**Data Analysis Report**

**Lund University**

**Master program in Psychology, autumn semester 2021**

**Nicole Evors**

**PSY14-HT20, 2021, autumn semester**

**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 the 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.

**Results and Analysis**

I started with cleaning the data sets in each part of the assignment, plus I filtered the IQ scores to keep only those that were greater than 80 for reliability reasons. I then constructed model one with age and sex as predictors of pain. For both sex and age p > 0.05. I rejected the null hypothesis which stated that sex and age have a significant effect on pain so in conclusion, both predictors do not statistically have a significant effect on pain. The coefficient of age was - 0.1194 indicating that an increase in unit age decreases the value of pain 0.1194 times. If one is male then pain increases by 0.9930 as compared to females.

*The equation for model 1 is pain = 0.993\*sex(male) – 0.1194\*age*

The p-value of model 1 is 0.1504, r-squared of 0.0271, and 138 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 the predictor variables. Only pain catastrophizing has a p < 0.05 meaning that it statistically significantly affects pain. When one is of the male gender the pain increases by 0.7052 compared to females. A unit increase in age results in a decrease in pain 0.0299 times. A unit increase in STAI results in a decrease in pain 0.0316 times. A unit increase in pain catastrophizing results equals an increase in pain 0.2362 times. A unit increase in mindfulness results equals an increase in pain 0.72074 times. A unit increase in cortisol serum results equals an increase in pain 0.65506 times.

*The equation of model 2 is pain = -5.1402 + 0.7052\*sex(male) - 0.0299\*age – 0.0316\*STAI\_trait + 0.2362\*pain\_cat + 0.7207\*mindfulness + 0.6651\*cortisol\_serum*

The p-value of model 2 is 0.0304, r-squared of 0.09734 and 134 degrees of freedom. The confidence interval for model one is shown in table 2.

Then, I checked to see if the assumptions of linear regression are true for both models. I began with model 1. The residuals of model 1 are not normally distributed as shown in the histogram in figure 1. There is linearity in model one as shown by figure 2. The red line is fairly straight indicating linearity. The variance of the residuals in model 1 increases as the predictor variables increases as shown in figure 3.

I checked to see if the assumptions of linear regression are true for model 2. The residuals of model 2 are not normally distributed as shown in the histogram in figure 4. There is linearity in model 2 as shown by figure 5. The red line is fairly straight showing linearity. The variance of the residuals in model 1 is constant as shown in figure 6. There is no multicollinearity in model 2 as all values have a VIF of less than 10.

**Discussion and Conclusion**

Comparing the two models, model 1 explains about three percent of the variability of pain while model 2 explains about eleven percent of the variability of pain. Model 1 has an AIC score of 825.81 while model 2 has an AIC score of 822.42. From the AIC scores, we can conclude that model 2 is the better fitting model. The F- test statistic of model 1 is 1.921 and its p-value is 0.1504 while the F-test statistic is 2.693 and its p-value is 0.0168.

In summary, I discovered pain is the only variable that statistically has significant effects on pain. There is a slight increase in variance explanation when psychological and hormone measures are added to the model.

**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 better adjusted R2 using my original data. Comparison 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 backward model on data file one with the following variables as the predictor variables; sex, age, STAI, pain catastrophizing, mindfulness, cortisol serum, weight, and IQ. The only variable that had a p <0.05 was pain catastrophizing thus it is the only variable that statistically has a significant effect on pain. If a person is of the male gender, then the pain increases by 0.6839 units compared to if one is of the female gender. A unit increase in age results in pain decreasing 0.0326 times. A unit increase in STAI\_trait results pain to decrease 0.0330 times. A unit increase in pain\_cat results pain to increase 0.2321 times. A unit increase in mindfulness results equals increase 0.7053 times. A unit increase in cortisol\_serum results pain increasing 0.6426 times. A unit increase in weight results pain to increase 0.0069 times. A unit increase in IQ results pain to increase 0.0106 times.

*The equation for this model is pain = -6.8169 + 0.6839\*sex(male) - 0.0325\*age – 0.0330\*STAI\_trait + 0.2321\*pain\_cat + 0.7053\*mindfulness + 0.6426\*cortisol\_serum + 0.0194\*weight + 0.0069\*IQ*

The p-value is 0.0756 and the F-test statistic is 1.837. The confidence intervals for the model are found in table 3.

Next, I run AIC scores of the backward model and the theory-based model and the results were, backward model had AIC score of 824.04 and the theory-based model had AIC score of 828.29. From the results above we can see that the backward model is the better fit out of those two.

**Discussion**

I made a prediction of the pain scores comparing the two models (backward model, the theory-based models. I found out that both models have an accuracy of 12.5%. The models are almost similar. I then calculated the accuracy of the models based on mean absolute error, the backward model had a mean absolute error of 1.1588 and the theory-based model had a mean absolute error of 1.1732. The backward model has a lower mean absolute error thus making it the better model. I would use the backward model to predict pain in an actual clinic context.

**Research question 3**

**Introduction**

The goal of the study was to increase the generalizability of the findings. Assessing the model coefficients and the overall predictive efficiency of the predictors in the model was done.

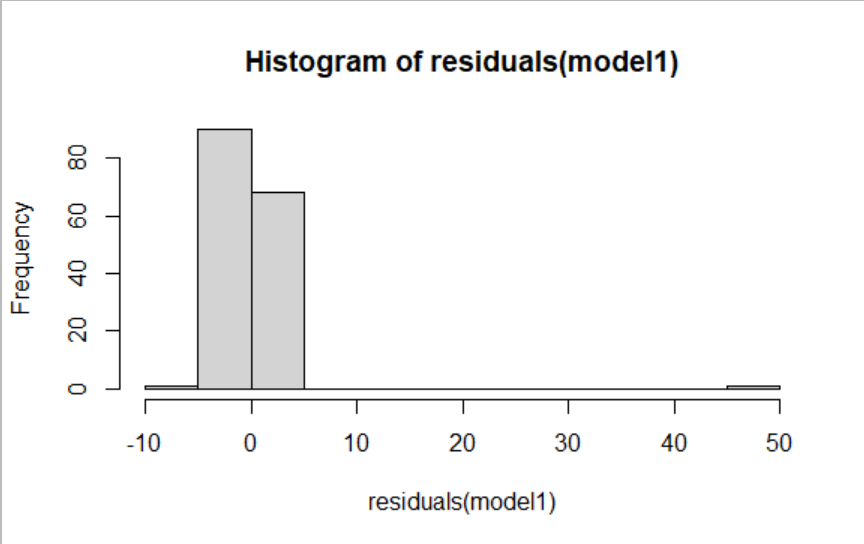
**Results and Analysis**

The first thing I did was filtering the IQ scores to those that were greater than 80 because those respondents seem to be more reliable. I then build a mixed linear regression model on data 3 with all the variables available as predictor variables. The results of the mixed linear regression model are as follows; if one is of the pain gender, pain increases by 0.7732 units, an increase unit in age results to a decrease in pain 0.0072 times, an increase unit in STAI\_trait results to a decrease in pain 0.0329 times, an increase unit in pain\_cat results to an increase in pain 0.2689 times, an increase unit in cortisol\_serum results to a decrease in pain 0.3839 times, an increase unit in cortisol\_saliva results to an increase in pain 1.17times, an increase unit in mindfulness results to an increase in pain 0.7635 times, an increase unit in weight results to an increase in pain 0.02465 times, an increase unit in IQ results to an increase in pain 0.01296 times and an increase unit in household\_income results to a decrease in pain 0.00001 times. The confidence intervals for the model are shown in table 4.

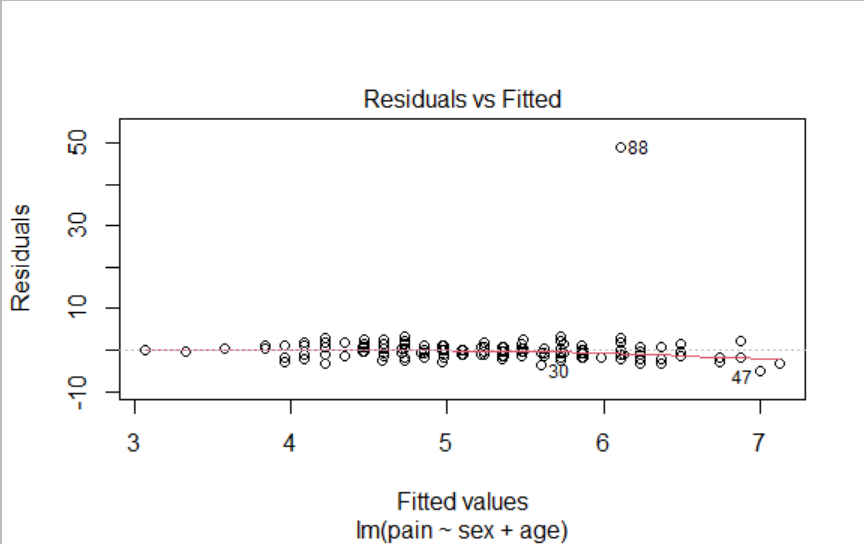
I predicted the pain values on data 4 with the model built earlier. The mean absolute error value for the prediction was 1.3741 and the plot for the predictions is seen in figure 7.

**Discussion**

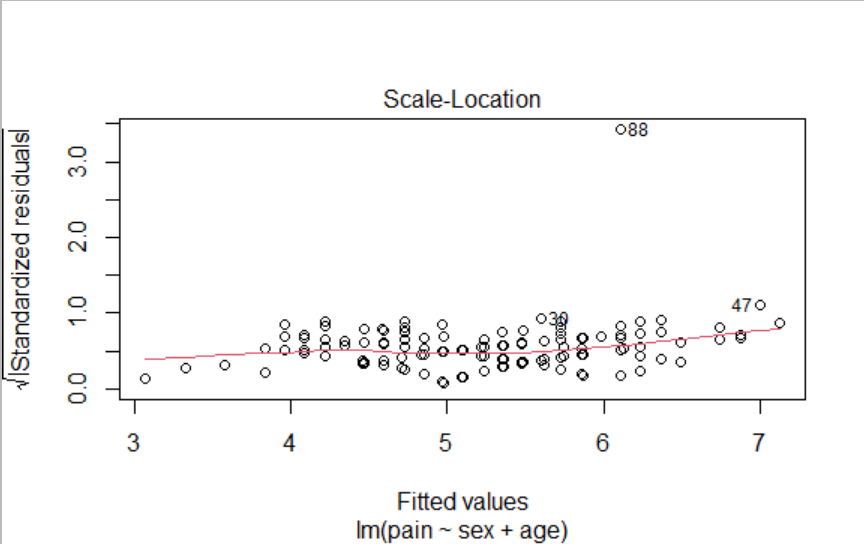
The coefficients of the new model decreased slightly as to compared to those of the fixed effects model. The decreasing of the values of the coefficients is due to the inclusion of other variables in the model. The r squared obtained in was closer to the conditional obtained in data 3. This is because 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.



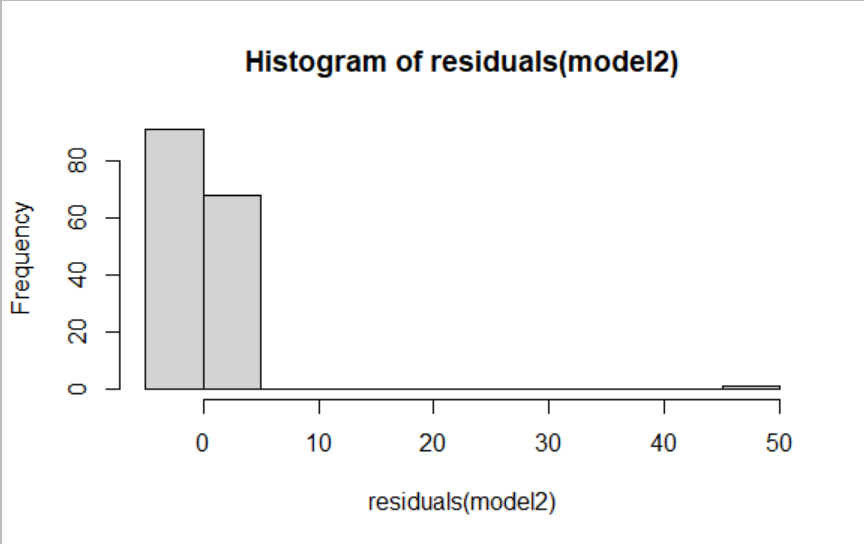
**Figure 1: Histogram of residuals of model 1**



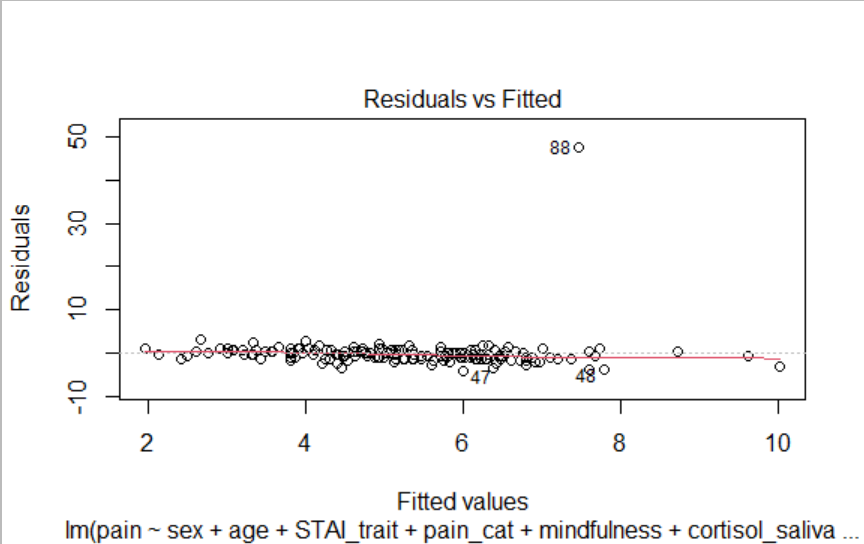
**Figure 2: Checking linearity of model 1**



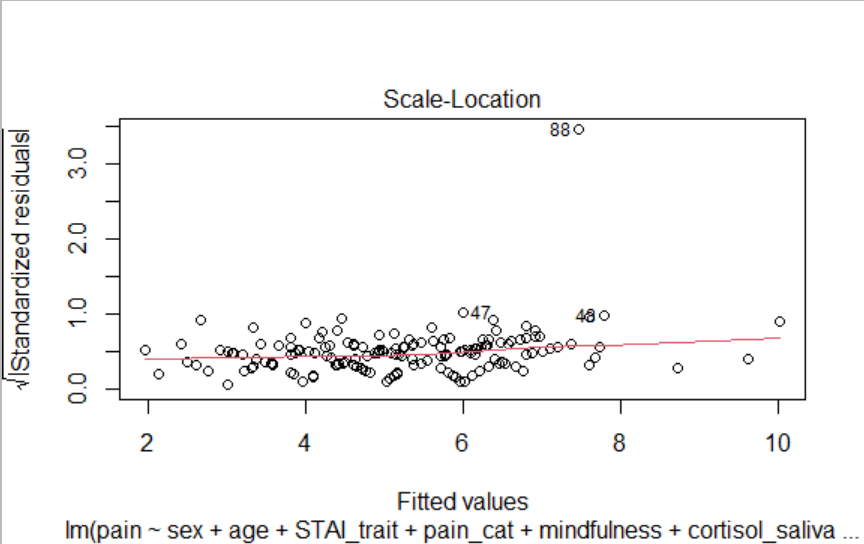
**Figure 3: Checking homoscedasticity of model 1**



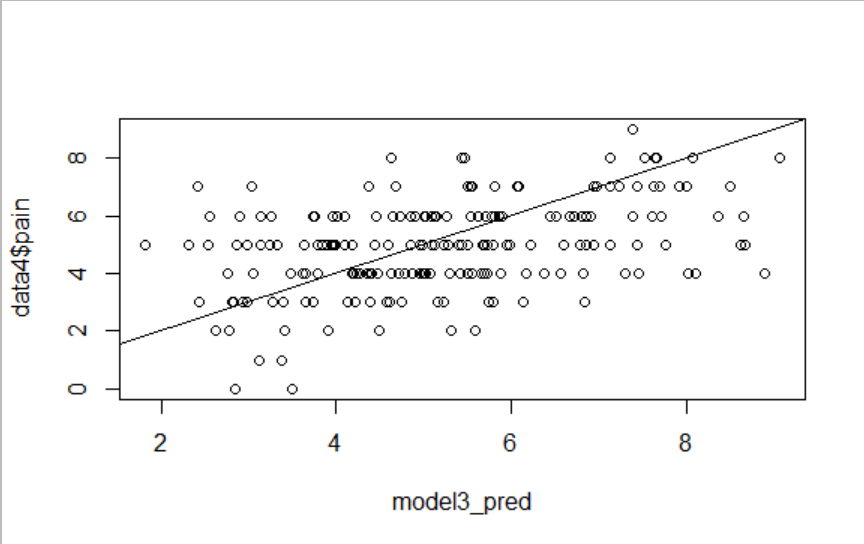
**Figure 4: Checking normality of residuals of model 1**



**Figure 5: Checking for linearity in model 2**



**Figure 6: Checking for homoscedasticity of model 2**



**Figure 7: Prediction of pain for model 3**

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

|  |  |  |
| --- | --- | --- |
|  | 2.5% | 7.5% |
| Intercept | 3.4220 | 15.5449 |
| Sex(male) | -0.4941 | 2.4802 |
| Age | -0.2713 | 0.0325 |

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

|  |  |  |
| --- | --- | --- |
|  | 2.5% | 7.5% |
| Intercept | -16.5178 | 6.2374 |
| Sex(male) | -0.8348 | 2.2452 |
| STAI\_trait | -0.1890 | 0.1376 |
| Pain\_cat | -0.0492 | 0.4231 |
| Mindfulness | -0.1978 | 1.6393 |
| Cortisol\_serum | -0.2678 | 1.5779 |

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

|  |  |  |
| --- | --- | --- |
|  | 2.5% | 7.5% |
| Intercept | -19.7960 | 6.1621 |
| Sex(male) | -0.9063 | 2.2740 |
| age | -0.2024 | 0.1373 |
| STAI\_trait | -0.1916 | 0.1256 |
| Pain\_cat | 0.0431 | 0.4211 |
| Mindfulness | -0.2281 | 1.6887 |
| Cortisol\_serum | -0.2903 | 1.5759 |
| Weight | -0.0476 | 0.0864 |
| IQ | -0.04789 | 0.0618 |

|  |  |  |
| --- | --- | --- |
|  | 2.5% | 7.5% |
| Intercept | -24.6963 | 5.2530 |
| Sex(male) | -0.8026 | 2.3489 |
| age | -0.1981 | 0.1838 |
| STAI\_trait | -0.1960 | 0.1301 |
| Pain\_cat | 0.0633 | 0.4745 |
| Mindfulness | -0.1664 | 1.6915 |
| Cortisol\_saliva | -0.6405 | 2.9958 |
| Weight | -0.0419 | 0.0912 |
| IQ | -0.0442 | 0.0676 |
| Household\_income | -0.00004 | 0.00001 |