LUIS FELIPE VILLOTA

LUND UNIVERSITY
QDA USING R - SIMM61

LAB ASSIGNMENT 2 - MIXED LINEAR MODEL: SURGERY DATA

Introduction

In the present report, we used linear mixed models in order to estimate the postoperative pain after wisdom tooth surgery with six fixed effect predictors, in the context of clustering groups of hospital of treatment (a random effect). Since no prior studies guided this investigation on the manifestation of the random effect we build two different models in Dataset A: one allowing the random intercept model and another combining the random slope and random intercept model. This, as mentioned, across the different hospitals. Based on the AIC criteria, using the *anova()* function we continued working with the random intercept model as the best fit for the data; - and later, without retraining it, we predicted "pain" in Dataset B. Finally, the most influential predictor is highlighted and visually analyzed across the clustering groups of hospitals.

Methods

Model description. First, we checked Dataset A and B in order to clean, transform and recode as necessary. No missing cases were found. Both have 200 observations and the same variables. We changed character variables into factors in Dataset A and B in variables "*ID*", "*sex*", and "*hospital*". We identified in Dataset A a level called "*woman*" in the "*Sex*" variable and we collapse it into "*female*". After this, in Dataset A we had 97 females and 103 males for a total of 200 patients. No further data cleaning was required in Dataset B.

After an exploratory data analysis and visualization of Dataset A, we saw that data points were clustered on the scatterplots (visually grouped by colors). This strongly suggested that group membership (in this case hospital of treatment) can explain variability in the outcome variable "*pain*", and therefore some observations might not be independent from each other. Therefore, in our model, "*hospital*" is the random effect predictor as it seemed to be a key element to take into account in order to make accurate predictions on postoperative pain on this dataset and outside it. The fixed effects were (6): "*age*", "*sex*", "*STAI_trait*", "*pain_cat*", "*mindfulness*" and "*cortisol_serum*". We used the *lmer ()* function to build one random slope model and another allowing only the random intercept.

Results

Goodness of fit and model coefficients. Although the random slope model had the smallest sum of residuals (217.34 against 224.31 of the random intercept model), the random intercept model proved to be a better fit according to the ANOVA (Chi^2 = 1.9647, df = 2, p = 0.3744, AIC for the random intercept model of 629.12 and for the random slope model 631.15). The cAIC showed 621.42 for the random intercept model and 623.73 for the random slope model). We report the complete model estimates in *Tables i and ii*.

Regarding the **explained variance** in Dataset B, applying our random intercept model from Dataset A, we have an *R^2 of* 0.3799 (37.99%) in Dataset B and *Marginal R^2* of 0.3852 (38.52%) and *Conditional R^2* of 0.4632 (46.32%) in Dataset A.

The most influential predictor was found to be *cortisol_serum* in the model with an R^2 of 0.123 and an upper.CL of 0.215 and a lower.CL of 0.052 at p<0.001. We build a **new model** for Dataset A now with only this predictor based on the random and intercept model using *lmer(pain ~ cortisol_serum + (cortisol_serum | hospital), control = lmerControl(optimizer = "Nelder_Mead"), data = Dataset_A).* We visualize its predictions for "*pain*" on different scatterplots for each hospital. Finally, we compare this new model it against the random intercept model, as seen in *Figure 1*. In the latter, we have the same formula for the random slope model here but we change to *(1 | hospital)* to restrict the slopes of *cortisol_serum* in each hospital.

Discussion

The R^2 obtained in Dataset B was 0.3799 (or 37.99% of explained variance in pain) and was closer to the *Marginal R^2* (38.52%) in Dataset A, rather than to the *Conditional R^2* in Dataset A (46.32%). This is due to the fact the we were applying the model that takes into account the fixed effects and not the random effects. *Marginal R^2* explains for fixed effects only. *Conditional R^2* explains the variance of the combined model (random slope and random intercept).

Based on *Figure 1* of the new model in which "*cortisol_serum*" is the only predictor based on the combined model (random slope and random intercept), the difference between the random intercept model and the random slope (which includes the random intercept) is hardly noticeable. Nevertheless, we compared model fit indices for both and we found (Chi^2 = 2.15, df = 2, p < 0.3411, AIC of the random slope model = 674.35, AIC of random intercept model = 672.50). So the random slope model has an AIC of less than 2 smaller than the one of the

random intercept model, and it is not at a significant level. Nevertheless, we also found that for the cAIC we do have a 661.37 for the random slope model and a 664.5371 for the random intercept model, making the random slope model slightly better fit. Yet, for our data, it was not fruitful to expect different slopes of "*cortisol_serum*" for each hospital, because we assumed that hospitals had effect on "*pain*" but no on "*cortisol_serum*". In conclusion, the random intercept model is a better fit for this data at explaining variation of pain in Dataset A.

R code: https://github.com/FelipeVillota/SIMM61_QDA-with-R/blob/main/wisdom.R

Appendix

Table i. Regression coefficients: random intercept model in Dataset A

	rand_intercept_model					
Predictors	*Estimates*	*std. Beta*	*CI*	*standardized CI*	*Statistic*	*p*
(Intercept)	3.85 **	-0.08	1.10 – 6.59	-0.31 – 0.15	2.77	**0.006**
age	-0.06 **	-0.19	-0.10 – -0.02	-0.31 – -0.07	-3.22	**0.002**
sex [male]	0.23	0.15	-0.09 – 0.55	-0.06 – 0.37	1.42	0.158
STAI trait	-0.02	-0.08	-0.06 – 0.02	-0.21 – 0.06	-1.14	0.256
pain cat	0.08 ***	0.26	0.04 – 0.13	0.12 – 0.41	3.61	**<0.001**
mindfulness	-0.23 *	-0.14	-0.43 – -0.02	-0.26 – -0.01	-2.16	**0.032**
cortisol serum	0.51 ***	0.34	0.33 – 0.69	0.22 – 0.46	5.58	**<0.001**
Random Effects						
σ^2	1.20					
$\tau_{00\ hospital}$	0.17					
ICC	0.13					
$N_{hospital}$	10					
Observations	200					
Marginal R^2 / Conditional R^2	0.385 / 0.463					

** p<0.05 ** p<0.01 *** p<0.001*

Table ii. Model comparison in Dataset A[1]

Model	RSS	AIC	cAIC	logLikelihood
Random intercept model	224.31	629.12	621.42	-305.56
Random slope model	217.34	631.15	623.73	-304.58

Table iii. Regression coefficients: model comparison in Dataset A

	rand_int				rand_slope_int			
Predictors	*Estimates*	*CI*	*Statistic*	*p*	*Estimates*	*CI*	*Statistic*	*p*
(Intercept)	3.85 **	1.10 – 6.59	2.77	**0.006**	3.46 *	0.72 – 6.20	2.49	**0.014**
age	-0.06 **	-0.10 – -0.02	-3.22	**0.002**	-0.06 **	-0.10 – -0.02	-3.02	**0.003**
sex [male]	0.23	-0.09 – 0.55	1.42	0.158	0.22	-0.10 – 0.53	1.34	0.181
STAI trait	-0.02	-0.06 – 0.02	-1.14	0.256	-0.02	-0.06 – 0.02	-1.08	0.281
pain cat	0.08 ***	0.04 – 0.13	3.61	**<0.001**	0.09 ***	0.04 – 0.13	3.86	**<0.001**
mindfulness	-0.23 *	-0.43 – -0.02	-2.16	**0.032**	-0.21 *	-0.42 – -0.01	-2.06	**0.041**
cortisol serum	0.51 ***	0.33 – 0.69	5.58	**<0.001**	0.51 ***	0.30 – 0.72	4.87	**<0.001**
Random Effects								
σ^2	1.20				1.18			
τ_{00}	0.17 $_{hospital}$				0.32 $_{hospital}$			
τ_{11}					0.03 $_{hospital.cortisol\_serum}$			
ρ_{01}					-0.88 $_{hospital}$			
ICC	0.13				0.15			
N	10 $_{hospital}$				10 $_{hospital}$			
Observations	200				200			
Marginal R^2 / Conditional R^2	0.385 / 0.463				0.385 / 0.479			
Deviance	611.151				609.244			
AIC	657.228				659.142			
AICc	658.176				660.546			
log-Likelihood	-319.614				-318.571			

** p<0.05 ** p<0.01 *** p<0.001*

[1] These results were obtained running the *anova()* function with both models (as shown on the R code). This RSS was obtained with the function *sum(residuals()^2).* We report that when producing *Table i* with the *tab_model()* command with the same models and include these indices we get different values but the same conclusion for having the random intercept model as the better fit for Dataset A: for the random intercept model (AIC: 657.22, cAIC: 658.17, logLikelihood: -319.61) and for the random slope model (AIC: 659.14, cAIC: 660.54, logLikelihood: -318.57), this is shown in *Table iii*.

Figure 1. Comparison of different fitted regression lines for the hospitals in Dataset A of the random slope model (black dots) and random intercept model (colored dots) with only the most influential predictor of both: cortisol_serum

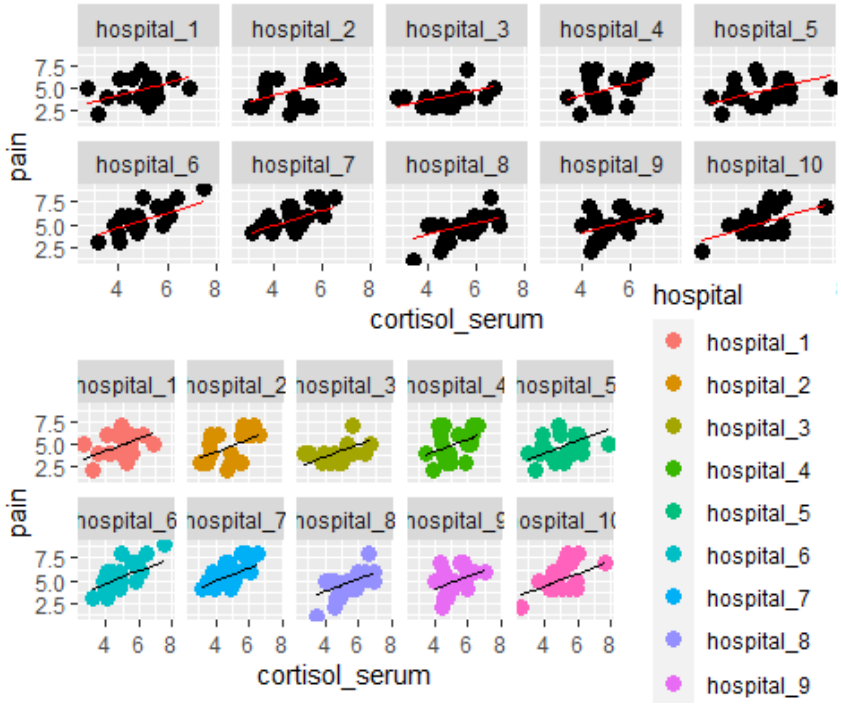

Figure 2. Graph of the random intercept model with only cortisol_serum as a predictor of pain in Dataset A

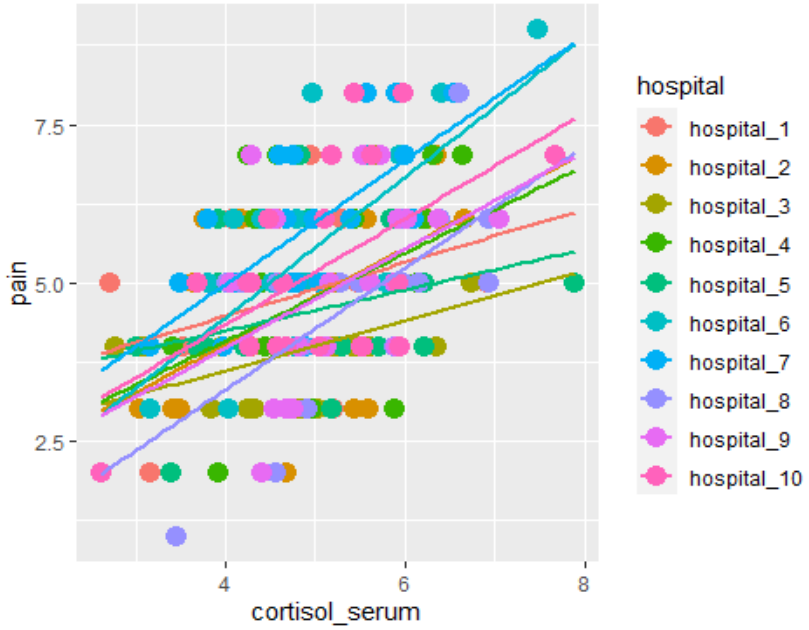

Figure 3. Graph of the random slope model with only cortisol_serum as a predictor of pain in Dataset A