Name: 林成鍇 Julius Ling

Student ID: D84085360

Department: Psychology 112

*** This assignment is performed with *Python* coding. For detailed display in IPython Notebook, click the link below.

https://colab.research.google.com/drive/1Bm6w75L9AwGf38ENUIoNAvs03KHdDgDa?usp=sharing

Homework 6 – Hierarchical Linear Model (HLM)

Outline

I. HLM – An investigation on the relationship of the homework completion and math scores of students from different schools.

II. Bayesian HLM – An inference integrating my personal prior knowledge (how much certainty I have about the effect of homework completion on the math scores) and the present evidence.

Data

Source: https://github.com/SuryaThiru/hierarchical-model-blog/blob/master/mlmldata.csv

Sample size, n = 260

Variables of interest

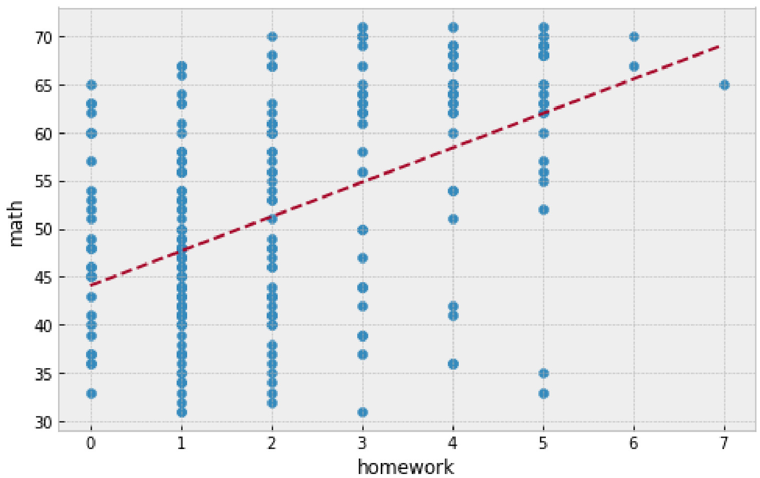
Independent variable (group level): school

Independent variable (individual level): homework completion

Dependent variable: math score

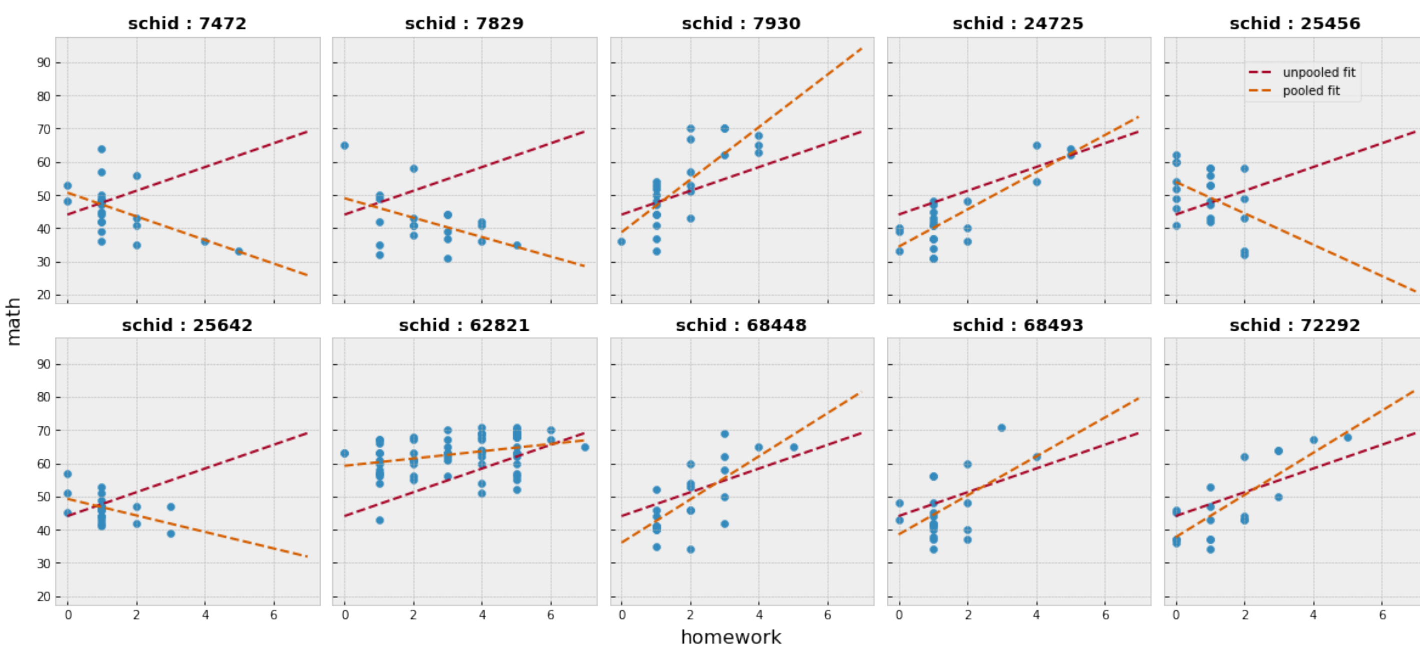
Right before starting up, to screen through the overall distribution and trend of the observed data, I made a pooled and unpooled linear model.

Pooled Linear Model:



The upward trend is apparent.

Unpooled Linear Model:



While the trend within each school is not necessarily upward sloping (each with a distinct set of slope and intercept), we believe that the distributions of the slopes and intercepts subject to normal distribution (the presumption of HLM).

The Intraclass Correlation Coefficient (ICC) is 0.32, indicating a rather high consistency within the schools.

So I perform a HLM analysis based on the observation, as I believe that the effect of school should be taken into account as well.

And here's the results presentation:

Model 1 (null model): math ~ 1 + (1 | school)

	math		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	48.86	45.08 – 52.64	**<0.001**
Random Effects			
σ^2	72.2		
τ_{00} school	34.01		
ICC	0.32		
N school	10		
Observations	260		

Model 2: math ~ homework + (1 + homework | school)

	math		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	44.77	39.39 – 50.15	**<0.001**
homework	2.04	-1.01 – 5.09	0.189
Random Effects			
σ^2	43.07		
τ_{00} school	69.31		
τ_{11} school.homework	22.45		
ρ_{01} school	-0.81		
ICC	0.67		
N school	10		
Observations	260		
Marginal R^2 / Conditional R^2	0.072 / 0.691		

The comparison of the random effects between the model 1 and 2 indicates that the students' homework completion level within the school has a significant effect in account for the drop of residual variance (72.2 – 43.07). The difference between marginal and conditional R-square does tell a consistent and convincing story that the difference between the school should have account for the improvement of the fit of the model (0.691-0.072). Hardly could we ignore such significant effect.

The big negative correlation effect between the fixed effect (the intercept and the effect of homework) is much more intriguing as it indicates that the student's level of homework completion has less (of even negative!) effect on the leverage of one's math score within the school with the greater average math score, while the hard work itself (level of homework completion) acts positively in the school with relatively lower startup point (the

It seems reasonable to me as to me, the good studying strategies itself are more reliable in accounting for better grades. Work hard "blindly" is not necessarily a strategy for good grades, especially among the students with higher grades. So, work smart instead!

However,

what if I know primarily that practice makes perfect, especially for the subject like Math (we all know how hard it is, aren't we?)

what if I'm the '愛拼才會贏'-kind of person believing in man's determination

what if I'm reluctant believe so for the excuse of me-lacking-of-determination-is-all-to-be-blame-for-my-poor-grades-not-because-of-that-I-am-stupid.

I'm having a doubt on such statistical result. Hence, I'm modelling, to show my degree of reluctance and uncertainty, with Bayesian HLM.

(The following section would be a mere 'playland'-kind of analysis. Enter if you please!)

The Bayesian analyses is conduct with probabilistic programming. It's a sampling approximation (a.k.a. Markov Chain Monte Carlo) method frequently used in Bayesian analyses.

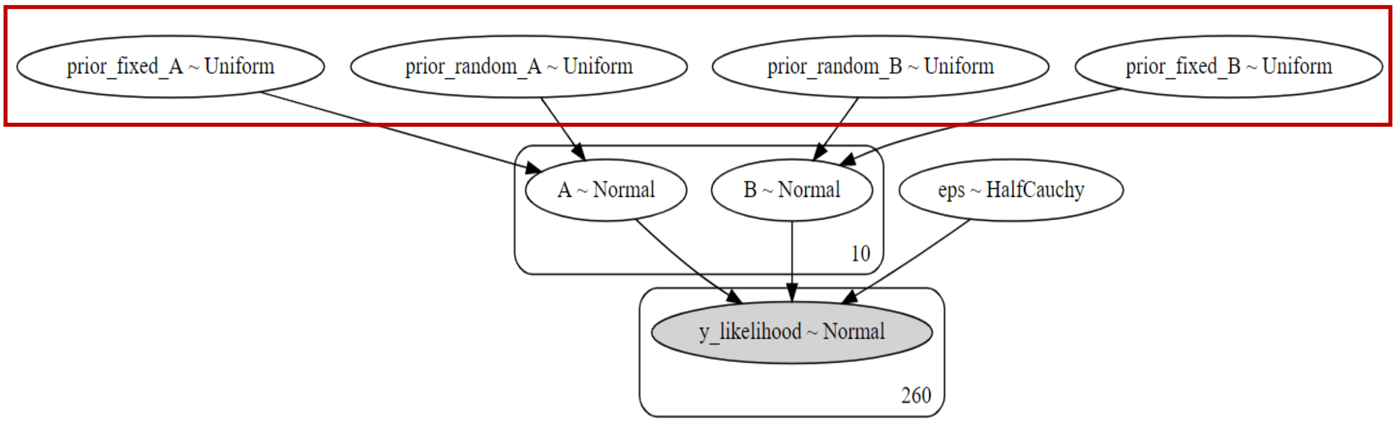
Just as what I've mention before, Bayesian analyses is an analysis integrating prior knowledge and present evidence, Hence, in Bayesian statistics, a parameter is not a mere value itself, but subjects to a prior distribution controlled by hyperparameters. At first, here's a will replicating the results in the previous section.

Here's the setup of my model to 'mimic' the way "Frequentist" does:

$$math_i = A_{ij} + B_{ij}homework + eps_i$$

$$math_i \sim Normal(A_{ij} + B_{ij}homework, eps)$$

The hyperparameters are programmed to have subjected to uniform distribution as a replication of Frequentist inference.



The posteriori distribution of the fixed & random effect of A & B and the residual variance converged with the one in the previous section, shown as below:

		Estimates	95% Credible Interval	
Fixed effect	Intercept	44.81	39.37	50.14
	Homework	2.22	0.01	4.22
Random effect	Residual	43.43		
	Intercept	71.91		
	Homework	25.00		

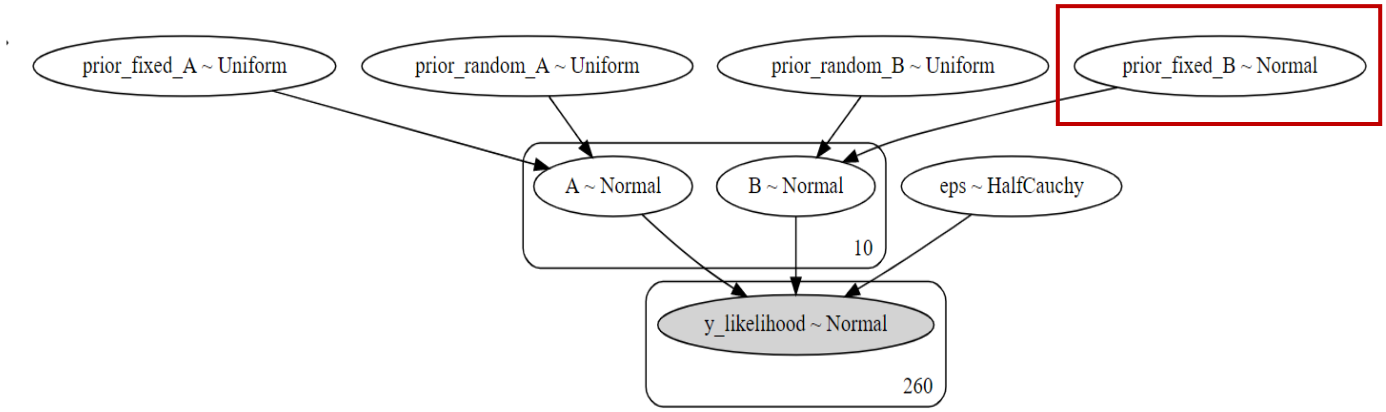
This is conceptually the Bayesian equivalent of confidence interval (long story short). However, it's not essentially the confidence interval, and the interpretation of such interval is different.

Predictors	*Estimates*	*CI*
(Intercept)	44.77	39.39 – 50.15
homework	2.04	-1.01 – 5.09
Random Effects		
σ^2	43.07	
τ_{00} school	69.31	
τ_{11} school.homework	22.45	

The result table in the previous section is up for your reference, the minute disparity in between is due the sampling error.

Just as I've confessed about my prior knowledge on the fixed effect of student's level of homework completion: such effect should be significantly positive. Hence, I'm setting up my model shown as below:

The fixed effect of B is presumed to have subjected to a normal distribution. I'm thereby having 2 hyperparameters to integrate up my prior knowledge.



$$prior\_fixed\_B \sim Normal(\mu, \sigma)$$

μ, acting as a parameter of my prior knowledge.

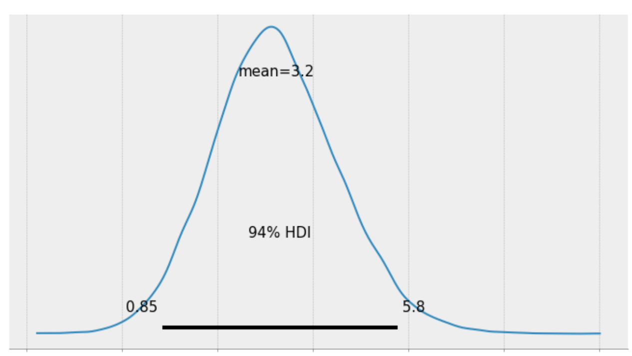
σ, acting as a parameter of my degree of certainty.

In this case, I know primarily a unit level increase in homework completion will lead to about 5 marks of improvement in math score averagely. And I'm quite certain (relatively small value of standard deviation) about it. Hence, to me, I'm convinced that:

$$prior\_fixed\_B \sim Normal(5, 2)$$

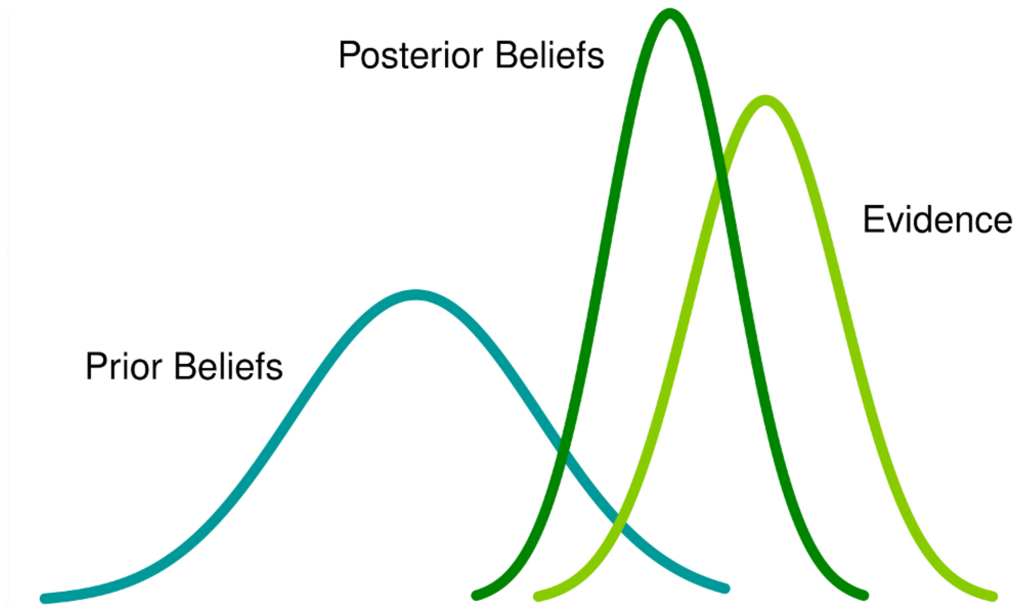
And here's the posterior distribution after integrating my prior knowledge:

		Estimates	95% Credible Interval	
Fixed effect	Intercept	44.67	39.36	50.18
	Homework	3.23	0.85	5.77
Random effect	Residual	43.56		
	Intercept	71.91		
	Homework	26.83		



The posterior distribution of the fixed effect is shown on the right. And note that the mean lies on the value 3.2, lower than my presumption.

Such result weights on my prior knowledge and present evidence via the σ, my certainty level.



The figure itself sum up the very fundamental spirit of Bayesian inference. Personally, I do appreciate it much as it takes the condition of uncertainty into accounts, as we all know that the measurement of human behavior itself could be somehow 'noisy'. Admit that. And integrate it up for refinement.

So, here's my very solution for the very last assignment.

Thousands appreciation & Happy Vacation!