**Data Analysis Report**

**Lund University**

**Master program in Psychology, autumn semester 2021 - re-exam II.**

**Nicole Evors**

**PSY14-HT21, 2021, autumn semester**

Logo

Description automatically generated

**Regression models with fixed and random effects**

**Assignment 1, Part 1**

**Introduction**

In this assignment, I worked with data related to perioperative pain and its psychological and hormonal predictors after wisdom tooth surgery. It was important to predict the amount of pain an individual would experience and make further comparisons of varying models in order to improve surgical pain management regimes. A study to assess the predictors and investigation of acquired results was carried out for a better understanding of perioperative pain.

After running the first model I checked if the assumptions of linearity normality and the homoscedasticity of the residuals are satisfied as shown in figures 1 and 2. Results of the QQ plot and Shapiro–Wilk test applied to the two models indicate that the residuals of models 1 and 2 are normally distributed as shown in figures 1 and 2 (B), the p-value of Shapiro Wilk test of the two models is p-value < 2.2e-16 and < 2.2e-16 respectively which confirmed that Non-normality of residuals was detected. Moreover, a p-values of < 2.22e-16 of ncvTest and figures 1 et 2 (C) shows that Heteroscedasticity (non-constant error variance) is detected in both models respectively. Cook's distance showed outliers in row 88. In contrast, linearity is more or less satisfied in the two models figure 1 et 2 (A) where The red line is fairly straight showing linearity. I checked possible multicollinearity with the functions *check\_collinearity* and *vif* of the packages ‘performance’ and ‘car’ respectively. I conclude that there is no multicollinearity in models 1 and two as all values have a VIF of less than 5.

**Dealing with non-normality and heteroscedasticity**

In order to deal with the non-normality and heteroscedasticity of residuals, I filter the data and I exclude row 88 from the data frame, then I build two new models (with the filtred data) and I test those models to examine normality and homoscedasticity of residuals. P-values of Shapiro Wilk test 0.9803 and 0.406 indicated that residuals appear as normally distributed in the two models. Moreover, ncvTest confirmed that error variance appears to be homoscedastic with P-values of 0.125 and 0.842 respectively. The plots of the two models (with filtred data) show that normality (Figure 3B and 4 B) and homoscedastic (Figure 3C and 4 C), and linearity (Figure 3A and 4 A) are respected

**Results and Analysis**

I constructed a model 1 with age and sex as predictors of pain. The P-Value of the predictor age is 0.000361, therefore I conclude that only age had a significant effect on pain. The coefficient of age was *-* 0.08541 indicating that an increase in unit age decreases the value of pain by 0.08541 times.

*The equation for model 1 is pain = 8.15772 0.30279\*sex(male) -0.08541 \*age*

The p-value of the (model 1) is 0.001061, r-squared of 0.08406, and 156 degrees of freedom. The confidence interval for model one is shown in table 1.

Then, I constructed a second model with sex, age, STAI, pain catastrophizing, mindfulness, and cortisol measures as predictor variables. Pain catastrophizing, mindfulness, cortisol serum, and age had a p-value of < 0.05 meaning that these predictors statistically significantly affect pain. A unit increase in age results in a decrease in pain -0.040803 times. A unit increase in pain catastrophizing results equals an increase in pain 0.108869times. A unit increase in mindfulness results equals a decrease in pain -0.272831 times. A unit increase in cortisol serum results equals an increase in pain 0.560298 times.

*The equation of model 2 is pain = 1.264919 + 0.169448 \* sex(male) - 0.040803\*age -0.001716\*STAI\_trait + 0.108869\*pain\_cat - 0.272831\*mindfulness + 0.560298\*cortisol\_serum*

The p-value of model 2 is < 2.2e-16, r-squared of 0.5221 and 152 degrees of freedom. The confidence interval for model one is shown in table 2.

**Discussion and Conclusion**

Comparing the two models, model 1 explains about three per cent of the variability of pain while model 2 explains about eleven per cent of the variability of pain. Model 1 has an AIC score of 576.9372 while model 2 has an AIC score of 481.5118. From the AIC scores, I can conclude that model 2 is the better fitting model, where the model with the lowest AIC and BIC score is preferred (model2). The F- test statistic of model 1 is 7.158 and its p-value is 0.001 while the F-test statistic is 27.67 and its p-value is < 2.2e-16.

In summary, I conclude that the best model is the second explaining that age mindfulness has a significantly negative effect on pain, while pain\_cat and cortisol\_serum have a very high significantly positive effect on pain.

**Assignment Part 2**

After publishing findings from the first assignment in a scientific journal, a fellow researcher’s commentary is published. She claimed that she got a better-adjusted R2 using my original data. A comparison of the two approaches against each other was done to prove which one was more effective in predicting pain.

**Results and Analysis**

I started by running a backward model on data file one (filtred) with the following variables as the predictor variables; sex, age, STAI, pain catastrophizing, mindfulness, cortisol serum, weight, and IQ. Hence, only 4 variables are retained (sex; age; pain\_cat; cortisol\_serum) in the model submitted to backward regression. A comparison of the initial model (the model submitted to backward regression), and the backward model reported that the backward model (with 4 predictors) is best fitted AIC => 528.2183 than the model submitted to backward regression AIC => 535.3447. Next, I run the AIC scores of the backward model and the theory-based model and the results were, that the backward model had an AIC score of 528.2183 and the theory-based model had an AIC score of 530.5249. From the results above, I can see that the backward model is the better fit of those two models.

*The equation for this model is pain = 0.32190 + 0.46339\*sex(male) -0.05654\*age+ 0.64975\*cortisol\_serum.*

The p-value is < 2.2e-16 and the F-test statistic is 33.66. The confidence intervals for the model are found in table 2.

**Discussion**

I made a prediction of the pain scores comparing the two models (the backward model, and the theory-based models. I found out that the backward model had an accuracy of 66.67% instead of 33.33% for the theory-based models. I then calculated the accuracy of the models based on AIC which revealed that the backward model is more accurate than the theory-based models with an AIC of 528.2183 for the backward model and AIC => 530.5249 for the theory-based model.

**Research question 3**

**Introduction**

The goal of the study was to build a mixed linear model with a hospital as a random effect in addition to predictors which are included as fixed effects. The LMM was built with the lmer function.

**Results and Analysis**

I build a mixed linear regression model on data 3 with fixed effect predictors. I used assignment part 1 as predictor variables, and hospital as random effects, and compared (the model coefficients and the confidence intervals of the coefficients) of the mixed model with those of the linear model (obtained in assignment part 1)

**The model coefficients of the LMM**

| (Intercept) | sexmale | Sex  woman | age | STAI\_  trait | pain\_  cat | mindfulness | cortisol\_  serum |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1.264918904 | 0.169447850 | 1.309 | -0.040803239 | -0.001716134 | 0.108869354 | -0.272831016 | 0.560298205 |

**The confidence intervals of the coefficients of the LMM**

|  | 2.50% | 97.50% |
| --- | --- | --- |
| (Intercept) | 1.0306 | 6.420186 |
| sexmale | -0.06259 | 0.569013 |
| sexwoman | -0.91336 | 3.52279 |
| age | -0.10062 | -0.02514 |
| STAI\_trait | -0.06128 | 0.020827 |
| pain\_cat | 0.035455 | 0.126767 |
| mindfulness | -0.42544 | -0.01654 |
| cortisol\_serum | 0.339685 | 0.693593 |

I predicted the pain values on data 4 with the model built earlier. The mean value for the prediction was 4.879 and the plot for the predictions is seen in figure 7.

R2 for Mixed Models

Conditional R2: 0.490

Marginal R2: 0.438

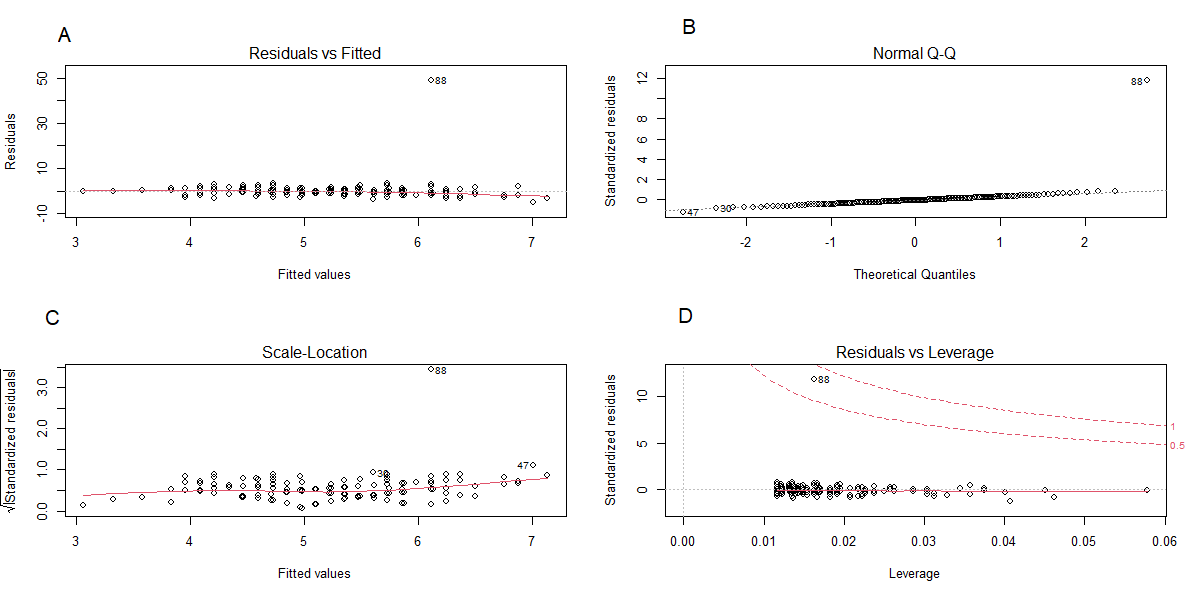
The variance explained by the model on data file 4

TSS = 222.5893

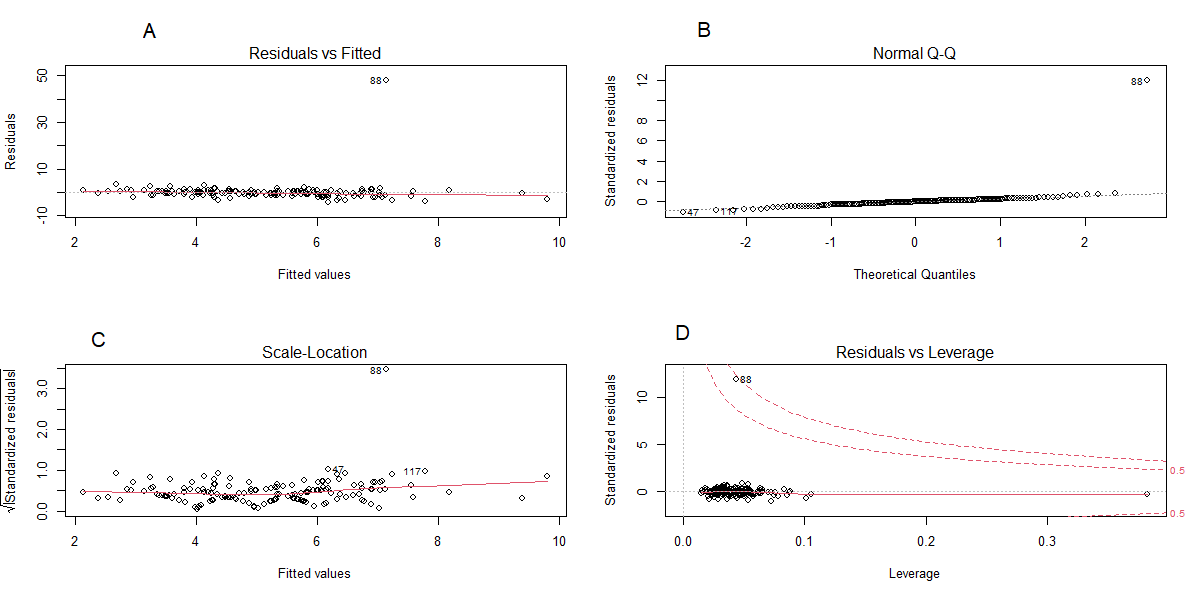
RSS = 222.5893

**Discussion**

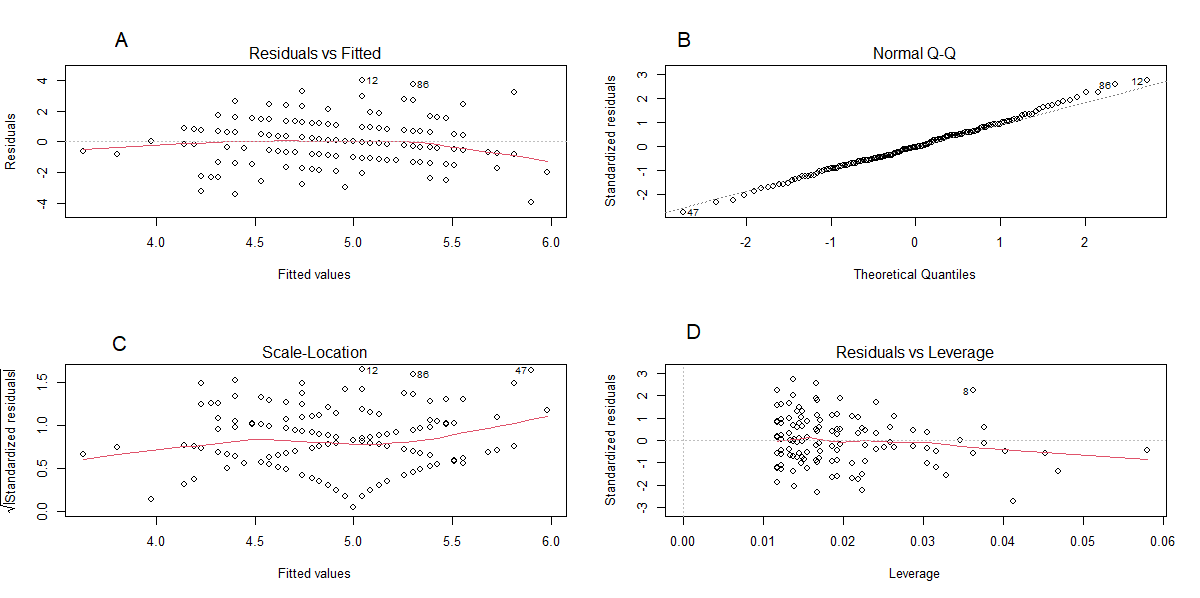
The coefficients of the new model decreased slightly as compared to those of the fixed-effects model. The decrease in the values of the coefficients is due to the inclusion of other variables in the model. The r-squared obtained was closer to the conditional obtained in data 3. This is because it is the data set on which the model was built thus a higher accuracy and r-squared. Based on the graph of predictions, the random intercept is the better of the two models because it measures different intercepts for each variable.

******

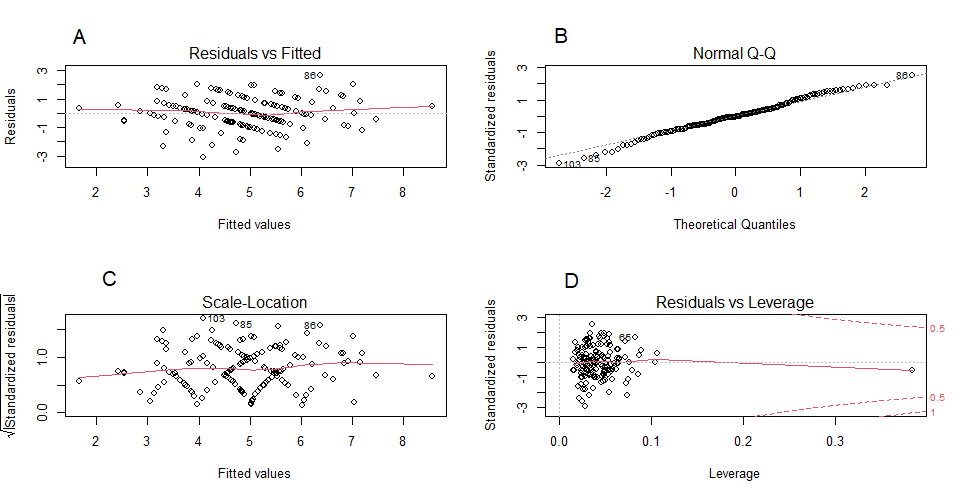
***Figure 1:*** Checking normality and homoscedasticity linearity of model 1



***Figure 2:*** Checking normality and homoscedasticity linearity of model 1

****

***Figure 3:*** Checking normality and homoscedasticity linearity of filtred model 1

****

***Figure 4:*** Checking normality and homoscedasticity linearity of filtred model 1

**Table 1:** Confidece interval for model 1

|  | 2.5% | 97.5 % |
| --- | --- | --- |
| Intercept | 6.3013795 | 10.0140511 |
| Sex(male) | -0.1582273 | 0.7638030 |
| Age | -0.1316715 | -0.0391477 |

**Table 2:** Confidence intervals for model 2

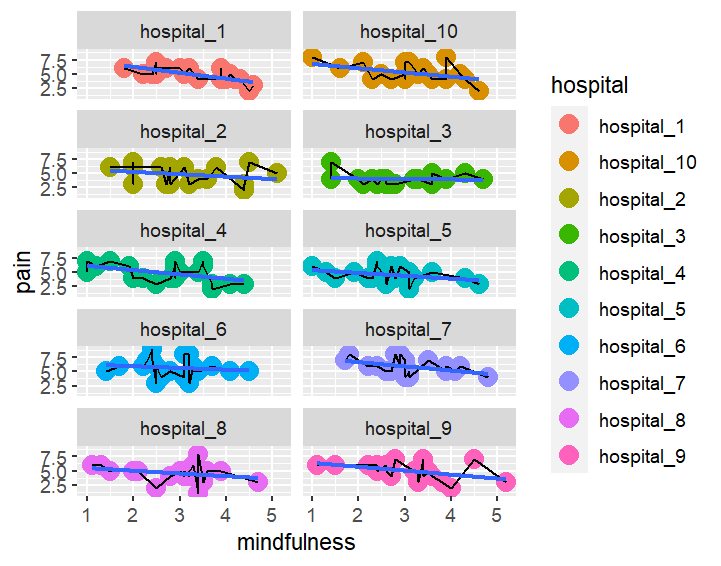
|  | 2.5% | 97.5 % |
| --- | --- | --- |
| Intercept | -1.35755000 | 3.887387813 |
| Sex(male) | -0.18487036 | 0.523766059 |
| Age | -0.07862423 | -0.002982245 |
| STAI\_trait | -0.03958360 | 0.036151330 |
| pain\_cat | 0.06568540 | 0.152053305 |
| mindfulness | -0.4802320 | -0.065429950 |
| cortisol\_serum | 0.34719621 | 0.773400197 |

**Table 3:** Confidence intervals for the backward model

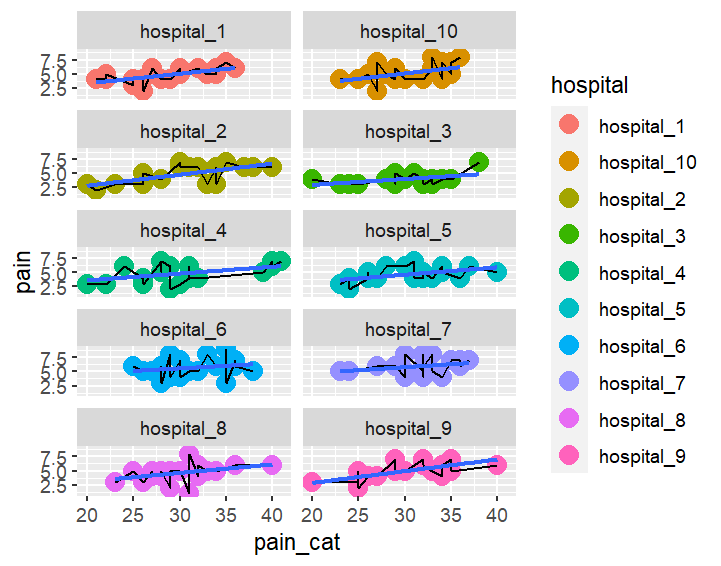
|  | 2.5% | 97.5 % |
| --- | --- | --- |
| Intercept | -2.07687548 | 2.7206783 |
| Sex(male) | 0.06420265 | 0.86257267 |
| age | -0.09672039 | -0.01636035 |
| Pain\_cat | 0.06514007 | 0.16157929 |
| Cortisol\_serum | 0.41932712 | 0.88017872 |

**Table 4:** Confidence intervals for the the model submitted to backward regression

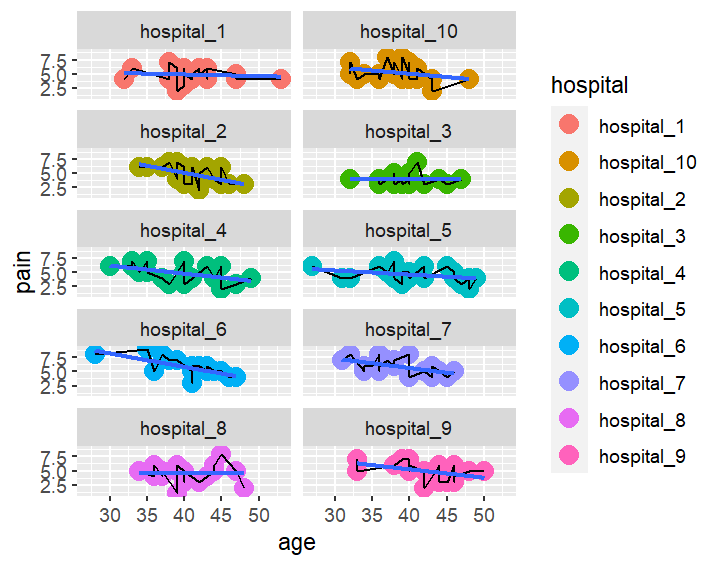
|  | 2.5% | 7.5% |
| --- | --- | --- |
| Intercept | -2.850539e+00 | 4.811281e+00 |
| Sex(male) | 2.668469e-02 | 8.674826e-01 |
| age | -9.744050e-02 | -1.066572e-02 |
| STAI\_trait | -6.592371e-02 | 4.045925e-02 |
| Pain\_cat | 5.731288e-02 | 1.675046e-01 |
| Mindfulness | -3.801973e-01 | 1.064706e-01 |
| Cortisol\_saliva | 3.755694e-01 | 8.908613e-01 |
| Weight | -1.915069e-02 | 2.042838e-02 |
| IQ | -1.142711e-02 | 2.092147e-02 |
| Household\_income | -1.078512e-05 | 4.297028e-06 |



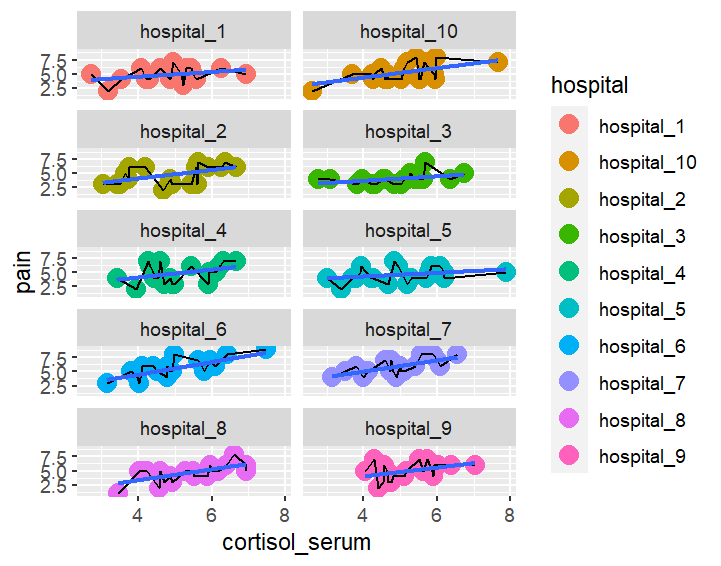
**Figure 5:** The fitted regression lines of the effect of mindfulness on pain for each hospital separately



**Figure 6:** The fitted regression lines of the effect of pain-cat on pain for each hospital separately



**Figure 7:** The fitted regression lines of the effect of age on pain for each hospital separately



**Figure 8:** The fitted regression lines of the effect of cortisol\_serum on pain for each hospital separately