Regression models with fixed and random effects PSYP14 (HT2020) Assignment 1

Karima Kandi

2020-12-11

## Assignment 1

#loading necessary packages  
suppressMessages(suppressWarnings(  
 library(psych, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(lsr, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(tidyverse, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(readr, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(sciplot, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(reshape2, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(MASS, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(dplyr, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(smacof, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(lm.beta, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(gridExtra, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(magrittr, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(ggplot2, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(lmtest, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(car, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(sandwich, quietly = TRUE, warn.conflicts = FALSE)))  
suppressMessages(suppressWarnings(  
 library(lmboot, quietly = TRUE, warn.conflicts = FALSE)))

#### Research question 1

Conduct a hierarchical regression to compare two models predicting postoperative pain. Model 1 uses age and sex as predictors, while model 2 uses age, sex, STAI, pain catastrophizing, mindfulness, and cortisol measures.

data\_sample\_1 = read.csv("https://tinyurl.com/ha-dataset1")  
  
view(data\_sample\_1) #to check is all is ok

Data management

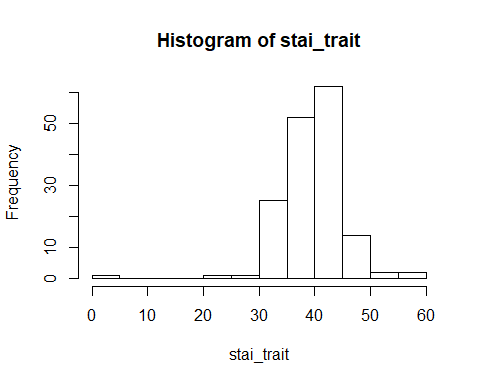
###data management  
  
#age, sex, stai\_trait, pain\_cat, mindfulness, cortisol\_serum, cortisol\_saliva, pain  
#create dataset with correct variables  
my.data <- data.frame(data\_sample\_1$age, data\_sample\_1$sex, data\_sample\_1$pain\_cat, data\_sample\_1$pain,  
 data\_sample\_1$mindfulness, data\_sample\_1$cortisol\_serum, data\_sample\_1$cortisol\_saliva,  
 data\_sample\_1$ID, data\_sample\_1$STAI\_trait, stringsAsFactors = TRUE)  
  
#high values indicate high postoperative pain  
pain <- as.numeric(my.data$data\_sample\_1.pain) #range 0-10  
  
#high values indicate high pain catastrophizing  
pain\_cat <- as.numeric(my.data$data\_sample\_1.pain\_cat) #range 0-52  
  
age <- as.numeric(my.data$data\_sample\_1.age)  
  
#high values indicate higher dispositional mindfulness  
mindful<- my.data$data\_sample\_1.mindfulness #1-6  
  
cortisol\_serum <- my.data$data\_sample\_1.cortisol\_serum  
cortisol\_saliva<- my.data$data\_sample\_1.cortisol\_saliva  
sex <- data\_sample\_1$sex  
  
#High values indicate high trait anxiety  
stai\_trait <- data\_sample\_1$STAI\_trait #range 20-80  
  
id <- my.data$data\_sample\_1.ID

###Data exploration  
summary(my.data)

## data\_sample\_1.age data\_sample\_1.sex data\_sample\_1.pain\_cat data\_sample\_1.pain  
## Min. : 27.00 female:90 Min. :19.00 Min. : 2.000   
## 1st Qu.: 37.00 male :70 1st Qu.:26.00 1st Qu.: 4.000   
## Median : 40.00 Median :30.00 Median : 5.000   
## Mean : 42.52 Mean :29.96 Mean : 5.037   
## 3rd Qu.: 43.00 3rd Qu.:34.00 3rd Qu.: 6.000   
## Max. :444.00 Max. :42.00 Max. :10.000   
##   
## data\_sample\_1.mindfulness data\_sample\_1.cortisol\_serum  
## Min. :1.000 Min. :1.920   
## 1st Qu.:2.357 1st Qu.:4.415   
## Median :2.950 Median :5.010   
## Mean :2.899 Mean :5.062   
## 3rd Qu.:3.500 3rd Qu.:5.680   
## Max. :5.630 Max. :8.140   
##   
## data\_sample\_1.cortisol\_saliva data\_sample\_1.ID data\_sample\_1.STAI\_trait  
## Min. :1.850 ID\_1 : 1 Min. : 3.90   
## 1st Qu.:4.442 ID\_10 : 1 1st Qu.:37.00   
## Median :5.050 ID\_100 : 1 Median :40.50   
## Mean :5.071 ID\_101 : 1 Mean :40.02   
## 3rd Qu.:5.832 ID\_102 : 1 3rd Qu.:43.25   
## Max. :7.490 ID\_103 : 1 Max. :56.00   
## (Other):154

Looking for outliers

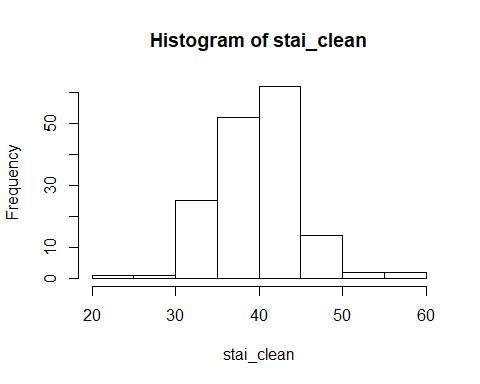
#outlier search  
#histogram show value under minimum (20)  
hist(stai\_trait) #range 20-80



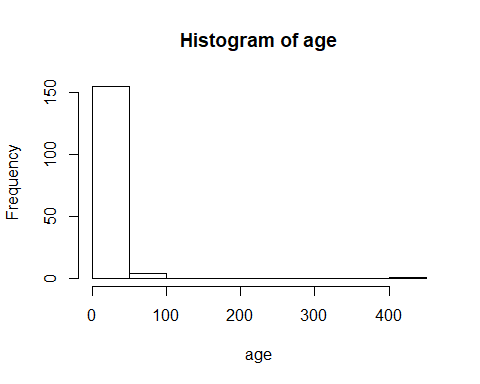
#finding row number for stai outlier  
my.data[my.data$data\_sample\_1.STAI\_trait <20,]

## data\_sample\_1.age data\_sample\_1.sex data\_sample\_1.pain\_cat  
## 150 40 female 30  
## data\_sample\_1.pain data\_sample\_1.mindfulness data\_sample\_1.cortisol\_serum  
## 150 5 2.94 4.75  
## data\_sample\_1.cortisol\_saliva data\_sample\_1.ID data\_sample\_1.STAI\_trait  
## 150 4.82 ID\_150 3.9

#setting stai outlier to NA  
stai\_clean<- na\_if(stai\_trait, 3.9)  
#new histogram for stai, clean variable  
hist(stai\_clean)



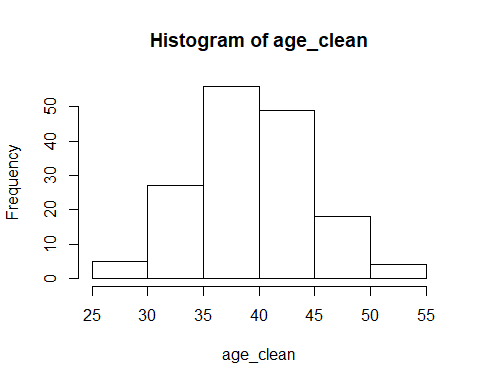
#histogram showed an age outlier  
hist(age)



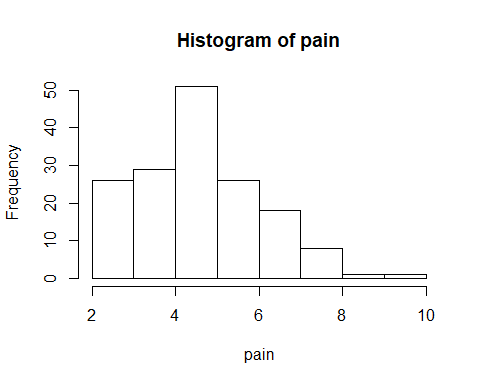
#finding row number for age outlier  
my.data[my.data$data\_sample\_1.age==444,]

## data\_sample\_1.age data\_sample\_1.sex data\_sample\_1.pain\_cat  
## 93 444 male 31  
## data\_sample\_1.pain data\_sample\_1.mindfulness data\_sample\_1.cortisol\_serum  
## 93 5 3.67 5.14  
## data\_sample\_1.cortisol\_saliva data\_sample\_1.ID data\_sample\_1.STAI\_trait  
## 93 4.42 ID\_93 39

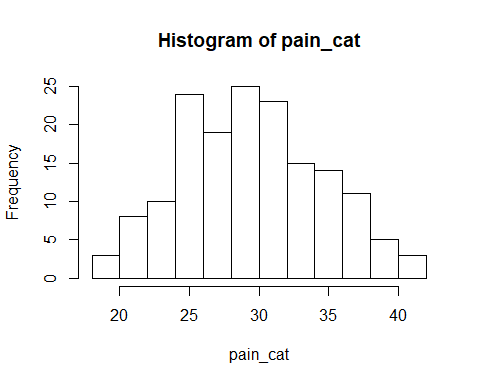
#setting age outlier to NA  
age\_clean<- na\_if(age, 444)  
  
#new histogram for age, clean variable  
hist(age\_clean) #min 27, max 53



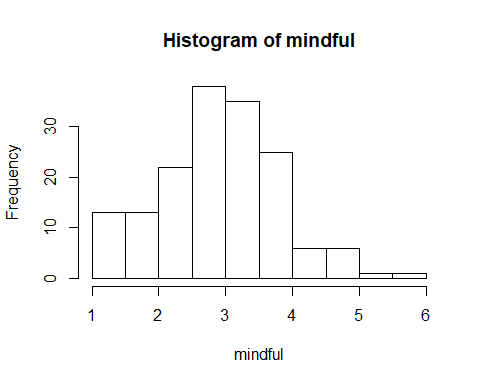
hist(pain) #range 0-10, normal



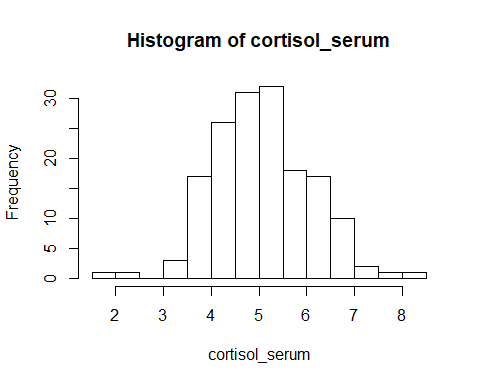
hist(pain\_cat) #range 0-52, normal



hist(mindful)#range 1-6, normal

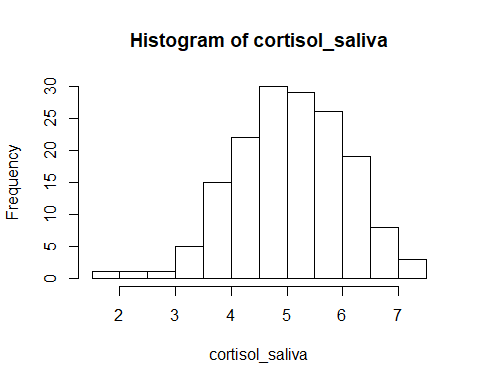


hist(cortisol\_serum) #min 1.92 max 8.14



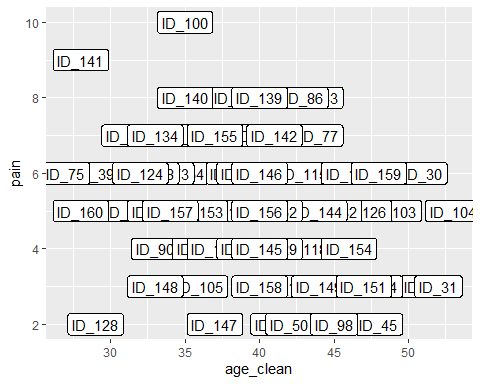
#overall normal, some smaller values compared by overall distribution

hist(cortisol\_saliva)#min 1.85, max 17.49, normal

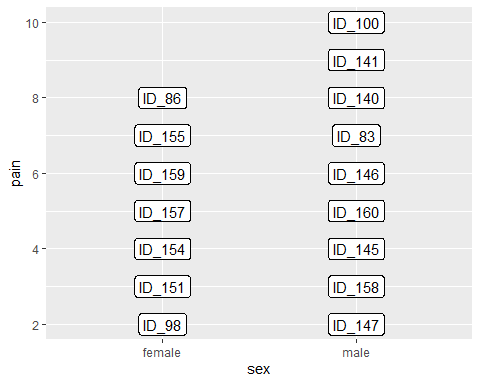


#new dataset with clean variables  
data\_ass1 <- data.frame(pain, pain\_cat, age\_clean, mindful, cortisol\_saliva,cortisol\_serum, sex, stai\_clean, id)  
  
#diagnostic 1: univariate outlier check  
data\_ass1 %>% ggplot()+  
 aes(x = age\_clean, y = pain, label = id)+  
 geom\_label() #id 100 sticks out slightly

## Warning: Removed 1 rows containing missing values (geom\_label).

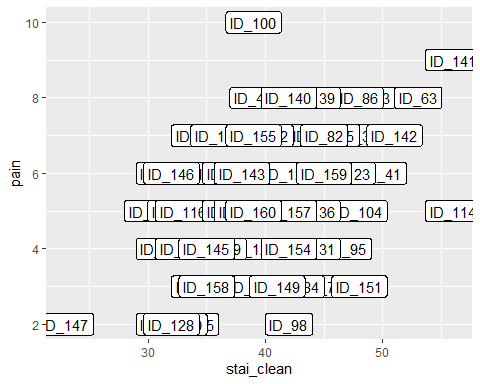


data\_ass1 %>% ggplot()+  
 aes(x = sex, y = pain, label = id)+  
 geom\_label() #id 100 and 141 sticks out slightly

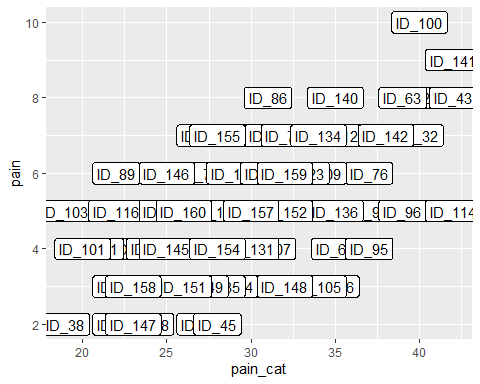


data\_ass1 %>% ggplot()+  
 aes(x = stai\_clean, y = pain, label = id)+  
 geom\_label()#id 100 sticks out a bit on the scale of pain (high), 114 sticks out a bit on the scale of pain (low for stai value)

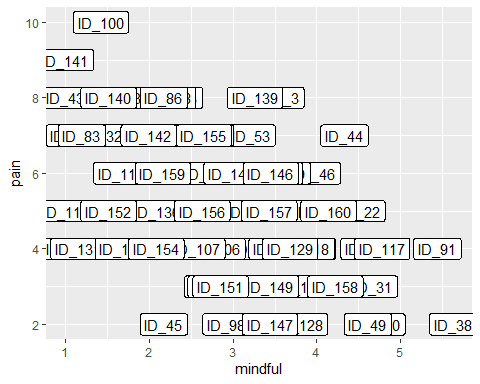
## Warning: Removed 1 rows containing missing values (geom\_label).



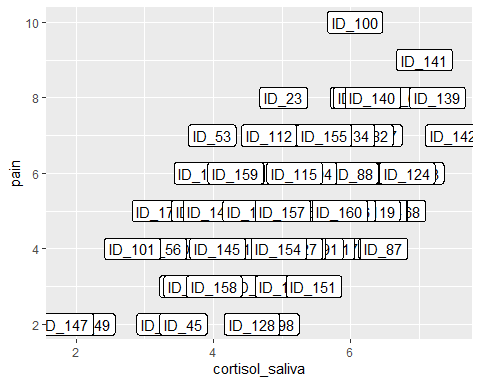
data\_ass1 %>% ggplot()+  
 aes(x = pain\_cat, y = pain, label = id)+  
 geom\_label() #looks good



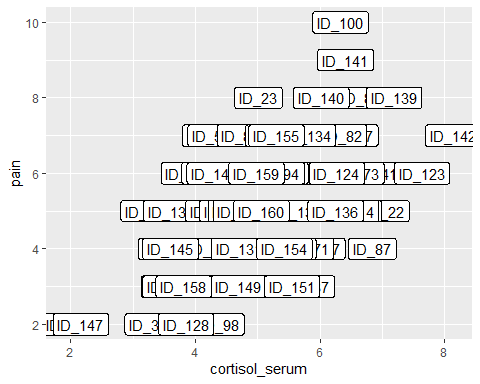
data\_ass1 %>% ggplot()+  
 aes(x = mindful, y = pain, label = id)+  
 geom\_label() #looks good



data\_ass1 %>% ggplot()+  
 aes(x = cortisol\_saliva, y = pain, label = id)+  
 geom\_label() #looks good



data\_ass1 %>% ggplot()+  
 aes(x = cortisol\_serum, y = pain, label = id)+  
 geom\_label() #looks good



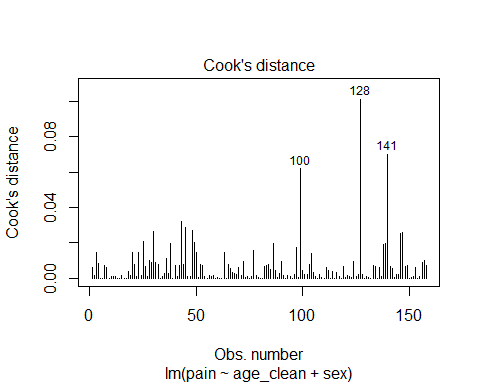
### Regression models and assumption checks

#omitting 2 participants with missing values, to ensure same number of observations in both models  
data\_ass1\_clean <- na.omit(data\_ass1)  
  
#build regression models  
mod.1 <- lm(pain ~ age\_clean+sex, data\_ass1\_clean) #Model from previous research  
mod.2 <- lm(pain~age\_clean+sex+stai\_clean+pain\_cat+mindful+cortisol\_saliva+cortisol\_serum, data\_ass1\_clean) #Zoltan's suggested model

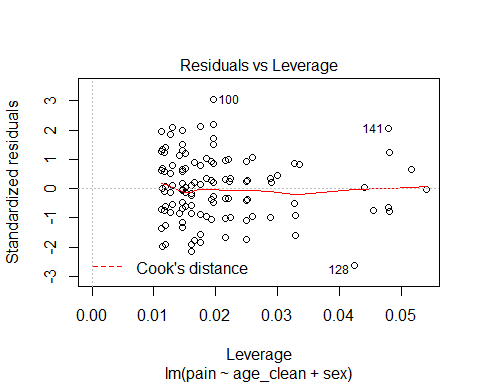
### Model diagnostics before I look at the results of each regression model

Leverage: looking for outliers among residuals that unduly influence the regression coefficients. Cut off values to look for are determined by the formula 4/N.

#cook's distance examined for model 1. According to cut off value >4/158 = 0.025 cases 100, 128, and 141 have high leverage.   
mod.1 %>% plot(which=4)



mod.1 %>% plot(which=5)

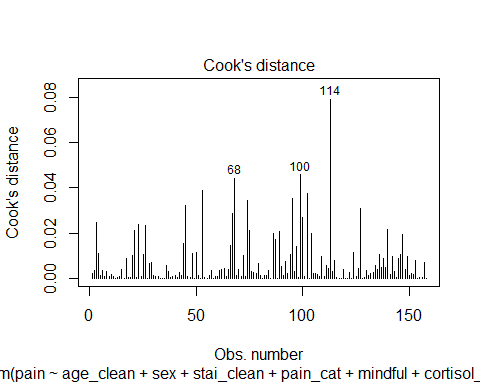
 Looking at the data for participants with higher leverage.

data\_ass1\_clean%>% slice(c(100,128,141))

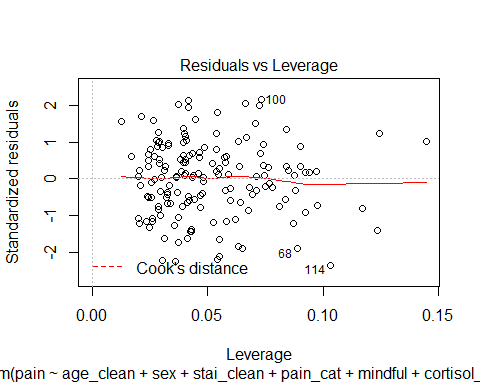
## pain pain\_cat age\_clean mindful cortisol\_saliva cortisol\_serum sex  
## 1 4 20 36 3.89 2.82 3.53 male  
## 2 4 26 41 3.68 4.50 4.28 male  
## 3 7 38 41 1.99 7.49 8.14 female  
## stai\_clean id  
## 1 33 ID\_101  
## 2 36 ID\_129  
## 3 51 ID\_142

Conclusion for model 1: Noticed that the values on other predictors seem reasonable for these outliers, so I decided to leave participants in. Considered also the rule of thumb for Cook’s distance values below 1, and no values were above 1.

#cook's distance examined for model 2. According to cut off value >4/158 = 0.025 cases 114, 100, 68 have high leverage.   
mod.2 %>% plot(which=4)



mod.2 %>% plot(which=5)

 Looking at the data for participants with higher leverage.

data\_ass1\_clean%>% slice(c(114, 100,68))

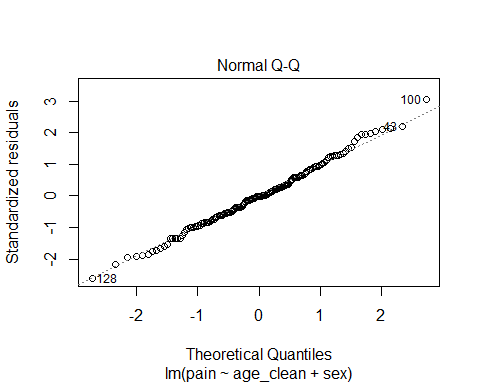
## pain pain\_cat age\_clean mindful cortisol\_saliva cortisol\_serum sex  
## 1 6 29 43 3.15 5.18 4.86 male  
## 2 4 20 36 3.89 2.82 3.53 male  
## 3 5 36 31 1.53 6.74 5.34 male  
## stai\_clean id  
## 1 38 ID\_115  
## 2 33 ID\_101  
## 3 44 ID\_68

Conclusion for model 2: Noticed that the values on other predictors seem reasonable for these outliers, so I decided to leave participants in. Considered also the rule of thumb for Cook’s distance values below 1, and no values were above 1.

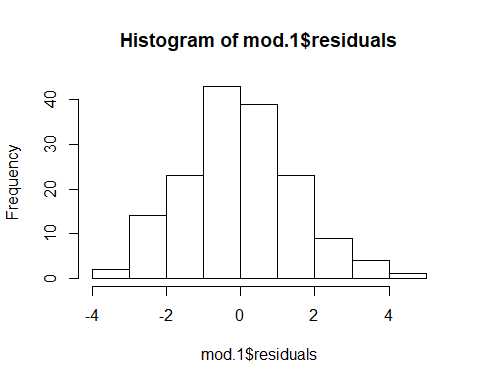
### Checking assumptions of linear regression for both models.

Step 1: normality

#Model 1   
mod.1%>% plot(which=2) #shows normality of residuals, 100 deviate from the line at the top



#Model 1  
hist(mod.1$residuals) #acceptably normal distribution



#Model 1  
psych::describe(residuals(mod.1)) #skew = 0.18, kurtosis = -0.04, no noticably large values

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 158 0 1.51 -0.02 -0.02 1.45 -3.89 4.59 8.48 0.18 -0.04 0.12

#Model 1, confidence intervals for predictors in model 1 with 95% significance level  
confint(mod.1, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) 6.2665620 10.18530671  
## age\_clean -0.1272040 -0.03145388  
## sexmale -0.5190811 0.44898978

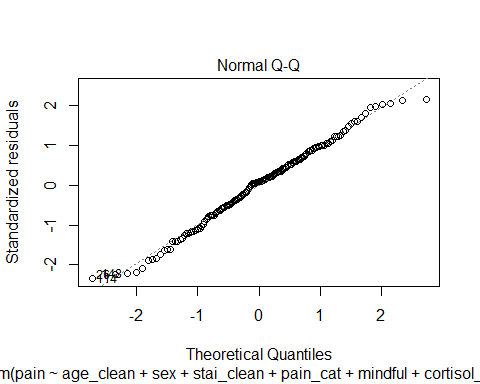
Restrictive confidence interval for regression estimated that take into account the small deviations in normality for residuals. Comparing it to 0.95 confidence intervals.

#Model 1, confidence intervals for predictors in model 1 with 99% significance level  
confint(mod.1, level = 0.99)

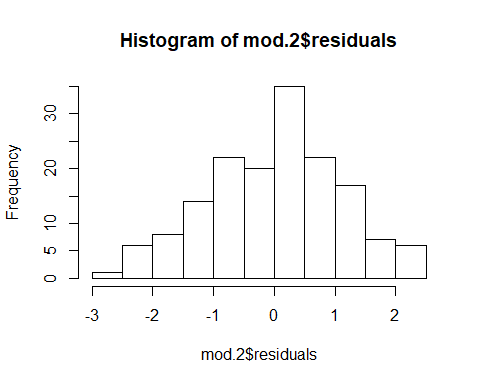
## 0.5 % 99.5 %  
## (Intercept) 5.6391572 10.81271152  
## age\_clean -0.1425339 -0.01612395  
## sexmale -0.6740726 0.60398133

No large differences between 0.95 and 0.99 level confidence intervals, 0.99 could be considered for reporting.

#Model 2  
mod.2%>% plot(which=2) #shows normality of residuals, 114, 26, 148 deviate from the line at the bottom



#Model 2  
hist(mod.2$residuals) #acceptably normal distribution



#Model 2  
psych::describe(residuals(mod.2)) #skew = -0.12, kurtosis = -0.42, no noticably large values

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 158 0 1.1 0.1 0.01 1.08 -2.5 2.35 4.85 -0.12 -0.42 0.09

Restrictive confidence interval for regression estimated that take into account the small deviations in normality for residuals. Comparing it to 0.95 confidence intervals.

#Model 2, confidence intervals for predictors in model 1 with 95% significance level  
confint(mod.2, level = 0.95)

## 2.5 % 97.5 %  
## (Intercept) -2.42351784 3.75241932  
## age\_clean -0.06212395 0.02854258  
## sexmale -0.13366468 0.63195311  
## stai\_clean -0.09182915 0.01891264  
## pain\_cat 0.07667173 0.18309013  
## mindful -0.48302607 -0.00766111  
## cortisol\_saliva -0.07934637 0.88459428  
## cortisol\_serum -0.21494266 0.68062588

#Model 2, confidence intervals for predictors in model 1 with 99% significance level  
confint(mod.2, level = 0.99)

## 0.5 % 99.5 %  
## (Intercept) -3.41292980 4.74183128  
## age\_clean -0.07664913 0.04306775  
## sexmale -0.25631997 0.75460840  
## stai\_clean -0.10957047 0.03665396  
## pain\_cat 0.05962304 0.20013882  
## mindful -0.55918160 0.06849442  
## cortisol\_saliva -0.23377385 1.03902176  
## cortisol\_serum -0.35841662 0.82409985

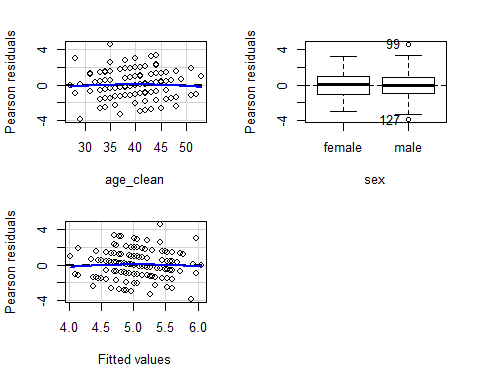
No large differences between 0.95 and 0.99 level confidence intervals, 0.99 could be considered

Conclusion for model 1 and 2: Both models pass checks for normality of residuals

### Step 2: linearity

Residual plot for model 1

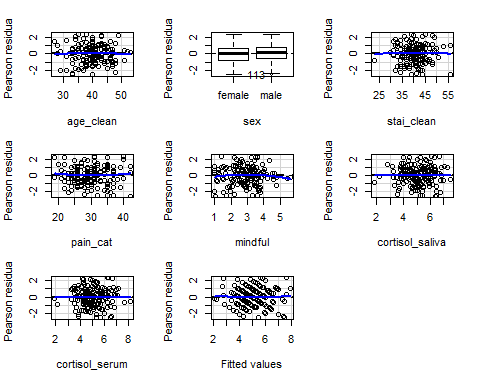
residualPlots(model = mod.1) #no large deviations from linearity



## Test stat Pr(>|Test stat|)  
## age\_clean -0.4618 0.6449  
## sex   
## Tukey test -0.4541 0.6498

Residual plot for model 2

residualPlots(model = mod.2) #no large deviations from linearity



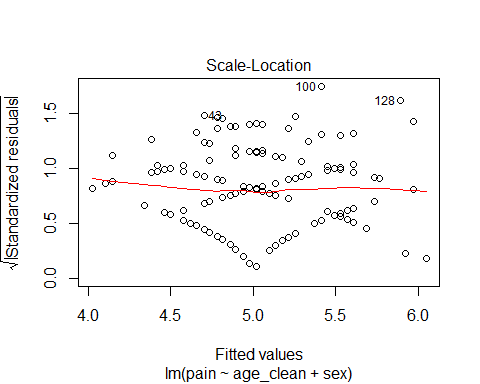
## Test stat Pr(>|Test stat|)  
## age\_clean -0.1377 0.8907  
## sex   
## stai\_clean -0.0469 0.9626  
## pain\_cat 0.6549 0.5135  
## mindful -0.9497 0.3438  
## cortisol\_saliva 0.0213 0.9831  
## cortisol\_serum -0.1142 0.9092  
## Tukey test 0.1660 0.8681

Conclusion for model 1 and 2: Both models pass checks for linearity for residuals.

### Step 3: Homoscedasticity

Heteroscedasticity tests for model 1

plot(mod.1, which = 3) #no visible heteroscedasticity



ncvTest(mod.1) #non-significant test, no heteroscedasticity indicated

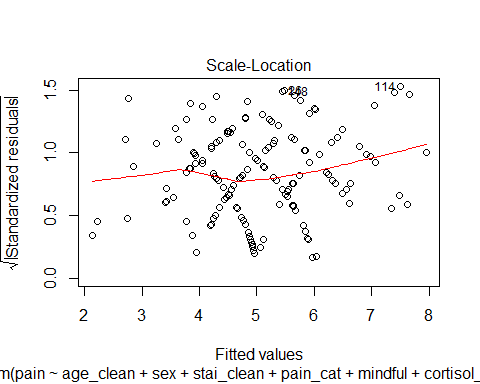
## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 0.9637738, Df = 1, p = 0.32624

#breusch test  
bptest(mod.1) #non-significant test, no heteroscedasticity indicated

##   
## studentized Breusch-Pagan test  
##   
## data: mod.1  
## BP = 1.9748, df = 2, p-value = 0.3725

Heteroscedasticity tests for model 2

plot(mod.2, which = 3) #no visible heteroscedasticity



ncvTest(mod.2) #non-significant test, no heteroscedasticity indicated

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 2.568665, Df = 1, p = 0.109

#breusch test  
bptest(mod.2) #non-significant test, no heteroscedasticity indicated

##   
## studentized Breusch-Pagan test  
##   
## data: mod.2  
## BP = 5.8436, df = 7, p-value = 0.5581

### Step 4: Multicollinearity

Multicollinearity test for model 1

vif(mod.1) #no vifs > 3

## age\_clean sex   
## 1.007797 1.007797

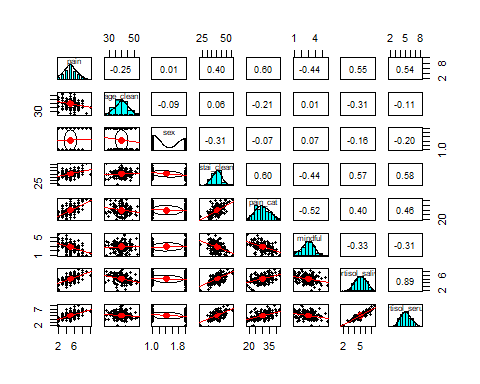
Multicollinearity test for model 2

vif(mod.2) #vifs for cortisol\_saliva and cortisol\_serum >3. The two cortisol measures correlate, not strange.

## age\_clean sex stai\_clean pain\_cat mindful   
## 1.643878 1.146737 2.529528 2.253118 1.491925   
## cortisol\_saliva cortisol\_serum   
## 7.468191 6.579508

Correlation matrix plot to check correlation between predictors in model 2, specifically cortisol\_saliva and cortisol\_serum

(plotcor <- data\_ass1\_clean %>% dplyr::select(pain, age\_clean, sex, stai\_clean, pain\_cat, mindful, cortisol\_saliva, cortisol\_serum) %>% pairs.panels(col = "red", lm = T))



## NULL

The two cortisol measures correlate at r = 0.89, which is a very high correlation.

Comparing models with both cortisol measures versus only unique measures

#looking at adjusted R^2 for mod.2, and relation between cortisol measures as predictors. When we have both of the cortisol measures as predictors, none of them are significant.   
summary(mod.2)

##   
## Call:  
## lm(formula = pain ~ age\_clean + sex + stai\_clean + pain\_cat +   
## mindful + cortisol\_saliva + cortisol\_serum, data = data\_ass1\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.50180 -0.73572 0.09629 0.72977 2.34787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.66445 1.56281 0.425 0.6713   
## age\_clean -0.01679 0.02294 -0.732 0.4654   
## sexmale 0.24914 0.19374 1.286 0.2004   
## stai\_clean -0.03646 0.02802 -1.301 0.1953   
## pain\_cat 0.12988 0.02693 4.823 3.44e-06 \*\*\*  
## mindful -0.24534 0.12029 -2.040 0.0431 \*   
## cortisol\_saliva 0.40262 0.24392 1.651 0.1009   
## cortisol\_serum 0.23284 0.22662 1.027 0.3059   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.128 on 150 degrees of freedom  
## Multiple R-squared: 0.5028, Adjusted R-squared: 0.4796   
## F-statistic: 21.67 on 7 and 150 DF, p-value: < 2.2e-16

When we have both of the cortisol measures as predictors, none of them are significant. The full model has adjusted R^2 of 0.48. Let’s compare with running summary for models with only cortisol\_saliva and cortisol\_serum as unique predictors.

#cortisol serum as predictor  
mod.serum <- lm(pain~age\_clean+sex+stai\_clean+pain\_cat+mindful+cortisol\_serum, data\_ass1\_clean)  
summary(mod.serum)

##   
## Call:  
## lm(formula = pain ~ age\_clean + sex + stai\_clean + pain\_cat +   
## mindful + cortisol\_serum, data = data\_ass1\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.57536 -0.64278 0.08773 0.68473 2.54676   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.83517 1.40053 1.310 0.1921   
## age\_clean -0.03786 0.01917 -1.974 0.0502 .   
## sexmale 0.27683 0.19411 1.426 0.1559   
## stai\_clean -0.02177 0.02672 -0.815 0.4166   
## pain\_cat 0.11553 0.02563 4.507 1.31e-05 \*\*\*  
## mindful -0.28204 0.11889 -2.372 0.0189 \*   
## cortisol\_serum 0.55836 0.11228 4.973 1.77e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.134 on 151 degrees of freedom  
## Multiple R-squared: 0.4937, Adjusted R-squared: 0.4736   
## F-statistic: 24.54 on 6 and 151 DF, p-value: < 2.2e-16

#cortisol saliva as predictor  
mod.saliv <- lm(pain~age\_clean+sex+stai\_clean+pain\_cat+mindful+cortisol\_saliva, data\_ass1\_clean)  
summary(mod.saliv)

##   
## Call:  
## lm(formula = pain ~ age\_clean + sex + stai\_clean + pain\_cat +   
## mindful + cortisol\_saliva, data = data\_ass1\_clean)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.48458 -0.75172 0.06268 0.74735 2.40932   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.115755 1.469006 0.079 0.9373   
## age\_clean -0.006779 0.020775 -0.326 0.7446   
## sexmale 0.228440 0.192723 1.185 0.2378   
## stai\_clean -0.038859 0.027931 -1.391 0.1662   
## pain\_cat 0.139007 0.025427 5.467 1.85e-07 \*\*\*  
## mindful -0.227184 0.119007 -1.909 0.0582 .   
## cortisol\_saliva 0.620719 0.120191 5.164 7.50e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.128 on 151 degrees of freedom  
## Multiple R-squared: 0.4993, Adjusted R-squared: 0.4794   
## F-statistic: 25.09 on 6 and 151 DF, p-value: < 2.2e-16

Conclusion: Adjusted R^2 for model with cortisol saliva performs close to model with cortisol serum.

#looking at collinearity of serum model, no issues  
vif(mod.serum) #vif < 3

## age\_clean sex stai\_clean pain\_cat mindful   
## 1.135194 1.138140 2.274430 2.018212 1.440958   
## cortisol\_serum   
## 1.596868

#looking at collinearity of saliva model, no issues   
vif(mod.saliv) #vif < 3

## age\_clean sex stai\_clean pain\_cat mindful   
## 1.347336 1.134331 2.511936 2.007966 1.459716   
## cortisol\_saliva   
## 1.812554

### Model comparison using AIC between model with both cortisol measures, model with only saliva, and model with only serum.

AIC(mod.2, mod.saliv, mod.serum)

## df AIC  
## mod.2 9 496.2111  
## mod.saliv 8 495.3192  
## mod.serum 8 497.0552

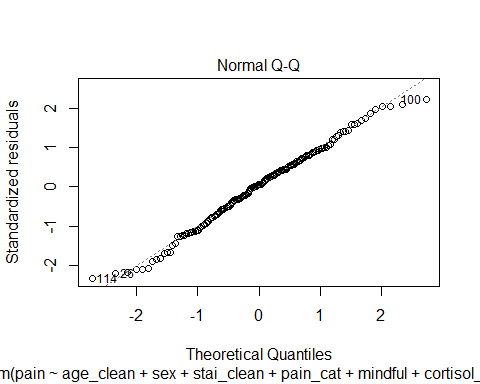
Conclusion: AIC value for model with cortisol saliva is lower than more model with cortisol serum, as well as for the full model.

Multicollinearity in full model (mod.2) results in neither of the cortisol measures being a significant predictor of pain. When I seperate the two cortisol measures, I can see that cortisol is a significant predictor for pain using either measure when they are in different models. But, cortisol saliva results in a slightly higher R^2 and lower AIC: so for future model comparisons, only cortisol saliva will be used as a predictor.

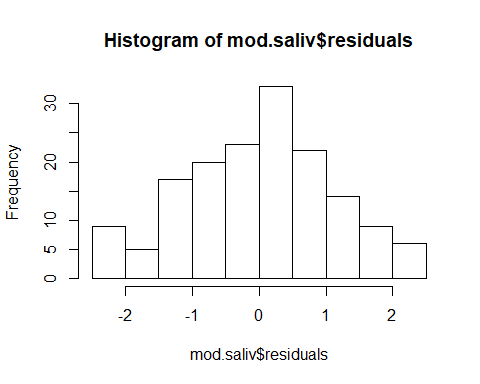
### Revisit diagnostics for mod.saliv

Diagnostics for mod.saliv (collinearity already checked) Normality of residuals:

mod.saliv%>% plot(which=2) #shows normality of residuals, 100 deviate from the line at the top, same participant as for mod.2



hist(mod.saliv$residuals) #acceptably normal distribution, some skew towards the negative but not large

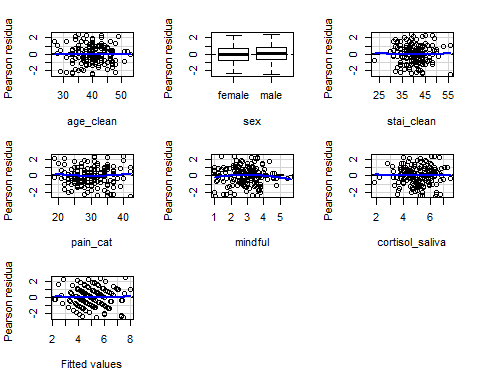


psych::describe(residuals(mod.saliv)) #skew = -0.13, kurtosis = -0.43, no noticably large values. A bit larger than for mod.2 but not concerning.

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 158 0 1.11 0.06 0.01 1.08 -2.48 2.41 4.89 -0.13 -0.43 0.09

Linearity Residual plot for mod.saliv

residualPlots(model = mod.saliv) #no large deviations from linearity



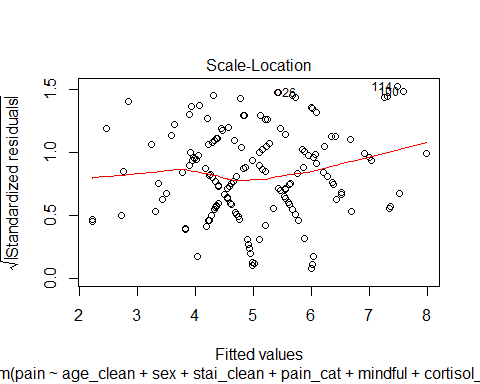
## Test stat Pr(>|Test stat|)  
## age\_clean -0.2368 0.8131  
## sex   
## stai\_clean 0.0992 0.9211  
## pain\_cat 0.7033 0.4830  
## mindful -0.9874 0.3251  
## cortisol\_saliva 0.2366 0.8133  
## Tukey test 0.2947 0.7683

Linearity for residuals looks fine

Linearity for residuals looks fine for mod.saliv, only minor changes compared to mod.2

Homoscedascicity Heteroscedasticity tests for mod.saliv

plot(mod.saliv, which = 3) #no visible heteroscedasticity



ncvTest(mod.saliv) #non-significant test, no heteroscedasticity indicated

## Non-constant Variance Score Test   
## Variance formula: ~ fitted.values   
## Chisquare = 2.499048, Df = 1, p = 0.11392

#breusch test  
bptest(mod.saliv) #non-significant test, no heteroscedasticity indicated

##   
## studentized Breusch-Pagan test  
##   
## data: mod.saliv  
## BP = 5.8491, df = 6, p-value = 0.4403

Diagnostics finished. Mod.saliv passes assumption checks.

### Model comparison, between mod.1:mod.saliv and mod.1:mod.2 (to confirm choice of predictor removal in mod.saliv)

AIC approach

AIC(mod.1, mod.2, mod.saliv)

## df AIC  
## mod.1 4 586.0217  
## mod.2 9 496.2111  
## mod.saliv 8 495.3192

AIC supports using mod.2 over mod.1, and model that contains cortisol measured using saliva has slightly lower AIC than full model.

Conclusion from AIC approach: Although mod 2 and mod.saliva explained quite similar levels of variance in the outcome variable (ca 48%) as determined by adjusted R squared, AIC indicates that using mod.saliva would be slightly preferable. Further, an additional added value is that mod-saliva also provides a more efficient model as it is built on a predictor less than in mod.2 whilst stll achieving almost the same Adjusted R^2 value.

Hierarchical regression using ANOVA

anova(mod.1, mod.saliv,mod.2 )

## Analysis of Variance Table  
##   
## Model 1: pain ~ age\_clean + sex  
## Model 2: pain ~ age\_clean + sex + stai\_clean + pain\_cat + mindful + cortisol\_saliva  
## Model 3: pain ~ age\_clean + sex + stai\_clean + pain\_cat + mindful + cortisol\_saliva +   
## cortisol\_serum  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 155 358.91   
## 2 151 192.17 4 166.744 32.7672 <2e-16 \*\*\*  
## 3 150 190.83 1 1.343 1.0556 0.3059   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Nested model using cortisol saliva adds significantly more explained variance than model 1, but the full model does not add significantly more explained variance than the model with only cortisol\_saliva.

### Model test statistics

Model statistics from ANOVA/summary  
mod.1 adjusted R^2 value = 0.05, F(2,155) = 5.37, p = .006.  
mod.saliv adjusted R^2 value = 0.48, F(6,151) = 25.09, p < .001

Loading custom function for regression results table

coef\_table = function(model){   
 require(lm.beta)   
 mod\_sum = summary(model)   
 mod\_sum\_p\_values = as.character(round(mod\_sum$coefficients[,4], 3))   
 mod\_sum\_p\_values[mod\_sum\_p\_values != "0" & mod\_sum\_p\_values != "1"] = substr(mod\_sum\_p\_values[mod\_sum\_p\_values != "0" & mod\_sum\_p\_values != "1"], 2, nchar(mod\_sum\_p\_values[mod\_sum\_p\_values != "0" & mod\_sum\_p\_values != "1"]))   
 mod\_sum\_p\_values[mod\_sum\_p\_values == "0"] = "<.001"   
   
   
 mod\_sum\_table = cbind(as.data.frame(round(cbind(coef(model), confint(model), c(0, lm.beta(model)$standardized.coefficients[c(2:length(model$coefficients))])), 2)), mod\_sum\_p\_values)   
 names(mod\_sum\_table) = c("b", "95%CI lb", "95%CI ub", "Std.Beta", "p-value")   
 mod\_sum\_table["(Intercept)","Std.Beta"] = "0"   
 return(mod\_sum\_table)   
}

Regression results table for mod.1

coef\_table(mod.1)

## b 95%CI lb 95%CI ub Std.Beta p-value  
## (Intercept) 8.23 6.27 10.19 0 <.001  
## age\_clean -0.08 -0.13 -0.03 -0.26 .001  
## sexmale -0.04 -0.52 0.45 -0.01 .886

Model interpretation for mod 1:  
Age is a significant predictor of postoperative pain, such that one unit change (year) in age corresponds to a 0.08 unit decrease in postoperative pain. For the standardised coefficients, a one standard deviation change in age (5.03 years) predicts a decrease by 0.26 of a standard deviation in postoperative pain, which is 0.41 units (SD for pain = 1.56, multiply this with standardised beta to get change in actual units). The relationship predicts a decrease in pain as age increases. Patient gender was not a significant predictor of postoperative pain.

Regression results table for mod.saliv

coef\_table(mod.saliv)

## b 95%CI lb 95%CI ub Std.Beta p-value  
## (Intercept) 0.12 -2.79 3.02 0 .937  
## age\_clean -0.01 -0.05 0.03 -0.02 .745  
## sexmale 0.23 -0.15 0.61 0.07 .238  
## stai\_clean -0.04 -0.09 0.02 -0.13 .166  
## pain\_cat 0.14 0.09 0.19 0.45 <.001  
## mindful -0.23 -0.46 0.01 -0.13 .058  
## cortisol\_saliva 0.62 0.38 0.86 0.4 <.001

Model interpretation for mod.saliv:

Age, gender, and STAI were not significant predictors of postoperative pain in this model. Mindfulness approached significance, such that when mindfulness increases by one standard deviation, postoperative pain decreases by 0.13 of a standard deviation. Pain catastrophizing was a significant predictor of postoperative pain, such that when pain catastrophizing increases by one standard deviation, postoperative pain increases by 0.45 of a standard deviation. Cortisol (measured by saliva) was a significant predictor of postoperative pain, such that when cortisol increases by one standard deviation, postoperative pain increases by 0.40 of a standard deviation.

Further, the relationships for these predictors with the outcome variables fits the theory based indication provided in the assignment description. Pain catastrophizing and cortisol have a positive relationship with the outcome variable as indicated by theory. In addition, mindfulness (as verging on being significant) has a negative relationship with the outcome variable which matches theory that has shown it may be a protective factor for postoperative pain.

#### Summary and answering the research question:

Unlike previous studies and meta-analyses, age did not present itself as a statistically significant predictor of postoperative pain in the second model tested (mod.saliv) when psychological and hormonal predictors were added.

Mod.2 and Mod.saliv performed quite closely on adjuster R^2 (ca 48% of the variance explained), yet indications were found in favor of mod.saliva on AIC. Mod.saliva also offers a model with fewer predictors.  
In essence, the study indicates that by adding psychological and hormonal predictors to the prediction model of postoperative pain, the variance explained increases. The expanded models (mod.2 and mod.saliva) including psychological and hormonal values explained around 48% of the variance in postoperative pain, while Mod.1 only explained around 5% of the variance. Further, running the analysis has expanded our understanding that pain catastrophizing and cortisol (in saliva) statistically significant explains variance in postoperative pain, demonstrating a positive relationship with the outcome wherein high levels on these predictors can be expected to relate to high levels on post-operative pain. Statistically it demonstrates a value of considering these two predictors in particular for predicting postoperative pain. In addition to this, mindfulness is worth noting as approaching statistical significance. Concluding, however, adding psychological and hormonal predictors to the model appear to have expanded our understanding of more predictors of postoperative pain, indicated new understanding concerning age as a predictor and added clear value in explaining variance in the outcome variable when adding these to our prediction model.

Regression equation written out

𝑌 = 0.12 + (-0.01 \* age\_clean) + (0.23 \* sexmale) +(-0.04 \* stai\_clean) + (0.14 \* pain\_cat)+ (-0.23 \* mindful)+ (0.62 \* cortisol\_saliva)