Exercise 20 - Model diagnostics of mixed models

Zoltan Kekecs

May 20, 2020

Table of Contents

Abstract

The aim of this exercise is to learn how to perform model diagnostics and checking assumptions of linear mixed models.

Data management and descriptive statistics

Wound healing data

In this exercise we will work with the wound healing data we used in the previous exercise.

https://github.com/kekecsz/SIMM32_2019_spring/blob/master/Wound_healing.sav

[*The following description is the same as in the previous exercise*]

This is simulated data about wound healing over time after a surgical procedure. We know that psychological factors, especially stress, can influence recovery after surgery, and the rate of wound healing. Let's say that we have a theory that it is important for hospitalized patients to have a connection with the outside world. So we may think that patients who have a window close to their hospital beds may have a better mood and thus, would show a faster recovery after surgery. This hypothesis is tested in a simple study looking at whether the distance of the patients' bed from the closest window would predict rate of wound healing. Distance is measured in meters, and wound healing is measured by rating the wound using a standardized wound rating measure taking into account the size of the wound, its inflammation and scarring. A physician rates the wound each day for seven days in the afternoon at the same time of the day. We will use this variable as our outcome measure.

Let's say that our hypothesis extends to the role of sunlight in this context, where we suppose that the more sunlight a patient gets the better their recovery would be. To test this hypothesis, our model will take into account whether the bed of the patient is in the north wing or the south wing of the hospital (since the hospital is in the northern hemisphere, we can assume that patients in the south wing would get more sunlight overall during their hospital stay).

We will run the model diagnostics on the model we considered final in the last exercise. In this model we predicted wound rating as an outcome using the fixed effect predictors of time, time^2, distance_window, and location, and we allowed for a random intercept for each participant, but did not include a random slope.

So here we start with reproducing the same dataframe and model what we had last time.

Loading and managing data

After downloading the dataset from GitHub, we create a dummy variable called south_wing to dummy code the variable "location".

```
RECODE location ('south_wing'=1) ('north_wing'=0) INTO south_wing.

EXECUTE.
```

We then restructure our dataset from wide to long format (see previous exercise on repeated measures mixed model analysis on how to do this).

```
VARSTOCASES
 /MAKE wound_healing FROM day_1 day_2 day_3 day_4 day_5 day_6 day_7
 /INDEX=time(7)
 /KEEP=ID distance_window south_wing
 /NULL=KEEP.
```

We center the variable ‘time’ to avoid problems with multicollinearity (see the previous exercise on Model Diagnostics for more information on this issue). Finally, we create a new variable which contains the squared value of the centered time.

```
DESCRIPTIVES VARIABLES=time
 /STATISTICS=MEAN STDDEV MIN MAX.
COMPUTE time_centered=time-4.
EXECUTE.


COMPUTE time_centered_sq=time_centered*time_centered.
EXECUTE.
```

Save this dataset to a new filename, so the raw data is not changed.

Building the model

Now we can build the final model from the previous exercise, wound_healing predicted by south_wing, distance_window, time_centered, and time_centered_sq, with a random intercept. We will need the **residuals and the predicted** values, so ask for them in the **Save** menu in the **Predicted values and residuals** box.

```
MIXED wound_healing BY south_wing WITH time_centered time_centered_sq
distance_window

  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100)
MXSTEP(10) SCORING(1)

    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0,
ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

  /FIXED=south_wing time_centered time_centered_sq distance_window |
SSTYPE(3)

  /METHOD=REML

  /PRINT=CORB  SOLUTION

  /RANDOM=INTERCEPT | SUBJECT(ID) COVTYPE(VC)

  /SAVE=PRED RESID.
```

Model diagnostics for linear mixed models

Assumptions of linear mixed models

Remember that the assumptions for linear models were the following:

- **Normality**: The residuals of the model must be normally distributed
- **Linearity**: Ther relationship between the outcome variable and the predictor(s) must be linear
- **Homoscedasticity**: The variance of the residuals are similar at all values of the predictor(s)
- **No Multicollinearity**: None of the predictors can be linearly determined by the other predictor(s)

The same assumptions need to be satisfied in the case of linear mixed models as well.

In addition, in mixed models we also assume that:

- we have modelled the **dependency structure** of random effects correctly (this is what we set in covariance type)
- the **within-unit residual** errors follow **normal distribution** and have **constant variance**
- **random effects** (the intercepts and slopes if any) are **normally distributed** centered around 0

How to effectively and consistently check these assumptions is still being worked out by researchers and statisticians. But there are already some trend emerging about how can we do model diagnostics. Because these techniques are constantly evolving, you should check new developments on this topic when you use mixed models a few years from now.

Influential outliers

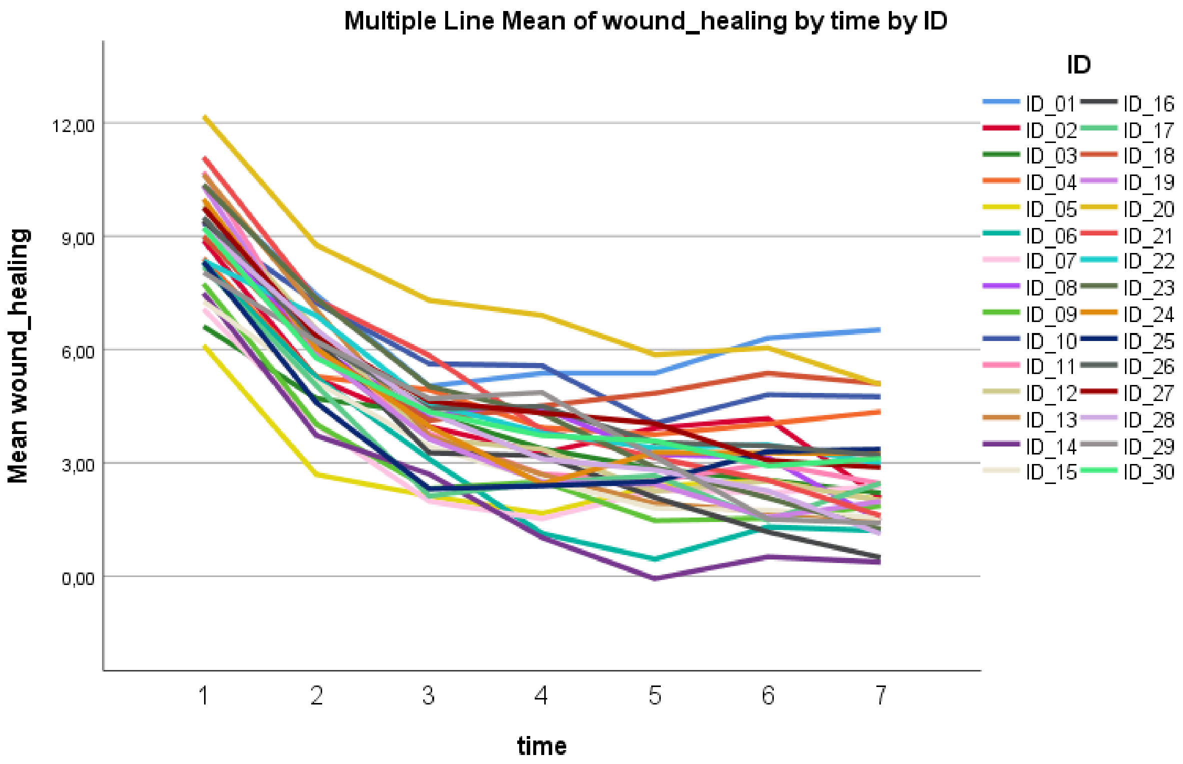
First, we need to check our data for influential outliers which have a large impact on our model.

One method used to check for influential outliers is to exclude observations or units (classes/participants) one-at-a-time and re-fit the model on the leave-one-out dataset. Then, you could compare the model coefficients of these leave-one-out models, and see if any exclusions made a big impact on the coefficients. There is no easy way of doing this in SPSS, but this can be done by using the influence() function from package influence.ME in R.

One alternative to this is to use visual examination of the data. We could for example look at the line charts we created in the previous exercise to look at each individual separately and determine if any of them or the individual observations belonging to them deviate extremely from the other observations or the prediction line.

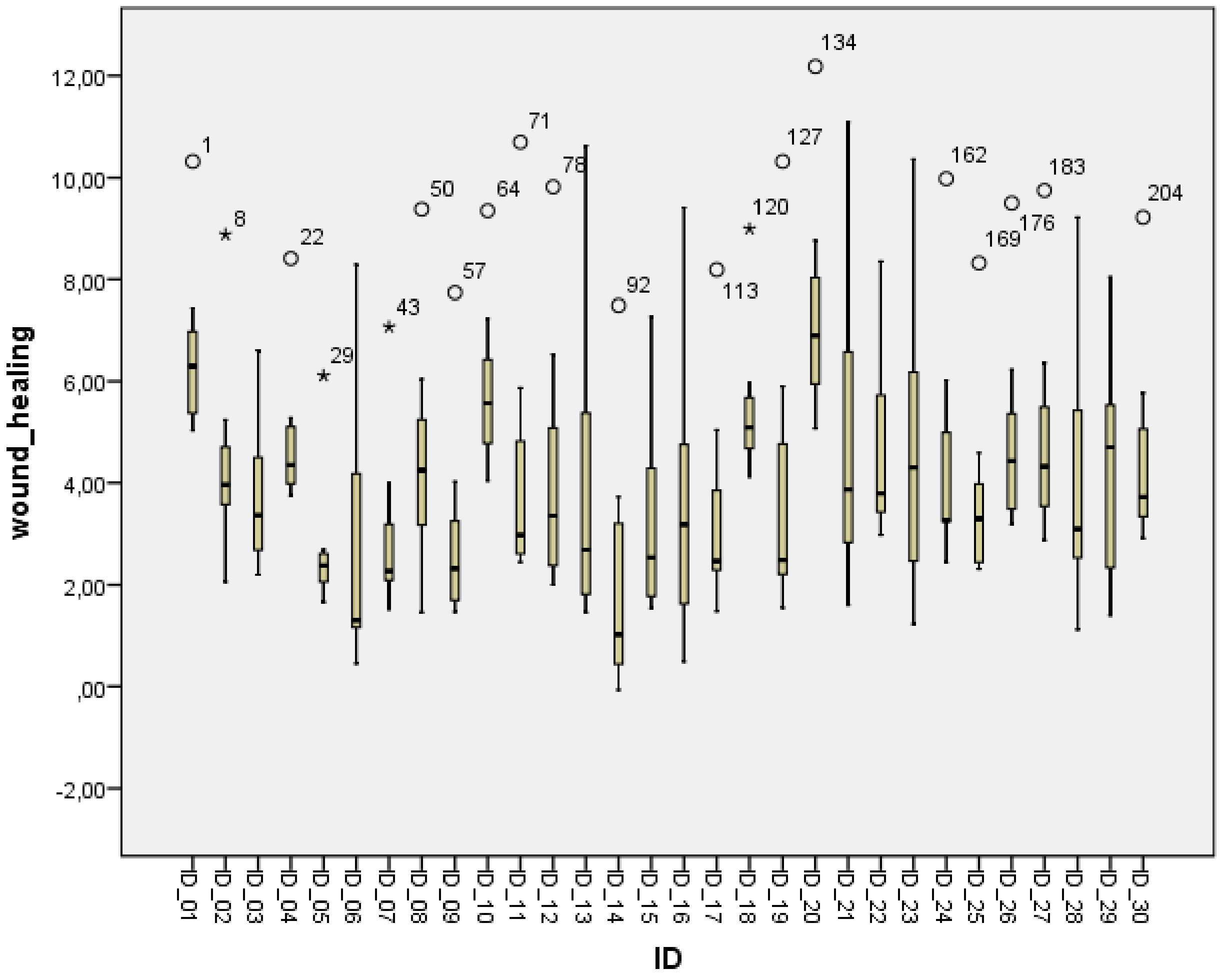
For a better comparison between participants, it is also possible to create a line chart where the different colored lines represent different participants (or other clusters). In our example, wound_healing can be on the Y axis, time can be on the X axis, and ID can determine color.

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=time MEAN(wound_healing)[name="MEAN_wound_healing"]
    ID MISSING=LISTWISE REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: time=col(source(s), name("time"), unit.category())
  DATA: MEAN_wound_healing=col(source(s), name("MEAN_wound_healing"))
  DATA: ID=col(source(s), name("ID"), unit.category())
  GUIDE: axis(dim(1), label("time"))
  GUIDE: axis(dim(2), label("Mean wound_healing"))
  GUIDE: legend(aesthetic(aesthetic.color.interior), label("ID"))
  GUIDE: text.title(label("Multiple Line Mean of wound_healing by time by ID"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: line(position(time*MEAN_wound_healing), color.interior(ID), missing.wings())
END GPL.
```



Also, you can call the **Analyze > Descritptive Statistics > Explore**, and look at the main outcome (wound_healing) while entering ID into the factor list. This will produce a boxplot divided by participant ID. Extreme cases can be spotted this way.

```
EXAMINE VARIABLES=wound_healing BY ID
 /PLOT BOXPLOT.
```

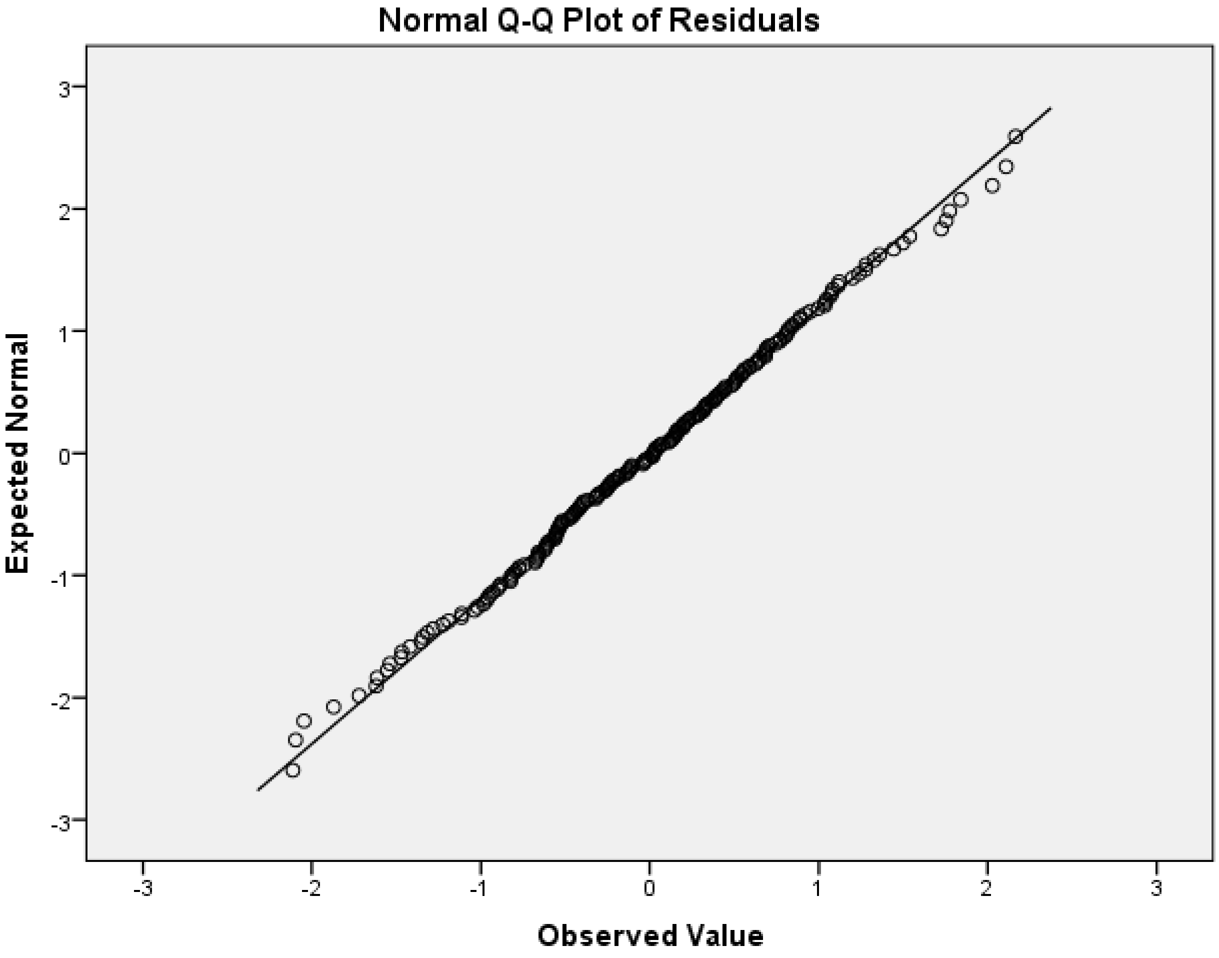


Normality

You can save the residuals in the Save menu when building the mixed effect model. The residuals can be inspected for deviations from normality using **Analyze > Descritptive Statistics > Explore** and looking at the skew, kurtosis, histogram and the QQ plot.

```
MIXED wound_healing WITH distance_window south_wing centered_time
centered_time_sq
  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100)
MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0,
ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=distance_window south_wing centered_time centered_time_sq |
SSTYPE(3)
  /METHOD=REML
  /PRINT=CORB  SOLUTION
  /RANDOM=INTERCEPT | SUBJECT(ID) COVTYPE(VC)
  /SAVE=RESID PRED.
```

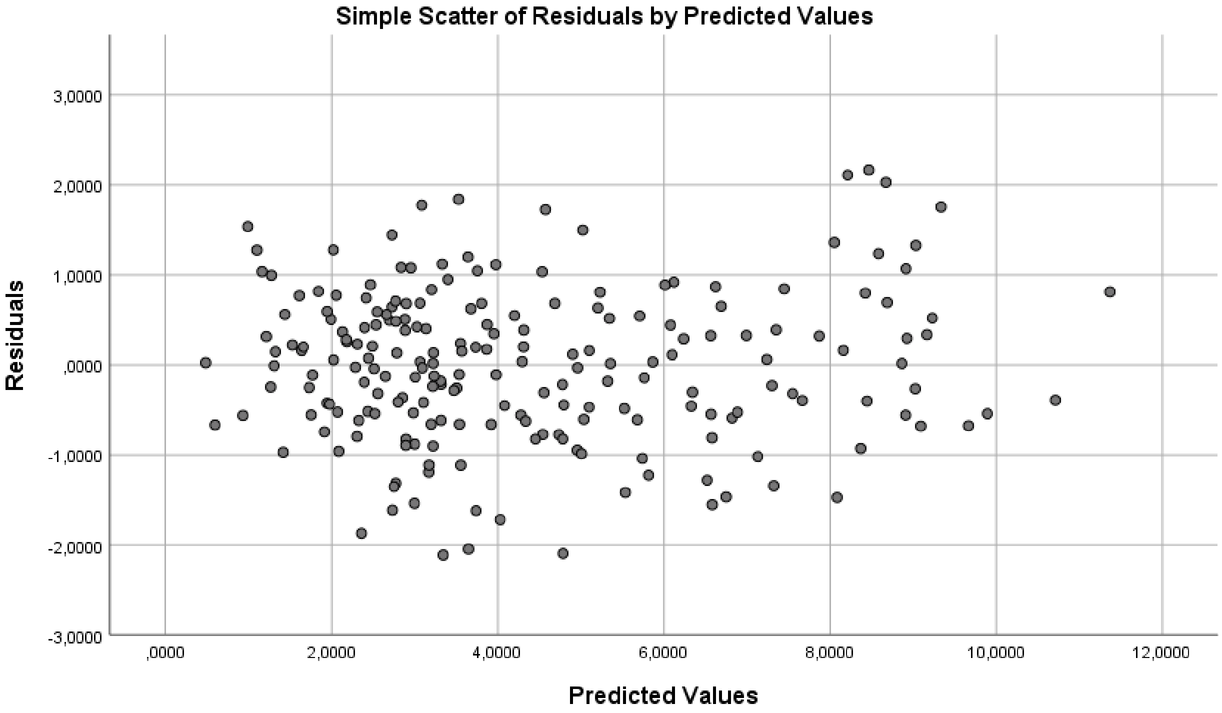
```
EXAMINE VARIABLES=RESID_1
  /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```



Linearity

The linearity of the relationship of the fixed effect predictors and the outcome can be explored by plotting the scatterplot of the residuals and the predicted values.

```
GGRAPH

  /GRAPHDATASET NAME="graphdataset" VARIABLES=PRED_1 RESID_1 MISSING=LISTWISE REPORTMISSING=NO

  /GRAPHSPEC SOURCE=INLINE.

BEGIN GPL

  SOURCE: s=userSource(id("graphdataset"))

  DATA: PRED_1=col(source(s), name("PRED_1"))

  DATA: RESID_1=col(source(s), name("RESID_1"))

  GUIDE: axis(dim(1), label("Predicted Values"))

  GUIDE: axis(dim(2), label("Residuals"))

  ELEMENT: point(position(PRED_1*RESID_1))

END GPL.
```



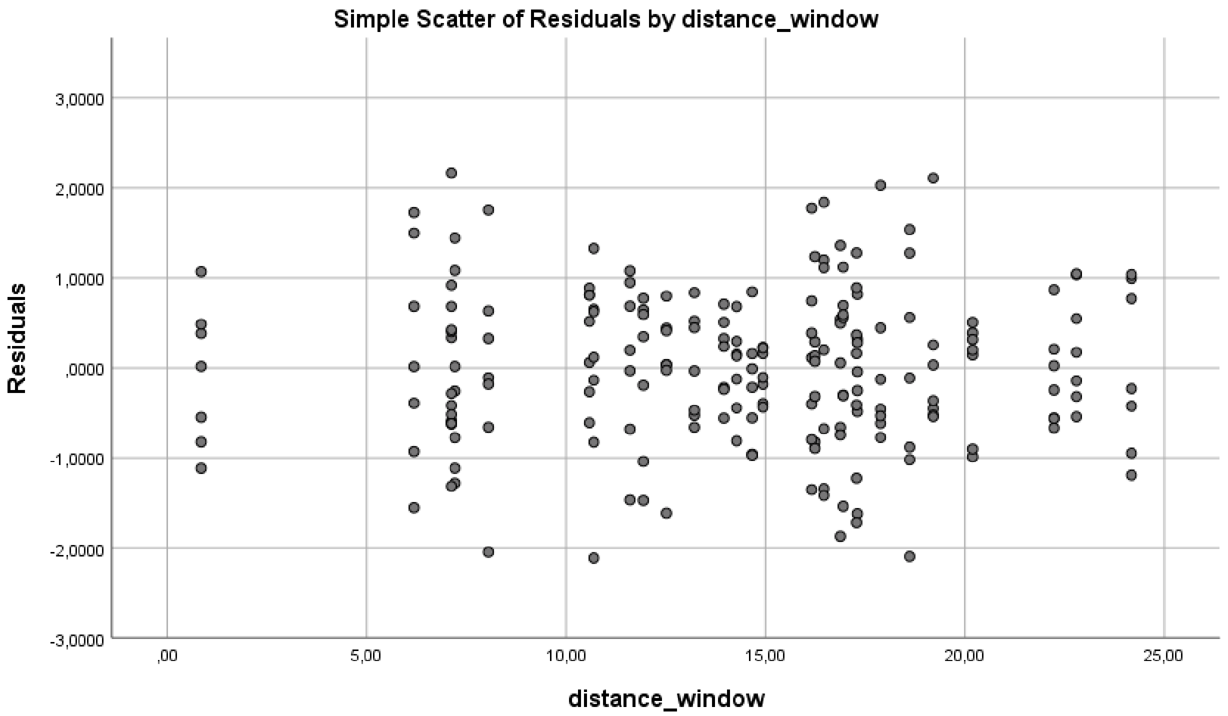
You should also look at the scatterplot of the residuals and the fixed predictors separately.

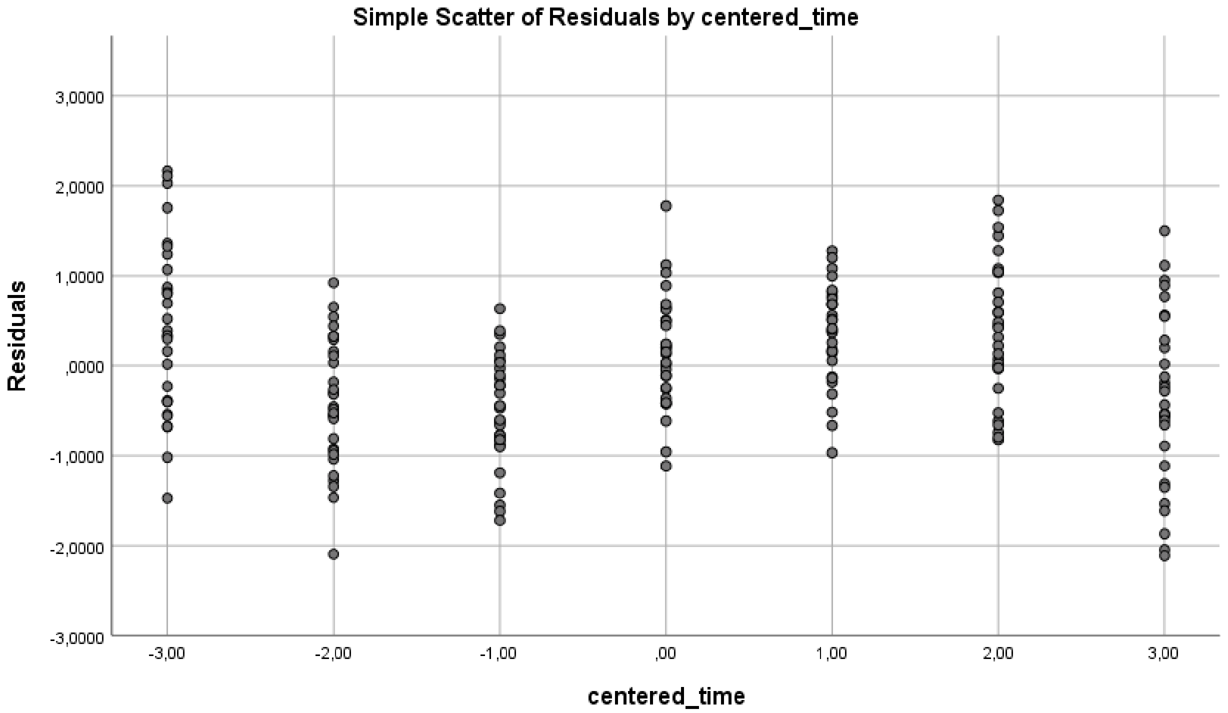
Here you can notice that time_centered still has a nonlinear relationship with the residuals. Maybe we could decrease this by adding the cubic term of time to the model as well. If this was a real paper, this might be done as an exploratory analysis or this might be discussed in the limitations or the future directions for research sections.

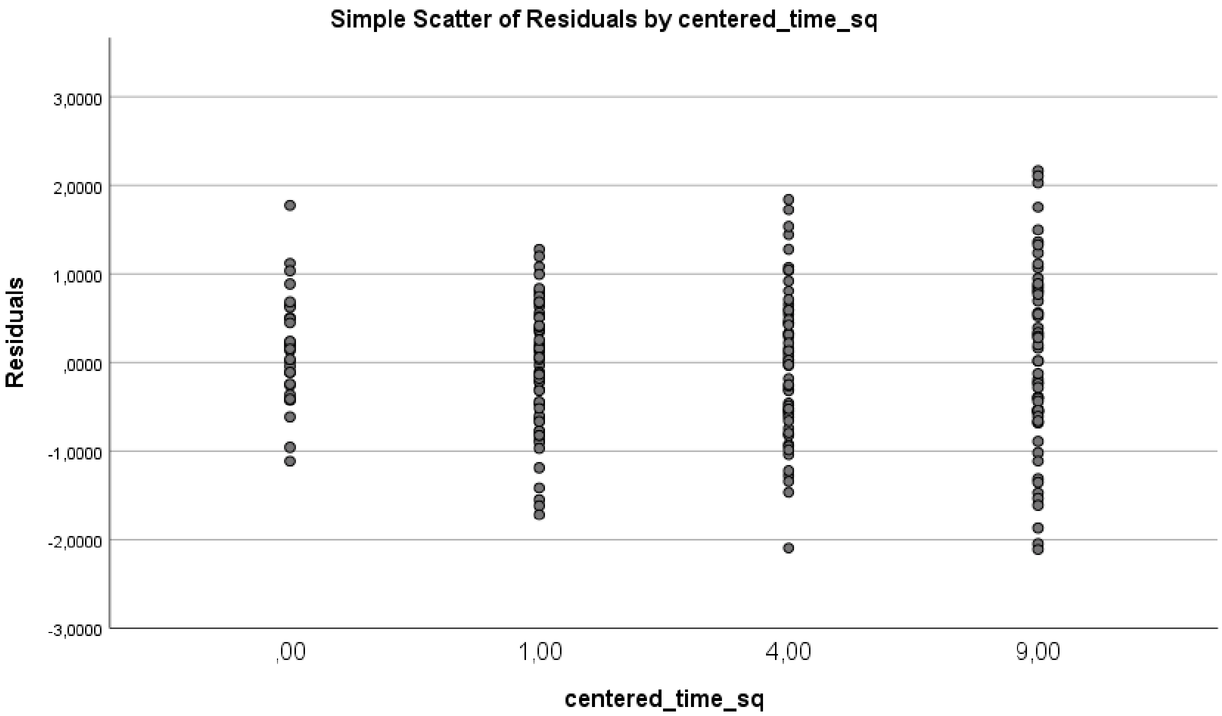
```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=distance_window RESID_1 MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: distance_window=col(source(s), name("distance_window"))
  DATA: RESID_1=col(source(s), name("RESID_1"))
  GUIDE: axis(dim(1), label("distance_window"))
  GUIDE: axis(dim(2), label("Residuals"))
  ELEMENT: point(position(distance_window*RESID_1))
END GPL.
```

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=time_centered RESID_1 MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: time_centered=col(source(s), name("time_centered"))
  DATA: RESID_1=col(source(s), name("RESID_1"))
  GUIDE: axis(dim(1), label("time_centered"))
  GUIDE: axis(dim(2), label("Residuals"))
  ELEMENT: point(position(time_centered*RESID_1))
END GPL.
```

```
GGRAPH
  /GRAPHDATASET NAME="graphdataset" VARIABLES=time_centered_sq RESID_1 MISSING=LISTWISE
    REPORTMISSING=NO
  /GRAPHSPEC SOURCE=INLINE.
BEGIN GPL
  SOURCE: s=userSource(id("graphdataset"))
  DATA: time_centered_sq=col(source(s), name("time_centered_sq"), unit.category())
  DATA: RESID_1=col(source(s), name("RESID_1"))
  GUIDE: axis(dim(1), label("time_centered_sq"))
  GUIDE: axis(dim(2), label("Residuals"))
  SCALE: linear(dim(2), include(0))
  ELEMENT: point(position(time_centered_sq*RESID_1))
END GPL.
```


Simple Scatter of Residuals by distance_window
Residuals
3,0000
2,0000
1,0000
,0000
-1,0000
-2,0000
-3,0000
,00
5,00
10,00
15,00
20,00
25,00
distance_window

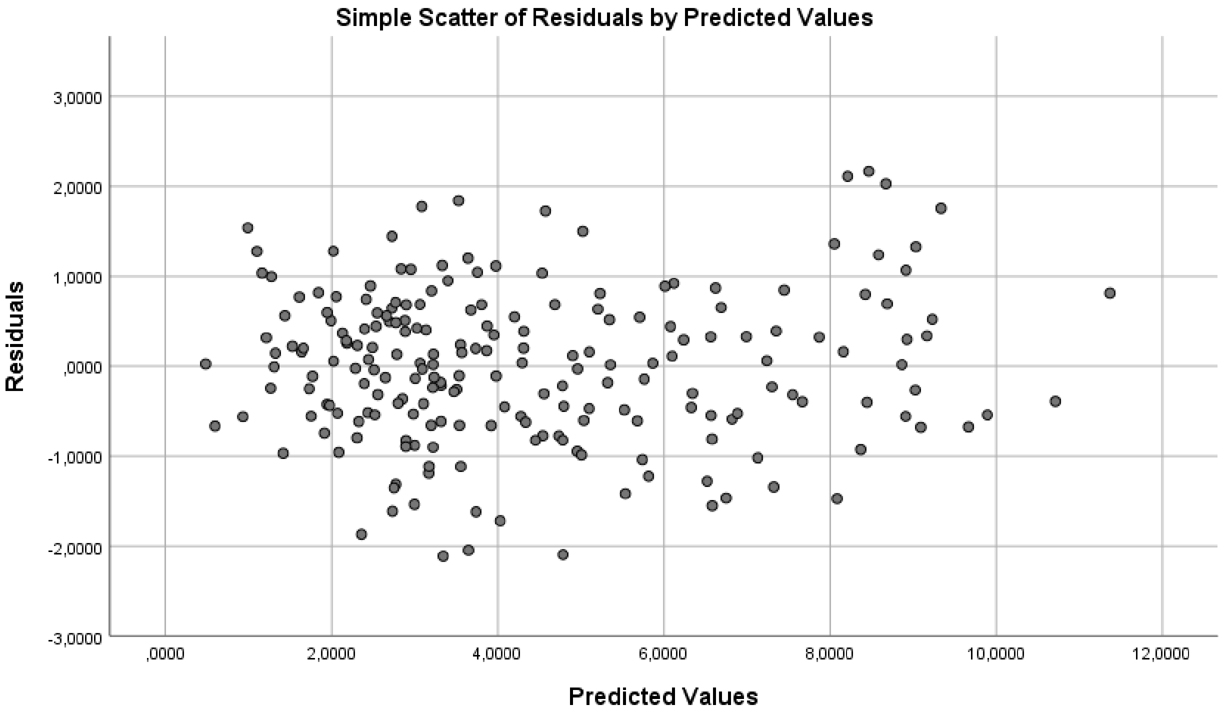

Simple Scatter of Residuals by centered_time
Residuals
3,0000
2,0000
1,0000
,0000
-1,0000
-2,0000
-3,0000
-3,00
-2,00
-1,00
,00
1,00
2,00
3,00
centered_time



Homoscedasticity

The homogeneity of variances on the observation level can be checked by viewing the same standardized residuals ~ predicted values plot as when checking linearity. Here, a funnel shape would indicate heteroscedasticity, but we don't see that in this plot.

```
GGRAPH

  /GRAPHDATASET NAME="graphdataset" VARIABLES=PRED_1 RESID_1
MISSING=LISTWISE REPORTMISSING=NO

  /GRAPHSPEC SOURCE=INLINE.

BEGIN GPL

  SOURCE: s=userSource(id("graphdataset"))

  DATA: PRED_1=col(source(s), name("PRED_1"))

  DATA: RESID_1=col(source(s), name("RESID_1"))

  GUIDE: axis(dim(1), label("Predicted Values"))

  GUIDE: axis(dim(2), label("Residuals"))

  ELEMENT: point(position(PRED_1*RESID_1))

END GPL.
```



Multicollinearity

Finally, we should check for multicollinearity of the fixed effect predictors. Without a well established way to extract the vif from mixed models, we can look at the pariwise correlations of the predictors.

The correlations don’t seem problematic.

```
CORRELATIONS
  /VARIABLES=south_wing time_centered time_centered_sq distance_window
  /PRINT=TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

Correlations

		south_wing	time_centered	time_centered_sq	distance_window
south_wing	Pearson Correlation	1	,000	,000	,049
	Sig. (2-tailed)		1,000	1,000	,479
	N	210	210	210	210
time_centered	Pearson Correlation	,000	1	,000	,000
	Sig. (2-tailed)	1,000		1,000	1,000
	N	210	210	210	210
time_centered_sq	Pearson Correlation	,000	,000	1	,000
	Sig. (2-tailed)	1,000	1,000		1,000
	N	210	210	210	210
distance_window	Pearson Correlation	,049	,000	,000	1
	Sig. (2-tailed)	,479	1,000	1,000	
	N	210	210	210	210

Notice that we see no correlation between time and its quadratic term, because we have centered time.

Correlation Matrix for Estimates of Covariance Parameters[a]

			Intercept + time [subject = ID]		
Parameter		Residual	UN (1,1)	UN (2,1)	UN (2,2)
Residual		1	-,302	,311	-,308
	UN (1,1)	-,302	1	-,848	,576

Intercept + time [subject = ID]	UN (2,1)	,311	-,848	1	-,850
	UN (2,2)	-,308	,576	-,850	1

a. Dependent Variable: wound_healing.

The correlation is .57, which might be used to verify that the Unstructured covariance type is the right covariance type.

If this correlation is small or close to zero, we might consider using another covariance type, such as Variance Components, which assumes no correlation between the random effects.

Constant variance of residuals across clusters

When working with mixed linear models we need to check for homoscedasticity across clusters as well.

We can run a significance test for that by fitting a linear model where we predict the squared residuals with the clustering variable (ID). Check the complete model F-test p-value. If it is < 0.05, heteroscedasticity on the cluster level might be problematic. To do this, you will have to dummy code ID. There are many levels for ID, so the fastest way to do this is to use SPSS's in-built function: **Transform > Create dummy variables.** Here, we need to specify that we want to create dummies from ID, that we want main effect dummies, and give a root name, after which SPSS will put a number to indicate, which level's dummy that particular variable is. I gave "ID_dummy" as a root name.

```
SPSSINC CREATE DUMMIES VARIABLE=ID

ROOTNAME1=ID_dummy

/OPTIONS ORDER=A USEVALUELABELS=YES USEML=YES OMITFIRST=NO.
```

We can also compute the square of the residuals in the usual way.

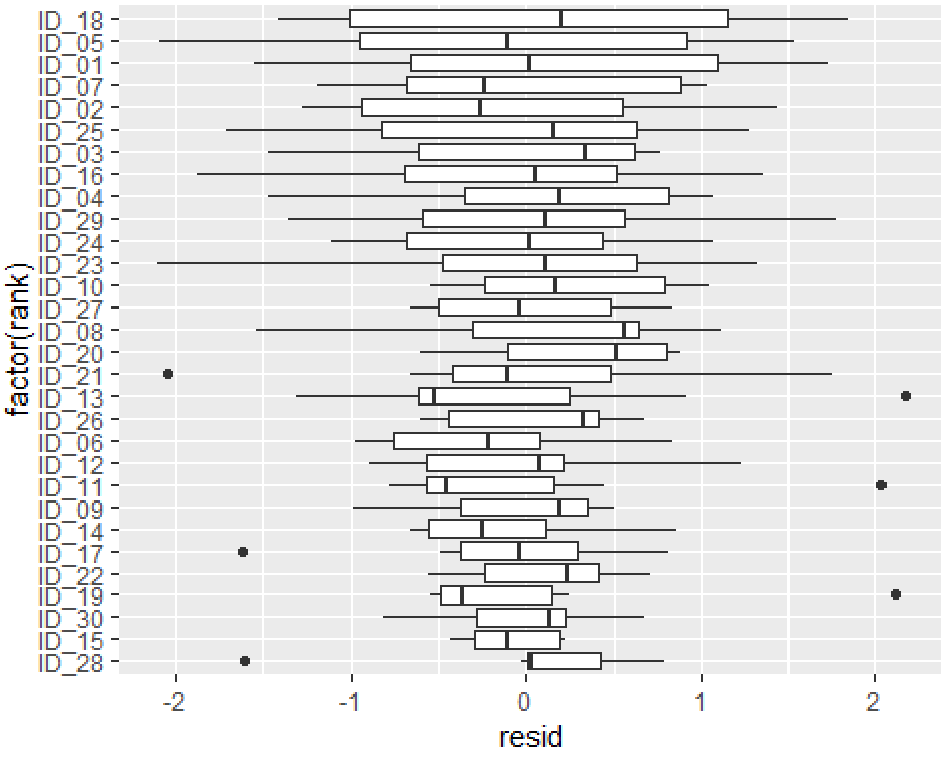
```
COMPUTE RESID_1_sq=RESID_1*RESID_1.

EXECUTE.
```

Finally, fit a regression model. Remember, that it is enough to give all but one of the dummy variables, because one of them will be the baseline level, so I did not include ID_dummy_1 as a predictor, but all other dummies.

```
REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT RESID_1_sq
 /METHOD=ENTER ID_dummy_2 ID_dummy_3 ID_dummy_4 ID_dummy_5
ID_dummy_6 ID_dummy_7 ID_dummy_8
  ID_dummy_9 ID_dummy_10 ID_dummy_11 ID_dummy_12 ID_dummy_13
ID_dummy_14 ID_dummy_15 ID_dummy_16
  ID_dummy_17 ID_dummy_18 ID_dummy_19 ID_dummy_20 ID_dummy_21
ID_dummy_22 ID_dummy_23 ID_dummy_24
  ID_dummy_25 ID_dummy_26 ID_dummy_27 ID_dummy_28 ID_dummy_29
ID_dummy_30.
```

(Another method used to investigate homoscedasticity is to inspect the cyclone plot. Here we plot the boxplot of the residuals for each participant, and order these boxes according to the interquartile range of the residual. Here we would expect a gradual increase of the variance of the residual from top to buttom instead of a sudden increase. If the increase is not consistent (some clusters have much larger variance than the previous one on the list), we can suspect heteroscedasticity across clusters/units. However, there is no easy way to do this in SPSS, you may be able to do it by using excel to calculate IQRs and order the IDs that way, then use SPSS for plotting.

The following figure was created in R)



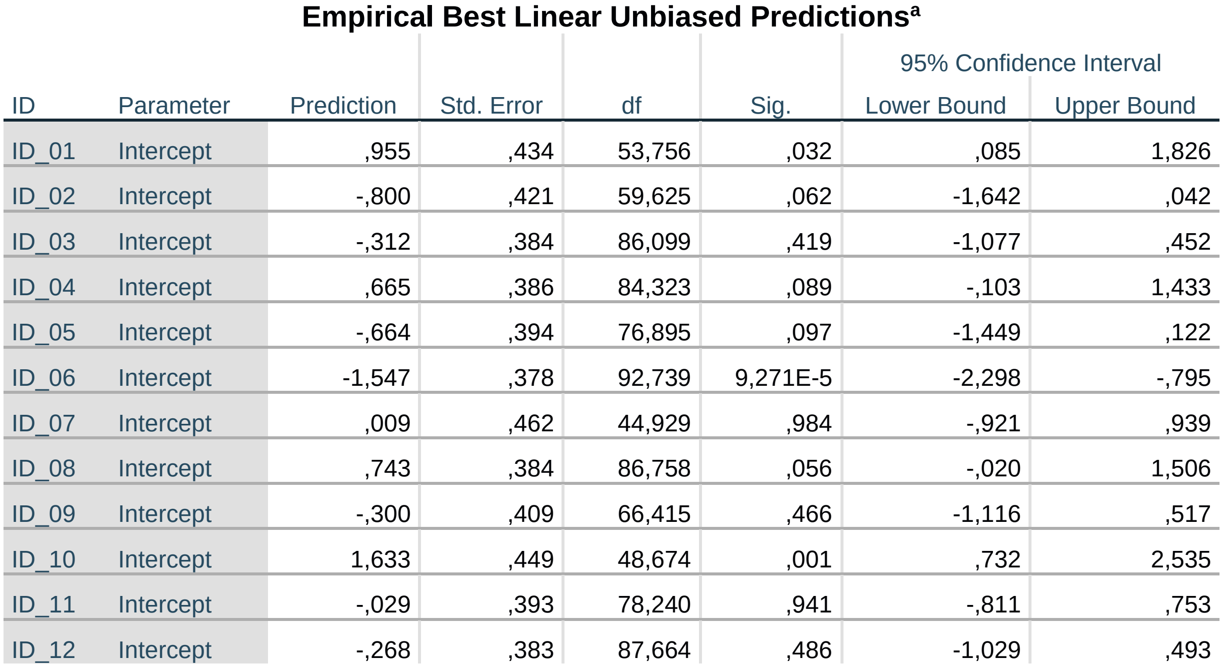
Normal distribution of the random effects

With mixed models we also assume that the random effects are normally distributed around 0. Starting with SPSS 25 it is possible to extract the random effects form the mixed procedure by editing the syntax.

When the specification of the model in the Mixed procedure is done, paste into syntax. In the syntax code add SOLUTION at the end of the line starting with RANDOM. In our case, the syntax for the random intercept model looks like this:

```
MIXED wound_healing WITH distance_window south_wing time_centered
time_centered_sq

  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100)
MXSTEP(10) SCORING(1)

    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0,
ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

  /FIXED=distance_window south_wing time_centered time_centered_sq |
SSTYPE(3)

  /METHOD=REML

  /PRINT=CORB  SOLUTION

  /RANDOM=INTERCEPT | SUBJECT(ID) COVTYPE(VC) SOLUTION

  /SAVE=RESID PRED.
```

In the output, this command creates a new table called Empirical Best Linear Unbiased Predictions. This table contains information about the random effects for each cluster. We need to copy the information contained in the “prediction” column to a new datafile in SPSS to be able to test the assumption of normality of random effects. This can be done by double clicking on the table and simply selecting the numbers in that column and **copying and pasting them into a new empty datafile** in SPSS. Alternatively, this table can be exported using SPSS’s **File > Export** feature into an excel file, and data can be copied from there.



Empirical Best Linear Unbiased Predictions[a]

						95% Confidence Interval	
ID	Parameter	Prediction	Std. Error	df	Sig.	Lower Bound	Upper Bound
ID_01	Intercept	,955	,434	53,756	,032	,085	1,826
ID_02	Intercept	-,800	,421	59,625	,062	-1,642	,042
ID_03	Intercept	-,312	,384	86,099	,419	-1,077	,452
ID_04	Intercept	,665	,386	84,323	,089	-,103	1,433
ID_05	Intercept	-,664	,394	76,895	,097	-1,449	,122
ID_06	Intercept	-1,547	,378	92,739	9,271E-5	-2,298	-,795
ID_07	Intercept	,009	,462	44,929	,984	-,921	,939
ID_08	Intercept	,743	,384	86,758	,056	-,020	1,506
ID_09	Intercept	-,300	,409	66,415	,466	-1,116	,517
ID_10	Intercept	1,633	,449	48,674	,001	,732	2,535
ID_11	Intercept	-,029	,393	78,240	,941	-,811	,753
ID_12	Intercept	-,268	,383	87,664	,486	-1,029	,493

ID_13	Intercept	-,365	,429	56,050	,398	-1,224	,494
ID_14	Intercept	-,844	,433	54,118	,057	-1,713	,024
ID_15	Intercept	-,459	,378	93,067	,228	-1,210	,292
ID_16	Intercept	-,742	,386	84,363	,058	-1,510	,025
ID_17	Intercept	-,885	,389	81,864	,025	-1,659	-,112
ID_18	Intercept	,837	,384	86,544	,032	,073	1,600
ID_19	Intercept	,471	,399	72,989	,242	-,325	1,267
ID_20	Intercept	2,009	,389	81,412	1,697E-6	1,235	2,783
ID_21	Intercept	-,252	,411	64,874	,542	-1,074	,570
ID_22	Intercept	,695	,378	92,929	,069	-,056	1,446
ID_23	Intercept	-,318	,389	82,077	,416	-1,091	,455
ID_24	Intercept	-,482	,532	32,955	,372	-1,564	,600
ID_25	Intercept	-,597	,389	81,985	,128	-1,370	,176
ID_26	Intercept	,335	,429	56,050	,438	-,524	1,194
ID_27	Intercept	,107	,379	92,596	,778	-,645	,859
ID_28	Intercept	,080	,382	88,841	,835	-,679	,838
ID_29	Intercept	,427	,381	90,184	,265	-,329	1,184
ID_30	Intercept	-,103	,378	93,183	,786	-,854	,648

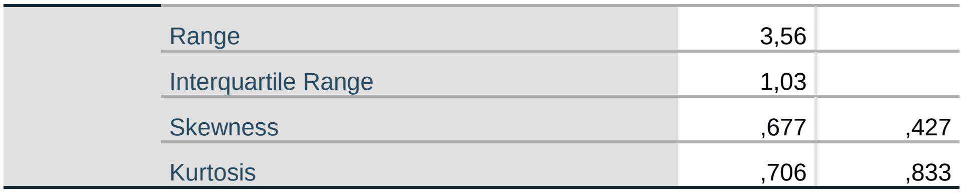
a. Dependent Variable: wound_healing

Now we can simply run Analyse > Descriptives > Explore and ask for normality plots and statistics to get information about the skew, kurtosis and the deviation from normal distribution.

In our example the random effects indeed seem to be distributed normally around 0, since the mean is 0 and the skew and kurtosis are within -1 and 1.

Descriptives

			Statistic	Std. Error
VAR00001	Mean		,0000	,14067
	95% Confidence Interval for Mean	Lower Bound	-,2877	
		Upper Bound	,2877	
	5% Trimmed Mean		-,0310	
	Median		-,1775	
	Variance		,594	
	Std. Deviation		,77047	
	Minimum		-1,55	
	Maximum		2,01	



	Range	3,56	
	Interquartile Range	1,03	
	Skewness	,677	,427
	Kurtosis	,706	,833

Dependency structure of the random effects

We assume that we have modeled the dependency (covariance) structure of the random effects correctly. This assumption is only important if we have multiple random effects. If we only have a single random intercept in our model, this is not an issue, since it has no correlation with other random effects. But if we have at least two random effects in the model (for example random intercept and the random slope of time), we need to think about how these are related to each other. If the random effects are potentially correlated, we often use the Unstructured covariance type to allow this correlation in the model: COVTYPE(UN) .

We can check whether the random effects are correlated in a random slope model if we ask for the "Correlation of Parameter Estimates" in the Statistics Menu in the Mixed model procedure.

```
MIXED wound_healing WITH distance_window south_wing time

 /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100)
MXSTEP(10) SCORING(1)

   SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0,
ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)

 /FIXED=distance_window south_wing time | SSTYPE(3)

 /METHOD=REML

 /PRINT=CORB  SOLUTION

 /RANDOM=INTERCEPT time | SUBJECT(ID) COVTYPE(UN).
```

Once we get the output, we should look for the Correlation Matrix for Estimates of Covariance Parameters table. In this table we will find data about the correlation of the random effects in the study. We should look at the correlation coefficient of UN(1,1), representing the random intercept, and UN(2,2), representing the random slope of time in this example.

Other resources

This document cannot be considered a complete guide for model diagnistics for mixed models.

Model diagnostics of mixed models is quite complex and methods and tools are still in development. The field has not caught up to these developments yet, and there is no real consensus on how to do model diagnostics on these models other than some basics mentioned above.

The sources I used to compile the above guide:

http://ademos.people.uic.edu/Chapter18.html

https://www.ssc.wisc.edu/sscc/pubs/MM/MM_DiagInfer.html

Loy, A., Hofmann, H., & Cook, D. (2017). Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners. Journal of Computational and Graphical Statistics, 26(3), 478-492.

Some more redings that can help those who are interested in the cutting edge on this topic:

http://thestatsgeek.com/2014/08/17/robustness-of-linear-mixed-models/

https://stat.ethz.ch/pipermail/r-sig-mixed-models/2014q2/022160.html

Loy, A., Hofmann, H., & Cook, D. (2017). Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners. Journal of Computational and Graphical Statistics, 26(3), 478-492.

For an alternative "consensus-based" approach on model diagnostics using plot lineups see Loy, Hofmann and Cook (2017).

Loy, A., Hofmann, H., & Cook, D. (2017). Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners. Journal of Computational and Graphical Statistics, 26(3), 478-492.