Social Data Institute

Continuing Professional Development (CPD) course

Introduction To Bayesian Inference & Modelling (2022/23)

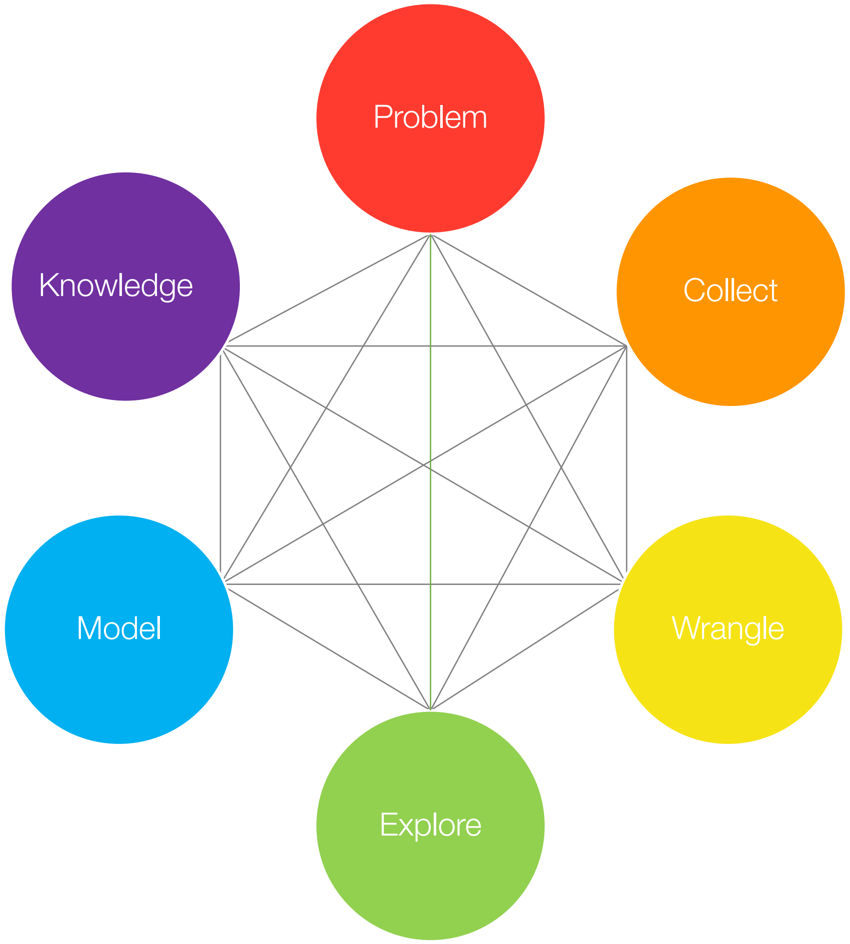
DAY 3: BAYESIAN HIERARCHICAL REGRESSION MODELS

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

Additional details:
https://www.ucl.ac.uk/social-data

Contents

- What are Hierarchical Regression Models?
  - Usage for exploring relationships with complex data structures (hierarchies, nesting, repeated measures, temporality etc.,)
  - Importance and why we use them to account for a specific data artefact
  - Wide applications of this specialized technique to other scientific domains
- Components of a hierarchical model
  - Statistical formulation
  - Random intercepts and Random slopes
  - Model types: Intercept-only, Random slope-only or both
- Model Specification from a Bayesian Framework
- Examples and interpretation



What are Hierarchical Regression Models?

Recall, we have extensively covered these various types of regression models

Distribution of dependent variable	**Suitable Model (GLM or GAM)**
Continuous measures: e.g., average income in postcode (£); concentrations of ambient particular matter (PM2.5); Normalised Vegetative Difference Index (NDVI) etc.,	Linear regression
Binary measures (1 = "present" or 0 = "absent"): e.g., Person's voting for a candidate, Lung cancer risk, house infested with rodents etc.,	Logistic Regression
Binomial measure (or proportion): e.g., prevalence of houses in a postcode infested with rodents, percentage of people in a village infected with intestinal parasitic worms, prevalence of household on a street segment victimised by crime etc.,	Logistic Regression
Counts or discrete measures: e.g., number of reported burglaries on a street segment, number of riots in a county etc.,	Poisson Regression

- Typically, the data structure or scenario we have been applying to these models are to single row records or unit observations (i.e., for an individual, or a geographical unit etc.,)

- What about data structures with repeated measurements, or unit observations nested within a group?

Data structures [1]

Imagine we have some dataset containing information on an individual-level.

ID	Name	Maths Performance	Maths TSR	OFSTED Grade
SCH01	Acton High School	29	20.9	5 (Worst)
SCH02	Brentside High School	40	18.6	5 (Worst)
SCH03	Greenford High School	51	11.7	3 (Below average)
SCH04	Northolt High School	60	9.9	2 (Good)
SCH05	Ellen Wilkinson School	88	14.6	0 (Excellent)
SCH06	Twyford Church of England	76	6.3	1 (Very Good)
SCH07	Featherstone High School	73	5.3	1 (Very Good)
SCH08	Drayton Manor High School	80	12.9	0 (Excellent)
SCH09	Dormers Wells High School	67	16.5	2 (Good)

Note 1: This dataset contain details for individual schools in Ealing Borough (inside London). Information on the overall maths performance of a school and the **maths teacher-student ratio (TSR)** in a class.

We want to understand what historical and sociodemographic factors have an impact on a school's performance when it comes to mathematics.

Note 2: We would typically fit a linear regression model if we wanted to see how just **Maths TSR** and **OFSTED Grade** are linked with **Maths Performance** variable.

Data structures [2]

Suppose we want to consider broader risk factors, not measured at an individual-level but at a group-level…

ID	LSOA	Name	Maths Performance	Maths TSR	OFSTED Grade	LSOA IMD Resources	LSOA IMD Income
SCH01	LSOA01	Acton High School***	29	20.9	5 (Worst)	5.2305	6.4734
SCH02	LSOA01	Brentside High School	40	18.6	5 (Worst)	5.2305	6.4734
SCH03	LSOA01	Greenford High School	51	11.7	3 (Below average)	5.2305	6.4734
SCH04	LSOA02	Northolt High School	60	9.9	2 (Good)	1.2353	0.3491
SCH05	LSOA02	Ellen Wilkinson School	88	14.6	0 (Excellent)	1.2353	0.3491
SCH06	LSOA03	Twyford Church of England	76	6.3	1 (Very Good)	0.2396	1.9843
SCH07	LSOA03	Featherstone High School	73	5.3	1 (Very Good)	0.2396	1.9843
SCH08	LSOA03	Drayton Manor High School	80	12.9	0 (Excellent)	0.2396	1.9843
SCH09	LSOA04	Dormers Wells High School	67	16.5	2 (Good)	3.1435	2.3679

For instance, other broader factors that might either be on an environmental, geopolitical, societal-level e.g., LSOA IMD public resource allocation for schools and average income scores. We have altered the structure of our dataset and made it far more complex…

Data structures [3]

A hierarchical or multi-level structure in the dataset is formed

ID	LSOA	Name	Maths Performance	Maths TSR	OFSTED Grade	LSOA IMD Resources	LSOA IMD Income
SCH01	LSOA01	Acton High School***	29	20.9	5 (Worst)	5.2305	6.4734
SCH02	LSOA01	Brentside High School	40	18.6	5 (Worst)	5.2305	6.4734
SCH03	LSOA01	Greenford High School	51	11.7	3 (Below average)	5.2305	6.4734
SCH04	LSOA02	Northolt High School	60	9.9	2 (Good)	1.2353	0.3491
SCH05	LSOA02	Ellen Wilkinson School	88	14.6	0 (Excellent)	1.2353	0.3491
SCH06	LSOA03	Twyford Church of England	76	6.3	1 (Very Good)	0.2396	1.9843
SCH07	LSOA03	Featherstone High School	73	5.3	1 (Very Good)	0.2396	1.9843
SCH08	LSOA03	Drayton Manor High School	80	12.9	0 (Excellent)	0.2396	1.9843
SCH09	LSOA04	Dormers Wells High School	67	16.5	2 (Good)	3.1435	2.3679

We have 9 records but we created an hierarchy…
3 school records nested in LSOA01;
2 school records nested in LSOA02;
3 school records nested in LSOA03;
1 school record nested in LSOA04

Data structures [4]

Individual-level data

ID	Name	Maths Performance	Maths TSR	OFSTED Grade
SCH01	Acton High School	29	20.9	5 (Worst)
SCH02	Brentside High School	40	18.6	5 (Worst)
SCH03	Greenford High School	51	11.7	3 (Below average)
SCH04	Northolt High School	60	9.9	2 (Good)
SCH05	Ellen Wilkinson School	88	14.6	0 (Excellent)
SCH06	Twyford Church of England	76	6.3	1 (Very Good)
SCH07	Featherstone High School	73	5.3	1 (Very Good)
SCH08	Drayton Manor High School	80	12.9	0 (Excellent)
SCH09	Dormers Wells High School	67	16.5	2 (Good)

Group-level data

LSOA	LSOA IMD Resources	LSOA IMD Income
LSOA01	5.2305	6.4734
LSOA02	1.2353	0.3491
LSOA03	0.2396	1.9843
LSOA04	3.1435	2.3679

We have 9 records but we created an hierarchy…

3 school records nested in LSOA01;
2 school records nested in LSOA02;
3 school records nested in LSOA03;
1 school record nested in LSOA04

A typical linear regression model would be severely inadequate for this problem due to the hierarchical structure that is formed in this dataset.

We would need a model that not only takes into account how the records are nested within a group; but one that would allow us to model the "within-group" variations formed in each group, as well as the "across-group" variations. Lastly, we will need a model that will let us fitted both individual-level and group-level variables, this is called a **hierarchical regression model**.

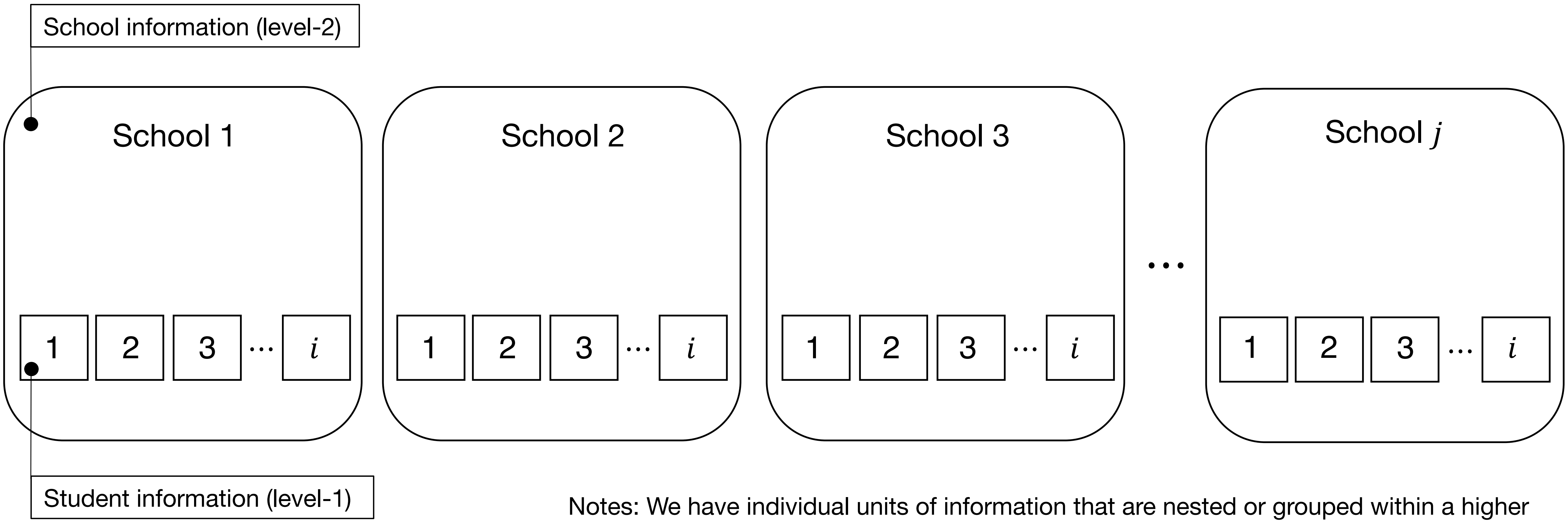
Definition:

A **hierarchical regression model**, are a specialised group of regression-based models that are able to recognise the existence of hierarchies within a data structure and account for them. It is a statistical model used for exploring the relationship between a dependent variable with one or more independent variables while accounting for these hierarchical structures.

Key characteristics of the hierarchical regression model:

- While it is commonly known as **hierarchical models**, it is also commonly interchangeable with the terms: **Multilevel models**; **Mixed-effect models**, **Nested data model** or even **Random-effects models**.
- The hierarchies formed by the natural structure of the dataset are treated as **levels** in the hierarchical model. There. can be more than one-level formed in the hierarchical regression model. A **two- or three-level hierarchical regression models** are often used a lot in research; however, more and more levels beyond 3 makes the regression incredibly complexed.
- The model structure is based on **levels** – the lowest level always correspond to individual units; while higher levels are the groupings. For example, a survey of a set of i number of students for their academic performance in j number of schools, across a set of years t. The students are level-1 (individual-level); schools are level-2 (grouping); and years are the level-3 (grouping or repeated measurements).

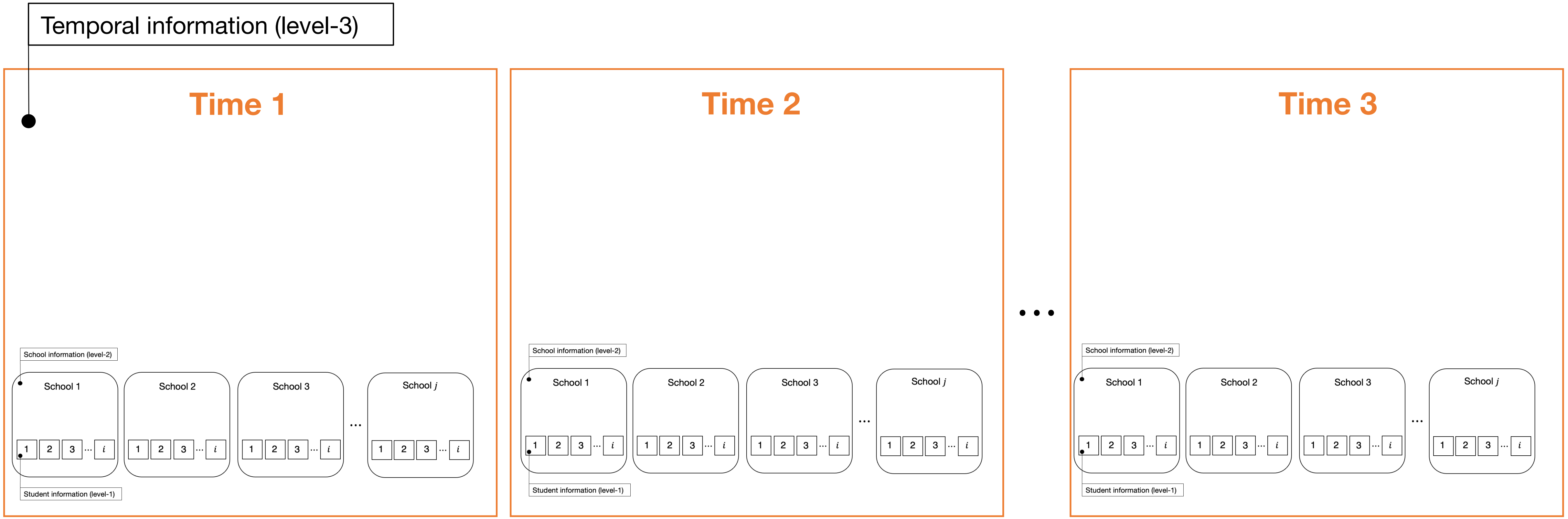
We are Illustrating concisely what we mean by two- or three-level model structure [1]



Notes: We have individual units of information that are nested or grouped within a higher measure. This is typically a **two-level structure** and a **two-level hierarchical regression** model must be used for this scenario.

Building on the example highlighted in point 3 (see slide 9): A survey of a set of i number of students for their academic performance in j number of schools, across a set of years t. The students are level-1 (individual-level); schools are level-2 (grouping); and years are the level-3 (grouping or repeated measurements).

We are Illustrating concisely what we mean by two- or three-level model structure [2]



Notes: We have individual units of information that are nested or grouped within a higher measure, where by the same individuals (from the same units) are repeated (i.e., longitudinal). This is typically a **three-level structure** and so a **three-level hierarchical regression** model must be used for this scenario.

Building on the example highlighted in point 3 (see slide 9): A survey of a set of i number of students for their academic performance in j number of schools, across a set of years t. The students are level-1 (individual-level); schools are level-2 (grouping); and years are the level-3 (grouping or repeated measurements).

Definition:

A **hierarchical regression model**, are a specialised group of regression-based models that are able to recognise the existence of hierarchies within a data structure and account for them. It is a statistical model used for exploring the relationship between a dependent variable with one or more independent variables while accounting for these hierarchical structures.

Why are hierarchical regression models important:

- It is an elegant way to model datasets that have varying scales in their measurements (- this artefact is caused by the multilevel or hierarchical structure in the dataset)
- It is an robust approach for accounting for **variations across individual units**, and at the same time, the "**within-group variations**" among groupings
- When we are modelling the direct relationship between the level-1 independent variables against the dependent variable, we can allow for direct interactions between level-1 and higher level independent variables that were measured at a group-level
- We can quantify group-specific differences as well as group-specific coefficients through the usage of "**varying-slopes**" or "**varying-coefficients**"

Components of a Hierarchical Regression Model

Recall the base model formula for a GLM

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \varepsilon$$

Variables

- y is the dependent variable
- $x_1, x_2, x_3, \ldots, x_k$ are the independent variables (which we have k number of them)

Parameters

- β_0 is the intercept
- $\beta_1, \beta_2, \beta_3, \ldots, \beta_k$ are the slopes (or coefficients) for the corresponding variables $x_1, x_2, x_3, \ldots, x_k$
- ε is the error term

Let's extend the above model into a hierarchical framework

Mathematical reformulation of the base GLM regression model using indexes

- When there is a hierarchical structure in the dataset, the base form of the GLM can be explicitly reformulated to show the hierarchies with indexes. For instance
  - ❖ We let i represent each individual unit or observation
  - ❖ We let j represent a group or cluster which an individual unit or observation i is from.
- Mathematical formulation of such scenario will be as follows:

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$

Breakdown of the above statistical model

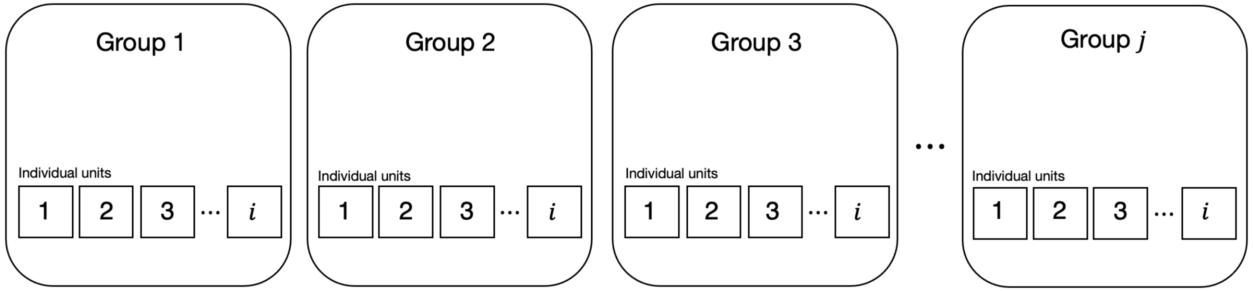
[1] Variables

- $y_{i,j}$ is the dependent variable. Is the observed outcome i in group j
- $x_1,\ x_2,\ x_3,\ \ldots,\ x_k$ are the k number independent variables
- Notation $x_{k,i,j}$ is the actual observation. It means that its the i observation in group j for the variable k

[2] Parameters

- $\beta_{0,j}$ is the intercept
- $\beta_{1,j}, \beta_{2,j}, \beta_{3,j}, \ldots, \beta_{k,J}$ are the coefficients corresponding to $x_1,\ x_2,\ x_3,\ \ldots,\ x_k$
- $\varepsilon_{i,j}$ is an error term

2-level hierarchical data drawn in picture form



2-level hierarchical data frame written in matrix algebraic form

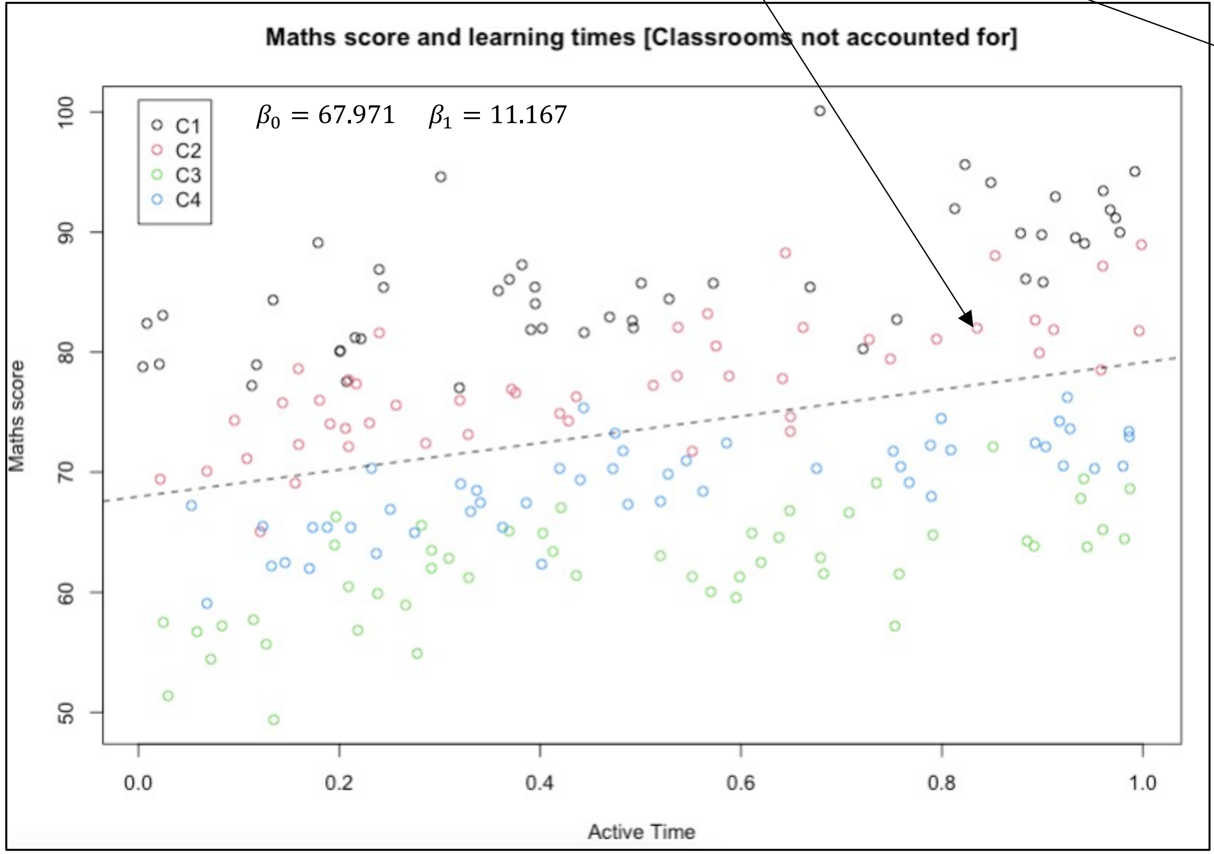
i	j	$y_{i,j}$	x_1	x_2	x_3	$\cdots$	x_k
1	1	$y_{1,1}$	$x_{1,1,1}$	$x_{2,1,1}$	$x_{3,1,1}$	$\cdots$	$x_{k,1,1}$
2	1	$y_{2,1}$	$x_{1,2,1}$	$x_{2,2,1}$	$x_{3,2,1}$	$\cdots$	$x_{k,2,1}$
3	1	$y_{3,1}$	$x_{1,3,1}$	$x_{2,3,1}$	$x_{3,3,1}$	$\cdots$	$x_{k,3,1}$
1	2	$y_{1,2}$	$x_{1,1,2}$	$x_{2,1,2}$	$x_{3,1,2}$	$\cdots$	$x_{k,1,2}$
2	2	$y_{2,2}$	$x_{1,2,2}$	$x_{2,2,2}$	$x_{3,2,2}$	$\cdots$	$x_{k,2,2}$
1	3	$y_{1,3}$	$y_{1,1,3}$	$x_{2,1,3}$	$x_{3,1,3}$	$\cdots$	$x_{k,1,3}$
2	3	$y_{2,3}$	$y_{1,2,3}$	$x_{2,2,3}$	$x_{3,2,3}$	$\cdots$	$x_{k,2,3}$
3	3	$y_{3,3}$	$y_{1,3,3}$	$x_{3,3,3}$	$x_{3,3,3}$	$\cdots$	$x_{k,3,3}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
i	j	$y_{i,j}$	$x_{1,i,j}$	$x_{2,i,j}$	$x_{3,i,j}$	$\cdots$	$x_{k,i,j}$

Notation for the intercept and coefficient i.e., $\beta_{0,j}$ and $\beta_{k,j}$ - what are they?

- Let us consider the following scenarios: 50 students in 4 classes, we are interested to know how active learning time impacts maths score

GLM model (not-indexed)

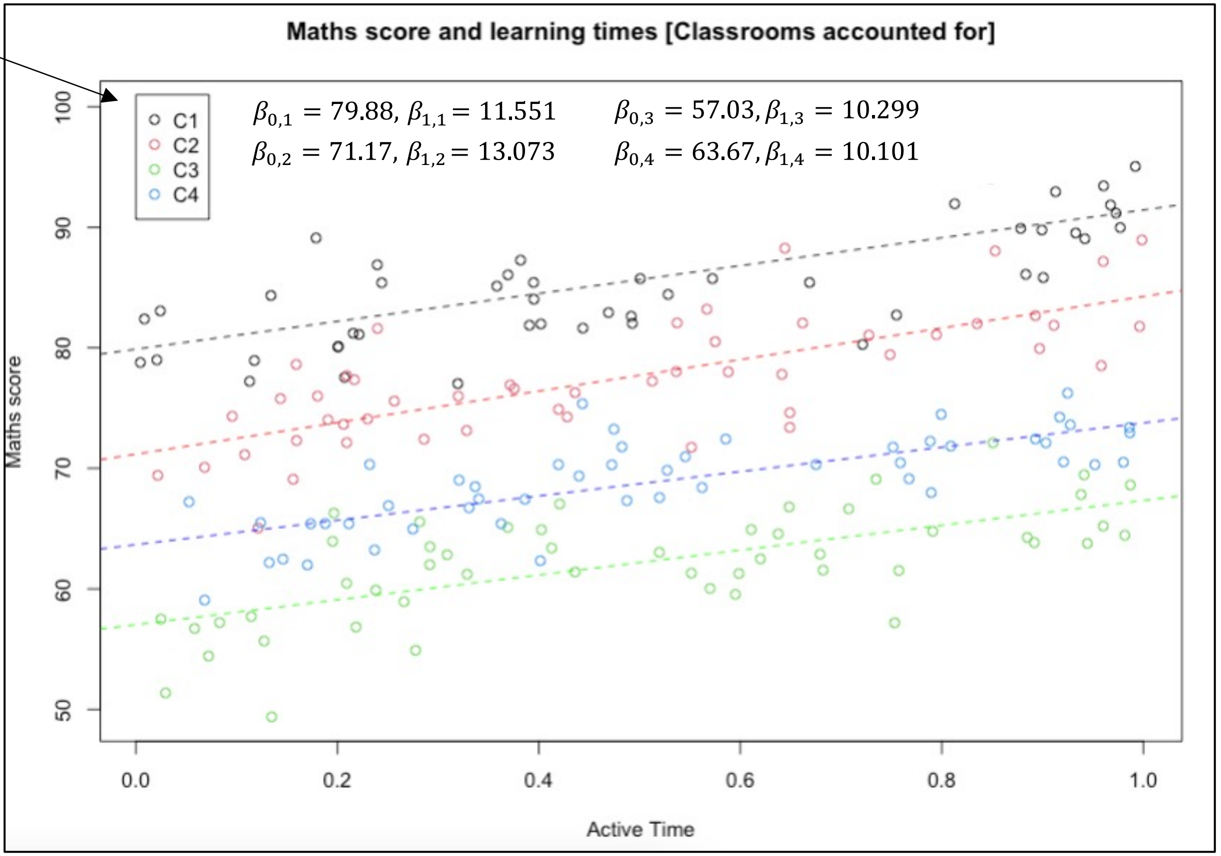
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$



Here, we can see that if we use a statistical model analyse this data without regards for the group structure. We get a single intercept and a single coefficient. Here, we are assuming that this relationship between active times and maths are similar across all the 200 students regardless of the classrooms they are in. This is what we term as Fixed Effects scenario

GLM model (indexed)

$$y_{i,j} = \beta_{0,j} + \beta_{1,j} x_{1,i,j} + \beta_{2,j} x_{2,i,j} + \cdots + \beta_{k,j} x_{k,i,j} + \varepsilon_{i,j}$$



However, we can see that this panel shows something different. Accounting for the classroom groups, **by fitting separate linear models** we get different intercepts (i.e., global mean) with different slope (or slope variation). There is an indication that some variation within the groups that's causing this pattern. This variation is known as a **Random Effects**, and it acting on our **intercepts and slopes**. This random effect must be accounted for, and so doing this would mean reformulated **indexed model with the random effects into a true hierarchical equation!**

Mathematical formulation for hierarchical regression model (full form) [1]

- Remember, we are **strictly** using a simple case of the 2-level model scenario
  - We let i represent each individual unit or observation
  - We let j represent a group or cluster which an individual unit or observation i is from.

- Mathematical formulation of such scenario will be as follows:

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$

Level 1 Equation

- For $\beta_{0,j}$ and for some $\beta_{k,j}$. Let us introduce some random effect (or random deviation) $u_{k,j}$ which causes this global intercept and slope coefficient to vary across some groups j and incorporate them to the indexed model. We would have these new equations specifically for the intercept $\beta_{0,j}$ and for the coefficients $\beta_{k,j}$ from the indexed model:

$$\left.\begin{aligned} \beta_{0,j} &= \gamma_{00} + u_{0,j} \\ \beta_{1,j} &= \gamma_{10} + u_{1,j} \\ \beta_{2,j} &= \gamma_{20} + u_{2,j} \\ &\vdots \\ \beta_{k,j} &= \gamma_{k0} + u_{k,j} \end{aligned}\right\} \text{Level 2 Equations}$$

Note 1: The above level 1 equation is a normal regression that we know. We makes it a hierarchical regression model is when we incorporate these random effects by nesting these level 2 equations to the intercept and coefficients in the level 1 equation.

Note 2: Notice how the intercept and slopes in level 1 equation are indeed a function of the components of the level 2 counterparts? To get the hierarchical regression in its full form, we simply substitute the level 2 equation into level 1

[1] Breakdown of components for the level 1 equation (individual units)

- $y_{i,j}$ is the dependent variable. Is the observed outcome i in group j
- $\beta_{0,j}$ is the intercept
- $\beta_{1,j}, \beta_{2,j}, \beta_{3,j}, \ldots, \beta_{k,J}$ are the coefficients that correspond to $x_1, x_2, x_3, \ldots, x_k$
- $\varepsilon_{i,j}$ is residual term

[2] Breakdown of the components of level 2 equation (group units)

- γ_{00} is a **fixed effect** (i.e., a constant term) associated with the intercept $\beta_{0,j}$
- $\gamma_{10}, \gamma_{20}, \cdots, \gamma_{k0}$ are **fixed effects** (i.e., constant terms) for the associated coefficients $\beta_{1,j}, \beta_{2,j}, \beta_{3,j}, \ldots, \beta_{k,J}$
- $u_{0,j}$ is the **random effects** (random variation caused by the groupings) on the intercept $\beta_{0,j}$
- $u_{1,j}, u_{2,j}, \cdots, u_{k,j}$ are **random effects** (i.e., random variation caused by the groupings) on the coefficients $\beta_{1,j}, \beta_{2,j}, \beta_{3,j}, \ldots, \beta_{k,J}$

Mathematical formulation for hierarchical regression model (full form) [2]

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$

Level 1 Equation

$$\begin{aligned}
\beta_{0,j} &= \gamma_{00} + u_{0,j} \\
\beta_{1,j} &= \gamma_{10} + u_{1,j} \\
\beta_{2,j} &= \gamma_{20} + u_{2,j} \\
&\vdots \\
\beta_{k,j} &= \gamma_{k0} + u_{k,j}
\end{aligned}$$

- 1st equation is a random-intercept
- 2nd, 3rd and 4th and so on equations are random-slopes
- Note that these equation does not have a two-level independent variable that impacts the outcome

Level 2 Equations

- Substitute the level 2 model equations into the level 1 model equation:

$$\Rightarrow y_{i,j} = (\gamma_{00} + u_{0,j}) + (\gamma_{10} + u_{1,j})x_{1,i,j} + (\gamma_{20} + u_{2,j})x_{2,i,j} + \cdots + (\gamma_{k0} + u_{1,j})x_{k,i,j} + \varepsilon_{i,j}$$

- After substitution, we expanding the expression and rearrange as follows:

$$\Rightarrow y_{i,j} = \underbrace{\gamma_{00} + \gamma_{10}x_{1,i,j} + \gamma_{20}x_{2,i,j} + \cdots + \gamma_{k0}x_{k,i,j}}_{\text{Fixed part}} + \underbrace{u_{0,j} + u_{1,j}x_{1,i,j} + u_{2,j}x_{2,i,j} + \cdots + u_{k,j}x_{k,i,j} + \varepsilon_{i,j}}_{\text{Random part}}$$

Model's true form

Note: There are model scenarios

- γ_{00} is the global intercept from the fixed part of the model we want to report
- $\gamma_{10}, \gamma_{20}, \ldots$ and γ_{k0} are the coefficients from the fixed part of the model we want to report now
- $u_{0,j}, u_{1,j}, u_{2,j}, \ldots$ and $u_{k,j}$ as well as $\varepsilon_{i,j}$ they have variances for random part of the model we want to report

Random-intercept-only, Random-slopes & Random coefficient scenarios [1]

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$

Level 1 Equation

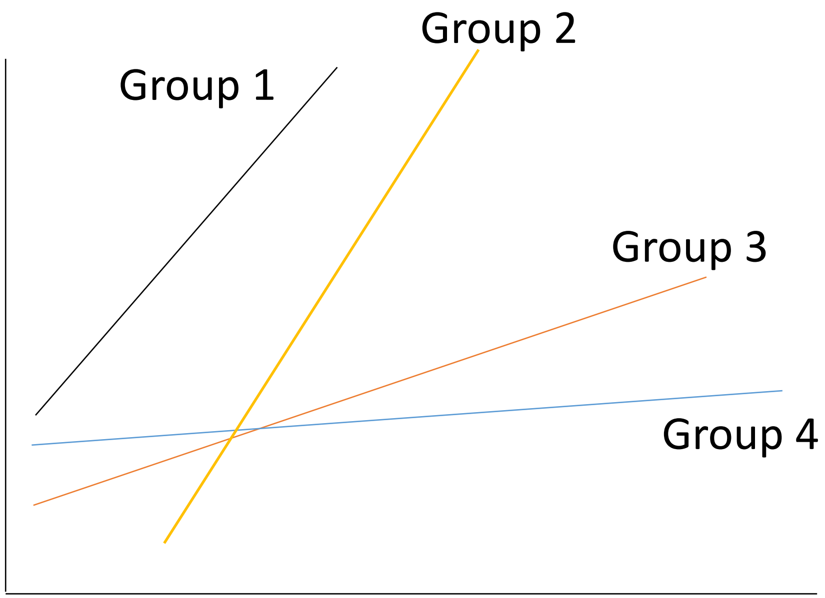
$$\left.\begin{aligned}
&\boxed{\beta_{0,j} = \gamma_{00} + u_{0,j}}\\
&\boxed{\begin{aligned}
\beta_{1,j} &= \gamma_{10} + u_{1,j}\\
\beta_{2,j} &= \gamma_{20} + u_{2,j}\\
&\vdots\\
\beta_{k,j} &= \gamma_{k0} + u_{k,j}
\end{aligned}}
\end{aligned}\right\}$$

Level 2 Equations

This is an example of a **random-slope model** which includes both a random-intercept and random-slopes. This means there group structures causes variation in the means across groups (i.e., intercepts) and slopes

$$y_{i,j} = \underbrace{\gamma_{00} + \gamma_{10}x_{1,i,j} + \gamma_{20}x_{2,i,j} + \cdots + \gamma_{k0}x_{k,i,j}}_{\text{Fixed part}} + \underbrace{u_{0,j} + u_{1,j}x_{1,i,j} + u_{2,j}x_{2,i,j} + \cdots + u_{k,j}x_{k,i,j} + \varepsilon_{i,j}}_{\text{Random part}}$$

Model's true form



Random-intercept-only, Random-slopes & Random coefficient scenarios [2]

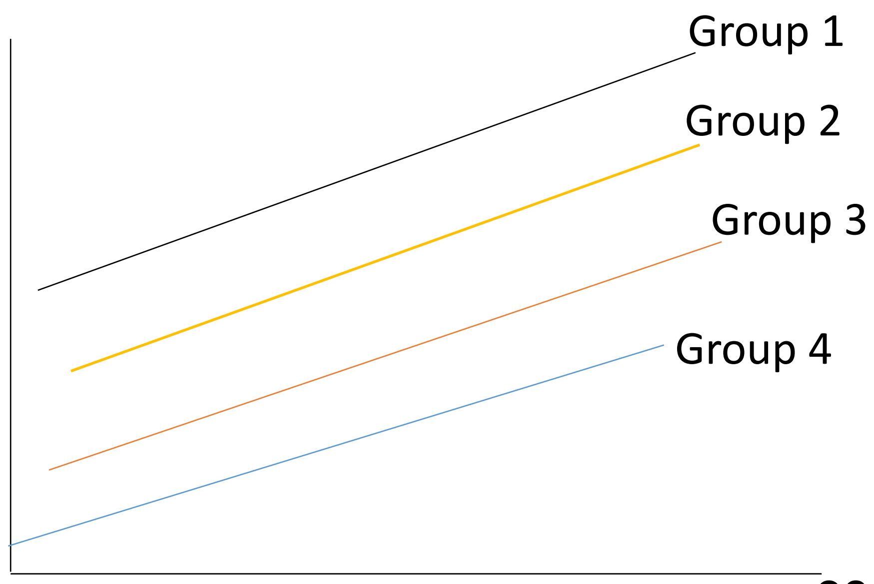
$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$ Level 1 Equation

$$\beta_{0,j} = \gamma_{00} + u_{0,j}$$ Level 2 Equation

Here, the model is much simpler:

$$y_{i,j} = \underbrace{\gamma_{00} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j}}_{\text{Fixed part}} + \underbrace{u_{0,j} + \varepsilon_{i,j}}_{\text{Random part}}$$ Model's true form

This is an example of a **random-intercept-only model** which only includes a random-intercept and excludes the random-slopes. This means that the group structure causes variation on the means (i.e., group-specific intercepts) but not on slopes



Random-intercept-only, Random-slopes & Random coefficient scenarios [1]

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \cdots + \beta_{k,j}x_{k,i,j} + \varepsilon_{i,j}$$

Level 1 Equation

$$\left.\begin{aligned}\beta_{0,j} &= \gamma_{00} + \gamma_{01}Z_1 + u_{0,j}\\ \beta_{1,j} &= \gamma_{10} + \gamma_{11}Z_1 + u_{1,j}\\ \beta_{2,j} &= \gamma_{20} + \gamma_{21}Z_1 + u_{2,j}\\ &\vdots\\ \beta_{k,j} &= \gamma_{k0} + \gamma_{k1}Z_1 + u_{k,j}\end{aligned}\right\}$$

Level 2 Equations

Suppose we have an independent variable measure on the group-level impacting our outcome on the individual-level.

- Substitute the level 2 model equations with the variables into the level 1 model equation:

$$\Rightarrow y_{i,j} = (\gamma_{00} + \gamma_{01}Z_I + u_{0,j}) + (\gamma_{10} + \gamma_{11}Z_I + u_{1,j})x_{1,i,j} + (\gamma_{20} + \gamma_{21}Z_I + u_{2,j})x_{2,i,j} + \cdots + (\gamma_{k0} + \gamma_{k1}Z_I + u_{1,j})x_{k,i,j} + \varepsilon_{i,j}$$

- After substitution, we expanding the expression and rearrange as follows:

$$\underbrace{\begin{aligned}\Rightarrow y_{i,j} = {} & \gamma_{00} + \gamma_{01}Z_1 + \\ & \gamma_{10}x_{1,i,j} + \gamma_{20}x_{2,i,j} + \cdots + \gamma_{k0}x_{k,i,j} + \\ & \gamma_{11}Z_1x_{1,i,j} + \gamma_{20}Z_1x_{2,i,j} + \cdots + \gamma_{k0}Z_1x_{k,i,j} + \\ & u_{0,j} + u_{1,j}x_{1,i,j} + u_{2,j}x_{2,i,j} + \cdots + u_{k,j}x_{k,i,j} + \varepsilon_{i,j}\end{aligned}}_{\text{Model's true form}}$$

γ_{00} is the global or population mean

γ_{01} is the random coefficient for Z_1

These are fixed effects coefficients for the variables in the level 1 equation

These are random coefficients for the interacting variables from the level 1 & 2 equation

These are the random effects

Advice – make life easy for yourself and use the random-intercept-only model. If you have a level-2 variable as you won't have to deal with any interactions!

An example and Interpretation

Example: Maths scores and active learning study [1]

GOAL: Assessing the impact of active learning on examine scores on mathematics among 200 school children total in 4 different class settings

$y_{i,j}$ = Maths scores attained by the student i in class j
$x_{1,i,j}$ = The proportion of active time spend studying by student i in class j
$x_{2,i,j}$ = Measure of feeling of support the student feels s/he receives

The students are clustered into 4 different classroom settings. For simplicity, here assume that the differences in classroom settings will cause the scores to vary from each other. Hence, we will use a **random-intercept-only model** to account for this

Model formulation

- Using a 2-level hierarchical model (random-intercept-only

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j} + \varepsilon_{i,j} \quad \text{(level-1 [students])}$$
$$\beta_{0,j} = \gamma_{00} + u_{0,j} \quad \text{(level-2 [classroom])}$$

- Specify likelihood function. The outcome is continuous – thus it normal (so no link function is need here).

$$y_{i,j} \sim \mathrm{Norm}(\mu_{i,j}, \sigma_\varepsilon)$$
$$\mu_{i,j} = \beta_{0,j} + \beta_{1,j}x_{1,i,j} + \beta_{2,j}x_{2,i,j}$$
$$\beta_{0,j} = \gamma_{00} + u_{0,j}$$

- Define the priors for the intercept, coefficients and random effects

$$\beta_{0,j} \sim \mathrm{Norm}(0, 20)$$
$$\beta_{1,j} \sim \mathrm{Norm}(0, 20)$$
$$\beta_{2,j} \sim \mathrm{Norm}(0, 20)$$
$$u_{0,j} \sim \mathrm{Norm}(0, \sigma_j)$$
$$\sigma_j \sim \mathrm{Uniform}()$$

- Build Bayesian model

Recall the Bayes' Rule: $P(\theta|Y) \propto P(Y|\theta)P(\theta)$

$$P(\boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{1,j}, \boldsymbol{\beta}_{2,j}, u_{0,j} \mid \mu_{i,j}) \propto P(\mu_{i,j} \mid \boldsymbol{\beta}_{0,j}, \boldsymbol{\beta}_{1,j}, \boldsymbol{\beta}_{2,j}, u_{0,j})\, P(\boldsymbol{\beta}_{0,j})P(\boldsymbol{\beta}_{1,j})P(\boldsymbol{\beta}_{2,j})P(u_{0,j})$$

Example: Maths scores and active learning study [2]

Stan code

```
data {
int<lower = 0> N;
int<lower = 0> CL;
int<lower = 0, upper = CL> ClassroomID[N];
int<lower = 0> k;
real<lower = 0> ActiveTime[N];
real<lower = 0> Supportive[N];
real<lower = 0> MathScore[N];
}
parameters {
real gamma00;
vector[k] beta;
vector[CL] u;
real<lower = 0> sigma_error;
real<lower = 0> group_error;
}
model {
real mu;
u ~ normal(0, group_error);
gamma00 ~ normal(0, 20);
beta ~ normal(0, 20);

for (i in 1:N) {
mu = gamma00 + u[ClassroomID[i]] + beta[1]*ActiveTime[i] + beta[2]*Supportive[i];
MathScore[i] ~ normal(mu, sigma_error);
                    }
}
```

Results (Random-intercept-only model)

Fixed effects

Intercept (γ_{00}) 33.24 (95% CrI: -18.18 to 68.28)
Active time ($\beta_{1,j}$) 11.44 (95% CrI: 9.90 to 12.98)
Supportive ($\beta_{2,j}$) 3.18 (95% CrI: 1.57 to 4.82)

Random effects

σ_e^2 (sigma error) 3.34
σ_u^2 (group error) 56.16
$u_{1,j}$ (random intercept Classroom 1) 44.92
$u_{2,j}$ (random intercept Classroom 2) 37.23
$u_{3,j}$ (random intercept Classroom 3) 21.59
$u_{4,j}$ (random intercept Classroom 4) 28.08

Any questions?