MEDB 5502, Module 12, Longitudinal data

Topics to be covered

- What you will learn
  - Mathematical formulation of random intercepts model
  - Description of HIV-intervention data
  - Random intercepts model using hiv-intervention data
  - Mathematical formulation of random slopes model
  - Assumptions and complications

Longitudinal data

- Measurements taken at different times
  - Emphasis in changes over time

Speaker notes

In the previous module, I talked about hierarchical models and mentioned a particular case, longitudinal data, that I want to talk more about in this presentation.

Longitudinal data is similar to repeated measures data. With both, you measure the same subject repeatedly. With longitudinal data, often the emphasis is in changes that occur over time. Repeated measurements, in contrast, emphasize different treatments with the hope that the time gaps between the measurements are small enough that you don’t see changes over time.

The differences between longitudinal data, repeated measures data, or hierarchical data are subtle. Perhaps these are distinctions without a difference. I decided to separate out longitudinal data for a different module perhaps more out of the desire to split a complex topic into smaller bite-sized pieces.

Random intercepts model, 1 of 3

- Simplest pattern for longitudinal data
- $Y_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, k$
  - n subjects, k time points
- t_j, time of jth measurement
  - First time is often zero

Speaker notes

The simplest longitudinal model has n subjects and k time points. The first time point is often set to zero. The times are often evenly spaced, but they don't have to be.

Random intercepts model, 2 of 3

- $Y_{ij} = \beta_0 + u_{0i} + \beta_1 t_j + \epsilon_{ij}$
  - β_0 and β_1 are unknown constants
  - u_{0i} and ϵ_{ij} are normally distributed
    - $SD(u_{0i}) = \sigma_{intercept}$
    - $SD(\epsilon_{ij}) = \sigma_{error}$

Speaker notes

There are two sources of random variation in the random intercepts model, u_{0i} and ϵ_{ij}.

Random intercepts model, 3 of 3

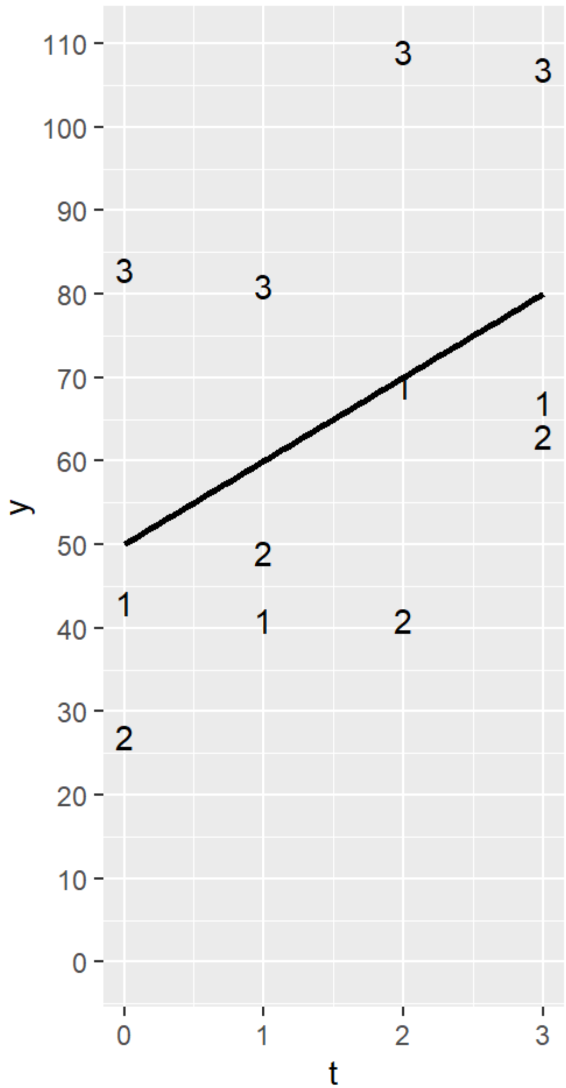
- $SD(Y_{ij}) = \sqrt{\sigma^2_{intercept} \ + \ \sigma^2_{error}}$
- $Corr(Y_{ij}, Y_{im}) = \frac{\sigma^2_{intercept}}{\sigma^2_{intercept} + \sigma^2_{error}}$

Speaker notes

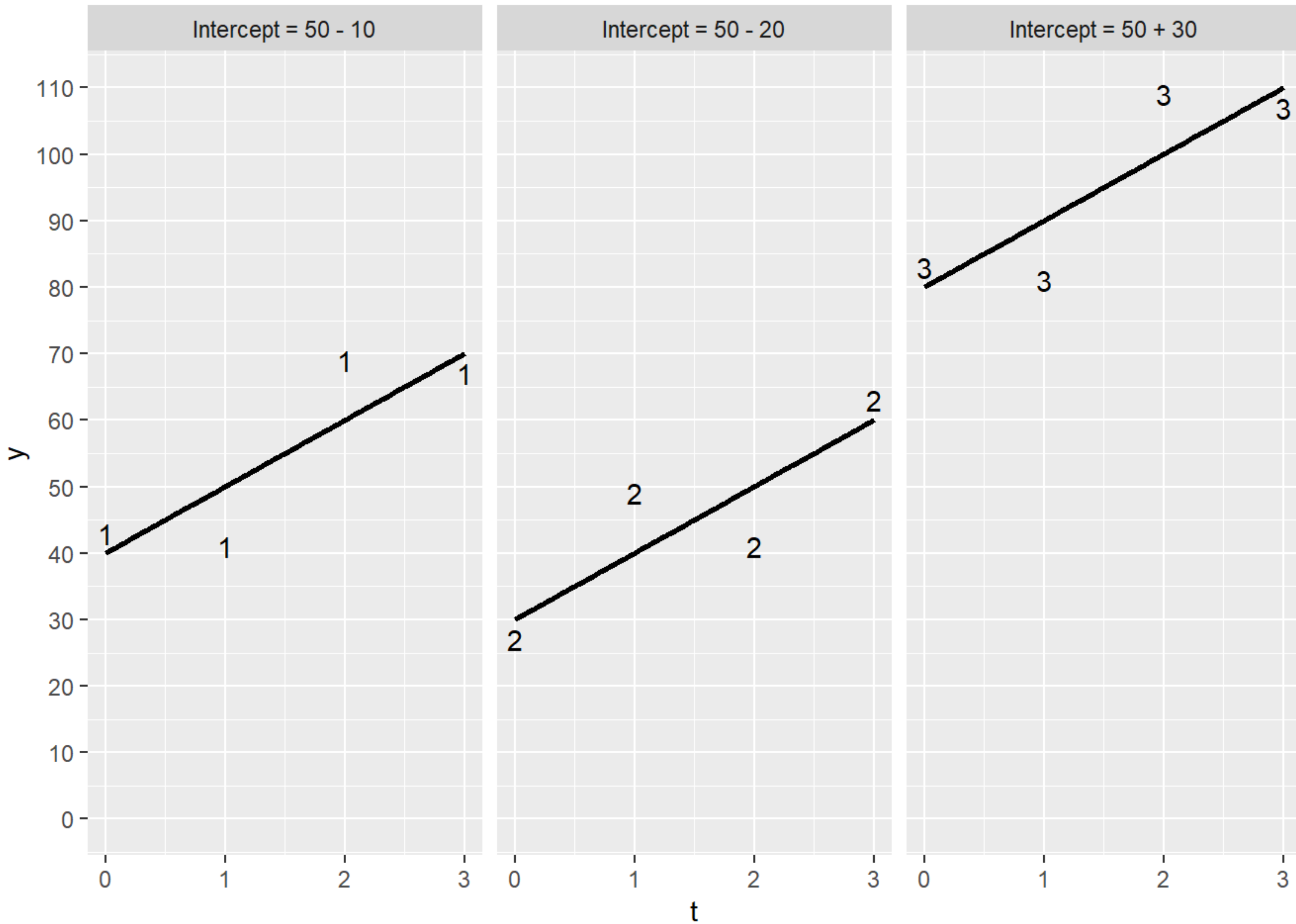
The standard deviation for any individual observation combines the standard deviation for the random intercepts and the standard deviation for the error terms. They combine in a Pythagorean way.

The correlation of two measurements on the same patient is comparable to a measure we defined in the last module, the intraclass correlation.

Random intercepts illustrated, 1 of 2



Random intercepts illustrated, 2 of 2



Break #1

- What you have learned
  - Mathematical formulation of random intercepts model
- What's coming next
  - Description of HIV-intervention data

Description of hiv-intervention data, 1 of 2

```
data_dictionary: hiv-intervention.txt

source: OzDASL website

description: |
  This is a longitudinal study of an intervention in 14-18 adolescents
intended to increase the frequency of condom protected sex. Subjects  were
allocated randomly to treatment or control groups. All were evaluated prior to
the intervention, immediately  after the intervention, 6 months and  12 months
after the intervention.The outcome variable is the logarithm-transformed
frequency of condom-protected sex ( log(Y+1) )."
```

Speaker notes

Here is a dataset I will use to illustrate the random intercepts model. It actually might require a more sophisticated model than the random intercepts, but it is always a good idea to start with the simplest model, even if you know it is an oversimplification. Slowly add layers of complexity, and don’t fit the final model too early. You want to wade in from the shallow end of the pool rather than jump right away into the deep end.

Description of hiv-intervention data, 2 of 2


```
BST:
  label: treatment group
  values:
    '1': BST intervention
    '0': control
Pre:
  label: Log-frequency of protected sex before the intervention
Post:
  label: Log-frequency of protected sex after the intervention
FU6:
  label: Log-frequency of protected sex reported at the 6 months follow-up
FU12:
  label: Log-frequency of protected sex reported at the 12 months follow-up
```

Speaker notes

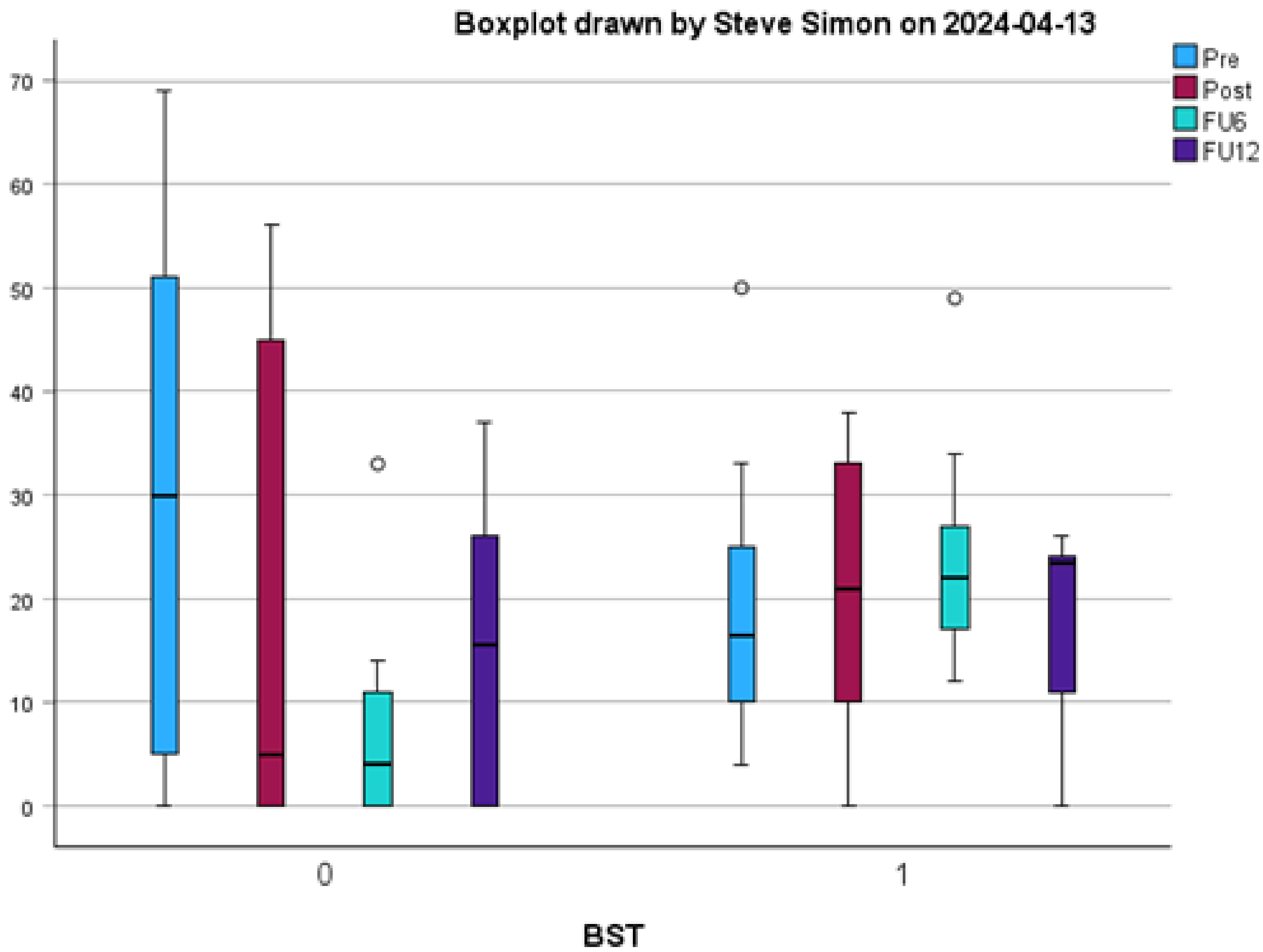
Here are the first three variables. The actual times associated with the Pre and Post measurements are unclear. It turns out that it will be best to hold the Pre measurement back for the time being and start the clock at time=0 for the Post measurement. Remember that you are wading in from the shallow end of the pool.

The remaining two variables FU6 and FU12 represent time=6 and time=12, respectively.

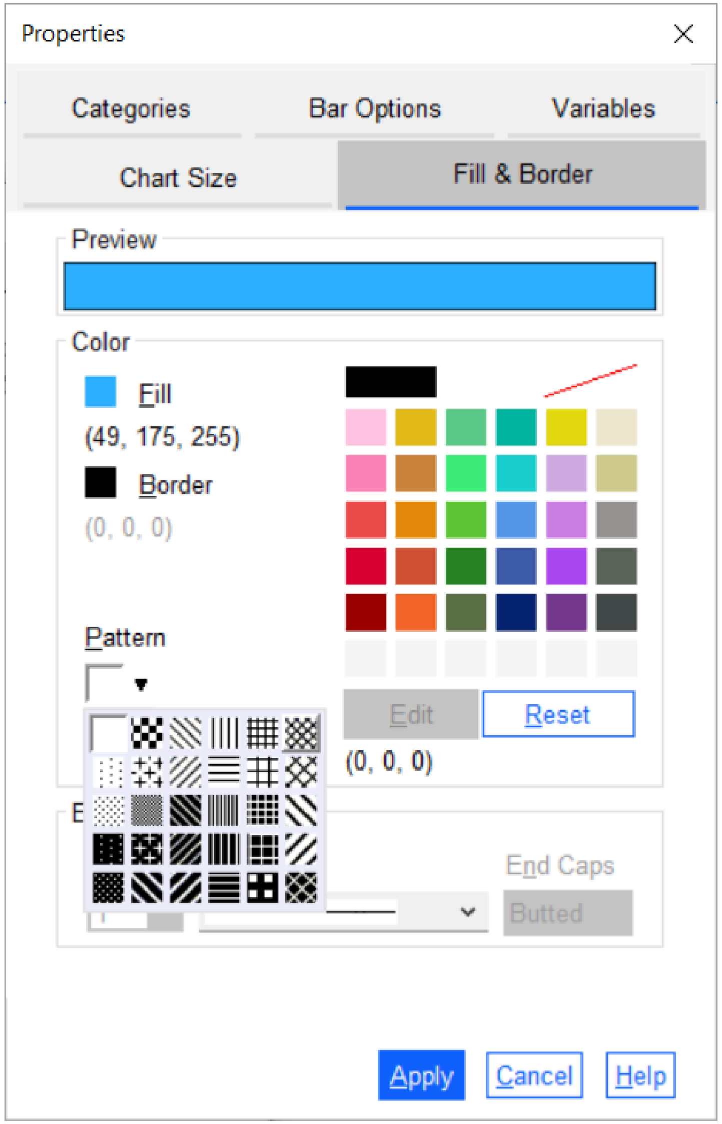
Wide format

	BST	Pre	Post	FU6	FU12
1	1	7	22	13	14
2	1	25	10	17	24
3	1	50	36	49	23
4	1	16	38	34	24
5	1	33	25	24	25
6	1	10	7	23	26
7	1	13	33	27	24
8	1	22	20	21	11
9	1	4	0	12	0
10	1	17	16	20	10
11	0	0	0	0	0
12	0	69	56	14	36
13	0	5	0	0	5

Boxplots


Boxplot drawn by Steve Simon on 2024-04-13
Pre
Post
FU6
FU12
70
60
50
40
30
20
10
0
0
1
BST

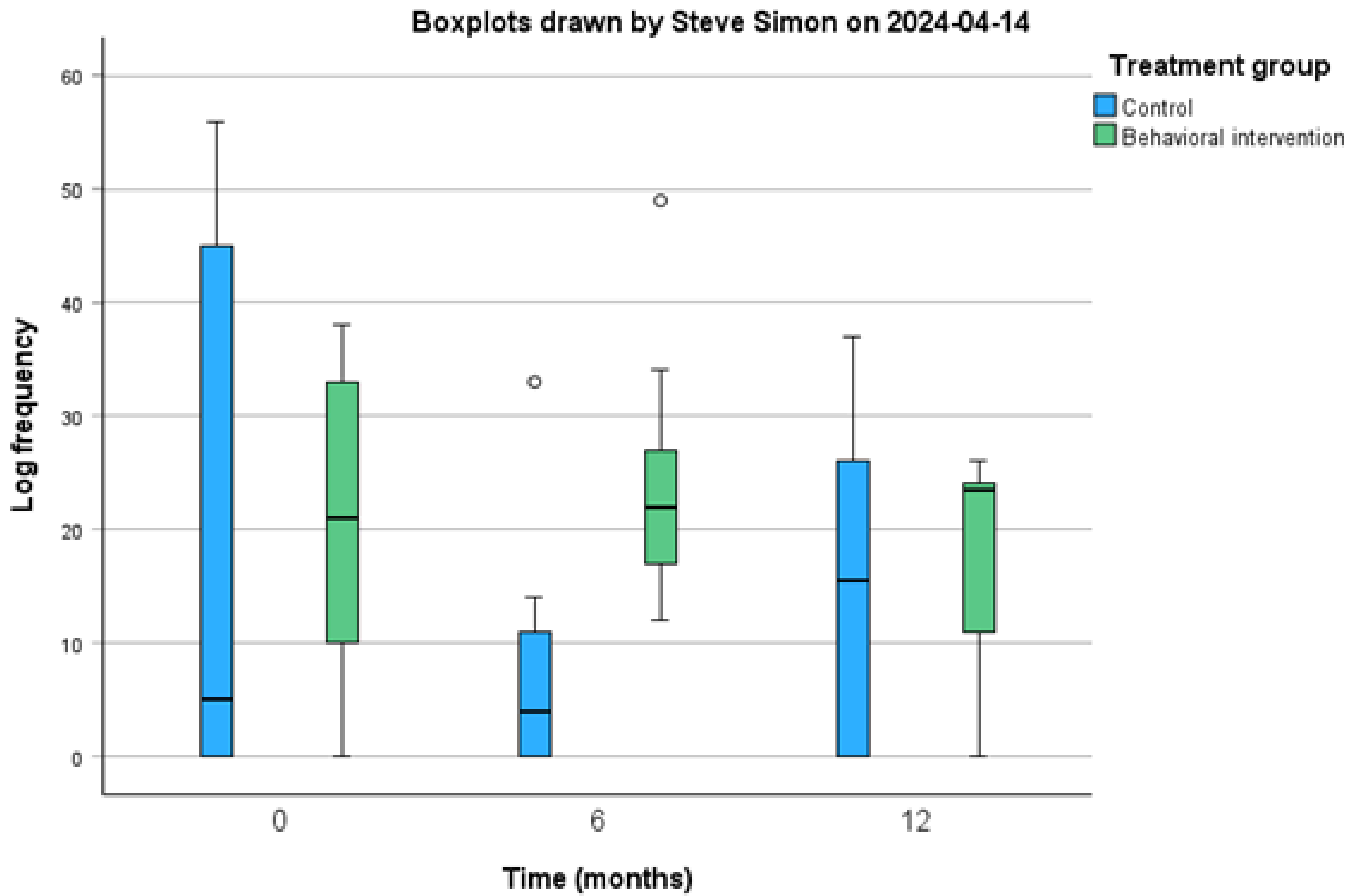
Colors and patterns



Tall format

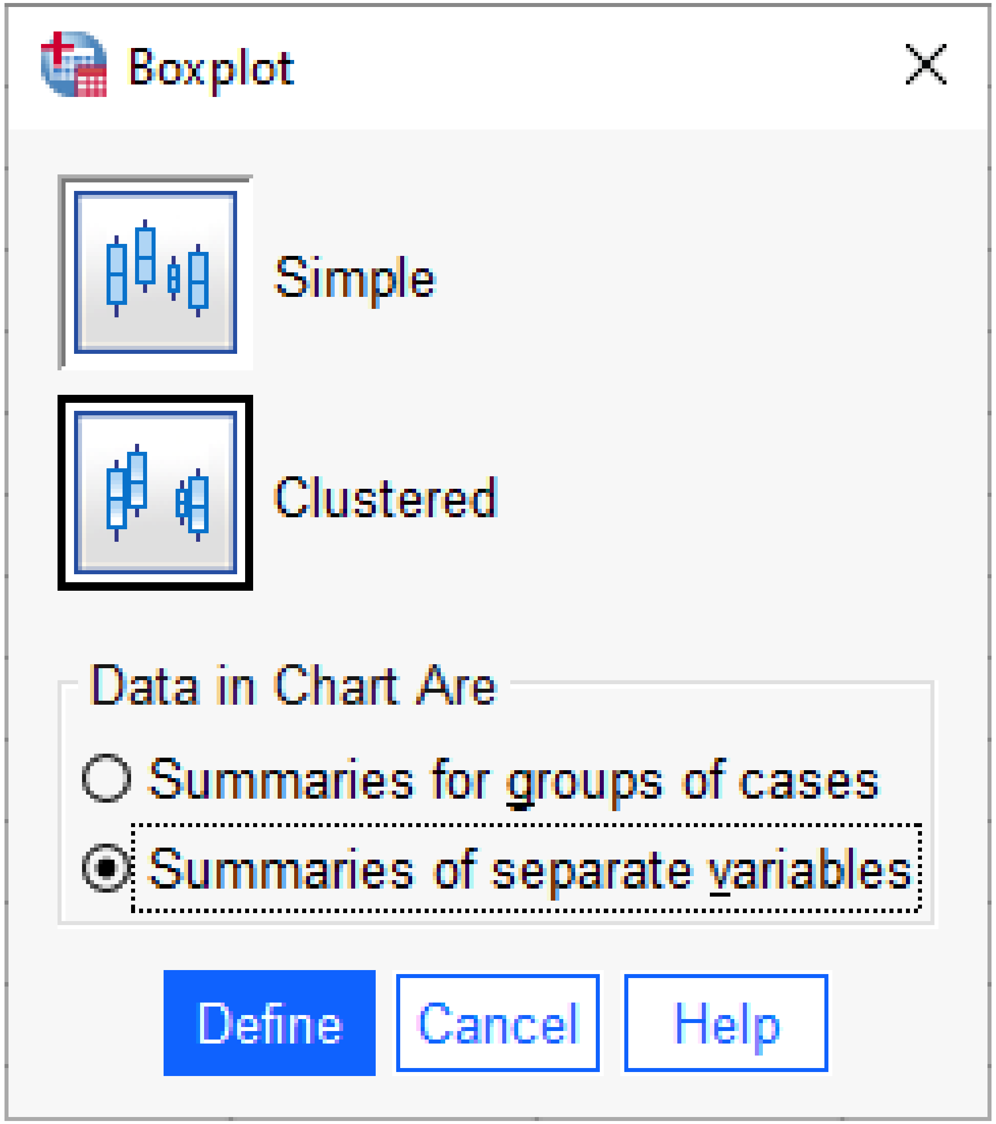
	id	BST	Pre	Index1	log_frequency	time
1	1	1	7	Post	22	0
2	1	1	7	FU6	13	6
3	1	1	7	FU12	14	12
4	2	1	25	Post	10	0
5	2	1	25	FU6	17	6
6	2	1	25	FU12	24	12
7	3	1	50	Post	36	0
8	3	1	50	FU6	49	6
9	3	1	50	FU12	23	12
10	4	1	16	Post	38	0
11	4	1	16	FU6	34	6
12	4	1	16	FU12	24	12

Alternate clustering of boxplots

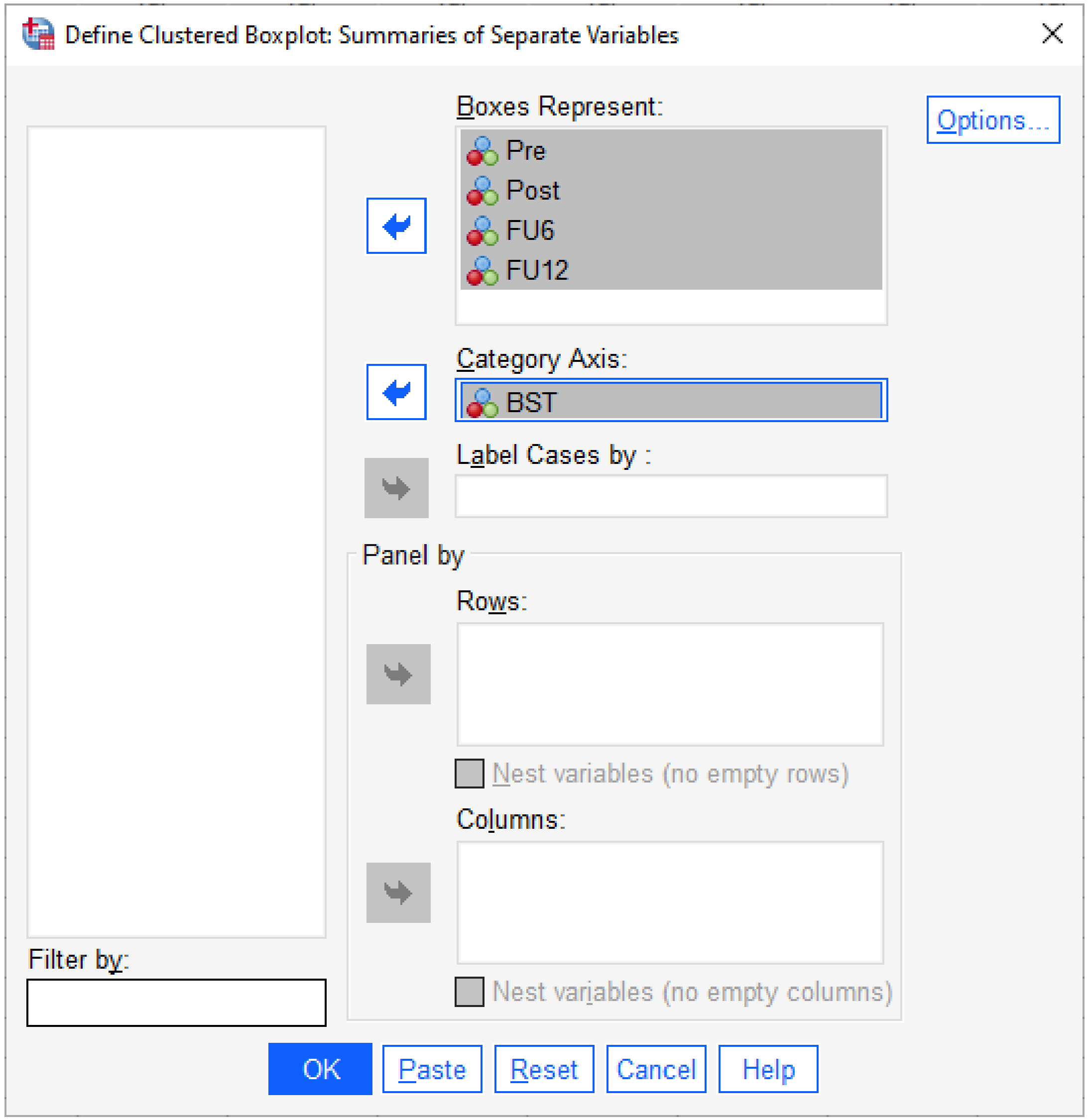


Live demo, restructuring and boxplots

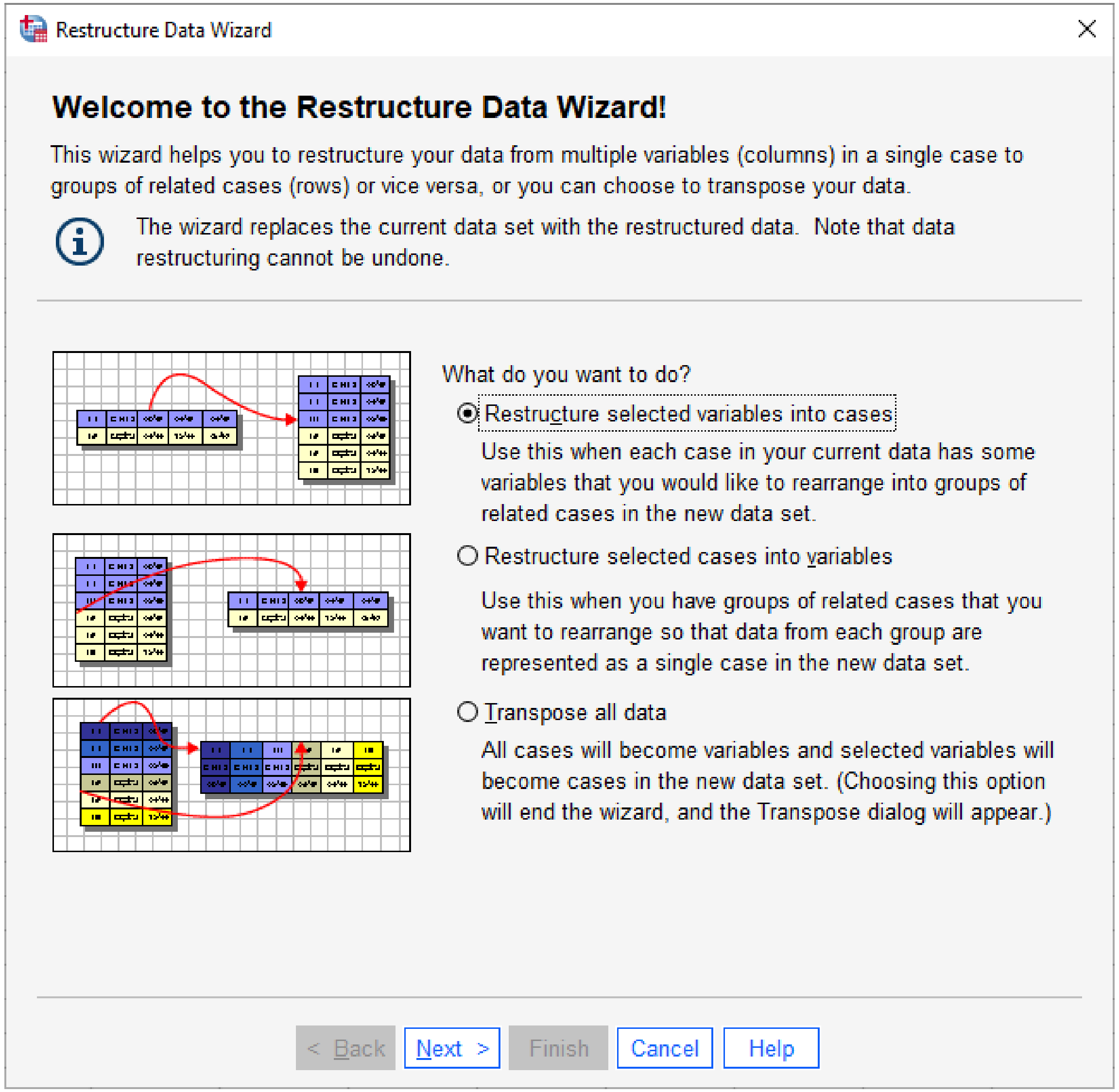
After menu Graph | Boxplot



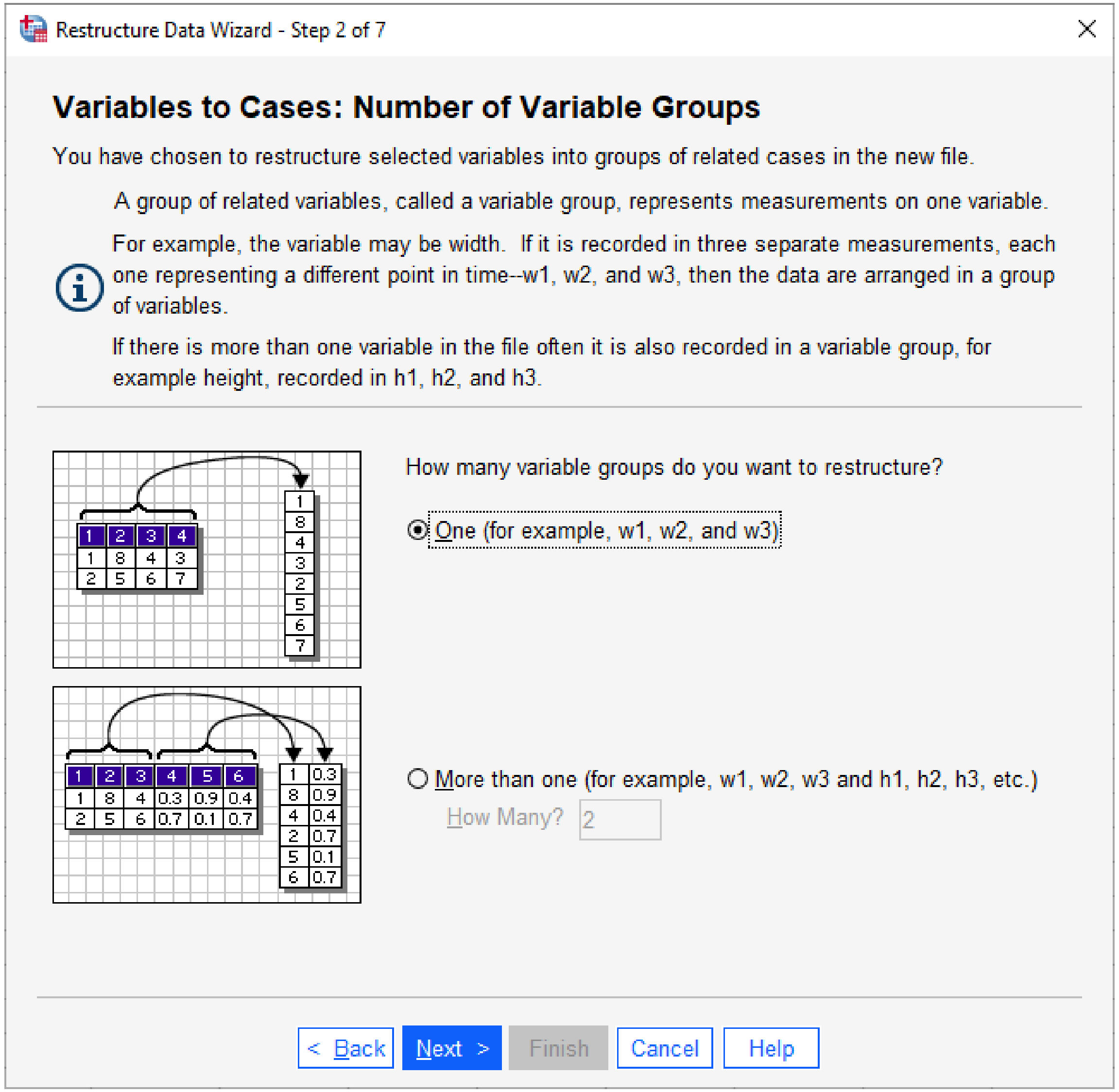
After button Define



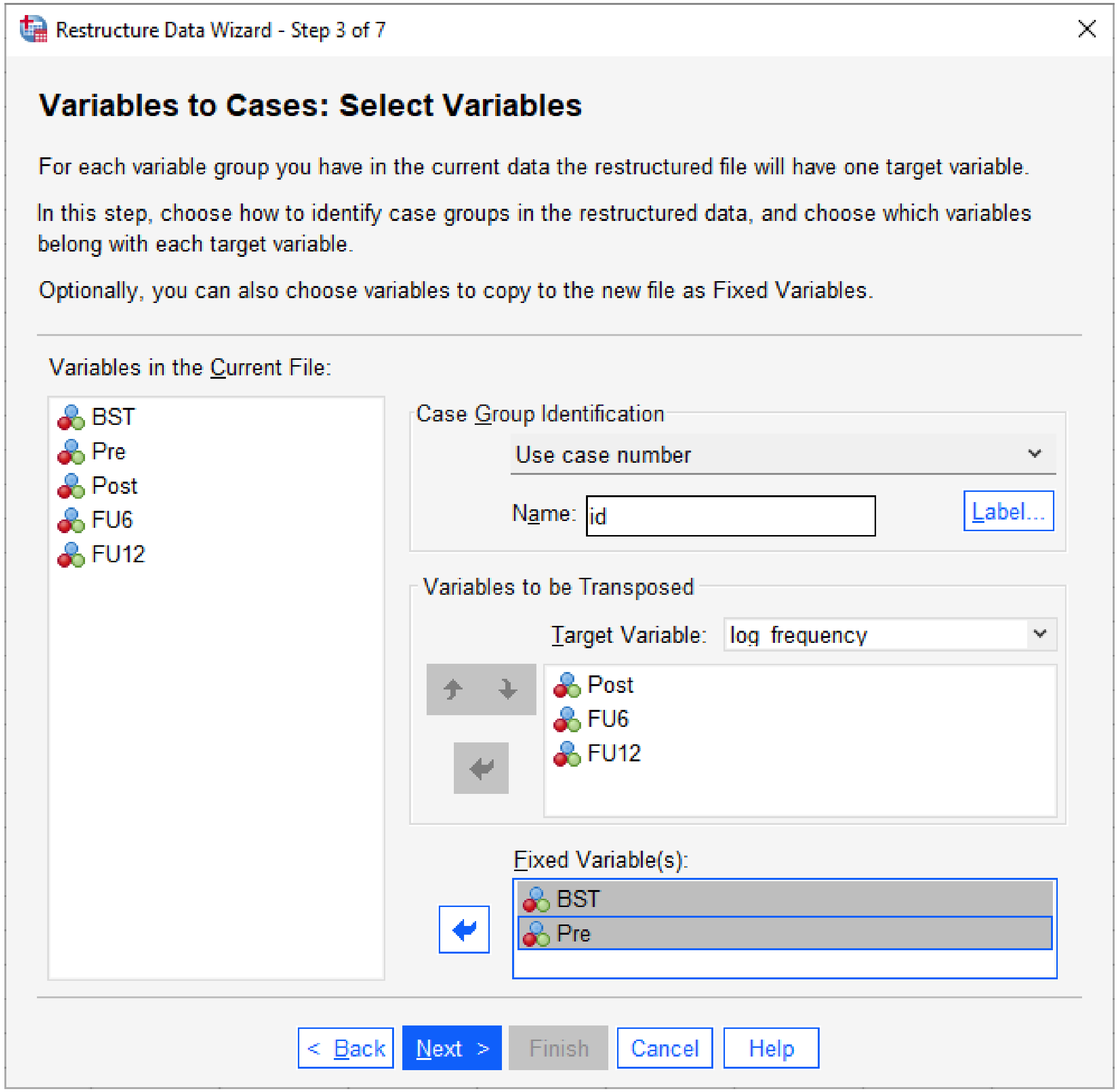
After menu Data | Restructure



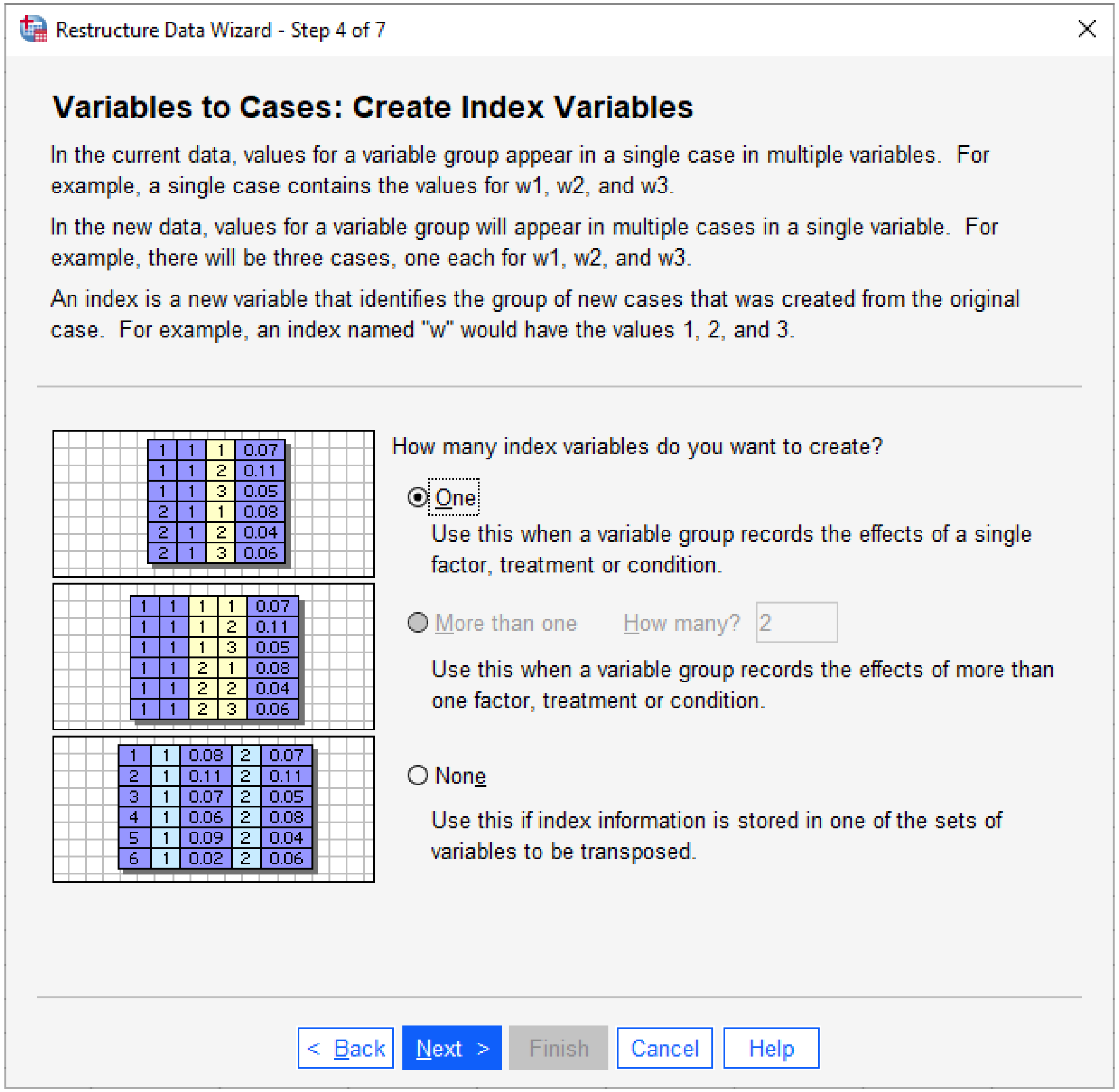
After button Next



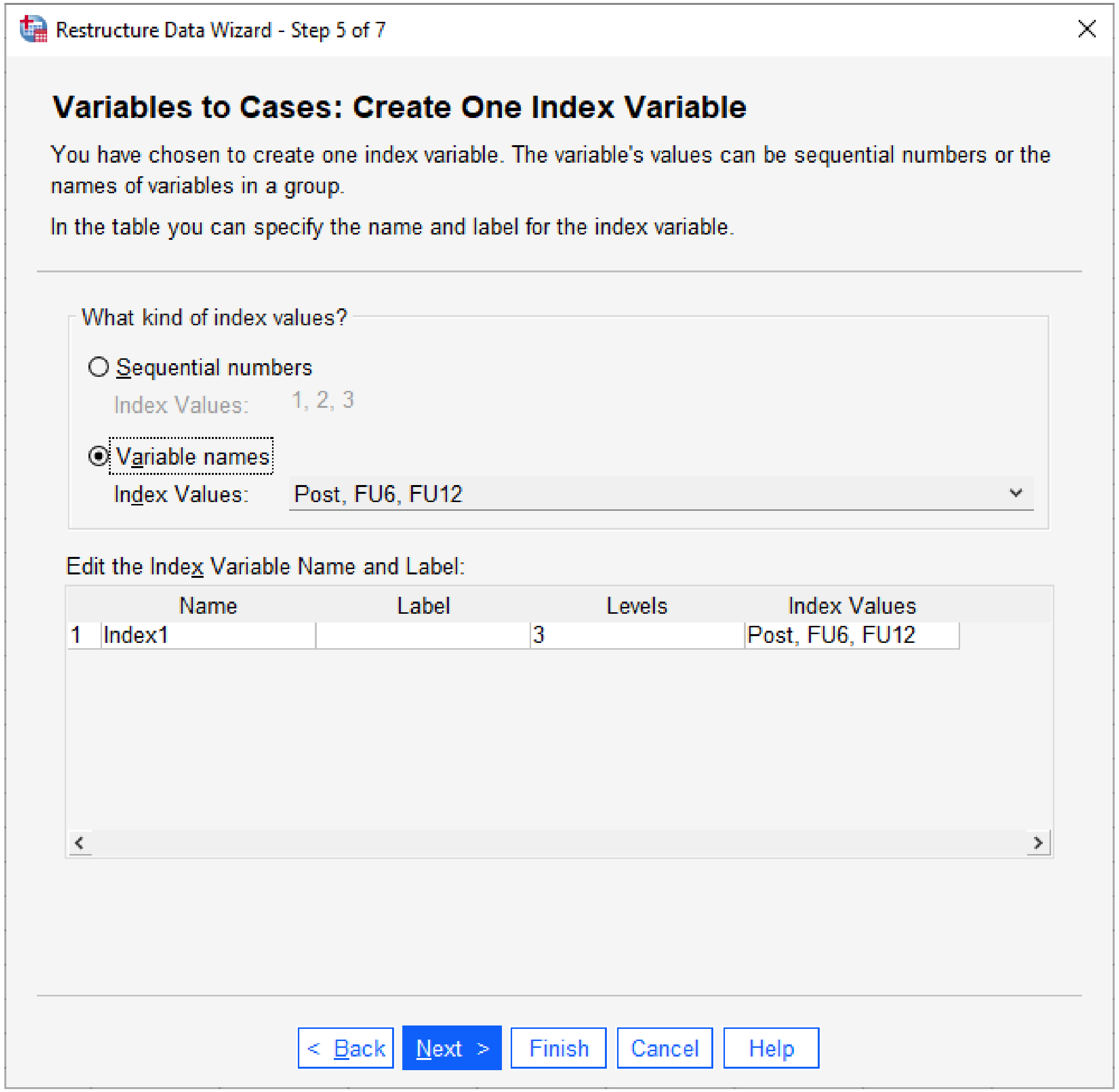
After button Next



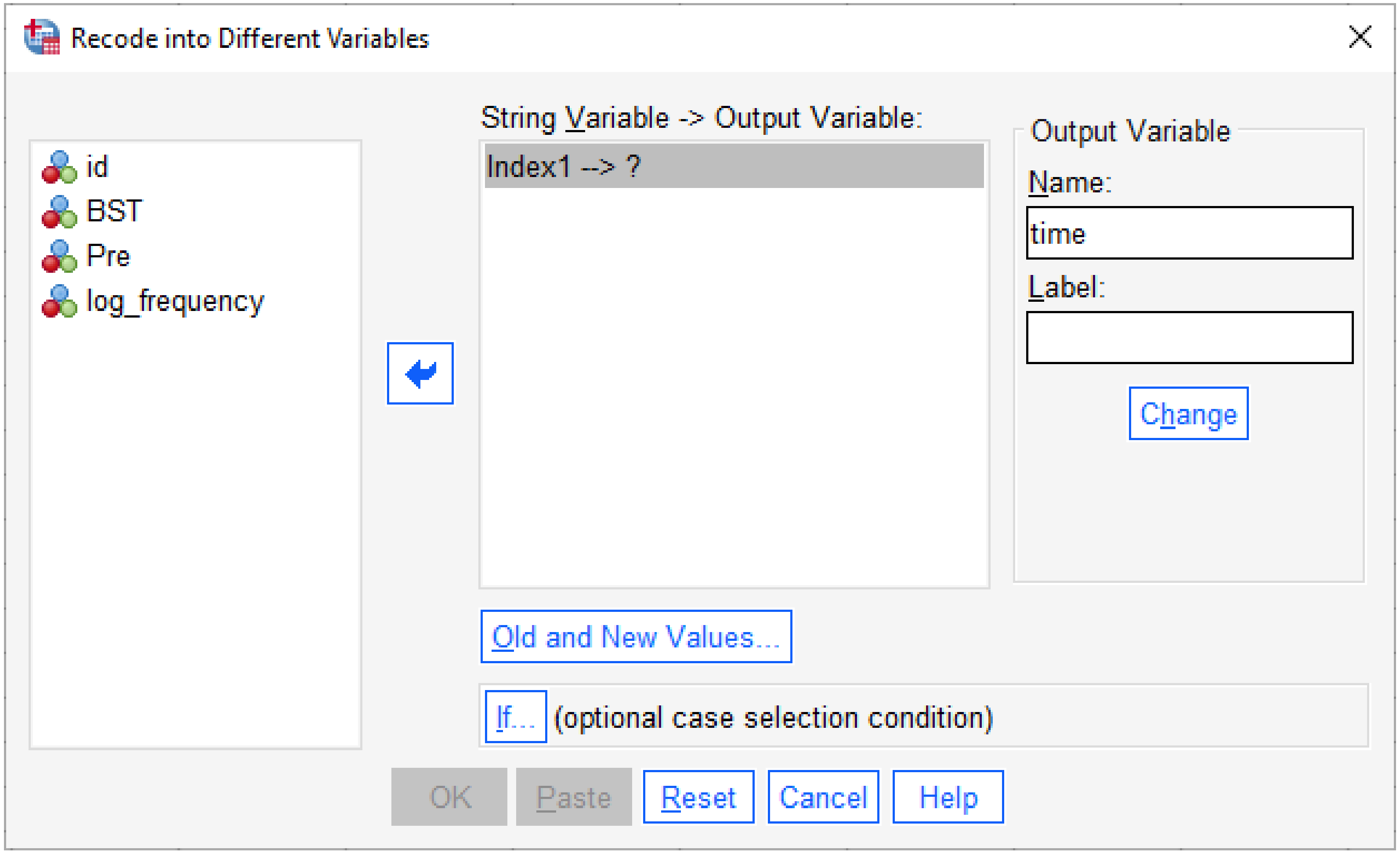
After button Next



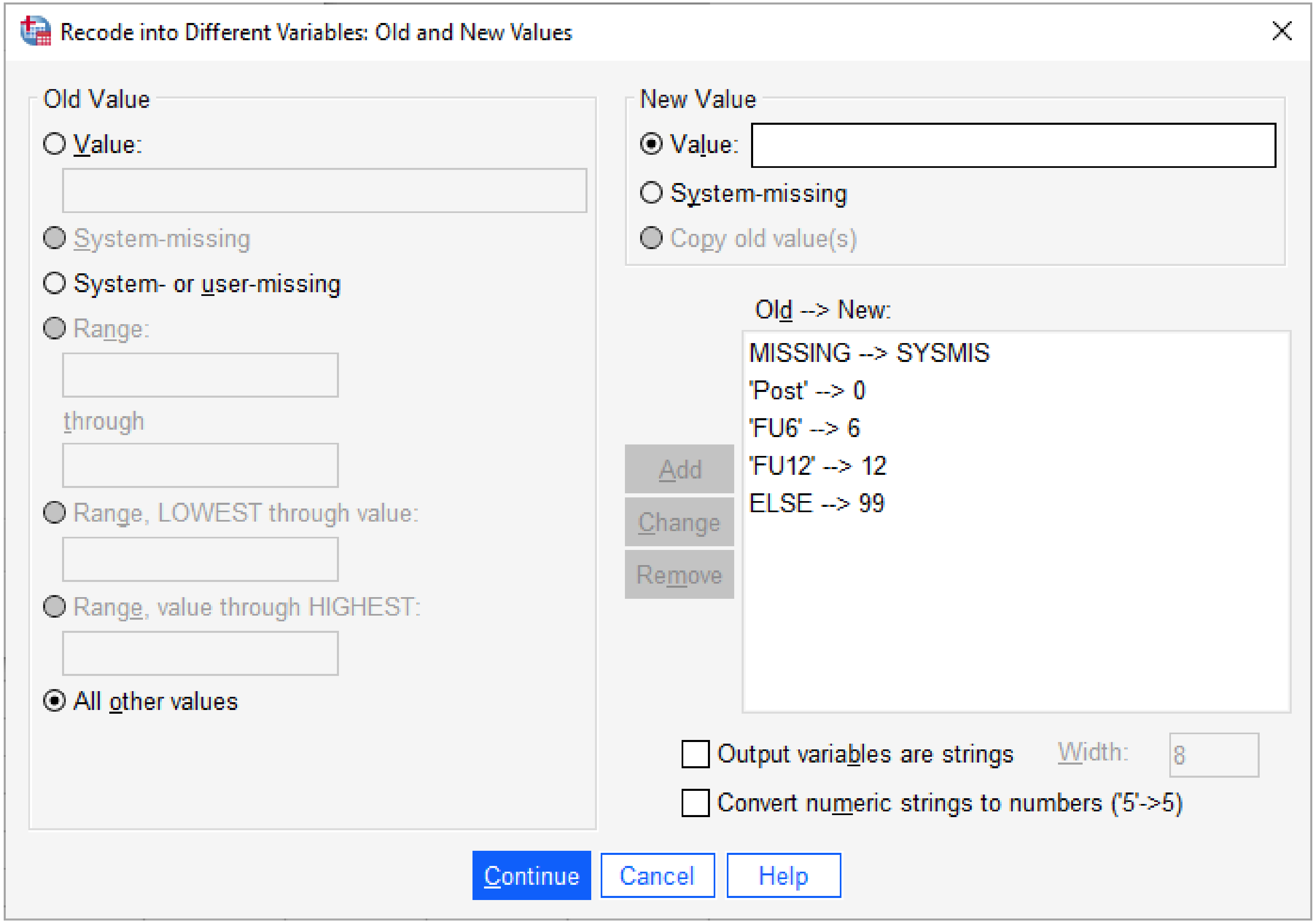
After button Next



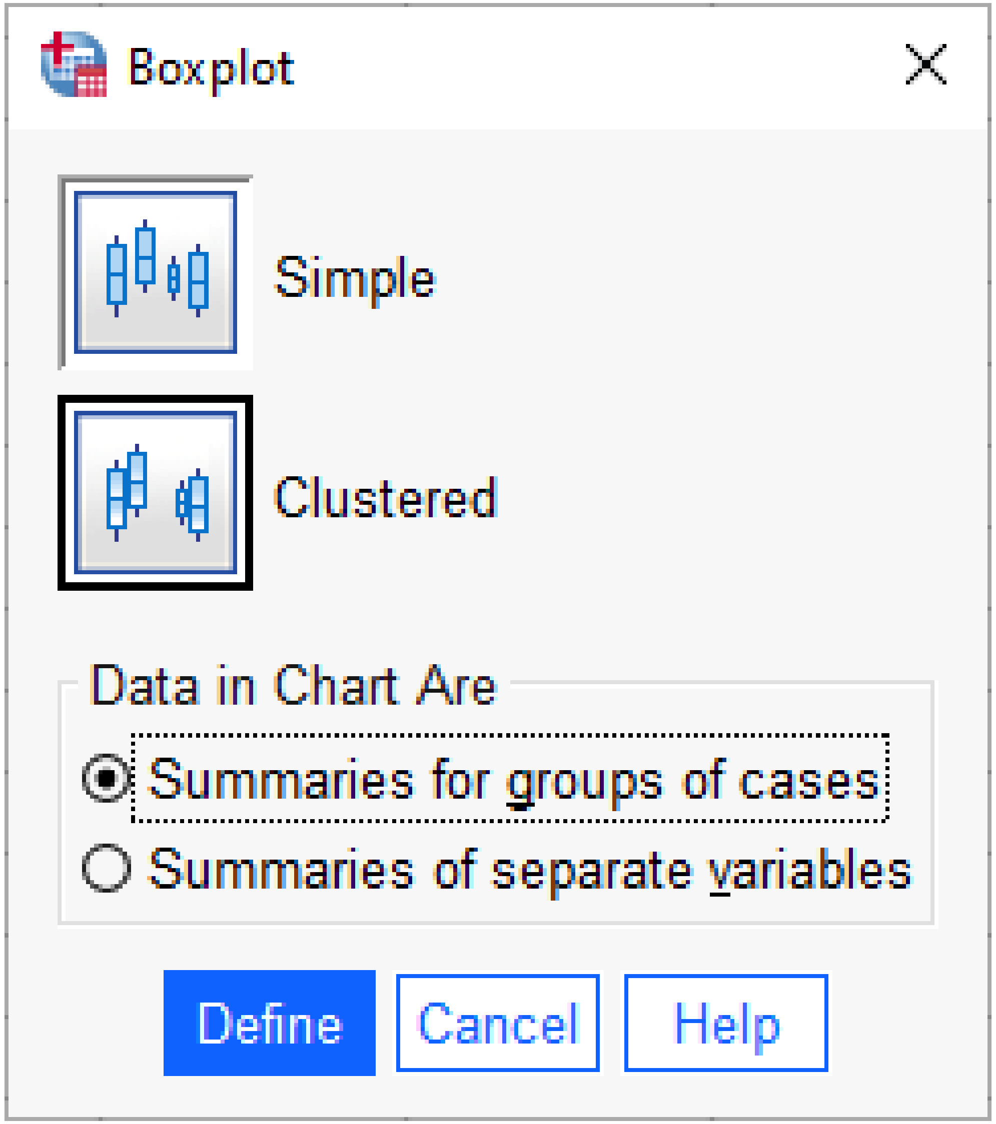
After menu Transform | Recode into Different Variables



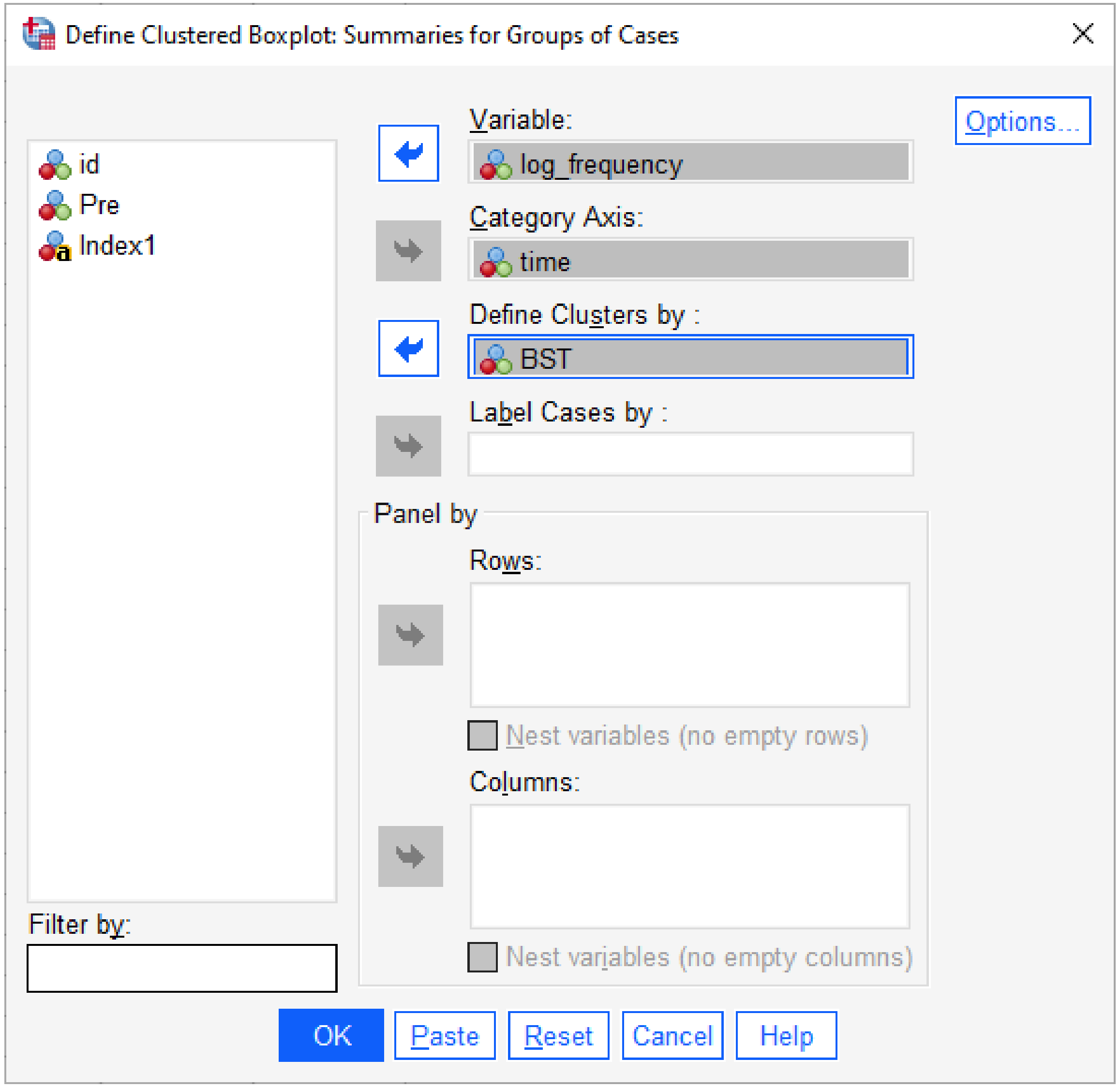
After button Old and New Values



After menu Graphs | Boxplot



After button Define



Break #2

- What you have learned
  - Description of HIV-intervention data
- What’s coming next
  - Random intercepts model using hiv-intervention data

Random intercepts analysis, 1 of 6

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	BST	2		1	
	time	1		1	
Random Effects	Intercept	1	Variance Components	1	id
Residual				1	
Total		5		5	

a. Dependent Variable: log_frequency.

Speaker notes

Random intercepts analysis, 2 of 6

Information Criteria[a]

-2 Restricted Log Likelihood	473.80460947
Akaike's Information Criterion (AIC)	477.80460947
Hurvich and Tsai's Criterion (AICC)	478.02683170
Bozdogan's Criterion (CAIC)	483.89071201
Schwarz's Bayesian Criterion (BIC)	481.89071201

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

a. Dependent Variable: log_frequency.

Speaker notes

Random intercepts analysis, 3 of 6

Coefficients of Determination

Pseudo-R Square Measures	Marginal	.064
	Conditional	.395

Speaker notes

Random intercepts analysis, 4 of 6

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	38.556	37.935	<.001
BST	1	18.002	2.110	.164
time	1	39.000	.723	.400

a. Dependent Variable: log_frequency.

Speaker notes

Random intercepts analysis, 5 of 6

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	22.508	3.935	28.633	5.720	<.001	14.455	30.561
[BST=0]	-7.133	4.910	18.002	-1.453	.164	-17.450	3.183
[BST=1]	0[b]	0	.	.	.	.	.
time	-.263	.309	39.000	-.850	.400	-.887	.362

a. Dependent Variable: log_frequency.

b. This parameter is set to zero because it is redundant.

Speaker notes

Random intercepts analysis, 6 of 6

Estimates of Covariance Parameters[a]

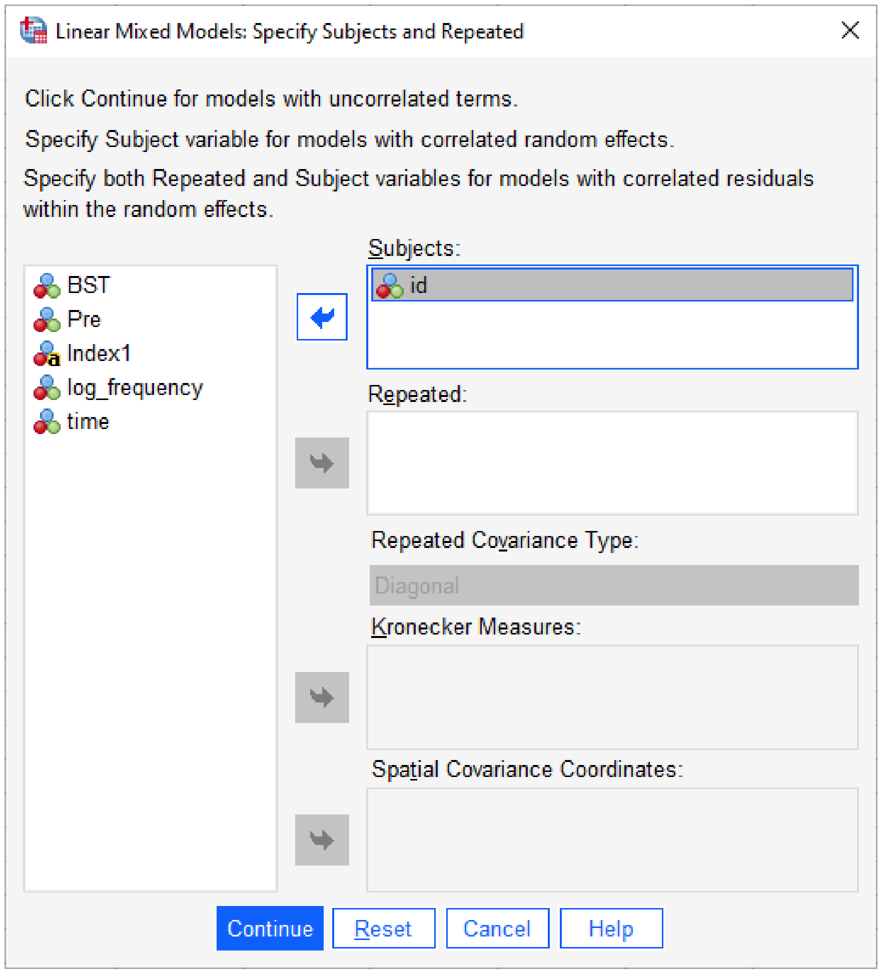
Parameter		Estimate	Std. Error
Residual		137.199	31.070
Intercept [subject = id]	Variance	74.827	41.497

a. Dependent Variable: log_frequency.

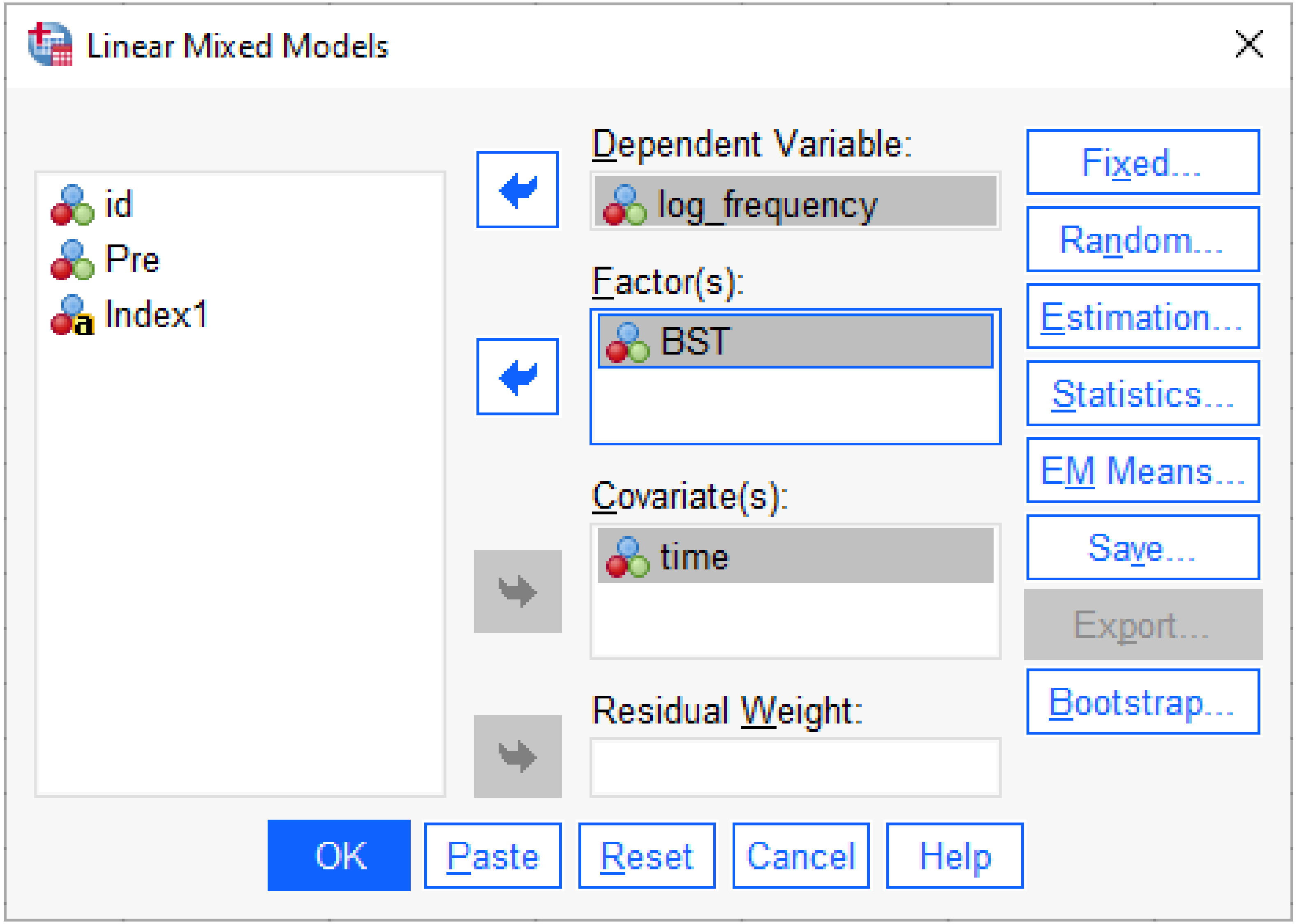
Speaker notes

Live demo, Random Intercepts Model

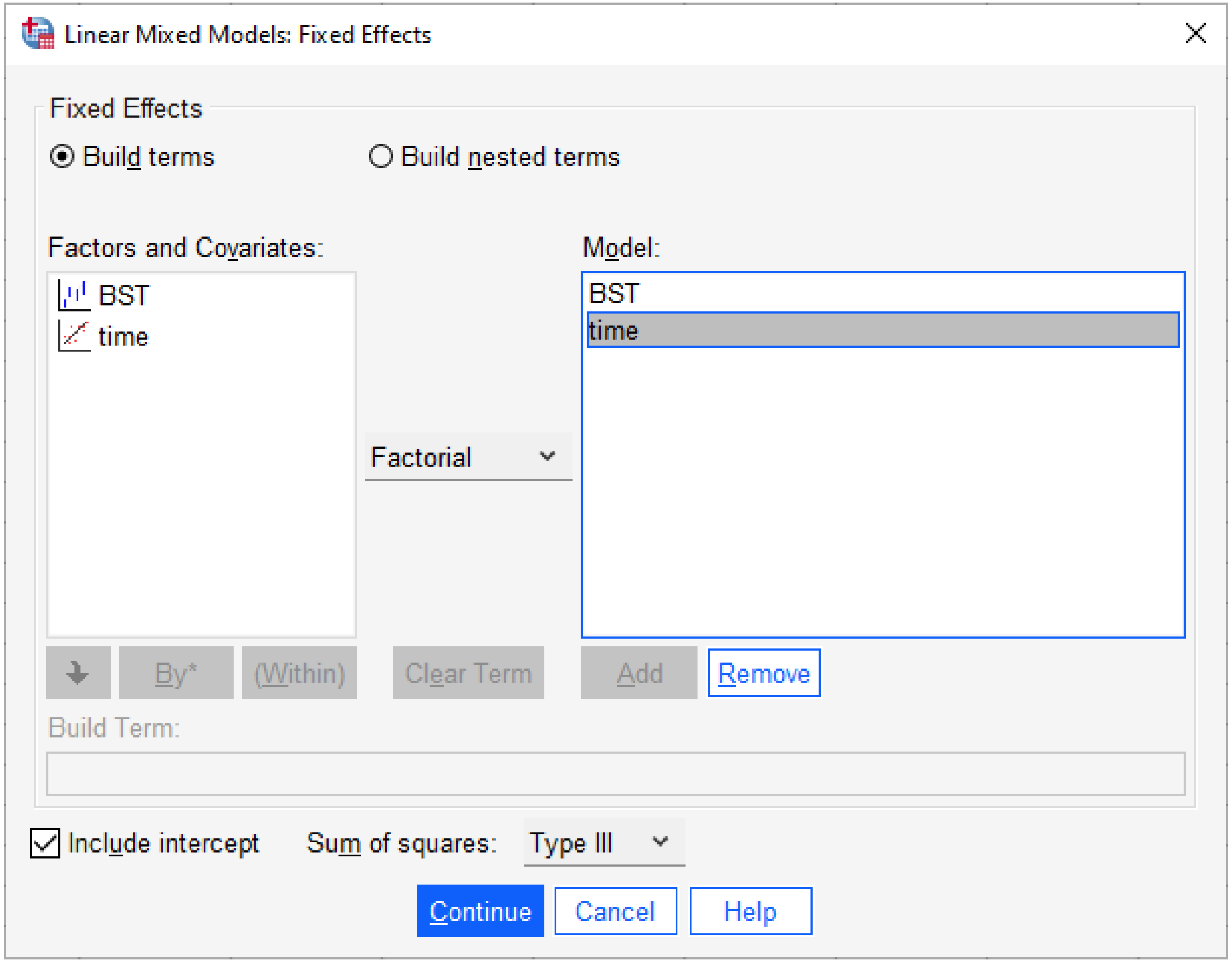
After menu Analyze | Mixed Models | Linear



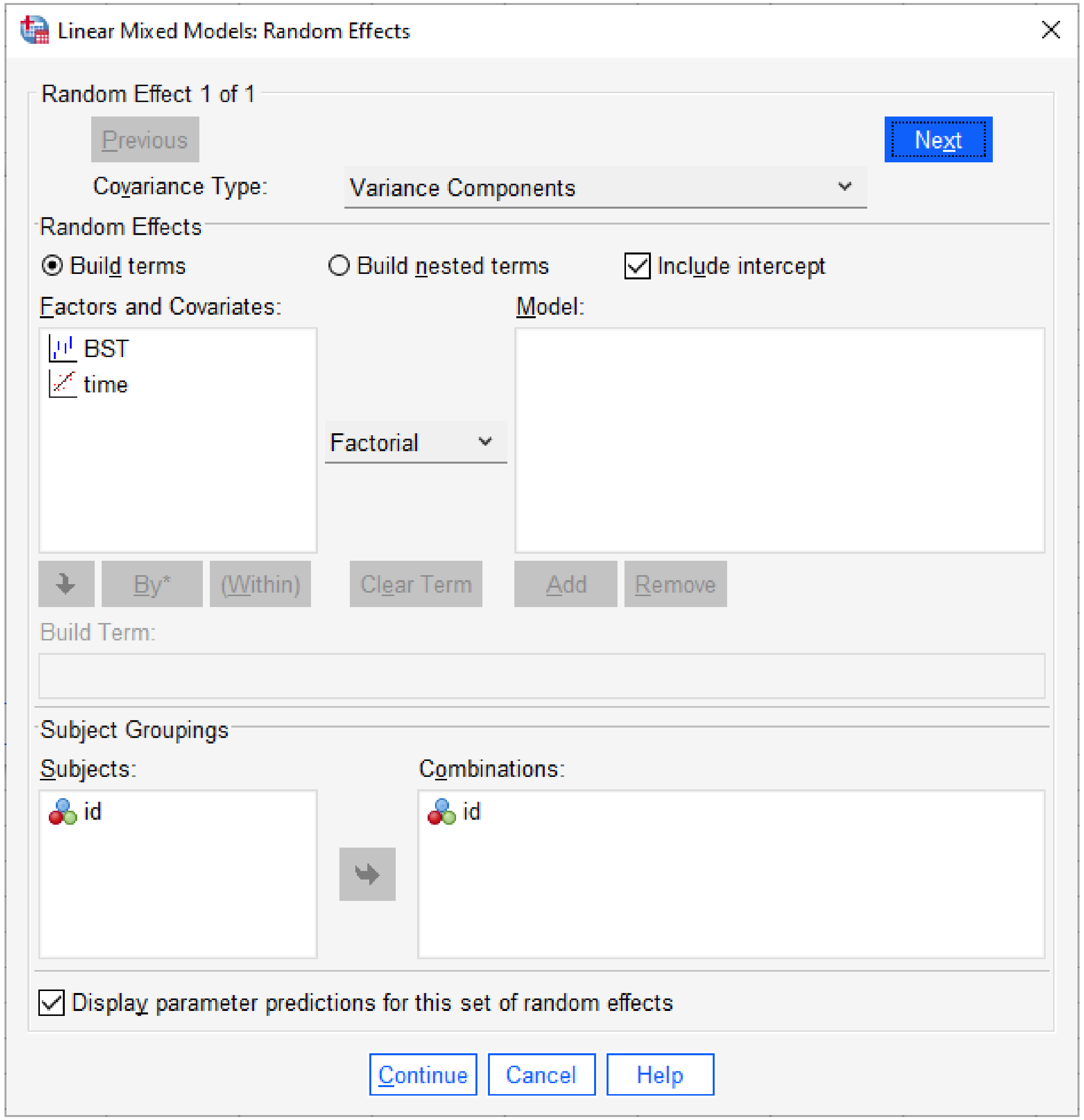
After button Continue



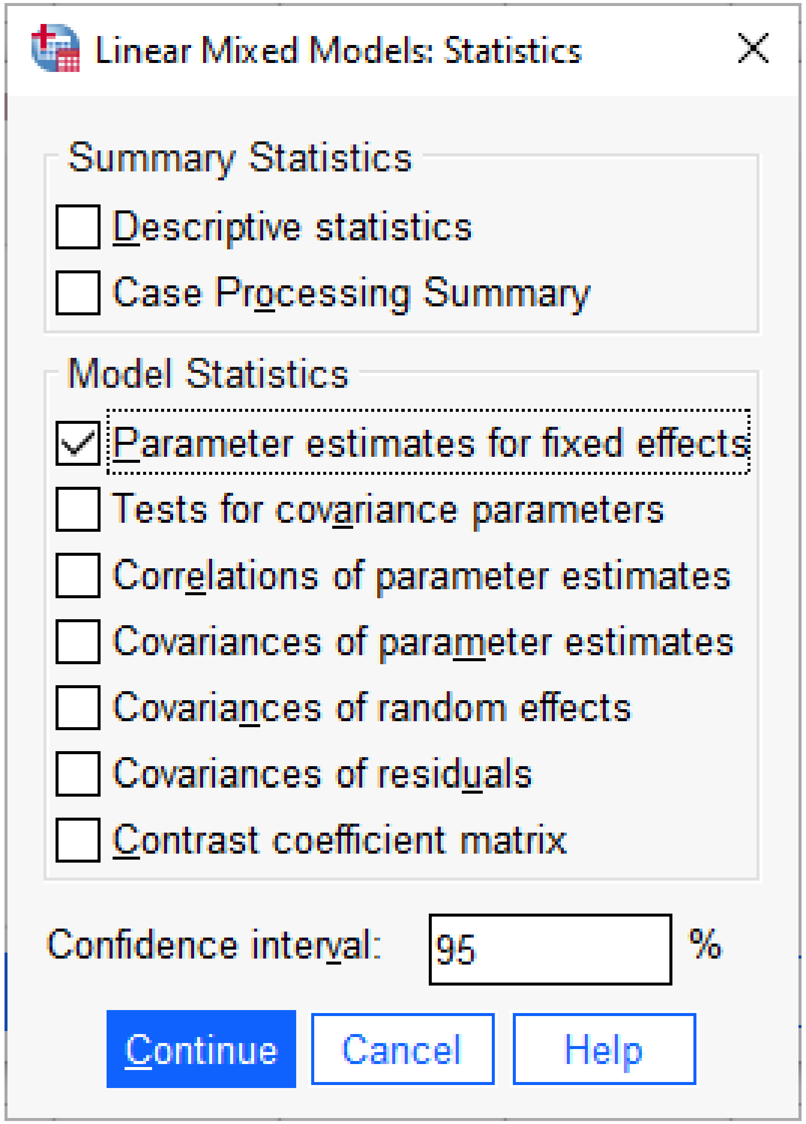
After button Fixed



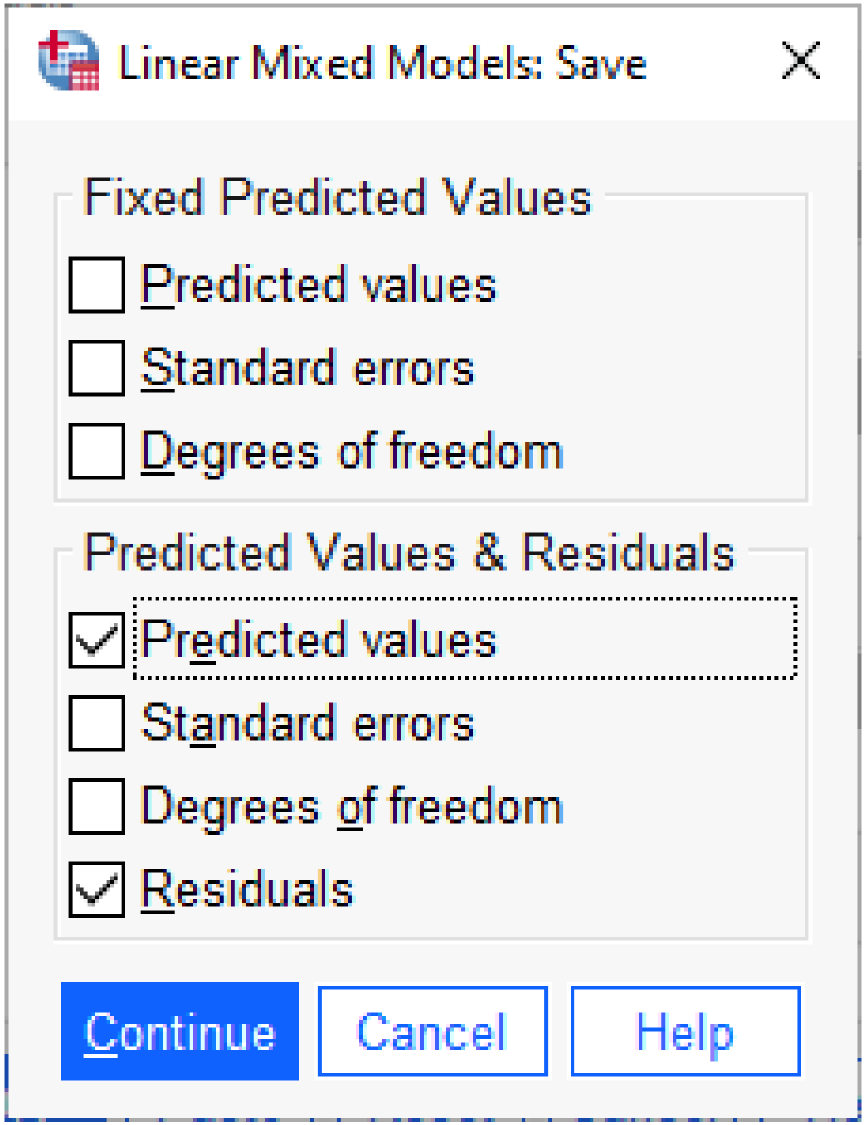
After button Random



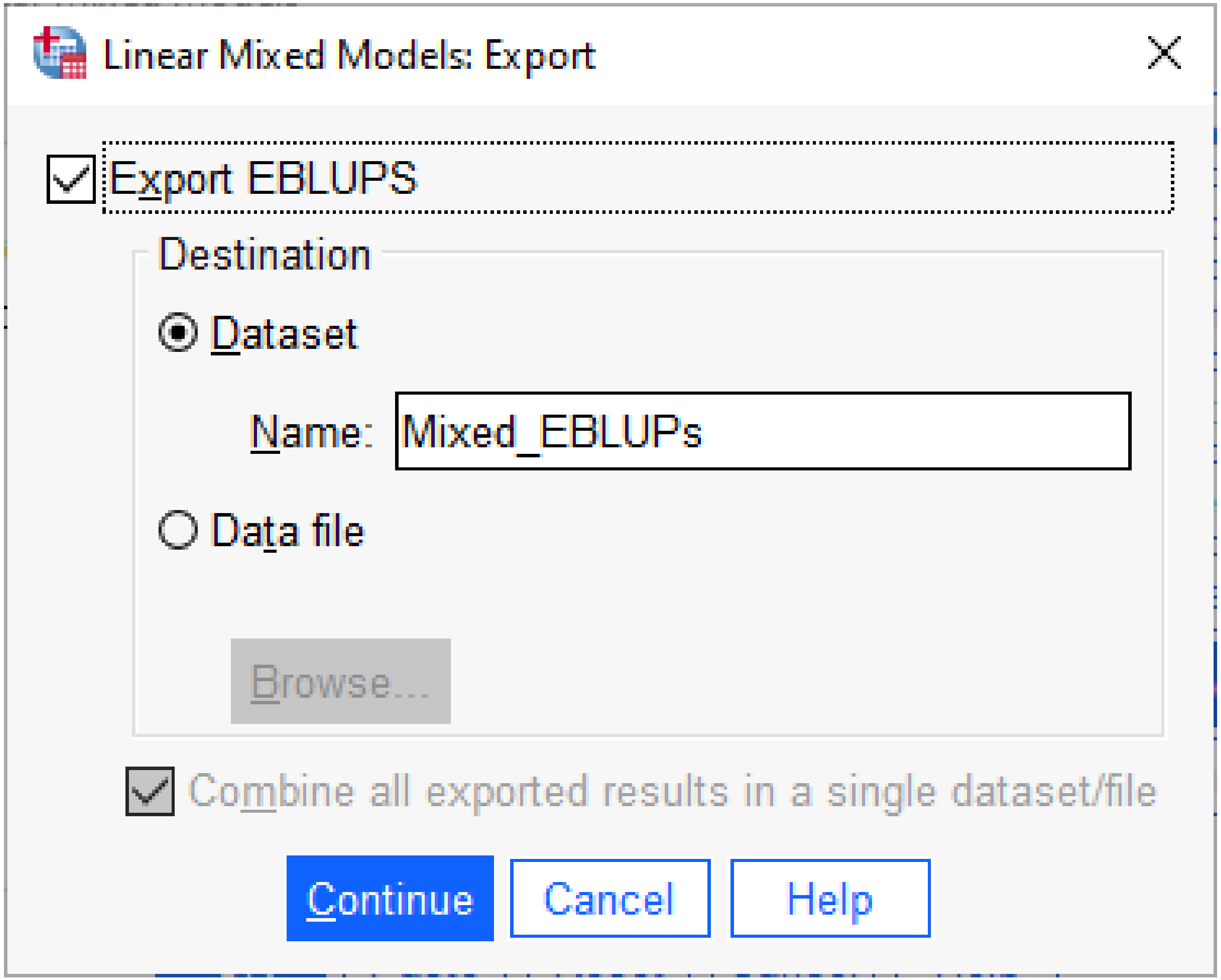
After button Statistics



After button Save



After button Export



Break #3

- What you have learned
  - Random intercepts model using hiv-intervention data
- What's coming next
  - Mathematical formulation of random slopes model

Random slopes model, 1 of 2

- Same notation for the time and outcome variables
- $Y_{ij},\ i = 1, \ldots, n;\ j = 1, \ldots, k$
  - n subjects, k time points
- t_j, time of jth measurement

Speaker notes

The random slopes model has the same basic notation for the time and outcome variables. The outcome variable has two subscripts on for the individual patient and one for each time measurement.

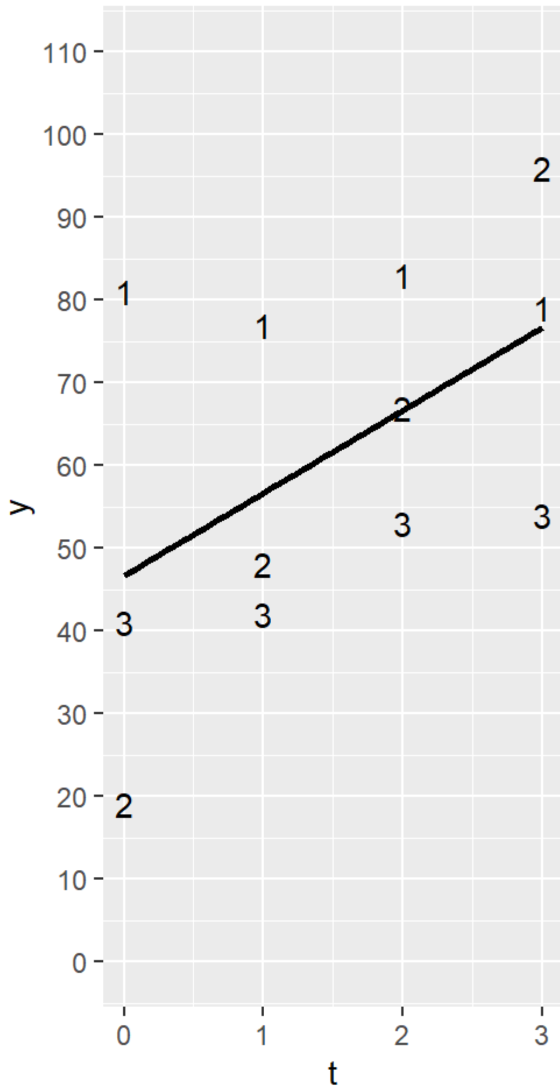
Random slopes model, 2 of 2

- $Y_{ij} = \beta_0 + u_{0i} + \beta_1 t_j + u_{0i} t_j + \epsilon_{ij}$
  - β_0 and β_1 are unknown constants
  - u_{0i}, u_{0i}, and ϵ_{ij} are normally distributed
    - $SD(u_{0i}) = \sigma_{intercept}$
    - $SD(u_{1i}) = \sigma_{slope}$
    - $SD(\epsilon_{ij}) = \sigma_{error}$

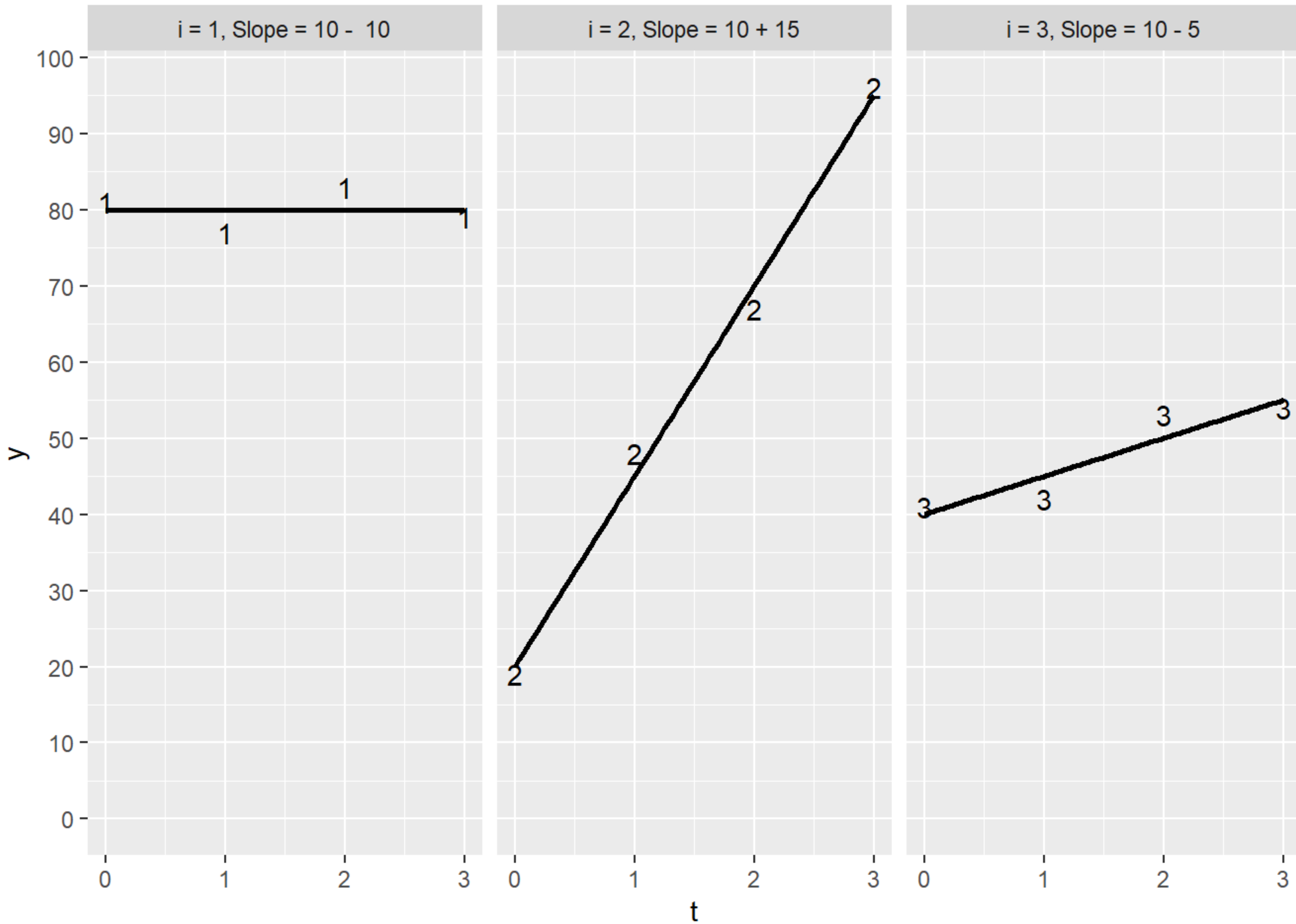
Speaker notes

There are three sources of random variation in the random slopes model, u_{0i}, u_{1i}, and ϵ_{ij}.

Random slopes illustrated, 1 of 2



Random slopes illustrated, 2 of 2



Break #4

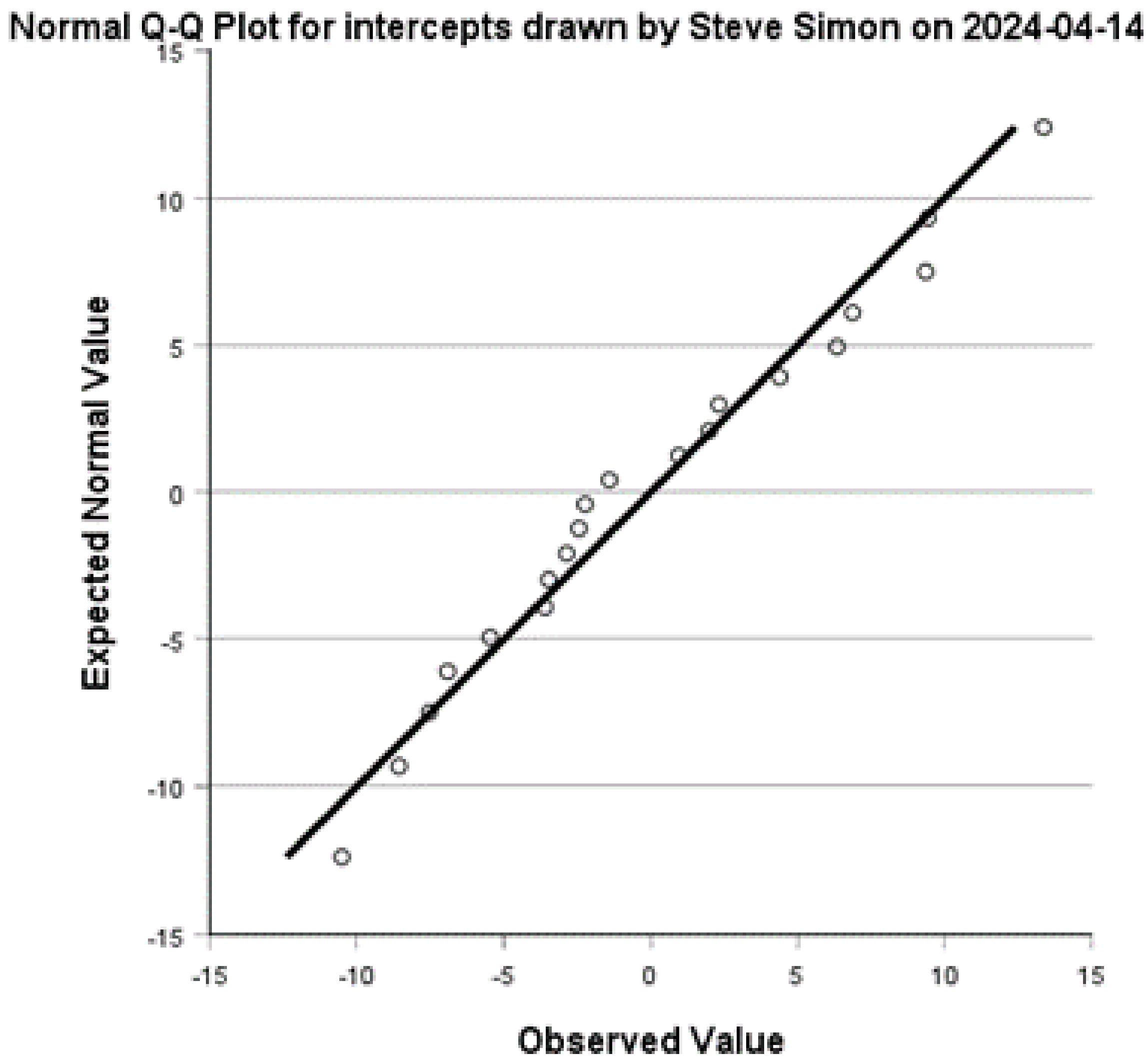
- What you have learned
  - Mathematical formulation of random slopes model
- What's coming next
  - Assumptions and complications

Assumptions

- Independence
  - Only between subjects
- Normality
  - Residuals
  - Random intercepts and/or slopes
- Linearity

Speaker notes

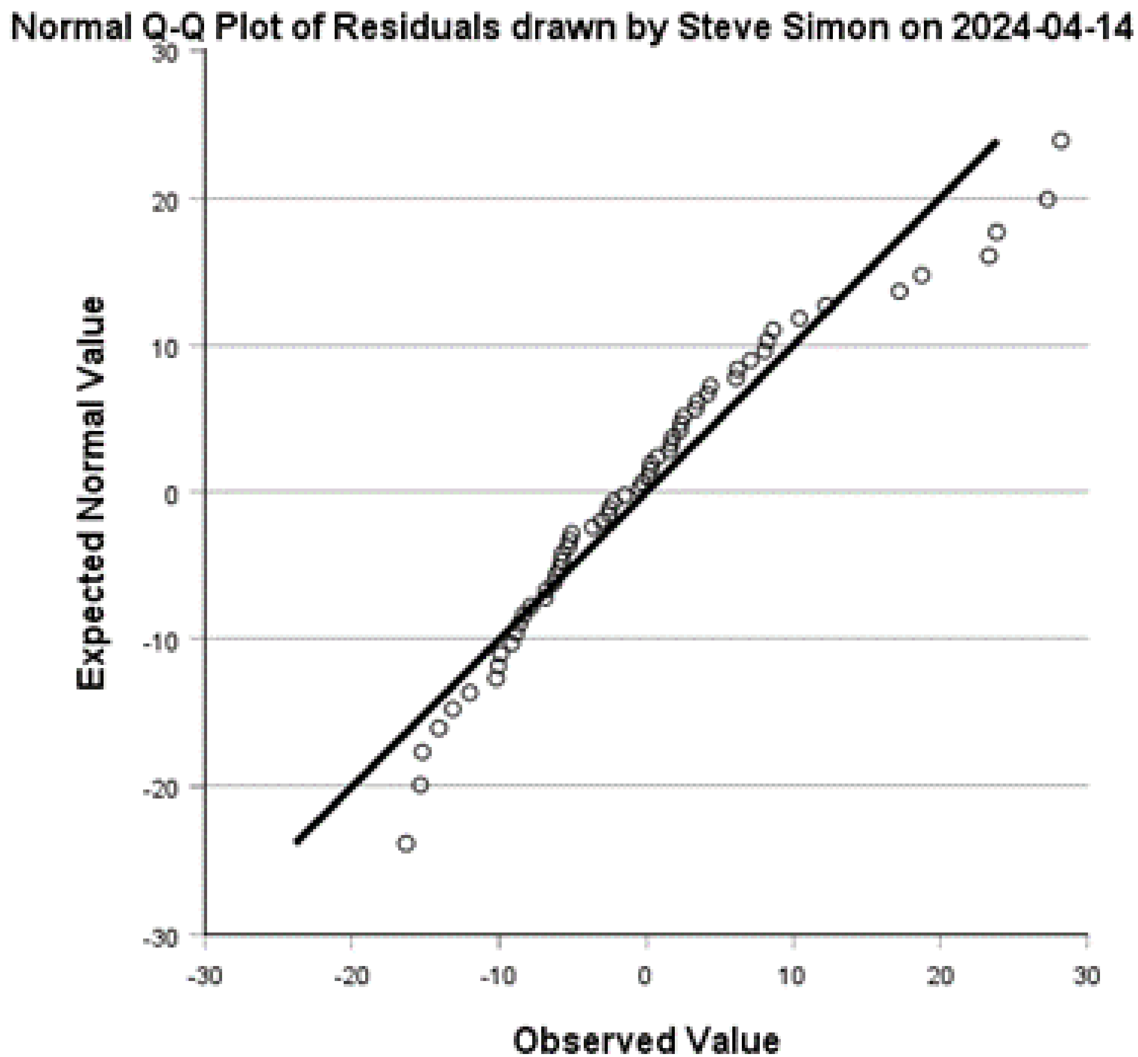
Normality check, 1 of 2



Speaker notes

The residuals represent the deviations of individual time measurements from the regression line for an individual patient. That means the deviation from a trend line that has been randomly shifted up or down (random intercepts) and/or a randomly steeper or flatter trend line (random slopes). In this dataset, the normal probability plot looks fairly close to a straight line.

Normality check, 2 of 2



Speaker notes

Linearity check

Complications

- Not a problem
  - Missing values
  - Better than Last Observation Carried Forward
- Problems (more tedious than difficult)
  - Interactions
  - Nonlinear trends
  - Covariates
    - Between patients
    - Within patients

Speaker notes

Missing values are normally a big headache, but less so for the random intercepts or slopes models. If a patient missed a visit at a particular time point, you can easily extrapolate from the visits before and after. You are assuming linearity, after all.

In some clinical trials, if a patient came at the intermediate evaluation but did not show up at the final evaluation, the research team would replace the missing final outcome with the intermediate outcome. This technique, called Last Observation Carried Forward (LOCF) was not very popular when it was introduced and has been pretty much discredited. The random intercepts or slopes model would extrapolate the trend from the intermediate value, which works much better.

Interactions are always difficult, and when the interactions involve time, it can get a bit messy. Nonlinear trends over time are also a bit of a problem. In both cases, though, the work is more tedious than difficult. Interactions and nonlinearity mean that your interpretation of the results will require a bit more thought and you can’t come up with as simple a story to tell.

A covariate is a variable which is not of direct interest in the research, but one that you must take account of in order to produce a credible analysis. In just about any cancer study, you should track whether the patient is a smoker. It’s not something you’re interested in testing. The role of smoking in lung cancer and most other types of cancer was established many decades ago. You still have to account for smoking though because it can explain so much of the variation in your outcome. Failure to account for smoking would greatly reduce your power and precision.

There are two types of covariates. The ones that are fixed and do not change over time are called time constant covariates or between subject covariates. Patient demographics are time constant. Measurements done at baseline to assess how ill the patient was at the start of the study are time constant.

Covariates that change over time are called time varying covariates or within subject covariates. The extent to which a patient complies with taking his/her medication is a time varying covariate. Seasonal changes in temperature, humidity, or pollen counts are time varying covariates.

There is one important distinction between time constant and time varying covariates. The latter are much better at removing variation from your outcome, and can greatly improve your power and precision.

Live demo, normality checks

Summary

- What you have learned
  - Mathematical formulation of random intercepts model
  - Description of HIV-intervention data
  - Random intercepts model using hiv-intervention data
  - Mathematical formulation of random slopes model
  - Assumptions and complications

Additional topics??

Speaker notes

- Learning objectives for mixed models
    - Set up a data set for a Repeated Measures ANOVA
    - Explain when a n-way RM ANOVA would be used to analyze data
    - Explain the data requirements for a n-way RM ANOVA
    - Write a research question for a n-way RM ANOVA
    - Write the Null and Research Hypotheses for a n-way RM ANOVA
    - Run a complete n-way RM ANOVA
    - List the assumptions for a n-way RM ANOVA
    - Explain the tests for the assumptions for a n-way RM ANOVA
    - Run the assumptions tests for a n-way RM ANOVA
    - Interpret the assumptions tests for a n-way RM ANOVA
    - Explain the issue that arises if the assumption of sphericity is violated
    - Explain how the Epsilon correction works
    - Explain the condition of a significant interaction effect
    - Explain the process to run simple main effects for a Repeated Measures n-way ANOVA
    - Run simple main effects for a Repeated Measures n-way ANOVA
    - Interpret simple main effects for a Repeated Measures n-way ANOVA
    - Explain the special considerations for simple main effects for a Repeated Measures n-way ANOVA with a factor with >2 levels
    - Explain the process to run main effects for a Repeated Measures n-way ANOVA
    - Run main effects for a Repeated Measures n-way ANOVA
    - Interpret main effects for a Repeated Measures n-way ANOVA
    - Summarize the output of a complete n-way RM ANOVA
    - Define the meaning of mixed model
    - Define Between Subjects in relation to the mixed model
    - Define Within Subjects in relation to the mixed model
    - Be familiar with the various names for a mixed model ANOVA
    - List the purpose of running a mixed model ANOVA
    - Identify Between and Within factors for a mixed model ANOVA
    - List the type of data that can be used to define the levels in a mixed model ANOVA
    - List the assumptions for a mixed model ANOVA
    - Successfully run the statistical analysis tests for the assumptions of a mixed model ANOVA

- Successfully run the statistical analysis tests for the assumptions of a mixed model ANOVA
- Interpret the statistical analysis tests for the assumptions of a mixed model ANOVA
- Explain how normality of the data is assessed
- Explain the rules for violation of the assumption of homogeneity of variances
- Define Homogeneity of Covariances
- Identify Box's Test
- Develop a research question for a mixed model ANOVA
- Write the Null and Alternative for a mixed model ANOVA
- Set up data for a mixed model ANOVA in the SPSS data editor
- Run a two-way mixed model ANOVA analysis using SPSS
- Interpret the output of a two-way mixed model ANOVA analysis
- Understand the process of determining the presence or not of an interaction effect
- Define Simple Main Effects
- Determine when Simple Main Effects would be run
- Define Main Effects
- Determine when Main Effects would be run
- Run the Simple Main Effects following a decision about the interaction effect
- Interpret the Simple Main Effects
- Run the Main Effects following a decision about the interaction effect
- Interpret the Main Effects
- Interpret Pair-Wise comparisons
- Create a complete write up of a mixed model ANOVA analysis