Applied Longitudinal Data Analysis. Chapter 3 Introducing the Multilevel Model for Change

Multi-level model for change allows researchers to answer within-person and between-person questions about change at the same time.

The multilevel model is specified by simultaneously postulating a pair of subsidiary models

1. a level-one submodel that describes how each person changes over time
2. a level-two submodel that describes how these changes differ across people

Statistical models are not statements about sample behaviour, they are statements about the **population process** that generated the data

Most methods of estimation provide a measure of goodness of fit such as an R-squared statistic or residual variance that quantifies the correspondence between fitted model and the sample data. If the model fits well you can draw conclusions about the population from which the sample was drawn.

For example if you fit the model

```
NEURO = 80 + 5(BWGT - 3)
```

(`NEURO` = neurological function and `BWGT` = birthweight in pounds). Which is predicting that an average 3-pound newwborn has a baseline function of 80 and that function is five points higher for each

This regression model is designed for single-observation cross-sectional datasets. To model longitudinal data we need to parse data into level-1 questions about within-person change and level-2 questions about between-person change. If the above study of neurological functioning by birthweight were longitudinal we might ask:

1. how does each child's neurological function change over time? (within-person question)
2. Do children's trajectories of change vary as a function of birthweight? (between-person question)

The distinction between the within-person and the between-person questions is more than cosmetic, it provides the core rationale for a statistical model for change and suggests that a model for change must include components at two levels.

- **Level-1 sub-model** describes how individuals change over time
- **Level-2 sub-model** describing how these change vary across individuals

Taken together these two components form what is known as a multilevel statistical model

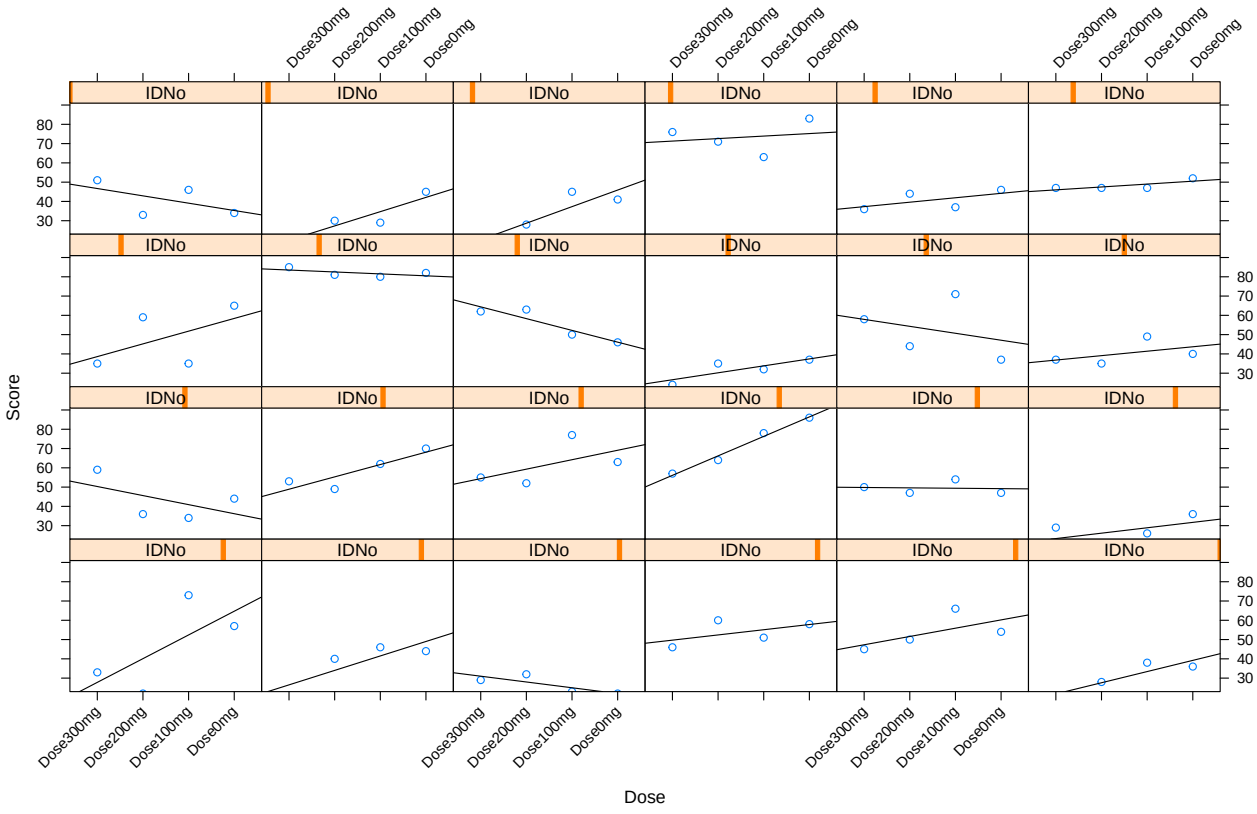
The dataset we will use is my own. This was a longitudinal study on the effects of instruction concerning Dose on caffeine withdrawal symptoms. Ps were randomly assigned to two instruction groups, blind (n=12), who were told falsely that their dose schedule was `300mg, 300mg, 300mg, 0mg` and the aware group (n = 12), who were told truthfully that their dose was `300mg, 200mg, 100mg, 0mg`. Each participant was assessed for level of withdrawawl symptoms 4 times each morning from Tuesday to Friday. Ps filled out a caffeine withdrawal symptom questionnaire. There are six columns: `ID`, `Group` (time-invariant), two covariates (`NumCups` and `NumQuit`) the time-variant factor `Dose`, and the score on the CWSQ at each time point.

The person-period (i.e. long-form) dataset, has three waves.

Level one submodel for individual change

This is the first level of the multi-level sub-model. This component is known as the individual growth model. This is the rate of growth we expect each member of the population to experience during the time-period under study. In the current example the level-1 submodel represents the hypothesised change in each each participant's withdrawal score.

(NOTE: The dataset used in the chapter is not available so I am going to use my own data)



We need to ask what type of population individual growth model might have predicted this sample data. Should caffeine withdrawal decrease linearly or in a curvilinear fashion. These plots confound true change with measurement error so try and look at the picture as a whole rather than individual direction of plots.

The level-1 submodel could be represented as:

Equation 3.1

$$Y_{ij} = [\beta_0 + \beta_1(\text{Dose} - 300)] + [\varepsilon_{ij}]$$

Here Y is caffeine withdrawal score. So what we are saying here is that, in the population from which the sample was drawn, Y_{ij}, caffeine withdrawal score for participant i at time j, is a linear function of Dose (you could equally call this variable Day).

This model assumes a straight line adequately represents each person's true change over time and that any deviations in from linearity observed in sample data result from random measurement error (ε_{ij})

This model uses two subscripts, i and j, to identify individuals and occasions respectively. For this dataset i runs through participants 1 through to 24 and j runs from 1 to 4 (for each wave of date)

Although everyone in this in this data set was assessed on the same three occasions the level-1 submodel is not limited to time-structured designs. The identical model could be applied to datasets where the timing and spacing of waves differs across people. For now we work with a time-structured design.

The square brackets in the equation above separates the two parts of the submodel, the structural part and the stochastic part. This is similar to the idea of true score and measurement error but the implications are broader.

The structural part of the equation embodies our hypotheses about each individual's true change trajectory over time. The equation stipulates that this trajectory is linear with age and has individual growth parameters β_0 and β_1 that characterise its shape for the ith child in the population. The individual growth parameters are the population parameters that lie beneath the intercepts and slopes when we fit OLS-estimated individual change trajectories in our exploratory analysis.

The intercept in the model in this case would be **true** level of caffeine withdrawal on day 1 (monday or start of Tuesday). Therefore the hypothesised trajectory intercepts the y-axis at Boi. Because we hypothesise that each child in the population has his or her own intercept, the growth parameter includes the subscript i. So Participant 1's intercept is $\beta_{0,1}$ and participant 2's is $\beta_{0,2}$ etc.

The OLS-fitted estimate equation has a special representation for the predictor Dose. Here we subtract the intercept value from the Dose parameter before fitting OLS change trajectories to the CWSQ Score data. This is known as *centering* and is a practice that facilitates interpretation of the intercept parameter.

By using Dose - 300 instead of Dose, the intercept in the equation represents the CWSQ for that participant at Dose of 300mg. If we hadn't centred it it would represent their Dose at 0mg, a dose that precedes the collection of data. This is not ideal because

- we would be predicting beyond the data's temporal limits
- we don't know whether the trajectory extends to zero linearly.

By using centering we are aligning $\beta_{0,i}$ with the first wave of data cellection. We are thus estimating the child's true **initial status**.

The slope is the most important parameter in a linear change model because it represents the **rate** at which the individual changes over time. So $\beta_{1,i}$ represents individual i's true rate of change in caffeine withdrawal per day. So participant 1's rate of change in CWSQ over time is $\beta_{1,1}$ and participant 2's is $\beta_{1,2}$ etc.

If $\beta_{1,2}$ is positive it means participant i's withdrawal scores have increased over time. If negative it means participant i's withdrawal has decreased.

In specifying the level-1 submodel that tries to describe everyone in the population we implicitly assume that all the true individual change trajectories have a common algebraic form. But we *don't* assume that everyone has the same trajectories. Because individuals have their own individual growth parameters (intercepts and slopes) different people alsa have their own unique change trajectories.

The level-1 submodel allows us to distinguish the trajectories of different people using just their individual growth parameters. *this leap is the cornerstone of individual growth modelling* because it means we can study *interindividual* differences in change by studying **interindividual differences in the growth parameters**.

Imagine a population in which each member dips into a well of *possible individual growth parameters* and selects a pair—a personal intercept and a slope.

These values then determine that individual's true change trajectory. Statistically we say that each person has drawn his or her individual growth parameter from an underlying bivariate distribution of intercepts and slopes. Because each individual draws his or her coefficients from an unknown *random* distribution of parameters, statisticians often call multi-level models *random coefficients models*.

Table 3.2: Definition and interpretation of parameters in the multilevel model for change

	Symbol	Definition	Illustrative interpretation
Level-1 Model (See Equation 3.1)			
Individual growth parameters	π_{0i}	*Intercept* of the true change trajectory for individual *i* in the population.	Individual *i*'s true value of *COG* at age 1 (i.e., his *true initial status*).
	π_{1i}	*Slope* of the true change trajectory for individual *i* in the population.	Individual *i*'s yearly rate of change in true *COG* (i.e., his *true annual rate of change*).
Variance component	σ_ε^2	*Level-1 residual variance* across all occasions of measurement, for individual *i* in the population.	Summarizes the net (vertical) scatter of the observed data around individual *i*'s hypothesized change trajectory.
Level-2 Model (See Equation 3.3)			
Fixed effects	γ_{00}	Population average of the level-1 intercepts, π_{0i} for individuals with a level-2 predictor value of 0.	Population average true initial status for nonparticipants.
	γ_{01}	Population average difference in level-1 intercept, π_{0i} for a 1-unit difference in the level-2 predictor.	Difference in population average true initial status between participants and nonparticipants.
	γ_{10}	Population average of the level-1 slopes, π_{1i} for individuals with a level-2 predictor value of 0.	Population average annual rate of true change for nonparticipants.
	γ_{11}	Population average difference in level-1 slope, π_{1i} for a 1-unit difference in the level-2 predictor.	Difference in population average annual rate of true change between participants and non-participants.
Variance components	σ_0^2	Level-2 residual variance in true intercept, π_{0i} across all individuals in the population.	Population residual variance of true initial status, controlling for program participation.
	σ_1^2	*Level-2 residual variance in true slope*, π_{1i} across all individuals in the population.	Population residual variance of true rate of change, controlling for program participation.
	σ_{01}	Level-2 residual covariance between true intercept, π_{0i} and true slope, π_{1i} across all individuals in the population.	Population residual covariance between true initial status and true annual rate of change, controlling for program participation.

(p.54) has his or her own individual growth parameters (intercepts and slopes), different people can have their own distinct change trajectories.

The Stochastic Part of the Multilevel Model

The *stochastic* part of the level-1 submodel appears in the second part of the equation above. It is the effect of random error ε_{ij} associated with the measurement of withdrawal via the CWSQ of participant i on occasion j. The level-1 error appears as $\varepsilon_{i,1}$, $\varepsilon_{i,2}$, $\varepsilon_{i,2}$ & $\varepsilon_{i,3}$. Each person's *true* change trajectory is determined by the structural component of the submodel. But each person's *observed* change trajectory also includes measurement errors.

The level-1 submodel accounts for this—the differences between the true and observed trajectories—by including **random errors**: $\varepsilon_{i,1}$ for measurement 1, $\varepsilon_{i,2}$ for measurement 2 etc. Psychometricisns consider random error a natural consequence of measurement fallibility and the vicissitudes of data collection. However it is probably wiser to name the $\varepsilon_{i,j}$ as *level-1 residuals*. Each residual in this dataset represents participant i's value of withdrawal **not predicted by Dose**. We adopt a more vague attitude because we know we can reduce the magnitude of these level-1 residuals by introducing time-varying predictors to the model other than Dose. That we can do this implies that the residuals are not just measurement error.

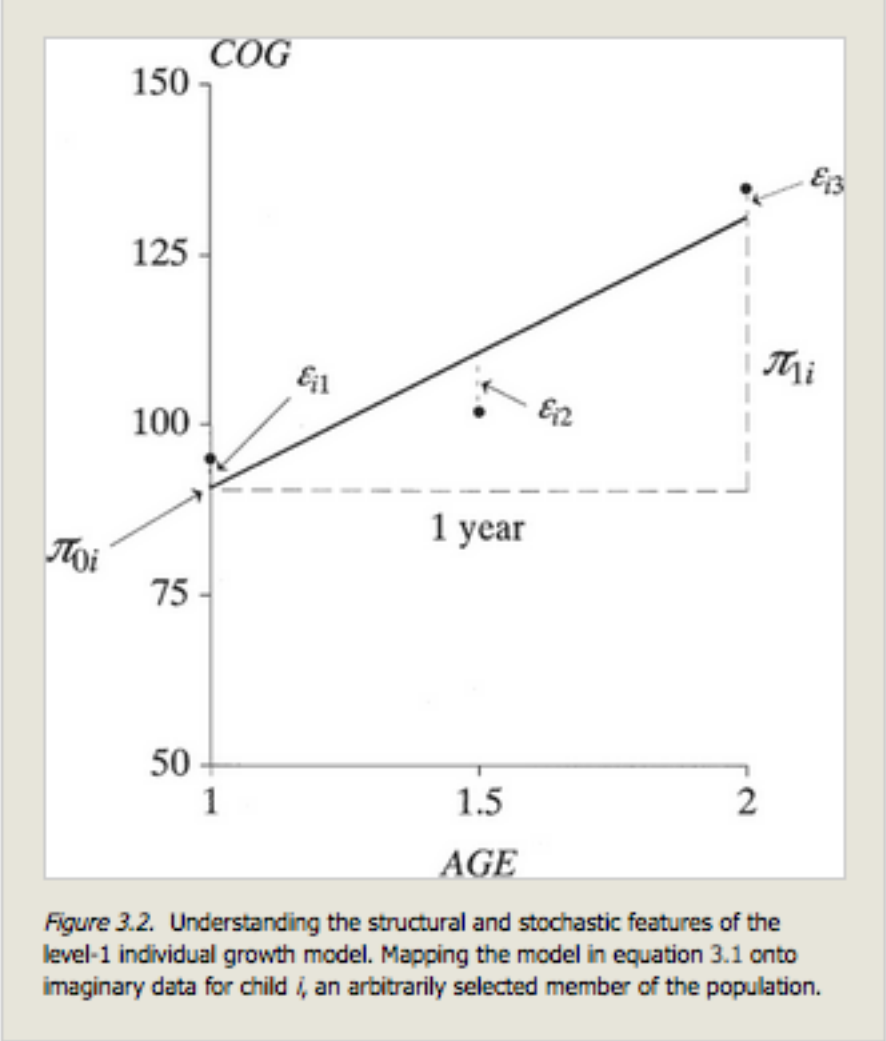


Figure 3.2. Understanding the structural and stochastic features of the level-1 individual growth model. Mapping the model in equation 3.1 onto imaginary data for child *i*, an arbitrarily selected member of the population.

One thing to be aware of is that these errors are ***unobserved***. Because of this we must make assumptions about them

- that the residuals are independently and identically distributed
- that the residuals have homoscedastic variance across occasions and individuals
- normally distributed

This means that irrespective of the individual and location, each error is drawn from a an *underlying distribution* with

- mean of 0
- unknown variance

Equation 3.2

$$\varepsilon_{i,j} \sim N(0, \sigma^2_\varepsilon)$$

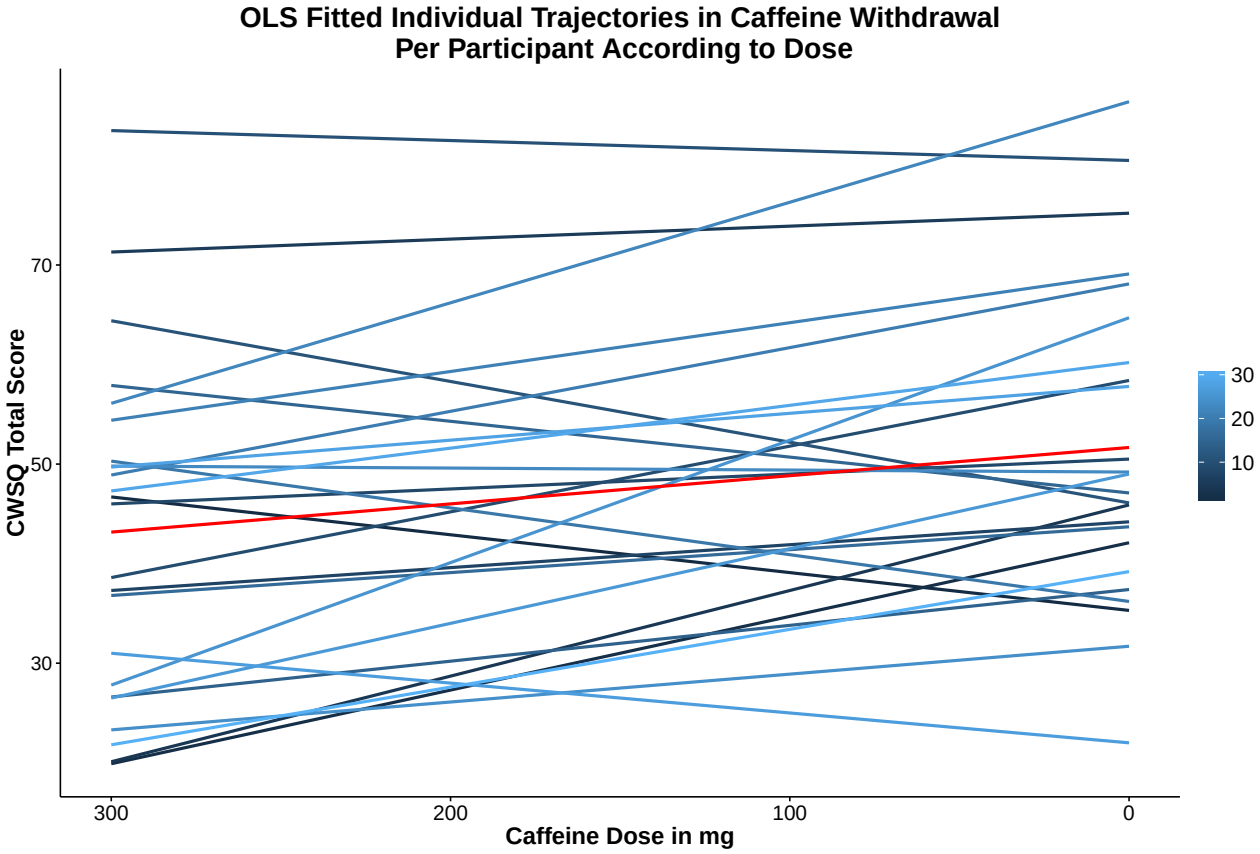
where the symbol ~ means 'is distributed as'. The N stands for normal distribution, with the first element inside the parentheses as the distribution's mean (here 0) and the second element identifying its variance.

This residual variance parameter captures the scatter of the level-1 residuals around each individual's true change trajectory.

Not all assumptions are necessarily true, but we will skip this for the time being.

Relating the level-1 submodel to the OLS exploratory methods of Chapter 2.

The following figure shows the results of using OLS methods to fit the level-1 submodel in equation 3.1 to the data for all 22 participants.



The next are stem and leaf displays for three summary statistics from these models: fitted intercepts, fitted slopes, and estimated residual variances.

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   0 | 8
##   2 | 0014993567
##   4 | 5670135789
##   6 | 26
##   8 | 5
```

Remember this data is centred, so the intercept it the score at 300mg. Each fitted intercept estimates the participant's intial level of caffeine withdrawal. Each fitted slope indicates each participant's true rate of change in withdrawal symptoms over the last 4 days of the study.

```
##
##   The decimal point is 1 digit(s) to the right of the |
##
##   -2 | 3
##   -1 | 8410
##   -0 | 5443320
##    0 | 135788
##    1 | 00149
##    2 | 4
```

The stem and leaf plots for intercept and slope reveal a lot of variation in intercept and slope across participants, suggesting that not all people have identical trajectories of change.

The between-person variation in estimated change trajectories is *necessarily inflated* over the underlying interindividual variability in the unknown change trajectories, becuase fitted trajectories are derived from observed data, and thus are *fallible* representations of true change. The size of the difference in true change trajectories depends on * the quality of your outcome measure * the efficacy of your hypothesised individual growth model

The skewed distribution of residual variances in the stem and leaf suggest great variation in the quality of the OLS summaries across participants (we expect the distribution of this statistic to be skewed because they are "squared" quantities and are therefor bounded by zero below)

When the residual variance is near 0 **the fitted trajectories are decent summaries of the observed data for those participants**. When the residual variances are poorer summaries of the observed data (the observed level of CWSQ Score are further away from the fitted lines, making the magnitude of the difference between observed and fitted data at each time point larger, and therefore also the residual variance larger).

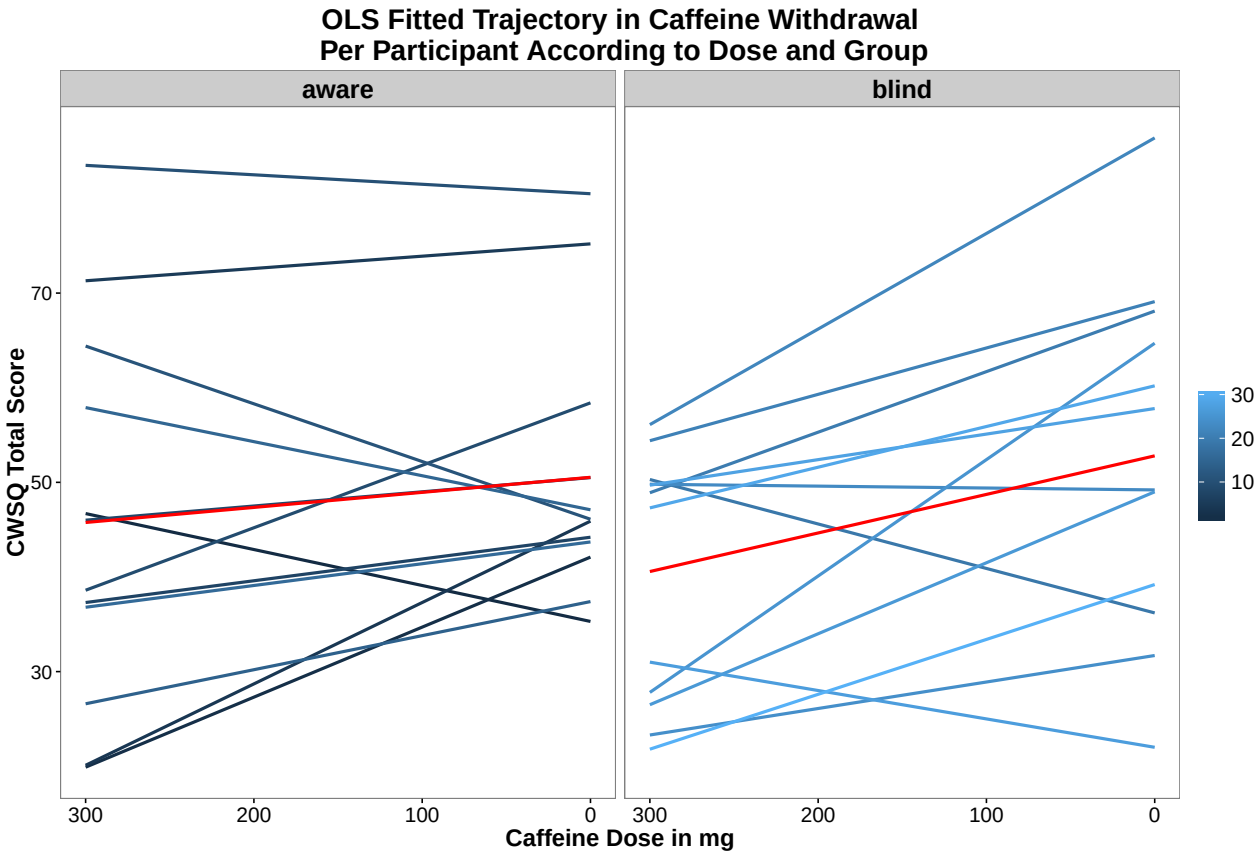
```
##
##   The decimal point is 2 digit(s) to the right of the |
##
##   0 | 000122222234445678
##   1 | 034
##   2 | 6
##   3 | 1
##   4 | 2
```

The Level-2 Submodel for Systematic Interindividual differences in Change

The level-2 submodel codifies the relationship between interindividual change trajectories and the time-invariant characteristics of the individual

The level-1 submodel forces people to differ only in the values of their individual growth parameters i.e. people can only differ in their intercepts and slopes. Thus we can recast vague questions about the relationship between 'change' and predictors as specific questions about individual growth parameters and predictors.

Like all statistical models the level-2 model describes hypothesised population processes, not sample behaviour. But insights gleaned from sample data can often provide valuable insight into model formation. See the next figure, which plots fitted OLS trajectories according to the information given to each participant about their dosing schedule over the four days. The average change trajectory is shown in bold. Participants in the blind group tended to have lower withdrawal scores on day 2 (which is where the graph starts) and their average rate of increase in withdrawal symptoms over the two days was steeper.



It is worth noting the large heterogeneity in intercept and slope *within* each group.

Thus our level two model must account for both the general patterns (between-group differences in intercept and slope) *and* interindividual heterogeneity.

What kind of population model might have given rise to these patterns? We have already discussed four specific features for the level-2 submodel:

1. **Its outcomes must be the individual growth parameters**, β_{0i} (individual intercept) and β_{1i} (slope) from equation 3.1 above. As in regular regression, where we model the population distribution of a random variable by making it an outcome, here, when we model the population distribution of the individual growth parameters they too must be the outcomes.

2. **The level-2 submodel must be written in separate parts, one for each level-1 growth parameter**. When we use a linear change individual growth model at level-1 (like we did in equation 3.1), we need *two level-2 submodels* one for intercept β_{0i} and one for slope β_{1i}.

3. **Each part must specify a relationship between an individual growth parameter and the predictor** (here it is Group or instruction). As you move across the two panels in the split-plot above the value of the predictor shifts from 1 to 0 (aware to blind). This suggests that each level-2 model should ascribe differences in either β_{0i} or β_{1i} to instruction (the fixed or grouping factor) just as we would in a regular regression model.

4. **Each model must allow individuals who share common change predictors (i.e. fixed factors like instruction/group) to vary in their individual change trajectories** This mean that each level-2 submodel must allow for stochastic variation in the indiviudal growth parameters.

These consideration lead us to posulate the following level-2 submodel for these data:

Equation 3.3

$$\beta_{0i} = \gamma_{00} + \gamma_{01} INSTRUCTION + \zeta_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} INSTRUCTION + \zeta_{1i}$$

Note: the β's here appear on the left side of the equation, whereas in the equation above for the observations of the ith individual at the jth time point the beta's appear on the left side of the eqaution. This is what we mean by a hierarchical model. Because each observation is a function of individual parameters, each of which are themselves a function of fixed parameters. This is why we call it a hierarchical model; parameters within parameters.

Like all level-2 submodels, equation 3.3 has more than one component. each resembling a regular regression model. Taken together the two components treat the intercept (β_{0i}) and the slope (β_{1i}) of an individua;'s growth trajectory as level-2 outcomes that may be associated with the predictor *INSTRUCTION*. So we are incorporating the fixed influence of *INSTRUCTION* on each participant's change score (i.e. as opposed the level-1 model alone which only incorporates the individual's true starting ans growth parameters).

Each component also has its own residual—here, ζ_{0i} and ζ_{1i}—that permits the level-1 parameters (the β's) of one person to differ stochastically from those of others.

The two components of this level-2 submodel have *seven* population parameters:

- the four regression parameters (the γ's shown in equation 3.3)
- three residual variance/covariance parameters we will soon define

All are estimated when we fit the multi-level model for change to data. We list, label, and define these in the second half of table 3.2 above.

Structural components of the level-2 submodel

The structural parts of the level-2 submodel contain four level-2 parameters—γ_{00}, γ_{01}, γ_{10}, and γ_{11}—known collectively as ***fixed effects***. Fixed effects capture systematic *interindivdual* differences in change trajectory according to values of the level-2 predictor(s). Two of the fixed effect in the equation above, γ_{00} and γ_{01} are level-2 intercepts; the other two γ_{10} and γ_{11} are level-2 slopes. As with regular regression the slopes are of greater interest because they are the *effects of predictors* (here, the effect of *INSTRUCTION*) on the individual growth parameters. You can interpet the level 2 parameters much as you do regular regression coefficients, except that you must remember **that they describe variation in 'outcomes' that are *themselves* level 1-growth individual growth parameters**. So they describe the growth in growth parameters.

The best way to unpack fixed effects is to identify a *prototypical individual* distinguished by particular predictor values, substitute those values into the level-2 submodel, and examine the consequences.

Level-2 Parameters for influence of Awareness on Individual Growth Trajectory

When postulating a prototypical individual who does is assigned to the 0 condition (the aware group) we set the value of instruction to 0. This

When $INSTRUCTION = 0$, the intercept parameter of the true change trajectory for individual i in the population is equal to the sum of the population average of level-1 intercepts (the 'general' starting point

for all individuals in the aware group) and the residual (i.e. stochastic departure) for all individuals in that group around that average starting point.

So:

$$\beta_{0i} = \gamma_{00} + \zeta_{0i}$$

Note: the first digit in the subscript designates intercept(0) or slope(1), the second designates the group-level parameter, in this case aware.

Also, when *INSTRUCTION* = 0 (i.e. if aware group are coded 0), the *slope* parameter of the true change trajectory for the ith individual is equal to the sum of the **population average of level-1 slopes for people in the aware group** and the average residual difference (i.e. SD) in slope from the average slope for individuals around that average rate of change.

So:

$$\beta_{1i} = \gamma_{10} + \zeta_{10}$$

This model hypothesises that, **in the population of people who are aware of dose reduction**, the values of initial status (at 300mg) and rate of change, so β_{0i} and β_{1i}, are centred around the level-2 (i.e. population) intercept and slope parameters for people in the population who are aware of dose reduction γ_{00} and $\gamma 10$. γ_{00} represents the average starting CWSQ score for those who are aware of dose reduction. $\gamma 10$ represents the **average true rate of *change* of CWSQ for those who are aware of dose reduction**. By fitting the multi-level model for change to the data we address the question:

> What is the average true trajectory of in the population for people who are aware of their caffeine dose reductions?

Level-2 parameters for influence of blindness on individual growth parameters

When we set *INSTRUCTION* = 1 (i.e. blind) group:

$$\beta_{0i} = (\gamma_{00} + \gamma_{01}) + \zeta_{0i}$$

and

$$\beta_{1i} = (\gamma_{10} + \gamma_{11}) + \zeta_{1i}$$

In the population of people allocated to the blind reduction group the values of intitial status and and annual rate of change β_{0i} and β_{1i} are centred around $(\gamma_{00} + \gamma_{01})$ and $(\gamma_{10} + \gamma_{11})$.

Comparing these centres to those for the aware group illustrates that the γ_{01} and γ_{11} capture the effects of *INSTRUCTION*.

- γ_{01} represents the hypothesised difference in average true initial status between groups
- γ_{11} represents the hypothesised difference in average true change trajectory between group.

This allows us to think of level-2 slopes γ_{10} and γ_{11} as "shifts" (i.e. in intercept and trajectory) associated with program participation.

See the figure below. In the figure on the left the thick line represents the average true change trajectory for individuals in one group. This is expressed as $\gamma_{00} + \gamma_{10}$ which is the theoretical influence of this group on intercept and slope. The thinner line represents the change trajectory for a particular individual (i) which is $(\gamma_{00} + \zeta_{0i}) + (\gamma_{10} + \zeta_{1i})$. This represents the influence of the *fixed elements*, the γ's (which would be the same irrespective of the number of times the study is run) and the stochastic elements, the ζ's, which represents the deflection from the group average caused by that individual being who they are. The right side of the graph is the same thing except for the other group. If γ_{10} and γ_{11} are non-zero then the average population trajectories in the two groups differ, if they are both 0 (i.e. no effect of fixed group factor) then they will not differ.

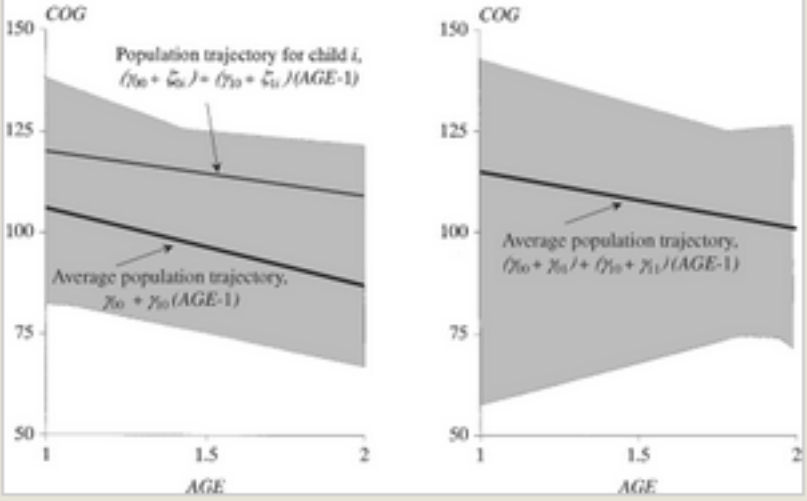
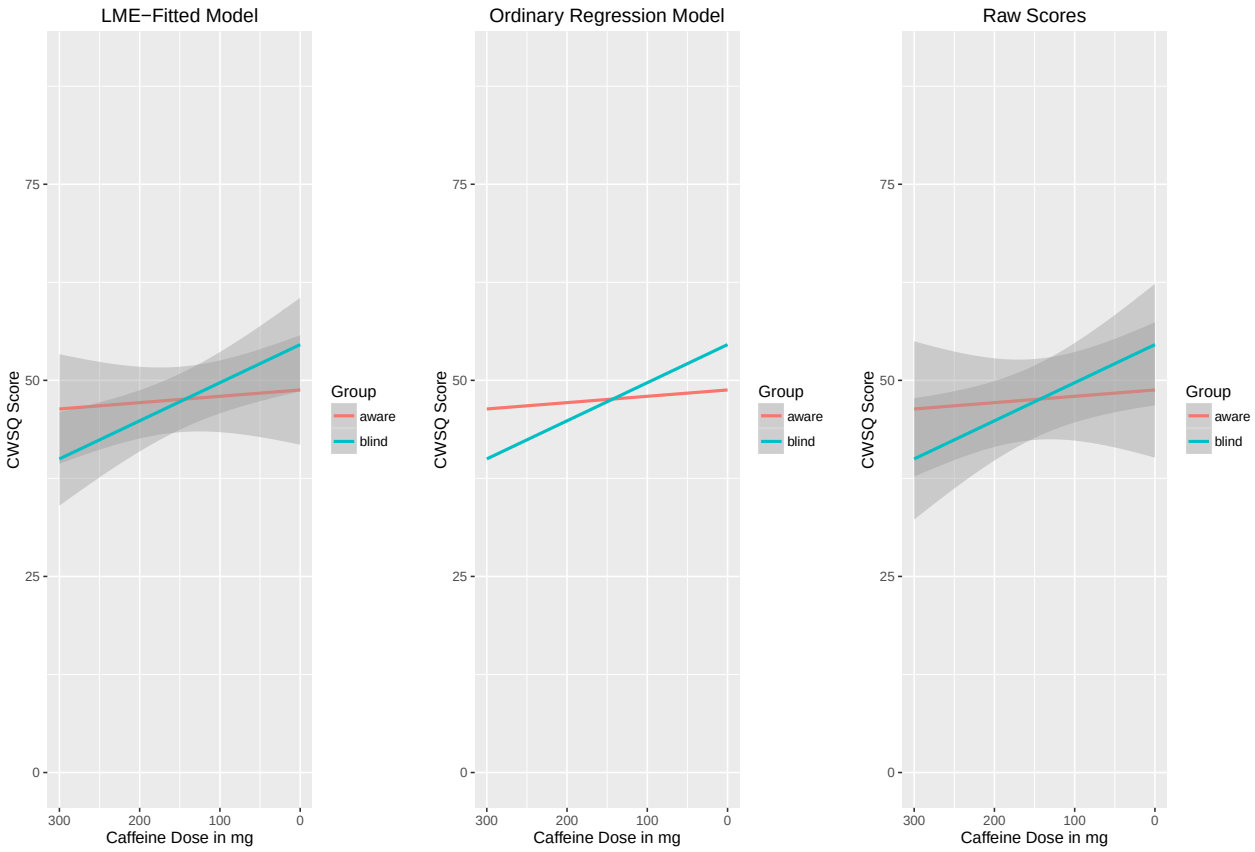


Figure 3.4. Understanding the structural and stochastic features of the level-2 submodel for inter-individual differences in change. Top panel presents fitted OLS trajectories separately by levels of the predictor *PROGRAM*. Bottom panel maps the model in equation 3.3 onto imaginary data for an arbitrary child *i* and the average population trajectory. The shaded portion in each of the lower panels is designed to suggest the existence of many distinct population trajectories for different children.

The two level-2 slope parameters γ_{10} and γ_{11} therefore address the questions: *What is the difference in the average trajectory of true change in caffeine withdrawal associated with instruction?'*

Below are three plots showing are the fitted regression slopes for each group. The left graph is the average population trajectory derived by obtaining fitted values for CWSQ Score for all individuals at all time points with the lme function and then plugging those fitted values into the "score' argument in ggplot2. Comparing the lme to the ordinary lm fitted model it appears there is not a lot of difference in the (*fixed*) effect of group on individual scores. The final graph is the same graph but using raw scores (but fit using lm).



In summary, the two components of this level-2 submodel have *seven* population parameters:

- the four regression parameters (the γ's shown in equation 3.3)
- three residual variance/covariance parameters we will soon define

All are estimated when we fit the multi-level model for change to data. We list, label, and define these in the second half of table 3.2 above.

Stochastic Components of the Level-2 Submodel

Each part of the level-2 submodel contains a residual that allows each person's individual growth parameters to be scattered around the relevant population averages. These residuals, ζ_{0i} and ζ_{1i}, represent the portion of the level-2 outcomes—the individual growth parameters—that remain **unexplained by the level-2 predictors.

We are less interested in their specific values than we are for their population variances, (σ_0^2 for residual variance in intercept and σ_1^2 for residual variance in slope) and covariances ($sigma_{01}$ residual covariance in intercept and slope).

If participant i is a member of the population of 'aware' participants then *INSTRUCTION* takes on the value of 0 and and the level-2 residuals in equation 3.3 represent deviations between his or her initial caffeine withdrawal and change in caffeine withdrawal from the population average initial withdrawal and change in withdrawal (i.e. γ_{00} and γ_{10}).

So the trajectory for child i in group 0 of the example experiment is contained in the left panel of figure 3.4. The child begins at $(\gamma_{00} + \zeta_{0i})$ and has a declining true rate of change $(\gamma_{10} + \zeta_{1i})$.

Other participants' parameters in the 0 group can be similarly constructed by combining fixed parameters γ_{00} and γ_{10} with residuals for each participant.

The lower right panel of figure 3.4 contains a shaded area to suggest the existence of many different true trajectories, all varying around the population mean intercept and slope for the 1 group.

So the variance of true individual intercept and slope parameters around the average represent the population variation around the average intercept and slope for each group. These are called *conditional* residual variances because they represent the portions of intercept and slope *left over* after accounting for for the effects of the model's predictors. The variance parameters allow us to address the question: 'How much heterogeneity in true change remains after accounting for the effects of instruction?'

Covariance Between Individual Intercept and Slope

In the level-2 submodel for change we also allow for a possible association between intercept and slope (e.g. people with higher starting rates of withdrawal may also show higher rates of increase in withdrawal over the five days), therefore we allow the level-2 residuals to be correlated. The population covariance σ_{01} summarises the association across all the individuals in the population between intercept and slope while controlling for group allocation. This parameter addresses the question: 'Controlling for program participation, are true initial status and rate of change related?'

To fit the multi-level model for change to actual data we must make some assumptions about the level-2 residuals (just as we did for the level-1 residuals in equation 3.2). But because we have *two* level-2 residuals we describe their underlying behaviour using a *bivariate distribution*. The standard assumption is that the two level-2 residuals ζ_{0i} and ζ_{1i} are **bivariate normal** with mean 0, unknown variances and unknown covariances σ_0^2 and σ_1^2 and unknown covariances σ_{01}. This can be expressed:

Equation 3.4

$$\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

We interpet this equation in much the same way as we interpret the assumptions about the level-1 residuals in equation 3.2. The first matrix on the right of the tilde in parentheses specifies the bivariate distributions mean vector; here assumed to be 0 for each residual. The second matric to the right of the tilde specifies the bivariate distribution's variance-covariance matrix, also known as the *level-2 error covariance matrix* because it captures the covariration among the level level-2 residuals (or errors). Two variances σ_0^2 and σ_1^2 appear on one diagonal and σ_{01} and σ_{10} on the other (these two are identical because the covariance between ζ_{0i} and ζ_{1i}). The level-2 error variance-covariance matrix (equation 3.4) and the level-1 residual variance (equation 3.2) are known collectively as the model's *variance components*.

Fitting the multilevel model of change to data

Before software for fitting multilevel models researchers used ad-hoc strategies like those presented in chapter 2:

1. They fitted individual growth trajectories in separate within-person OLS regression analyses
2. They then regressed the individual growth parameter estimates obtained on selected level-2 predictors

This approach has two flaws.

1. it ignores information about the individual growth parameters' precision, even though we knows this varies (see varying residual variances in stem and leaf plot)

2. it replaces *true* individual growth parameters—the real outcomes in a level-2 submodel—with their fallible estimates.

The level-2 submodel do not describe the difference between the parameter *estimates* and predictors, but between the parameters *true values* and predictors.

For now we focus on one particular type of estimation—*maximum likelihood*.

Advantages of Maximum Likelihood Estimation

ML is currently the most popular approach to statistical estimation.

As N increases ML estimates have three desirable properties

1. they are *asymptotically unbiased (consistent)*—they converge on the on the unknown true values of population parameters;
2. they are *asypmtotically normally distributed*—their sampling distributions are approximately normal with known variance;
3. they are *asymptotically efficient*—their standard errors are smaller than those derived by others methods.

The attractive properties of ML estimates are *asymptotic*. This means that in practice—in any analysis of a real sample—the properties hold only *approximately*. To enjoy these advantages you need a relatively large sample.

Using ML methods to fit a Multilevel Model

Conceptually maximum likelihood:

> are those guesses for the values of the unknown population parameters that maximise the probability of observing a particular sample of data.

In the context of our study, this means the estimates of the fixed effects and variance components **that make it most likely that we would have observed the specific patterns of change found for these 24 individuals**

To derive an ML estimate for a population parameter a statistician must first construct a *likelihood function*—an expression that describes the probability of observing the sample data as a function of the model's unknown parameters.

The computer then numerically examines the relative performance of potentially competing estimates until those that maximise the likelihood are found.

The likelihood function for the caffeine withdrawal data is a function of the probablility that we would observe the particular temporal pattern of CWSQ scores found in the person-period dataset. We seek estimates of the fixed effects and variance components whose values maximise the probability of observing this observed pattern.

All likelihood functions are expresses as the product of probabilities (or probability densities). For cross-sectional each sample menber usually contributes just one term, related ot the probability that *that* person has his or her observed data. But because longitudinal data consist of several observations, one per measurment occassion, each person contributes several terms to thhe likelihood function, which contains as many terms as there are records in the person-period data set.

Ther particular term that each person contributes on each occasion depends on the specification and assumptions of the hypothesised model. The multilevel model contains structural parts (as shown in, for example, in equations 3.1 and 3.3) and stochastic parts (whose behaviour is described in equations 3,2 adn 3.4). The structural portion describes the true outcome value for person i on occasion j for his or her particular predictor values. It depends on the *unknown* values of the fixed effects. The stochastic portion—the level-1 and level-2 residuals—introduce an element of randomness into the proceedings, scattering the observations for person i on occasion j from the structurally specified value.

To derive a ML estimate we must also make assumptions about the *distribution* of the residuals. Both the level-1 residuals ε_{ij} and level-2 residuals ζ_{0i} and ζ_{1i} are:

- are assumed to be normally distributed
- with mean 0
- unknown variance
- independent of one another
- independent of the model's predictors

Given a model and its underlying assumptions a statistician can write a mathematical expression for the distribution or *probability density* of the the outcome. This expression has

- a mean determined bu the model's structural parts
- a variance determined by its stochastic parts

As a probability density function it also desctibes the likelihood that a person with particular values of the prdictors—only *INSTRUCTION* in equation 3.3—could have particular outcome values using a set of unknown fixed effects and and variance components whose values we would like to estimate. That is, it also contains the actual data values observed for that person on that occasion.

It is a short step from here to the full sample likelihoof, which we reach by exploiting the well-known multiplicative property of independent probabilities (i.e. p(event1) = 0.5 and p(event2) = 0.5 then p(event1and2) = 0.5 x 0.5 = 0.25). Statisticians use this principle to ceate a full sample likelihood from the separate person-period likelihoods just developed.

1. First they write down the value of the probability density of the outcome for each person in the data set on every occasion, thereby describing the likelihood that he/she obtained their particular value of the outcome on that occasion.
2. then they multiply these terms together
3. this yields an expression for the likelihood of simultaneously observing all the data in the person period dataset

Because each person-period likelihood is a function of the data and the unknown parameters, so is their product the full sample likelihood.

To find estimates of the unknown population parameters, we identify those values of the unknown parameters that maximise this product of probabilities. Conceptually, imagine a computer trying out billions of alternative estimates, multiplying them together as specified in the sample likelihood function to yield a numeric value for the likelihood, and comparing those numeric values across all of the billions of tries until those estimates yield the maximum value of the likelihood function are found. These would be the maximum likelihood estimates for this particular problem.

The enormity of this search is daunting even with fast computers. To reduce the search for maximum likelihoods statisticians, instead of finding the value of unknown parameters that maximise the likelihood function, they find the value of parameters that maximise its likelihood. Working with the *log-likelihood function* sacrifices nothing because the values that maximise it also maximise the raw likelihood function. The transformation to logarithms simplifies the intensive numerical calculations involved because

a. the logarith of a product is the sum of the separate logarithms
b. the logarithm of a term raised to a power is the power of the multiplies by the logarithm of the term

And so, since the sample likelihood contains boht multiplicative and exponentiated terms, the logarithmic transformation moves the numerical maximisation into a more workable form, computationally.

Though simpler than maximising the likelihood function itself, maximising the log-likelihood function also involves iteration. All software programs that provide maximum likelihood estimates for the multilevel model for change use an iterative procedure.

1. first the program generates reasonable starting values for *all model parameters*, usually by applying something like the OLS methods we rejected in chapter 2
2. in successive iterations the program gradually refines its estimates as it searches for the log-likelihood function's maximum
3. when this search converges—and the distance between successive iterations is trivially small—the resultant estimates are output. If the algorithm does not converge you must repeat the search with more iterations or improve the model specification

Once the maximum likelihood estimation is complete it is reasonable easy to estimate their associated sampling variation in the form of *asymptotic standard errors (ase)*. We use the term asymoptotic because, as noted earlier, ML standard error are accurate only in large samples. Like any standard error the *ase* estimates the precision with which the estimate has been obtained—the smaller the *ase* the more precise the estimate.

Interpreting Estimated Fixed Effects

The fixed effects parameters of the level-2 submodel—the γ's of equation 3.3—quantify the effects of predictors on the individual change trajectories. In our example they quantify the realtionship between the individual growth parameters and the effect of instruction. We interpret these estimates much as we do any regression coefficient, with one key difference: the level-2 outcomes that these fixed growth parameters describe *are the level-1 individual growth parameters themselves.*

Output of multilevel model of caffeine withdrawal according to instruction and dose reduction.

Below is the output from the lme package fitting a multilevel model. The fixed effects are Days (1 to 4 corresponding to Doses 300-200-100-0 mg) and Group aware(0) vs blind(1). The random effects are the Days by subject interaction.

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: caff
##        AIC      BIC   logLik
##   769.2819 789.7967 -376.641
##
## Random effects:
##  Formula: ~Days | IDNo
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 14.969886 (Intr)
## Days         1.593453 -0.428
## Residual     8.970553
```

```
##
## Fixed effects: Score ~ Days * Group
##                     Value Std.Error DF   t-value p-value
## (Intercept)      45.54167  5.475667 70  8.317099  0.0000
## Days              0.80833  1.272903 70  0.635031  0.5275
## Groupblind      -10.41667  7.743763 22 -1.345169  0.1923
## Days:Groupblind   4.05000  1.800157 70  2.249803  0.0276
##  Correlation:
##                 (Intr) Days   Grpbln
## Days            -0.629
## Groupblind      -0.707  0.445
## Days:Groupblind  0.445 -0.707 -0.629
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3708388 -0.5957788  0.0379000  0.5919955  2.6607279
##
## Number of Observations: 96
## Number of Groups: 24
```

It is important to write down the structural portion of the fitted model before attempting to interpet the fixed effects.

We begin first with the first part of the fitted submodel, for initial status.

$$\beta_{0i} = 45.54 + (-10.42)INSTRUCTION$$

In the population from which the sample was drawn we estimate the *true* initial status (i.e. on Day 1 when Caffeine = 300mg) for participants in the aware group to be 45.54. For participants in the blind group we estimate it to be 10.42 points lower. i.e. the withdrawal score for people in the aware group is 10.42 points higher when at Dose 300mg than it is for those in the blind group. (Note: these values can be found in the (Intercept) and Groupblind rows of the fixed effects component in the summary of the lme model above)

Now we will examine the second part of the fitted model: the rate of change.

$$\beta_{1i} = 0.81 + 4.05INSTRUCTION$$

In the population from which this sample was drawn, we estimate the true average rate of change for the aware group to be 0.81 and for the average person in the population subjected to the conditions of the blind group , we estimate the true rate of change to be 4.05 points greater. The average aware participant's withdrawal score (when INSTRUCTION = 0) went up by only 0.8 points over the course of a week, whereas the average blind participant's score (i.e. when INSTRUCTION = 1) went up by 4.86 (add 4.05 to change trajectory for aware group). Thus our hypothesis seems to be correct; blindness leads to larger increase in withdrawal symptoms. (Note: these values can be found in the Days and Days:Groupblind rows of the fixed effects portion of the summary of the lme model)

Another way to do this is to plot fitted trajectories for prototypical individuals. Even in a simple analysis like this, which involves one dichotomous predictor, it is invaluable to inspect prototypical trajectories visually.

For this multilevel model only two prototypes are possible: a blind ($INSTRUCTION = 1$) and a aware ($INSTRUCTION = 0$). Substituting these values into equation 3.5 yields the estimated initial status and daily growth rates for each of these groups

When Program = 0 (aware)

$$\beta_{0i} = 45.54 + (-10.42)(0) = 45.54$$
$$\beta_{1i} = 0.81 + 4.05(0) = 0.81$$

When Program = 1 (blind)

$$\beta_{0i} = 45.54 + (-10.42)(1) = 35.12$$
$$\beta_{1i} = 0.81 + 4.05(1) = 4.86$$

So here we have the average initial value and change parameters β_{0i} and β_{1i} expressed in terms of the fixed effects of each group $(\gamma_{00} + \gamma_{01})$ and $(\gamma_{10} and \gamma_{11})$.

Single Parameter Test for the Fixed Effects

As in regular regression you can conduct a hypothesis test on each fixed effect (i.e. each γ) using a single parameter test. Relevant p-values for each of these parameters are displayed in the output for the lme model under 'fixed effects'.

Most commonly the null hypothesis examined is that the value of the parameter, the parameter's influence on the average individual scores in the population, is 0 (i.e. $H_0 : \gamma = 0$) against the two-sided alternative that it is not $H_1 : \gamma \neq 0$.

Examining Estimated Variance Components

Estimated variance and covariance components ae trickier to interpret as their numeric values have little absolute meaning and there are no graphic aids to fall back on. Interpretations for a single fitted model is especially difficult as there are no benchmark for evaluating the components' magnitudes. The increases the value of hypothesis testing, for at least the tests provide sone benchmark (against the null value of 0) for comparison.

Interpreting Estimated Variance Components

Variance components assess the amount of outcome variability left—at either level-1 or level—after fitting the multi-level model. They quanitfy the amount of *residual variation* in the true initial status remaining after we control for program participation.

Tests for variance components evaluate whether there is any remaining *residual* outcome variation that could potentially be explained by other predictors. The level of the particular variance component—level-1 or level-2—dictates the tyoe of predictor that might be added. In genera, all the tests are similar in that they assess the evidence concerning the null hypothesis that that the parameter's population value is 0 $(H_0 : \sigma^2 = 0)$, against the alternative that it is not $H_1 : \sigma^2 \neq 0$