A mix within

Linear Mixed Models and their applications

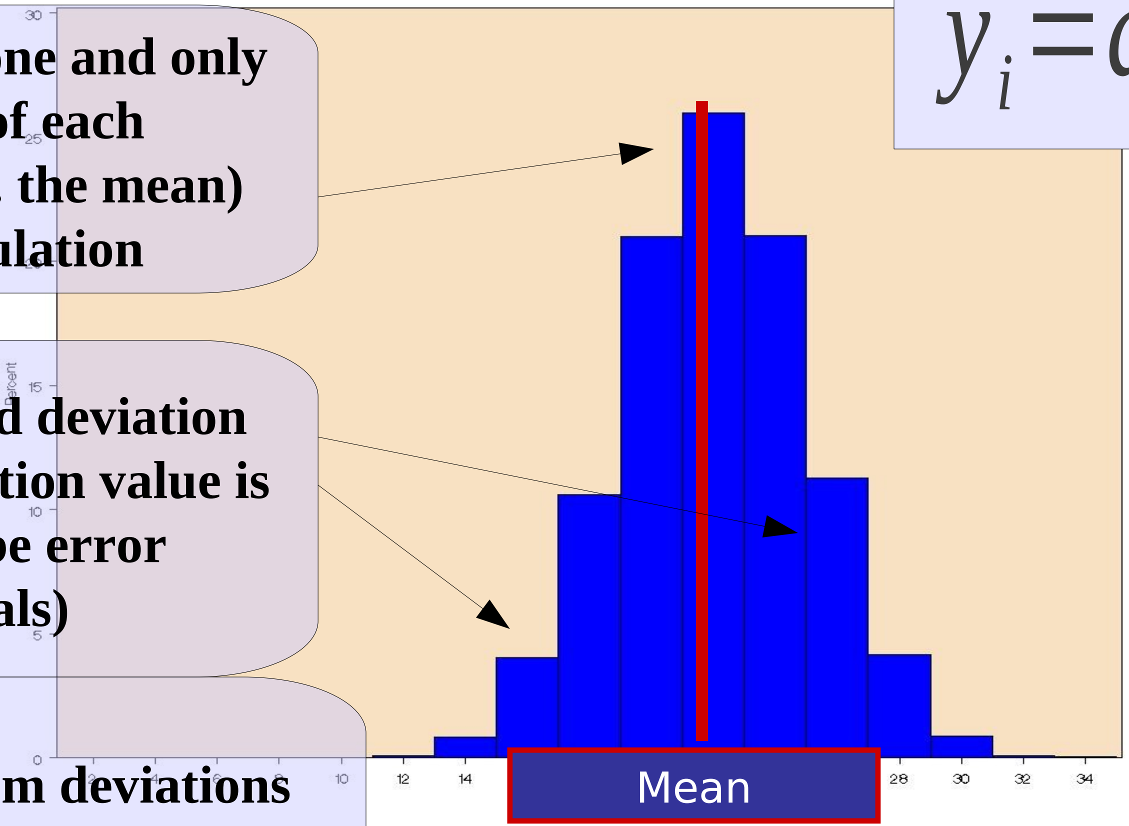
Marcello Gallucci
Univerisity of Milano-Bicocca

Some GLM Assumptions

$$y_i = a + e_i$$

1) There exists one and only one value of each parameter (e.g. the mean) in the population

2) Any observed deviation from the population value is deemed to be error (residuals)



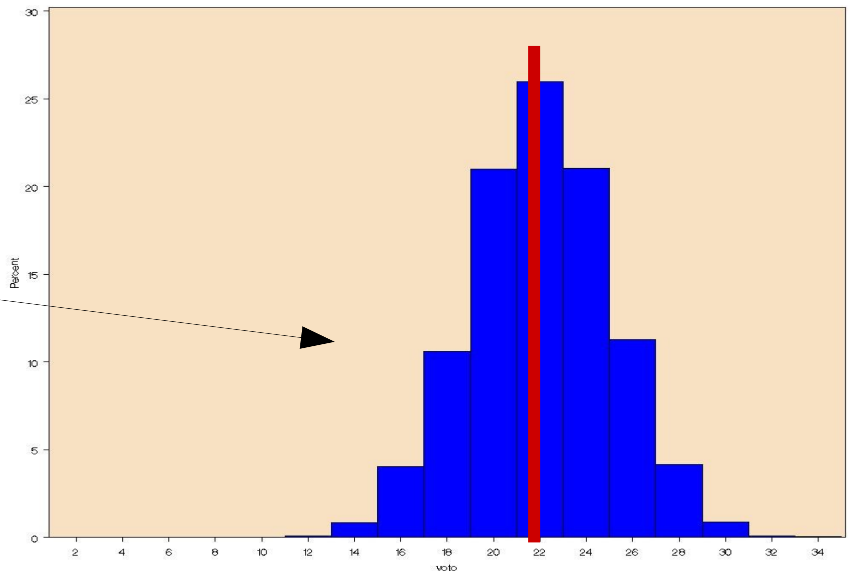
3) The random deviations from the model are normally distributed

GLM Assumptions

$$y_i = a + e_i$$

$$corr(e_i, e_j) = 0$$

Random variations are independent and normally distributed



$$e_i \sim N(0, \sigma)$$

GLM

When the assumptions are NOT met because the data, and thus the errors, have more complex structures, we generalize the GLM to the Linear Mixed Model

The Linear Mixed Model is the statistical model underlying **multilevel models** and **repeated measures** analysis

Linear Mixed Model

GLM

Regression

T-test

ANOVA

ANCOVA

Moderation

Mediation

Path Analysis

LMM

Random coefficients models

Random intercept regression models

One-way ANOVA with random effects

One-way ANCOVA with random effects

Intercepts-and-slopes-as-outcomes models

Multi-level models

Example “beers”

Let's consider the case where the beer-smile research was conducted by gathering data in several different bars

For each participant we measured # of beers and # of smiles

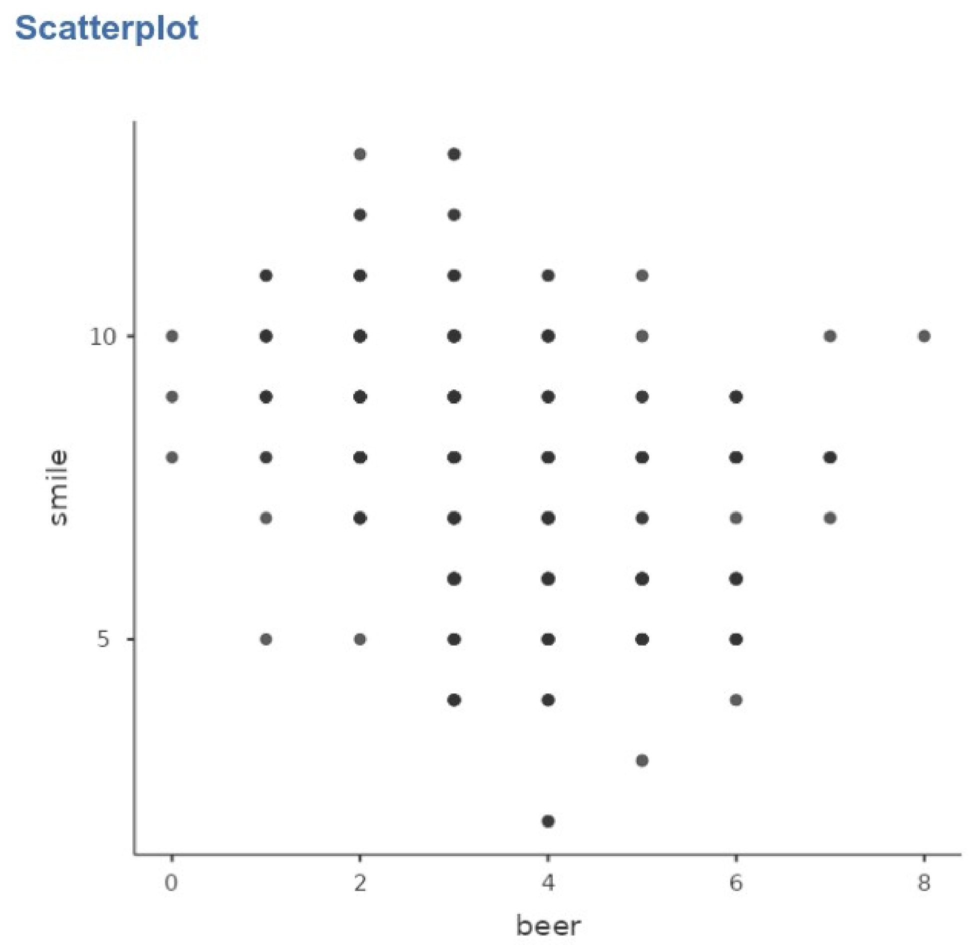
bar

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	a	3	1.3	1.3	1.3
	b	14	6.0	6.0	7.3
	c	22	9.4	9.4	16.7
	d	21	9.0	9.0	25.6
	e	14	6.0	6.0	31.6
	f	20	8.5	8.5	40.2
	g	24	10.3	10.3	50.4
	h	12	5.1	5.1	55.6
	i	16	6.8	6.8	62.4
	l	22	9.4	9.4	71.8
	m	21	9.0	9.0	80.8
	n	15	6.4	6.4	87.2
	o	16	6.8	6.8	94.0
	p	11	4.7	4.7	98.7
	q	3	1.3	1.3	100.0
	Total	234	100.0	100.0	

For a total of 234 participants

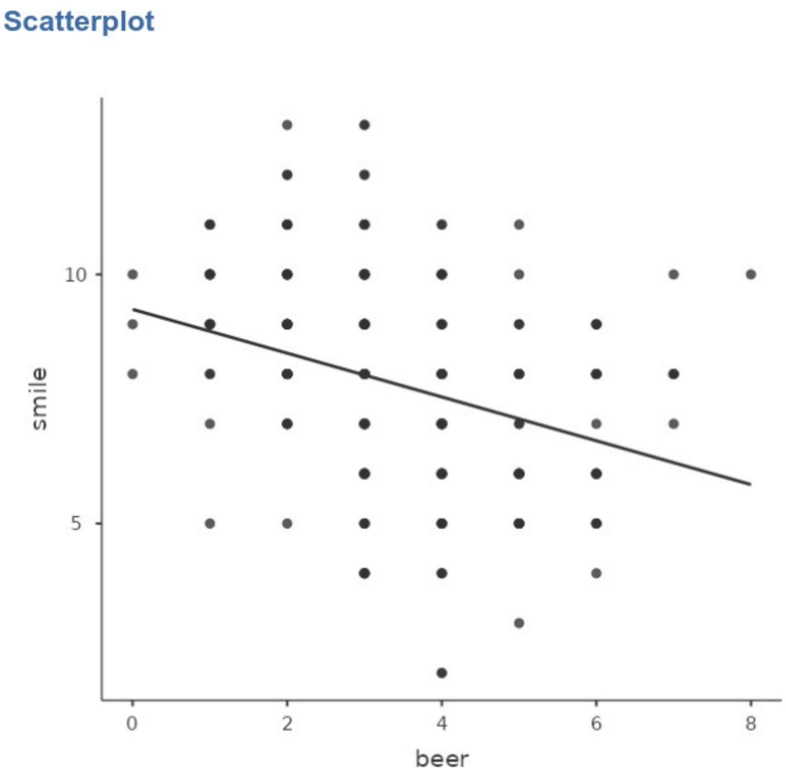
Example "beers" 2

As compared with the example with a few participants, now we have a very different scatterplot



Example “beers” 2

A simple regression confirms that results are indeed different



Negative effect

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval Lower	95% Confidence Interval Upper	β	df	t	p
(Intercept)	7.765	0.130	7.508	8.022	0.000	232	59.503	< .001
beer	−0.440	0.085	−0.608	−0.271	−0.320	232	−5.147	< .001

Why

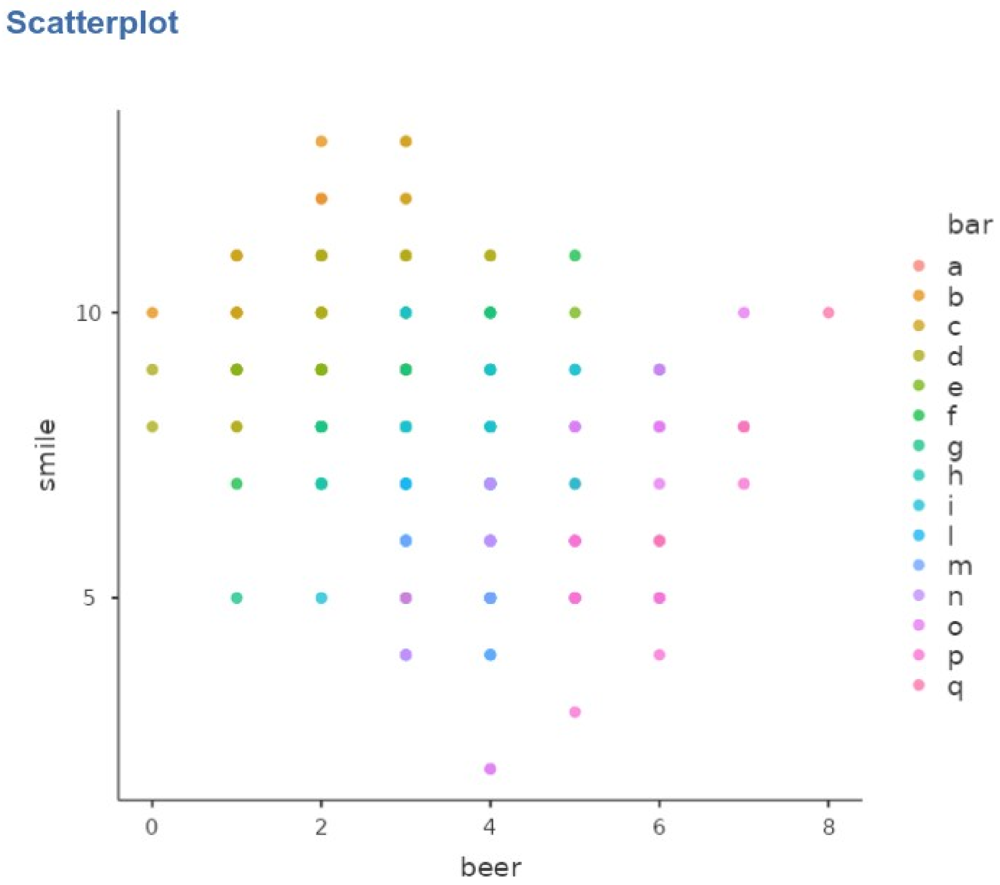
Results may be biased by a mis-specification of the model, where the structure of the data is not taken into account

- In fact:
  - Subjects are sampled in clusters specified by **bars**
  - Each bar may have specific characteristics (quality, entertainment, etc) that may affect the measured variables
  - Subjects within the same bar may be more similar than across bars

Scatterplot by Bar

Let's see the data broken down by bar

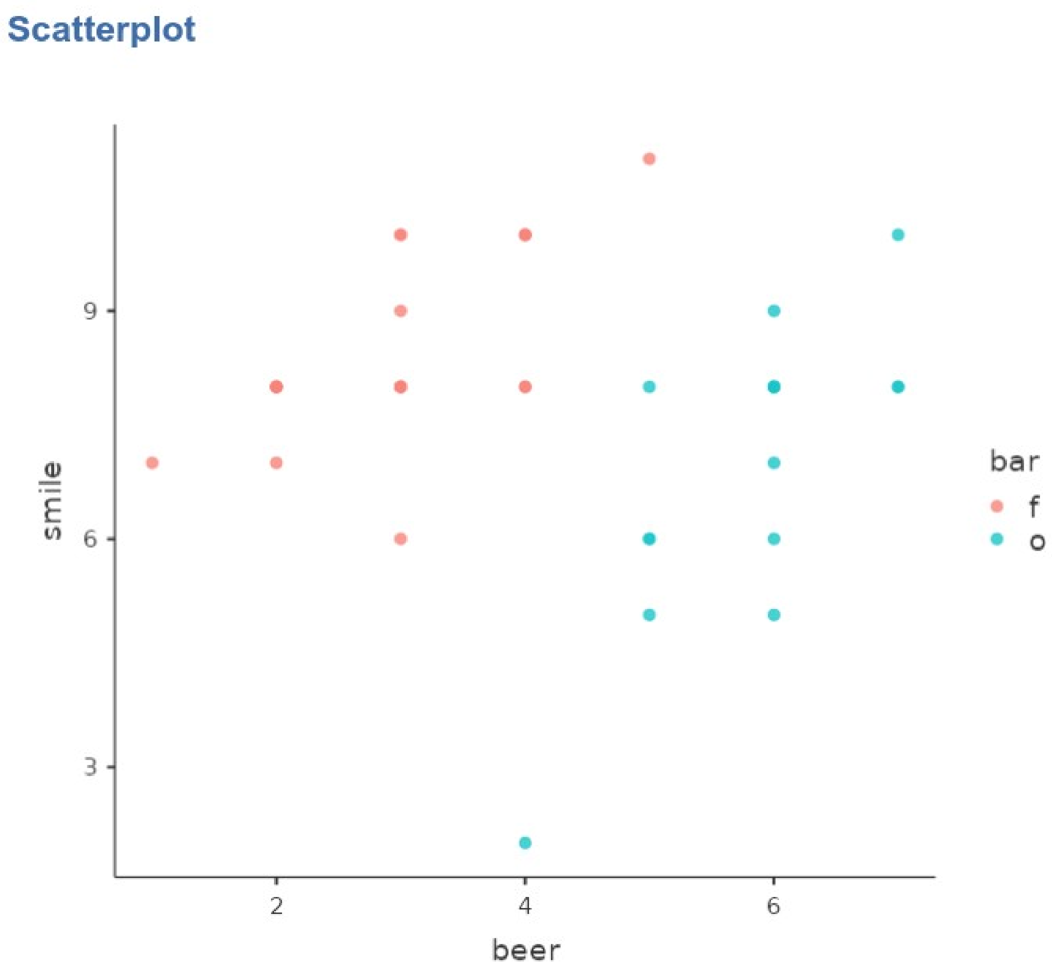
Bar



Scatterplot by Bar

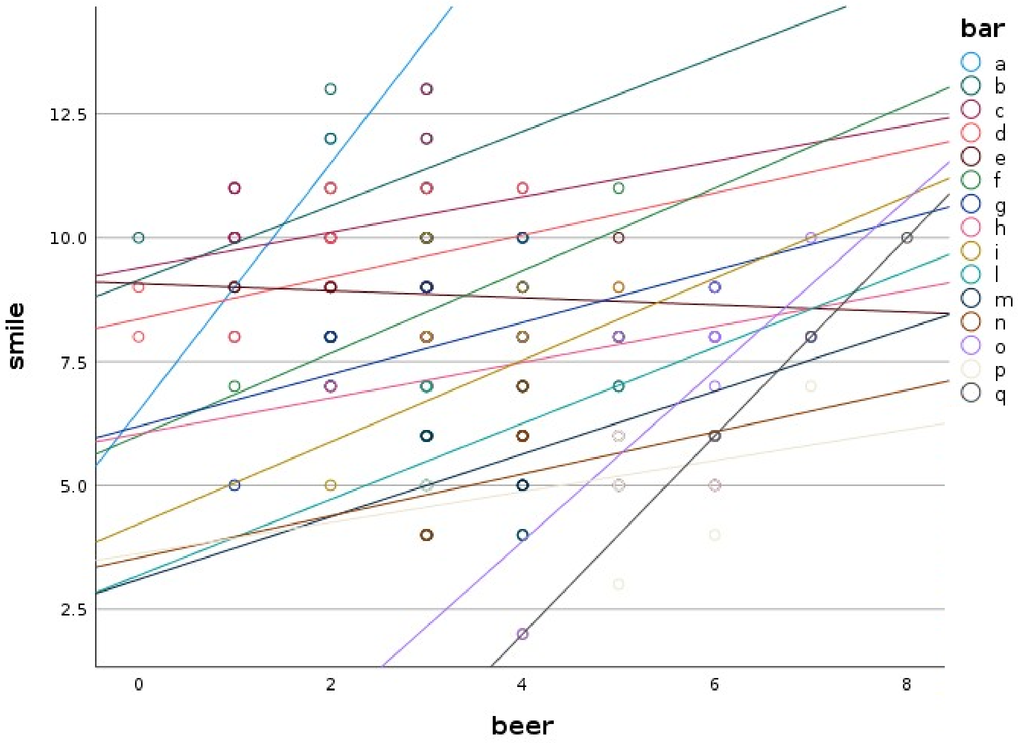
Let's see the data only for bar "f" and "o"

Bar



Scatterplot by Bar

It seems that the relations between IV and DV is positive, but within each bar

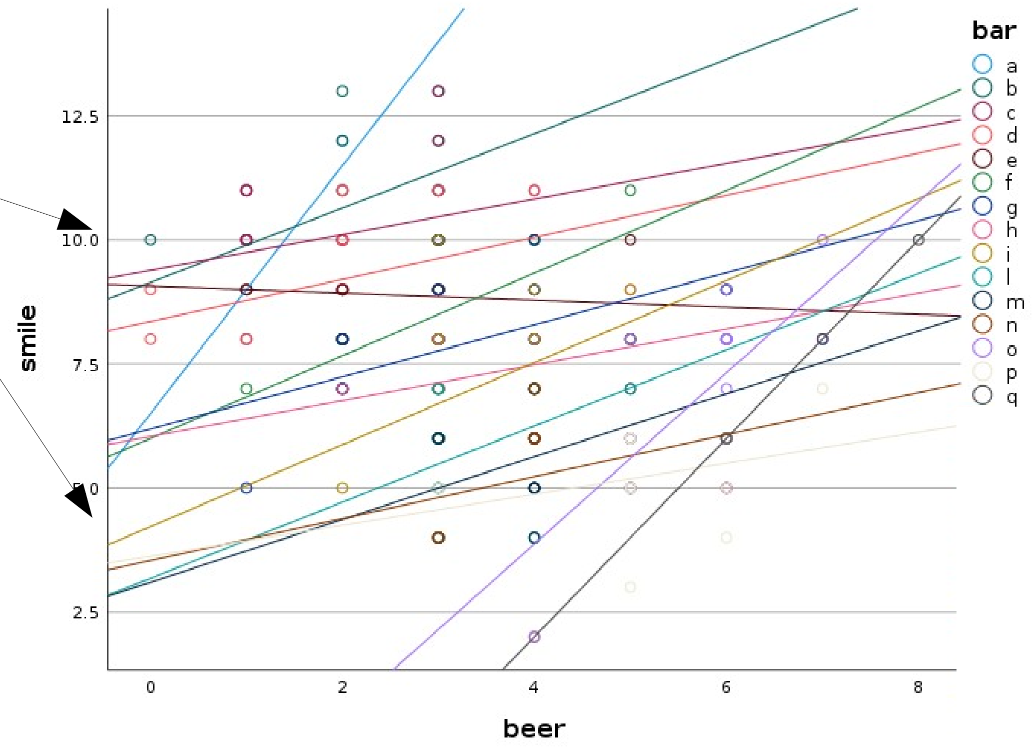


Bar

Scatterplot by Bar

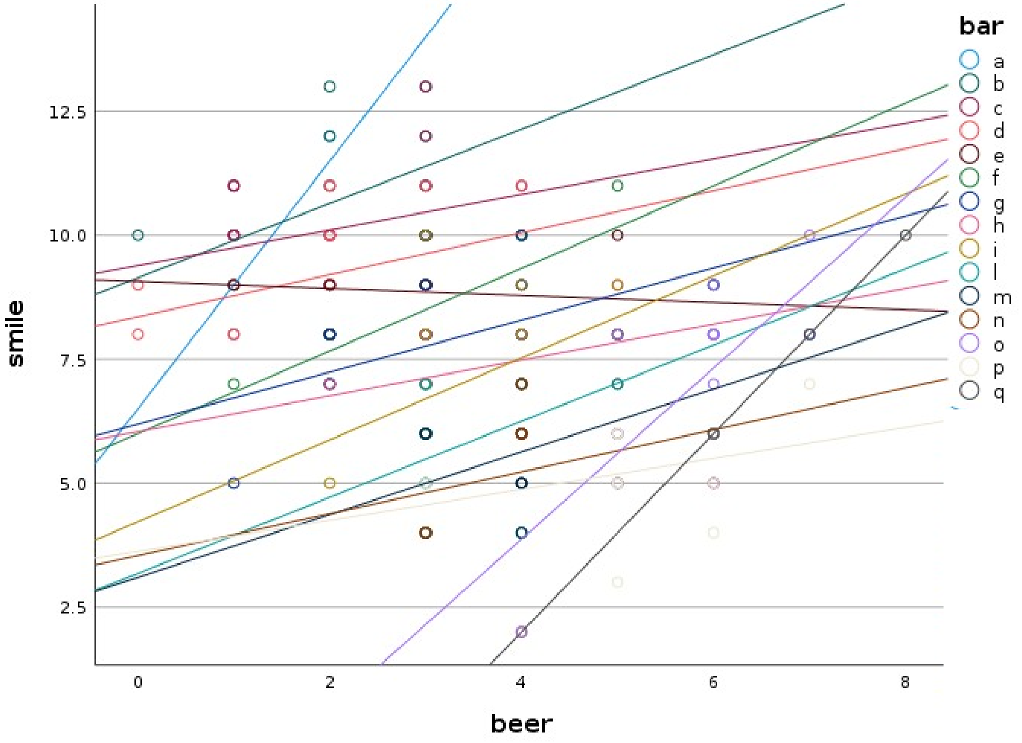
Bar

Intercepts seem to be different from bar to bar



Scatterplot by Bar

Bar



Slopes are all positive

Slopes seem to vary across bars

The Model

- It seems that considering the participants as all equivalent and independent one each other (GLM assumption) does not fit our data

- It seems that a better model should allow each bar (each cluster) to have a different regression line (a different intercept and **b** coefficient)

The Model

- Let's define a model with a regression line for each cluster

y_{ij} Smiles of subject i in the cluster j

$$\hat{y}_{ia} = a_a + b_a \cdot x_{ia}$$
$$\hat{y}_{ib} = a_b + b_b \cdot x_{ib}$$
$$\hat{y}_{ic} = a_c + b_c \cdot x_{ic}$$

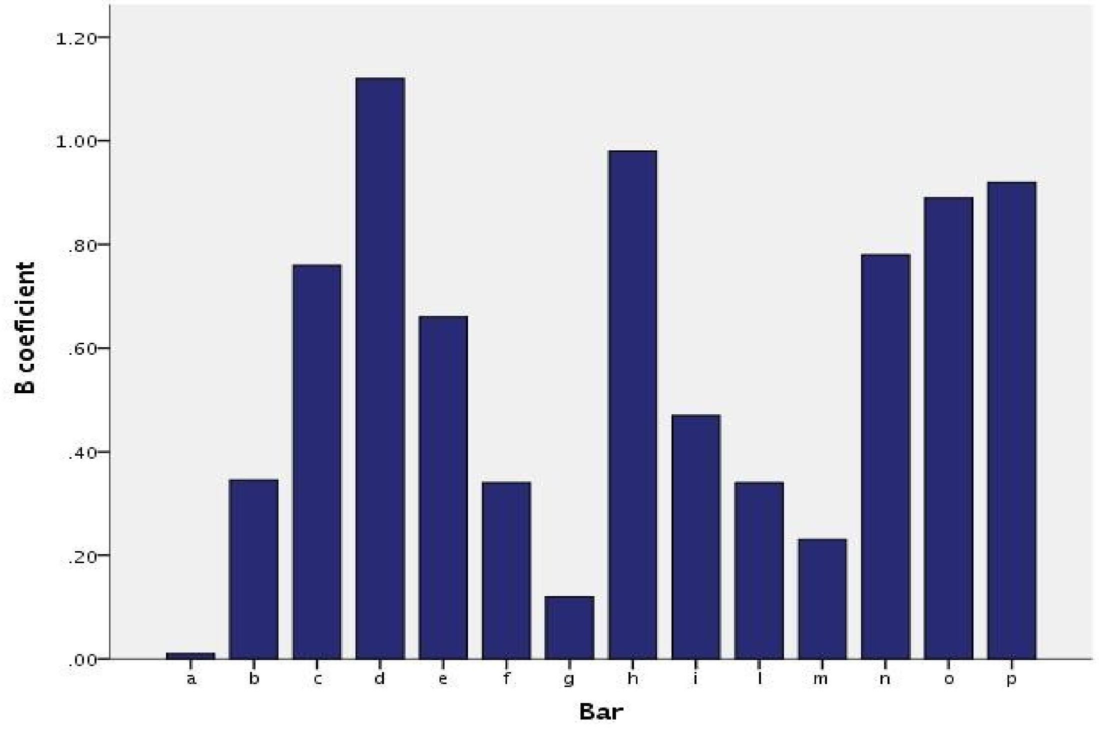
$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

In these regressions the coefficients may vary from cluster to cluster: they are not Fixed

Varying coefficients

- If coefficients may vary, they will have a distribution

A possible distribution of coefficients b estimated for different clusters

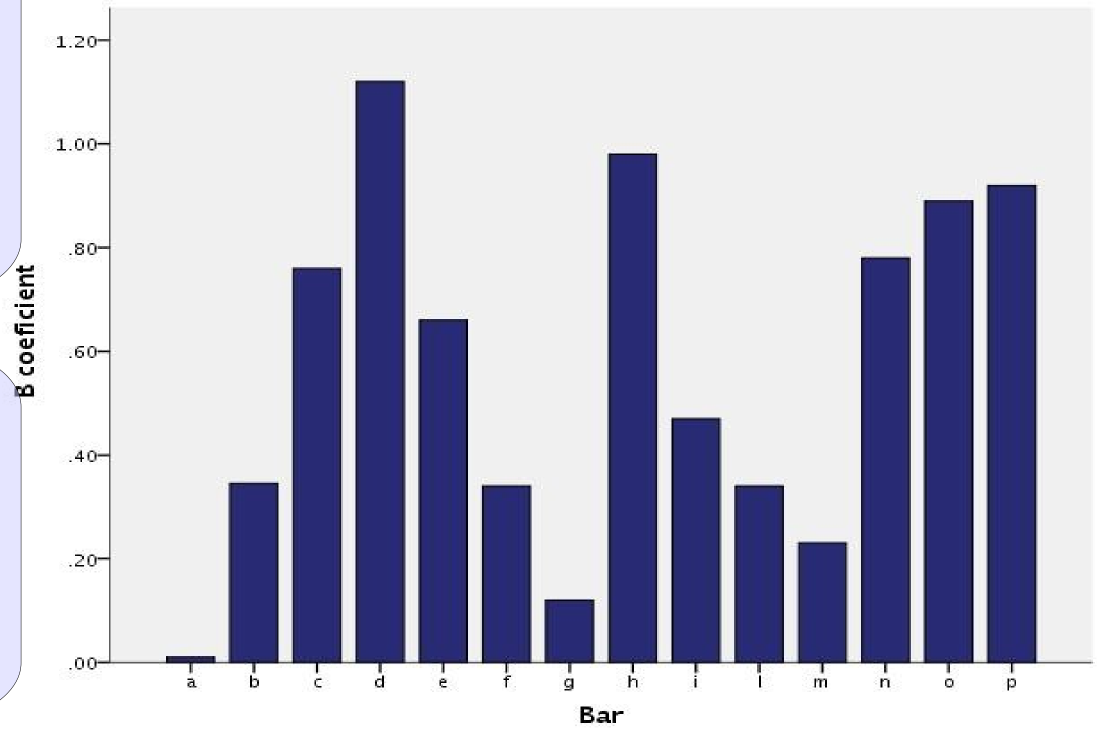


Random coefficients

- Varying coefficients are called random coefficients

Coefficients will exhibit variability

That is: in the population there exist different coefficients, a sample of which we estimated using the clustered data

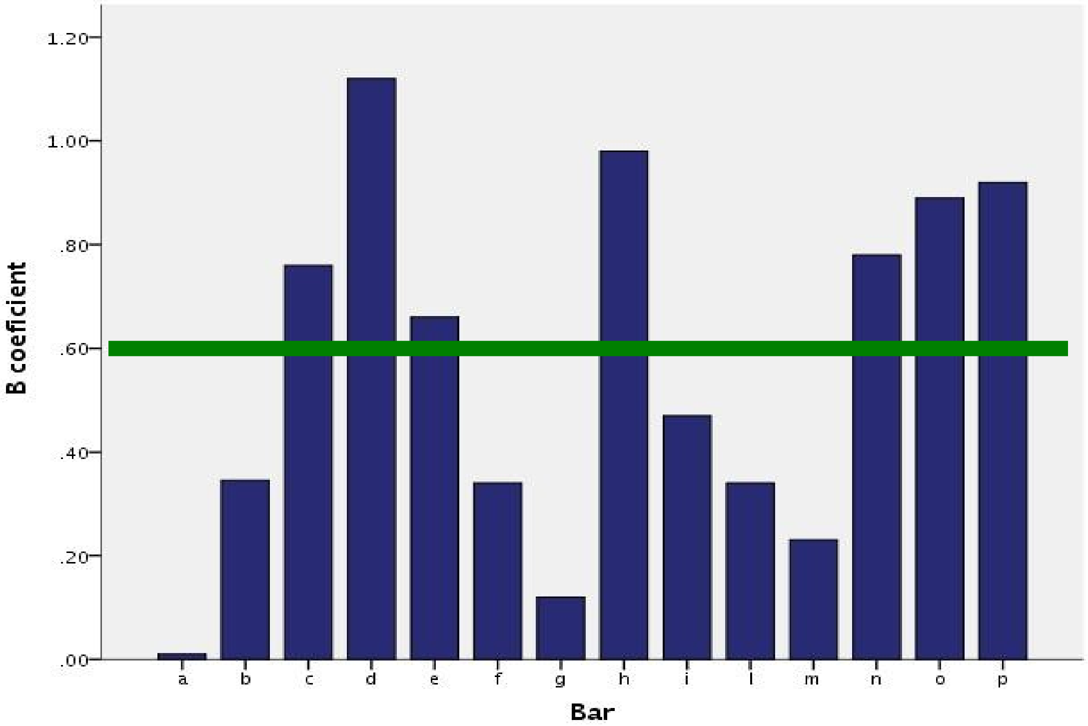


Average of the coefficients

- If coefficients vary as a variable in the population, they will have a mean and a variance, that we can estimate in our data

Mean

$$\bar{b} = \frac{\sum_j b_j}{k}$$



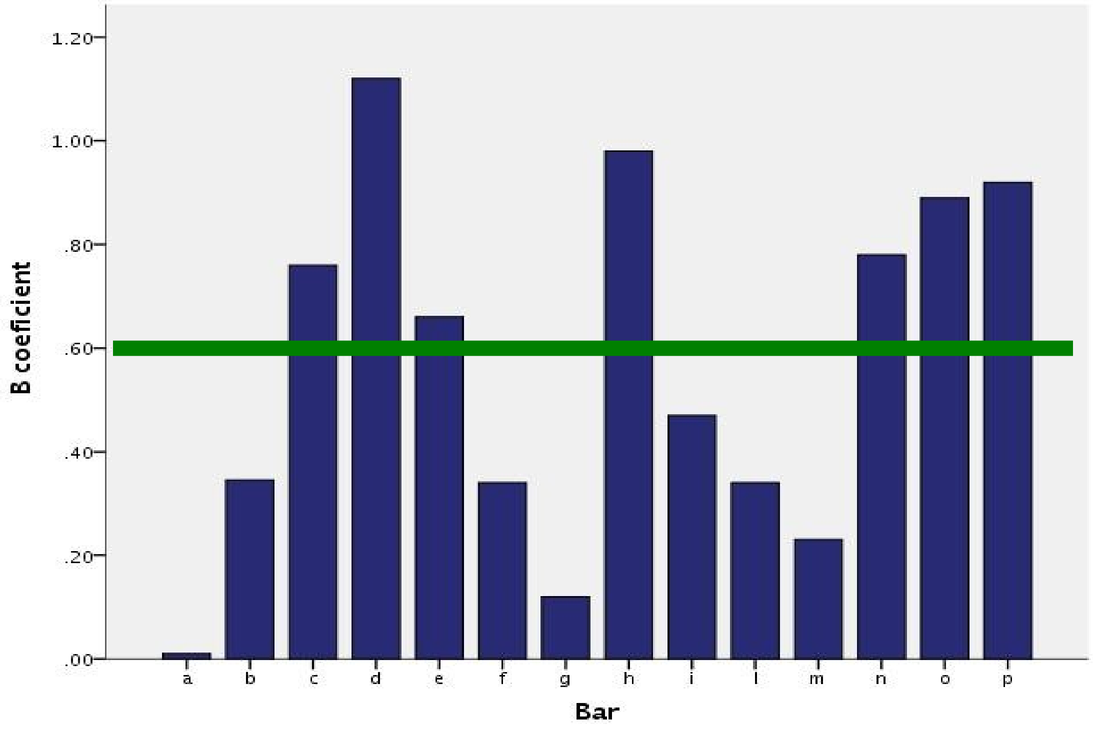
Fixed coefficients

- If coefficients vary as a variable in the population, they will have a mean and a variance, that we can estimate in our data

Mean

$$\bar{b} = \frac{\sum_j b_j}{k}$$

Recall the mean is a fixed parameter for a distribution, and so is the mean of the coefficients: it is a fixed effect



The Model

- We can now define a model with a regression for each cluster and the mean values of coefficients

One regression per cluster

$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

Each coefficient is defined as the deviation from the mean coefficient

$$b'_j = b_j - \bar{b}$$

Overall model

$$\hat{y}_{ij} = a_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

The Model

- We can now define a model with a regression for each cluster and the mean value of coefficients

Overall model

$$\hat{y}_{ij} = a_j + b'_j \cdot x_{ij} + \bar{b} \cdot x_{ij}$$

Random coefficients

Fixed coefficient

The mixed model

- The same goes for the intercepts

One regression per cluster

$$\hat{y}_{ij} = a_j + b_j \cdot x_{ij}$$

Intercepts as deviations from the average intercept

$$a'_j = a_j - \bar{a}$$

Overall model

$$\hat{y}_{ij} = \bar{a} + a'_j + \bar{b} \cdot x_{ij} + b'_j \cdot x_{ij}$$

The mixed model

- We can now define a model with a regression for each cluster and the mean values of coefficients

Overall model

$$\hat{y}_{ij} = \bar{a} + a'_j + \bar{b} \cdot x_{ij} + b'_j \cdot x_{ij}$$

Random coefficients

Fixed coefficients

A GLM which contains both fixed and random effects is called a Linear Mixed Model

GLM as a special case

It is clear that everything we know for the GLM applies here: the GLM is in fact a special case of the LMM, where there are not random effects

LMM

$$\hat{y}_{ij} = \bar{a} + a'_j + \bar{b} \cdot x_{ij} + b'_j \cdot x_{ij}$$

GLM

$$\hat{y}_{ij} = \bar{a} + \bar{b} \cdot x_{ij}$$

The mixed model

- In practice, mixed models allow to estimate the kind of effects we can estimate with the GLM, but they allow the effects to vary across clusters.
- Effects that vary across clusters are called **random effects**
- Effects that do not vary (the ones that are the same across clusters) are said to be **fixed effects**

The mixed model

- To specify a correct model, we only need to understand if there are **clusters of cases** (measures or subjects) and decide which coefficients (intercepts or b coefficients) may vary across those clusters

- The fixed effects of the model are interpreted like in the GLM (regression/ANOVA)

- **Random effects** are generally not interpreted, but we can look at their variance to decide to keep them as random (variance>0) or fix them.

- In this way we take into the account the dependence among data

Building a model

To build a model in a simple way, we need to answer very few questions:

- What is (are) the cluster variable(s)?
- What are the fixed effects?
- What are the random effects?

A clustering variable

- **What is (are) the cluster variable(s)?**
- What are the fixed effects?
- What are the random effects?

    - Any variable that groups observations (cases or measurements) such that scores may be more similar within each group than across groups
    - Any variable whose levels (groups) are a sample of a larger population of levels (groups)
    - Example: bars created groups of scores (participants) that may be more similar within the bar that across bars

Fixed effects

- What is (are) the cluster variable(s)?
- **What are the fixed effects?**
- What are the random effects?

  - Any effect that we are interested in on average (as in a standard ANOVA/Regression)
  - Example: the effect of beer on smiles in general

Fixed effects

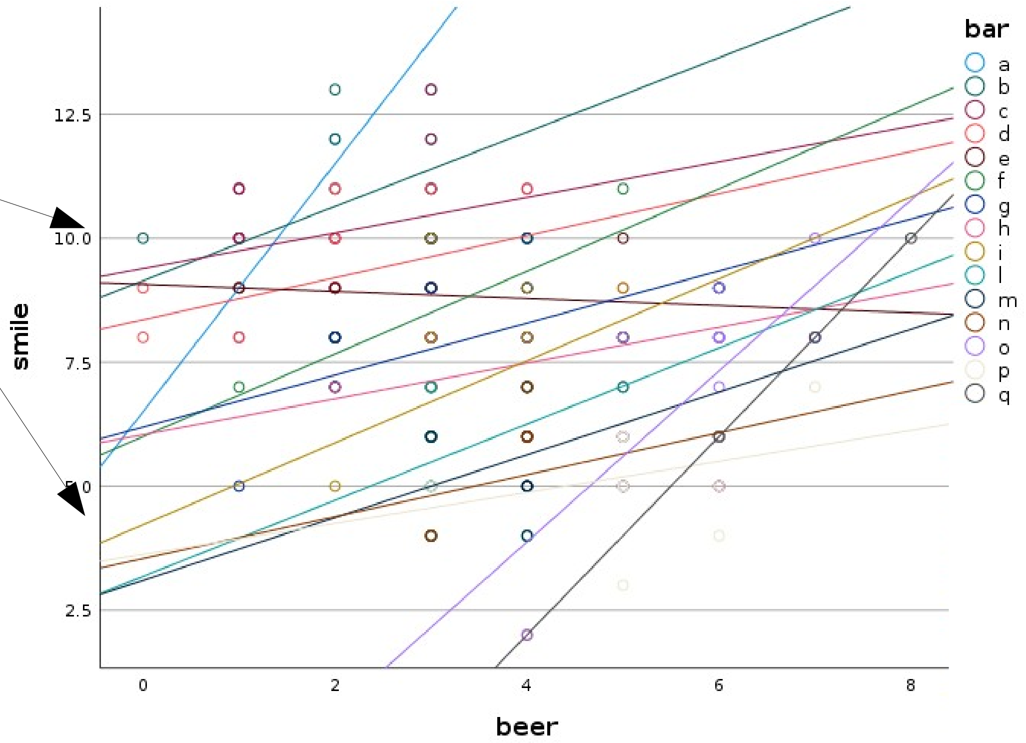
- What is (are) the cluster variable(s)?
- What are the fixed effects?
- **What are the random effects?**
  - Any effect that may vary from cluster to cluster
  - (Thus:) **Any effect that can be computed within each cluster**
  - Example: the intercepts and the effect of beer on smiles each bar

Beers at the bar

We start with a simple model

Bar

Intercepts seem to be different from bar to bar



Beers at the bar

We define a model where each cluster is allow to have a different intercept, the rest of the model is like a standard regression

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

- Fixed effects? Intercept and beer effect
- Random effects? Intercepts
- Clusters? bar

Authors and books may call this model:
Random-intercepts regression
or
Intercepts-as-outcomes model

Beers at the bar

We define a model where each cluster is allow to have a different intercept, the rest of the model is like a standard regression

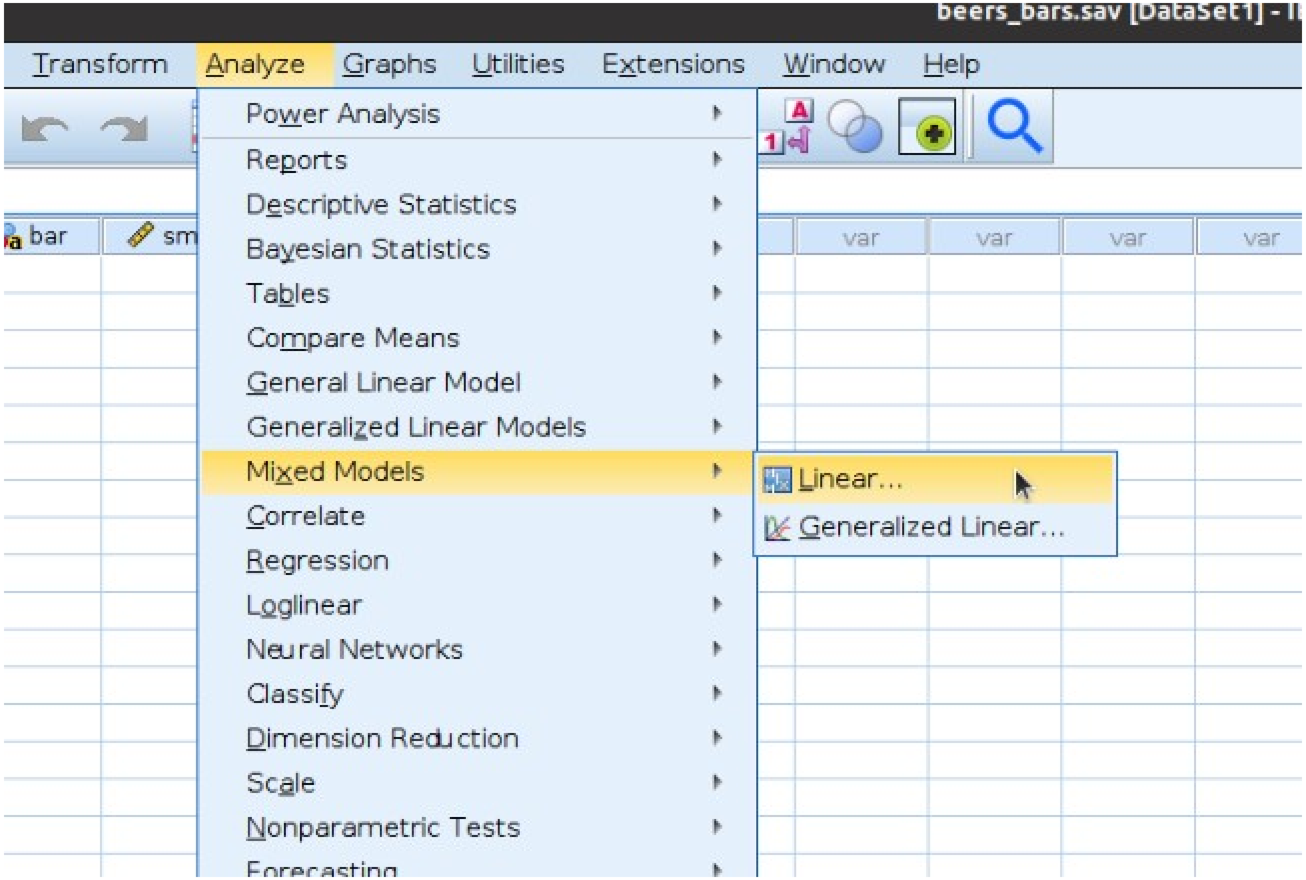
$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

- Fixed effects? Intercept and beer effect
- Random effects? Intercepts
- Clusters? bar

Authors and books may call this model:
Random-intercepts regression
or
Intercepts-as-outcomes model

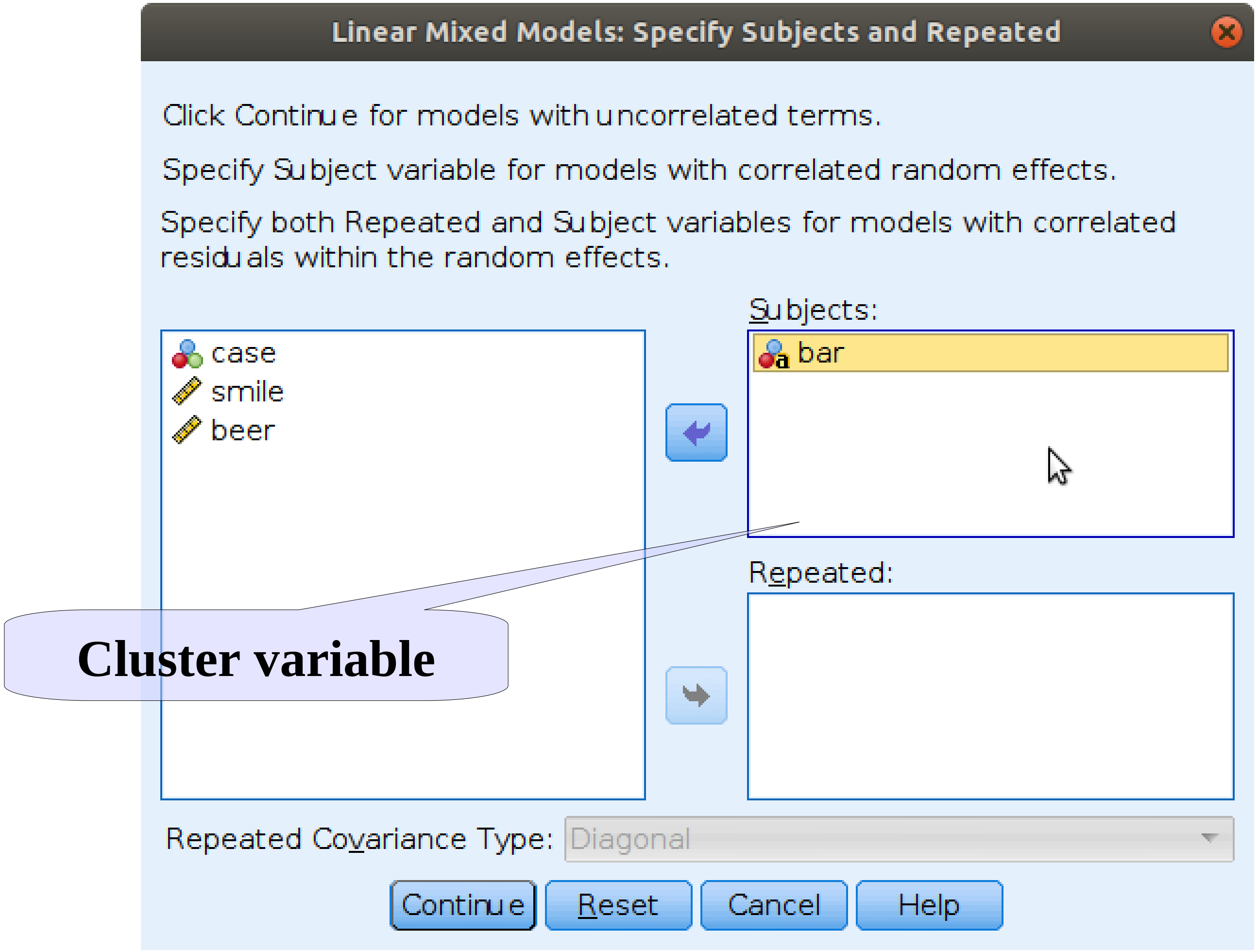
SPSS Input

Analyze → Mixed Models → Linear



SPSS Input

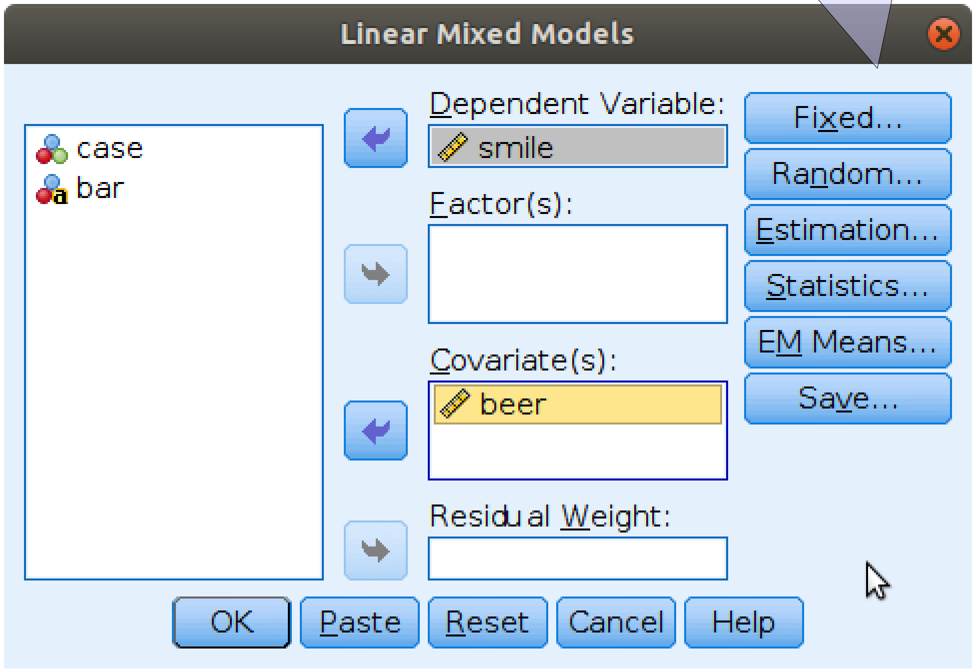
Analyze → Mixed Models → Linear



SPSS Input

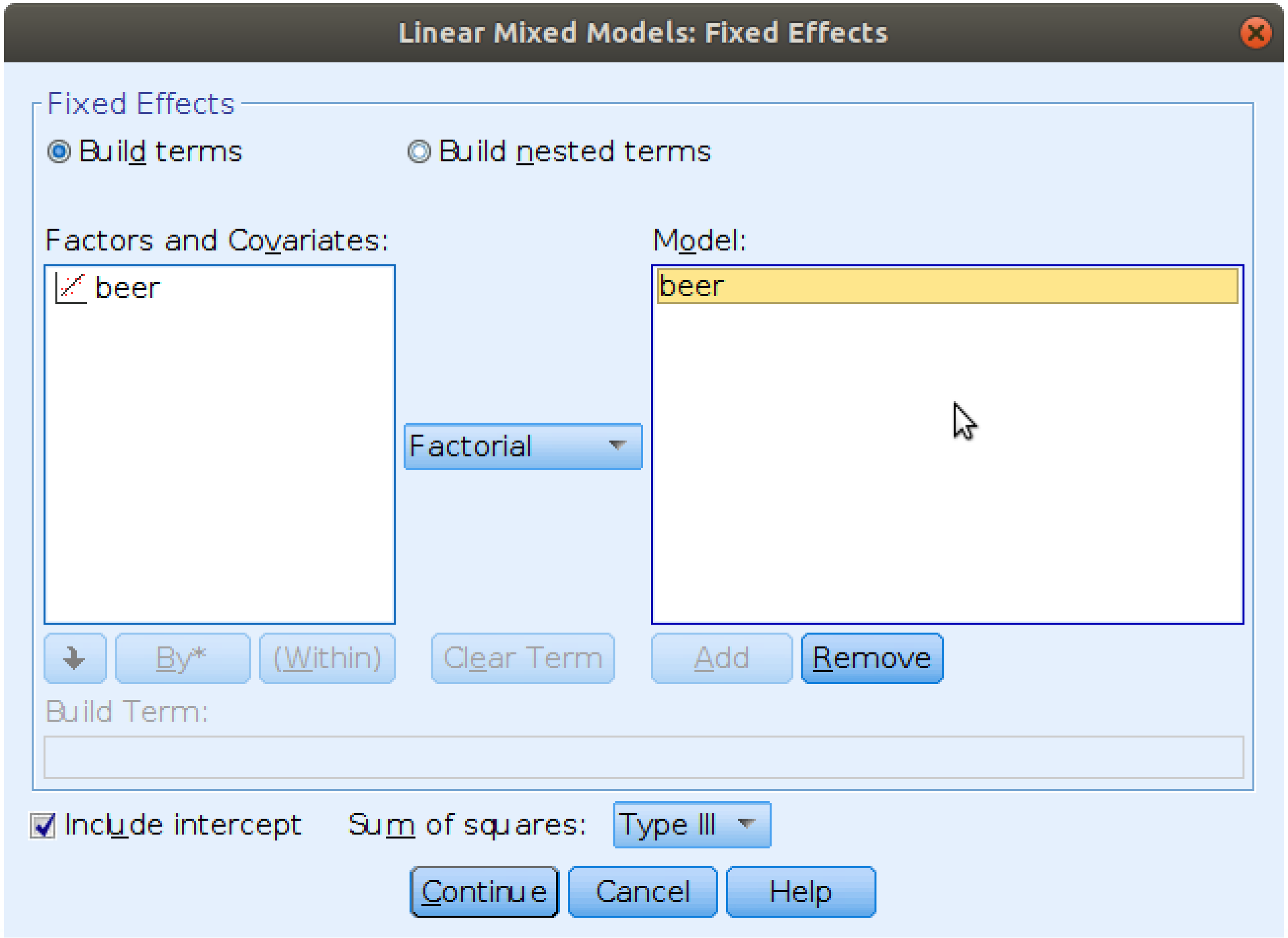
Analyze → Mixed Models → Linear

Then select Fixed



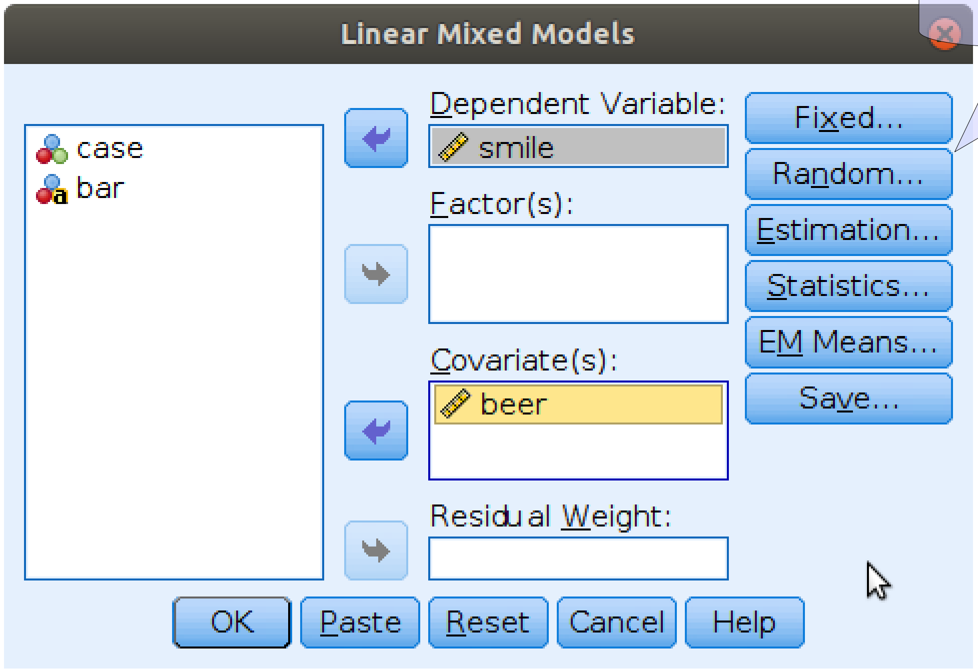
SPSS Input

Analyze → Mixed Models → Linear



SPSS Input

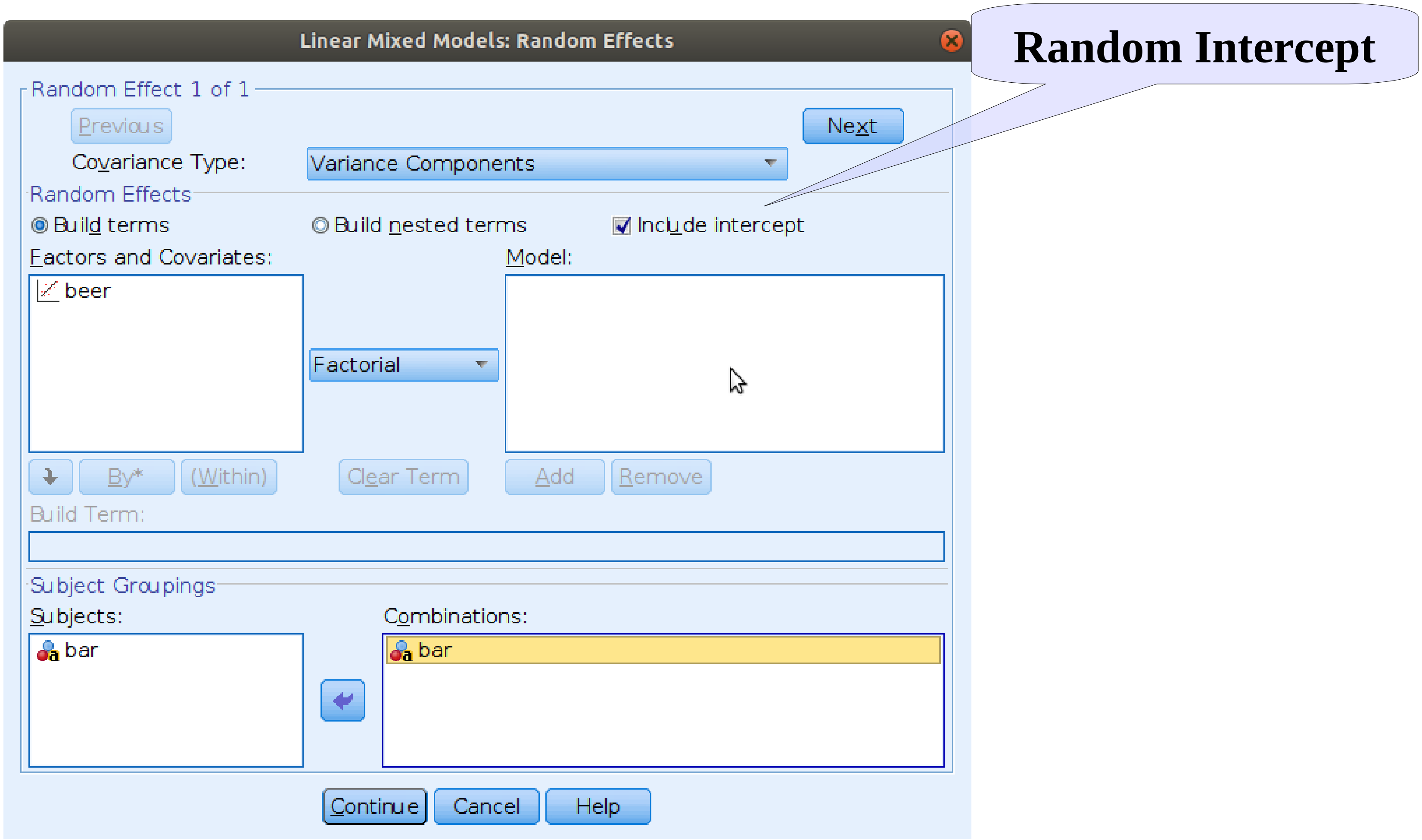
Analyze → Mixed Models → Linear



Then select Random

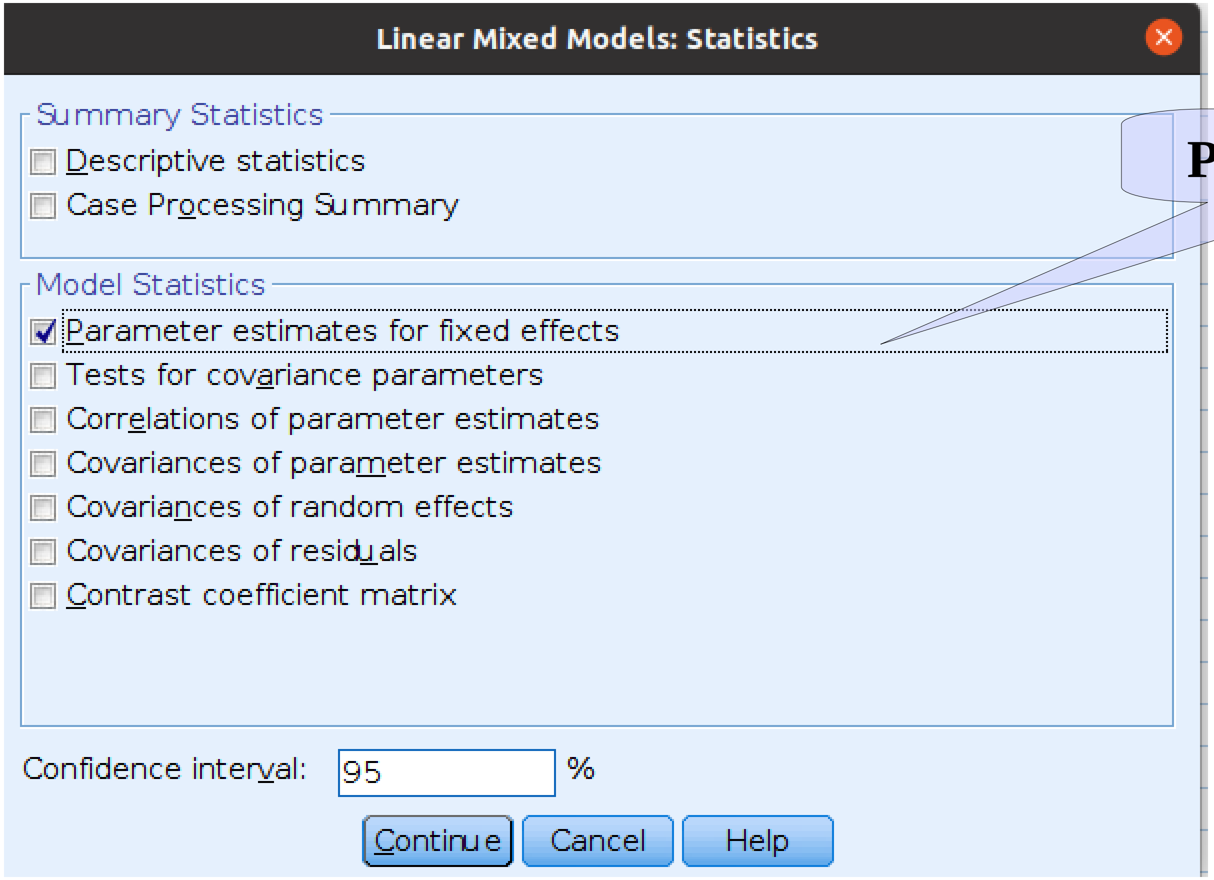
SPSS Input

Analyze → Mixed Models → Linear



SPSS Input

Analyze → Mixed Models → Linear



Print coefficients

SPSS syntax

$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

```
MIXED smile WITH beer by bar
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001) HCONVERGE(0,
    ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=beer| SSTYPE(3)
  /METHOD=REML
 /print solution TESTCOV
  /random intercept   | SUBJECT(bar) COVTYPE(un).
```

Fixed effects (intercept is included by default)

Random effects

Cluster variable

SPSS Output

Let's see if the model is how intended

Model Dimension[a]

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	beer	1		1	
Random Effects	Intercept[b]	1	Variance Components	1	bar
Residual				1	
Total		3		4	

a. Dependent Variable: smile.

b. 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.

OK!

SPSS Output

We then check the variability of the random effects. If there is variability across bars, it means we were right to model the coefficients as random

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter		Estimate	Std. Error
Residual		1.451189	.139257
Intercept [subject = bar]	Variance	5.816514	2.320621

a. Dependent Variable: smile.

Variance greater than 0

SPSS Output

If everything is fine, we interpret the fixed effects as in any other GLM (regression)

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	5.841778	.695573	19.205	8.399	<.001	4.386978	7.296578
beer	.552973	.080740	229.072	6.849	<.001	.393884	.712061

a. Dependent Variable: smile.

Intercept: On average, for zero beers we expect 5.8 smiles

SPSS Output

If everything is fine, we interpret the fixed effects as in any other GLM (regression)

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	5.841778	.695573	19.205	8.399	<.001	4.386978	7.296578
beer	.552973	.080740	229.072	6.849	<.001	.393884	.712061

a. Dependent Variable: smile.

Intercept: On average, as beers increase on 1 unit, we expect smile to increase of .552 smiles

SPSS Output

We also get the overall omnibus tests

Fixed Effects

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	19.205	70.535	<.001
beer	1	229.072	46.906	<.001

a. Dependent Variable: smile.

These are equivalent to the GLM F-tests

Beers at the bar 2

We can now try a model where also the **b** coefficients are allow to vary across clusters

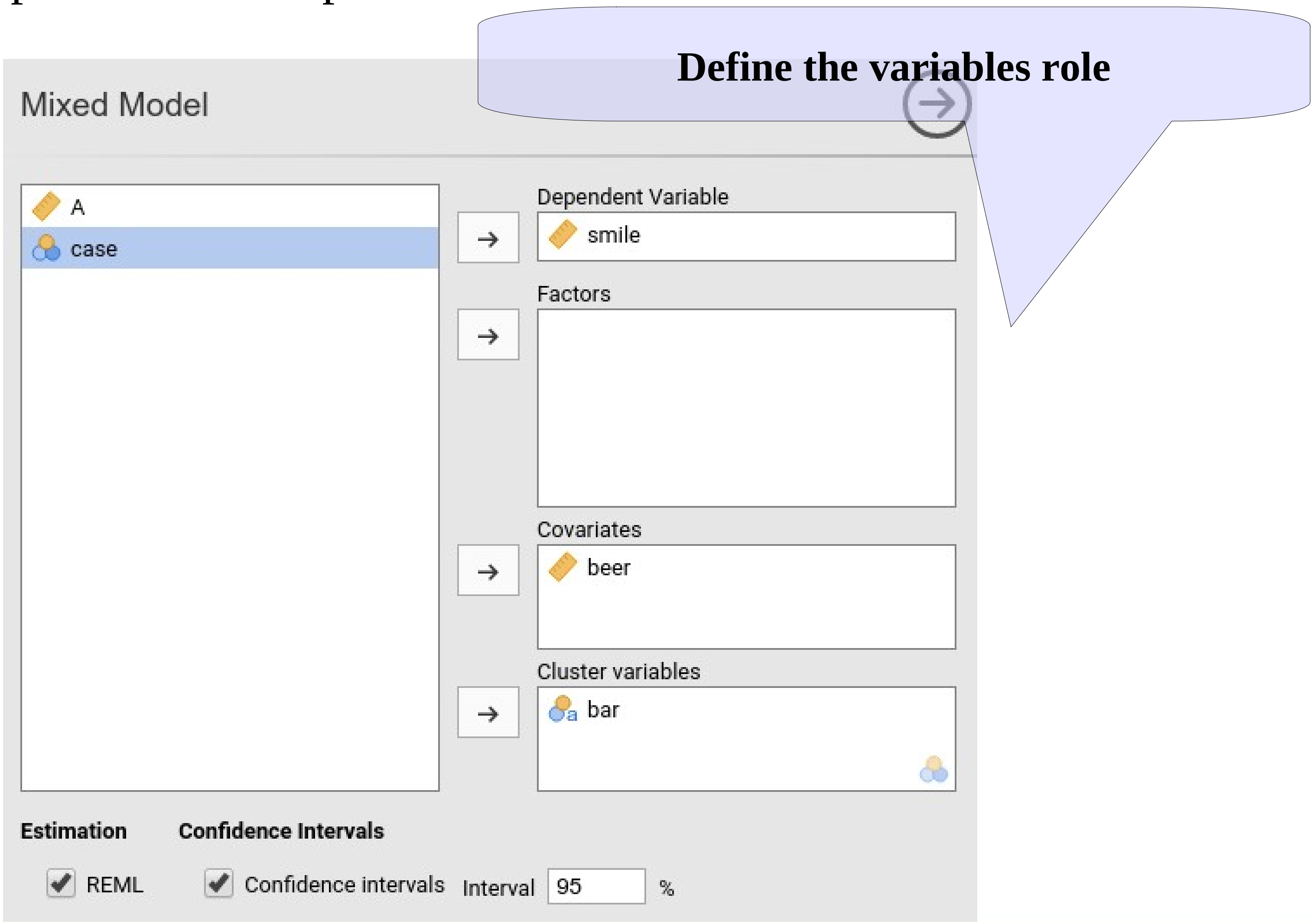
$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + b \cdot x_{ij} + e_{ij}$$

- Fixed effects? Intercept and beer effect
- Random effects? Intercepts and b coefficients
- Clusters? bar

Some authors may call this model:
Random-coefficients regression
or
Intercepts- and Slopes-as-outcomes model

Jamovi

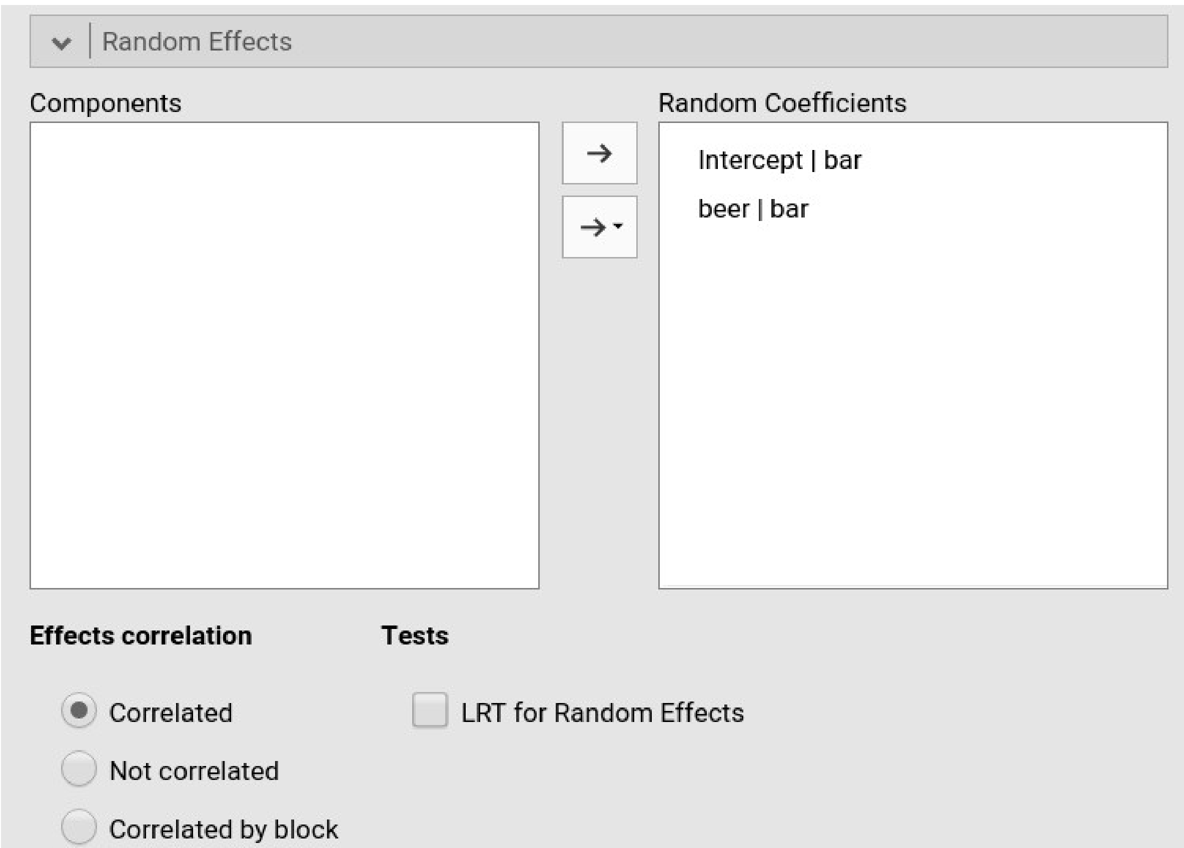
- We keep the same setup for the variables



Jamovi

- We add the effect of beer as a random coefficient

Define the random component



Jamovi

- As soon as you define the random component, you get the results

Model Results

Model Fit

Type	R^2	df	LRT X^2	p
Conditional	0.822	4	203.003	< .001
Marginal	0.090	1	17.016	< .001

[4]

>

R-squared Marginal: How much variance can the fixed effects alone explain of the overall variance

R-squared Conditional: How much variance can the fixed and random effects together explain of the overall variance

Jamovi

- As soon as you define the random component, you get the results

F-test for the main effect of beer

Model Results

Fixed Effect Omnibus tests

	F	Num df	Den df	p
beer	36.057	1	7.234	< .001

Note. Satterthwaite method for degrees of freedom

Jamovi

- As soon as you define the random component, you get the results

coefficients for the main effect of beer

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval Lower	95% Confidence Interval Upper	df	t	p
(Intercept)	7.610	0.633	6.368	8.851	12.928	12.013	< .001
beer	0.555	0.093	0.374	0.737	7.234	6.005	< .001

Jamovi

- As soon as you define the random component, you get the results

Random Components

Groups	Name	SD	Variance	ICC
bar	(Intercept)	2.417	5.842	0.803
	beer	0.167	0.028	
Residual		1.196	1.431	

Note. Number of Obs: 234 , groups: bar 15

Random coefficients variances

Random Parameters correlations

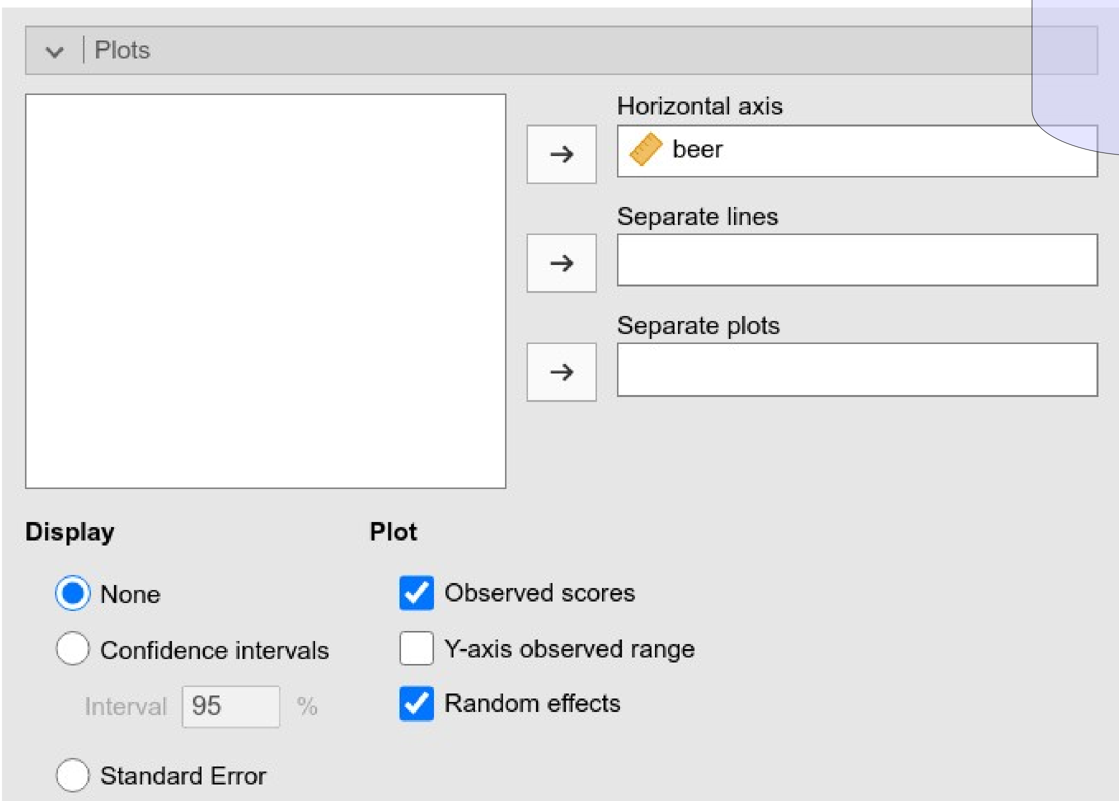
Groups	Param.1	Param.2	Corr.
bar	(Intercept)	beer	−0.766

Random coefficients correlation

Jamovi

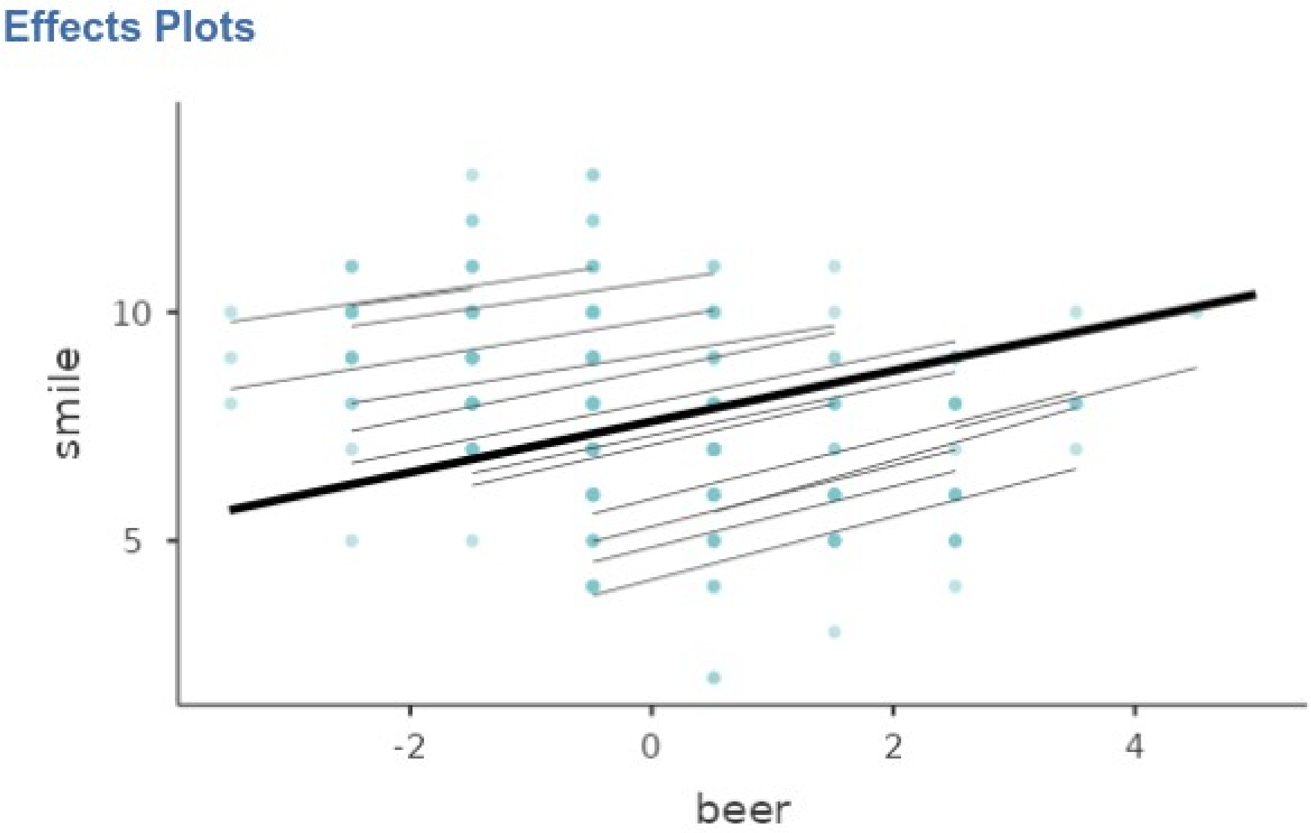
- Jamovi can plot up to a 3-way interaction

Plot



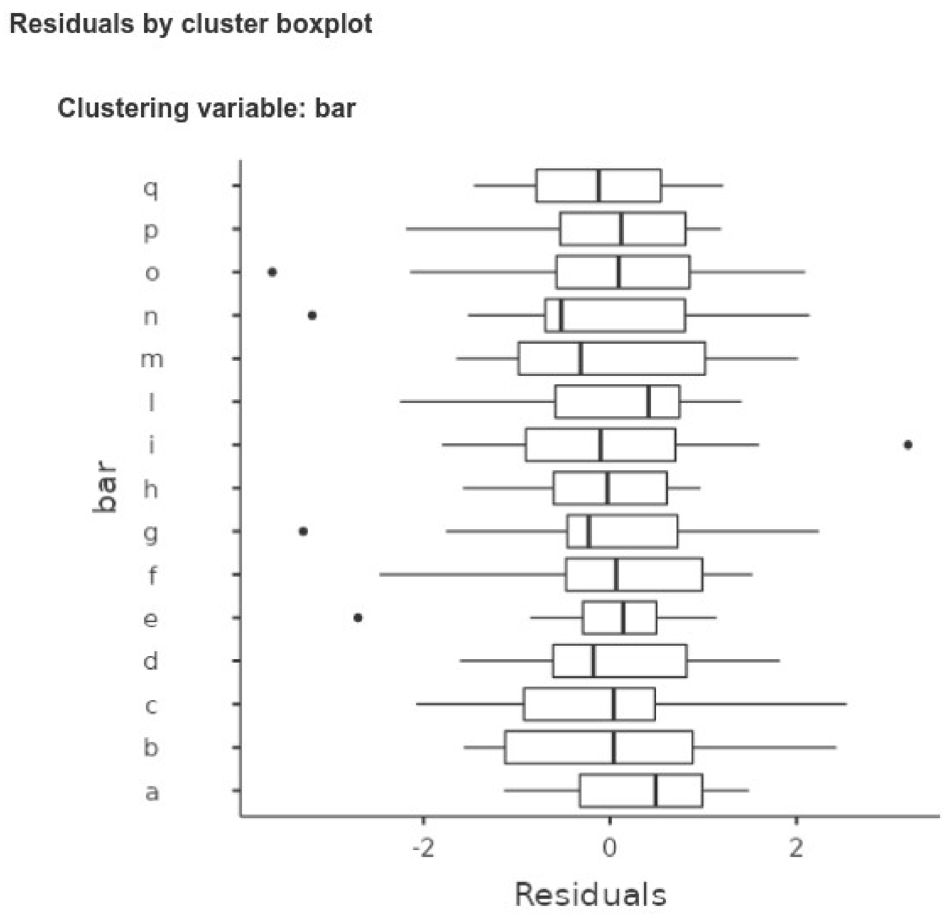
Jamovi

- In this case is fixed and random effects regression lines



Residuals plot

- We can also look at the residual plot by cluster



Mixed Linear Models

- With the mixed model one can take into the account dependency among measures (within clusters) almost in any situation
- It allows applying the GLM logic to a broader range of designs
- Any kind of independent variables
- Generalizes to the generalized linear model (logistic etc)
- Efficient handling of missing values
- **Multi-level research designs**
- **Repeated measures designs**

Multi-level models

Multi-level models

- The Multi-level model i**s not** a “statistical technique”!
- The Multi-level model is **an approach** to analyze multi-level designs
- **The multi-level model is estimated using a mixed model**
- What is peculiar**:**
  - The importance of the clustering variables (higher levels)
  - The research questions
  - The cluster level is called group level (*group=cluster in this terminology*)

Example

- Assume we have a multi-country research, in which we measured individuals (people) *charity contribution* and their *income* (*individual level*).
- We can have information about country taxes regulations (*country level*)
- We are interested on the relationship between *contribution* and *income* at the **individual level and at the country level**

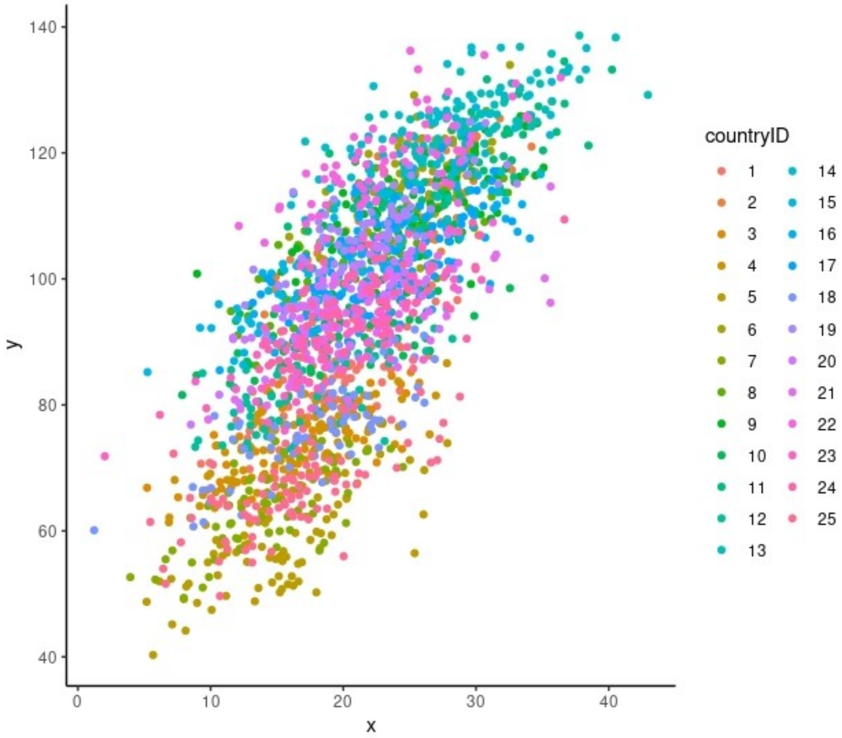
Questions

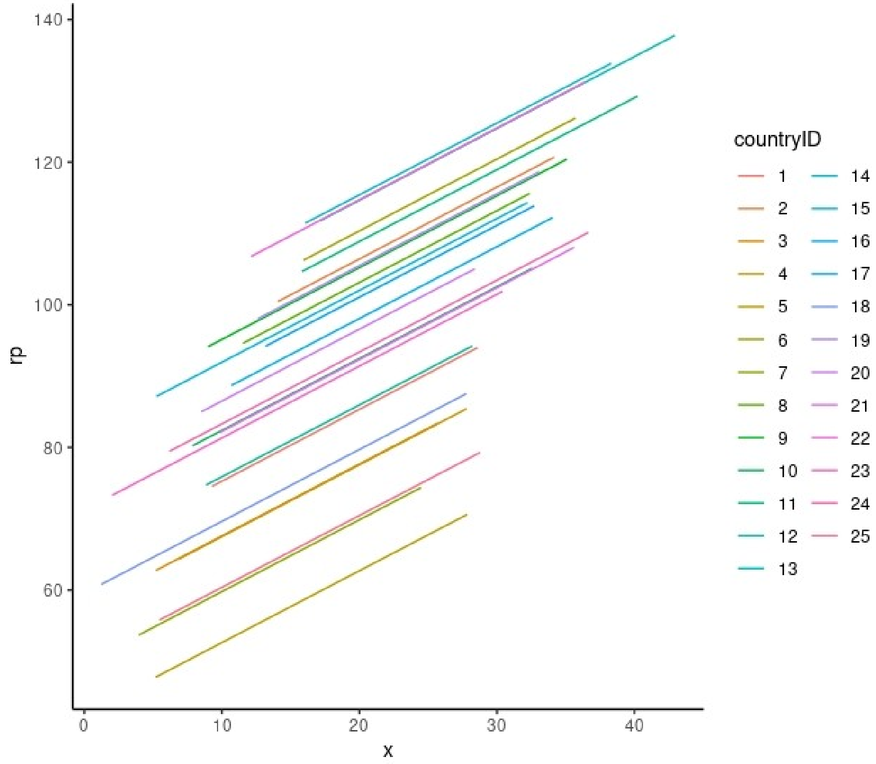
- We are interested on the relationship between cooperation and income at the **individual levels and at the country level**
- Independently of the country: Does people with higher income contribute more? (*individual level effect*)
- Independently of the people: Do countries with average higher income show higher average contribution? (*country level effect*)

Structure vs aim

- If we only look at the data structure, we are in the beers at bars example

- But the research aim is different

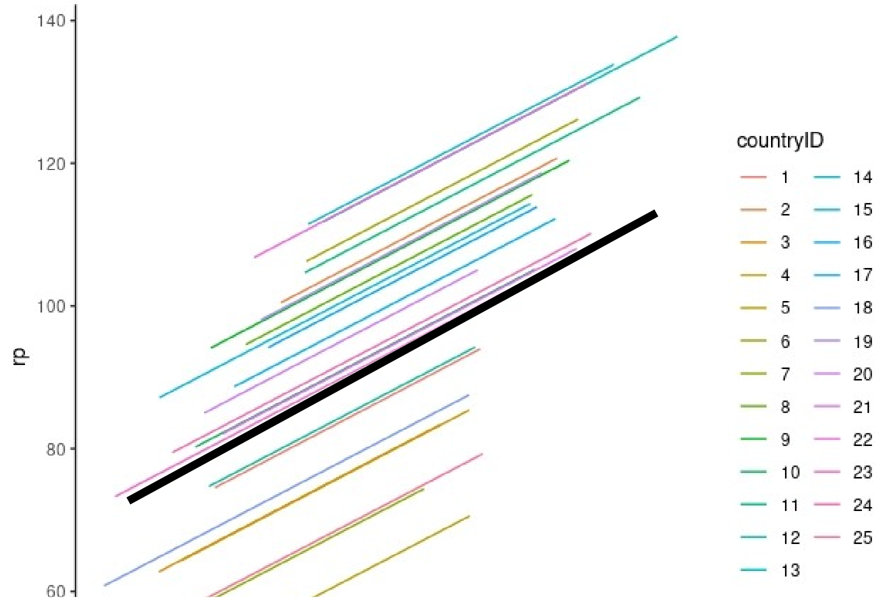




Individual level

- If we fit a model like *beer at bars*, we only get the individual level effect, averaged across countries

Independently of the country: Does people with higher income contribute more? (*individual level effect*)



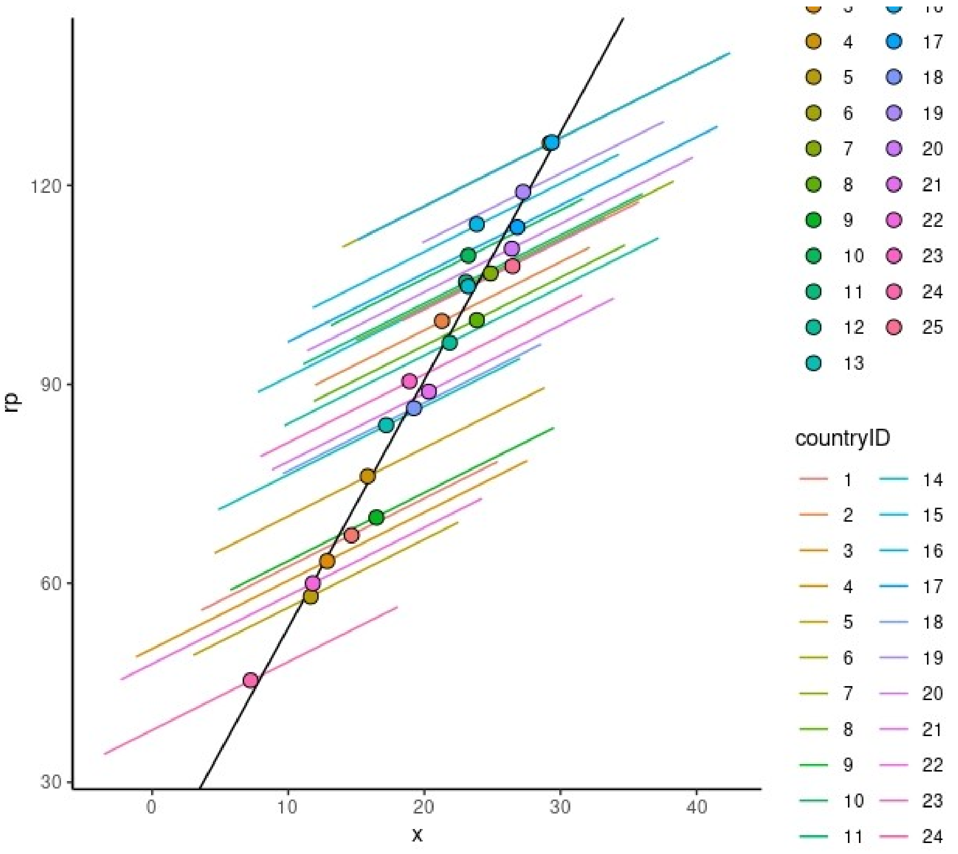
Fixed Effects Parameter Estimates

			95% Confidence Interval				
Names	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	95.73	3.0129	89.830	101.64	24.0	31.8	< .001
x	1.01	0.0235	0.964	1.06	1928.6	43.0	< .001

Country level

- But income may have an effect also at the country (second) level

Independently of the people: Do countries with average higher income show higher average contribution? (*country level effect*)



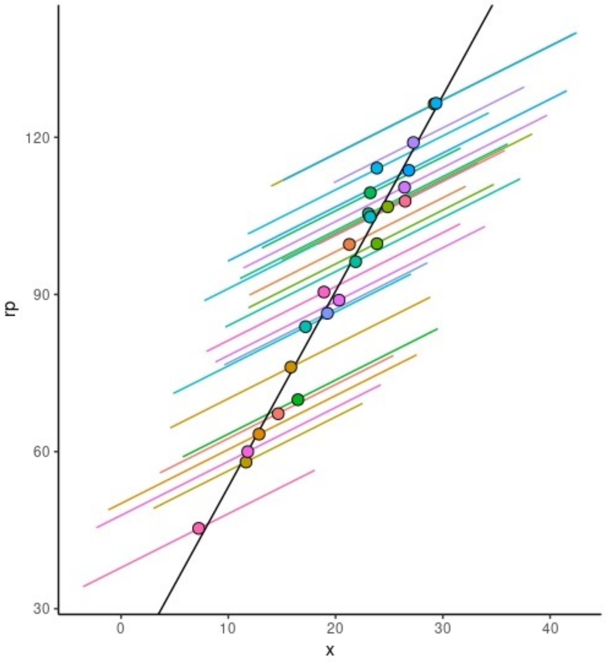
Country vs individual level

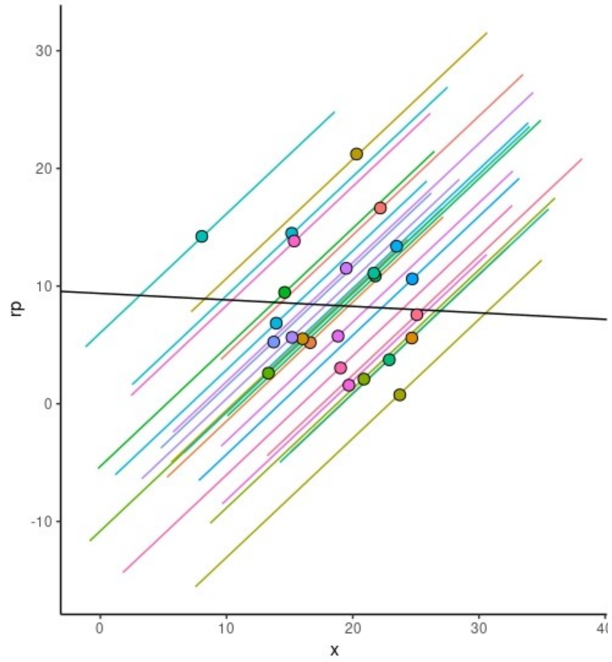
- The effect of a variable at each level is **independent** of the effect at any other level

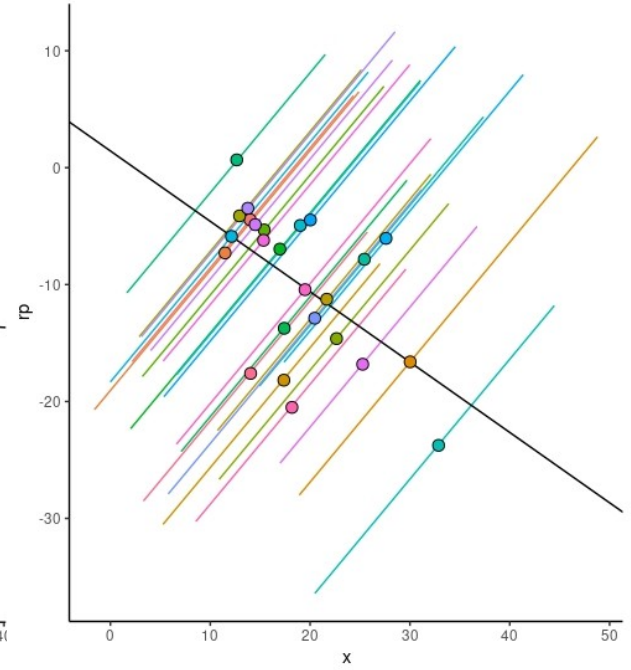
Individual +, country +

individual +, country 0

individual +, country -







The mixed model

- To capture the effect of countries (second level) we should include the country levels means
- To make it independent of people levels, we group-center individual level **x**

$$\hat{y}_{ij} = \bar{a} + a'_j + b_1 \cdot (x_{ij} - \bar{x}_j) + b_2 \cdot \bar{x}_j$$

Group centered x (country centered)

Group mean (country mean)

Coefficients

- The model returns the effects at level 1 (individuals) and level 2 (country)

$$\hat{y}_{ij} = \bar{a} + a'_j + b_1 \cdot (x_{ij} - \bar{x}_j) + b_2 \cdot \bar{x}_j$$

Independently of the country: Does people with higher income contribute more?

Do countries with average higher income show higher average contribution?

Data

- We simply compute two new variables: the group centered x and the group means

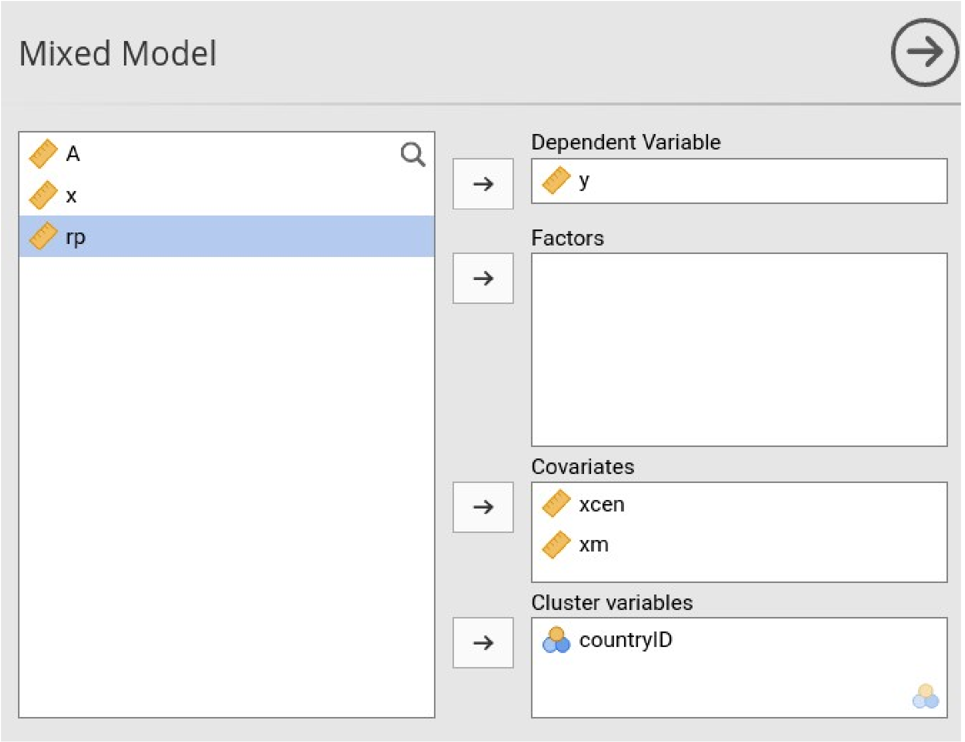
$$xcen = \left(x_{ij} - \bar{x}_j \right)$$

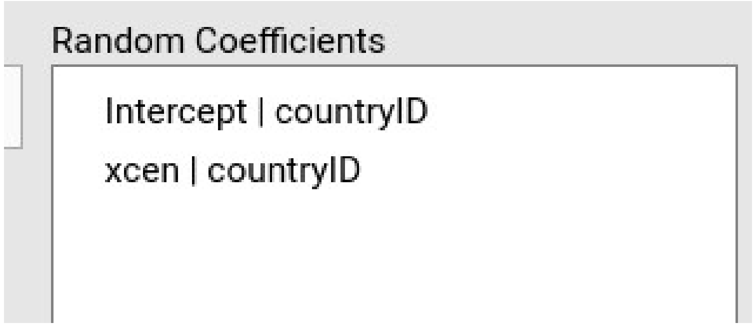
$$xm = \bar{x}_j$$

countryID	x	y	xm	xcen
10	27.128	98.083	20.547	6.581
10	21.709	99.612	20.547	1.163
10	20.586	94.523	20.547	0.040
10	14.580	81.559	20.547	-5.967
10	28.697	106.248	20.547	8.150
10	8.937	84.683	20.547	-11.609
10	17.241	94.658	20.547	-3.306
10	18.252	84.677	20.547	-2.295
10	18.913	90.324	20.547	-1.633
11	33.202	129.347	28.333	4.870
11	23.598	114.079	28.333	-4.735
11	23.746	109.661	28.333	-4.587
11	26.870	113.275	28.333	-1.463
11	25.305	116.171	28.333	-3.028
11	31.789	114.460	28.333	3.456
11	18.956	119.912	28.333	-9.377
11	32.524	123.181	28.333	4.191

Model

- We use them as independent variables





Results (fixed effect)

- And interpret the coefficients accordingly

Fixed Effects Parameter Estimates

Names	Estimate	SE	95% Confidence Interval Lower	Upper	df	t	p
(Intercept)	95.65	1.3013	93.104	98.21	23.1	73.5	< .001
xcen	1.01	0.0278	0.954	1.06	21.8	36.3	< .001
xm	4.37	0.3126	3.757	4.98	23.1	14.0	< .001

Within each country: as people income increases of 1 unit, contribution increases of **1.01**

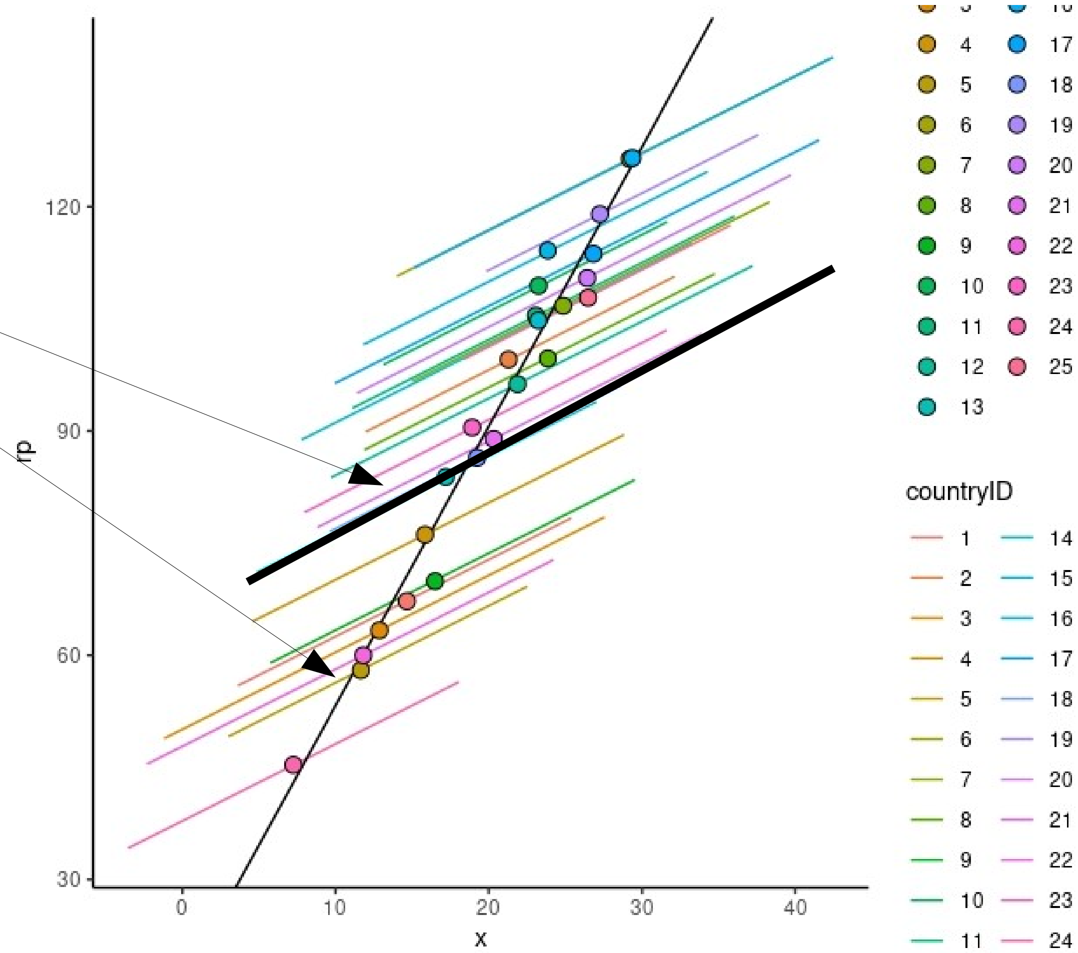
Across countries: As the average income of a country increases 1 unit, the average contribution increases of **4.37** units

Results (fixed effect)

- And interpret the coefficients accordingly

Fixed Effects Parameter Estimates

Names	Estimate	SE
(Intercept)	95.65	1.3013
xcen	1.01	0.0278
xm	4.37	0.3126



Multi-level models

- **The multi-level model is estimated using a mixed model**
- What is peculiar**:**
  - We want to estimate predictors effects at each level
  - We want to estimate higher level effect over and beyond lower level effect (contextual effects)

Mixed Linear Models

- With the mixed model one can take into the account dependency among measures (within clusters) almost in any situation
- It allows applying the GLM logic to a broader range of designs
- Any kind of independent variables
- Generalizes to the generalized linear model (logistic etc)
- Efficient handling of missing values
- Multi-level research designs
- **Repeated measures designs**

Repeated Measures Anova as a linear mixed model

A repeated measures design

- Consider now a classical repeated measure design (within-subjects) the levels of the WS IV (5 different trials) are represented by different measures taken on the same person

trial

		1	2	3	4	5
Participants	1	Y11	Y21	Y31	Y41	Y51
	2	Y12	Y22	Y32	Y42	Y52
	3	Y13	Y23	Y33	Y43	Y53
	….					
	N	Y1n	Y2n	Y3n	Y4n	Y5n

Standard file format

- As for many applications of the repeated-measure design, each level of the WS-factor is represented by a column in the file

One participant, one row

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

1 : group

	group	err_t0	err_t1	err_t2	err_t3	err_t4	x
1	1	.14	.22	.439	.27	.01	-.04
2	1	.43	.52	.492	.48	.43	-.36
3	1	.61	.43	.446	.51	.57	-1.77
4	0	.29	.70	1.000	.89	.75	1.63
5	1	.16	.49	.500	.56	.29	-.32
6	0	.70	.36	.573	.57	.69	-1.16
7	0	.35	.51	.572	.46	.77	-.87
8	1	.45	.49	.545	.41	.43	-1.79
9	1	.05	.55	.333	.54	.53	1.01
10	1	.10	.35	.358	.57	.67	.58
11	0	.14	.45	.373	.25	.29	-.88
12	0	.04	.74	.541	.53	.35	-.27
13	1	.62	.73	.529	.31	.48	1.36
14	1	.15	.22	.101	.17	.17	-.32
15	0	.72	.55	.568	.53	.57	-.98

Long file format

- For the mixed model we need to tabulate the data as if they came from a between-subject design

One measure, one row

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities

	id	group	x	trial	error	va
1	1	1	-.04	1	.14	
2	1	1	-.04	2	.22	
3	1	1	-.04	3	.44	
4	1	1	-.04	4	.27	
5	1	1	-.04	5	.01	
6	2	1	-.36	1	.43	
7	2	1	-.36	2	.52	
8	2	1	-.36	3	.49	
9	2	1	-.36	4	.48	
10	2	1	-.36	5	.43	
11	3	1	-1.77	1	.61	
12	3	1	-1.77	2	.43	
13	3	1	-1.77	3	.45	
14	3	1	-1.77	4	.51	

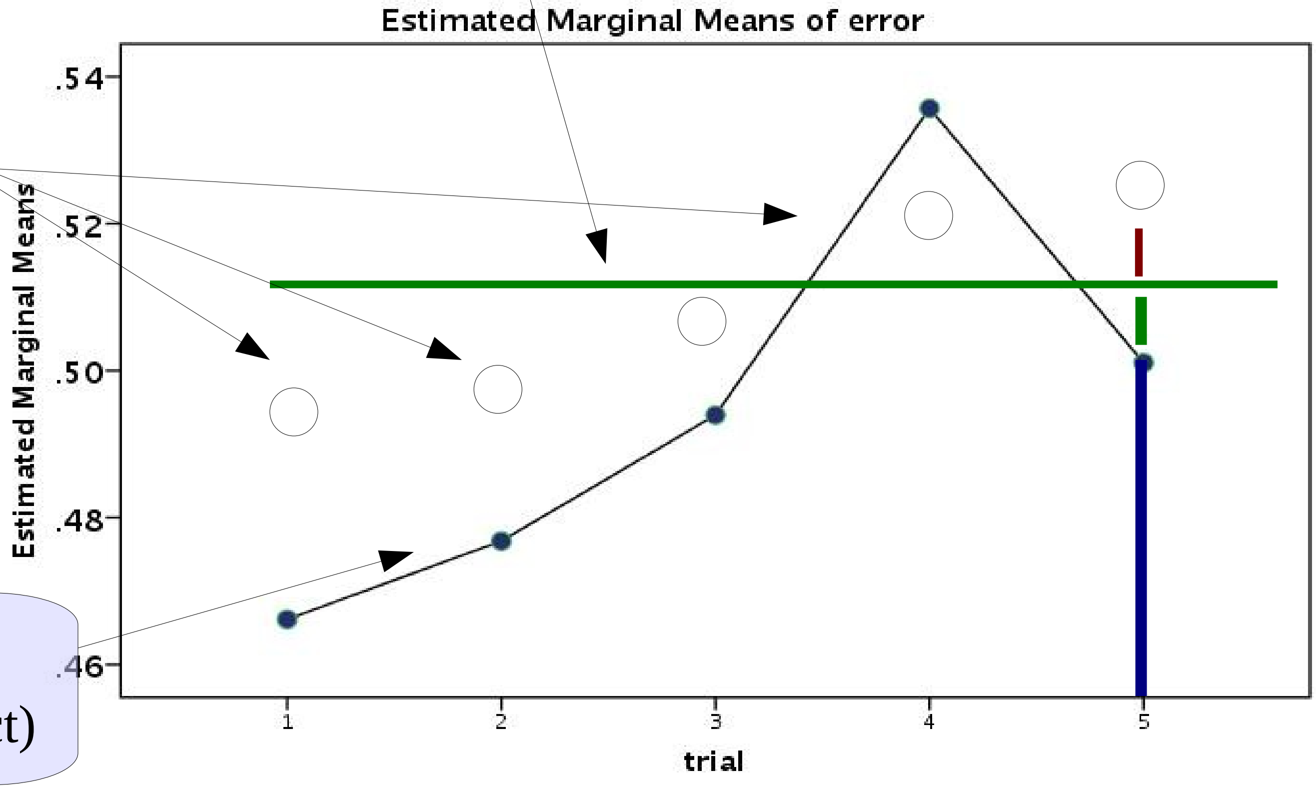
Participant scores

Plot for 1 participant

Participant average trait

Participant scores

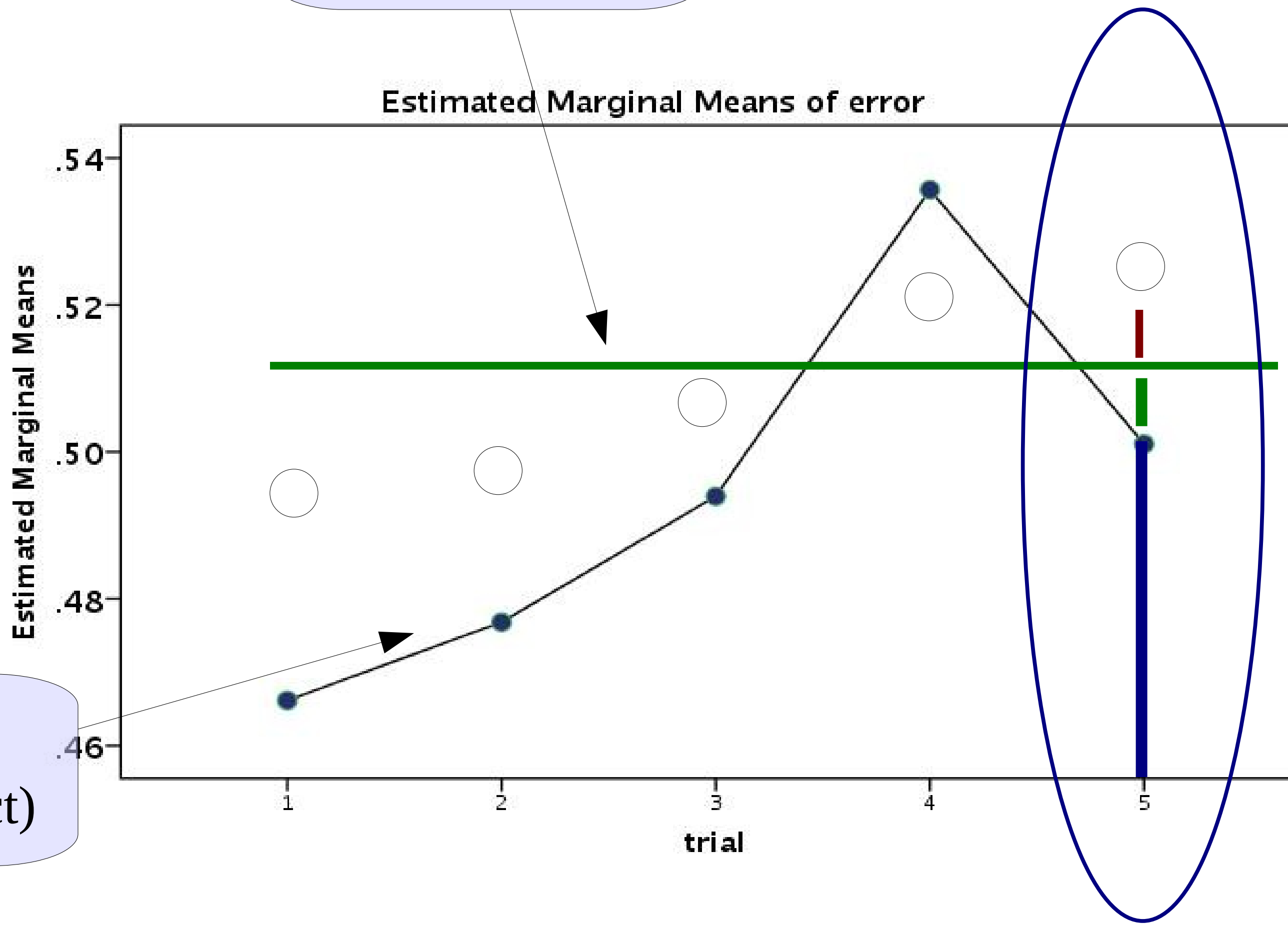
Averages of the sample (fixed effect)



Where does the score come from?

Plot for 1 participant

Participant average trait



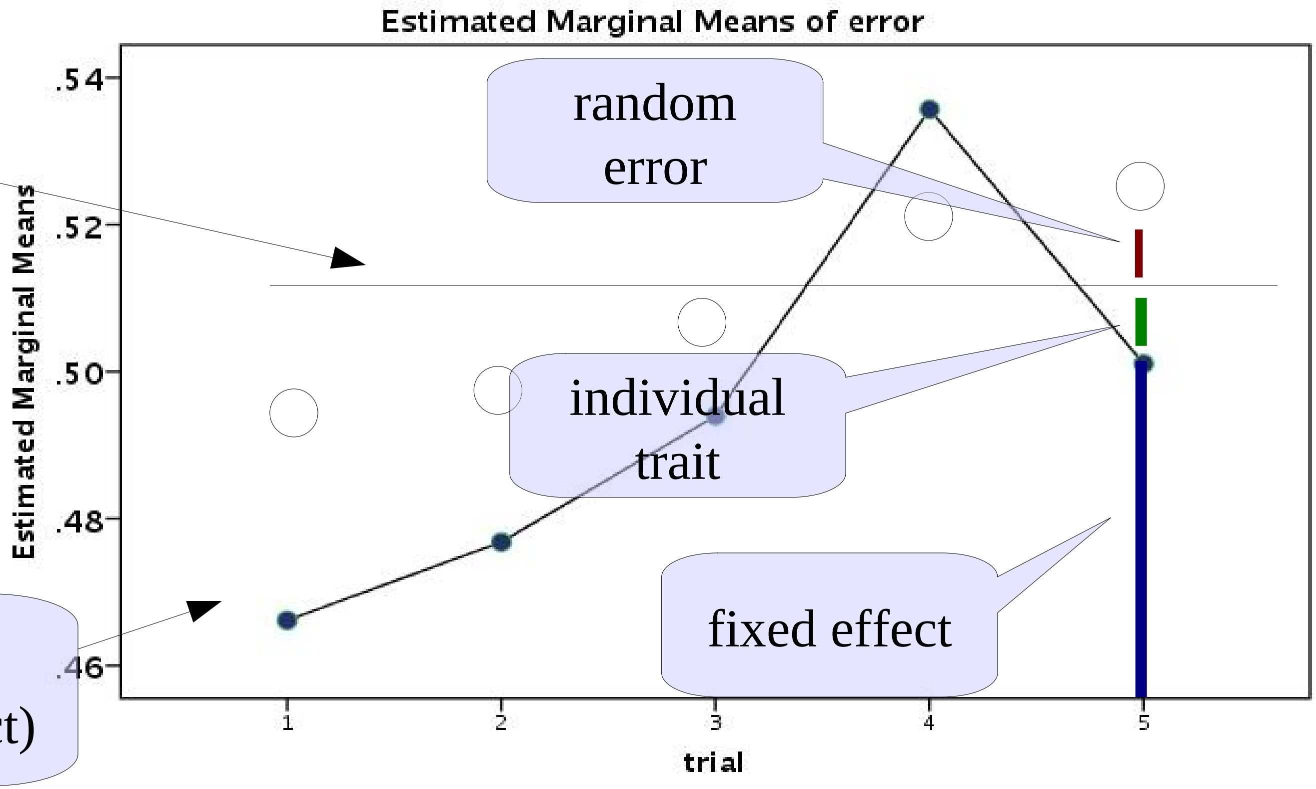
Averages of the sample (fixed effect)

Participant component

Plot for 1 participant

Participant individual trait

Averages of the sample (fixed effect)



Solution

Thus, we should consider an extra residual term which represents participants individual characteristic. This term is the same within each participant

one participant
one trait

$$
\left.\begin{aligned}
Y_{11} &= a + b_1 \cdot T_1 + u_1 + e_{11} \\
Y_{21} &= a + b_2 \cdot T_2 + u_1 + e_{21} \\
Y_{31} &= a + b_3 \cdot T_3 + u_1 + e_{31}
\end{aligned}\right\}
$$

Each score,
one residual

Average effects
of trials

$\ldots$

$$
\left.\begin{aligned}
Y_{1j} &= a + b_1 \cdot T_1 + u_j + e_{1j} \\
Y_{2j} &= a + b_2 \cdot T_2 + u_j + e_{2j} \\
Y_{3j} &= a + b_3 \cdot T_3 + u_j + e_{3j}
\end{aligned}\right\}
$$

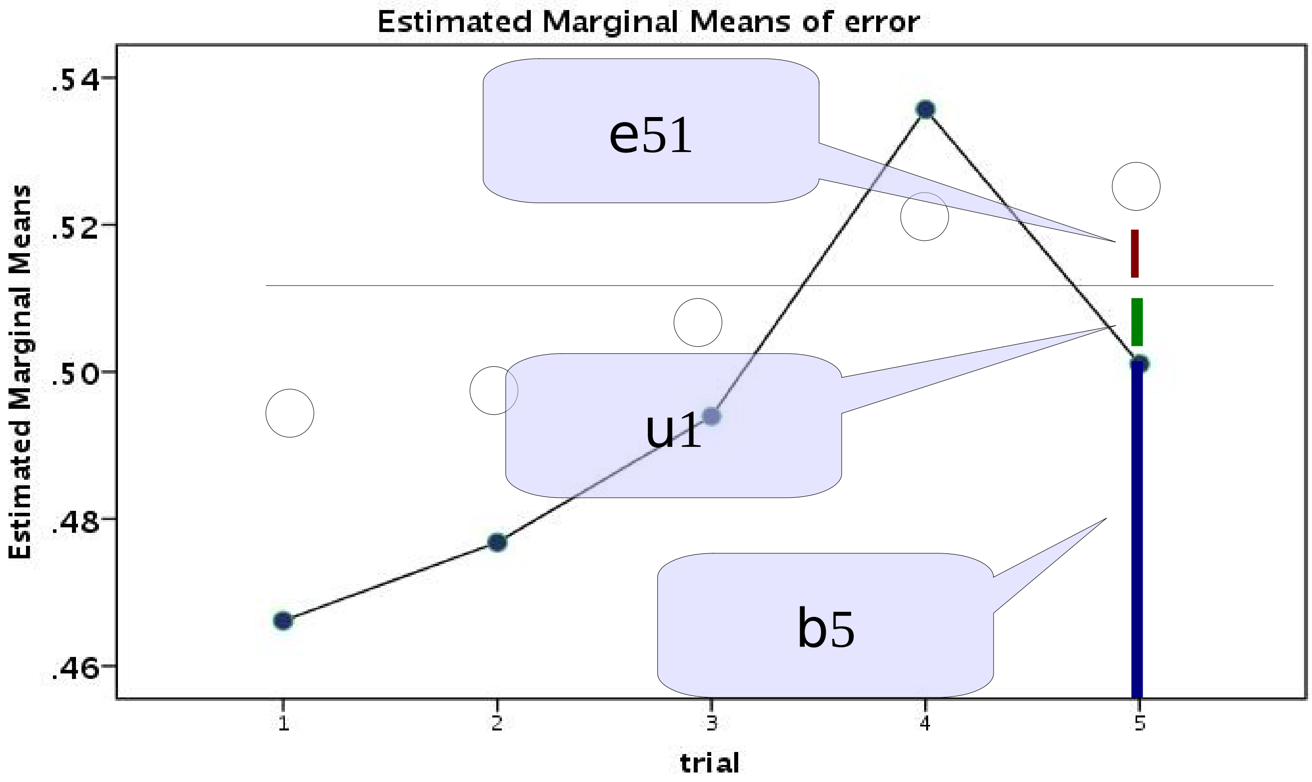
Each score,
one error

One participant
one trait

We assume the 5 trials are dummy coded

Participant component

$$Y_{51} = a + b \cdot T_5 + u_1 + e_{51}$$



Building the model

We translate this in the standard mixed model

$$Y_{ij} = a + b' \cdot T_i + u_j + e_{ij}$$

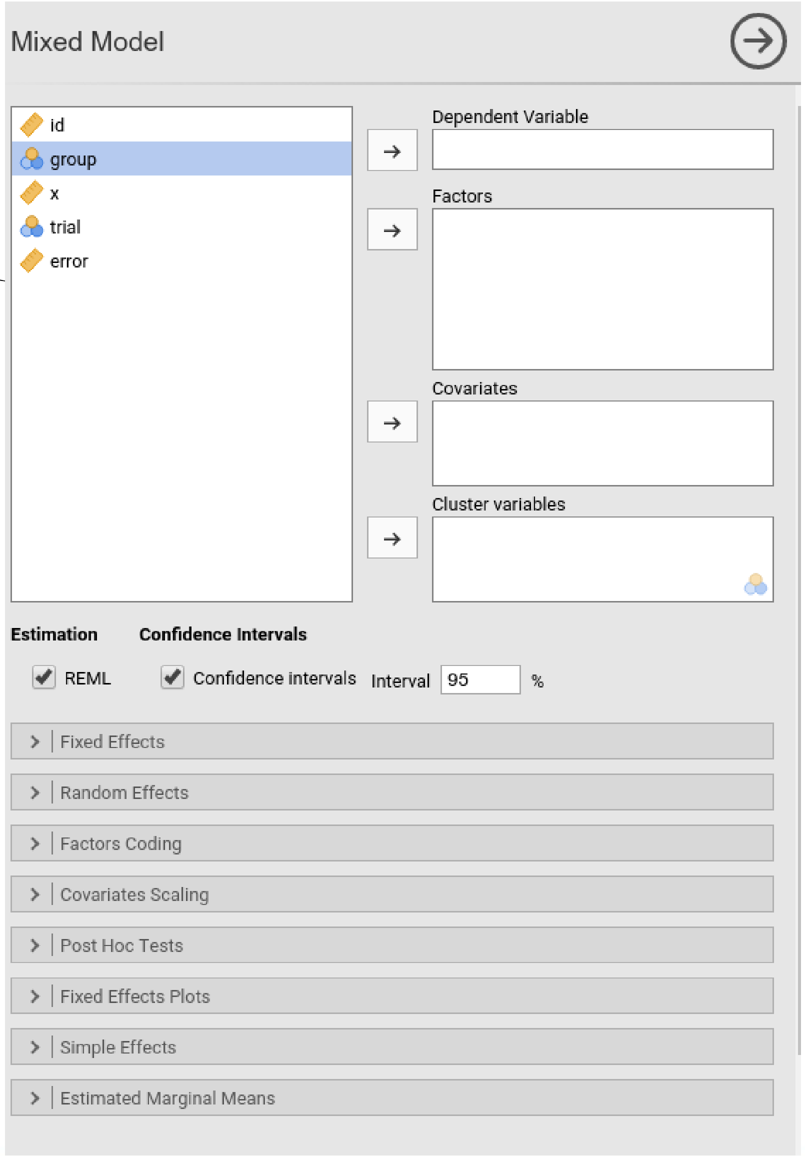
$$y_{ij} = \bar{a} + a_j + \bar{b} \cdot x_{ij} + e_{ij}$$

- Fixed effects? Intercept and trial effect
- Random effects? Intercepts
- Clusters? participants

GAMLj: General mixed models

Variables

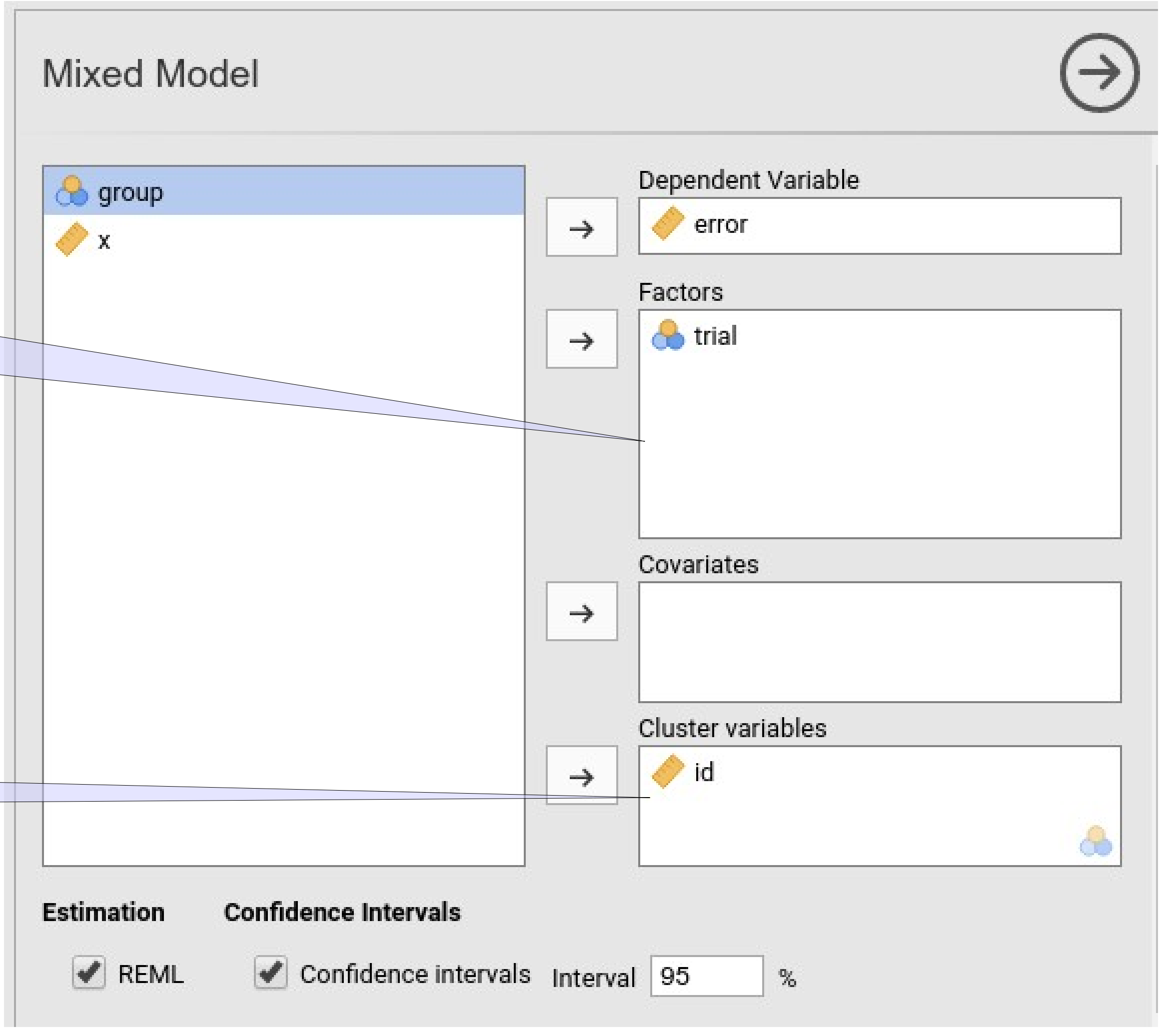
Options



GAMLj: General mixed models

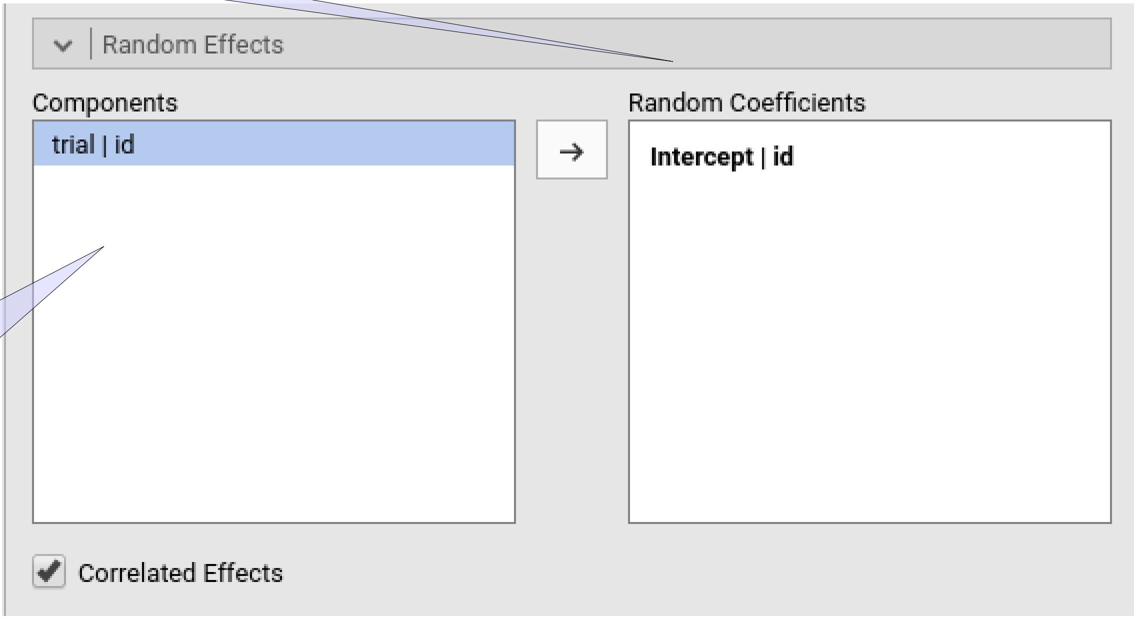
Categorical independent variable

Clustering variable(s)



GAMLj: random coefficients

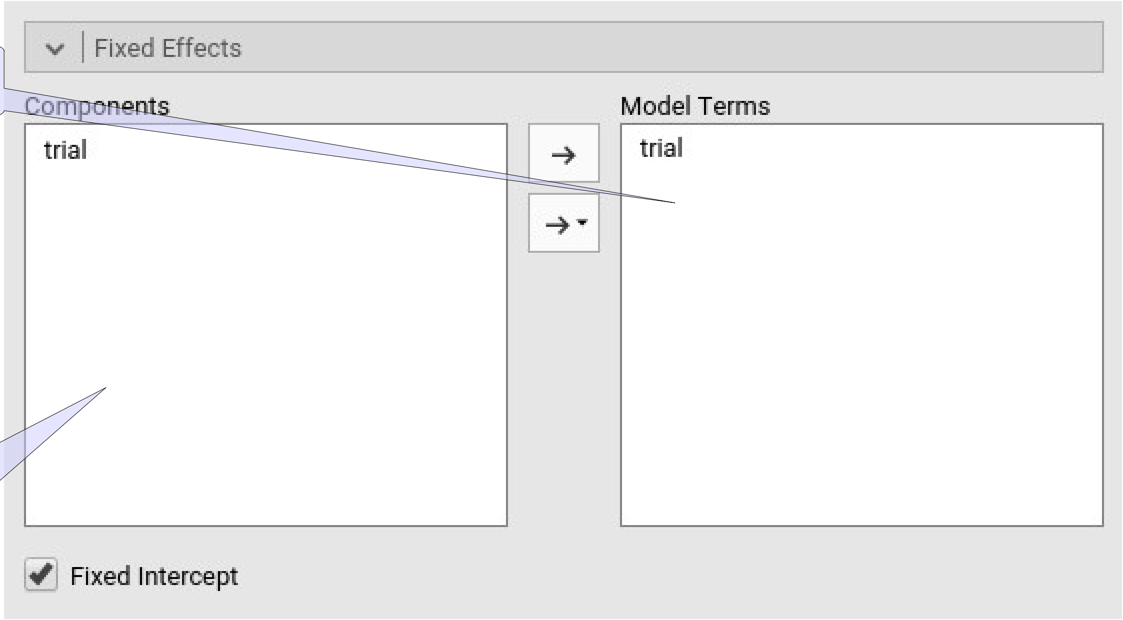
Random intercepts



All possible random coefficients

GAMLj: fixed coefficients

Fixed effect



All possible fixed coefficients

GAMLj: Results: model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	error ~ 1 + (1 | id) + trial
AIC	−463.8270
R-squared Marginal	0.0148
R-squared Conditional	0.2171

R-squared

R-squared Conditional: How much variance can the fixed and random effects together explain of the overall variance

R-squared Marginal: How much variance can the fixed effects alone explain of the overall variance

GAMLj: Results: random

Variance of intercepts

Random Components

Groups	Name	SD	Variance
id	(Intercept)	0.0883	0.00780
Residual		0.1738	0.03020

Note. Numer of Obs: 1000 , groups: id , 200

As long as the variance is non-zero, we are fine

GAMLj: Results: fixed

F-tests

Fixed Effect ANOVA

	F	Num df	Den df	p
trial	4.72	4	796	< .001

Note. Satterthwaite method for degrees of freedom

Coefficients

Fixed Effects Parameter Estimates

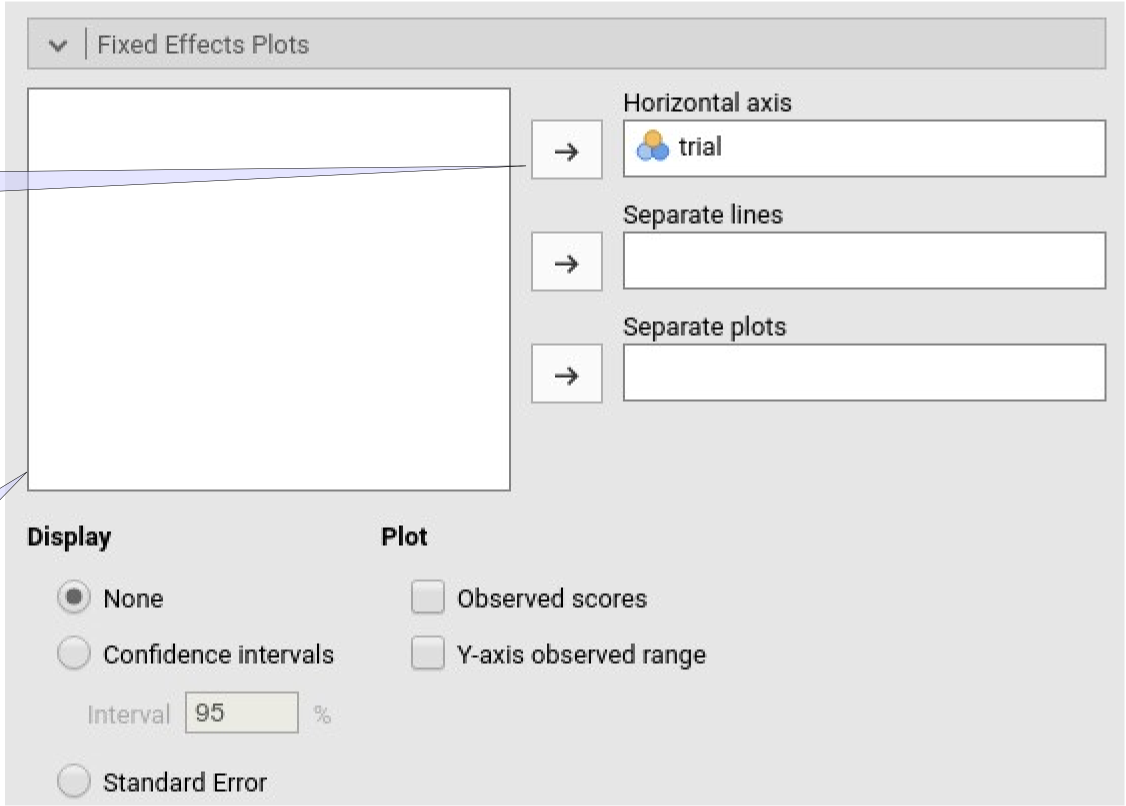
				95% Confidence Interval				
Effect	Contrast	Estimate	SE	Lower	Upper	df	t	p
(Intercept)	Intercept	0.49474	0.00832	0.4784	0.51104	199	59.4620	< .001
trial1	2 - (1, 2, 3, 4, 5)	−0.01791	0.01099	−0.0395	0.00363	796	−1.6296	0.104
trial2	3 - (1, 2, 3, 4, 5)	−7.92e−4	0.01099	−0.0223	0.02075	796	−0.0720	0.943
trial3	4 - (1, 2, 3, 4, 5)	0.04094	0.01099	0.0194	0.06248	796	3.7246	< .001
trial4	5 - (1, 2, 3, 4, 5)	0.00634	0.01099	−0.0152	0.02788	796	0.5764	0.564

Contrasts used to cast the categorical IV

GAMLj: plot

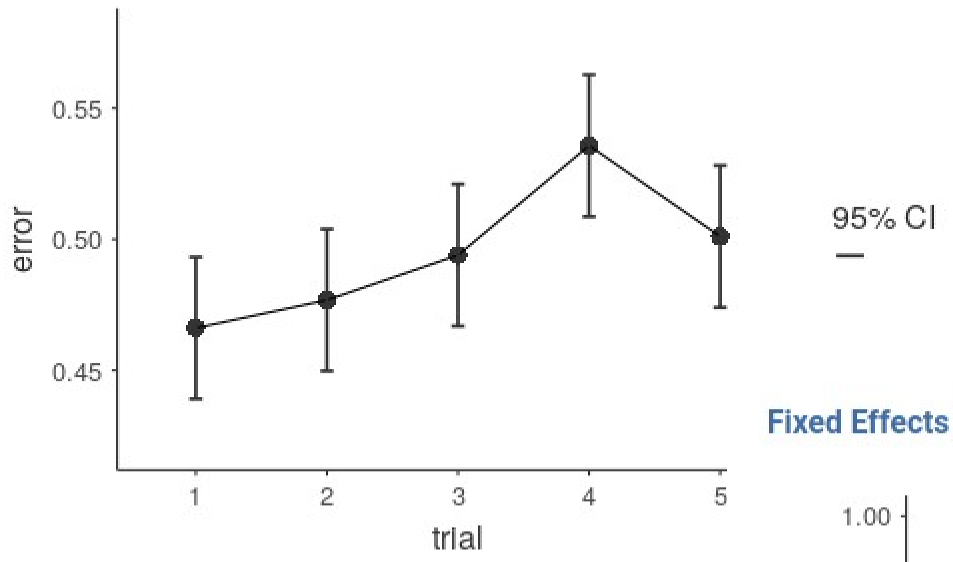
Fixed effect to plot

Options

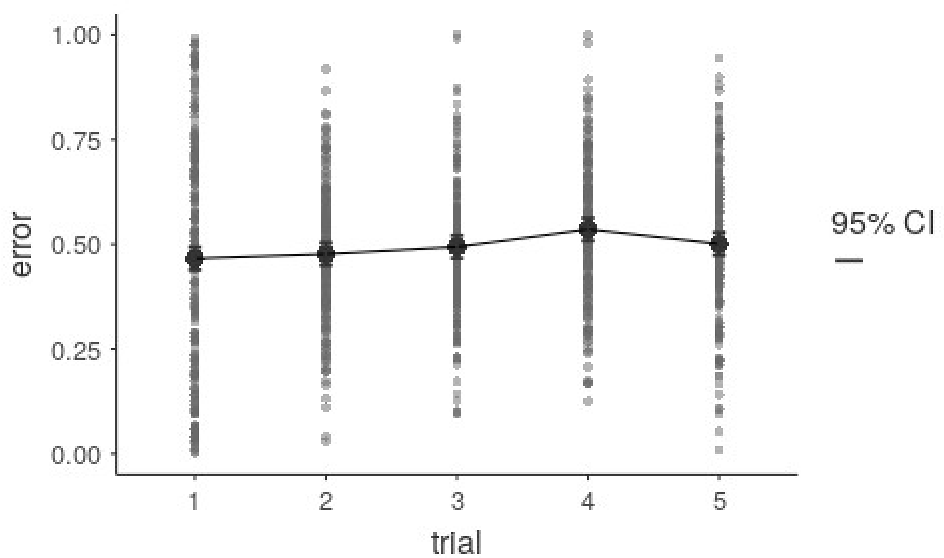


GAMLj: plot

Fixed Effects Plots



Fixed Effects Plots



Between and Repeated Measures Anova

linear mixed model

Standard design

- There are two groups - a Control group and a Treatment group, measured at 4 times. These times are labeled as 1 (pretest), 2 (one month posttest), 3 (3 months follow-up), and 4 (6 months follow-up).

- The dependent variable is a depression score (e.g. Beck Depression Inventory) and the treatment is drug versus no drug. If the drug worked about as well for all subjects the slopes would be comparable and negative across time. For the control group we would expect some subjects to get better on their own and some to stay depressed, which would lead to differences in slope for that group (*)

*) https://www.uvm.edu/~dhowell/StatPages/More_Stuff/Mixed-Models-Repeated/Mixed-Models-for-Repeated-Measures1.html

Standard design

- There are two groups - a Control group and a Treatment group, measured at 4 times. These times are labeled as 1 (pretest), 2 (one month posttest), 3 (3 months follow-up), and 4 (6 months follow-up).

Contingency Tables

Contingency Tables

	group		
time	1	2	Total
0	12	12	24
1	12	12	24
3	12	12	24
6	12	12	24
Total	48	48	96

96 observations
24 subjects

Standard design: data

- Data are in the long format

One subject 4 rows

	subj	time	group	dv
1	1	0	1	296
2	1	1	1	175
3	1	3	1	187
4	1	6	1	192
5	2	0	1	376
6	2	1	1	329
7	2	3	1	236
8	2	6	1	76
9	3	0	1	309
10	3	1	1	238
11	3	3	1	150
12	3	6	1	123
13	4	0	1	222
14	4	1	1	60
15	4	3	1	82
16	4	6	1	85
17	5	0	1	150
18	5	1	1	271
19	5	3	1	250

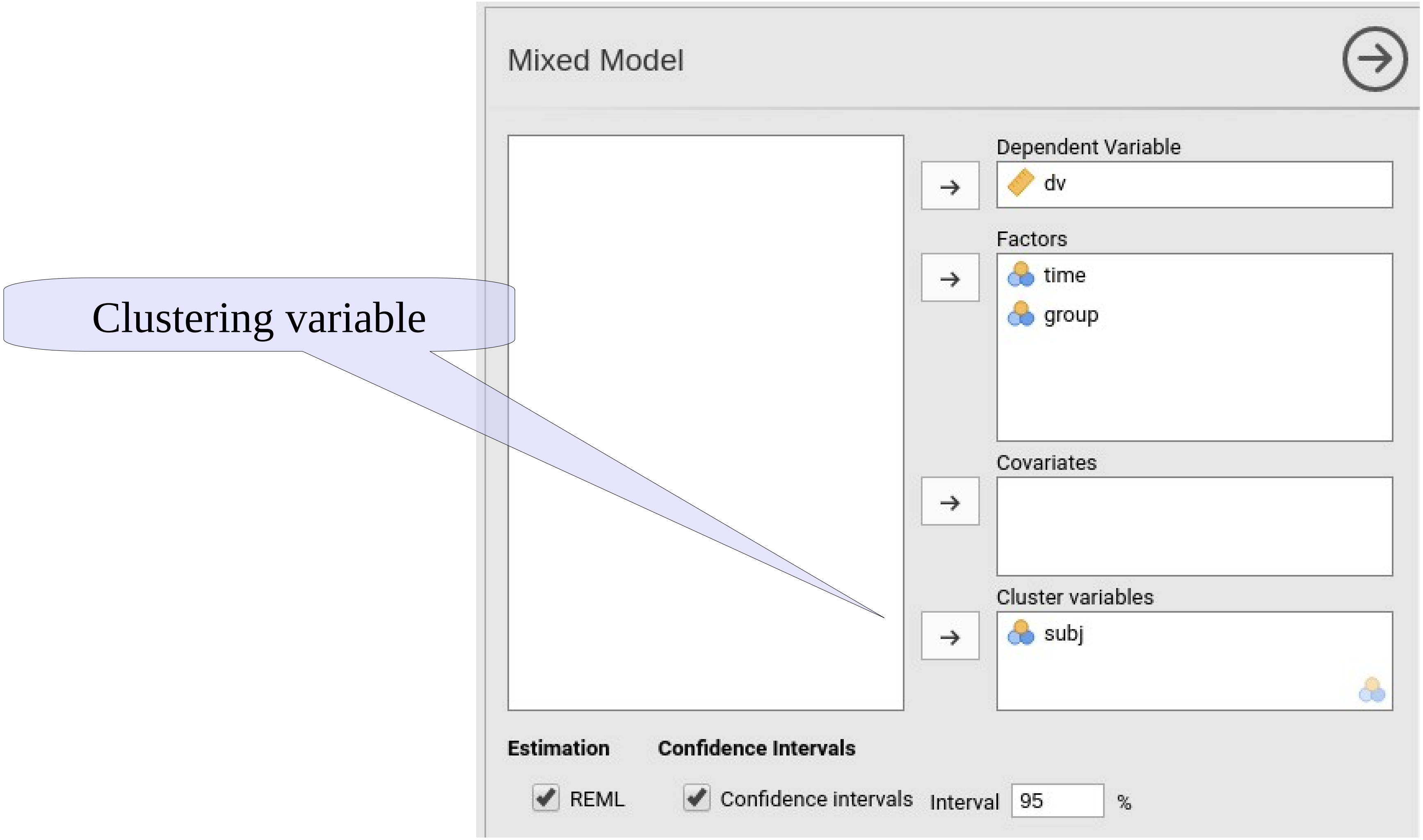
Mixed model

We can translate this in a standard mixed model

- Fixed effects? Intercept and group,time, and interaction effect
- Random effects? Intercepts
- Clusters? subjects

Variables

- Definition of the analysis

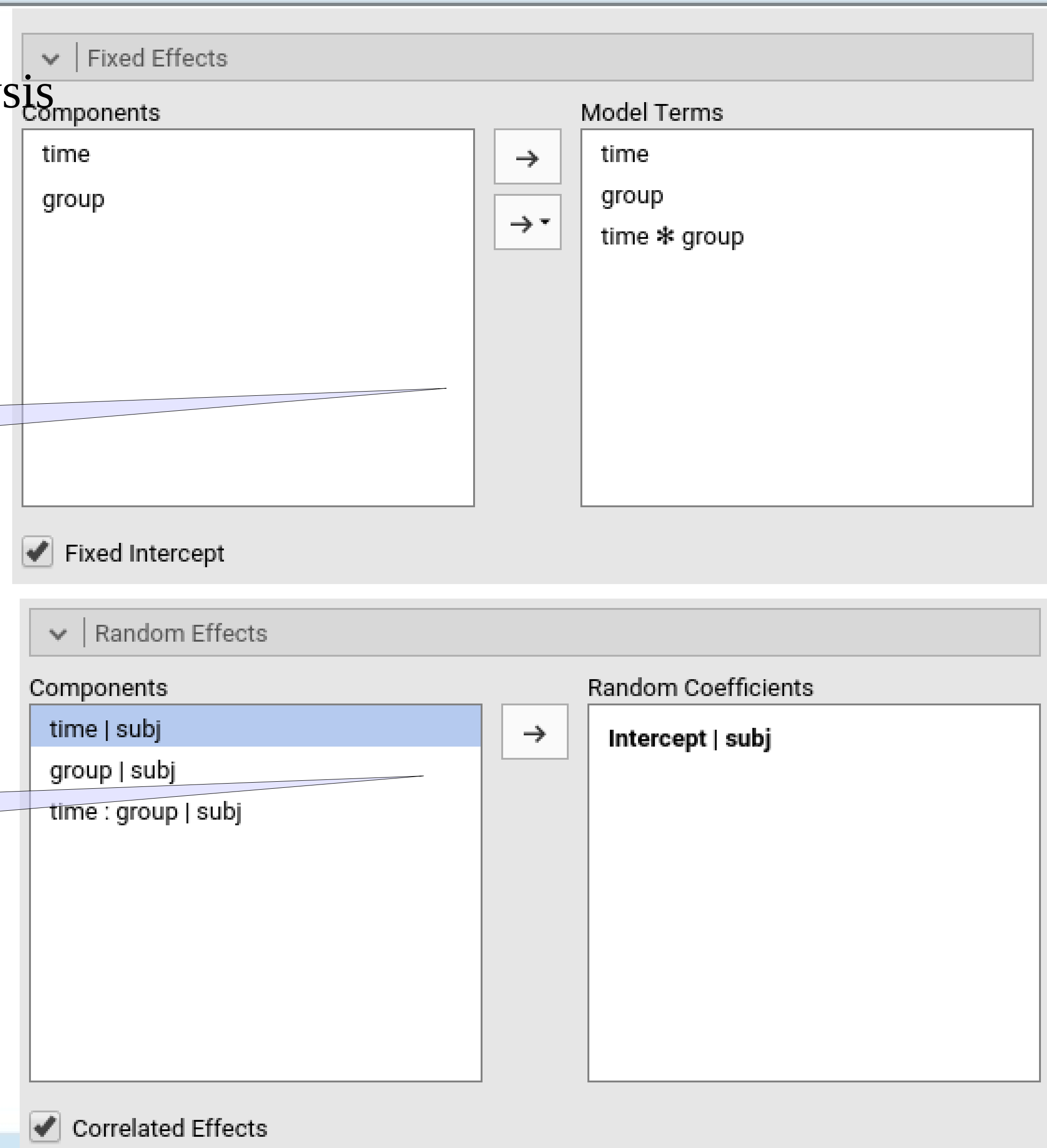


Model

- Definition of the analysis

Fixed effects

Random effects



Results

- Interpretation of results

Model

Mixed Model

Model Info

Info	
Estimate	Linear mixed model fit by REML
Call	dv ~ 1 + (1 | subj) + time + group + time:group
AIC	1011.895
R-squared Marginal	0.554
R-squared Conditional	0.768

Random effects

Random Components

Groups	Name	SD	Variance
subj	(Intercept)	50.4	2539
Residual		52.5	2761

Note. Numer of Obs: 96 , groups: subj , 24

Results

- Interpretation of results

Fixed F-tests

Fixed Effect ANOVA

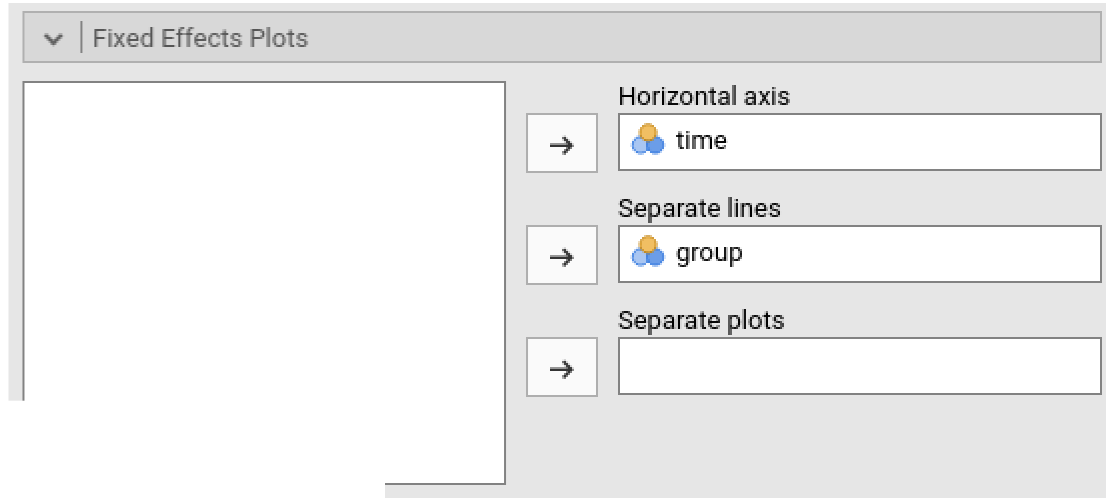
	F	Num df	Den df	p
time	45.14	3	66.0	< .001
group	13.71	1	22.0	0.001
time:group	9.01	3	66.0	< .001

Note. Satterthwaite method for degrees of freedom

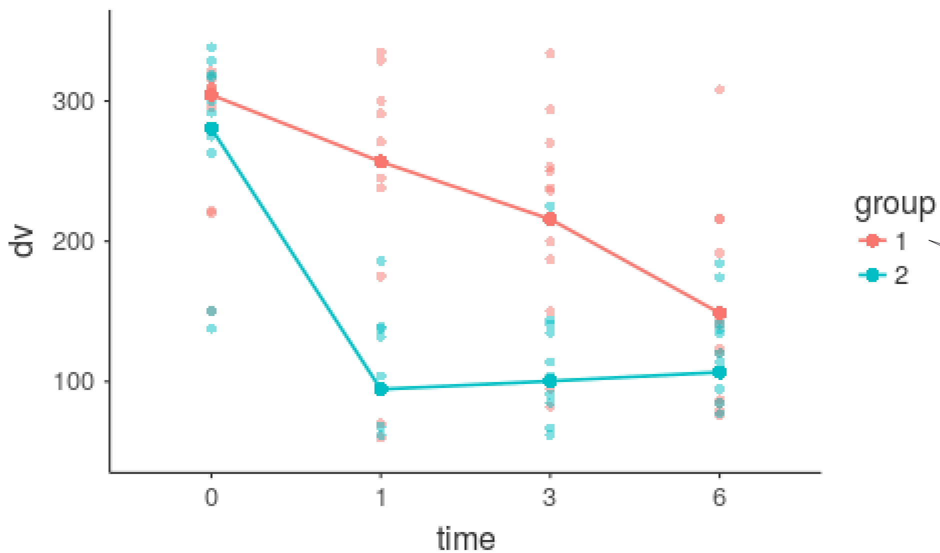
- For the moment we ignore the coefficients of the parameter estimates

Results: plot

- Interpretation of results



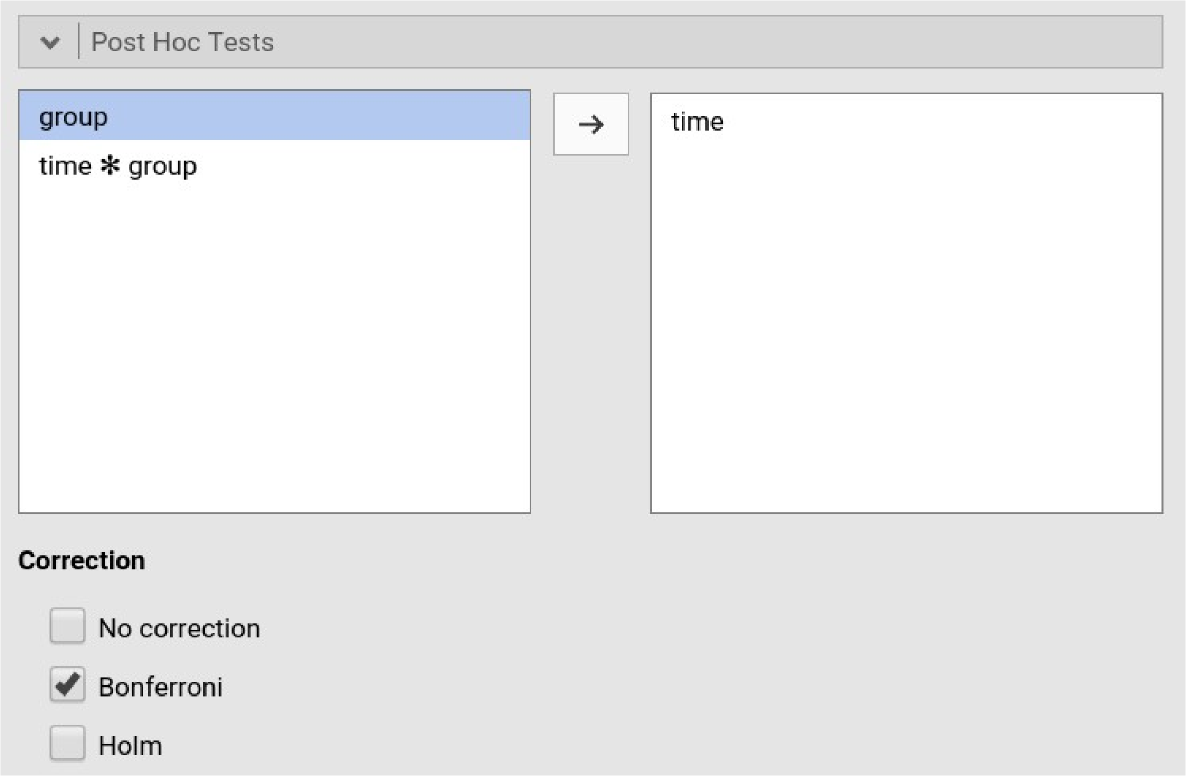
Fixed Effects Plots



Red is control group

Post-hoc tests

- As for the GLM, post-hoc tests compare all possible pairs of means and correct for inflated Type-I error



Post-hoc tests

- As for the GLM, post-hoc tests compare all possible pairs of means and correct for inflated Type-I error

Post Hoc Tests

Post Hoc Comparisons - time

Comparison							
time		time	Difference	SE	test	df	$p_{bonferroni}$
0	-	1	116.8	15.2	7.70	66.0	< .001
0	-	3	134.3	15.2	8.86	66.0	< .001
0	-	6	164.6	15.2	10.85	66.0	< .001
1	-	3	17.5	15.2	1.16	66.0	1.000
1	-	6	47.8	15.2	3.15	66.0	0.015
3	-	6	30.3	15.2	2.00	66.0	0.300

The end