Assignment\_3\_mode comparison\_AM

Marton Kovacs

2022-05-08

In this lab assignment you are going to work with (simulated) data related to perioperative pain and its psychological and hormonal predictors. In the assignment you will assess the added benefit of including some psychological and hormonal predictors to the already established demographic predictors of pain.

In this assignment you will set up a hierarchical regression model to predict postoperative pain after wisdom tooth surgery.

# Research problem

The amount of pain experienced around and after surgeries are highly variable between and within individuals. In order to improve surgical pain management regimens we need to understand what influences pain around surgical procedures and predict the amount of pain an individual will experience.

Your first study in this area is related to assessing the influence of trait and state psychological measures on pain, and to see whether taking into account these variables can improve our understanding of postoperative pain.

# Procedures and measures

Use the data file called ‘assignment\_3\_dataset’, from the ‘data/’ folder.

You have collected data from 160 adults who were scheduled to undergo surgical extraction of the third mandibular molar (wisdom tooth surgery). Patients filled out a form in the waiting room before their surgery. The form contained questions about their sex, age, and weight, and psychological questionnaires assessing anxiety, pain catastrophizing, and mindfulness (see descriptions below). You also got blood samples and saliva samples from participants in the waiting room 5 minutes before their operations to determine the serum (a component of the blood) and salivary cortisol levels of participants. Participants were contacted 5 hours after the surgery to see how much pain they were experiencing. The **level of pain** at that moment was recorded using a numerical rating scale using a **scale of 0 to 10**, where 0 means “no pain” and 10 means “worst pain I can imagine”.

**The State Trait Anxiety Inventory:** T measures trait anxiety on a scale of 20 to 80, higher scores mean higher anxiety. Anxiety has been found in many studies to positively correlate with the level of pain experienced. This is **variable STAI\_trait** in the dataset.

**The Pain Catastrophizing Scale** measures the extent of pain catastrophizing, which is characterized by a tendency to magnify the threat value of a pain stimulus and to feel helpless in the presence of pain, as well as by a relative inability to prevent or inhibit pain-related thoughts in anticipation of, during, or following a painful event. The total score on this scale ranges from 0 to 52, higher scores mean higher catastrophizing. Pain catastrophizing is one of the well-established predictors of clinical pain. This is **variable pain\_cat** in the dataset.

**The Mindful Attention Awareness Scale (MAAS)** measures dispositional mindfulness, which may be described as a tendency to turn attention to present-moment experiences in an open, non-judgmental way. The MAAS total score ranges from 1 to 6 (an average of the item scores), with higher scores representing higher dispositional mindfulness. Trait mindfulness has been theorized to serve as a protective factor against pain, as the individual would be more objective about their pain experience and tend to associate less discomfort, despair, and hopelessness to the pain-related sensations. This is **variable mindfulness** in the dataset.

**Cortisol** is a stress hormone associated with acute and chronic stress. Cortisol levels are thought to be positively associated with pain experience. Cortisol can be **measured from both blood and the saliva**, although, serum cortisol is often regarded in medical research as more reliably related to stress (serum is a component of the blood plasma). These are **variables cortisol\_serum**, and **cortisol\_saliva** in the dataset.

# Research question

Previous studies and meta-analyses showed that age and sex are often predictors of pain (age is negatively associated with pain, while sex is a predictor more dependent on the type of the procedure). You would like to determine the extent to which taking into account psychological and hormonal variables aside from the already used demographic variables would improve our understanding of postoperative pain.

To answer this research question you will **need to compare two models** (with a hierarchical regression). The **simpler model** should contain **age and sex as predictors of pain**, while the **more complex model** should contain the **predictors: age, sex, STAI, pain catastrophizing, mindfulness, and cortisol measures**. Notice that the predictors used in the simpler model are a subset of the predictors used in more complex model. **You will have to do model comparison to assess whether substantial new information was gained about pain in the more complex model compared to the simpler model.**

# What to report

As usual, before you can interpret your model, you will need to run data and model diagnostics. First, check the variables included in the more complex model (age, sex, STAI, pain catastrophizing, mindfulness, and cortisol measures as predictors, and pain as an outcome) for **coding errors**, and the model itself for **influential outliers** (for example using Cook’s distance). Furthermore, check the final model to see if the **assumptions of linear regression hold true**, that is, **normality** (of the residuals), **linearity** (of the relationship), **homogeneity of variance** (also called homoscedasticity) and that there is no excess **multicollinearity** (“uncorrelated predictors” in Navarro’s words). If you find anything amiss during these checks, make the appropriate decision or correction and report your findings and actions in your report.

**Note:** If you do any changes, such as exclude cases, or exclude predictors from the model, you will have to re-run the above checks for your final data and model.

Report the results of the simpler model and the more complex model. For both models you should report the model test statistics (adj.R2, F, df, and p value). Also, report the statistics describing the coefficients of the predictors in a table format (unstandardized regression coefficients and 95% confidence intervals, standardized regression coefficients (B and Beta values), and p values).

Write up the regression equation of the more complex model in the form of 𝑌 = 𝑏0 + 𝑏1 ∗ X1 + 𝑏2 ∗ X2 +…+ bn \* Xn, in which you use the actual regression coefficients of your models. (b0 stands for the intercept and b1, b2 … bn stand for the model coefficients for each of the predictors, and X1, X2, … Xn denote the predictors).

Compare the two models in terms of how much variance they explain of pain’s variability in the sample. Report Akaike information criterion (AIC) for both models and the F test statistic and p value of the likelihood ratio test comparing the two models.

# What to discuss

In your discussion of thefindings, briefly interpret the results of the above analyses, and indicate whether you think that anything was gained by including the psychological and hormone measures in the model. # Solution

## Read the data

Read the dataset used in this assignment. Pay attention to the extension of the datafile.

library(readxl)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.5 ✔ purrr 0.3.4  
## ✔ tibble 3.1.6 ✔ dplyr 1.0.8  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

library(broom)  
  
  
pain <- read\_excel("C:/Users/Lenovo/Downloads/assignment\_3\_dataset (1).xlsx")

## Data and model diagnostics

### Data diagnostics

#### Descriptives of the variables

Run an exploratory data analysis (EDA) to investigate the dataset.

glimpse(pain)

## Rows: 160  
## Columns: 12  
## $ ID <chr> "ID\_1", "ID\_2", "ID\_3", "ID\_4", "ID\_5", "ID\_6", "ID\_7…  
## $ pain <dbl> 5, 4, 5, 7, 4, 6, 4, 5, 4, 5, 6, 3, 4, 5, 7, 4, 4, 4,…  
## $ sex <chr> "female", "female", "female", "female", "male", "fema…  
## $ age <dbl> 38, 36, 51, 39, 48, 45, 45, 35, 39, 33, 46, 42, 36, 4…  
## $ STAI\_trait <dbl> 39, 46, 49, 48, 36, 37, 34, 39, 31, 38, 37, 35, 37, 3…  
## $ pain\_cat <dbl> 25, 31, 32, 41, 26, 28, 22, 32, 27, 36, 26, 29, 26, 4…  
## $ cortisol\_serum <dbl> 4.67, 6.01, 5.18, 6.65, 2.95, 4.32, 5.32, 5.26, 3.36,…  
## $ cortisol\_saliva <dbl> 4.78, 6.71, 4.75, 6.68, 3.20, 4.36, 5.19, 5.53, 3.30,…  
## $ mindfulness <dbl> 3.83, 2.82, 3.93, 1.90, 4.39, 3.01, 4.97, 2.95, 3.43,…  
## $ weight <dbl> 62.14, 97.13, 78.96, 78.50, 76.88, 47.22, 55.10, 68.2…  
## $ IQ <dbl> 99, 71, 97, 97, 129, 115, 74, 98, 93, 76, 81, 103, 10…  
## $ household\_income <dbl> 52505, 64993, 68598, 47822, 100580, 43960, 65548, 436…

summary(pain)

## ID pain sex age   
## Length:160 Min. : 1.000 Length:160 Min. :24.00   
## Class :character 1st Qu.: 4.000 Class :character 1st Qu.:38.00   
## Mode :character Median : 5.000 Mode :character Median :41.00   
## Mean : 5.194 Mean :40.74   
## 3rd Qu.: 6.000 3rd Qu.:44.00   
## Max. :50.000 Max. :53.00   
## STAI\_trait pain\_cat cortisol\_serum cortisol\_saliva  
## Min. :26.00 Min. :14.00 Min. :2.700 Min. :2.530   
## 1st Qu.:37.00 1st Qu.:27.00 1st Qu.:4.268 1st Qu.:4.120   
## Median :40.00 Median :30.00 Median :4.995 Median :5.035   
## Mean :40.09 Mean :29.88 Mean :4.941 Mean :4.952   
## 3rd Qu.:43.00 3rd Qu.:32.25 3rd Qu.:5.633 3rd Qu.:5.753   
## Max. :52.00 Max. :42.00 Max. :7.110 Max. :7.190   
## mindfulness weight IQ household\_income  
## Min. :1.000 Min. :33.80 Min. : 52.00 Min. : 5548   
## 1st Qu.:2.595 1st Qu.:62.15 1st Qu.: 91.75 1st Qu.: 53489   
## Median :3.115 Median :68.77 Median :101.00 Median : 72149   
## Mean :3.204 Mean :68.57 Mean :100.74 Mean : 70426   
## 3rd Qu.:3.830 3rd Qu.:74.65 3rd Qu.:109.00 3rd Qu.: 87734   
## Max. :6.220 Max. :97.13 Max. :144.00 Max. :139268

colSums(is.na(pain))

## ID pain sex age   
## 0 0 0 0   
## STAI\_trait pain\_cat cortisol\_serum cortisol\_saliva   
## 0 0 0 0   
## mindfulness weight IQ household\_income   
## 0 0 0 0

#### Correct coding errors

If you find values in the dataset during the EDA, that are not correct based on the provided descriptions of the variables of the dataset please correct them here.

Based on the data structure, the sex variable is labeled as character, and this is incorrect, since variable sex is considered as categorical data which is factor not character variable. Therefore we need to change the sex variable to be factor variable.

pain <- pain %>%  
 mutate(sex = as.factor(sex)) %>%   
 mutate(sex = recode\_factor(sex, "woman" = "female"))

## Double check using correct data set

view(pain)  
summary(pain)

## ID pain sex age   
## Length:160 Min. : 1.000 female:84 Min. :24.00   
## Class :character 1st Qu.: 4.000 male :76 1st Qu.:38.00   
## Mode :character Median : 5.000 Median :41.00   
## Mean : 5.194 Mean :40.74   
## 3rd Qu.: 6.000 3rd Qu.:44.00   
## Max. :50.000 Max. :53.00   
## STAI\_trait pain\_cat cortisol\_serum cortisol\_saliva  
## Min. :26.00 Min. :14.00 Min. :2.700 Min. :2.530   
## 1st Qu.:37.00 1st Qu.:27.00 1st Qu.:4.268 1st Qu.:4.120   
## Median :40.00 Median :30.00 Median :4.995 Median :5.035   
## Mean :40.09 Mean :29.88 Mean :4.941 Mean :4.952   
## 3rd Qu.:43.00 3rd Qu.:32.25 3rd Qu.:5.633 3rd Qu.:5.753   
## Max. :52.00 Max. :42.00 Max. :7.110 Max. :7.190   
## mindfulness weight IQ household\_income  
## Min. :1.000 Min. :33.80 Min. : 52.00 Min. : 5548   
## 1st Qu.:2.595 1st Qu.:62.15 1st Qu.: 91.75 1st Qu.: 53489   
## Median :3.115 Median :68.77 Median :101.00 Median : 72149   
## Mean :3.204 Mean :68.57 Mean :100.74 Mean : 70426   
## 3rd Qu.:3.830 3rd Qu.:74.65 3rd Qu.:109.00 3rd Qu.: 87734   
## Max. :6.220 Max. :97.13 Max. :144.00 Max. :139268

##Based on the double check, there is no more category of women in the variable sex and all the data is clear. ### Model diagnostics #### Build the more complex model

In order to test the more complex model for outliers and to test the assumptions first build the model. #checking the regression model with preductors variables are sex, age, STAI- trait, pain category,mindifullness, cortisol serum, and cortisol saliva

complex\_reggmod <- lm(pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness + cortisol\_serum + cortisol\_saliva, data = pain)  
  
nrow(model.frame(complex\_reggmod))

## [1] 160

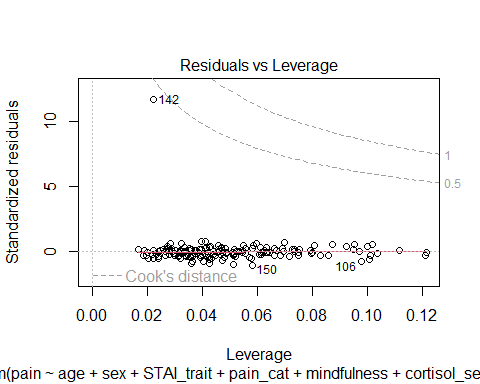
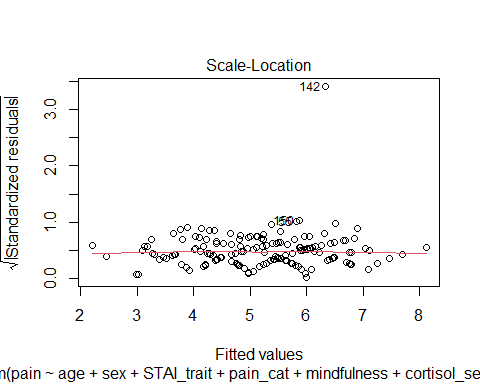
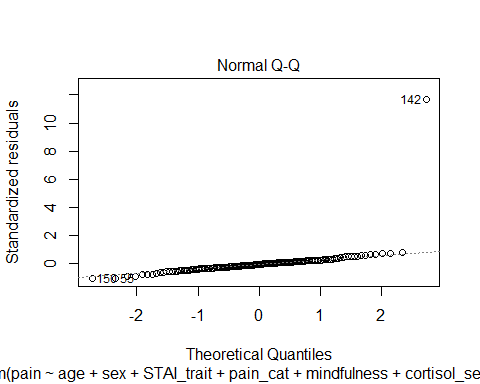
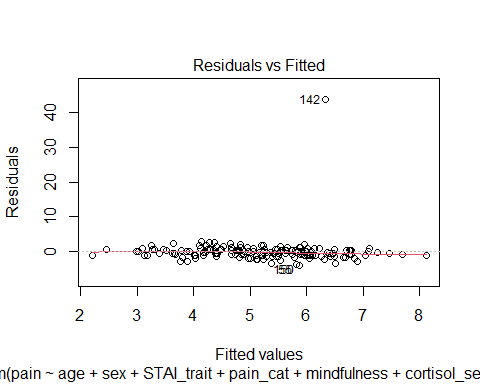
summary(complex\_reggmod)

##   
## Call:  
## lm(formula = pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness +   
## cortisol\_serum + cortisol\_saliva, data = pain)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.874 -1.072 -0.209 0.527 43.678   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.009962 5.253537 -0.383 0.703  
## age 0.001216 0.073615 0.017 0.987  
## sexmale -0.065604 0.641907 -0.102 0.919  
## STAI\_trait -0.012388 0.085089 -0.146 0.884  
## pain\_cat 0.094006 0.087834 1.070 0.286  
## mindfulness 0.073098 0.380279 0.192 0.848  
## cortisol\_serum -0.048785 0.748362 -0.065 0.948  
## cortisol\_saliva 0.985497 0.767880 1.283 0.201  
##   
## Residual standard error: 3.791 on 152 degrees of freedom  
## Multiple R-squared: 0.08499, Adjusted R-squared: 0.04285   
## F-statistic: 2.017 on 7 and 152 DF, p-value: 0.05639

#### Checking for influential outliers

Check for outlier values in the model.

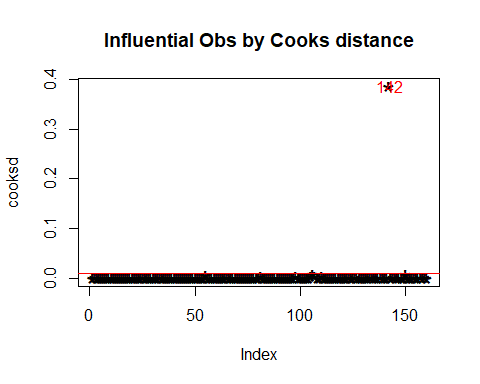
plot(complex\_reggmod)



## Based on the plot residual vs leverage, there is a potential outlier from data number 142.Therefore for further detection, we will check by cook's distance command.

cooksd <- cooks.distance(complex\_reggmod)  
## Plot the cook's distance using the traditional criterion

cooksd <- cooks.distance(complex\_reggmod)  
plot(cooksd, pch="\*", cex=2, main="Influential Obs by Cooks distance")  
abline(h = 3\*mean(cooksd, na.rm=T), col="red")  
text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>4\*mean(cooksd, na.rm=T),names(cooksd),""), col="red")



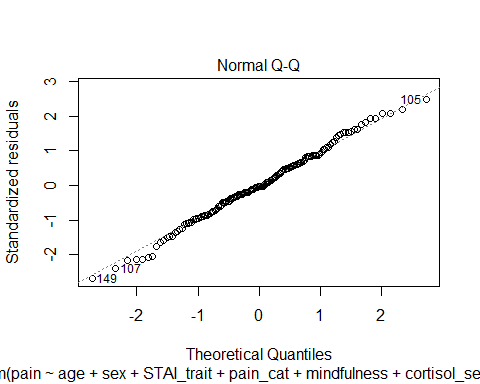
pain\_final <- pain[-142, ]  
complex\_mod\_clean <- lm(pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness + cortisol\_serum + cortisol\_saliva, data = pain\_final)  
summary(complex\_mod\_clean)

##   
## Call:  
## lm(formula = pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness +   
## cortisol\_serum + cortisol\_saliva, data = pain\_final)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.2331 -0.7638 -0.0435 0.7736 2.9558   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.35209 1.72405 0.784 0.43412   
## age -0.02395 0.02413 -0.992 0.32256   
## sexmale 0.31689 0.21061 1.505 0.13451   
## STAI\_trait -0.02554 0.02788 -0.916 0.36119   
## pain\_cat 0.09491 0.02878 3.298 0.00122 \*\*  
## mindfulness -0.12942 0.12474 -1.038 0.30114   
## cortisol\_serum 0.16662 0.24529 0.679 0.49801   
## cortisol\_saliva 0.43773 0.25209 1.736 0.08452 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.242 on 151 degrees of freedom  
## Multiple R-squared: 0.3648, Adjusted R-squared: 0.3354   
## F-statistic: 12.39 on 7 and 151 DF, p-value: 1.768e-12

#### Checking assumptions

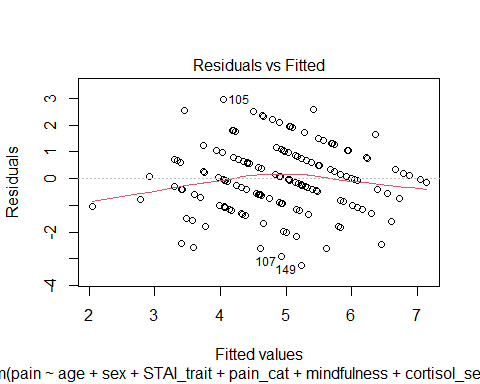
# Check the normality assumption.

plot(complex\_mod\_clean, which = 2)

 The residuals plot looks normally distributed based on the above plot. It can be concluded that the regression meets the assumption of normality.

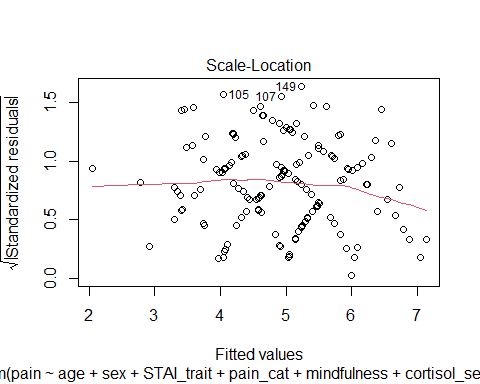
#Check the linearity assumption.

plot(complex\_mod\_clean, which = 1)

 The plot of residuals vs fitted values indicate that there is no pattern in the residual plot. This result indicate that there is a linear relationship between the predictors and the outcome.

#Check the homoscedasticty assumption (homogeneity of variance).

plot(complex\_mod\_clean, which = 3)



library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

bptest(complex\_mod\_clean)

##   
## studentized Breusch-Pagan test  
##   
## data: complex\_mod\_clean  
## BP = 6.4386, df = 7, p-value = 0.4896

The result of Breusch-pagan test indicates that there is no heterocedastity since the variants of residual is homogenize as the result of p-value (0.4) is higher than 0.05.

#Check the multicollinearity assumption.

vif(complex\_mod\_clean)

## age sex STAI\_trait pain\_cat mindfulness   
## 1.526619 1.140617 2.117744 1.948680 1.530694   
## cortisol\_serum cortisol\_saliva   
## 6.065266 6.622923

In general, a VIF above 10 indicates high correlation and is cause for concern. There are some other arguments about the threshold score for VIF that Some authors suggest a more conservative level of 2.5 or above. A rule for interpreting the variance inflation factor: 1 = not correlated. Between 1 and 5 = moderately correlated. Greater than 5 = highly correlated. Further information related to the VIF interpretation can be checked in this link: #(VIF above 5), or a VIF threshold of 3 is recommended the following links: <http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2009.00001.x/full>

<https://statisticalhorizons.com/multicollinearity> <http://blog.minitab.com/blog/understanding-statistics/handling-multicollinearity-in-regression-analysis>

The VIF test result indicated that there are two variables have VIF score above 5 means they are highly correlated. They are cortisol\_serum and cortisol saliva.

Due to this result, then we have to check the correlation between tose two variables.

#checking for the correlation between cortisol serum and cortisol saliva

cor.test(pain\_final$cortisol\_serum, pain\_final$cortisol\_saliva)

##   
## Pearson's product-moment correlation  
##   
## data: pain\_final$cortisol\_serum and pain\_final$cortisol\_saliva  
## t = 27.255, df = 157, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.8769531 0.9323752  
## sample estimates:  
## cor   
## 0.908583

Pearson’s correlation coeficient between those variables are high, above 0.90. Thus, we have to drop one variable which has lower effect from previous regression analysis which is cortisol\_serum and we shoul keep cortisol\_saliva in the model.

### Making decision based on model diagnostics

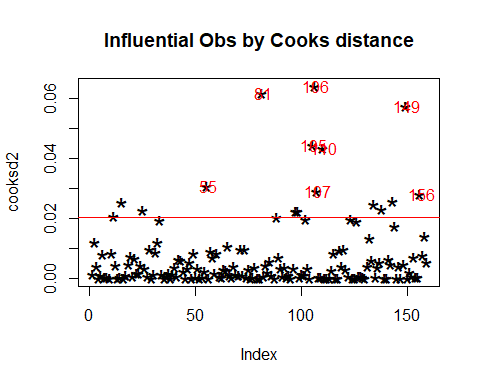
If based on the assumption tests you decide to drop a predictor variable you should do that here. Create your updated model.

complex\_mod\_updated <- lm(pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness + cortisol\_saliva, data = pain\_final)  
summary(complex\_mod\_updated)

##   
## Call:  
## lm(formula = pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness +   
## cortisol\_saliva, data = pain\_final)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.3045 -0.7315 -0.0800 0.7903 3.0861   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.31715 1.72022 0.766 0.445052   
## age -0.02236 0.02398 -0.933 0.352549   
## sexmale 0.31188 0.21011 1.484 0.139783   
## STAI\_trait -0.02736 0.02771 -0.987 0.325008   
## pain\_cat 0.09927 0.02800 3.545 0.000522 \*\*\*  
## mindfulness -0.12800 0.12450 -1.028 0.305536   
## cortisol\_saliva 0.58590 0.12613 4.645 7.3e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.24 on 152 degrees of freedom  
## Multiple R-squared: 0.3629, Adjusted R-squared: 0.3377   
## F-statistic: 14.43 on 6 and 152 DF, p-value: 5.475e-13

#### Checking outliers of the updated model

cooksd2 <- cooks.distance(complex\_mod\_updated)  
plot(cooksd2, pch="\*", cex=2, main="Influential Obs by Cooks distance")  
abline(h = 3\*mean(cooksd2, na.rm=T), col="red")  
text(x=1:length(cooksd2)+1, y=cooksd2, labels=ifelse(cooksd2>4\*mean(cooksd2, na.rm=T),names(cooksd2),""), col="red")

 The above plot shows that there are some outliers : number 81, 106,105, 110, and 149. All the outliers will be deleted from the data. #new data without outliers

pain\_updated <- pain\_final[c(-81, -105, -106, -149, -81),]

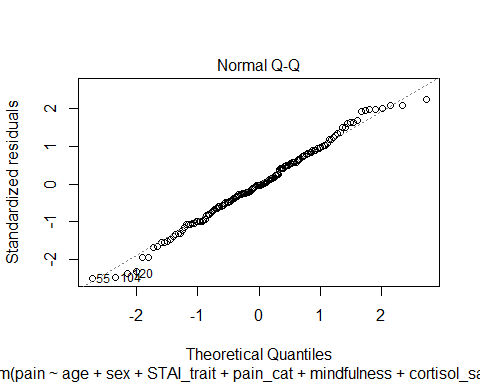
complex\_mod\_final <- lm(pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness + cortisol\_saliva, data = pain\_updated)  
summary(complex\_mod\_final)

##   
## Call:  
## lm(formula = pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness +   
## cortisol\_saliva, data = pain\_updated)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.81310 -0.70548 -0.05038 0.75213 2.52672   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3460135 1.6301662 0.212 0.832199   
## age -0.0234512 0.0229198 -1.023 0.307890   
## sexmale 0.3472562 0.1992415 1.743 0.083430 .   
## STAI\_trait -0.0006069 0.0272043 -0.022 0.982233   
## pain\_cat 0.1060612 0.0268573 3.949 0.000121 \*\*\*  
## mindfulness -0.0840213 0.1167905 -0.719 0.473017   
## cortisol\_saliva 0.5016771 0.1231631 4.073 7.53e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.157 on 148 degrees of freedom  
## Multiple R-squared: 0.3981, Adjusted R-squared: 0.3737   
## F-statistic: 16.32 on 6 and 148 DF, p-value: 2.274e-14

#### Checking assumptions of the updated model

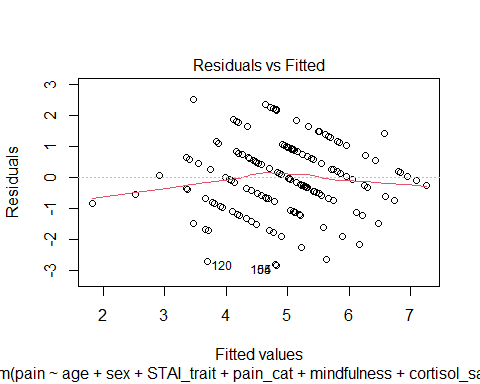
Normality assumption

plot(complex\_mod\_final, which = 2)



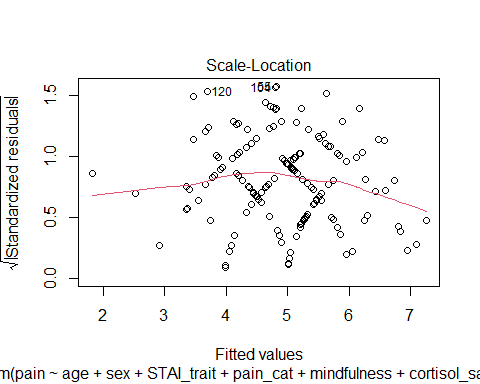
Linearity assumption

plot(complex\_mod\_final, which = 1)



Homoscedasticty assumption (homogeneity of variance)

plot(complex\_mod\_final, which = 3)



library(lmtest)  
bptest(complex\_mod\_final)

##   
## studentized Breusch-Pagan test  
##   
## data: complex\_mod\_final  
## BP = 9.2288, df = 6, p-value = 0.1611

Multicollinearity assumption

vif(complex\_mod\_final)

## age sex STAI\_trait pain\_cat mindfulness   
## 1.550720 1.147090 2.226380 1.885904 1.534431   
## cortisol\_saliva   
## 1.780388

The updated regression model has met the assumption of linearity, normality of residuals, homoscedasticity (the p value is greater than 0.5) , and multicolinearity (there are no VIF score above 5).

## Model comparison

Report the results of the simpler model and the more complex model. For both models you should report the model test statistics (adj.R2, F, df, and p value). Also, report the statistics describing the coefficients of the predictors in a table format (unstandardized regression coefficients and 95% confidence intervals, standardized regression coefficients (B and Beta values), and p values).

Write up the regression equation of the more complex model in the form of 𝑌 = 𝑏0 + 𝑏1 ∗ X1 + 𝑏2 ∗ X2 +…+ bn \* Xn, in which you use the actual regression coefficients of your models. (b0 stands for the intercept and b1, b2 … bn stand for the model coefficients for each of the predictors, and X1, X2, … Xn denote the predictors).

Compare the two models in terms of how much variance they explain of pain’s variability in the sample. Report Akaike information criterion (AIC) for both models and the F test statistic and p value of the likelihood ratio test comparing the two models.

#Create the simple model and get the results of the model that needs to be reported based on the What to report section.

simple\_mod<- lm(pain ~ age + sex, data = pain\_final)  
summary(simple\_mod)

##   
## Call:  
## lm(formula = pain ~ age + sex, data = pain\_final)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.6617 -0.9974 0.1464 1.0508 3.5026   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.49145 0.94828 8.955 9.64e-16 \*\*\*  
## age -0.08906 0.02304 -3.865 0.000163 \*\*\*  
## sexmale 0.10287 0.23268 0.442 0.659028   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.464 on 156 degrees of freedom  
## Multiple R-squared: 0.08784, Adjusted R-squared: 0.07614   
## F-statistic: 7.511 on 2 and 156 DF, p-value: 0.0007684

All predictors simultaneously can predict pain, with F(2, 156) = 7.511, p < 0.01. All predictors could explain 8% variance of pain (R-squared=0.08). The equation for the model is: pain = 8.49 - 0.09 \* age + 0.10 \* sex

#Create the more complex model based on the results of the model diagnostics. Also, get the results that needs to be reported based on the What to report section.

complex\_mod\_final <- lm(pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness + cortisol\_saliva, data = pain\_updated)  
summary(complex\_mod\_final)

##   
## Call:  
## lm(formula = pain ~ age + sex + STAI\_trait + pain\_cat + mindfulness +   
## cortisol\_saliva, data = pain\_updated)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.81310 -0.70548 -0.05038 0.75213 2.52672   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.3460135 1.6301662 0.212 0.832199   
## age -0.0234512 0.0229198 -1.023 0.307890   
## sexmale 0.3472562 0.1992415 1.743 0.083430 .   
## STAI\_trait -0.0006069 0.0272043 -0.022 0.982233   
## pain\_cat 0.1060612 0.0268573 3.949 0.000121 \*\*\*  
## mindfulness -0.0840213 0.1167905 -0.719 0.473017   
## cortisol\_saliva 0.5016771 0.1231631 4.073 7.53e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.157 on 148 degrees of freedom  
## Multiple R-squared: 0.3981, Adjusted R-squared: 0.3737   
## F-statistic: 16.32 on 6 and 148 DF, p-value: 2.274e-14

All predictors simultaneously can predict pain, with F(6, 148) = 16.32, p < 0.001. All predictors could explain 39,81% variance of pain (R-squared=0.398). The equation for the model is: pain = 0.34 - 0.02 \* age + 0.34 \* sex - 0.01 \* STAIT\_trait *+ 0.10*  pain\_cat - 0.08 \* mindfulness \* + 0.50 \* cortisol\_saliva

# Compare the two models.

comparing two model with anova error models were not all fitted to the same size of dataset