Data Manipulation Report

Achal Gupta

5 December 2018

### Loading Required libraries

knitr::opts\_chunk$set(echo = FALSE)  
library(lubridate)  
library(dplyr)  
library(eeptools)

### Reading CSV file and selecting required columns

Pain <- read.csv("Pain\_Full.csv")  
  
Pain <- Pain %>%   
 select (ID, Centre, Assignment, Nosessions, Gender, DOB, Abdomen, painsites, otherillness, illnessdetails, painmgt, Education, Marital, T1\_presentpain, T1\_worstpain, T1\_averagepain, T1Painintensity, bestdescribes, T1McGillTot, appetitie, T1totRolandMorris, T1Employment\_final, T2McGillTot, T2totRM, Medicalcard, Treatment)  
  
head(Pain)

## ID Centre Assignment Nosessions Gender DOB Abdomen painsites  
## 1 1 Cork 1 10 Female 01/07/1986 no 5  
## 2 2 Galway 0 15 Female 20/12/1972 yes 5  
## 3 3 Dublin 1 4 Female 11/07/1980 no 1  
## 4 4 Donegal 1 6 Female 25/09/1964 no 1  
## 5 5 Limerick 0 15 Female 20/06/1970 no 7  
## 6 6 Mayo 1 4 Female 11/01/1961 no 1  
## otherillness illnessdetails painmgt Education Marital  
## 1 yes asthma no Degree Single  
## 2 yes crohns disease no Primary Single  
## 3 no 1 no Leaving Cert Living with partner  
## 4 no no Leaving Cert Married  
## 5 no 1 no Degree Separated  
## 6 no 1 no Leaving Cert Married  
## T1\_presentpain T1\_worstpain T1\_averagepain T1Painintensity bestdescribes  
## 1 9 9 8 8.666667 Horrible  
## 2 8 10 8 8.600000 Horrible  
## 3 8 10 10 9.333333 Excruciating  
## 4 4 8 6 6.000000 Distressing  
## 5 8 10 8 8.666667 Horrible  
## 6 7 10 8 8.333333 Horrible  
## T1McGillTot appetitie T1totRolandMorris  
## 1 30 no 14  
## 2 29 yes 18  
## 3 35 yes 21  
## 4 12 no 7  
## 5 38 yes 19  
## 6 13 no 17  
## T1Employment\_final T2McGillTot T2totRM Medicalcard  
## 1 Unable to work due to pain 27 15 1  
## 2 Unable to work due to pain 26 17 1  
## 3 At home not looking for paid work 28 12 1  
## 4 Unable to work due to pain 11 5 1  
## 5 Working PT due to pain NA 20 1  
## 6 Unable to work due to pain 12 15 1  
## Treatment  
## 1 PGAP Group  
## 2 MTAU Control Group  
## 3 PGAP Group  
## 4 PGAP Group  
## 5 MTAU Control Group  
## 6 PGAP Group

### Data Cleaninig

Verify the structure of data and see if it looks good

str(Pain)

## 'data.frame': 110 obs. of 26 variables:  
## $ ID : int 1 2 3 4 5 6 8 11 14 21 ...  
## $ Centre : Factor w/ 7 levels "Cork","Donegal",..: 1 4 3 2 5 6 4 4 4 7 ...  
## $ Assignment : int 1 0 1 1 0 1 1 0 1 1 ...  
## $ Nosessions : int 10 15 4 6 15 4 10 15 2 14 ...  
## $ Gender : Factor w/ 2 levels "Female","Male": 1 1 1 1 1 1 2 1 1 1 ...  
## $ DOB : Factor w/ 110 levels "01/03/1973","01/05/1976",..: 4 77 38 92 74 37 82 50 14 80 ...  
## $ Abdomen : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 1 1 1 1 ...  
## $ painsites : int 5 5 1 1 7 1 5 2 4 6 ...  
## $ otherillness : Factor w/ 2 levels "no","yes": 2 2 1 1 1 1 1 1 1 1 ...  
## $ illnessdetails : Factor w/ 34 levels " ","?","1","asthma",..: 4 9 3 1 3 3 3 3 3 3 ...  
## $ painmgt : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Education : Factor w/ 7 levels "Degree","Junior Cert",..: 1 6 3 3 1 3 5 1 3 2 ...  
## $ Marital : Factor w/ 5 levels "Divorced","Living with partner",..: 5 5 2 3 4 3 2 3 5 3 ...  
## $ T1\_presentpain : int 9 8 8 4 8 7 8 7 4 9 ...  
## $ T1\_worstpain : int 9 10 10 8 10 10 9 10 10 10 ...  
## $ T1\_averagepain : int 8 8 10 6 8 8 7 8 7 10 ...  
## $ T1Painintensity : num 8.67 8.6 9.33 6 8.67 ...  
## $ bestdescribes : Factor w/ 4 levels "Discomforting",..: 4 4 3 2 4 4 1 4 2 2 ...  
## $ T1McGillTot : int 30 29 35 12 38 13 22 10 21 37 ...  
## $ appetitie : Factor w/ 2 levels "no","yes": 1 2 2 1 2 1 1 1 1 1 ...  
## $ T1totRolandMorris : int 14 18 21 7 19 17 17 19 11 12 ...  
## $ T1Employment\_final: Factor w/ 10 levels "At home not looking for paid work",..: 6 6 1 6 10 6 6 6 6 6 ...  
## $ T2McGillTot : int 27 26 28 11 NA 12 10 13 17 NA ...  
## $ T2totRM : int 15 17 12 5 20 15 6 15 10 NA ...  
## $ Medicalcard : int 1 1 1 1 1 1 2 2 1 NA ...  
## $ Treatment : Factor w/ 2 levels "MTAU Control Group",..: 2 1 2 2 1 2 2 1 2 2 ...

We noticed column DOB is in character format but it should be date, converting DOB to date

Pain$DOB <- dmy(Pain$DOB)

Identifying NA values

summary(Pain)

## ID Centre Assignment Nosessions   
## Min. : 1.00 Cork :20 Min. :0.0000 Min. : 1.00   
## 1st Qu.: 63.75 Donegal :21 1st Qu.:0.0000 1st Qu.: 7.50   
## Median :137.50 Dublin : 5 Median :0.0000 Median :14.00   
## Mean :141.38 Galway :37 Mean :0.4636 Mean :11.48   
## 3rd Qu.:210.75 Limerick: 8 3rd Qu.:1.0000 3rd Qu.:15.00   
## Max. :291.00 Mayo :10 Max. :1.0000 Max. :15.00   
## Sligo : 9   
## Gender DOB Abdomen painsites otherillness  
## Female:76 Min. :1948-10-10 no :103 Min. : 0.000 no :68   
## Male :34 1st Qu.:1959-12-20 yes: 7 1st Qu.: 2.000 yes:42   
## Median :1969-09-22 Median : 4.000   
## Mean :1968-07-14 Mean : 4.718   
## 3rd Qu.:1977-01-12 3rd Qu.: 6.000   
## Max. :1993-08-11 Max. :17.000   
##   
## illnessdetails painmgt Education   
## 1 :61 no :98 Degree :44   
## asthma : 6 yes:12 Junior Cert :15   
## depression : 6 Leaving Cert :25   
## : 5 Other : 6   
## high blood pressure: 2 Postgrad :10   
## underactive thyroid: 2 Primary : 6   
## (Other) :28 Trade Qualification: 4   
## Marital T1\_presentpain T1\_worstpain   
## Divorced : 3 Min. : 2.000 Min. : 6.000   
## Living with partner: 9 1st Qu.: 5.000 1st Qu.: 8.000   
## Married :60 Median : 6.000 Median : 9.000   
## Separated : 5 Mean : 6.245 Mean : 9.009   
## Single :33 3rd Qu.: 8.000 3rd Qu.:10.000   
## Max. :10.000 Max. :10.000   
##   
## T1\_averagepain T1Painintensity bestdescribes T1McGillTot   
## Min. : 1.000 Min. : 3.667 Discomforting:31 Min. : 4.00   
## 1st Qu.: 6.000 1st Qu.: 6.667 Distressing :38 1st Qu.:14.25   
## Median : 7.000 Median : 7.333 Excruciating : 8 Median :21.00   
## Mean : 7.083 Mean : 7.439 Horrible :31 Mean :21.85   
## 3rd Qu.: 8.000 3rd Qu.: 8.333 NA's : 2 3rd Qu.:29.00   
## Max. :10.000 Max. :10.000 Max. :42.00   
## NA's :1 NA's :1   
## appetitie T1totRolandMorris T1Employment\_final  
## no :91 Min. : 2.00 Unable to work due to pain :67   
## yes:19 1st Qu.:10.00 Working PT due to pain :20   
## Median :13.00 Retired due to pain :11   
## Mean :13.19 At home not looking for paid work: 4   
## 3rd Qu.:17.00 Student : 2   
## Max. :22.00 Unemployed not due to pain : 2   
## (Other) : 4   
## T2McGillTot T2totRM Medicalcard Treatment   
## Min. : 2.00 Min. : 0.00 Min. :1.00 MTAU Control Group:59   
## 1st Qu.:13.00 1st Qu.: 8.00 1st Qu.:1.00 PGAP Group :51   
## Median :19.00 Median :12.00 Median :1.00   
## Mean :19.76 Mean :11.58 Mean :1.39   
## 3rd Qu.:26.00 3rd Qu.:15.00 3rd Qu.:2.00   
## Max. :42.00 Max. :21.00 Max. :2.00   
## NA's :21 NA's :17 NA's :10

Removing all rows where McGill Score is NA at time interval T2 (T2McGillTot)

Pain <- Pain %>%  
 filter(!is.na(T2McGillTot))

Fill in average pain and pain intensity of NA rows with average values

Pain$T1\_averagepain[which(is.na(Pain$T1\_averagepain))] <- mean(Pain$T1\_averagepain, na.rm = TRUE)  
Pain$T1Painintensity[which(is.na(Pain$T1Painintensity))] <- mean(Pain$T1Painintensity, na.rm = TRUE)

Fill in 0 for MedicalCard where valur is NA

Pain$Medicalcard[which(is.na(Pain$Medicalcard))] <- 0

Verify the above processing

summary(Pain)

## ID Centre Assignment Nosessions Gender   
## Min. : 1 Cork :20 Min. :0.0000 Min. : 2.00 Female:63   
## 1st Qu.: 63 Donegal :18 1st Qu.:0.0000 1st Qu.: 6.00 Male :26   
## Median :136 Dublin : 4 Median :0.0000 Median :15.00   
## Mean :135 Galway :29 Mean :0.4045 Mean :11.37   
## 3rd Qu.:198 Limerick: 4 3rd Qu.:1.0000 3rd Qu.:15.00   
## Max. :291 Mayo : 9 Max. :1.0000 Max. :15.00   
## Sligo : 5   
## DOB Abdomen painsites otherillness  
## Min. :1948-10-10 no :82 Min. : 0.00 no :54   
## 1st Qu.:1960-02-26 yes: 7 1st Qu.: 2.00 yes:35   
## Median :1969-03-10 Median : 3.00   
## Mean :1968-07-05 Mean : 4.36   
## 3rd Qu.:1976-09-08 3rd Qu.: 6.00   
## Max. :1993-08-11 Max. :15.00   
##   
## illnessdetails painmgt Education   
## 1 :48 no :81 Degree :32   
## depression : 6 yes: 8 Junior Cert :11   
## asthma : 5 Leaving Cert :21   
## : 4 Other : 6   
## high blood pressure: 2 Postgrad : 9   
## underactive thyroid: 2 Primary : 6   
## (Other) :22 Trade Qualification: 4   
## Marital T1\_presentpain T1\_worstpain   
## Divorced : 1 Min. : 2.000 Min. : 7.000   
## Living with partner: 8 1st Qu.: 5.000 1st Qu.: 8.000   
## Married :51 Median : 6.000 Median : 9.000   
## Separated : 3 Mean : 6.135 Mean : 9.011   
## Single :26 3rd Qu.: 8.000 3rd Qu.:10.000   
## Max. :10.000 Max. :10.000   
##   
## T1\_averagepain T1Painintensity bestdescribes T1McGillTot   
## Min. : 1.000 Min. : 3.667 Discomforting:27 Min. : 4.00   
## 1st Qu.: 6.000 1st Qu.: 6.667 Distressing :28 1st Qu.:14.00   
## Median : 7.000 Median : 7.333 Excruciating : 7 Median :20.00   
## Mean : 6.943 Mean : 7.356 Horrible :25 Mean :21.03   
## 3rd Qu.: 8.000 3rd Qu.: 8.333 NA's : 2 3rd Qu.:29.00   
## Max. :10.000 Max. :10.000 Max. :41.00   
##   
## appetitie T1totRolandMorris T1Employment\_final  
## no :75 Min. : 2.00 Unable to work due to pain :55   
## yes:14 1st Qu.:10.00 Working PT due to pain :15   
## Median :13.00 Retired due to pain : 9   
## Mean :12.85 At home not looking for paid work: 4   
## 3rd Qu.:16.00 FT job on sick leave due to pain : 1   
## Max. :21.00 Redundant : 1   
## (Other) : 4   
## T2McGillTot T2totRM Medicalcard Treatment   
## Min. : 2.00 Min. : 0.00 Min. :0.000 MTAU Control Group:53   
## 1st Qu.:13.00 1st Qu.: 8.00 1st Qu.:1.000 PGAP Group :36   
## Median :19.00 Median :12.00 Median :1.000   
## Mean :19.76 Mean :11.52 Mean :1.326   
## 3rd Qu.:26.00 3rd Qu.:15.00 3rd Qu.:2.000   
## Max. :42.00 Max. :21.00 Max. :2.000   
##

### Derived COlumns

1. Calculate age from the DOB column

Pain$Age <- floor(age\_calc(Pain$DOB, units = "years"))

1. Calculate T1AvgScore And T2AvgScore

Pain$T1AverageScore <- (Pain$T1McGillTot + Pain$T1totRolandMorris)/2  
Pain$T2AverageScore <- (Pain$T2McGillTot + Pain$T2totRM)/2

1. Calculate Average Reduction Pain Score reduction

Pain$AvgReduction <- Pain$T1AverageScore - Pain$T2AverageScore

### Write output to CSV file to be imported into Minitab

write.csv(Pain, file="Pain\_Clean.csv")