SAKA003 Quantitative Methods: Multivariate Analysis
Alvaro von Borries
810207-8618

Lab Report 2 – Model Comparison

Introduction

This report contains the comparison of two models to predict the level of pain experienced around and after surgeries. The first model includes only demographic variables as predictors: age and gender. The second model adds psychological and hormonal predictors.

Data Exploration and Preparation

Table 1 shows the descriptive statistics for the variables involved in the analysis. It is possible to see that there seems to be coding errors in *pain* and *mindfulness*. After further exploration of the dataset we find one value over 10 for *pain* and one value over 6 for *mindfulness*. New variables were coded excluding those observations (pain_2 and mindfulness_2).

The sex variable is a string and females are coded both "female" and "woman". To use it in our models we recode it as a numerical dichotomic variable (1 for female and woman, 0 for male). Also pain level is converted from nominal to scale.

Further examination of frequency tables and histograms show that all scale variables are fairly normally distributed and that the sample contains 52.5% females. Table 2 presents the updated descriptive statistics considering all mentioned changes and corrections.

Data Analysis

The simple model considering only gender and age as predictors is significantly better than the null model as F=6.777 and p-value=0.002. However, the model explains only 8% of the variance in pain ($R^2 = 0.08$).

Figure 1 shows that there are no influential cases considering the threshold of 1, but there are quite a few considering the 4/N one (in this case 0.025). However, they do not seem to be affecting the assumptions. Figures 2 to 4 show that residuals are fairly normally distributed, there seems to be some heteroscedasticity (the Breush-Pagan test gives an F-statistic of 2.996 and p-value of 0.053, non-significant at 5% level), and the relation between pain and age is fairly linear. Finally, the VIF values (1.001 for both variables) indicate that there is not multicollinearity. All in all, the assumptions are considered met.

This model is now compared with a more complex one that also considers psychological and hormonal predictors: *anxiety* through the STAI, pain catastrophizing, the tendency to turn attention to present-moment (Mindfulness), and the cortisol level (from serum). The models are compared with a hierarchical regression, and the statistics are presented in Table 5. The complex model is significantly better than the null model (F=12.854). It is also significantly better than the simple one as the AIC has dropped from 122 to 78 and R^2 has increase to 0.338 (this model is able to explain almost 34% of the variance in pain).

Assumptions are also generally met in this model. As seen in Figures 5 to 10 residuals are normally distributed, heteroscedasticity is less a problem than before (Breush-Pagan test gives an F-statistic of 1.178 and p-value of 0.321), and the relationships between pain and the new predictors are fairly linear. Also, there is no problematic multicollinearity as the highest VIF reported is 1.873 for STAI (lower than the threshold of 3). There are influential cases at the 4/N threshold, but they do not seem to affect the main assumptions, therefore they are not treated nor removed from the analysis (Figure 11).

In the simple model both variables are negatively related to pain (Table 3). Every extra year is associated with 0.08 points decrease in pain, while women experience on average 0.13 points of pain less than men. Only the constant and the coefficient for age are statistically significant. Since sex is a predictor dependent on the type of procedure, we can conclude that it does not influence the pain experienced after this particular surgery.

In the complex model (Table 4) every extra year is associated with 0.03 points decrease in pain, women experience on average 0.32 points of pain less than men, every extra point in the anxiety measure is associated with 0.01 points less in pain, and every extra point in the pain catastrophizing scale is related to 0.08 points more in pain, one extra point of cortisol (serum) is associated with 0.54 more points of pain, and each extra point in the MAAS score is related to 0.14 more points in pain. Only pain catastrophizing and cortisol level are statistically significant at the 5% level. Moreover, looking at the standardized B, it is clear that the variables with the largest effect on pain are cortisol level and pain catastrophizing followed by mindfulness and gender. Finally, the regression equation for the complex models is as follows:

*Pain level = 2.26 - 0.03*Age - 0.32*Gender - 0.01*STAI +* ***0.08*Pain_cat*** *+*
0.54*Cortisol *- 0.14*Mindfulness*

All in all, the more complex model significantly improves our predictive capacity. However, an exploration of two extra models, a first one only with pain catastrophizing and cortisol as predictors (simple_2), and a second one adding age (complex _2), reveals that although there seems to be a small gain by adding age, AIC drops less than a point. However, the model complex_2 is significantly better than the complex model (AIC drops by more than 2 points), while model simple_2 is not. Therefore, the selected model should be model complex_2 (no diagnostics were run). Table 6 presents the results which translate in the following equation:

Pain level = 0.94 ***+ 0.10*Pain_cat*** *+* ***0.5*Cortisol*** *– 0.03*Age*

Appendix

Table 1. Initial Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pain level	160	1.00	50.00	5.19	3.875
Age	160	24.00	53.00	40.74	5.044
STAI	160	26.00	52.00	40.09	5.150
Pain catastrophizing	160	14.00	42.00	29.88	4.778
Cortisol (serum)	160	2.70	7.11	4.94	0.992
Cortisol (saliva)	160	2.53	7.19	4.95	1.010
Mindfulness	160	1.00	6.22	3.20	0.977
Valid N (listwise)	160				

Table 2. Final Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Pain level	159	1.00	8.00	4.91	1.524
Gender	160	0.00	1.00	0.53	0.501
Age	160	24.00	53.00	40.74	5.044
STAI	160	26.00	52.00	40.09	5.150
Pain catastrophizing	160	14.00	42.00	29.88	4.778
Cortisol (serum)	160	2.70	7.11	4.94	0.992
Cortisol (saliva)	160	2.53	7.19	4.95	1.010
Mindfulness	159	1.00	5.63	3.19	0.950
Valid N (listwise)	158				

Table 3. Simple Model Results

	Unstandarized Coefficients				Standarized B	t	p-value
	B	Std. Error	95% CI lb for B	95% CI ub for B			
Constant	8.44	0.96	6.54	10.33		9.79	< .001
Age	-0.08	0.02	-0.13	-0.04	-0.28	-3.65	< .001
Gender	-0.13	0.23	-0.59	0.33	-0.04	-0.57	0.573

Table 4. Complex Model Results

	Unstandarized Coefficients				Stand. B	t	p-value
	B	Std. Error	95% CI lb for B	95% CI ub for B			
Constant	2.26	1.70	-1.10	5.62		1.33	0.186
Age	-0.03	0.02	-0.08	0.01	-0.11	-1.44	0.151
Gender	-0.32	0.21	-0.74	0.10	-0.11	-1.49	0.140
STAI	-0.01	0.03	-0.07	0.04	-0.04	-0.49	0.626
Pain cat.	0.08	0.03	0.03	0.14	0.26	2.93	0.004
Cortisol (serum)	0.54	0.13	0.30	0.79	0.36	4.35	< .001
Mindfulness	-0.14	0.13	-0.39	0.11	-0.09	-1.08	0.284

Table 5. Models Comparison

Model	R-Square	Adjusted R-Square	F	df reg.	df res.	p-value	AIC
Simple	0.08	0.07	6.78	2	155	0.002	122.09
Complex	0.34	0.31	12.85	6	151	< .001	78.15
Simple_2	0.33	0.32	37.55	2	156	< .001	76.41
Complex_2	0.34	0.32	26.20	3	155	< .001	75.68

Table 6. Complex_2 Model Results

	Unstandarized Coefficients				Stand. B	t	p-value
		Std. Error	95% CI lb for B	95% CI ub for B			
Constant	0.94	1.29	-1.61	3.49		0.73	0.469
Pain cat.	-0.10	0.02	0.51	0.14	0.31	4.16	< .001
Cortisol (serum)	-0.50	0.11	0.28	0.72	0.33	4.47	< .001
Age	-0.03	0.02	-0.08	0.01	-0.11	-1.64	0.103

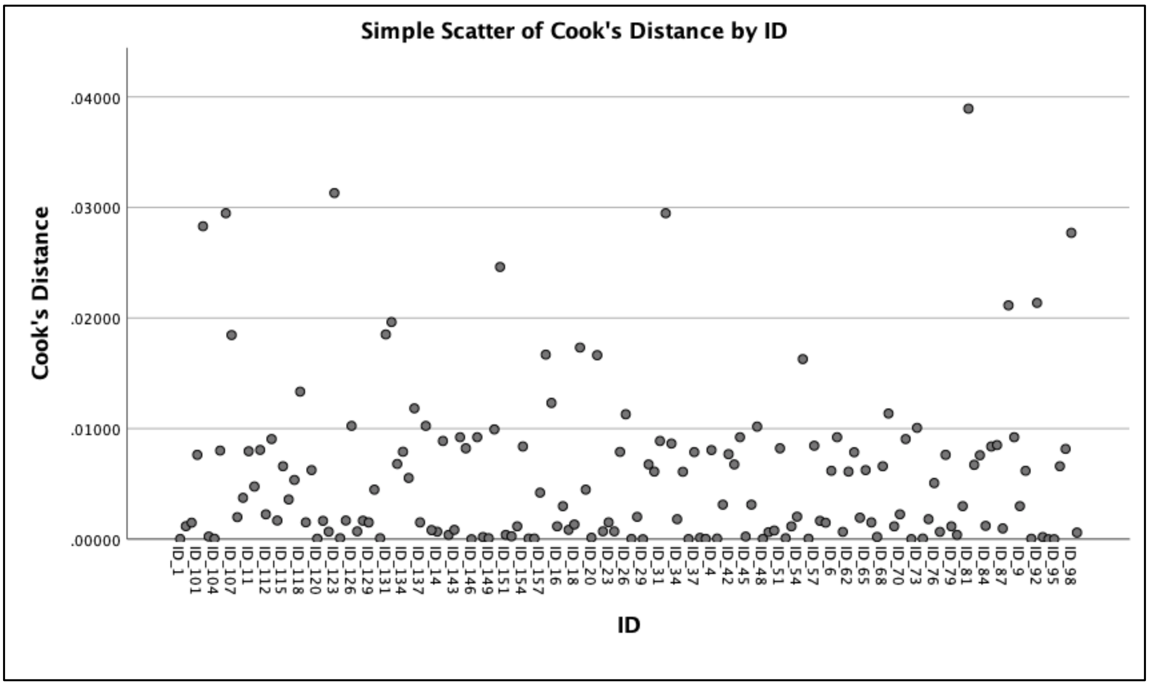

Figure 1. Cook’s Distances Scatterplot – Simple Model

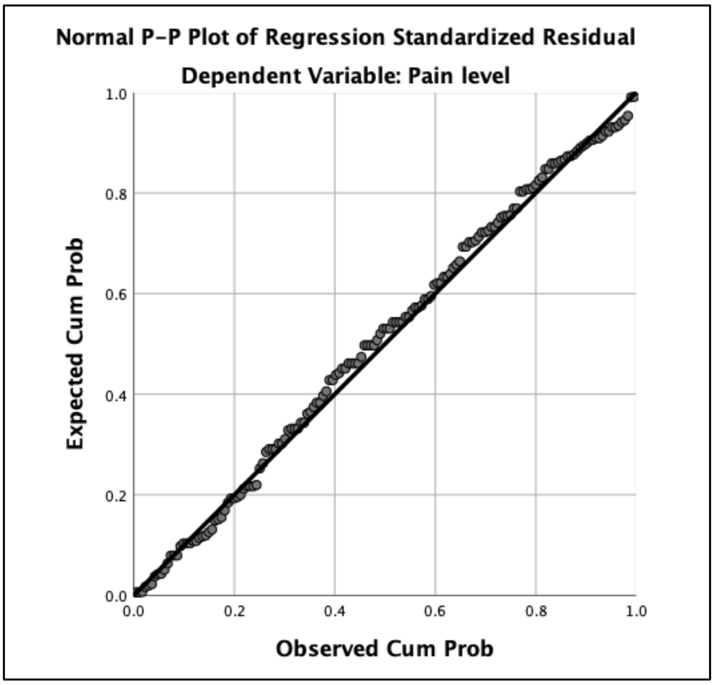

Figure 2. Normality of Residuals – Simple Model

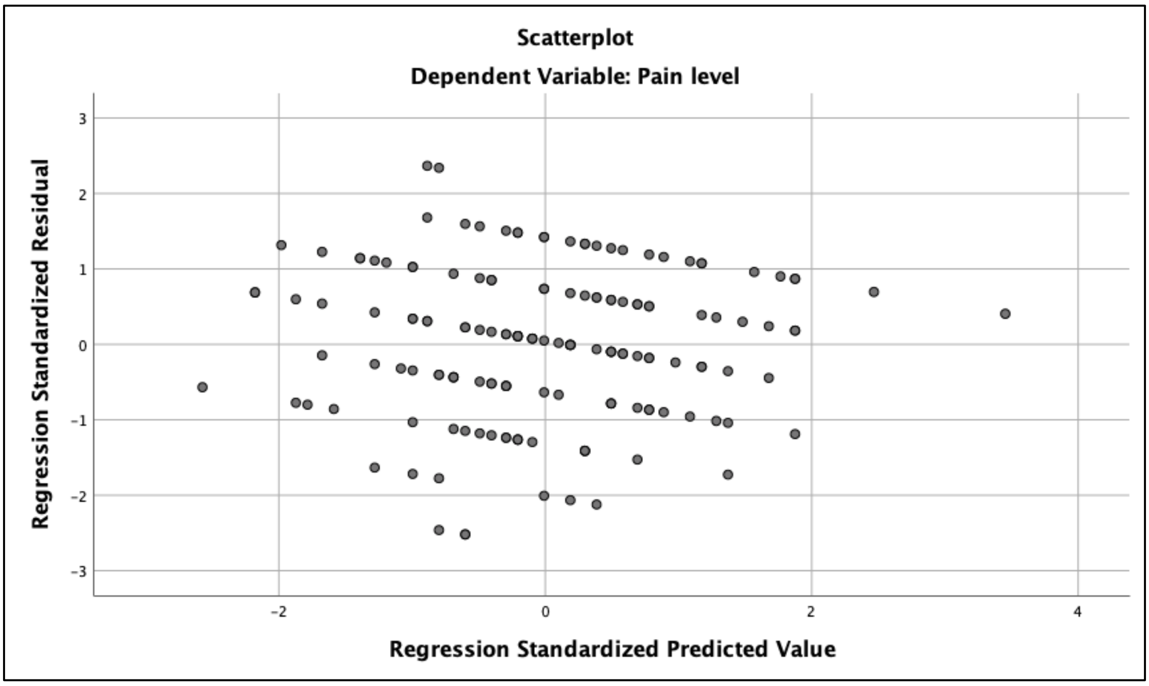

Figure 3. Homogeneity of Variance – Simple Model

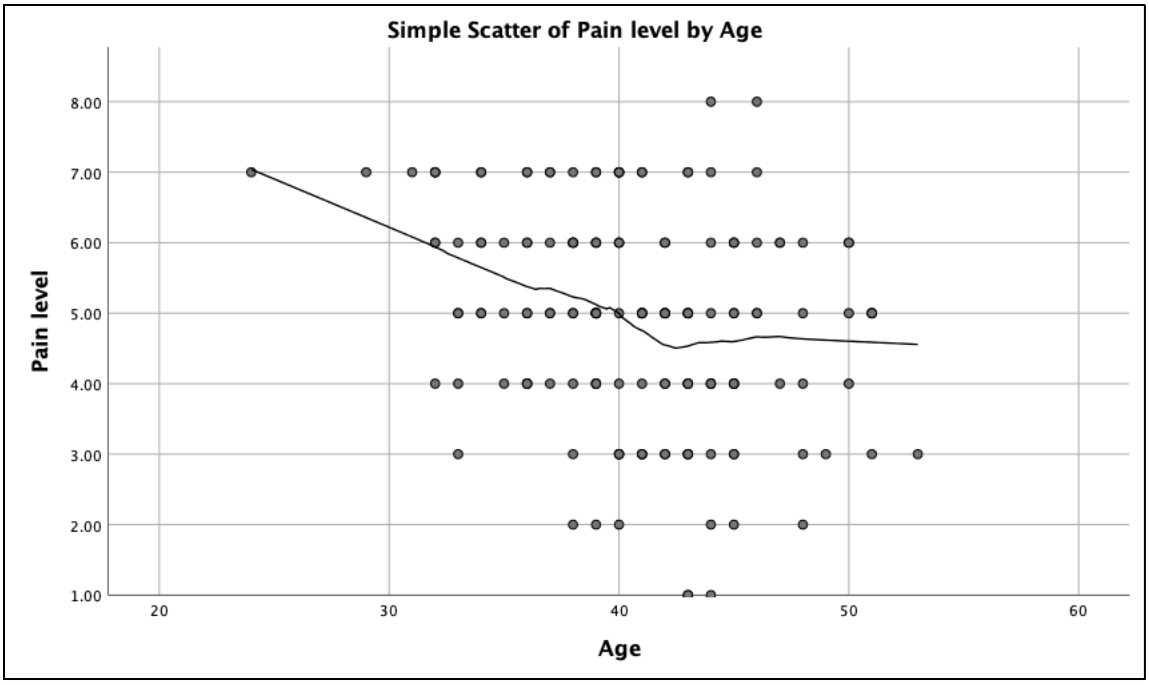

Figure 4. Scatterplot Pain level vs Age

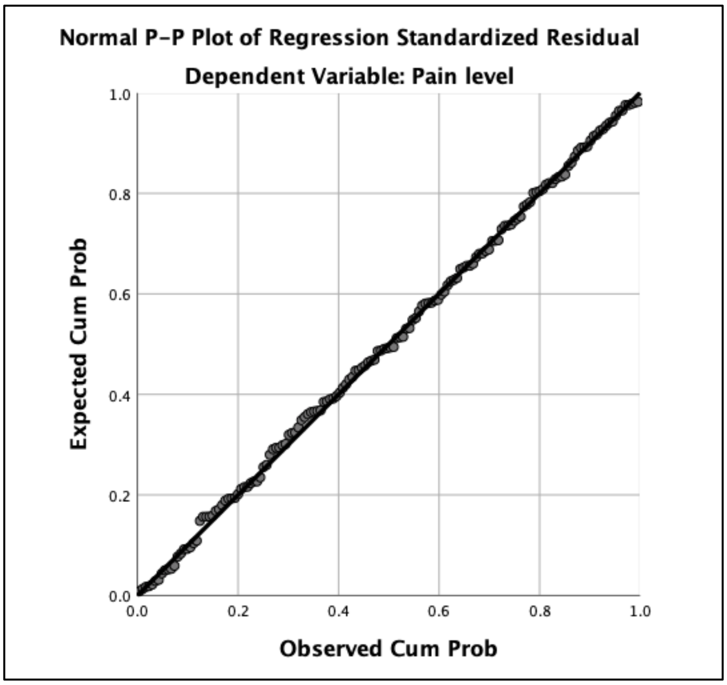

Figure 5. Normality of Residuals - Complex Model

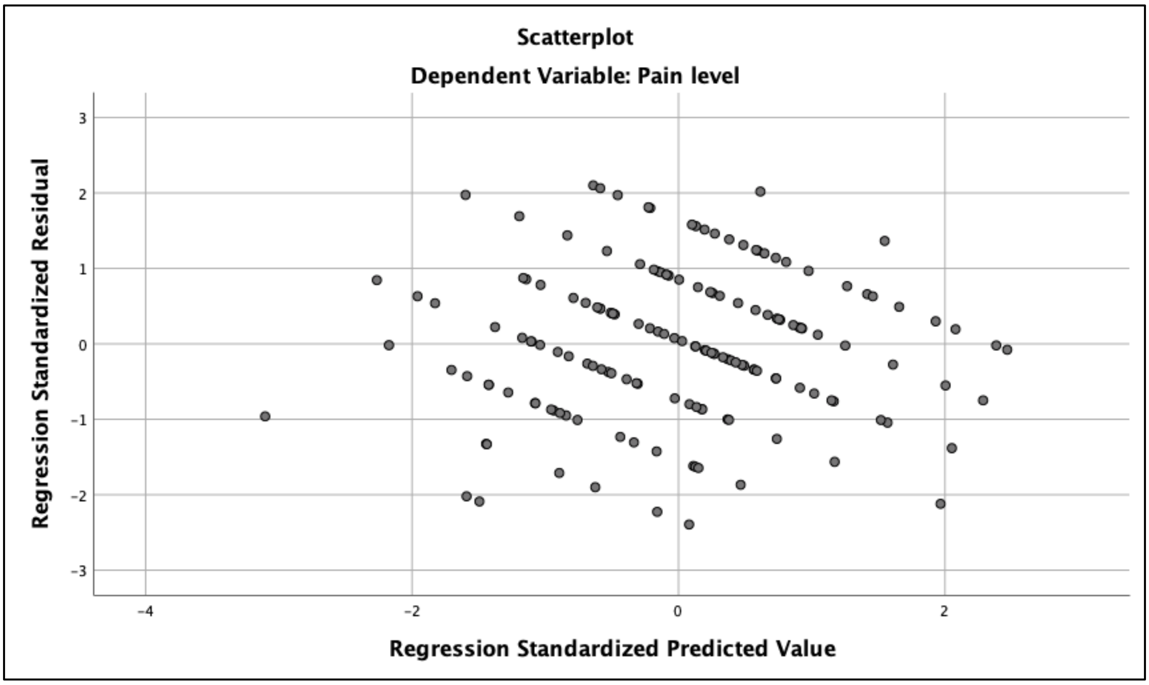

Figure 6. Homogeneity of Variance – Complex Model

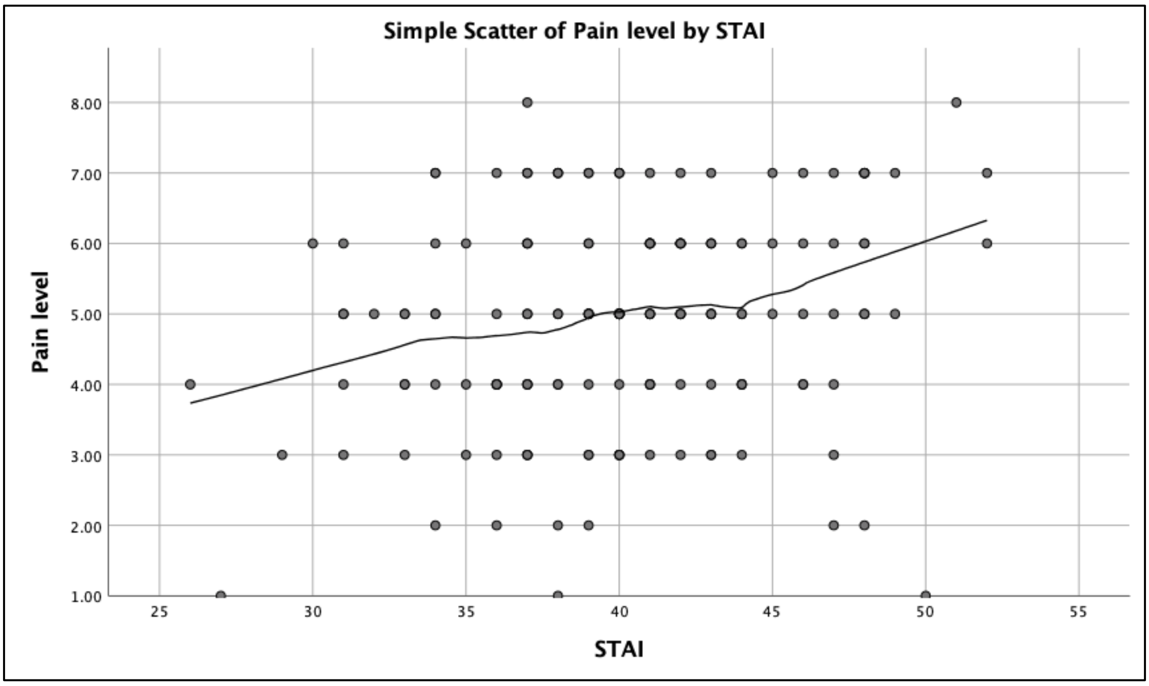

Figure 7. Scatterplot Pain level vs STAI

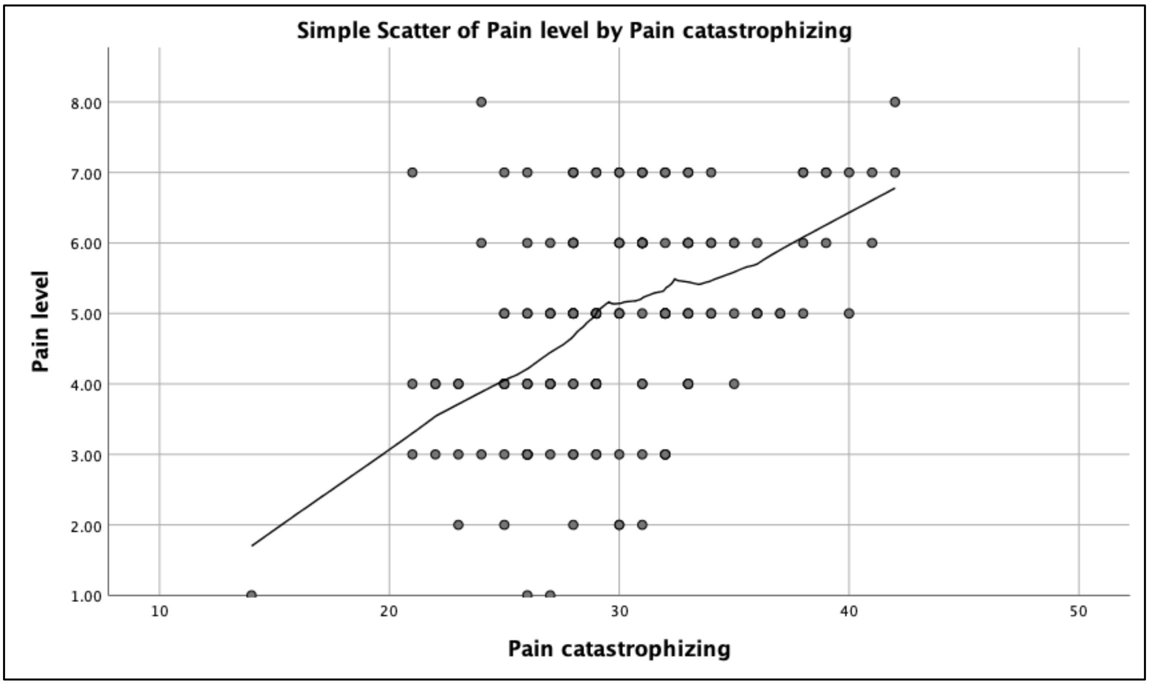

Figure 8. Scatterplot Pain level vs Pain Catastrophizing

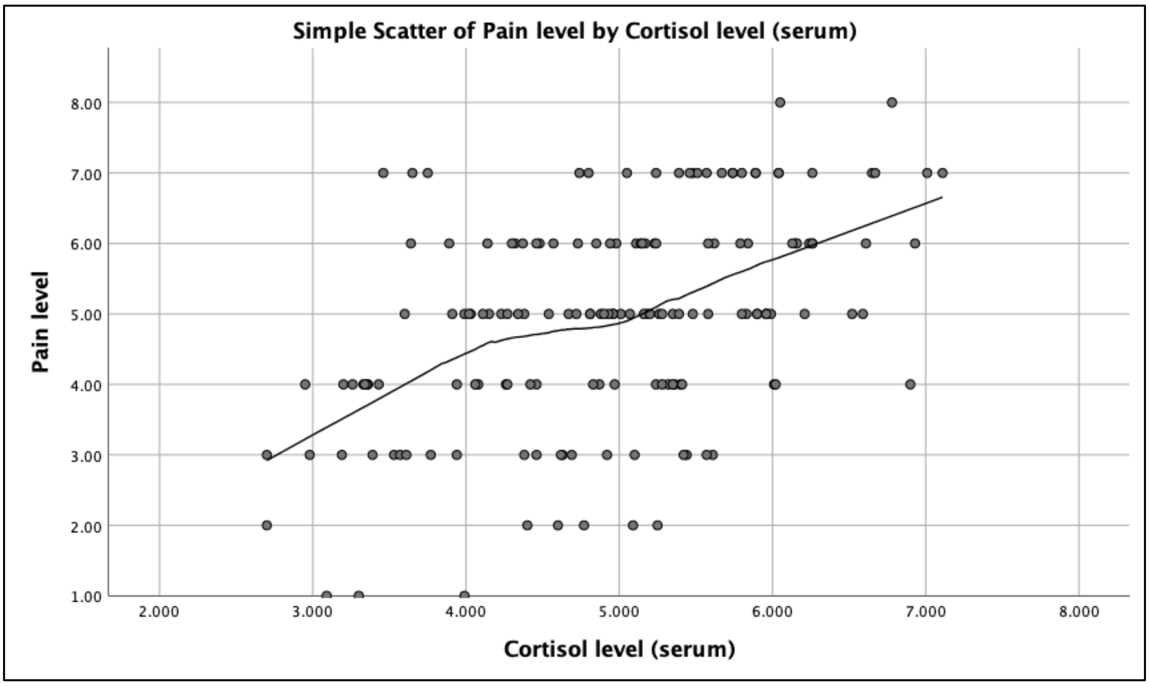

Figure 9. Scatterplot Pain level vs Cortisol (serum)

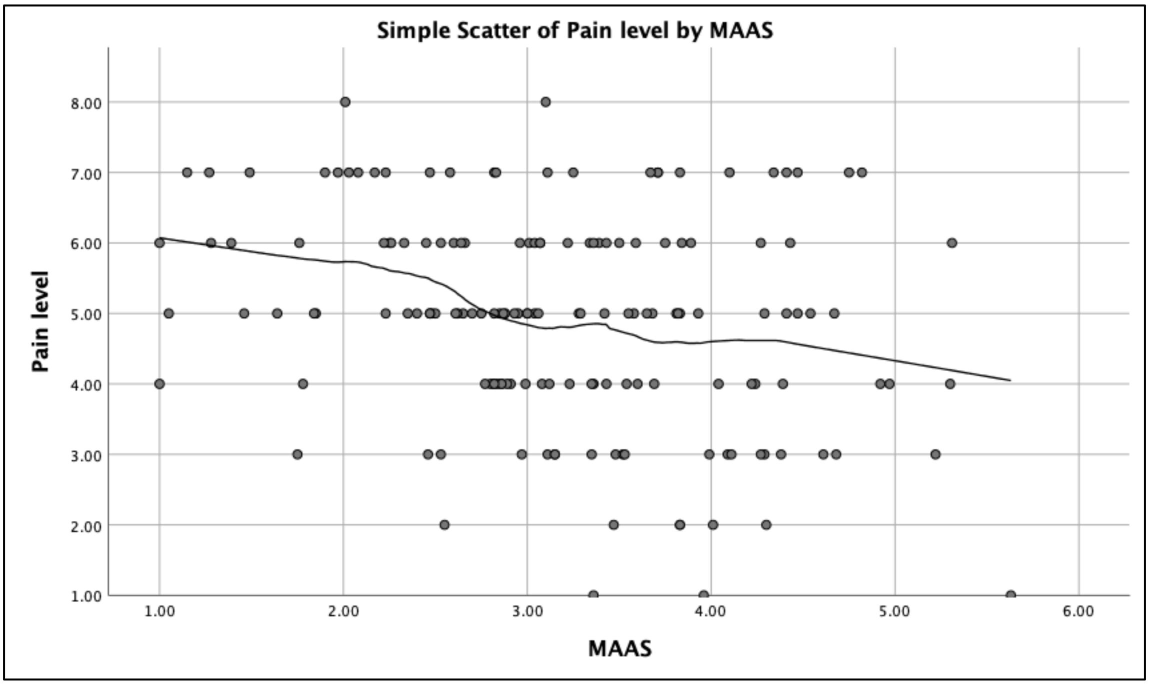

Figure 10. Scatterplot Pain level vs Mindfulness

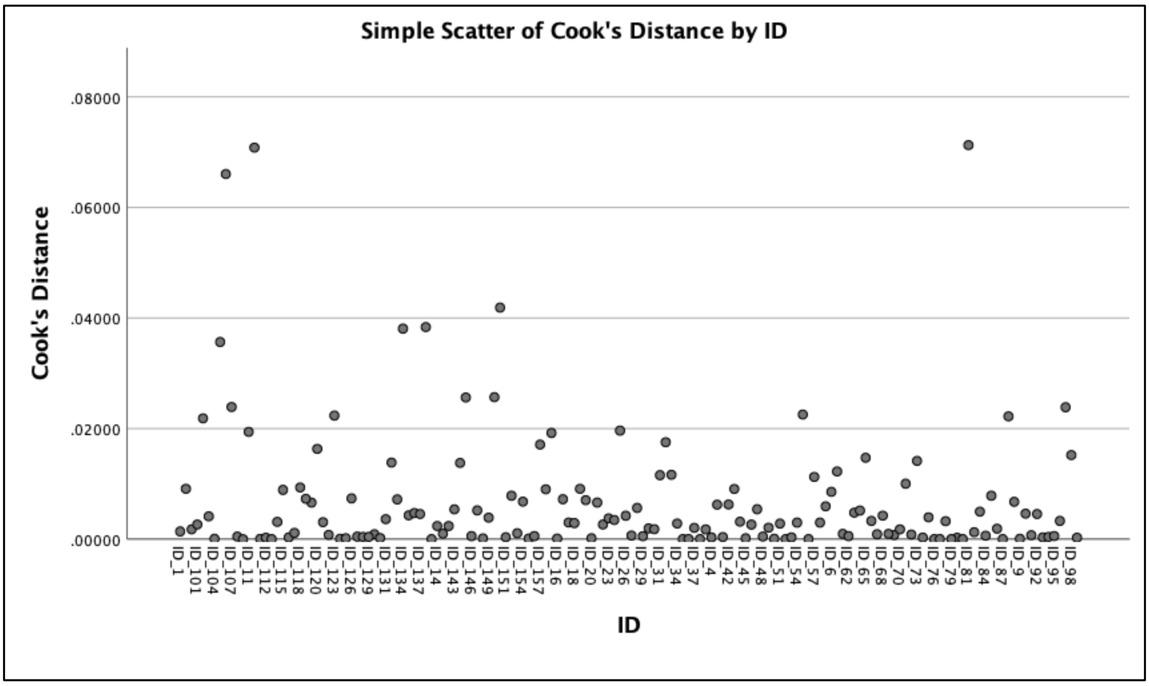

Figure 11. Cook’s Distances Scatterplot – Complex Model

Links
Syntax: https://github.com/avonborries/SAKA003-VT20/tree/master/Lab%202