Analysing repeated measures with Linear Mixed Models

(Random Effects Models)

(2)

4 repeated measures - one group

Written by: Robin Beaumont e-mail: robin@organplayers.co.uk

Date last updated 06 January 2012

Version: 1

". . . these methods provide powerful and flexible tools to analyse, what until relatively recently, have been seen as almost intractable data"

Everitt 2010 ,p.250

Types of model covariance structures		
name	**Variance at each time point**	**Correlation (covariance) between measurement times**
Scaled identity	constant	No correlation (i.e. independent)
Compound symmetry	constant	constant
Diagonal	different at each time	No correlation (i.e. independent)
Unstructured	different at each time	Different for each time pairing
AR(1) =Autoregressive	constant	Correlation gets less as time points get further apart (i.e. t_1, $t_2 = \rho$ but t_1, $t_3 = \rho^2$)

How this document should be used:
This document has been designed to be suitable for both web based and face-to-face teaching. The text has been made to be as interactive as possible with exercises, Multiple Choice Questions (MCQs) and web based exercises.

If you are using this document as part of a web-based course you are urged to use the online discussion board to discuss the issues raised in this document and share your solutions with other students.

This document is part of a series see:
http://www.robin-beaumont.co.uk/virtualclassroom/contents.htm

Who this document is aimed at:
This document, in one in a series at the website above, aimed at those people who want to learn more about statistics in a practical way.

Good luck and do let me know what you think. Robin Beaumont

Acknowledgment

My sincere thanks go to Claire Nickerson for not only proofreading several drafts but also providing additional material and technical advice.

Videos to support material

There are a set of YouTube videos to accompany this chapter which can be found at:
http://www.youtube.com/view_play_list?p=05FC4785D24C6E68

Contents

1. Introduction

Much of this chapter is based upon Twisk 2006 chapter 6 Multilevel analysis in Longitudinal studies. This has been done deliberately so that those who are interested can compare this analysis using SPSS with that using the purpose build multilevel modelling software MLwinN. MLwinN is developed by the Centre for Multilevel modelling at Bristol university (UK) http://www.bristol.ac.uk/cmm/ and is free to academics, they also produce some excellent teaching materials.

id	health	lifestyle	time
1	4.2	2.51	1
1	3.9	2.10	2
1	3.9	2.16	3
1	3.6	2.26	4
2	4.4	2.48	1
2	4.2	2.34	2
2	4.6	2.45	3
2	4.1	2.57	4
3	3.7	2.17	1
3	4.0	2.39	2

The dataset we are going to consider consists of four measurements on two variables taken on 147 individuals, The two measures are a lifestyle (time dependent input variable) and health status measure (outcome, independent variable).

From the screen dump opposite you will note that the data is in long format, with 4 columns and 588 rows.

We will analyse this data in five stages and at each stage we will consider one or more questions we are trying to answer from the analysis

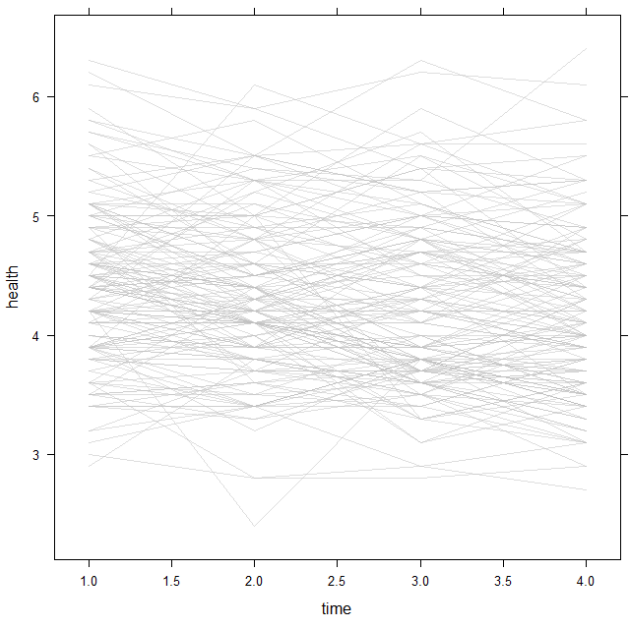
- Graphical inspection of the data, are there any discernable patterns either over time for the repeated measures and/or across subjects?
- Naive regression analysis not taking into account the clustering - Are the repeated measures sufficiently clustered for us to consider a multilevel approach?
- A random intercept model (i.e. taking into account each persons baseline measure) - - Are the baseline scores sufficiently varied across individuals to warrant us taking this into account by including a random intercept parameter?
- A random slope model (i.e. taking into account each person's slope and ignoring the baseline). - Do the angles of the regression lines (i.e. 'trajectories ') vary across individuals to such a degree as to warrant us taking this into account by including a random slope parameter?
- A random intercept + slope model (i.e. taking into account each person's baseline and slope). - Do the angles of the regression lines (i.e. 'trajectories ') and their initial levels interact with each other. Meaning do we need to also include both random intercept/slope parameters?

2. Graphical inspection of the data

We can inspect the data several ways either focusing on each individuals trajectory (i.e. profile/spaghetti plots) or the differences between individuals, by superimposing them, so in effect we are either focusing on time changes at the individual or those across the individual level at a particular time point of over the time of the study.

Exercise

In the previous chapter you were asked to draw profile plots for the the Twisk health lifestyle dataset (twisk_longitudinal_long.sav) dataset, which you may have realised is what we have been discussing above. If you did not carry out that exercise please do so now. The result is shown opposite.

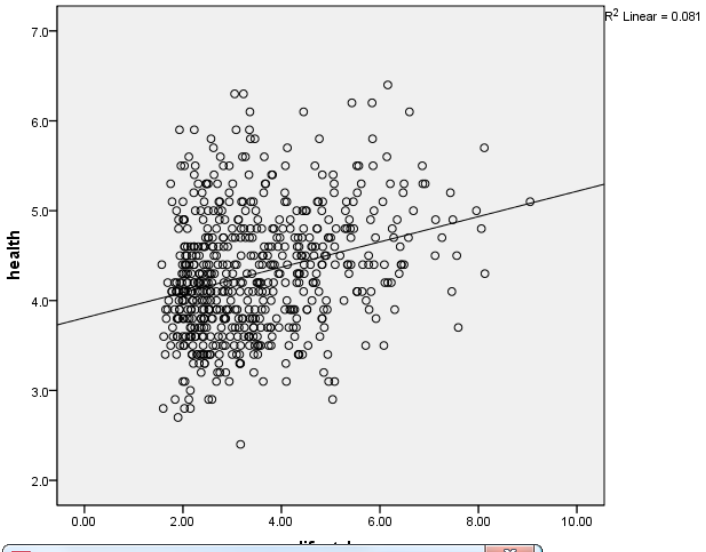


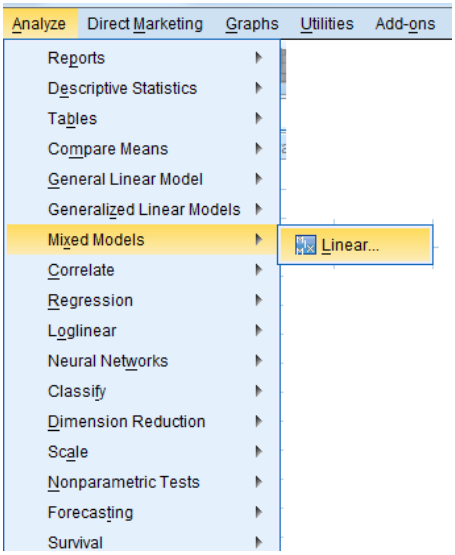
There looks like a wide discrepancy between initial and also subsequent health scores. To be able to comment more we would need to know what was the possible minimum and maximum scores for the health variable.

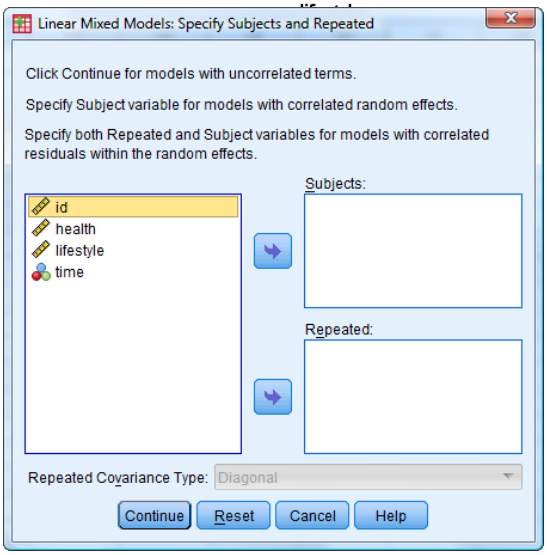
Over time there is possibly a slight downward linear trend, with the more shaded area moving from 4.5 initially to 3.5 at time 4. But this might just be my eyes hence the need for some statistical analysis to help unravel this!

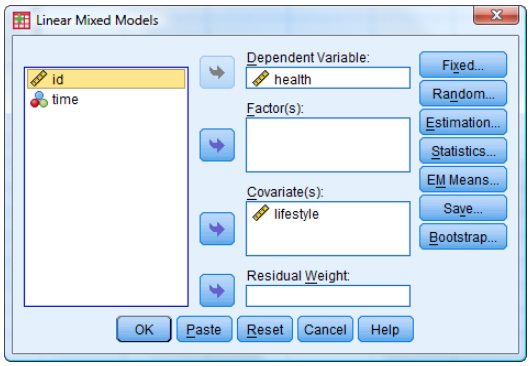
3. Ignoring the clustering - standard regression

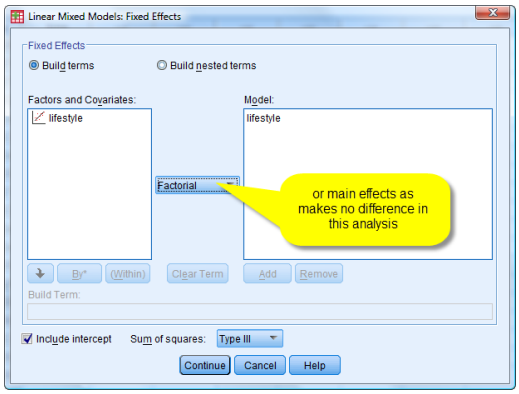
The first analysis we will carry out will ignore the fact that the four observations for each subject are dependent upon one another. Therefore this is just a standard regression analysis with a scatterplot, the results of which are shown below, along with the steps demonstrating how it was achieved in the Mixed Linear Models dialog boxes.

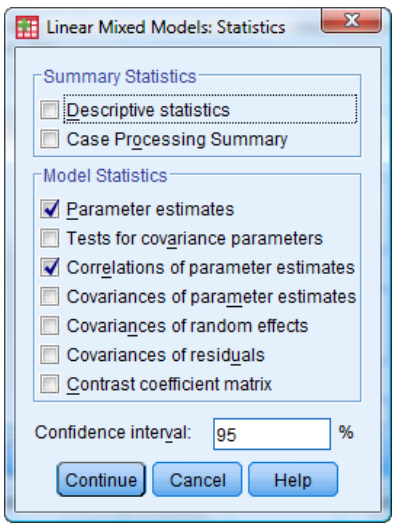


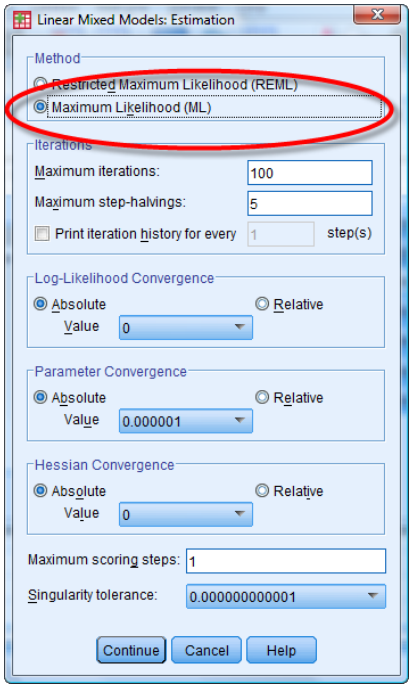






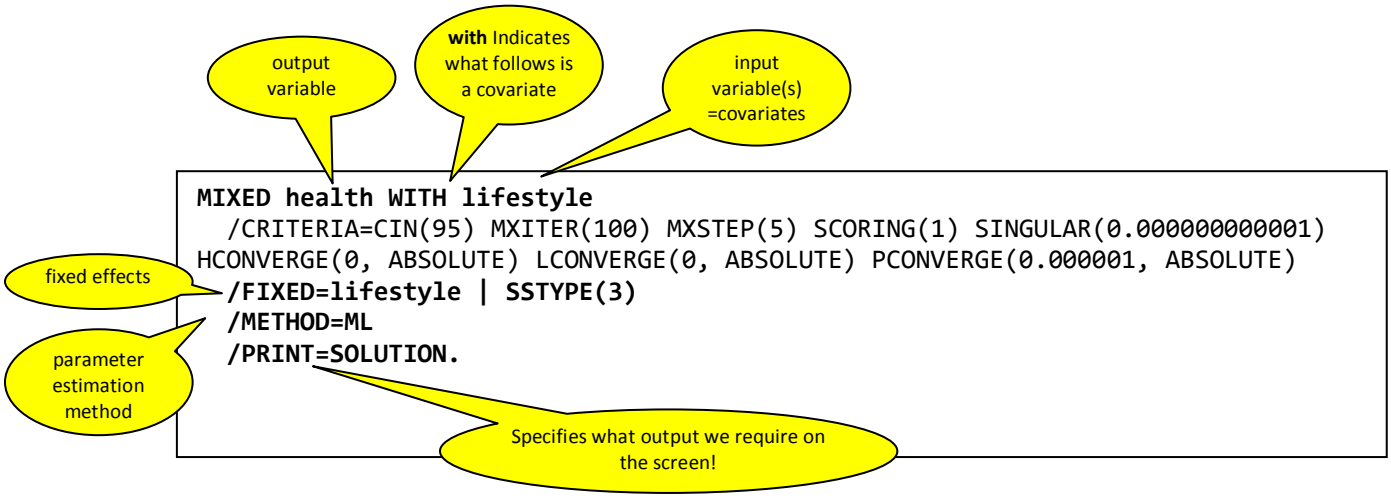






All the above dialog boxes should be familiar to you now from the last chapters' exercises and explanations.

By clicking on the paste button in the Linear Mixed Models dialog box you can see the SPSS syntax that is created:



As you have used R in the past you probably now see that it is probably easier to use SPSS syntax rather than the dialog boxes as we progress but for now I will provide both.

Model Dimension[a]

		Number of Levels	Number of Parameters
Fixed Effects	Intercept	1	1
	lifestyle	1	1
Residual			1
Total		2	3

a. Dependent Variable: health.

Information Criteria[a]

-2 Log Likelihood	1184.665
Akaike's Information Criterion (AIC)	1190.665
Hurvich and Tsai's Criterion (AICC)	1190.706
Bozdogan's Criterion (CAIC)	1206.795
Schwarz's Bayesian Criterion (BIC)	1203.795
The information criteria are displayed in smaller-is-better forms.	

a. Dependent Variable: health.

Estimates of Fixed Effects[a]

Parameter		Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
	Intercept	3.808062	.073363	588	51.907	.000	3.663976	3.952148
	lifestyle	.140831	.019518	588	7.215	.000	.102497	.179164

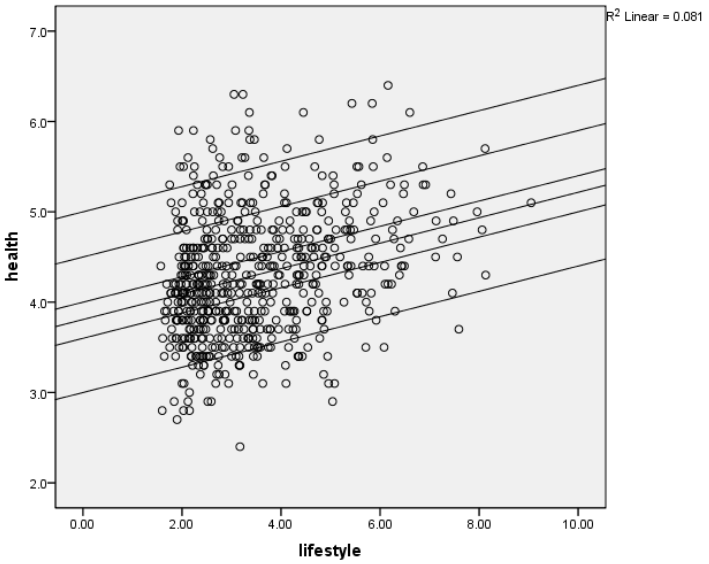
a. Dependent Variable: health.

Correlation Matrix for Estimates of Fixed Effects[a]

Parameter		Intercept	lifestyle
	Intercept	1	-.928
	lifestyle	-.928	1

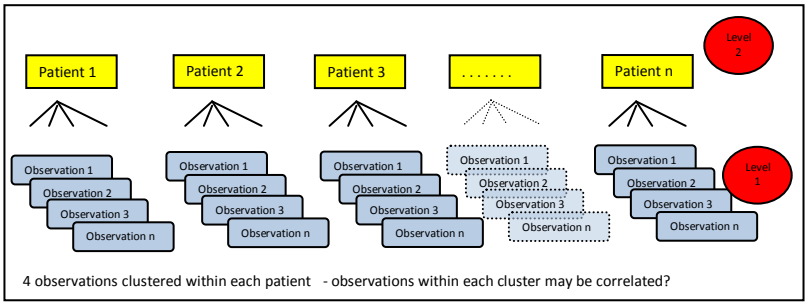
a. Dependent Variable: health.

While the p value for the lifestyle parameter estimate is reported to be zero (in actual fact it is not, SPSS is just indicating that it is less than one in a thousand) the correlation between lifestyle and health (from the scatterplot above) is .286 (√0.081 from scatterplot). Remember that the coefficient of determination equal .286, indicates that just less than a third of the variability in health can be accounted for by the lifestyle measure if we accept that the repeated measures are independent, which we have a strong suspicion is not the case. Also interestingly, if we carry out a ordinary least squares regression we get a unstandardised B of 0.141 identical to that obtained above by the Maximun Likelihood method.



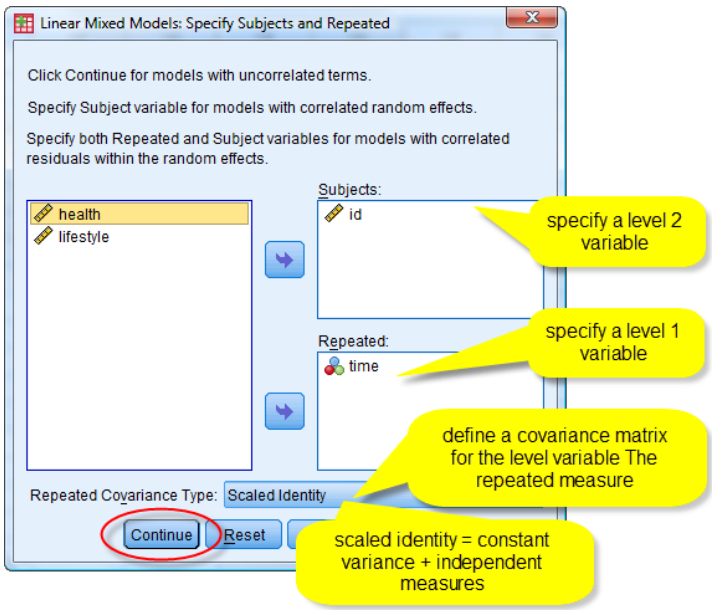
Before we can take into account the dependency over time for the 4 repeated measures for each subject we must first consider the fact that each subject has a different baseline measure, and take this into account in our model. I have tried to indicate in the scatter plot opposite what we are now going to do. Imagine an infinite number of parallel regression lines through the data, with the greatest number going through the mean value of health (and lifestyle) and their distribution being normal.

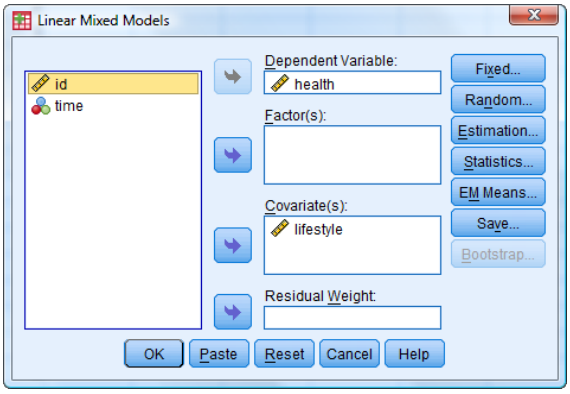
4. Adding a random intercept (adding a level 2 component)

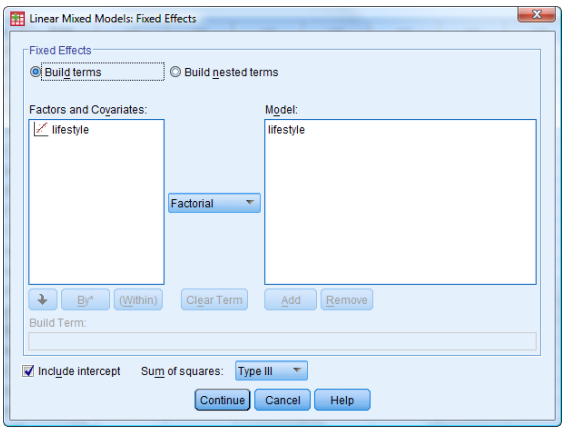


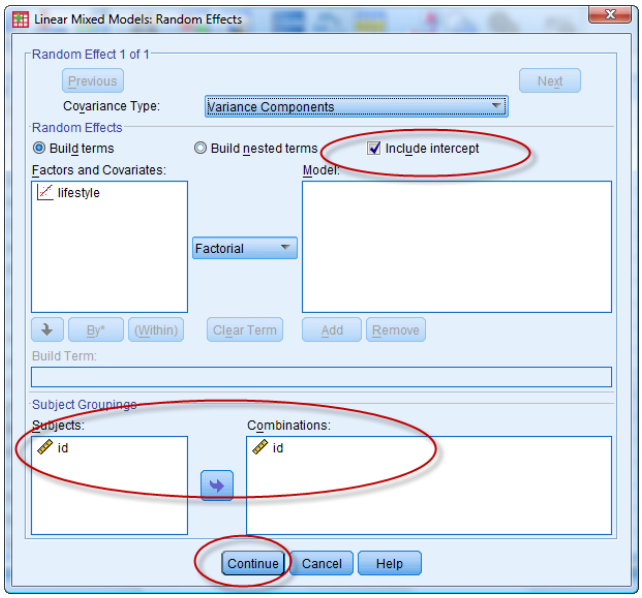
This time we add the id variable to the Subjects box in the initial Linear mixed models dialog box, along with the time variable to the repeated measures box (in effect specifying a random variable at the lowest level). This is now what is called a **multilevel model**. The Linear Mixed Models variables box and fixed effects boxes stay the same.

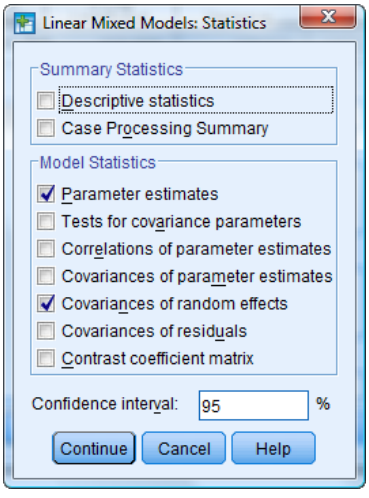
Note that we have used the scaled identity repeated covariance matrix as we are assuming at the moment that the 4 repeated measures are independent and share a common variance over the 4 measurements in this model. We now click on the random effects button and specify our random effect, the intercept. We also specify the subject grouping = ID.

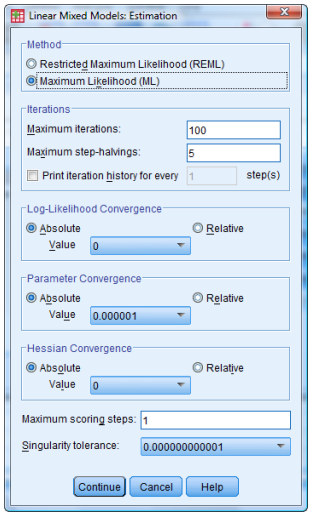












The above configuration of dialog box settings produces the following SPSS syntax:

```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001)
HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(VC)
  /REPEATED=time | SUBJECT(id) COVTYPE(ID).
```

fixed effects

random effects

Specifies the residual covariance matrix

identifies independent units

VC (variance components) when specified in RANDOM subcommand, a **scaled identity (ID)** structure is assigned to each of the effects specified.

If VC specified on the REPEATED subcommand, it will be replaced by the diagonal (**DIAG**) structure. Note that the diagonal structure has the same interpretation as the variance components structure *(p.1223 SPSS 19 syntax guide)*.

The output is reproduced below

Model Dimension[b]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables	Number of Subjects
Fixed Effects	Intercept	1		1		
	lifestyle	1		1		
Random Effects	Intercept[a]	1	Variance Components	1	id	
Repeated Effects	time	4	Identity	1	id	147
Total		7		4		

a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

b. Dependent Variable: health.

Information Criteria[a]

-2 Log Likelihood	812.006
Akaike's Information Criterion (AIC)	820.006
Hurvich and Tsai's Criterion (AICC)	820.075
Bozdogan's Criterion (CAIC)	841.513
Schwarz's Bayesian Criterion (BIC)	837.513

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: health.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	4.053661	.094250	395.162	43.010	.000	3.868366	4.238956
lifestyle	.070424	.023081	538.548	3.051	.002	.025085	.115763

a. Dependent Variable: health.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error
Repeated Measures	Variance	.127774	.008648
Intercept [subject = id]	Variance	.320993	.041875

a. Dependent Variable: health.

Random Effect Covariance Structure (G)[a]

	Intercept | id
Intercept | id	.320993
Variance Components	

a. Dependent Variable: health.

We can see that the fit of the model has changed by looking at the Information criteria table, taking the simplest measure first, the -2LL is now 812.006 compared to 1184.665 a difference of 372.659, we know that this follows a chi chi distribution if the two models are the same (basically the difference due to random sampling) and the degrees of freedom is equal to the difference of the degrees of freedom for the two models. Inspecting the Model dimension tables we see that the difference is 4-3=1 . Using the expression `1-pchisq(372.659,1)` in R as described in the previous chapter and the youtube video, we find that it is smaller than R can manage to compute returning the value 0. Therefore we can say that this model is definitely an improvement over the previous one. We also notice that all the other information criteria values have dropped, where smaller numbers indicate better fit.

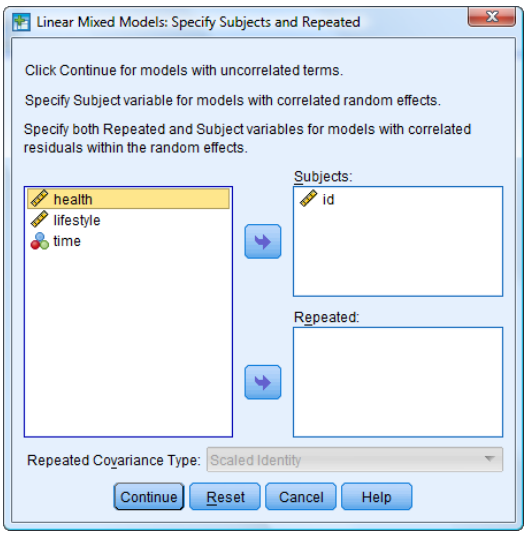
Having said all of the above Twisk (2006 p.93) states that **the inclusion of a random intercept is a conceptual necessity for a repeated measures design**, after all is not taking into account each subjects baseline measure and how the repeated measure vary in relation to it one of the defining characteristics of this type of analysis.

Inspecting the Fixed parameter estimates we can see that while both are still statistically significant, the values have changed slightly.

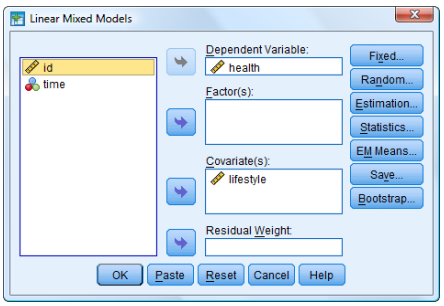
One further aspect needs to be considered and that is how much of the models variance is distributed within the set of four repeated measures and also between the subjects. Think about this in terms of level one and level 2 variance, where level one is the repeated measures. We simply do this by dividing the between subject variance (.3209) by the total variance (.3209 + .1277). The values are obtained from the Estimates of covariance parameters table, producing a result of 0.3209/(.3209 + .1277) = 0.7153 This value is called the **Intraclass correlation coefficient (ICC)**. And once again Fisher was there first, devoting a whole chapter to it in his statistical methods for research workers.

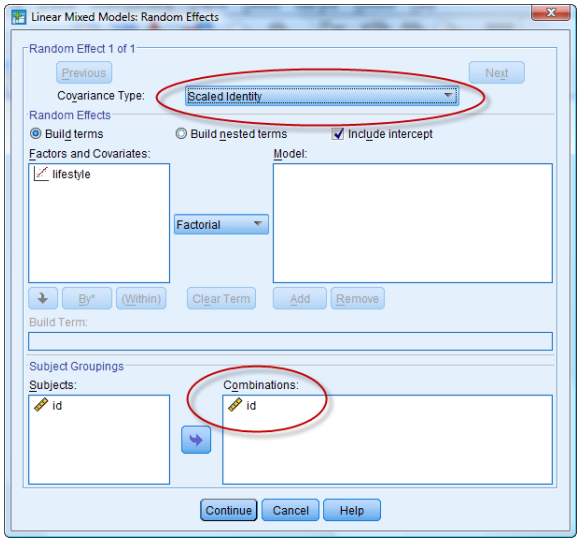
Techie note: When you select Variance component in SPSS it applies either a ID or DIAG covariance structure depending upon it being specified in the random effect dialog box (applies ID) or the repeated covariance type dialog box (applies DIAG).

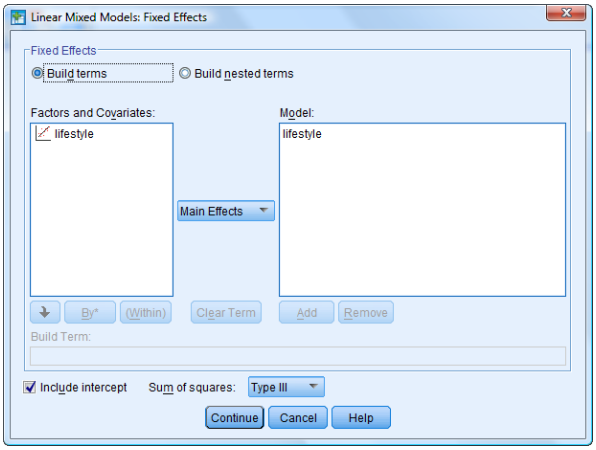
4.1 Equivalent analysis



As I mentioned before SPSS tends to have mind of its own concerning the Linear Mixed Models dialog boxes and you can define the above model another way, instead of defining time as a repeated measure you can drop it from the model entirely and end up with the same results as given on the previous page, basically this is because we have in both models created two random effects one being the intercept and the other the repeated measures.








```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(ID).
```

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	lifestyle	1		1	
Random Effects	Intercept	1	Identity	1	id
Residual				1	
Total		3		4	

a. Dependent Variable: health.

Information Criteria[a]

-2 Log Likelihood	812.006
Akaike's Information Criterion (AIC)	820.006
Hurvich and Tsai's Criterion (AICC)	820.075
Bozdogan's Criterion (CAIC)	841.513
Schwarz's Bayesian Criterion (BIC)	837.513

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: health.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	4.053661	.094250	395.162	43.010	.000	3.868366	4.238956
lifestyle	.070424	.023081	538.548	3.051	.002	.025085	.115763

a. Dependent Variable: health.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error
Residual		.127774	.008648
Intercept [subject = id]	Variance	.320993	.041875

a. Dependent Variable: health.

Random Effect Covariance Structure (G)[a]

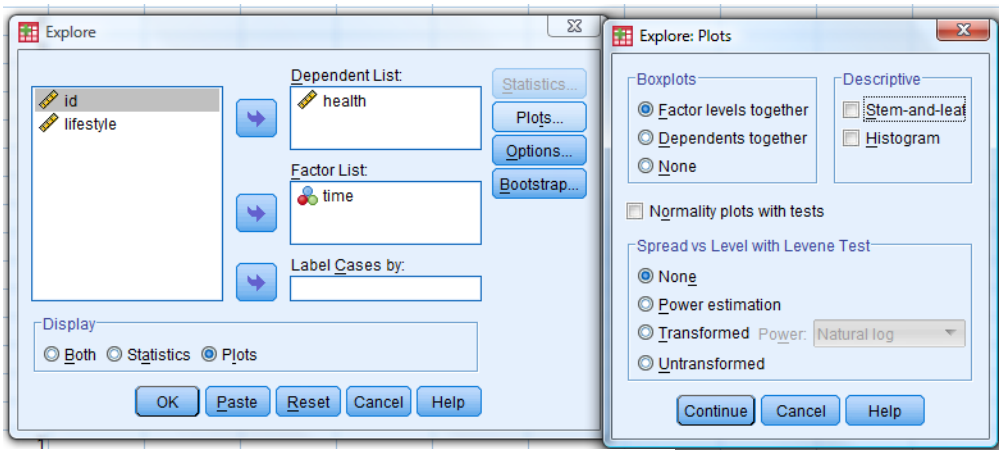
	Intercept | id
Intercept | id	.320993

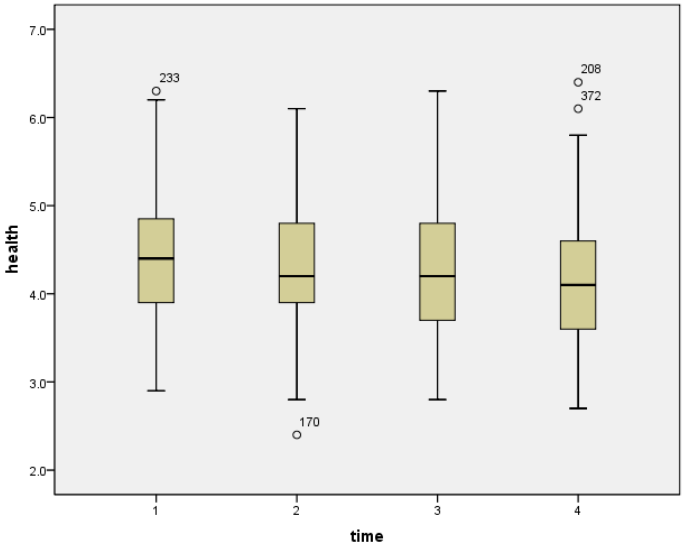
Identity

a. Dependent Variable: health.

5. Inspecting the variance and correlations at each time point

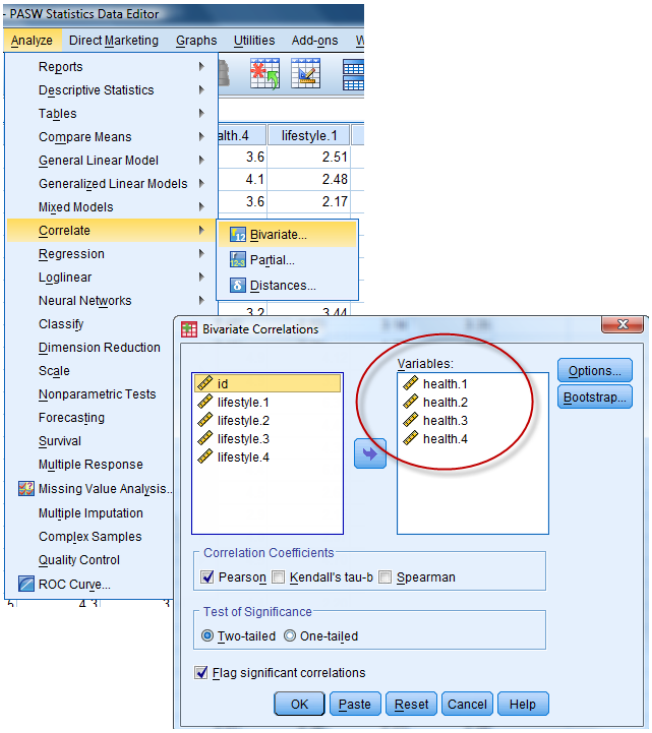
It would be sensible to visually inspect the variance at each time point to see if it is sensible to assume that the error covariance matrix we have specified (i.e. constant variance over time) is tenable. The dialog boxes below are self explanatory, called up from the menu option Analyse -> descriptive statistics -> explore.





Clearly the variance at each time appears very similar (we could also have gleamed this from the profile plot presented earlier).

Whereas the correlations between the various time points for the outcome variable (i.e. health measure), seem to get less as the intervals get further apart (suggesting a autoregressive covariance structure, see the previous chapter). I found it is easy to obtain the table of correlations using the data in wide format.



Correlations

		health.1	health.2	health.3	health.4
health.1	Pearson Correlation	1	.756**	.704**	.670**
	Sig. (2-tailed)		.000	.000	.000
	N	147	147	147	147
health.2	Pearson Correlation	**.756****	1	.774**	.779**
	Sig. (2-tailed)	.000		.000	.000
	N	147	147	147	147
health.3	Pearson Correlation	**.704****	**.774****	1	.847**
	Sig. (2-tailed)	.000	.000		.000
	N	147	147	147	147
health.4	Pearson Correlation	**.670****	**.779****	**.847****	1
	Sig. (2-tailed)	.000	.000	.000	
	N	147	147	147	147

**. Correlation is significant at the 0.01 level (2-tailed).

Clearly the observations are highly correlated and variance for each time point is fairly constant. This exploratory data analysis lets use decide which is the most appropriate error covariance matrix for the model which we will make use of latter.

6. A random slope only model

Usually we would now move onto adding a random slope parameter to the model, however for pedagical reasons I would like now to create a random slope only model.

This time rather than using the dialog boxes I have just edited the SPSS syntax in the last example.

```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION
  /RANDOM= lifestyle | SUBJECT(id) COVTYPE(ID).
```

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	lifestyle	1		1	
Random Effects	lifestyle	1	Identity	1	id
Residual				1	
Total		3		4	

a. Dependent Variable: health.

Information Criteria[a]

-2 Log Likelihood	863.296
Akaike's Information Criterion (AIC)	871.296
Hurvich and Tsai's Criterion (AICC)	871.364
Bozdogan's Criterion (CAIC)	892.803
Schwarz's Bayesian Criterion (BIC)	888.803

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: health.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	4.126805	.095068	495.206	43.409	.000	3.940019	4.313590
lifestyle	.035758	.033847	343.415	1.056	.292	-.030816	.102332

a. Dependent Variable: health.

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error
Residual		.137157	.009432
lifestyle [subject = id]	Variance	.033431	.004643

a. Dependent Variable: health.

Random Effect Covariance Structure (G)[a]

	lifestyle | id
lifestyle | id	.033431

Identity

a. Dependent Variable: health.

Not focusing too much on this analysis we note by looking at Akaike's Information Criterion (AIC) that this model, with a value of 871, is a better fit than the naive regression analysis approach where the AIC was 1190 but poorer in contrast to the random intercept model with a AIC of 820

A logical next step is now to test a model with both random effects included.

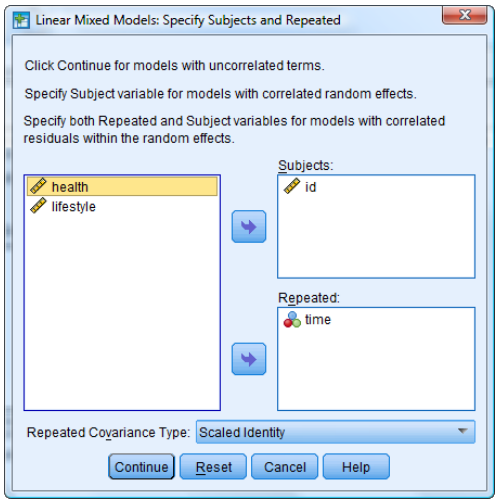
7. Random intercept and slope

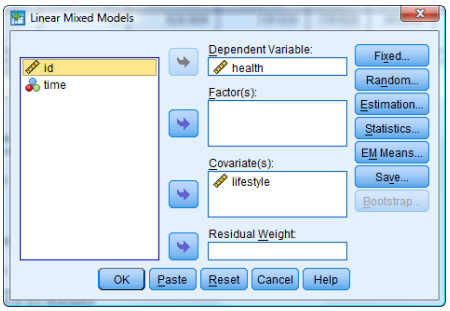
Again using the SPSS syntax above and modifying produces the required model.

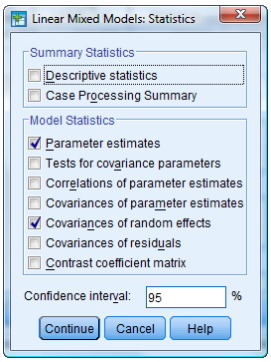
```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION
  /RANDOM= intercept lifestyle | SUBJECT(id) COVTYPE(un).
```

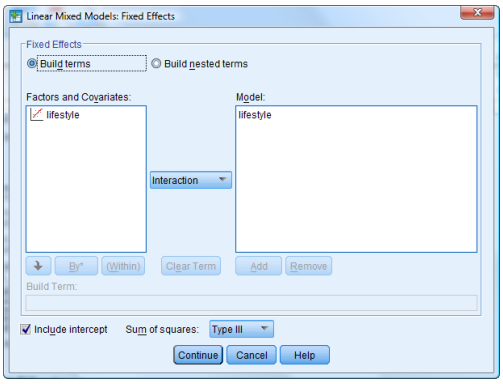
Interestingly using the dialog boxes described on the next page, including specifying the repeated measures produces identical output (except for the model dimension table), basically because we identified the error covariance matrix at the bottom level that is the repeated measures as being independent and having a shared variance.

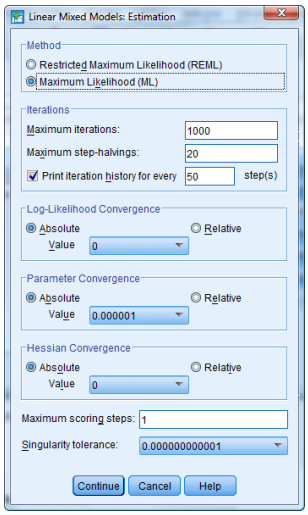
```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION
  /RANDOM=INTERCEPT lifestyle | SUBJECT(id) COVTYPE(UN)
  /REPEATED=time | SUBJECT(id) COVTYPE(ID).
```

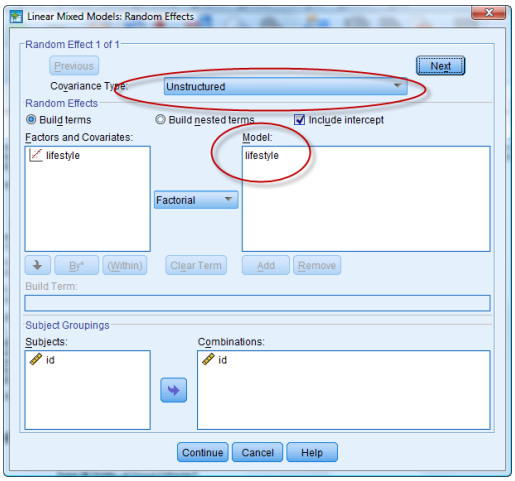












The results are given below.

Warnings
Iteration was terminated but convergence has not been achieved. The MIXED procedure continues despite this warning. Subsequent results produced are based on the last iteration. Validity of the model fit is uncertain.

Information Criteria[a]	
-2 Log Likelihood	817.807
Akaike's Information Criterion (AIC)	829.807
Hurvich and Tsai's Criterion (AICC)	829.952
Bozdogan's Criterion (CAIC)	862.067
Schwarz's Bayesian Criterion (BIC)	856.067
The information criteria are displayed in smaller-is-better forms.	
a. Dependent Variable: health.	

Model Dimension[b] Using dialog boxes specifying subjects and repeated levels						
		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables	Number of Subjects
Fixed Effects	Intercept	1		1		
	lifestyle	1		1		
Random Effects	Intercept + lifestyle[a]	2	Unstructured	3	id	
Repeated Effects	time	4	Identity	1	id	147
Total		8		6		
a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.						
b. Dependent Variable: health.						

Model Dimension[b] Using syntax on previous page only specifying subjects not repeated levels					
		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	lifestyle	1		1	
Random Effects	Intercept + lifestyle[a]	2	Unstructured	3	id
Residual				1	
Total		4		6	
a. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.					
b. Dependent Variable: health.					

Estimates of Covariance Parameters[b]			
Parameter		Estimate	Std. Error
Repeated Measures	Variance	.115421	.008041
Intercept + lifestyle [subject = id]	UN (1,1)	.921763	.129628
	UN (2,1)	-.172021	.017870
	UN (2,2)	.044685[a]	.000000
a. This covariance parameter is redundant.			
b. Dependent Variable: health.			

Random Effect Covariance Structure (G)[a]		
	Intercept | id	lifestyle | id
Intercept | id	.921763	-.172021
lifestyle | id	-.172021	.044685
Unstructured		
a. Dependent Variable: health.		

Estimates of Fixed Effects[a]							
Parameter		Estimate	Std. Error	df	t	Sig.	95% Confidence Interval
							Lower Bound / Upper Bound
	Intercept	4.057631	.117377	491.193	34.569	.000	3.827008 / 4.288253
	lifestyle	.066013	.031126	617.984	2.121	.034	.004887 / .127139
a. Dependent Variable: health.							

Firstly notice that there was a problem with the computer estimating the parameter values, normally we would ignore the subsequent results, however it is educational to inspect the above. While the fixed coefficients have changed little the lifestyle parameter has lost its statistical significance. However inspecting the statistical significance of the random component of lifestyle, obtained by selecting the Tests for covariance parameters option in the Linear mixed models

statistics dialog box P values are provided all of which are very small (highly significant), the equivalent in SPSS syntax is TESTCOV.

Estimates of Covariance Parameters[b]

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		.115421	.008041	14.353	**.000**	.100689	.132308
Intercept + lifestyle [subject = id]	UN (1,1)	.921763	.129628	7.111	**.000**	.699704	1.214294
	UN (2,1)	-.172021	.017870	-9.626	**.000**	-.207045	-.136997
	UN (2,2)	.044685[a]	.000000	.	**.**	.	.

a. This covariance parameter is redundant. The test statistic and confidence interval cannot be computed.

b. Dependent Variable: health.

Twisk (2006 p.94) using WLwiN manages to obtain a convergent result for this model with a -2ll value of 810 compared to the non convergent 817.8 value above, he compares the value with the previous -2ll and finds that it is not a better fit than the random intercept model. You can also obtain a convergent result in SPSS if you change the maximum scoring steps in the estimation dialog box to 10.

What are the key points from the above analysis, using the random intercept model:

There is a statistically significant (p=0.002) relationship between health and lifestyle with a magnitude (β) of 0.07 (95% CI 0.025 to 0.115) when allowing for variability across individual health measures.

We can interpret the parameter value from both a between patient perspective or from a temporal/patient perspective:

- Two subjects whose lifestyle differs by one unit will have a predicted health difference of 0.07
- A subjects whose lifestyle improves by one unit will have a predicted health improvement of 0.07

8. Changing the (level one) error covariance matrix

You may remember that in the previous section we noted that the repeated measures for the dependent variable appeared to have equal variance for the 4 time periods whereas the correlations tailed off as the time intervals got further apart suggestions a autoregressive covariance structure, taking this into account should make our model fit better so lets try it.

```
MIXED health WITH lifestyle
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(5) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=lifestyle | SSTYPE(3)
  /METHOD=ML
  /PRINT=G  SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(ID)
  /REPEATED=time | SUBJECT(id) COVTYPE(AR1).
```

down from 812 (df=4) ☺

down from 812 (df=4) ☺

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables	Number of Subjects
Fixed Effects	Intercept	1		1		
	lifestyle	1		1		
Random Effects	Intercept	1	Identity	1	id	
Repeated Effects	time	4	First-Order Autoregressive	2	id	147
Total		7		5		

a. Dependent Variable: health.

Information Criteria[a]

-2 Log Likelihood	782.660
Akaike's Information Criterion (AIC)	792.660
Hurvich and Tsai's Criterion (AICC)	792.763
Bozdogan's Criterion (CAIC)	819.543
Schwarz's Bayesian Criterion (BIC)	814.543

The information criteria are displayed in smaller-is-better forms.

a. Dependent Variable: health.

Estimates of Fixed Effects[a]

Parameter		Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
	Intercept	4.0133	.0965	381.533	41.588	.000	3.8236	4.2031
	lifestyle	.0822	.0238	500.842	3.450	.001	.0353	.1290

a. Dependent Variable: health.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	381.533	1729.589	.000
lifestyle	1	500.842	11.899	.001

a. Dependent Variable: health.

Random Effect Covariance Structure (G)[a]

	Intercept | id
Intercept | id	.259627
Identity	

a. Dependent Variable: health.

Estimated variance at each time

Estimates of Covariance Parameters[a]							
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Repeated Measures	AR1 diagonal	.188278	.031003	6.073	.000	.136343	.259994
	AR1 rho	.459551	.093044	4.939	.000	.259485	.621796
Intercept [subject = id]	Variance	.259627	.048515	5.351	.000	.180007	.374464
a. Dependent Variable: health.							

Estimated correlation between any two consecutive time points for health

down from.32 on page 8

The Autoregresive covariance pattern matrix: $\sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$

The autoregressive covariance matrix requires two parameters, the variance σ^2 constant at each level (time) and the correlation ρ^2 (rho) at each level (time). So in the above table the row labelled AR1 diagonal provides an estimate of the variance for each repeated measure and the row labelled AR1 rho is the estimate of the correlation between each repeated measure.

We can see that the model now has a -2LL of 782.66 (df=5) compared to 812.006 (df=4) the difference being 29.346 (df=1) and if the two are only exhibiting sampling variability the value follows a chi square distribution with df=1 therefore using the expression `1-pchisq(29.346,1)` in R as described in the previous chapter and the youtube video, this gives a P value of 6.054135e-08 = .00000006541 which is highly significant.

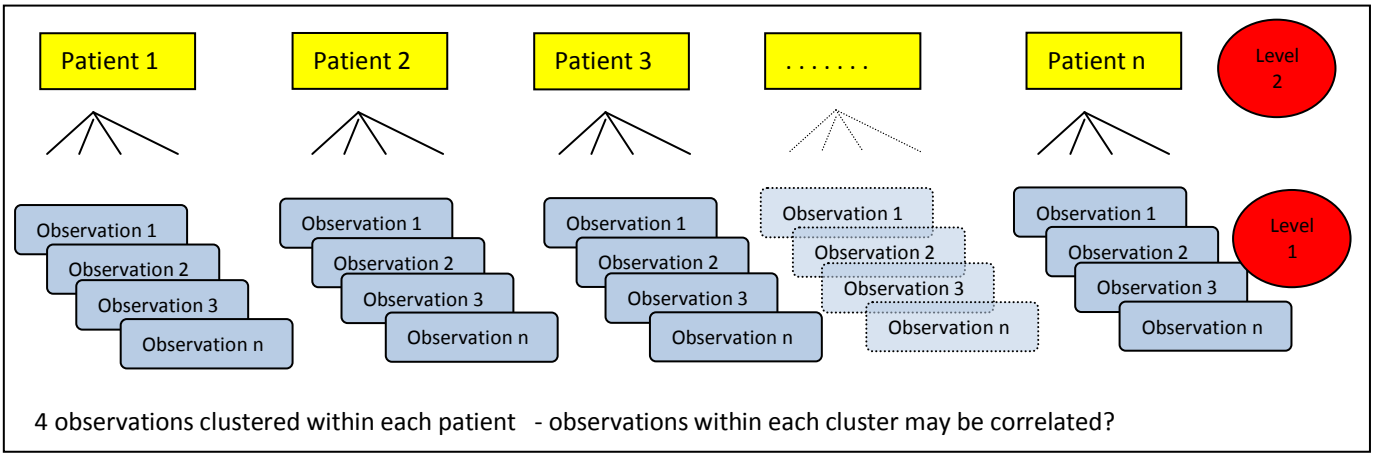
While we may have gained a greater degree of fit, nothing else has really changed, now the statistically significance (p=0.002) between health and lifestyle has gone from 0.002 to 0.001 and the magnitude (β) from .07 to .08 which is nothing to write home about!

9. Summary

In this chapter we investigated using a random intercept and random slope parameter to model repeated measures. We deliberately took an example from a book that uses different software so you are able to compare results, and also Twisks more elegant thorough narrative, I hope you will read Twisks book.

The model could have been more complex, such as having two group of patients (designated by some treatment or characteristic and similarly we could have also included other repeated measures, some of these refinements will be considered in the next chapter.

The important message to take home with that we are in effect modelling data at two levels and taking into account the correlated nature of the repeated measures within patients and at the same time the independence between patients.



4 observations clustered within each patient - observations within each cluster may be correlated?

There has been much research in this area of statistics in the last decade and one commentator said 'the number of books concerning repeated measures is rapidly approaching infinity' sorry I can't find the reference but I certainly feel that this is the case putting together this chapter.

On a more positive note Everitt 2010 ,p.250 lighlights that "these methods provide powerful and flexible tools to analyse, what until relatively recently, have been seen as almost intractable data"

10. References

Bickel R 2007 Multilevel Analysis for Applied Research: It's Just Regression! The Guilford Press.

Bollen K A, Curran P J. 2006 Latent Curve Models: A Structural Equation Perspective. Wiley

Bollen K. A. 1989 Structural Equations with Latent Variables. Wiley.

Brown H, Prescott R 2006 (2nd ed.) Applied Mixed models in Medicine. Wiley.

Campbell 2006 (2nd ed.) Statistics at square 2. BMJ books.

Crawley M J 2005 Statistics: An introduction using R. Wiley

Crowder M J.Hand D J. 1990 Analysis of repeated measures. London: Chapman and Hall.

Diggle P J Hegaerty P Liang K-Y. Zeger S. 2002 (2nd ed.) Analysis of Longitudinal Data. Oxford Science publications.

Duncan T E, Duncan S C, Strycker L A. 2006 (2nd ed.) An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications: Lawrence Erlbaum.

Everitt B S, Hothorn T 2006 (1st ed) A handbook of statistical analyses using R. Chapman & Hall.

Everitt B S, Hothorn T 2010 (2st ed) A handbook of statistical analyses using R. Chapman & Hall.

Field A 2009 Discovering Statistics Using SPSS. Sage [Chapter nineteen describes multilevel modelling including restructuring data].

Fitzmaurice G M, Laird N M, Ware J M, 2004 Applied Longitudinal Analysis. Wiley

Goldstein H. 2011 (4th ed.) Multilevel statistical models. Wiley.

Hand D, Crowder M 1996 Practical Longitudinal data analysis. Chapman & hall

Heck R H, Thomas S L & Tabata L N. 2010 Multilevel and Longitudinal modelling with IBM SPSS. Routledge.

Hedeker, D., and R. D. Gibbons. 2006. Longitudinal Data Analysis. Wiley. Website: tigger.uic.edu/~hedeker/ml.html

Hox J 2010 (2nd ed) Multilevel analysis: Techniques and applications. Lawrence Erlbaum

Landau S, Everitt B S 2004 A handbook of statistical analyses using SPSS. Chapman & Hall.

Leeuw J D,·Meijer E 2008 (eds.) Handbook of Multilevel Analysis. Springer.

Matthews J N S, Altman D G, Campbell M J, Royston J P 1990 Analysis of serial measurements in medical research. B M J 300 230 – 235

Norman G R, Streiner D L. 2008 (3rd ed) Biostatistics: The bare Essentials.

Sarkar D 2008 Lattice: Multivariate Data Visualization with R. Springer.

Sheather S J 2009 A Modern approach to regression with R. Springer. Website, includes R tutorials and datasets: http://www.stat.tamu.edu/~sheather/book/

Singer, J. D., and J. B. Willett. 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press Inc, USA. gseacademic.harvard.edu/alda/

Twisk J W R 2003 Applied Longitudinal analysis for Epidemiology: A practical Approach. Cambridge University Press.

Twisk J W R 2006 Applied multilevel analysis. Cambridge University Press.

Weiss R 2005 Modeling Longitudinal Data. Springer.

West B T, Welch K B, Galecki A T. 2007 Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman & Hall. Book website: http://www.umich.edu/~bwest/almmussp.html

Wu L 2010 Mixed Effects Models for Complex Data. Chapman & Hall.

Zuur A F, ·Ieno E N, Walker N J, Saveliev A A,·Smith G M 2009 Mixed Effects Models and Extensions in Ecology with R. Springer