are somewhat reduced, more so the level 1 variance which is expected because age is time-varying (i.e. a level 1 variable). There is a reduction in the likelihood statistic, which is now 280335.

We would expect the linear growth rate to vary from person to person around its mean value, rather than be fixed, and so we make the coefficient of age random at level 2 and continue iterations until convergence to give:
Note that the coefficient for age now has a subscript j, indicating that it varies at level 2 (i.e. between individuals).

The deviance, that is the reduction in the likelihood statistic, is 280335 - 279216 = 1119; this is large and is clearly statistically highly significant (comparing to a chi-squared distribution on 2 degrees of freedom). Hence there is considerable variation between people in their linear growth rates. We can get some idea of the size of this variation by taking the square root of the slope variance \( \sigma_{\beta_j}^2 \) to give the estimated standard deviation \( \sigma_{\beta_j} \). Assuming Normality, about 95% of people will have growth rates within two standard deviations of the mean growth rate (= -0.246), giving a 95% coverage interval of -0.85 to 0.35 for the ‘growth rate’. This suggests that physical health functioning improves with age for some people.

We can also look at various plots of the level 2 residuals. To obtain a plot of the standardised level 2 residuals, slope (\( \hat{u}_{ij} \)) versus intercept (\( \hat{u}_{0j} \)):

- From the Model menu, select Residuals
- Next to level at the bottom of the Residuals window, select 2:person
- Click Calc
- Click on the Plots tab and, under pairwise, check standardised residuals
- Click Apply
We see from the above plot that the two level 2 residuals are positively correlated. From the Estimates window we see that the model estimate is 0.21 (from the Model menu, select Estimate tables, change from FIXED PART to level 2: person and check C). A positive correlation implies that the greater the expected score at age 50, the faster the growth. However, this statistic needs to be interpreted with great caution: it can vary according to the scale adopted, and is relevant only for linear growth models.

Allowing the growth rate to vary across individuals, by fitting a random coefficient at level 2 to age, implies that the between-individual (level 2) variance depends on age. To calculate level 2 variance function:

- From the Model menu, select Variance function
- Next to level at the bottom of the Variance function window, select 2:person
- Click on Name to see the form of the level 2 variance function (a quadratic function in age)
- Next to variance output to, select c35
- Click Calc
- In the Names window, assign the name l2var to c35

To plot the level 2 variance against age:

- From the Graphs menu, select Customised Graph(s)
- At the top left of the window, change from dataset D10 to an empty dataset D1 which has no graph settings
- From the drop-down list next to y, select l2var
- From the drop-down list next to x select age
- From the drop-down list next to plot type, select line
The variance plot is shown below, after adding axis labels. We can see that the between-individual variance in physical functioning increases with age.

6 Complex level 1 variation

Before going on to further elaborate the level 2 variation we can allow for complex, that is non-constant, variation at level 1. So far we have allowed the between-individual (level 2) variance to depend on age, which was achieved by fitting a random coefficient for age at level 2. Suppose we believe that the within-individual (level 1) variance might also depend on age. For example, we might expect greater variance in physical functioning over time for older people than for younger people. We allow the level 1 variance to depend on age by declaring the coefficient of age to be random at level 1.

- In the Equations window, click on age and check i(occasion)
- Click Done
- You should find that an i subscript has been added to the coefficient for age,
and two extra terms have been added to the level 1 covariance matrix
- Click More to fit the new model

The model estimates are:

$$\begin{align*}
\text{pcs}_{ij} & \sim N(\bar{X}B, \Omega) \\
\text{pcs}_{ij} &= \beta_{0ij}\text{cons} + \beta_{1ij}\text{age}_{ij} \\
\beta_{0ij} &= 51.720(0.073) + \mu_{0ij} + \epsilon_{0ij} \\
\beta_{1ij} &= -0.243(0.006) + \mu_{1ij} + \epsilon_{1ij} \\
\begin{bmatrix}
\mu_{0ij} \\
\mu_{1ij}
\end{bmatrix} & \sim N(0, \Omega_\mu) : \Omega_\mu = \\
&= \begin{bmatrix}
31.006(0.676) \\
0.351(0.040) & 0.087(0.005)
\end{bmatrix} \\
\begin{bmatrix}
\epsilon_{0ij} \\
\epsilon_{1ij}
\end{bmatrix} & \sim N(0, \Omega_\varepsilon) : \Omega_\varepsilon = \\
&= \begin{bmatrix}
26.067(0.304) \\
0.136(0.028) & 0.007(0.004)
\end{bmatrix}
\end{align*}$$

$$-2\text{loglikelihood}(ICLS Deviance) = 279069.930(41958 \text{ of } 52890 \text{ cases in use})$$

Note that, because age is a level 1 variable, it does not make sense to say that the effect of age varies between measurement occasions. Rather, the parameters in the level 1 variance matrix should be thought of as coefficients of the level 1 variance function, a quadratic function in age. To see the equation of the variance function, and to obtain an estimate of it:

- From the Model menu, select Variance function
- Next to level at the bottom of the Variance function window, select 1:occasion
- Click on Name to see the form of the level 1 variance function (a quadratic function in age - see below)
- Next to variance output to, select c36
- Click Calc
- In the Names window, assign the name l1var to c36

From the Variance function window we see that the level 1 variance is the following function of the level 1 parameters, whose estimates are obtained by running the model to convergence:

$$\text{var}(\epsilon_{0ij}\text{cons} + \epsilon_{1ij}\text{age}_{ij}) = \sigma_{e0}^2\text{cons}^2 + 2\sigma_{e01}\text{cons} \times \text{age}_{ij} + \sigma_{e1}^2\text{age}_{ij}^2$$
The following plot shows \textit{l1var} versus \textit{age}:

As a result of allowing the level 1 variance to depend on age, there is a statistically significant decrease in the likelihood statistic of $279216 - 279070 = 146$ with 2 degrees of freedom. We shall see later that some of this level 1 variation can be explained by further modelling of the level 2 variation.

7 Repeated measures modelling of non-linear polynomial growth

‘Growth’ in functioning may not be linear for all people over this age range. One simple way of inducing non-linearity is to add a quadratic term in age, which is achieved by including age-squared as an additional explanatory variable in the model. We can ask MLwiN to calculate $age^2$ and add it to model as follows:

- In the \textbf{Equations} window, click on \textit{age} then \textbf{Modify Term}
- Check polynomial, then change poly degree from 1 to 2
- Click Done and respond OK to the message that appears
- The predictor age has been replaced by age^1 and age^2, and variables with these names have also been added to the worksheet
- Fit random coefficients at level 2 to both age^1 and age^2
- Fit a random coefficient at level 1 to age^1
- Click Start to fit the model

At convergence we have:

\[
\begin{align*}
\text{pcs}_{ij} & \sim N(\mu_{ij}, \Omega) \\
\text{pcs}_{ij} &= \beta_0 + \beta_{1ij} \text{age}_{ij} + \beta_{2ij} \text{age}^2_{ij} \\
\beta_0 &= 51.780(0.074) + \mu_0 + \epsilon_{0ij} \\
\beta_{1ij} &= -0.197(0.008) + \mu_1 + \epsilon_{1ij} \\
\beta_{2ij} &= -0.004(0.001) + \mu_2 \\

\begin{bmatrix} 
\mu_0 \\
\mu_1 \\
\mu_2 \\
\end{bmatrix} & \sim N(0, \Omega_\mu) : \Omega_\mu = \\
& \begin{bmatrix} 
34.097(0.726) & 0.959(0.050) & 0.105(0.007) \\
0.959(0.050) & 0.077(0.004) & -0.001(0.000) \\
0.105(0.007) & -0.001(0.000) & 0.000(0.000) \\
\end{bmatrix} \\

\begin{bmatrix} 
\epsilon_{0ij} \\
\epsilon_{1ij} \\
\end{bmatrix} & \sim N(0, \Omega_\epsilon) : \Omega_\epsilon = \\
& \begin{bmatrix} 
25.837(0.308) & 0.059(0.030) \\
0.059(0.030) & 0.005(0.004) \\
\end{bmatrix}
\end{align*}
\]

\[-2\log\text{likelihood(ML Deviance)} = 278522.965(41958 \text{ of 52890 cases in use})\]

The likelihood statistic shows a further drop, this time by 547 with 4 degrees of freedom (one fixed parameter and three random parameters), so there is strong evidence that a quadratic term, which varies from person to person, improves the model.

We also find that the parameter estimates in the level 1 variance-covariance matrix have all decreased. What has happened is that the more complex level 2 variation, which we have introduced in order to model non-linear growth in individuals, has absorbed some of the residual level 1 variation in the earlier model. We can view this final model for the random variation as a convenient and reasonably parsimonious description of how the overall variance is partitioned between the levels.

We can use the Variance function window to calculate the variance at both level 1 and level 2 for each record in the dataset, and these can be added to obtain the total predicted variance.
From the Model menu, select Variance function
Next to level at the bottom of the Variance function window, select 1:occasion
Next to variance output to, select l1var
Click Calc
Now change level to 2:person
Next to variance output to, select l2var
Click Calc
From the Data Manipulation menu select Command interface
In the box at the bottom of the Command interface window, type:
\texttt{calc c37=`l1var'+'l2var'}
Press return, then type:
\texttt{name c37 `totvar'}

We will now plot the level 1 variance, level 2 variance, and total variance against age.

From the Graphs menu, select Customised Graph(s)
At the top left of the window, make sure that D1 is selected
You should find that a plot of l1var versus age has already been specified. If not, select l1var for y, age for x, and select line for plot type. By default the line will be plotted in blue
Now click on the 2\textsuperscript{nd} row under ds #. Select l2var for y, age for x, and select line for plot type. Click on plot style and change the colour to green
Now click on the 3\textsuperscript{rd} row under ds #. Click plot what? Select totvar for y, age for x, and select line for plot type. Click on plot style and change the colour to red
Click Apply

The plot below shows the estimated level 1 variance (blue), level 2 variance (green) and total variance (red) in physical functioning as functions of age. While the level 1 variance is almost constant across the age range, the level 2 variance (and therefore the total variance) increases with age.
Adding person-level explanatory variables

We will now add employment grade (occupation) and gender (fem) to the model. Before adding occupation to the model, we need to declare it as a categorical variable so that MLwiN knows to create and add dummy variables. The gender variable, fem, is already coded as a binary (0,1) variable so can be added in its current form.

- In the Names window, highlight occupation then click Toggle Categorical so that the entry in the categorical column changes to True
- Go to the Equations window, click Add Term. Under variable select fem and click Done
- Click on Add Term again and select occupation. Retain the default occupation_1 (high grade) as the reference category. Click Done
- Click More to fit the model
Both gender and employment grade are significantly associated with physical functioning. Women and employees in lower grades have poorer physical functioning than men and high-grade employees. Note that the intercept now refers to the reference group, i.e. men in high employment grade, of age 50.

9 Does growth differ by group? (cross-level interaction between age and grade)

Does physical functioning decline faster in people from the low employment grades compared with those in the high employment grades? We will add a cross-level interaction between age and occupation to explore this.

- In the Equations window, click Add Term. Change order to 1 and select age and occupation from the two drop-down lists that appear under variable. Click Done
- Click More to fit the model
The reduction in the likelihood statistic is 33 with 4 degrees of freedom (four additional parameters), so there is strong evidence that the growth rate differs by employment grade. However, the interactions with the quadratic age terms appear not to be significant (based on a comparison of the coefficients with their standard errors). We will therefore see if we can simplify the model by removing these terms.

In the Equations window, click on any of the four interaction terms followed by Modify Term.

Next to poly degree, change from 2 to 1.

Click Done.

Variables age^2.occupation_2 and age^2.occupation_3 will be removed from the model.

Click More to fit the model.
The reduction in the likelihood statistic is only 2 with 2 degrees of freedom. We therefore conclude that the interactions with the age-squared term are not needed.

The estimate of the average decline in physical functioning by age in the top employment grade is -0.150. For the low employment grade, it is -0.150 - 0.093 = -0.243.

We can plot the predicted average growth curve for each grade as follows:

- From the Model menu, select Customised Predictions
- Click on age then Change Range. Click Range. Next to Upper Bound, type 25. Next to Lower, type -10. Next to Increment, type 1. This will produce predictions for ages 40 to 75 years (because age was centred about 50). Click Done
- Click on occupation then Change Range. Check category then each of occupation_1, occupation_2 and occupation_3. Click Done
- Click Fill Grid
- Click on the Predictions tab. The grid contains a row for every combination of occupation grade and age for each year in the range -10 to 25. Click Predict to compute the predictions (ignore the message about a -ve definite covariance matrix)
- Click on Plot Grid. Next to x, check age.pred. Under Grouped by, check occupation.pred
- Click Apply

The predicted average growth curve for each occupation grade is plotted. Note that the gender dummy, fem, has been fixed at its sample mean of 0.31 which for a (0,1) variable is equal to the proportion in category 1. We could have fixed this at 0 or 1 to obtain the curves for one gender.