Cultura y Educación
Culture and Education

ISSN: 1135-6405 (Print) 1578-4118 (Online) Journal homepage: http://www.tandfonline.com/loi/rcye20

Applying the hierarchical linear model to longitudinal data / La aplicación del modelo lineal jerárquico a datos longitudinales

Ryan W. Walters & Lesa Hoffman

To cite this article: Ryan W. Walters & Lesa Hoffman (2017) Applying the hierarchical linear model to longitudinal data / La aplicación del modelo lineal jerárquico a datos longitudinales, Cultura y Educación, 29:3, 666-701, DOI: 10.1080/11356405.2017.1367168

To link to this article: https://doi.org/10.1080/11356405.2017.1367168

Published online: 20 Oct 2017.

Submit your article to this journal

Article views: 21

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=rcye20

Download by: [66.45.159.208] **Date:** 27 November 2017, At: 07:54

Cultura y Educación / Culture and Education, 2017
Vol. 29, No. 3, 666–701, https://doi.org/10.1080/11356405.2017.1367168

Routledge
Taylor & Francis Group

Check for updates

Applying the hierarchical linear model to longitudinal data / *La aplicación del modelo lineal jerárquico a datos longitudinales*

Ryan W. Walters[a] and Lesa Hoffman[b]

[a]*Creighton University;* [b]*University of Kansas*

(*Received 2 November 2016; accepted 26 February 2017*)

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Abstract: Educational researchers and school administrators frequently evaluate academic outcomes collected from cross-sectional sampling designs with overt nested structures, such as when students are nested within schools. More recently, interest has focused on the longitudinal collection of academic outcomes to evaluate a student's growth across time. In a longitudinal context, the repeatedly measured academic outcomes are nested within a student. Proper analysis of longitudinal data requires the hierarchical linear model to quantify the extra correlations within students created by the nested sampling structure. In this article, we introduce the hierarchical linear model used to quantify and predict between-student differences in a repeatedly measured continuous maths achievement outcome. This introduction is presented as a conversation representative of those we have frequently with individuals who lack statistical training in hierarchical linear models for longitudinal data. Specifically, we cover why repeated-measures ANOVA may not always be appropriate, how the hierarchical linear model can be used to quantify between-student differences in change and how student- and occasion-level predictors can be properly modelled and interpreted.

Keywords: longitudinal; hierarchical linear model; random effects; within-person change

Resumen: A menudo, los investigadores en educación y los administradores de centros escolares evalúan resultados académicos recopilados a partir de diseños de muestreo transversales con estructuras anidadas evidentes, como cuando el alumnado está anidado en escuelas. Recientemente, el interés se ha centrado en la recopilación longitudinal de resultados académicos con el fin de evaluar el progreso de un alumno a lo largo del tiempo. En un contexto longitudinal, los resultados académicos medidos repetidamente están anidados intra-alumno. Un análisis apropiado de datos longitudinales exige que el modelo lineal jerárquico cuantifique las correlaciones adicionales intra-alumno creadas por la estructura de muestreo anidada. Este artículo presenta el modelo lineal jerárquico que se emplea para cuantificar y predecir diferencias inter-alumno en un resultado continuo de rendimiento en

English version: pp. 666–681 / *Versión en español*: pp. 682–698
References / *Referencias*: p. 698
Translated from English / *Traducción del inglés*: Joaquim Siles i Borràs
Authors' Address / *Correspondencia con los autores*: Ryan W. Walters, Creighton University, Omaha, NE, USA. E-mail: ryanwalters@creighton.edu

© 2017 Fundacion Infancia y Aprendizaje

matemáticas medido repetidamente. El estudio se presenta en forma de conversación, representativa de aquellas que frecuentemente mantenemos con personas que carecen de formación estadística, con relación a los modelos lineales jerárquicos para datos longitudinales. Concretamente, el estudio explica por qué los modelos de análisis de varianza (ANOVA) con medidas repetidas pueden no siempre ser apropiados, cómo puede utilizarse el modelo lineal jerárquico para cuantificar las diferencias inter-alumno que se producen en el cambio, y cómo pueden modelarse e interpretarse adecuadamente los predictores de nivel de alumno y ocasión de la medición.

Palabras clave: longitudinal; modelo lineal jerárquico; efectos aleatorios; cambio intra-persona

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Sydney Farasyn was recently hired as a data scientist for a school district consisting of more than 15,000 third- through eighth-grade students across 52 elementary and middle schools. Six years ago, her district began collecting yearly maths, reading and science achievement data, but up to this point the data have only been evaluated on a yearly basis to retain a consistent level of state and federal funding. Sydney recently met with the district superintendent, who became aware of the potential utility of evaluating longitudinal academic achievement data, and she was charged with analysing these data to evaluate how student achievement changed over time. Because Sydney has limited experience of modelling longitudinal data, she decided to meet with Dominik Altmann, a professor at a local university whose didactic course offerings include modelling longitudinal data. Sydney has just knocked on Dominik's office door.

"Please come in, Sydney. Have a seat," Dominik said.

"Thank you for taking the time to meet with me, Dr. Altmann," Sydney began.

Dominik interrupted: "Please, call me Dominik."

"Okay. Thank you," Sydney continued, noticeably tense. "Our district begins collecting academic achievement data once a student enters the third grade. Currently, I have six years of maths achievement data for our first cohort of 2,700 students who have recently completed eighth grade. I've been tasked with evaluating how maths achievement in this cohort of students changed from third to eighth grade."

"Is the maths outcome measured on a continuous scale?" Dominik asked.

"Yes," Sydney responded. "The scores range from zero to 200. In case you were wondering, the test is vertically scaled, meaning that each grade-specific test includes enough grade-to-grade content to effectively indicate increasing achievement, or growth."

"Do you have experience modelling longitudinal data?" Dominik asked.

"Some, but I have much more experience modelling cross-sectional data, such as when students are nested within classrooms."

Sydney continued: "I've run a few repeated-measures ANOVAs (RM-ANOVA) on these data, but my statistics program said I only had data for about 50% of the students. I am not sure what was going wrong," she said, slightly embarrassed. "One of my colleagues told me that I need to use a hierarchical linear model (HLM), so I started reading the longitudinal chapters in textbooks dedicated primarily to HLMs for cross-sectional data. However, I had a hard time making the connection between cross-sectional and longitudinal HLMs, so I decided to contact you."

Dominik smiled. "Although there is considerable overlap in modelling cross-sectional and longitudinal data, seeing that overlap can be difficult. As for textbooks, I recommend you consult Hoffman (2015), which is dedicated completely to modelling longitudinal data. I also recommend her textbook because it maps directly onto how I will explain modelling of longitudinal data to you during this and any subsequent meetings."

"Okay. Thanks!" Sydney took a pen and paper out of her bag and jotted a few notes.

Dominik continued: "Did you know that RM-ANOVA is subsumed within the HLM?"

"No," sighed Sydney.

"That's okay," Dominik assured. "We will start by describing nested data structures and how RM-ANOVA, and other forms of the HLM, go about handling longitudinal data."

"Sounds good to me. I'm ready," she said enthusiastically.

"Great. So, let me start by describing the overlap in modelling cross-sectional and longitudinal data. With cross-sectional data, you had to account for the nesting of students within classrooms because the achievement scores from students within the same classroom inevitably shared some extra association, or correlation, due to classroom factors, such as their teacher. This same concept applies to longitudinal data because the maths scores observed from the same student will be more highly correlated than maths scores observed from different students. I like to say that you will always be more correlated with yourself than with another person. It is this extra correlation, or *dependency*, that you need to be concerned about. Further, a longitudinal design can be phrased using traditional hierarchical terminology by saying that the repeatedly measured maths scores at level 1 are nested within students at level 2. RM-ANOVA is one way to account for the dependency that was created as a result of nested or the hierarchical structure. Did you know that there are two types of RM-ANOVA models?"

"No," Sydney said modestly.

"Don't worry," Dominik assured calmly: "we'll talk about the types and how they address dependency in a few minutes. Initially, though, I want to highlight four reasons why estimating the HLM is preferred over a RM-ANOVA. Although we'll cover this briefly today, a complete description is provided in chapters one and three in Hoffman (2015)."

Sydney shifted slightly in her chair as Dominik continued.

"Okay. First, RM-ANOVA can only quantify *constant* between-student differences in maths scores over time. Although this effectively acknowledges that students have different levels of maths ability, a RM-ANOVA model would

Downloaded by [66.45.159.208] at 07:54 27 November 2017

constrain these between-student differences to remain constant from third to eighth grade."

"I think that is a little too simplistic because we know that students' ability levels mature at different rates," Sydney said pointedly.

"Indeed. This was true for my own children. Earlier, you said that you were tasked to evaluate change in maths scores across years, but RM-ANOVA cannot estimate change directly because time is modelled as categorical, not continuous. As a result, it only provides between-student *mean* differences in maths scores between each occasion. Although the HLM can quantify these mean differences, it also quantifies *random* between-student differences in *change across time*, thereby allowing you to *predict* why some students change differently from other students."

"What do you mean by random differences?" Sydney asked.

"In your situation, *random* means that individual students change differently over time." Sydney nodded approvingly while Dominik continued. "Second, RM-ANOVA requires that the repeated maths occasions be *balanced* such that every student is measured at the same occasions. If, however, students are measured at different time points, then time is *unbalanced* and the exact time at which students were measured must be rounded to fit into those common times. But rounding will introduce bias into your analysis and may produce inaccurate results."

"Well, I'm glad we measure the majority of students at approximately the same time every academic year," Sydney said, relieved.

"Should you ever have unbalanced occasions, though, know that the HLM models time as a continuous variable, allowing you to use exact time to ensure that change across occasions is modelled as accurately as possible."

Sydney energetically wrote notes while Dominik continued. "The third reason the HLM is preferred over RM-ANOVA is that RM-ANOVA is a *complete data technique* that uses something called least-squares estimation. What this means for you is that every student is required to have their maths score reported at all occasions, and if they do not, they will be excluded from analysis."

Sydney's eyes widened. "So, if a student is missing even one maths score, they will be excluded from the entire analysis?"

"Yes. This is called *listwise deletion* and is the reason why 50% of students were excluded in your initial RM-ANOVA. By contrast, the HLM uses maximum likelihood estimation that makes use of all available data. If you're interested, you can read section three within chapter five in Hoffman (2015) for more information about maximum likelihood estimation."

"Thanks. It is true that we do not capture all students at every occasion for many reasons, most commonly due to students transferring schools," said Sydney.

Dominik nodded. "Given that attrition is likely in any longitudinal design, the ability to use all available data is critical. Finally, there are two types of RM-ANOVA for describing the pattern of *variability* in achievement across occasions

Downloaded by [66.45.159.208] at 07:54 27 November 2017

— univariate and multivariate. In *univariate* RM-ANOVA, the variances and covariances of the maths scores are constrained into a pattern called *compound symmetry*, meaning that the variances and covariances of maths scores are constrained to be *constant* across occasions. If, however, students' maths scores change at different rates across years, compound symmetry cannot hold because the variances would necessarily change across years, too. The bottom line is that compound symmetry is likely to be overly restrictive. You can, however, get around the compound symmetry assumption by estimating a *multivariate* RM-ANOVA, in which all variances and covariances are estimated as observed in your data using an *unstructured* variance pattern."

Sydney took a moment to think. "Multivariate RM-ANOVA seems like a pretty good option."

"Although a multivariate RM-ANOVA might seem like a decent option for you, least squares estimation still requires complete data, so you would still be missing 50% of students. As a result, the HLM is a better option because it uses all available data and has many options available to describe the pattern of variability over time, including the compound symmetry and unstructured patterns. I encourage you to consult chapters four and five in Hoffman (2015) for a more complete description of available variance-covariance patterns."

Sydney shifted in her seat. "Okay, I understand why the HLM is preferred over RM-ANOVA for my data, but what did you mean when you said earlier that the HLM can *quantify* and *predict* random between-student differences in amount and change?"

Dominik smiled. "That's a great question, Sydney, and to answer it we first need to describe the two sides of the HLM. The first side consists of the *model for the means*, which includes the *fixed* intercept and *fixed* regression slopes used to describe the *average* starting point and *average* change in maths scores across students. These fixed effects are interpreted similarly to effects in a traditional linear regression model."

Sydney added confidently: "You mean that the fixed intercept would indicate the average maths score when all predictors equal 0, the fixed regression slope for a continuous predictor would indicate the average change in maths score per one-unit increase in the predictor, and the fixed regression slope for a categorical predictor would represent the average difference in maths score between two groups."

"Exactly!" exclaimed Dominik, delighted. "The second side of the HLM consists of the *model for the variance* that includes how the *random errors* in prediction are distributed and related across students and occasions. In a linear regression model, the only error term is the residual representing the *deviation* from a student's predicted outcome value and their observed outcome value; it is the variance of those residuals that is explained, or reduced, by predictor variables. However, given the hierarchical structure of longitudinal data, prediction errors will occur both between and within students, with between-student errors termed *random effects* and within-student errors termed *residuals*. These errors still represent deviations, so with that in mind, a *random intercept* is the student-

Downloaded by [66.45.159.208] at 07:54 27 November 2017

specific deviation from the average or fixed intercept, a *random slope* is the student-specific deviation from the average or fixed regression slope, and a *residual* is the occasion-specific deviation from of a student's maths score at a given occasion from their own predicted growth trajectory. It is the *variance* of the random effects and residuals that quantify between- and within-student differences and tell you whether student- and/or occasion-level predictors should be included in the model for the means."

"Okay. How will I know how much variance is at the student and occasion levels?" Sydney asked.

"Great question! The proportion of variability at each level can be quantified by calculating the *intra-class correlation* (ICC) from an estimated HLM that includes no predictor variables. This no-predictor HLM is known as an *unconditional, random intercept model*, and has three estimated values — a fixed intercept, a random intercept variance and a residual variance. In longitudinal data with two levels of nesting, such as when repeated observations are nested within students, random effects are partitioned out of residual variance. Let me draw this for you (see Figure 1). Because the residual variance and random intercept variance are assumed independent of each other, the total variance in the maths outcome in a two-level HLM is approximately the *sum* of these two variance components. The ICC is the proportion of *total variability* that is between students and ranges between 0 and 1. It is calculated as:

$$\frac{\text{random intercept variance}}{(\text{random intercept variance} + \text{residual variance})} = \frac{\text{random intercept variance}}{\text{total variance}}$$

"I think I understand," Sydney added. "So, an ICC of .60 indicates that approximately 60% of the variability in maths scores is due to between-student factors."

"An alternative interpretation would be that 40% of the variability in maths scores is due to occasion-level factors. It is this unconditional, random intercept model that serves as the starting point in your model-building process. I tell you what, instead of talking about these models, we should learn by doing. Did you bring the maths data with you?"

"Yes," Sydney said as she reached into her tote bag and pulled out her USB drive, handing it to Dominik. "I typically use SPSS software to do most of my statistical analyses, but I also have access to SAS software. I assume either program can be used to run the HLM, right?"

"Absolutely! I'll make sure you have both SPSS and SAS syntax (see Appendix), but today we'll work primarily in SPSS." Dominik plugged the USB drive into his computer and opened Sydney's data (*Data.csv*) using SPSS. He then typed the required SPSS syntax to estimate the unconditional, random intercept model.

"The output shows that the fixed intercept is 120.9, that the random intercept variance is 286.1, and the residual variance is 280.9. So, the ICC is: 286.1 /(286.1 + 280.9) = .50. Care to interpret the ICC?" Dominik asked.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

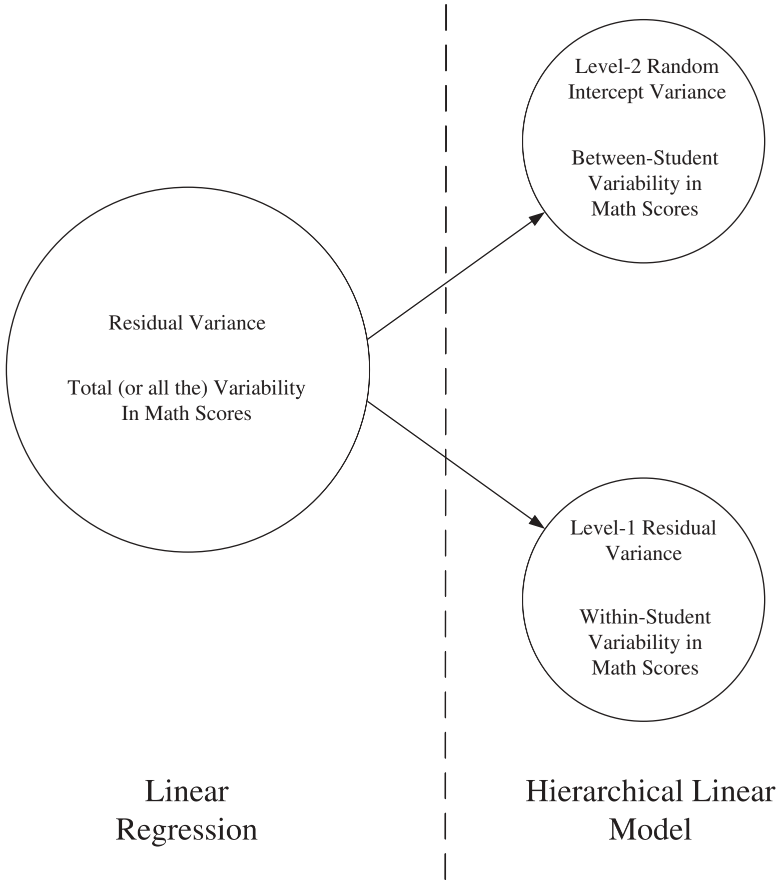

Figure 1. How variance is partitioned using the unconditional, random intercept hierarchical linear model. The residual variance estimate representing total variability from linear regression is partitioned into between- and within-student variance in a hierarchical linear model.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

"So, approximately 50% of the variability in maths scores is due to between-student factors and the remaining 50% is at the occasion level," Sydney said astutely.

"Is there a better way to describe the random intercept variance? I mean, can we give 286.1 some context?"

Dominik raised both eyebrows, impressed by Sydney's intuitiveness. "Absolutely! We can calculate a *95% random effects confidence interval*, which quantifies the range of student-specific intercepts for 95% of students in your sample. This confidence interval is calculated as:

$$\text{fixed effect} \pm 1.96\sqrt{\text{random effect variance}}$$

Here, $120.9 \pm 1.96\sqrt{286.1}$ indicates that 95% of the students had average maths scores between 87.7 and 154.1."

"That's very useful! So, what do we do next?" Sydney asked.

"Next, we consider the model for time. Remember, you are concerned with *change* over time, so we need to make sure that the model for time includes all relevant fixed and random effects of time to accurately capture any between-student differences in change."

"How do we do that?" Sydney asked politely.

"First, we want to look at the mean trajectory of maths scores across occasions using a line chart (see Figure 2)." Dominik typed and ran the SPSS syntax required to produce the line chart.

"As you can see, average maths scores are increasing across time in a linear trajectory. Our next step would be to determine whether the fixed, or average, effect of linear time is statistically significant, but before we do that we need to *centre* the time effect to ensure it has a meaningful zero point."

"Centre?" asked Sydney.

"Indeed. Centring requires that we create a new time variable by subtracting a meaningful value from the original time variable, such that the meaningful value we subtract becomes the 0-point for this new variable. Without centring, some results could be nonsense. For example, if we were to use your original, un-centred occasion variable in the HLM (*OccasionID* in the SPSS and SAS syntax), the intercept value and any interactions with this occasion variable would be interpreted

Downloaded by [66.45.159.208] at 07:54 27 November 2017

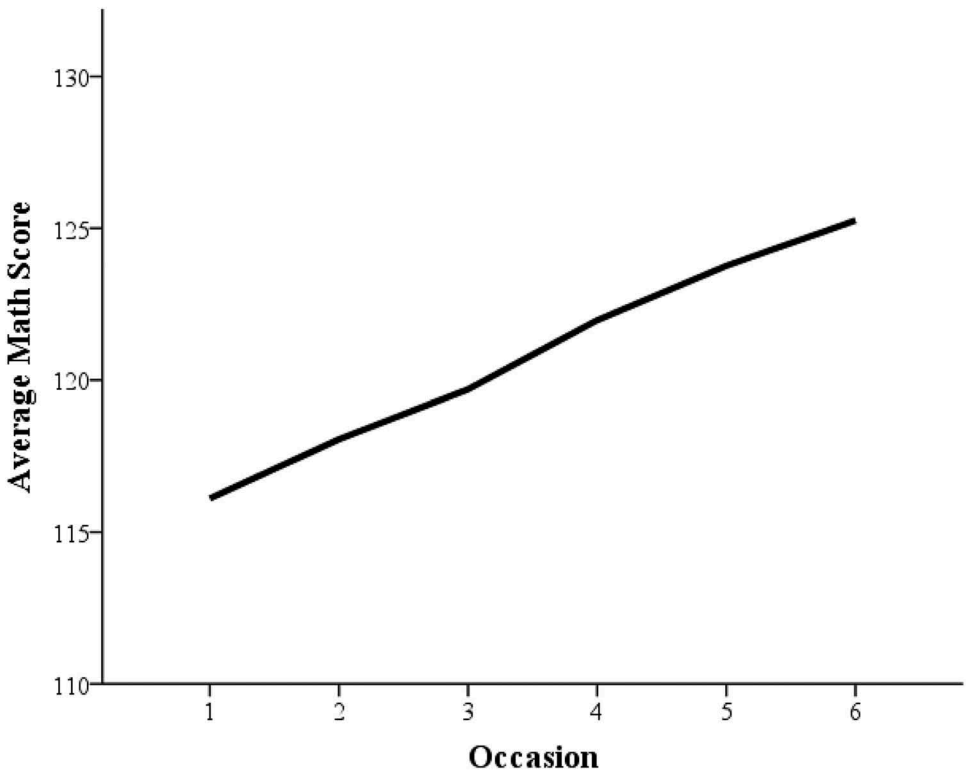

Figure 2. Average maths scores across the repeated occasions.

specifically at occasion 0, which does not exist given that the occasion variable ranges from 1 to 6! So, to centre this variable, we'll create a new variable that subtracts 1 from the original occasion variable (called *Time0* in the SPSS and SAS syntax), such that a value of 0 on this new time variable represents the first occasion, a value of 1 represents the second occasion, and so forth. Does that make sense?"

"I think so. So, the sixth occasion would be a value of 5 on the new centred time variable?" Sydney questioned.

"Correct. It is this new centred time variable that will be used in all subsequent HLMs." Dominik created the centred time variable in SPSS and estimated the HLM that included the fixed effect of linear time. He continued: "The HLM we just estimated is called a *fixed linear time, random intercept model*. The fixed intercept value is 116.2 and the fixed linear time effect of 1.9. Would you like to try to interpret these values?"

"Sure," Sydney said, shifting into a more upright position. "The intercept is the average maths score when all predictors equal 0. In this case a 0 on the centred time variable indicates the first occasion, so the average maths score at the first occasion (in which students were in third grade) is about 116 points."

"Correct!" praised Dominik.

Sydney continued: "Because the HLM models time as a continuous variable, the maths scores increased by an average of 1.9 points at each subsequent occasion, and this increase of 1.9 points is statistically significant given $p < .05$."

"Nice work!" exclaimed Dominik.

"Is an average change of 1.9 points a lot?" asked Sydney.

Dominik grinned slightly. "Whether an effect has any practical significance is up to the researcher, not the statistician. Remember, statistics are just tools. With that said, we can quantify how much of the *reason* for a student's maths score is due to the linear time effect by calculating pseudo-R^2."

"Pseudo-R^2?" questioned Sydney. "When I hear R^2, I think effect size. Is pseudo-R^2 similar to the R^2 effect size estimate from linear regression?"

"Well," chuckled Dominik: "in linear regression there is only one error term, the residual, and the variance of these residuals quantifies all of the reasons why an individual had a specific outcome value. The purpose of including predictors was to explain, or reduce, the variance of these residuals. By contrast, in the HLM there are *multiple* error terms, so there are multiple variance components available to be explained. This is why R^2 from the HLM is labelled as *pseudo-R^2*. 'Pseudo' in this case quite literally means 'one of many'. In the analyses so far, we have estimated two error terms — a random intercept and a residual. As I mentioned earlier, random effects are level 2, between-student effects whose variances are explained by student-level predictors, whereas residuals are level 1, within-student effects whose variance is explained by occasion-level predictors."

"Okay, so because the fixed linear time effect is an occasion-level predictor, it will explain residual variance."

"Yes." Dominik continued: "To calculate pseudo-R^2 for the fixed linear time effect, we compare the residual variance before and after we included the fixed effect of linear time."

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Dominik returned to his computer. "The residual variance was 280.9 from the unconditional, random intercept model, and it was 268.6 from the fixed linear time, random intercept model. Similar to linear regression, pseudo-R^2 represents the proportion *reduction* in variance comparing the model with *fewer* parameters (the unconditional, random intercept model) to the model with *more* parameters (the fixed linear time, random intercept model). It is calculated as:

$$\frac{\text{Variance}_{\text{fewer}} - \text{Variance}_{\text{more}}}{\text{Variance}_{\text{fewer}}}$$

For your data, 280.9–268.6 = 12.3, and 12.3 /280.9 = .04, so the linear time predictor explained about 4% of the level 1, residual variance."

"That is not a lot of variance explained," Sydney stated, disappointed.

"Although that may be true, the fixed effect of linear time represents a starting point that we will use to determine whether students change differently over time. That is, we need to determine whether this fixed, or average, effect adequately describes every student in the sample."

"Oh, you mean we also need to test for random between-student differences in the rate of change across years?" asked Sydney.

"Yes," said Dominik. "We do this by including a random effect of linear time in the HLM. This additional random effect is also known as a *random slope* and is partitioned out of residual variance just like the random intercept variance. Upon including this random effect, the HLM is now termed a *random linear time model*."

Dominik added the random effect of time to the SPSS syntax, estimated the model and scrolled to find the relevant output. "You can see here, Sydney, that the random effect variance for linear time is approximately 4.7."

"Does that indicate a significant amount of random between-student differences?" asked Sydney.

"Great question! Unlike fixed effects, for which we have *p*-values given by default, we need to use *model comparison* to assess the significance of new random effects. Specifically, model comparison requires the *likelihood ratio test* to determine whether the random linear time variance — and the covariance between the random intercept and random linear time effects — make the random linear time model fit the data significantly better than the fixed linear time, random intercept model. The likelihood ratio test is calculated by subtracting the –2 log-likelihood (–2LL) between the two comparison models. In general, when adding fixed or random effects, the –2LL will decrease, as is the case here as the –2LL decreased from 129,666.0 to 129,509.3 between models, a difference of 156.7. This difference of 156.7 actually follows a chi-square distribution, so to determine the statistical significance of this difference, we need to find the *critical value*, or cutoff value, from a chi-square distribution with degrees of freedom equal to the number of parameters we *added* when estimating the random linear time model."

"We added random linear time effect," added Sydney.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Downloaded by [66.45.159.208] at 07:54 27 November 2017

"Yes, but in doing so, we also added the random intercept-random linear time covariance, so the random linear time model added two parameters. As such, we have two degrees of freedom. You can look up critical values for the chi-square distribution with two degrees of freedom using any internet browser by searching for 'chi square critical'. I have conducted enough likelihood ratio tests to know that the critical value for two degrees of freedom is 5.99. Because 156.7 is greater than 5.99, our random linear time model fits the data significantly better than the fixed linear time, random intercept model."

"So, this means that the maths scores are increasing at different rates between students," remarked Sydney.

"Indeed, but your assumption that they are solely increasing might be too simplistic. We can see by calculating a 95% random effects confidence interval as we did earlier for the random intercept. In this case, $1.9 \pm 1.96\sqrt{4.7}$ results in a confidence interval ranging from −2.3 to 6.1, meaning that, although the maths scores for some students are predicted to increase more than average, the maths scores for other students are actually predicted to decrease."

"Well, that's not good!" exclaimed Sydney. "Does a *linear* effect of time sufficiently describe how students' maths scores change?"

"Probably," Dominik added. "Given the linear trajectory we saw in the line chart we created earlier and that the average maths scores are not close to the maximum, or ceiling, of 200 points, nonlinear effects of time are highly unlikely. However, the HLM can be used to test the significance of fixed and random polynomial (e.g., square, cubic), exponential or other kinds of nonlinear effects of time."

For completeness, Dominik walked Sydney through the estimation of both the *fixed quadratic, random linear time model* and the *random quadratic time model*, both of which resulted in nonsignificant fixed or random quadratic time effects. "As expected, the random linear time model is our final model for time. If, however, you find nonlinear effects in future HLMs, you will want to consult chapter six in Hoffman (2015) for interpretation and subsequent model-building advice."

"Thanks. So, why are some students increasing more than others, or worse, why are some students decreasing?" Sydney asked curiously.

"This is where other predictor variables serve their purpose. What predictor variables do you collect?"

"We collect free/reduced lunch status (*FRL* in the SPSS and SAS syntax) as a proxy for socio-economic status, which is a dichotomous variable in which a value of 1 indicates the student receives FRL and 0 indicates they do not. Unfortunately, we typically see that a student's FRL status does not change from year to year. We also administer a brief questionnaire that measures a student's self-efficacy before giving each achievement test."

"The FRL variable is what we call a *time-invariant* predictor given its values *do not change* within a student across occasions. By contrast, the self-efficacy variable is a *time-varying* predictor given that its value *can change* within a student from occasion-to-occasion. Including time-invariant predictors in the HLM is more direct, so that is where we will begin, but before doing so I want to know whether you expect the difference in average maths scores between students receiving or not receiving FRL to change across occasions. Alternatively, do you expect the rate of change in maths scores to differ between students who receive or do not receive FRL?"

"I'm not exactly sure what you mean," Sydney stated openly.

"If we include the FRL variable *by itself* in the HLM, the estimated difference in maths scores between students who receive or do not receive FRL will be considered *constant across all occasions*. Likewise, the rate at which maths scores change across occasions will be considered *identical* for students who receive or do not receive FRL."

Sydney contemplated briefly. "Similar effects across these two groups of students seem unlikely, but I cannot hypothesize a direction for this effect."

"We can address both by estimating a *cross-level interaction* effect."

"A cross-level interaction?" asked Sydney.

"Yes. A cross-level interaction is an interaction between a level 2, student-level predictor and a level 1, occasion-level predictor. I'll describe this effect further after we've modelled it." Dominik turned to his computer, added FRL and the linear time-by-FRL cross-level interaction, and estimated the HLM in SPSS. "The results indicate the cross-level interaction effect of −0.4 is statistically significant given $p < .05$. How much experience do you have interpreting interactions, Sydney?"

"Very little," Sydney remarked candidly.

"That's okay. Interpreting interaction effects can be confusing, but all can be remedied with practice. I recommend you consult chapter two in Hoffman (2015) for help interpreting interactions. Briefly, because there are two predictors in this cross-level interaction, it is considered a *two-way* interaction, meaning this effect modifies, or *moderates*, the main effects of the predictors included in the interaction. As a result, the fixed effects for linear time and FRL are now termed *simple main effects* interpreted specifically *when the interacting predictor equals 0*. So, with that in mind, the simple main effect for linear time of 2.0 is specifically for whom?"

Sydney began thinking out loud. "You said that the simple main effect of linear time is interpreted specifically when the interacting predictor, FRL in this case, equals 0. FRL equals 0 for students who do not receive FRL, so maths scores specifically for students who do not receive FRL increased by an average of 2.0 points at each subsequent occasion."

"Nice work! I really enjoyed your thought process," Dominik said, smiling widely. "The simple main effect of linear time for students who do not receive FRL is quite literally: 2.0–0.4*(FRL = 0) = 2.0–0 = 2.0."

"Okay, so if the linear time effect for students who *do not* receive FRL is 2.0, how do we get the linear time effect for students who *do* receive FRL?"

"Great question. Remember, the cross-level interaction effect indicates how a simple main effect is moderated, or changed, with changes in the value of the interacting predictor. So, given the fixed cross-level interaction effect of −0.4, if a student receives FRL, the linear time effect is *lower* by 0.4 points: 2.0–0.4*(FRL = 1) = 2.0–0.4 = 1.6. The fact that the interaction effect was statistically

Downloaded by [66.45.159.208] at 07:54 27 November 2017

significant indicates that this difference in linear time slopes between these two groups of students, 2.0 vs. 1.6, is statistically different."

Dominik continued: "Alternatively, the cross-level interaction effect also indicates that the *difference* in maths scores between students who receive or do not receive FRL *is becoming larger* across occasions. We can tell this by considering the simple main effect of FRL of −4.3, which indicates that, specifically at the first occasion, when time equals 0, maths scores for students who receive FRL were an average of 4.3 points *lower* compared to students who do not receive FRL: −4.3–0.4*(Time = 0) = −4.3–0 = −4.3. The *negative* cross-level interaction effect indicates that at each subsequent occasion, this difference became more negative, or *larger*, by an average of 0.4 points. So, at the second occasion, students who receive FRL average 4.7-point lower maths scores: −4.3–0.4*(Time = 1) = −4.3–0.4 = −4.7, compared to students who do not receive FRL. By the sixth occasion, students who receive FRL average 6.3-point lower maths scores: −4.3–0.4*(Time = 5) = −4.3–2.0 = −6.3."

"Interesting," remarked Sydney.

"Similar to the other fixed effects we've modelled, we also want to know how much variance was explained by FRL status."

"I assume you mean we can calculate pseudo-R^2, but we added FRL *and* the linear time-by-FRL cross-level interaction. You said earlier that student-level predictors explain random effect variances, but which random effect variance does FRL explain — random intercept variance, random linear time variance, or both?" Sydney asked.

"Great question, Sydney. The FRL effect explains between-student differences in maths scores represented by random intercept variance, whereas the linear time-by-FRL cross-level interaction effect explains between-student differences in the rate of change in maths scores across occasions represented by random linear time variance — we can calculate a pseudo-R^2 for each estimated effect. For the fixed effect of FRL, we compare the random intercept variance from the random linear time model before and after including FRL."

Dominik turned back to the SPSS output. "The random intercept variance was 249.1 before including FRL and 245.4 after, so using the same calculation we used earlier for the fixed linear time effect, 249.1–245.3 = 3.8, and 3.8 /249.1 = .02."

"This means that the FRL predictor explained about 2% of the random intercept variance," Sydney added correctly. "So, given that the cross-level interaction explains between-student differences in change, we need to consider the amount random linear time variance before and after adding this cross-level interaction."

"Correct!" Dominik praised as he returned to his computer. "We can see that random linear time variance decreased trivially from 4.72 to 4.69 after including the cross-level interaction, so pseudo-R^2 is: (4.72–4.62) /4.72 = .006."

"So, the statistically significant cross-level interaction explained less than 1% of the between-student differences in linear change," Sydney said, slightly

Downloaded by [66.45.159.208] at 07:54 27 November 2017

disappointed. "Can we test for *random* between-student differences in the FRL effect by estimating a random FRL effect?"

"Good question, Sydney, but no. Random between-student differences in FRL cannot be estimated because your data has only two levels. However, random FRL effects could be estimated if you considered additional *higher-level* nested structures, such as when repeatedly-measured students are nested within classrooms or schools. This type of *clustered-longitudinal analysis* would require a three- or four-level HLM and is a discussion for another day, but I would recommend consulting chapter 11 in Hoffman (2015) if you encounter this type of data."

"Okay, thanks. So, how does the self-efficacy predictor (labelled *SelfEff* in SPSS and SAS) explain additional variance?" Sydney asked.

"Self-efficacy is a *time-varying predictor* measured before every maths test. Modelling self-efficacy will be considerably more complicated than modelling time-invariant predictors, such as FRL, because time-varying predictors generally contain both between- and within-person effects. To correctly model a time-varying predictor, you first need to evaluate *how* the time-varying predictor changes over time. For your data, this means that you would initially model self-efficacy *as an outcome*, ignoring maths scores for the time being, to evaluate whether self-efficacy has any *random* time effects." Dominik noticed that Sydney looked slightly perplexed. He paused for a second and then continued. "Let's take a step back. It might help to consider that the maths scores we've been modelling as an outcome are also time-varying, so modelling fixed and random effects of time for self-efficacy as an outcome will follow the exact same sequence of steps we already discussed for those maths scores."

Sydney nodded and Dominik continued. "As I mentioned earlier, the appropriate modelling strategy for self-efficacy as a time-varying predictor of maths scores is dictated specifically by whether self-efficacy has any random time effects." Sydney relaxed. "Are you comfortable modelling that?" Dominik asked.

"I believe so. I have my notes and will have the SPSS code you've written."

"Good. If self-efficacy *does not* have any random time effects, then self-efficacy does not change differently between students. This means is that you can explicitly partition the between- and within-student effects of self-efficacy by creating two new variables that uniquely represents each effect using a procedure called *person-mean-centring*."

"By centring, I assume you mean that we are going to be subtracting some meaningful value like we did earlier for the occasion variable?" Sydney asked astutely.

"Yes, indeed," Dominik said, impressed. "Briefly, person-mean-centring requires that you first calculate average self-efficacy for each student *using the student's own self-efficacy measured across occasions*; this new variable will represent the *between-student* effect of self-efficacy. Next, you will subtract a student's average self-efficacy from their observed self-efficacy at each occasion; this variable will represent the *within-student* effect of self-efficacy. We are centring the within-student effect at each student's *own* average level of self-

Downloaded by [66.45.159.208] at 07:54 27 November 2017

efficacy, hence the name person-mean-centring, or in your case, student-mean-centring. You will *not* use the original self-efficacy variable in any subsequent HLM; instead, you will use the two newly-created variables as predictors."

"I think that makes sense," remarked Sydney. "So, how do I interpret these effects?"

"The between-student effect of self-efficacy represents *differences* in maths scores between students who *average* one-point higher self-efficacy. This student-level effect will explain random intercept variance if modelled by itself, but could also explain random linear time variance if included in a cross-level interaction with linear time."

"Similar to the linear time-by-FRL cross-level interaction," added Sydney.

"Exactly! By contrast, the within-student effect represents the change in maths score at an occasion when a student's self-efficacy is one-point higher *than their usual level*. This occasion-level effect explains residual variance. Does that make sense?"

Sydney finished writing her notes. "I believe so."

"It's okay if you're struggling a bit, Sydney. I definitely struggled when I was learning about person-mean-centring. I recommend that you consult section two in chapter nine of Hoffman (2015) for more information about person-mean-centring," Dominik added calmly.

"Thanks. So, what happens if I do find significant random time effects when modelling self-efficacy as an outcome?"

"In that case, using person-mean-centring is inappropriate because the HLM we've been estimating has no mechanism to account for the observed between-student differences in how self-efficacy changed over time. That is, if you were to ignore the random effects of time for self-efficacy and person-mean-centre self-efficacy, then the estimated between- and within-student fixed effects for self-efficacy would be incorrect."

"Incorrect fixed effects? Well, that's not good," remarked Sydney, slightly sarcastically. "So, if self-efficacy changes differently between students, what are my options?"

"In that situation, you would need to estimate a *multivariate* HLM in which both maths and self-efficacy variables are modelled as outcomes in a single model. As you might expect, a multivariate HLM is more complicated than the univariate HLM we have been talking about, in terms of both structuring your data for analysis and interpreting the results."

"So, the HLM we had been talking about for most of this meeting is considered *univariate* because we had only considered one outcome, maths?" Sydney asked.

"Exactly. By contrast, the multivariate HLM would allow you to include all necessary random effects for *both* maths and self-efficacy outcomes in the same model allowing you to evaluate how the outcomes relate to each other at each level: the between-student relationships between outcomes through the estimated *correlations of random effects* and the within-student relationships between outcomes through the estimated *correlations of residuals*."

Downloaded by [66.45.159.208] at 07:54 27 November 2017

“How would those correlations would be interpreted?” Sydney asked politely.

“A positive between-student random intercept correlation indicates that if a student averages higher time-0 maths scores *compared to other students*, they are likely to average higher time-0 self-efficacy compared to other students (or lower self-efficacy if the correlation is negative). A random slope correlation is interpreted similarly, but instead relates the student-specific slope when compared to other students. And a positive within-student residual correlation indicates that if a student has a higher maths score *than predicted at a given occasion*, their self-efficacy is predicated to be higher than predicted at that occasion (or lower than predicted if the correlation is negative).” Dominik glanced briefly at his watch. “I know the multivariate HLM is a lot to think about, but if your data requires a multivariate HLM please consult section two of chapter nine in Hoffman (2015).”

“Thanks,” Sydney said while finishing up her notes. Looking at her own watch, she exclaimed, “I’ve taken up a lot of your time today!”

“That is not a problem,” Dominik said with a laugh. “I really enjoy working with people who want to learn the material. As you start modelling your district’s data, you’ll inevitably come up with more questions, so if you ever want to meet again, please let me know.” Dominik saved the SPSS syntax file and handed Sydney her USB drive.

“I’ll email you later today after I write the SAS syntax.”

Sydney placed her notepad and pen in her tote bag and stood to leave.

“Thank you, Dominik. I really appreciate your time and have a great rest of your day.”

Dominik nodded and smiled. “You have a great day, too.”

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Downloaded by [66.45.159.208] at 07:54 27 November 2017

La aplicación del modelo lineal jerárquico a datos longitudinales

Sydney Farasyn fue contratada recientemente como especialista en gestión de datos en un distrito escolar formado por más de 15,000 alumnos de tercer a octavo curso[1], distribuidos en 52 centros de educación primaria y secundaria. Seis años atrás, el distrito empezó a recopilar datos anuales de rendimiento en las áreas de matemáticas, lectura y ciencias, aunque hasta ahora estos datos solo se habían evaluado anualmente con el fin de mantener un nivel coherente de financiación a nivel estatal y federal. Sydney se reunió recientemente con el superintendente del distrito, el cual, tras darse cuenta de la utilidad potencial de la evaluación de los datos académicos, le encomendó la tarea de analizar dichos datos con el fin de evaluar el cambio en el rendimiento académico de los alumnos a lo largo del tiempo. Dado que Sydney tiene una experiencia limitada en los modelos de datos longitudinales, decidió reunirse con Dominik Altmann, un profesor de una universidad de la zona que imparte cursos sobre modelos de datos longitudinales. Sydney acababa de llamar a la puerta del despacho de Dominik.

— Entra, por favor, Sydney. Toma asiento, dijo Dominik.

— Gracias por aceptar reunirse conmigo, Dr. Altmann, empezó a decir Sydney

Dominik la interrumpió — Por favor, llámame Dominik.

— De acuerdo. Gracias, continúo diciendo Sydney, notablemente tensa.

— Nuestro distrito empieza a recopilar datos sobre el rendimiento académico de un alumno una vez que este empieza el tercer curso. En estos momentos tengo seis años de datos de rendimiento académico en el área de matemáticas; los datos pertenecen a nuestro primer grupo de 2,700 alumnos que recientemente ha terminado octavo. Me han encomendado la tarea de evaluar cómo ha cambiado el rendimiento en el área de las matemáticas en este cohorte de alumnos entre los cursos de tercero y octavo.

— ¿Los resultados de matemáticas se miden en una escala continua?, preguntó Dominik.

— Sí, respondió Sydney. Las puntuaciones van de 0 a 200. Antes de que me lo preguntes, deja que te diga que la prueba se ha escalado verticalmente, lo que significa que cada prueba específica a cada uno de los grados incluye el contenido 'grado a grado' suficiente como para indicar de manera efectiva el incremento del rendimiento o crecimiento.

— ¿Tienes experiencia en modelos de datos longitudinales?, le preguntó Dominik.

— Un poco, aunque tengo mucha más experiencia en modelos de datos transversales, como cuando los alumnos están anidados intra-aula.

Sydney continuó diciendo: — He aplicado a estos datos unos cuantos modelos de análisis de varianza (ANOVA) con medidas repetidas (RM), pero mi programa de estadística me dice que solo tengo los datos de un 50% de los alumnos. No sé qué está fallando, terminó por decir, ligeramente confusa. Uno de mis compañeros me ha dicho que debo utilizar un modelo lineal jerárquico (HLM), así que he empezado a leer los capítulos sobre datos longitudinales en los libros de texto que tratan los HLM para datos transversales. No obstante, me ha costado mucho establecer una relación entre los HLM longitudinales y transversales; este es el motivo que me ha llevado a ponerme en contacto contigo.

Dominik sonrió. — Aunque el modelado de datos transversales y de datos longitudinales se solapan de manera considerable, ver dicho solapamiento puede resultar difícil. Y, en cuanto a los libros de texto, te recomiendo que consultes Hoffman (2015), que está íntegramente dedicado a modelos de datos longitudinales. También te recomiendo su libro de texto porque coincide directamente con cómo te explicaré la modelación de datos longitudinales durante nuestra reunión de hoy y las posteriores.

— ¡De acuerdo, gracias! Sydney sacó papel y bolígrafo de su bolsa y empezó a tomar notas.

Dominik continuó: — ¿Sabías que el modelo de análisis de varianza con medidas repetidas (RM-ANOVA) está incluido dentro del HLM?

— No, suspiró Sydney.

— Está bien, la tranquilizó Dominik, empezaremos por describir las estructuras de datos anidados y cómo RM-ANOVA, y otras formas del HLM, tratan los datos longitudinales.

— Me parece perfecto. Estoy preparada, respondió Sydney entusiasmada.

— Fantástico. Empecemos por describir el solapamiento del modelado de datos transversales y longitudinales. Con los datos transversales, era necesario explicar el anidado de alumnos intra-aulas porque, debido a los factores del aula, como por ejemplo el profesor, el rendimiento de los alumnos dentro de una misma aula compartía inevitablemente alguna asociación o correlación adicional. El mismo concepto se aplica a los datos longitudinales, ya que las puntuaciones obtenidas en matemáticas por un mismo alumno estarán más altamente correlacionadas que las puntuaciones en matemáticas obtenidas por diferentes alumnos. Dicho de otro modo, siempre estarás más correlacionada contigo misma que con otra persona. Esta correlación adicional, o dependencia, es precisamente aquello en lo que debes centrarte. Es más, es posible expresar un diseño longitudinal mediante terminología jerárquica tradicional diciendo que las puntuaciones de matemáticas medidas repetidamente a nivel 1 están anidadas intra-alumnos a nivel 2. RM-ANOVA es una manera de explicar la dependencia que se crea como resultado de una estructura jerárquica o anidada. ¿Sabías que hay dos tipos de modelos RM-ANOVA?

— No, respondió Sydney con modestia.

— No te preocupes, la tranquilizó Dominik, en breve hablaremos de los dos tipos y de cómo gestionan la dependencia. Primero, no obstante, quiero destacar cuatro razones por las cuales se estima que el HML es preferible al RM-ANOVA.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Aunque hoy abordaremos estas cuestiones de forma más bien resumida, en los capítulos uno y tres de Hoffman (2015) encontrarás una descripción completa al respecto.

Sydney cambió ligeramente de posición en su silla mientras Dominik continuaba.

— Muy bien. Primero, RM-ANOVA solo puede cuantificar diferencias *constantes* inter-alumno en las puntuaciones obtenidas en matemáticas a lo largo del tiempo. Aunque, efectivamente, esto reconoce que los alumnos tienen diferentes niveles de capacidad en matemáticas, un modelo RM-ANOVA limitaría dichas diferencias inter-alumno para que se mantuviesen constantes desde tercero a octavo.

— Creo que esto es un tanto simplista porque sabemos que los niveles de capacidad de los alumnos maduran a ritmos diferentes, puntualizó Sydney.

— Efectivamente. Este fue el caso de mis hijos. Al principio has dicho que te habían encomendado la tarea de evaluar el cambio en las puntuaciones obtenidas en matemáticas a lo largo de los años; no obstante, RM-ANOVA no puede estimar el cambio directamente porque el tiempo está modelado como categórico y no como continuo, y por lo tanto solo nos proporciona las diferencias *medias* inter-alumno de las puntuaciones obtenidas en matemáticas entre cada una de las ocasiones. Aunque, por un lado, el HLM puede cuantificar estas diferencias medias, por el otro también cuantifica diferencias *aleatorias* inter-alumno en el *cambio a lo largo del tiempo*, lo que te permitirá *predecir* por qué algunos alumnos cambian de manera diferente a otros alumnos.

— ¿A qué te refieres cuando hablas de diferencias aleatorias?, le preguntó Sydney.

— En tu caso, *aleatorio* significa que los alumnos cambian individualmente de manera diferente a lo largo del tiempo. Sydney asintió con la cabeza en signo de conformidad mientras Dominik continuaba. En segundo lugar, RM-ANOVA requiere que las ocasiones repetidas de matemáticas se *equilibren* de modo que se mida a cada uno de los alumnos en las mismas ocasiones. No obstante, si se mide a los alumnos en momentos de tiempo diferentes, el tiempo estará *desequilibrado* y el tiempo exacto cuando se midió a los alumnos deberá redondearse para adaptarlo a esos tiempos comunes. Sin embargo, el hecho de redondear introducirá un sesgo en tu análisis que podría generar resultados imprecisos.

— Bien, estoy contenta de que midamos a la mayoría de los alumnos aproximadamente al mismo tiempo en cada año académico, dijo Sydney aliviada.

— No obstante, tienes que saber que, si alguna vez te encuentras con ocasiones desequilibradas, el HLM modela el tiempo como una variable continua; esto te permitirá usar el tiempo exacto para asegurarte de que el cambio que se produce entre las diferentes ocasiones se modele con la máxima precisión posible.

Sydney se puso a tomar notas enérgicamente mientras Dominik continuaba. — La tercera razón por la cual el HLM es preferible al RM-ANOVA es que RM-ANOVA es una técnica de datos completa que utiliza algo llamado estimación de mínimos cuadrados. En tu caso, esto significa que será necesario disponer de

Downloaded by [66.45.159.208] at 07:54 27 November 2017

todas las puntuaciones de matemáticas de todos los alumnos en todas las ocasiones o, de lo contrario, los alumnos en cuestión quedarán excluidos del análisis.

Sydney levantó las cejas en signo de sorpresa. — Así, ¿si a un alumno le falta ni que sea una puntuación de matemáticas, quedará excluido de todo el análisis?

— Sí. Esto recibe el nombre de *listwise deletion* y es el motivo por el cual el 50% de los alumnos quedaron excluidos en tu RM-ANOVA inicial. En cambio, el HLM usa una estimación de máxima verosimilitud que emplea todos los datos disponibles. Si estás interesada en ello, puedes leer la sección tres del capítulo cinco de Hoffman (2015), donde encontrarás más información acerca de la estimación de máxima verosimilitud.

— Gracias. Es cierto que, por muchos motivos diferentes, no recopilamos los datos de todos los alumnos en todas las ocasiones; principalmente debido al hecho de que los alumnos cambian de centro, explicó Sydney.

Dominik asintió con la cabeza. — Como es probable que se produzcan bajas en cualquier diseño longitudinal, será fundamental poder utilizar todos los datos disponibles. Por último, hay dos tipos de RM-ANOVA que permiten describir el patrón de *variabilidad* en el rendimiento a lo largo de las diferentes ocasiones: el univariante y el multi-variante. En el RM-ANOVA *univariante*, las varianzas y las covarianzas de las puntuaciones de matemáticas están restringidas en un patrón llamado *simetría compuesta*, lo que significa que dichas varianzas y covarianzas de puntuaciones de matemáticas están limitadas a ser *constantes* a lo largo de las ocasiones. No obstante, si las puntuaciones de matemáticas del alumnado cambian a ritmos diferentes a lo largo de los años, la simetría compuesta será insostenible porque las varianzas también cambiarían necesariamente a lo largo de los años. En definitiva, es probable que la simetría compuesta sea excesivamente restrictiva. Así pues, para evitar la asunción de la simetría compuesta deberemos estimar un RM-ANOVA multi-variante en el que se estimen todas las varianzas y covarianzas, tal como se observan en tus datos, utilizando un patrón de varianza *desestructurado*.

Sydney pensó durante unos momentos. — El RM-ANOVA parece ser una muy buena opción.

— Aunque pueda parecer que el RM-ANOVA multi-variante es una opción bastante buena para tu caso, la estimación de mínimos cuadrados continuará necesitando datos completos, por lo que te continuaría faltando el 50% del alumnado. Como resultado de ello, el HLM es una opción mejor porque utiliza todos los datos disponibles y cuenta con muchas opciones a la hora de describir el patrón de variabilidad a lo largo del tiempo, incluidos la simetría compuesta y los patrones desestructurados. Te recomiendo que consultes los capítulos cuatro y cinco de Hoffman (2015), donde encontrarás una descripción completa de los patrones de varianza-covarianza disponibles.

Sydney cambió ligeramente de posición en la silla. — Muy bien, aunque por un lado entiendo por qué, para mis datos, el HLM es preferible al RM-ANOVA, por el otro no acabo de comprender qué querías decir antes cuando has afirmado que el HLM puede *cuantificar* y *predecir* diferencias inter-alumno aleatorias respecto a la cantidad y el cambio.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Dominik sonrió. — Es una muy buena pregunta, Sydney, y para responderla primero será necesario describir las dos vertientes del HLM. La primera vertiente consiste en el *modelo para las medias*, que incluye las pendientes de intercepto *fijo* y de regresión *fija* que se emplean para describir el punto de inicio *promedio* y el cambio *promedio* en las puntuaciones de matemáticas del alumnado. Estos efectos fijos se interpretan de manera parecida a los efectos en un modelo de regresión lineal tradicional.

Sydney añadió entonces, segura de sí misma: — Quieres decir que el intercepto fijo indicaría el promedio de la puntuación de matemáticas cuando todos los predictores equivalieran a 0; la pendiente de regresión fija para un predictor continuo indicaría el cambio promedio en la puntuación de matemáticas por el incremento de una unidad en el predictor; y la pendiente de regresión fija para un predictor categórico representaría la diferencia promedio en la puntuación de matemáticas entre dos grupos.

— ¡Exactamente!, exclamó Dominik, satisfecho. — La segunda vertiente del HLM consiste en el *modelo para la varianza*, que incluye cómo los *errores aleatorios* en la predicción se distribuyen y se relacionan con el alumnado y las ocasiones. En un modelo de regresión lineal, el único término de error es el residual que representa la *desviación* a partir de un valor de resultado predicho de un alumno y el valor de su resultado observado; es la varianza de dichos residuales lo que se explica, o se reduce, mediante las variables predictoras. No obstante, dada la estructura jerárquica de los datos longitudinales, los errores de predicción se producirán tanto inter-alumnos como intra-alumnos, recibiendo los errores inter-alumnos el nombre de *efectos aleatorios* y los errores intra-alumnos el de *residuales*. Estos errores continúan representando desviaciones, así que teniendo esto en mente, un *intercepto aleatorio* será la desviación específica del alumno a partir del promedio o del intercepto fijo; una *pendiente aleatoria* será la desviación específica del alumno a partir del promedio o la pendiente de regresión fija; y un *residual* será las desviación específica de la ocasión en función de una puntuación de matemáticas del alumno en una ocasión dada a partir de su propia trayectoria de crecimiento predicha. La *varianza* de los efectos aleatorios y los residuales es lo que cuantifica las diferencias inter-alumno e intra-alumno y nos dice si los predictores de nivel de alumno y/u ocasión deberían incluirse en el modelo para las medias.

— De acuerdo. Pero, ¿cómo sabré cuánta varianza hay a nivel de alumno y de ocasión?, preguntó Sydney.

— ¡Buena pregunta! La proporción de variabilidad a cada nivel puede cuantificarse calculando la *correlación intra-clase* (ICC) a partir de un HLM estimado que no incluya variables predictoras. Este HLM sin predictor recibe el nombre de *modelo de intercepto aleatorio incondicional* y tiene tres valores estimados: un intercepto fijo, una varianza de intercepto aleatorio y una varianza residual. En datos longitudinales con dos niveles de anidado, como cuando se anidan observaciones repetidas intra-alumno, los efectos aleatorios están separados de la varianza residual. Permíteme que te lo dibuje [véase la Figura 1]. Dado que asumimos que la varianza residual y la varianza de

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Downloaded by [66.45.159.208] at 07:54 27 November 2017

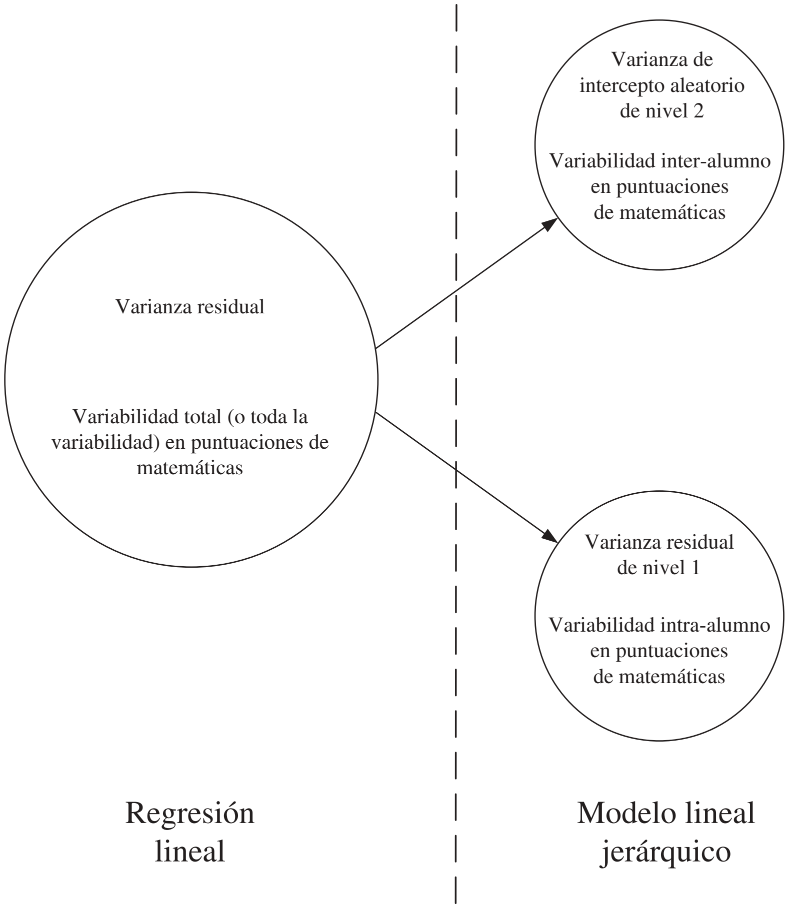

Figura 1. Cómo se separa la varianza utilizando el modelo lineal jerárquico de intercepto aleatorio. La estimación de la varianza residual que representa la variabilidad total a partir de la regresión lineal se separa en varianza inter-alumno e intra-alumno en un modelo lineal jerárquico.

intercepto aleatorio son independientes la una de la otra, la varianza total en los resultados de matemáticas en un HLM de dos niveles equivale, aproximadamente, a la *suma* de estos dos componentes de la varianza. La ICC es la proporción de *variabilidad total* existente entre alumnos y rangos entre 0 y 1. Esto se calcula del siguiente modo:

$$\frac{\text{Varianza del intercepto aleatorio}}{(\text{Varianza del intercepto aleatorio} + \text{varianza residual})} = \frac{\text{Varianza del intercepto aleatorio}}{\text{Varianza total}}.$$

— Creo que lo entiendo, dijo Sydney. — Es decir, una ICC de .60 indica que aproximadamente el 60% de la variabilidad en las puntuaciones de matemáticas se debe a factores inter-alumno.

— Una interpretación alternativa sería que el 40% de la variabilidad en las puntuaciones de matemáticas se debe a factores de nivel de ocasión. Este modelo de intercepto aleatorio incondicional sirve como punto de partida para tu proceso de construcción de modelo. Creo que, en vez de hablar de estos modelos, deberíamos utilizarlos. ¿Has traído los datos de matemáticas?

— Sí, respondió Sydney mientras extraía un USB de la bolsa y se lo entregaba a Dominik. — Suelo utilizar el software SPSS para realizar la mayoría de mis análisis estadísticos, pero también tengo acceso a SAS. Supongo que podemos utilizar cualquiera de los dos programas para ejecutar el HLM, ¿no?

— ¡Por supuesto! Aunque hoy trabajaremos principalmente con SPSS, me aseguraré de que tengas la sintaxis tanto de SPSS como de SAS [véase el Apéndice]. Dominik insertó el USB en su ordenador y abrió los datos de Sydney (*Data.csv*) con SPSS. A continuación introdujo la sintaxis de SPSS necesaria para estimar el modelo de intercepto aleatorio incondicional.

— El resultado nos muestra que el intercepto fijo es 120.9, que la varianza del intercepto aleatorio es 286.1 y que la varianza residual es 280.9. Esto significará que la ICC es: 286.1 /(286.1 + 280.9) = .50. ¿Quieres interpretar la ICC?, le preguntó Dominik.

— Así pues, aproximadamente el 50% de la variabilidad en las puntuaciones de matemáticas se debe a los factores inter-alumno, mientras que el 50% restante se sitúa a nivel de ocasión, dijo Sydney con cierta astucia.

— ¿Hay mejor manera de describir la varianza del intercepto aleatorio? Quiero decir, ¿podemos dotar el 286.1 de algún contexto?

Dominik alzó ambas cejas mostrándose impresionado por la intuición de Sydney. — ¡Por supuesto! Podemos calcular un *intervalo de confianza en efectos aleatorios del 95%*, lo cual cuantifica el rango de interceptos específicos del alumno en el 95% del alumnado de tu muestra. El intervalo de confianza se calcula del siguiente modo:

$$\text{Efecto fijo} \pm 1.96\sqrt{\text{efecto aleatorio de la varianza}}.$$

Aquí, $120.9 \pm 1.96\sqrt{286.1}$ indica que el 95% de los alumnos ha obtenido unas puntuaciones de matemáticas promedio de entre 87.7 y 154.1.

— ¡Esto me es de gran utilidad! Pero, ¿ahora qué hacemos?, preguntó Sydney.

— Ahora tomamos el modelo y lo aplicamos al tiempo. Recuerda que lo que te interesa es el *cambio* a lo largo del tiempo, lo que significa que necesitamos asegurarnos de que el modelo relativo al tiempo incluya todos los efectos fijos y aleatorios relevantes de tiempo para, de este modo, captar con exactitud cualquier diferencia inter-alumno en el cambio.

— ¿Cómo podemos hacerlo?, preguntó Sydney de manera educada.

— Primero, con la ayuda de un gráfico de líneas, observamos la trayectoria media de las puntuaciones de matemáticas en todas las ocasiones [véase la Figura 2]. Dominik introdujo y ejecutó la sintaxis de SPSS para elaborar el gráfico de líneas.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Downloaded by [66.45.159.208] at 07:54 27 November 2017

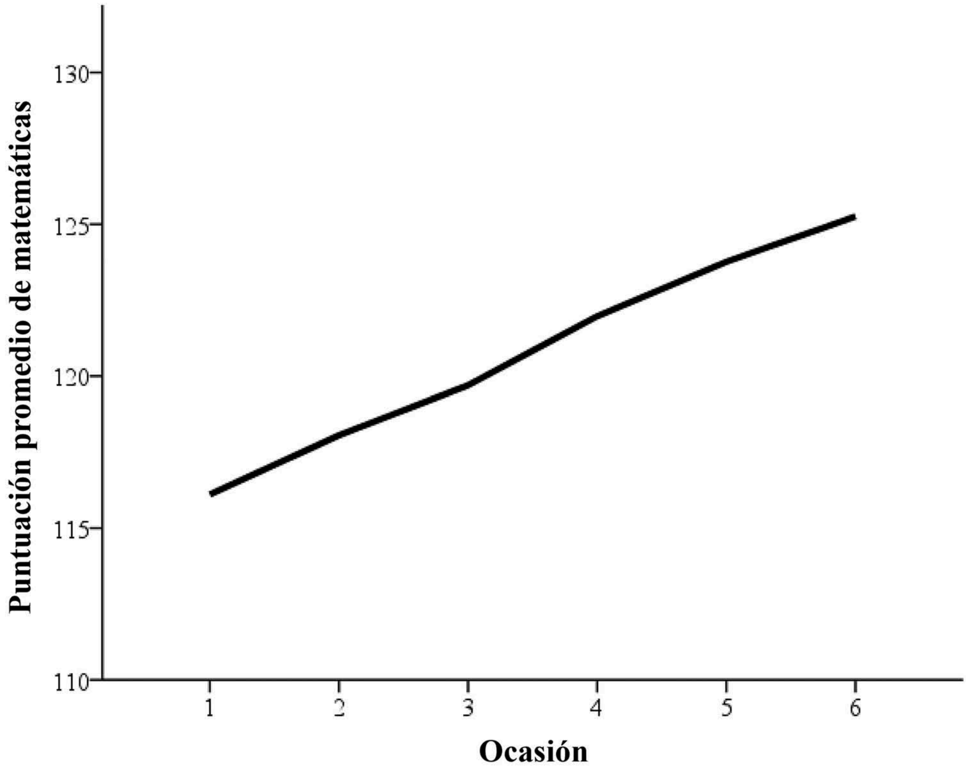

Figura 2. Puntuaciones promedio de matemáticas a lo largo de repetidas ocasiones.

— Como puedes ver aquí, las puntuaciones promedio de matemáticas aumentan a lo largo del tiempo en una trayectoria lineal. Nuestro próximo paso será determinar si el efecto fijo o promedio de tiempo lineal es estadísticamente significativo; no obstante, antes de eso tenemos que *centrar* el efecto del tiempo para asegurarnos de que tenga un punto cero significativo.

— ¿Centrar?, preguntó Sydney.

— Efectivamente. El centrado requiere que creemos una nueva variable de tiempo a través de restar un valor significativo de la variable de tiempo original, de modo que el valor significativo que restemos se convierta en el punto 0 para esta nueva variable. Sin el centrado, algunos resultados podrían no tener ningún sentido. Por ejemplo, si usáramos tu variable original de la ocasión descentrada en el HLM (*OcassionID* en la sintaxis de SPSS y SAS), el valor del intercepto y cualquier interacción con esta variable de la ocasión se interpretaría concretamente en la ocasión 0, la cual no existe porque la variable de la ocasión oscila entre 1 y 6. Así pues, para centrar esta variable crearemos una nueva variable que reste 1 de la variable original de la ocasión (que en la sintaxis de SPSS y de SAS recibe el nombre de *Time0*), de modo que un valor de 0 en esta nueva variable represente la primera ocasión, un valor de 1 represente la segunda ocasión, y así sucesivamente. ¿Lo entiendes?.

— Creo que sí. Es decir, ¿la sexta ocasión equivaldría a un valor de 5 en la nueva variable de tiempo centrada?, preguntó Sydney.

— Correcto. Esta nueva variable de tiempo centrada es la que utilizaremos en todos los HLM subsiguientes. Dominik creó la variable de tiempo centrada en SPSS y estimó el HLM que incluía el efecto fijo del tiempo lineal. Y prosiguió: — El HLM que acabamos de estimar recibe el nombre de *modelo de intercepto aleatorio de tiempo lineal fijo*. El valor del intercepto fijo es de 116.2, mientras que el efecto del tiempo lineal fijo es de 1.9. ¿Quieres intentar interpretar estos valores?

— Claro, dijo Sydney a la vez que adoptaba una posición más erguida sobre la silla. El intercepto es la puntuación promedio de matemáticas cuando todos los predictores son igual a 0. En este caso, un 0 en la variable de tiempo centrada indica la primera ocasión, de modo que la puntuación promedio de matemáticas en la primera ocasión (en la que el alumnado en cuestión cursaba tercero) es de unos 116 puntos.

— ¡Correcto!, la felicitó Dominik.

Sydney continuó: — Gracias a que el tiempo de los modelos HLM es una variable continua, las puntuaciones de matemáticas aumentaron en un promedio de 1.9 puntos en cada ocasión subsiguiente, y este aumento de 1.9 puntos es estadísticamente significativo dado que $p < .05$.

— ¡Buen trabajo!, exclamó Dominik.

— ¿Es mucho, un cambio promedio de 1.9 puntos?, preguntó Sydney.

Dominik sonrío. Que un efecto tenga o no una significatividad parcial es algo que dependerá del investigador, no del estadístico. Recuerda, la estadística es solo una herramienta. Dicho eso, podemos cuantificar qué cantidad del *motivo* de la puntuación de matemáticas de un alumno se debe a un efecto de tiempo lineal calculando el pseudo-R^2.

— ¿El pseudo-R^2?, preguntó Sydney. Cuando oigo R^2 pienso en el tamaño del efecto. ¿Es pseudo-R^2 similar a la estimación del tamaño del efecto de R^2 de la regresión lineal?

— Bueno…, rió Dominik, en la regresión lineal solo hay un término de error, el residual, y la varianza de estos residuales cuantifica todos los motivos por los cuales una persona tenía un valor de resultado específico. El propósito de incluir predictores era explicar, o reducir, la varianza de estos residuales. Por otro lado, en el HLM hay *múltiples* términos de error, lo que quiere decir que existe una multiplicidad de componentes de la varianza que será necesario explicar. Este el motivo por el cual la R^2 del HLM recibe el nombre de *pseudo-R^2*. 'Pseudo', en este caso, significa casi literalmente 'uno de muchos'. En los análisis realizados hasta ahora hemos estimado dos términos de error: un intercepto aleatorio y un residual. Como he mencionado anteriormente, los efectos aleatorios son efectos inter-alumno de nivel 2 cuyas varianzas se explican mediante los predictores de nivel de alumno, mientras que los residuales son efectos intra-alumno de nivel 1 cuya varianza se explica mediante predictores de nivel de ocasión.

— Muy bien, esto quiere decir que el hecho de que el efecto de tiempo lineal fijo sea un predictor de nivel de ocasión nos explicará la varianza residual.

— Sí, y continuó diciendo Dominik: Para calcular la pseudo-R^2 relativa al efecto de tiempo lineal fijo, comparamos la varianza residual antes y después de incluir el efecto fijo del tiempo lineal.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Dominik volvió a fijar la vista en la pantalla de su ordenador. — La varianza residual era 280.9 en modelo de intercepto aleatorio incondicional, y de 268.6 en modelo de intercepto aleatorio de tiempo lineal fijo. De manera parecida a la regresión lineal, la pseudo-R^2 representa la reducción proporcional en la varianza, si comparamos el modelo con *menos* parámetros (el modelo de intercepto aleatorio incondicional) con el modelo con *más* parámetros (el modelo de intercepto aleatorio de tiempo lineal fijo). Esto se calcula del siguiente modo:

$$\frac{\text{Varianza}_{\text{menos}} - \text{Varianza}_{\text{más}}}{\text{Variance}_{\text{menos}}}$$

Con relación a tus datos, 280.9–268.6 = 12.3, y 12.3 /280.9 = .04; es decir, el predictor de tiempo lineal explica sobre un 4% de la varianza residual de nivel 1.

— Eso no explica gran cosa de la varianza, afirmó Sydney decepcionada.

— Aunque por un lado tienes razón, el efecto fijo del tiempo lineal representa un punto de partida que usaremos para determinar si los alumnos cambian de manera diferente a lo largo del tiempo. Es decir, necesitamos determinar si el efecto fijo, o promedio, describe de manera adecuada a cada uno de los alumnos de la muestra.

— Ah, ¿quieres decir que también tenemos que probar las diferencias aleatorias inter-alumno en el ritmo de cambio a lo largo de los años?, preguntó Sydney.

— Sí, respondió Dominik. Para ello, deberemos incluir un efecto aleatorio del tiempo lineal en el HLM. Este efecto aleatorio adicional también recibe el nombre de *pendiente aleatoria* y está separada de la varianza residual, igual que la varianza de intercepto aleatorio. Después de haber incluido este efecto aleatorio, el HLM recibe ahora el nombre de *modelo de tiempo lineal aleatorio*.

Dominik agregó el efecto aleatorio del tiempo a la sintaxis del SPSS, estimó el modelo y se fue desplazando hasta encontrar el resultado relevante. — Aquí puedes ver, Sydney, que la varianza del efecto aleatorio relativo al tiempo lineal es de aproximadamente 4.7.

— ¿Indica eso una cantidad significativa de diferencias aleatorias inter-alumno?, preguntó Sydney.

— ¡Buena pregunta! A diferencia de los efectos fijos, para los cuales tenemos valores *p* dados por defecto, lo que aquí necesitamos es usar una *comparación de modelos* que nos permita evaluar la significatividad de nuevos efectos aleatorios. Concretamente, la comparación de modelos requiere la *prueba de razón de verosimilitud* para determinar si la varianza de tiempo lineal aleatorio, y la covarianza entre el intercepto aleatorio y los efectos de tiempo lineal aleatorios, consigue que el modelo de tiempo lineal aleatorio se adapte significativamente mejor a los datos que el modelo del intercepto aleatorio de tiempo lineal fijo. La prueba de razón de verosimilitud se calcula restando el –2 log-verosimilitud (–2LL) entre los dos modelos de comparación. En general, cuando añadimos efectos fijos o aleatorios, el –2LL disminuye, como sucede aquí cuando el –2LL disminuye de 129,666.0 a 129,509.3 entre modelos, una diferencia de 156.7. De hecho, la diferencia de 156.7 sigue una distribución ji cuadrada, por lo que para

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Downloaded by [66.45.159.208] at 07:54 27 November 2017

determinar la significatividad estadística de esta diferencia necesitaremos encontrar el *valor crítico*, o valor límite, a partir de una distribución ji cuadrada con grados de libertad equivalentes a la cantidad de parámetros que hemos *añadido* cuando estimábamos el modelo de tiempo lineal aleatorio.

— Añadimos el efecto de tiempo lineal aleatorio, dijo Sydney.

— Sí, pero al hacerlo también añadimos la covarianza de tiempo lineal aleatorio-intercepto aleatorio, por lo que el modelo de tiempo lineal aleatorio añade dos parámetros. Esto significa que tenemos dos grados de libertad. Puedes buscar valores críticos para la distribución ji cuadrada con dos grados de libertad a través de cualquier navegador de Internet buscando 'crítico ji cuadrado'. He realizado suficientes pruebas de razón de verosimilitud como para saber que el valor crítico relativo a dos grados de libertad es 5.99. Dado que 156.7 es mayor que 5.99, nuestro modelo de tiempo lineal aleatorio se adapta significativamente mejor a los datos que el modelo del intercepto aleatorio de tiempo lineal fijo.

— Esto significa que las puntuaciones de matemáticas aumentan en índices diferentes entre alumnos, observó Sydney.

— Efectivamente, aunque tu asunción de que únicamente aumentan puede que sea un poco simplista. Esto puede apreciarse si calculamos un intervalo de confianza de efectos aleatorios del 95%, como hemos hecho antes con relación al intercepto aleatorio. En este caso, $1.9 \pm 1.96\sqrt{4.7}$ resulta en un intervalo de confianza que oscila entre −2.3 y 6.1, lo que significa que, aunque se predice que las puntuaciones obtenidas en matemáticas para algunos alumnos aumentan más del promedio, también se predice que las puntuaciones de matemáticas de otros alumnos disminuyen.

— ¡Vaya, pues no vamos bien!, se exclamó Sydney. ¿Un efecto de tiempo *lineal* describe cómo cambian las puntuaciones de matemáticas de los alumnos?

— Probablemente, respondió Dominik. Dado que la trayectoria lineal que hemos visto en el gráfico de líneas que hemos creado antes y que las puntuaciones promedio de matemáticas no se acercan mucho al máximo, o techo, de 200 puntos, los efectos de tiempo no lineales serán altamente improbables. No obstante, el HLM puede utilizarse para probar la significatividad de los efectos de tiempo polinómicos (e.g., cuadrado o cúbico), exponenciales u otros tipos de efectos de tiempo no lineales.

Para mayor exactitud, Dominik guió a Sydney a través de la estimación del *modelo de tiempo lineal aleatorio de cuadrático fijo* y del *modelo de tiempo cuadrático aleatorio*, y ambos obtuvieron como resultado efectos de tiempo cuadrático aleatorio o fijo no significativos. — Como cabía esperar, el modelo de tiempo lineal es nuestro modelo final para el tiempo. No obstante, si encuentras efectos no lineales en futuros HLM, puedes consultar el capítulo seis de Hoffman (2015) dedicado a la interpretación y los subsiguientes consejos de construcción de modelo.

— Gracias. Así, pues, ¿por qué algunos alumnos aumentan más que otros, o lo que es peor, por qué algunos disminuyen?, preguntó Sydney con curiosidad.

— Aquí es cuando entran en juego otras variables predictoras. ¿Qué variables predictoras recoges?

— Recopilamos estados de almuerzos con descuento/gratuitos (*FRL* en la sintaxis de SPSS y SAS) como indicador del estado socio-económico, que es una variable dicótomica en la que un valor 1 indica que el alumno recibe el FRL (almuerzo con descuento/gratuito) y un valor 0 indica que no lo recibe. Lamentablemente, solemos ver que el estado de FRL de un alumno no cambia de año a año. También administramos un breve cuestionario que mide la auto-eficacia de un alumno antes de entregarle las pruebas de rendimiento.

— La variable FRL es lo que llamamos un predictor *invariante con el tiempo*, ya que sus valores *no cambian* en un mismo alumno en ocasiones diferentes. Por otro lado, la variable de auto-eficacia es un predictor que *varía con el tiempo* dado que su valor *puede cambiar* en un mismo alumno de una ocasión a otra. Incluir los predictores invariantes con el tiempo en el HLM es más directo, así que es ahí donde empezaremos; pero antes de ello, quiero saber si esperas que la diferencia en las puntuaciones promedio de matemáticas cambie de una ocasión a otra entre los alumnos que reciben el FRL y los que no lo reciben. De manera alternativa, ¿esperas que el índice de cambio en puntuaciones de matemáticas sea diferente entre los alumnos que reciben el FRL y los que no?

— No estoy segura de entenderte, admitió Sydney.

— Si incluimos la variable FRL *por sí sola* en el HLM, se considerará que la diferencia estimada en las puntuaciones de matemáticas entre los alumnos que reciben o no el FRL es *constante en todas las ocasiones*. Del mismo modo, el índice en el que cambian las puntuaciones matemáticas en todas las ocasiones se considerará como idéntico en todos los alumnos independientemente de que reciban o no el FRL.

Sydney pensó en silencio durante unos instantes — La existencia de efectos similares en estos dos grupos de alumnos parece improbable, aunque no puedo especular sobre qué dirección podría tener este efecto.

— Podemos tratarlos a los dos a través de estimar un efecto de *interacción multi-nivel*.

— ¿Una interacción multi-nivel?, preguntó Sydney.

— Sí. Una interacción multi-nivel es una interacción entre un predictor de nivel de alumnos, de nivel 2 y un predictor de nivel de ocasión, de nivel 1. Ahora, una vez modelado, describiré este efecto más detalladamente. Dominik volvió a clavar vista en la pantalla de su ordenador, añadió la interacción multi-nivel de tiempo lineal-por-FRL, y el FRL, y realizó una estimación de HLM en SPSS. — El resultado indica que el efecto de la interacción multi-nivel de −0.4 es estadísticamente significativa dado que $p < .05$. ¿Tienes experiencia en la interpretación de interacciones, Sydney?

— Muy poca, respondió Sydney con cierta candidez.

— Está bien. La interpretación de los efectos de la interacción puede ser desconcertante, pero no hay nada que la práctica no solucione. Te recomiendo que consultes el capítulo dos de Hoffman (2015), el cual te ayudará a interpretar las interacciones. Dicho en pocas palabras, dado que en esta interacción multi-nivel hay dos predictores, se considera que esta es una interacción *bidireccional*, lo que significa que este efecto modifica, o *modera*, los efectos principales de los

Downloaded by [66.45.159.208] at 07:54 27 November 2017

predictores incluidos en la interacción. Como resultado de ello, los efectos fijos relativos al tiempo lineal y el FRL reciben ahora el nombre de *efectos simples principales* y se interpretan concretamente *cuando el predictor de interacción equivale a 0*. Por lo tanto, y teniendo esto presente, ¿a quién hace referencia específicamente el efecto simple principal relativo al tiempo lineal de 2.0?

Sydney empezó a pensar en voz alta: — Antes has dicho que el efecto simple principal del tiempo lineal se representa específicamente cuando el predictor de interacción, en este caso el FRL, es equivalente a 0. FRL es igual a 0 para alumnos que no reciben el FRL, por lo que las puntuaciones de matemáticas específicas de los alumnos que no reciben el FRL aumentaron en un promedio de 2.0 puntos en cada ocasión subsiguiente.

— ¡Buen trabajo! Me gusta mucho como razonas, dijo Dominik mientras sonreía. El efecto simple principal del tiempo lineal relativo a los alumnos que no reciben el FRL es, literalmente, el siguiente: 2.0–0.4*(FRL = 0) = 2.0–0 = 2.0.

— Así, si el efecto del tiempo lineal relativo a los alumnos que *no* reciben el FRL es 2.0, ¿cómo obtenemos el efecto del tiempo lineal relativo a los alumnos que *sí* que reciben el FRL?.

— ¡Buena pregunta! Recuerda que el efecto de interacción multi-nivel indica cómo se modera o cambia un efecto simple principal con cambios en el valor del predictor de interacción. Dado el efecto de la interacción multi-nivel fijo de −0.4, si un alumno recibe el FRL, el efecto del tiempo lineal es 0.4 puntos *inferior*: 2.0–0.4*(FRL = 1) = 2.0–0.4 = 1.6. El hecho de que el efecto de la interacción fuera estadísticamente significativo indica que esta diferencia en pendientes de tiempo lineal entre estos dos grupos de alumnos, 2.0 vs. 1.6, es estadísticamente diferente.

Dominik continuó: — De manera alternativa, el efecto de interacción multi-nivel también indica que la *diferencia* en puntuaciones de matemáticas entre alumnos que reciben o no reciben el FRL *se está ensanchando* a lo largo de las ocasiones. Esto es algo que podemos afirmar después de haber tenido en cuenta el efecto simple principal del FRL de −4.3, el cual indica que, específicamente en la primera ocasión, en que el tiempo es igual a 0, las puntuaciones de matemáticas de los alumnos que reciben el FRL tenían un promedio de 4.3 puntos *menos* en comparación con las de los alumnos que no reciben el FRL: −4.3–0.4*(Time = 0) = −4.3–0 = −4.3. El efecto de interacción multi-nivel *negativo* indica que en cada ocasión subsiguiente, esta diferencia se convierte en más negativa, o *mayor*, en un promedio de 0.4 puntos. Así pues, en la segunda ocasión, los alumnos que reciben el FRL presentan un promedio de 4.7 puntos menos en sus puntuaciones de matemáticas: −4.3–0.4*(Time = 1) = −4.3–0.4 = −4.7, comparado con los alumnos que no reciben el FRL. Cuando llegamos a la sexta ocasión, los alumnos que reciben el FRL presentan un promedio de 6.3 puntos menos en sus puntuaciones de matemáticas: −4.3–0.4*(Time = 5) = −4.3–2.0 = −6.3.

— Muy interesante, comentó Sydney.

— De manera parecida a otros efectos fijos que hemos modelado, ahora también queremos saber cómo se explica tanta varianza a través del estado del FRL.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

— Asumo que quieres decir que podemos calcular la pseudo-R^2, pero hemos añadido el FRL *y* la interacción multi-nivel de tiempo lineal-por-FRL. Antes has dicho que los predictores de nivel de los alumnos explican varianzas de efectos aleatorios; no obstante, ¿qué varianza de efecto aleatorio explica el FRL: la varianza de intercepto aleatorio, la varianza de tiempo lineal aleatorio, o ambas?, preguntó Sydney.

— ¡Buena pregunta, Sydney! El efecto FRL explica las diferencias en puntuaciones de matemáticas entre alumnos, representadas mediante la varianza de intercepto aleatorio, mientras que el efecto de interacción multi-nivel de tiempo lineal-por-FRL explica las diferencias en el índice de cambio entre las puntuaciones de matemáticas entre alumnos de unas ocasiones a otras, representadas mediante la varianza de tiempo lineal aleatorio, pudiéndose calcular una pseudo-R^2 para cada efecto estimado. Con relación al efecto fijo del FRL, comparamos la varianza de intercepto aleatorio a partir del modelo de tiempo lineal aleatorio antes y después de haber incluido el FRL.

Dominik volvió al resultado del SPSS. — La varianza de intercepto aleatorio era de 249.1 antes de incluir el FRL y de 245.4 después de incluir el FRL, por lo que usando el mismo cálculo que utilizamos anteriormente para el efecto de tiempo lineal fijo, tenemos que 249.1–245.3 = 3.8, y 3.8 /249.1 = .02.

— Esto significa que el predictor FRL nos has explicado aproximadamente el 2% de la varianza de intercepto aleatorio, añadió correctamente Sydney. Por lo tanto, dado que la interacción multi-nivel explica las diferencias inter-alumno en el cambio, ahora será necesario tener en cuenta la varianza de tiempo lineal aleatorio antes y después de añadir esta interacción multi-nivel.

— ¡Correcto!, respondió Dominik mientras volvía a su ordenador. Podemos ver que la varianza de tiempo lineal disminuía de manera poco significativa de 4.72 a 4.69 después de incluir la interacción multi-nivel, por lo que la pseudo-R^2 es: (4.72–4.62) /4.72 = .006.

— Esto significa que la interacción multi-nivel, estadísticamente significativa, explicaba menos de un 1% de las diferencias inter-alumno en cambio lineal, dijo Sydney ligeramente decepcionada. ¿Podemos verificar las diferencias *aleatorias* inter-alumno en el efecto FRL a través de estimar un efecto FRL aleatorio?

— Buena pregunta, Sydney, pero no. Las diferencias aleatorias inter-alumno en FRL no pueden estimarse porque tus datos solo tienen dos niveles. No obstante, los efectos FRL aleatorios podrían estimarse si tuvieses en cuenta las estructuras anidadas adicionales de *nivel superior*, como cuando los alumnos medidos repetidamente están anidados en aulas o centros. Este tipo de *análisis longitudinal agrupado* requeriría un HLM de tres o cuatro niveles, lo cual es un debate para otro día; no obstante, si te encuentras con este tipo de datos, yo te recomendaría que consultases el capítulo 11 de Hoffman (2015).

— Muy bien, gracias. Así pues, ¿cómo explica la varianza adicional el predictor de auto-eficacia (etiquetado como *SelfEff* en SPSS y SAS)?, preguntó Sydney.

— La auto-eficacia es un *predictor que varía con el tiempo* medido antes de cada prueba de matemáticas. Modelar la auto-eficacia será considerablemente más complejo que modelar los predictores invariantes con el tiempo, como el FRL,

Downloaded by [66.45.159.208] at 07:54 27 November 2017

porque por norma general los predictores que varían con el tiempo contienen efectos inter e intra-personales. Para modelar correctamente un predictor que varía con el tiempo, primero será necesario evaluar *cómo* cambia a lo largo del tiempo el predictor que varía con el tiempo. Con respecto a tus datos, esto significa que inicialmente modelarías la auto-eficacia *como si fuera un resultado*, ignorando por el momento las puntuaciones de matemáticas, con el fin de evaluar si la auto-eficacia tiene efectos de tiempo *aleatorio*.

Dominik notó cierta expresión de perplejidad en el rostro de Sydney, así que hizo una breve pausa y entonces continuó: — Demos un paso atrás. Esto podría ayudarnos a tener en cuenta que las puntuaciones de matemáticas que hemos modelado como si fueran un resultado también varían con el tiempo, lo que significa que la modelación de efectos de tiempo fijos y aleatorios para la auto-eficacia a modo de resultado seguirá exactamente la misma secuencia de pasos que ya hemos abordado anteriormente para dichas puntuaciones de matemáticas.

Sydney asintió con la cabeza y Dominik continuó: — Como he mencionado antes, la estrategia de modelación apropiada para la auto-eficacia como predictor de puntuaciones de matemáticas que varía con el tiempo viene dictada específicamente por si la auto-eficacia presenta o no efectos de tiempo aleatorios. Sydney se relajó. — Te sientes cómoda modelando esto?, le preguntó Dominik.

— Creo que sí. Tengo mis propias notas y además también tendré el código de SPSS que tú has escrito.

— Perfecto. Si la auto-eficacia *no* tiene ningún efecto de tiempo aleatorio, eso querrá decir que la auto-eficacia no cambia de manera diferente entre alumnos. Esto significa que podrás separar de manera explícita los efectos de auto-eficacia inter-alumno e intra-alumno a través de crear dos nuevas variables que representen exclusivamente a cada uno de los efectos empleando un procedimiento llamado *persona-media-centrado*.

— Asumo que, por centrado, quieres decir que restaremos algún valor significativo como hicimos antes con relación a la variable de ocasión…, preguntó Sydney astutamente.

— Sí, claro, dijo Dominik, impresionado. — Expresado de manera sintética, persona-media-centrado requiere que primero calcules el promedio de auto-eficacia con relación a cada alumno *utilizando la auto-eficacia propia de cada alumno medida a lo largo de las ocasiones*; esta nueva variable representará el efecto *inter-alumno* de auto-eficacia. A continuación restarás el promedio de auto-eficacia de un alumno de su auto-eficacia observada en cada ocasión; esta variable representará el efecto *intra-alumno* de auto-eficacia. Estamos centrando el efecto intra-alumno en el nivel promedio de auto-eficacia *propio* de cada alumno; de ahí el nombre de persona-media-centrado o, en tu caso, de alumno-media-centrado. *No* utilizarás la variable de auto-eficacia original en ninguno de los HLM subsiguientes, sino que usarás las dos variantes de reciente creación como predictores.

— Creo que tiene sentido, dijo Sydney. Así, entonces, ¿cómo interpreto estos efectos?

Downloaded by [66.45.159.208] at 07:54 27 November 2017

— El efecto inter-alumno de auto-eficacia representa las *diferencias* entre las puntuaciones de matemáticas obtenidas por alumnos que presentan un *promedio* de auto-eficacia un punto más alto. Este efecto a nivel de alumno explicará la varianza de intercepto aleatorio siempre que se modele por sí misma, aunque también podrá explicar la varianza de tiempo lineal aleatorio si esta está incluida en una interacción multi-nivel con tiempo lineal.

— De manera parecida a las interacciones multi-nivel de tiempo lineal-por-FRL, añadió Sydney.

— ¡Exactamente! Por otro lado, el efecto intra-alumno representa el cambio en puntuaciones de matemáticas en una ocasión cuando la auto-eficacia de un alumno es un punto superior *a su nivel habitual*. Este efecto de nivel de ocasión explica la varianza residual. — ¿Lo entiendes?

Sydney terminó de tomar notas. — Creo que sí.

— No te preocupes, si lo encuentras un poco difícil, Sydney. Te aseguro que a mí me costó cuando estudié la persona-media-centrado. Te recomiendo que, para obtener más información sobre la persona-media-centrado, leas la sección dos del capítulo nueve de Hoffman (2015), añadió Dominik con calma.

— Gracias. Así, entonces, ¿qué sucede si encuentro unos efectos de tiempo aleatorios significativos cuando modele la auto-eficacia como resultado?

— En ese caso, el uso de persona-media-centrado será inadecuado porque el HLM que hemos estado estimando no dispone de ningún mecanismo que explique las diferencias inter-alumno observadas respecto a cómo ha cambiado la auto-eficacia a lo largo del tiempo. Es decir, si ignorases los efectos aleatorios de tiempo relativos a la auto-eficacia y la auto-eficacia de persona-media-centrado, entonces los efectos fijos estimados inter-alumno e intra-alumno relativos a la auto-eficacia serían incorrectos.

— ¿Efectos fijos incorrectos? ¡Vaya, pues vamos bien!, exclamó Sydney en un tono ligeramente sarcástico. Así, si la auto-eficacia cambia de manera diferente entre alumnos, ¿qué opciones tengo?

— En una situación así, necesitarías estimar un HLM *multi-variante* en el que las variables de matemáticas y auto-eficacia estuvieran modeladas como resultados en un único modelo. Como cabe esperar, un HLM multi-variante es más complicado que el HLM univariado del que hemos estado hablando, tanto en términos de estructurar tus datos para el análisis como de interpretar los resultados.

— Así, ¿el HLM del que hemos estado hablando durante la mayor parte de esta reunión está considerado como *univariante* porque solo habíamos considerado un único resultado, las matemáticas?, preguntó Sydney.

— ¡Exactamente! Por otro lado, el HLM multi-variante te permitiría incluir en el mismo modelo todos los efectos aleatorios necesarios para los resultados de *tanto* las matemáticas *como* la auto-eficacia, con lo que podrás evaluar cómo los resultados se relacionan entre sí a cada nivel: las relaciones inter-alumno entre resultados a través de las *correlaciones de efectos aleatorios* estimados y las relaciones intra-alumno entre resultados a través de las *correlaciones de residuales* estimadas.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

— ¿Cómo se interpretarían estas correlaciones?, preguntó Sydney educadamente.

— Una correlación positiva de intercepto aleatorio inter-alumno indica que si un alumno presenta un promedio de puntuaciones de matemáticas tiempo-0 superiores *en comparación con otros alumnos*, es probable que presente un promedio de auto-eficacia tiempo-0 superior comparado con otros alumnos (o una auto-eficacia inferior si la correlación fuera negativa). Una correlación de pendiente aleatoria se interpreta de manera parecida, pero en relación con la pendiente específica del alumno cuando se compara con otros alumnos. Y una correlación residual inter-alumno positiva indica que si un alumno presenta una puntuación de matemáticas superior *a la predicha en una ocasión dada*, se predecirá que su auto-eficacia es superior a la predicha en dicha ocasión (o inferior a la predicha si la correlación es negativa). Dominik dio un vistazo rápido a su reloj. Sé que el HLM multi-variante requiere tiempo, pero si tus datos necesitan un HLM multi-variante, consulta la sección dos del capítulo nueve de Hoffman (2015).

— Gracias, dijo Sydney mientras terminaba de tomar notas. Sydney también miró su reloj y exclamó: — Te he robado mucho tiempo!

— No hay ningún problema, respondió Dominik sonriendo. Me encanta trabajar con personas a las que les gusta aprender. Cuando empieces a modelar los datos de tu distrito, seguro que te surgirán más preguntas, así que si quieres que nos volvamos a reunir, solo tienes que decírmelo. Dominik guardó el archivo de la sintaxis de SPSS y le entregó a Sydney su USB.

— Hoy mismo, una vez haya escrito la sintaxis de SAS, te enviaré un correo electrónico.

Sydney colocó su bloc de notas y su bolígrafo en la bolsa y se levantó para marcharse.

— Gracias, Dominik. Te estoy muy agradecida por el tiempo que me has dedicado. Espero que pases un buen día.

Dominik asintió con la cabeza y sonrió. — Tú también.

Notas

1. Corresponde a tercero de Primaria y segundo de ESO en el sistema educativo español.

Disclosure statement

No potential conflict of interest was reported by the authors / *Los autores no han referidoningún potencial conflicto de interés en relación con este artículo*.

References / *Referencias*

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. New York, NY: Routledge.

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Appendix A / *Apéndice A*

SPSS SYNTAX

```
* Encoding: UTF-8.

*Import that data.
PRESERVE.
SET DECIMAL DOT.
GET DATA /TYPE = TXT
  /FILE = 'I:\Invited Manuscripts\Culture & Education\Data.csv'
  /ENCODING = 'UTF8′
  /DELCASE = LINE
  /DELIMITERS = ','
  /ARRANGEMENT = DELIMITED
  /FIRSTCASE = 2
  /DATATYPEMIN PERCENTAGE = 95.0
  /VARIABLES =
  StudentID AUTO
  OccasionID AUTO
  FRL AUTO
  SelfEff AUTO
  Math AUTO
  /MAP.
RESTORE.
CACHE.
EXECUTE.
DATASET NAME DataSet1 WINDOW = FRONT.

TITLE 'Unconditional Random Intercept Model for Math ICC'.
MIXED Math
    /METHOD = REML
    /PRINT = SOLUTION TESTCOV
    /RANDOM = INTERCEPT |SUBJECT(StudentID).

TITLE 'Mean Math Scores across Observations'.
GGRAPH
    /GRAPHDATASET NAME = 'graphdataset' VARIABLES = OccasionID MEAN
    (Math)[name = 'MEAN_Math'] MISSING = LISTWISE
  REPORTMISSING = NO
 /GRAPHSPEC SOURCE = INLINE.
BEGIN GPL
  SOURCE: s = userSource(id('graphdataset'))
  DATA: OccasionID = col(source(s), name('OccasionID'), unit.category())
  DATA: MEAN_Math = col(source(s), name('MEAN_Math'))
  GUIDE: axis(dim(1), label('OccasionID'))
  GUIDE: axis(dim(2), label('Mean Math'))
  SCALE: linear(dim(2), min(110), max(130))
  ELEMENT: line(position(OccasionID*MEAN_Math), missing.wings())
END GPL.

*Centre Time at the First Occasion.
COMPUTE Time0 = OccasionID — 1.
FORMATS Time0(f1.0).
EXECUTE.

TITLE 'Fixed Linear Time, Random Intercept Model'.
```

Downloaded by [66.45.159.208] at 07:54 27 November 2017

```
MIXED Math WITH Time0
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV
  /FIXED = Time0
  /RANDOM = INTERCEPT |SUBJECT(StudentID).
TITLE 'Random Linear Time Model'.
MIXED Math WITH Time0
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV
  /FIXED = Time0
  /RANDOM = INTERCEPT Time0 |SUBJECT(StudentID) COVTYPE(UN).

TITLE 'Fixed Quadratic, Random Linear Time Model'.
MIXED Math WITH Time0
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV
  /FIXED = Time0 Time0*Time0
  /RANDOM = INTERCEPT Time0 |SUBJECT(StudentID) COVTYPE(UN).

TITLE 'Random Quadratic Time Model'.
MIXED Math WITH Time0
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV
  /FIXED = Time0 Time0*Time0
  /RANDOM = INTERCEPT Time0 Time0*Time0 |SUBJECT(StudentID) COVTYPE
(UN).

TITLE 'Random Linear Time: FRL*Linear Time'.
MIXED Math WITH Time0 FRL
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV
  /FIXED = Time0 FRL Time0*FRL
  /RANDOM = INTERCEPT Time0 |SUBJECT(StudentID) COVTYPE(UN).
```

Downloaded by [66.45.159.208] at 07:54 27 November 2017

Appendix B /*Apéndice B*

SAS SYNTAX

```
*Import that data;
DATA DataSet1;
        INFILE 'I:\Invited Manuscripts\Culture & Education\Data.csv'
                DELIMITER = ','
                MISSOVER
                DSD
                LRECL = 32,767
                FIRSTOBS = 2;
        INFORMAT StudentID OccasionID FRL SelfEff Math best32.;
        FORMAT StudentID OccasionID FRL SelfEff Math best12.;
        INPUT StudentID OccasionID FRL SelfEff Math;
RUN;

TITLE 'Unconditional Random Intercept Model for Math ICC';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;

        CLASS StudentID;
```

```
        MODEL Math = /SOLUTION DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT /SUBJECT = StudentID;
RUN; TITLE;

*Get Means Plot;
PROC SORT DATA = DataSet1; By OccasionID; RUN;
PROC MEANS DATA = DataSet1 NOPRINT; VAR Math; BY OccasionID; OUTPUT
OUT = Means; RUN;
DATA Means; SET Means; IF _STAT_ NE 'MEAN' THEN DELETE; RUN;
PROC GPLOT DATA = Means;
        SYMBOL1 INTERPOL = JOIN;
        AXIS1 LABEL = ('OccasionID') OFFSET = (10,10);
        AXIS2 LABEL = (ANGLE = 90 'Mean Math') ORDER = (110 TO 130 BY 5);
        PLOT Math*OccasionID /HAXIS = AXIS1 VAXIS = AXIS2;
RUN; TITLE;

*Centre Time at the First Occasion;
DATA DataSet1; SET DataSet1;
Time0 = OccasionID — 1;

RUN;

TITLE 'Fixed Linear Time, Random Intercept Model';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;
        CLASS StudentID;
        MODEL Math = Time0 /SOLUTION DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT /SUBJECT = StudentID;
RUN; TITLE;

TITLE 'Random Linear Time Model';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;
        CLASS StudentID;
        MODEL Math = Time0 /SOLUTION DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT Time0 /SUBJECT = StudentID;
RUN; TITLE;

TITLE 'Fixed Quadratic, Random Linear Time Model';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;
        CLASS StudentID;
        MODEL Math = Time0 Time0*Time0 /SOLUTION DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT Time0 /SUBJECT = StudentID;
RUN; TITLE;

TITLE 'Random Quadratic Time Model';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;
        CLASS StudentID;
        MODEL Math = Time0 Time0*Time0 /SOLUTION DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT Time0 Time0*Time0 /SUBJECT = StudentID;
RUN; TITLE;

TITLE 'Random Linear Time Model: Add FRL*Linear Time';
PROC MIXED DATA = DataSet1 METHOD = REML NOCLPRINT COVTEST;
        CLASS StudentID;
        MODEL Math = Time0 FRL Time0*FRL /SOLUTION
        DDFM = SATTERTHWAITE;
        RANDOM INTERCEPT Time0 /SUBJECT = StudentID;
RUN; TITLE;
```

Downloaded by [66.45.159.208] at 07:54 27 November 2017