codingRMD_MNK

MNK

2024-07-26

Contents

1 Packages

```r
if(!requireNamespace("haven")) install.packages("haven"); library(haven)
if(!requireNamespace("dplyr")) install.packages("dplyr"); library(dplyr)
if(!requireNamespace("visdat")) install.packages("visdat"); library(visdat)
if(!requireNamespace("naniar")) install.packages("naniar"); library(naniar)
if(!requireNamespace("psych")) install.packages("psych"); library(psych)
if(!requireNamespace("mice")) install.packages("mice"); library(mice)
if(!requireNamespace("Hmisc")) install.packages("Hmisc"); library(Hmisc)
if(!requireNamespace("knitr")) install.packages("knitr"); library(knitr)
if(!requireNamespace("kableExtra")) install.packages("kableExtra"); library(kableExtra)
if(!requireNamespace("lavaan")) install.packages("lavaan"); library(lavaan)
```

2 Loading files (and predefined transformation)

2.1 EMC data

```r
setwd("V:/Research/Dementie/Studenten/Studenten/Max/Databeheer")

dataEMC <- haven::read_sav("data_SCTQ_merged_16072024.sav") #load EMC collected data
dataEMC[dataEMC == 999] <- NA

dataEMC_copy <- dataEMC
```

2.2 UMCG data

```r
setwd("V:/Research/Dementie/Studenten/Studenten/Max/Databeheer")
dataUMCG <- haven::read_sav("data_SC_UMCG_mnk.sav") #load UMCG collected data
dataUMCG[dataUMCG == 999] <- NA

dataUMCG_copy <- dataUMCG
```

3 Data preprocessing

3.1 Pre defined transformation

```r
dataEMC$TAS20_fac1_tf <- 35 - dataEMC$TAS20_fac1_Identificeren_Gevoelens
dataUMCG$TAS20_fac1_tf <- 35 - dataUMCG$TAS20_fac1_Identificeren_Gevoelens
```

3.2 Synchronizing names and compute totals

```r
dataEMC$FP_1_6_total <- dataEMC$FP_1t6_ToM + dataEMC$FP_1t6_empathy
colnames(dataUMCG)[which(colnames(dataUMCG) == "SET_UMCG_Totaal")] <- "SET_UMCG_total"
dataEMC$SET_UMCG_Cognitief_Totaal <- dataEMC$SET_UMCG_1 +dataEMC$SET_UMCG_2 + dataEMC$SET_UMCG_3
dataEMC$SET_UMCG_Affectief_Totaal <- dataEMC$SET_UMCG_4 +dataEMC$SET_UMCG_5 + dataEMC$SET_UMCG_6
```

3.3 Creating ID values for UMCG sample

```
dataUMCG$ID <- paste0("UMCG", seq(1, nrow(dataUMCG)))
```

3.4 Data pooling

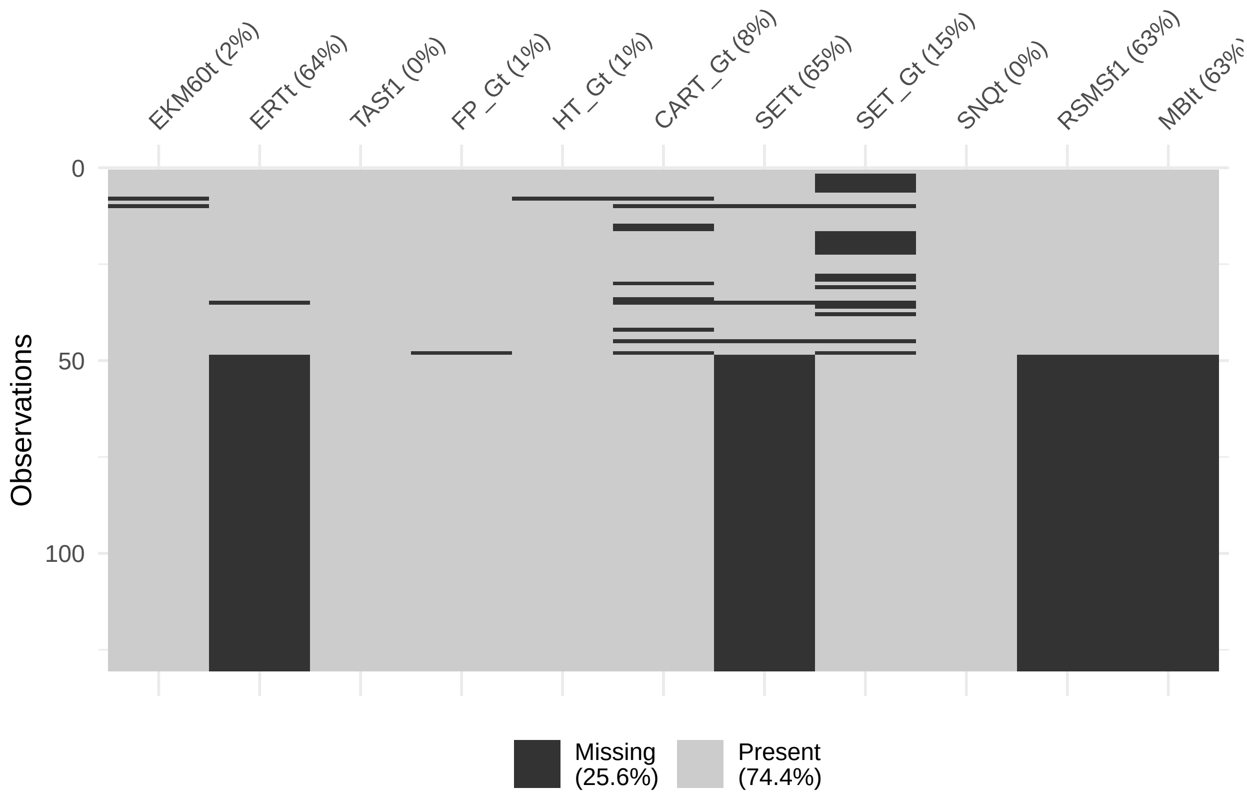
```
data_pooled <- dplyr::bind_rows(dataEMC, dataUMCG) # pooling the data
data_pooled_copy <- data_pooled
```

3.5 Renaming and var name sets

```
names_relevant_items <- c("Ekman60_total", "ERT_total", "TAS20_fac1_tf",
                          "FP_1_6_total", "hintingtask_total", "cartoons_total", "SET_total_mc", "SET_UI
                          "SNQ_NL_total", "RSMS_AMSP", "MBI_total")
names_rel_items_abbr <- c('EKM60t', 'ERTt', 'TASf1',
                          'FP_Gt', 'HT_Gt', 'CART_Gt', 'SETt', 'SET_Gt',
                          'SNQt', 'RSMSf1', 'MBIt')
names(data_pooled)[match(names_relevant_items, names(data_pooled))] <- names_rel_items_abbr #renaming
names_covariates <- c("age", "sex", "education_level")
```

3.6 Missing values part 1; visualisation

```
vis_miss(data_pooled[,names_rel_items_abbr])
```



In the above plot it is visible that we miss >=63% of data for four items (ERT, SET, RSMS, MBI). These items were not included in the data collection of UMCG, and therefor we assume that these missing values are missing completely at random (aka the characteristics of these participants did not influence whether this data is present/absent). Data for the UMCH sample will be imputed for those four items using stochastic imputation based the two variables that have the highest and significant correlations with the items.

15% of data is missing for SET_Gt, all these missings are in the EMC data. For those people, SET was conducted with another testing protocol (solely multiple choise question, no open questions). We will impute SET_Gt scores for these people based on means in the EMC sample corrected for their scores on the MC questions.

The 8% missing values for the CART_Gt variable can probably be explained by lack of time (slower participants) because it was the last test in the EMC protocol. We will impute the scores using stochastic imputation based on th two variables that have the highest and significant correlation with this test.

Imputation will be continued after checks for outliers and non-normality.

4 Data checks

4.1 EMC sample data checks

Descriptes before outlier deletion and/or tranformations

```
# Descriptives EMC start
psych::describe(data_pooled[data_pooled$centerID == "EMC", c(names_covariates, "MoCA_total", names_rel_
  select(n, min, max, mean, median, sd, skew, kurtosis)
```

```
##                    n  min max  mean median    sd  skew kurtosis
## age               48 21.0  79 48.15   54.5 19.48 -0.13    -1.50
## sex               48  0.0   1  0.67    1.0  0.48 -0.69    -1.56
## education_level 48  2.0   7  5.69    6.0  1.11 -1.01     1.13
## MoCA_total        48 19.0  30 25.94   26.0  2.62 -0.69     0.04
## EKM60t            46 34.0  58 47.00   47.5  5.37 -0.49     0.09
## ERTt              47 34.0  76 56.74   57.0  9.74 -0.16    -0.68
## TASf1             48  5.0  28 20.04   21.0  5.42 -0.60    -0.05
## FP_Gt             47 10.5  24 20.40   21.0  2.48 -1.29     3.17
## HT_Gt             47  6.0  12 11.21   12.0  1.12 -2.31     7.58
## CART_Gt           38  2.5  12  8.00    8.0  2.63 -0.30    -0.89
## SETt              45 13.0  18 16.89   17.0  1.39 -1.36     1.02
## SET_Gt            28  9.0  12 11.46   12.0  0.79 -1.39     1.33
## SNQt              48 11.0  21 19.08   19.0  1.64 -2.53    10.04
## RSMSf1            48 11.0  32 22.60   23.0  4.73 -0.50    -0.21
## MBIt              48 42.0  70 54.44   55.0  7.16  0.16    -0.74
```

4.1.1 Outliers - EMC

```r
dataEMC_subset_scaled <- cbind(data_pooled[data_pooled$centerID == "EMC",][, c("ID", names_covariates,
moca_below26_EMC <- dataEMC_subset_scaled$MoCA_total < 26

outlierCheck_EMC <- data.frame(cbind(moca_below26_EMC, abs(dataEMC_subset_scaled[,names_rel_items_abbr]
#outlierCheck_EMC <- cbind(moca_below26_EMC, sapply(dataEMC_subset_scaled[,names_rel_items_abbr], FUN =

rowSums(outlierCheck_EMC, na.rm = TRUE)
```

```
##  [1] 0 1 1 0 0 1 0 0 2 1 0 0 0 0 0 1 1 1 0 0 1 0 0 0 0 1 0 1 0 0 0 0 1 1 1 0 0 1
## [39] 1 0 0 1 0 0 1 0 1 2
```

```r
meanitemSS_EMC <- rowMeans(dataEMC_subset_scaled[,c(
                          'EKM60t', 'ERTt', 'TASf1',
                          'FP_Gt', 'CART_Gt', 'SETt',
                          'SNQt', 'RSMSf1', 'MBIt')], na.rm = TRUE)
SSmeanitemSS_EMC <- scale(meanitemSS_EMC)
outlier_EMC <- dataEMC_subset_scaled[which(abs(SSmeanitemSS_EMC)>3), "ID"]
length(outlier_EMC)
```

```
## [1] 1
```

We have one participant that has an absolute scaled mean item scaled score of over 3, we will exclude this participant from analyses.

```r
data_pooled <- data_pooled[-(which(data_pooled$ID==outlier_EMC)),]
```

4.1.2 Normality - EMC

```
psych::describe(data_pooled[data_pooled$centerID == "EMC", c(names_covariates, "MoCA_total", names_rel_
  select(n, min, max, mean, median, sd, skew, kurtosis)
```

```
##                  n  min max  mean median    sd  skew kurtosis
## age             47 21.0  79 47.64   54.0 19.37 -0.09    -1.49
## sex             47  0.0   1  0.66    1.0  0.48 -0.65    -1.61
## education_level 47  2.0   7  5.72    6.0  1.10 -1.10     1.49
## MoCA_total      47 19.0  30 26.09   26.0  2.44 -0.56    -0.15
## EKM60t          45 34.0  58 47.29   48.0  5.06 -0.39     0.12
## ERTt            46 34.0  76 57.11   57.5  9.52 -0.17    -0.59
## TASf1           47  5.0  28 20.04   21.0  5.48 -0.59    -0.11
## FP_Gt           47 10.5  24 20.40   21.0  2.48 -1.29     3.17
## HT_Gt           46  6.0  12 11.26   12.0  1.08 -2.56     9.46
## CART_Gt         38  2.5  12  8.00    8.0  2.63 -0.30    -0.89
## SETt            44 13.0  18 16.98   17.0  1.27 -1.37     1.22
## SET_Gt          28  9.0  12 11.46   12.0  0.79 -1.39     1.33
## SNQt            47 16.0  21 19.26   19.0  1.13 -0.50    -0.18
## RSMSf1          47 11.0  32 22.74   23.0  4.67 -0.56    -0.05
## MBIt            47 42.0  70 54.30   55.0  7.17  0.20    -0.71
```

```
normalityCheck_EMC <- psych::describe(data_pooled[data_pooled$centerID == "EMC", c(names_covariates, "Mo
  select(skew, kurtosis)

any(abs(normalityCheck_EMC$skew)>3)
```

```
## [1] FALSE
```

```
any(abs(normalityCheck_EMC$kurtosis)>10)
```

```
## [1] FALSE
```

No absolute skew>3 and no absolute kurtosis>10 is EMC sample after deleting 1 outlier.

4.1.3 Imputing the SET_Gt variable some EMC missings due to different testing protocol

```
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_2), "SET_open_2"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_2"], na.rm = TRUE))
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_3), "SET_open_3"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_3"], na.rm = TRUE))
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_6), "SET_open_6"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_6"], na.rm = TRUE))
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_14), "SET_open_14"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_14"], na.rm = TRUE))
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_16), "SET_open_16"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_16"], na.rm = TRUE))
data_pooled[data_pooled$centerID == "EMC" & is.na(data_pooled$SET_open_17), "SET_open_17"] <-
  as.numeric(colMeans(data_pooled[data_pooled$centerID == "EMC", "SET_open_17"], na.rm = TRUE))

data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_1"] <-
```

```
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_2 + data_pooled[data_pooled$centerID == "EMC",]$S
data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_2"] <-
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_3 + data_pooled[data_pooled$centerID == "EMC",]$S
data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_3"] <-
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_6 + data_pooled[data_pooled$centerID == "EMC",]$S
data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_4"] <-
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_14 + data_pooled[data_pooled$centerID == "EMC",]$
data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_5"] <-
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_16 + data_pooled[data_pooled$centerID == "EMC",]$
data_pooled[data_pooled$centerID == "EMC", "SET_UMCG_6"] <-
  data_pooled[data_pooled$centerID == "EMC",]$SET_open_17 + data_pooled[data_pooled$centerID == "EMC",]$

data_pooled$SET_Gt <- rowSums(data_pooled[,c(which(names(data_pooled)=="SET_UMCG_1") : which(names(data_
```

4.2 UMCG sample data checks

Descriptives before outlier deletion and/or tranformations

```
# Descriptives UMCG start
psych::describe(data_pooled[data_pooled$centerID == "UMCG", c(names_covariates, "MoCA_total", names_rel_
  select(n, min, max, mean, median, sd, skew, kurtosis)
```

```
##                  n min  max  mean median    sd  skew kurtosis
## age             82  18   66 28.71     22 14.07  1.36     0.16
## sex             82   0    1  0.63      1  0.48 -0.55    -1.72
## education_level 82   3    7  6.00      6  0.61 -1.30     6.07
## MoCA_total       0 Inf -Inf   NaN     NA    NA    NA       NA
## EKM60t          82  38   56 46.99     48  4.52 -0.26    -0.88
## ERTt             0 Inf -Inf   NaN     NA    NA    NA       NA
## TASf1           82   7   31 21.09     21  4.62 -0.68     0.53
## FP_Gt           82  13   24 20.17     21  2.51 -0.81     0.21
## HT_Gt           82   6   12 11.40     12  1.00 -2.66     9.62
## CART_Gt         82   0   12  8.51      9  2.51 -0.49    -0.05
## SETt             0 Inf -Inf   NaN     NA    NA    NA       NA
## SET_Gt          82   8   12 11.33     12  0.94 -1.38     1.28
## SNQt            82  16   22 19.38     19  1.27 -0.40    -0.21
## RSMSf1           0 Inf -Inf   NaN     NA    NA    NA       NA
## MBIt             0 Inf -Inf   NaN     NA    NA    NA       NA
```

4.2.1 Outliers - UMCG

```
dataUMCG_subset_scaled <- cbind(data_pooled[data_pooled$centerID == "UMCG",][, c("ID", names_covariates
moca_below26_UMCG <- dataUMCG_subset_scaled$MoCA_total < 26

outlierCheck_UMCG <- data.frame(cbind(moca_below26_UMCG, abs(dataUMCG_subset_scaled[,names_rel_items_abb
#outlierCheck_UMCG <- cbind(moca_below26_UMCG, sapply(dataUMCG_subset_scaled[,names_rel_items_abbr], FU

rowSums(outlierCheck_UMCG, na.rm = TRUE)
```

```
##  [1] 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0
```

```
## [39] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [77] 0 0 0 0 1 0
```

```
meanitemSS_UMCG <- rowMeans(dataUMCG_subset_scaled[,c(
                          'EKM60t', 'ERTt', 'TASf1',
                          'FP_Gt', 'CART_Gt', 'SET_Gt',
                          'SNQt', 'RSMSf1', 'MBIt')], na.rm = TRUE)
SSmeanitemSS_UMCG <- scale(meanitemSS_UMCG)
outlier_UMCG <- dataUMCG_subset_scaled[which(abs(SSmeanitemSS_UMCG)>3), "ID"]
length(outlier_UMCG)
```

```
## [1] 1
```

We have one participant that has an absolute scaled mean item scaled score of over 3, we will exclude this participant from analyses.

```
data_pooled <- data_pooled[-(which(data_pooled$ID==outlier_UMCG)),]
```

4.2.2 Normality - UMCG

```
psych::describe(data_pooled[data_pooled$centerID == "UMCG", c(names_covariates, "MoCA_total", names_rel
  select(n, min, max, mean, median, sd, skew, kurtosis)
```

```
##                  n min  max  mean median    sd  skew kurtosis
## age             81  18   66 28.70     22 14.16  1.35     0.12
## sex             81   0    1  0.63      1  0.49 -0.53    -1.74
## education_level 81   3    7  6.01      6  0.60 -1.36     6.61
## MoCA_total       0 Inf -Inf   NaN     NA    NA    NA       NA
## EKM60t          81  38   56 47.06     48  4.50 -0.29    -0.83
## ERTt             0 Inf -Inf   NaN     NA    NA    NA       NA
## TASf1           81   7   31 21.20     21  4.54 -0.70     0.72
## FP_Gt           81  14   24 20.26     21  2.40 -0.72     0.03
## HT_Gt           81   6   12 11.40     12  1.01 -2.64     9.49
## CART_Gt         81   0   12  8.55      9  2.49 -0.53     0.05
## SETt             0 Inf -Inf   NaN     NA    NA    NA       NA
## SET_Gt          81   8   12 11.35     12  0.94 -1.44     1.50
## SNQt            81  16   22 19.41     19  1.25 -0.41    -0.12
## RSMSf1           0 Inf -Inf   NaN     NA    NA    NA       NA
## MBIt             0 Inf -Inf   NaN     NA    NA    NA       NA
```

```
normalityCheck_UMCG <- psych::describe(data_pooled[data_pooled$centerID == "UMCG", c(names_covariates,
  select(skew, kurtosis)

any(abs(normalityCheck_UMCG$skew)>3, na.rm = TRUE)
```

```
## [1] FALSE
```

```
any(abs(normalityCheck_UMCG$kurtosis)>10, na.rm = TRUE)
```

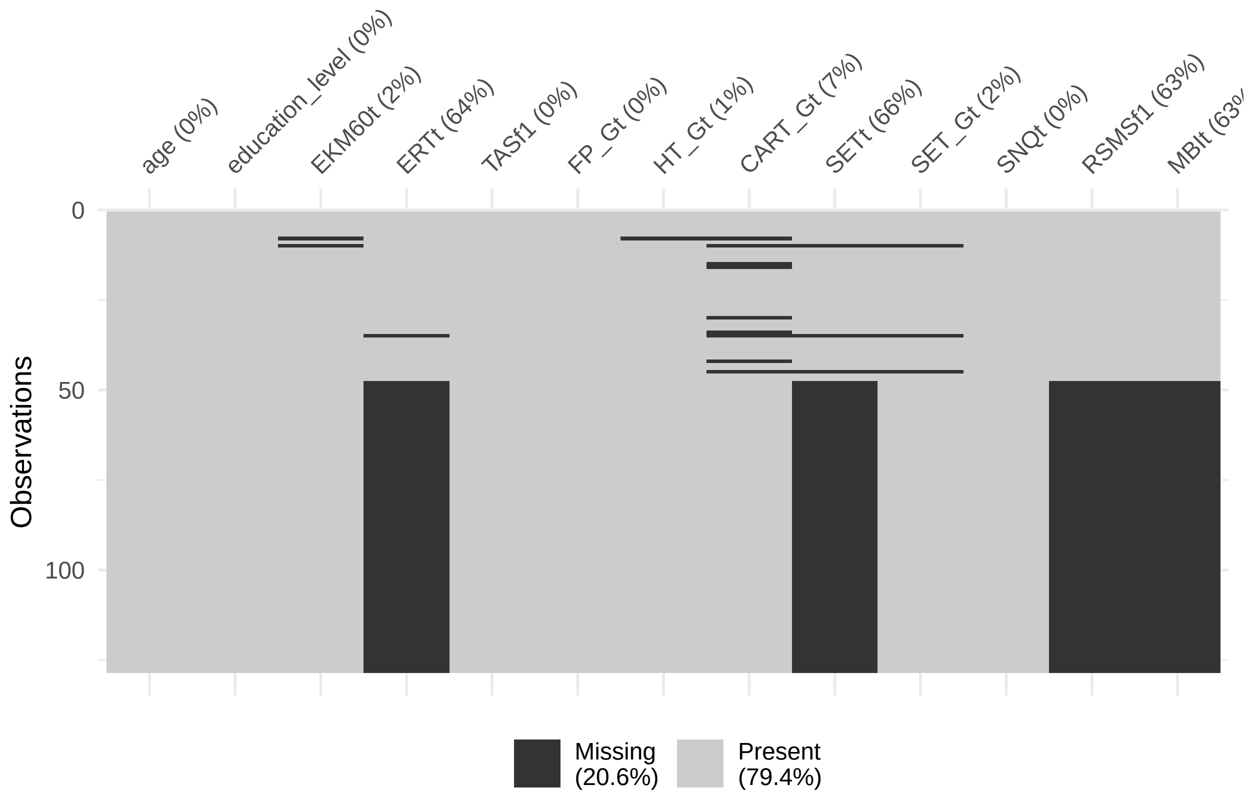
```
## [1] FALSE
```

No absolute skew>3 and no absolute kurtosis>10 is EMC sample after deleting 1 outlier.

5 Stochastic imputation

At this point, we have the following missing values.

```r
vis_miss(data_pooled[,c("age", "education_level",names_rel_items_abbr)])
```



And the following zero order paired correlation (respectively: corr, N, p-values).

```r
rcorr(as.matrix(data_pooled[, c("age", "education_level",
                                'EKM60t', 'ERTt', 'TASf1',
                                'FP_Gt', 'CART_Gt', 'SET_Gt',
                                'SNQt', 'RSMSf1', 'MBIt')]))
```

```
##                   age education_level EKM60t  ERTt TASf1 FP_Gt CART_Gt SET_Gt
## age              1.00           -0.28  -0.06 -0.42  0.04 -0.10    0.02  -0.20
## education_level -0.28            1.00   0.18  0.45  0.09  0.14    0.34   0.18
## EKM60t          -0.06            0.18   1.00  0.64  0.17  0.30    0.24   0.12
## ERTt            -0.42            0.45   0.64  1.00  0.17  0.35    0.25   0.35
## TASf1            0.04            0.09   0.17  0.17  1.00  0.11    0.02   0.04
## FP_Gt           -0.10            0.14   0.30  0.35  0.11  1.00    0.22   0.18
## CART_Gt          0.02            0.34   0.24  0.25  0.02  0.22    1.00   0.19
## SET_Gt          -0.20            0.18   0.12  0.35  0.04  0.18    0.19   1.00
## SNQt            -0.09            0.26   0.10  0.14  0.06  0.23    0.13   0.10
## RSMSf1          -0.07            0.00   0.29  0.09  0.35  0.24    0.29   0.13
## MBIt             0.33           -0.04   0.13  0.18  0.22  0.01    0.32   0.07
##                  SNQt RSMSf1  MBIt
```

```
## age                -0.09  -0.07  0.33
## education_level  0.26   0.00 -0.04
## EKM60t           0.10   0.29  0.13
## ERTt             0.14   0.09  0.18
## TASf1            0.06   0.35  0.22
## FP_Gt            0.23   0.24  0.01
## CART_Gt          0.13   0.29  0.32
## SET_Gt           0.10   0.13  0.07
## SNQt             1.00  -0.14  0.05
## RSMSf1          -0.14   1.00 -0.09
## MBIt             0.05  -0.09  1.00
##
## n
##                 age education_level EKM60t ERTt TASf1 FP_Gt CART_Gt SET_Gt SNQt
## age             128             128    126   46   128   128     119    125  128
## education_level 128             128    126   46   128   128     119    125  128
## EKM60t          126             126    126   44   126   126     119    124  126
## ERTt             46              46     44   46    46    46      38     44   46
## TASf1           128             128    126   46   128   128     119    125  128
## FP_Gt           128             128    126   46   128   128     119    125  128
## CART_Gt         119             119    119   38   119   119     119    119  119
## SET_Gt          125             125    124   44   125   125     119    125  125
## SNQt            128             128    126   46   128   128     119    125  128
## RSMSf1           47              47     45   46    47    47      38     44   47
## MBIt             47              47     45   46    47    47      38     44   47
##                 RSMSf1 MBIt
## age                 47   47
## education_level     47   47
## EKM60t              45   45
## ERTt                46   46
## TASf1               47   47
## FP_Gt               47   47
## CART_Gt             38   38
## SET_Gt              44   44
## SNQt                47   47
## RSMSf1              47   47
## MBIt                47   47
##
## P
##                 age    education_level EKM60t ERTt   TASf1  FP_Gt  CART_Gt
## age                    0.0015          0.5077 0.0036 0.6536 0.2670 0.8348
## education_level 0.0015                 0.0425 0.0016 0.2938 0.1080 0.0002
## EKM60t          0.5077 0.0425                 0.0000 0.0588 0.0007 0.0087
## ERTt            0.0036 0.0016          0.0000        0.2556 0.0182 0.1263
## TASf1           0.6536 0.2938          0.0588 0.2556        0.2054 0.7995
## FP_Gt           0.2670 0.1080          0.0007 0.0182 0.2054        0.0168
## CART_Gt         0.8348 0.0002          0.0087 0.1263 0.7995 0.0168
## SET_Gt          0.0269 0.0509          0.1999 0.0217 0.6607 0.0420 0.0409
## SNQt            0.2871 0.0028          0.2846 0.3457 0.4734 0.0107 0.1681
## RSMSf1          0.6426 0.9846          0.0516 0.5463 0.0161 0.1084 0.0797
## MBIt            0.0233 0.7803          0.3871 0.2414 0.1332 0.9620 0.0490
##                 SET_Gt SNQt   RSMSf1 MBIt
## age             0.0269 0.2871 0.6426 0.0233
## education_level 0.0509 0.0028 0.9846 0.7803
```

```
## EKM60t          0.1999 0.2846 0.0516 0.3871
## ERTt            0.0217 0.3457 0.5463 0.2414
## TASf1           0.6607 0.4734 0.0161 0.1332
## FP_Gt           0.0420 0.0107 0.1084 0.9620
## CART_Gt         0.0409 0.1681 0.0797 0.0490
## SET_Gt                 0.2611 0.4116 0.6504
## SNQt            0.2611        0.3499 0.7418
## RSMSf1          0.4116 0.3499        0.5658
## MBIt            0.6504 0.7418 0.5658
```

5.1 Imputation predictor selection

5.1.1 Predictor selection ERTt

Lets check the correlations of ERTt with other relevant variables.

```r
matcor <- Hmisc:: rcorr(as.matrix(data_pooled[, c("age", "education_level",
                                        'EKM60t', 'ERTt', 'TASf1',
                                        'FP_Gt', 'HT_Gt', 'CART_Gt', 'SET_Gt',
                                        'SNQt', 'RSMSf1', 'MBIt')]))

#Check whether sex is also a significant predictor
sexERTt <- lm(ERTt ~ sex, data = data_pooled)
summary(sexERTt)$coefficients["sex", "Pr(>|t|)"] #Not significant
```

```
## [1] 0.1775048
```

```r
sigpred_ERTt <- matcor$r["ERTt",matcor$P["ERTt",]<.05]; sigpred_ERTt #significant correlations of ERTt
```

```
##             age education_level          EKM60t            <NA>           FP_Gt
##      -0.4211357       0.4526528       0.6353631              NA       0.3468226
##          SET_Gt
##       0.3452348
```

```r
# corrplot(matcor$r, p.mat = matcor$P, sig.level = 0.05, method = "number", type = "lower", tl.col = "b
#           #insig = "blank",
#           addCoef.col = "black")
```

We will impute the missing ERTt values using predictive mean matching (stochastic) based on the significant predictors.

5.1.2 Predictor selection RSMSf1

Lets check the correlations of RSMSf1 with other relevant variables.

```r
#Check whether sex is also a significant predictor
sexRSMSf1 <- lm(RSMSf1 ~ sex, data = data_pooled)
summary(sexRSMSf1)$coefficients["sex", "Pr(>|t|)"] #Not significant
```

```
## [1] 0.08581445
```

```r
#Check numerical correlations
sigpred_RSMSf1 <- matcor$r["RSMSf1",matcor$P["RSMSf1",]<.05]; sigpred_RSMSf1 #significant correlations
```

```
##     TASf1      <NA>
## 0.3494505        NA
```

5.1.3 Predictor selection MBIt

Lets check the correlations of MBIt with other relevant variables.

```r
#Check whether sex is also a significant predictor
sexMBIt <- lm(MBIt ~ sex, data = data_pooled)
summary(sexMBIt)$coefficients["sex", "Pr(>|t|)"] #Not significant
```

```
## [1] 0.33373
```

```r
#Check numerical correlations
sigpred_MBIt <- matcor$r["MBIt",matcor$P["MBIt",]<.05]; sigpred_MBIt #significant correlations of MBIt
```

```
##       age    CART_Gt      <NA>
## 0.3305121 0.3216157        NA
```

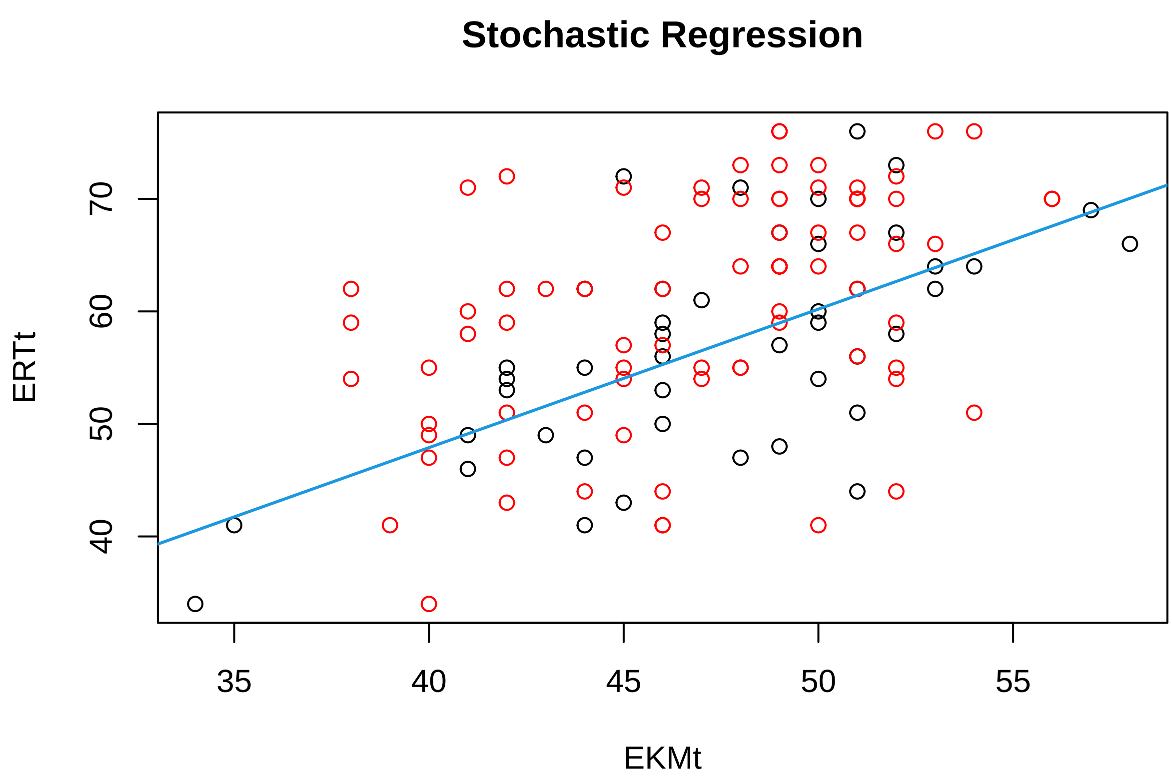
5.2 Imputation of ERTt, RSMSf, MBIt

```r
data_pooled_subset <- data_pooled[,c("ID", "centerID", "sex", "age", "education_level",
                                      'EKM60t', 'ERTt', 'TASf1',
                                      'FP_Gt', 'CART_Gt', 'SET_Gt',
                                      'SNQt', 'RSMSf1', 'MBIt')]

predictor_matrix <- make.predictorMatrix(data_pooled_subset)
predictor_matrix[] <- 0

#assign the significant predictors to the prediction matrix
predictor_matrix["ERTt", ] <- colnames(predictor_matrix) %in% names(sigpred_ERTt)[!is.na(names(sigpred_
predictor_matrix["RSMSf1", ] <- colnames(predictor_matrix) %in% names(sigpred_RSMSf1)[!is.na(names(sigp
predictor_matrix["MBIt", ] <- colnames(predictor_matrix) %in% names(sigpred_MBIt)[!is.na(names(sigpred_

#assign imputation medhods
imputation_methods <- make.method(data_pooled_subset)
imputation_methods[] <- ""  # Set all methods to "" initially
imputation_methods[c("ERTt", "RSMSf1", "MBIt")] <- "pmm"  # Use "pmm" for ERTt

#imputation
data_pooled_subset_imp <- complete(mice(data_pooled_subset,
                                method = imputation_methods,
                                predictorMatrix = predictor_matrix,
                                m = 1,
                                seed = 42), 1)
```

```
##
##  iter imp variable
##   1   1  ERTt  RSMSf1  MBIt
##   2   1  ERTt  RSMSf1  MBIt
##   3   1  ERTt  RSMSf1  MBIt
##   4   1  ERTt  RSMSf1  MBIt
##   5   1  ERTt  RSMSf1  MBIt
```

5.2.1 Visualisation ERTt impuations

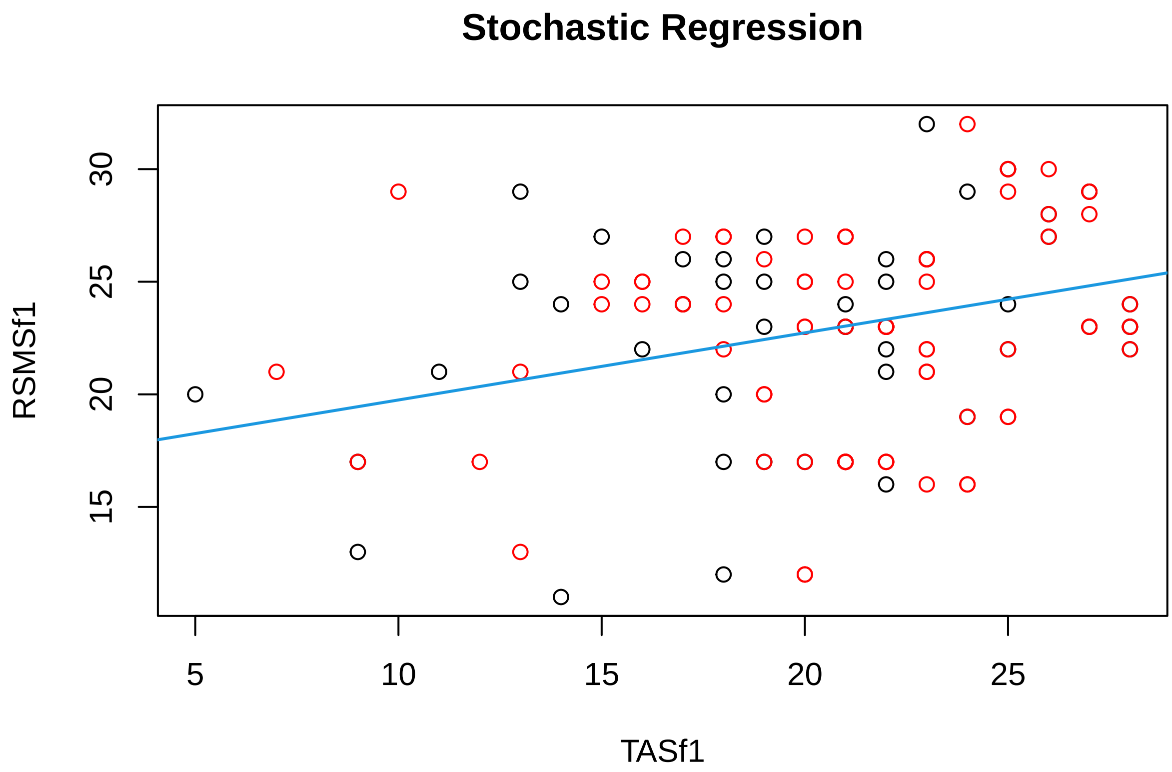
```
# Stochastic regression imputation plot for ERTt
plot(data_pooled_subset$EKM60t[!is.na(data_pooled_subset$ERTt)], data_pooled_subset_imp$ERTt[!is.na(data
     main = "Stochastic Regression",
     xlab = "EKMt", ylab = "ERTt")
points(data_pooled_subset$EKM60t[is.na(data_pooled_subset$ERTt)], data_pooled_subset_imp$ERTt[is.na(data
       col = "red")
abline(lm(data_pooled_subset$ERTt ~ data_pooled_subset$EKM60t, data_pooled_subset_imp), col = "#1b98e0"
```



5.2.2 Visualisation RSMSf1 impuations

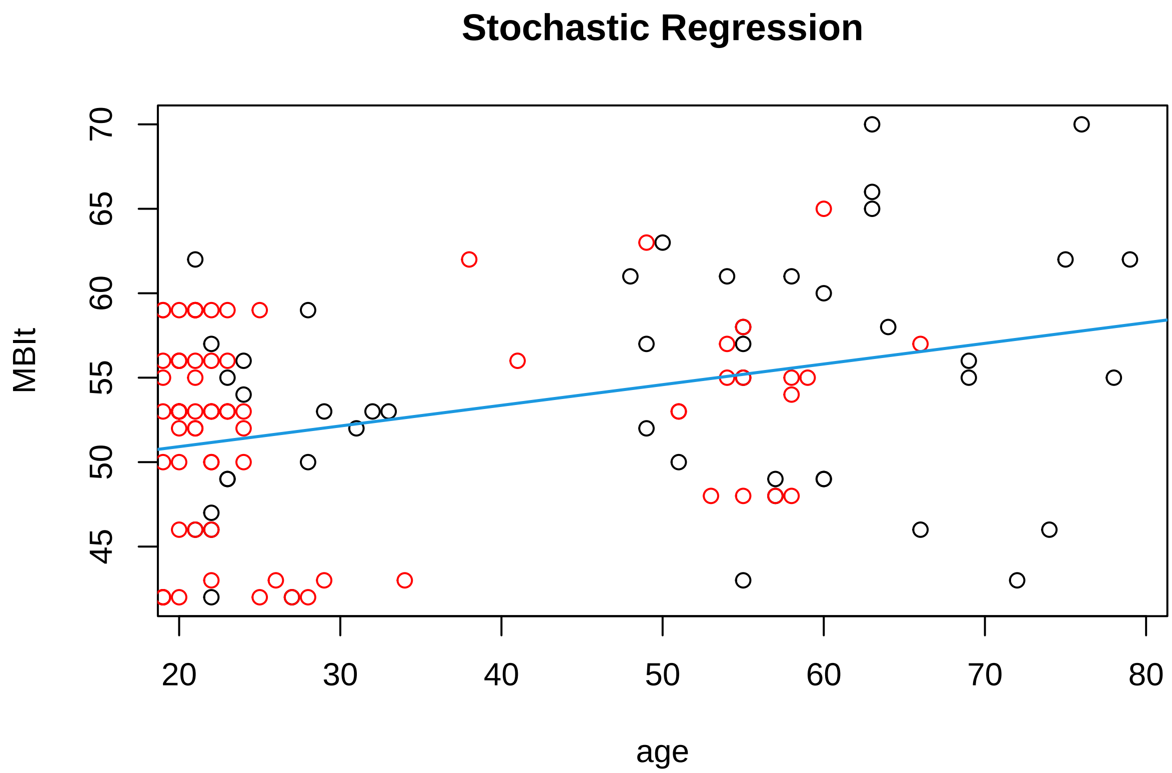
```
# Stochastic regression imputation plot for RSMSf1
plot(data_pooled_subset$TASf1[!is.na(data_pooled_subset$RSMSf1)], data_pooled_subset_imp$RSMSf1[!is.na(d
```

```
     main = "Stochastic Regression",
     xlab = "TASf1", ylab = "RSMSf1")
points(data_pooled_subset$TASf1[is.na(data_pooled_subset$RSMSf1)], data_pooled_subset_imp$RSMSf1[is.na(d
       col = "red")
abline(lm(data_pooled_subset$RSMSf1 ~ data_pooled_subset$TASf1, data_pooled_subset_imp), col = "#1b98e0
```



5.2.3 Visualisation MBIt impuations

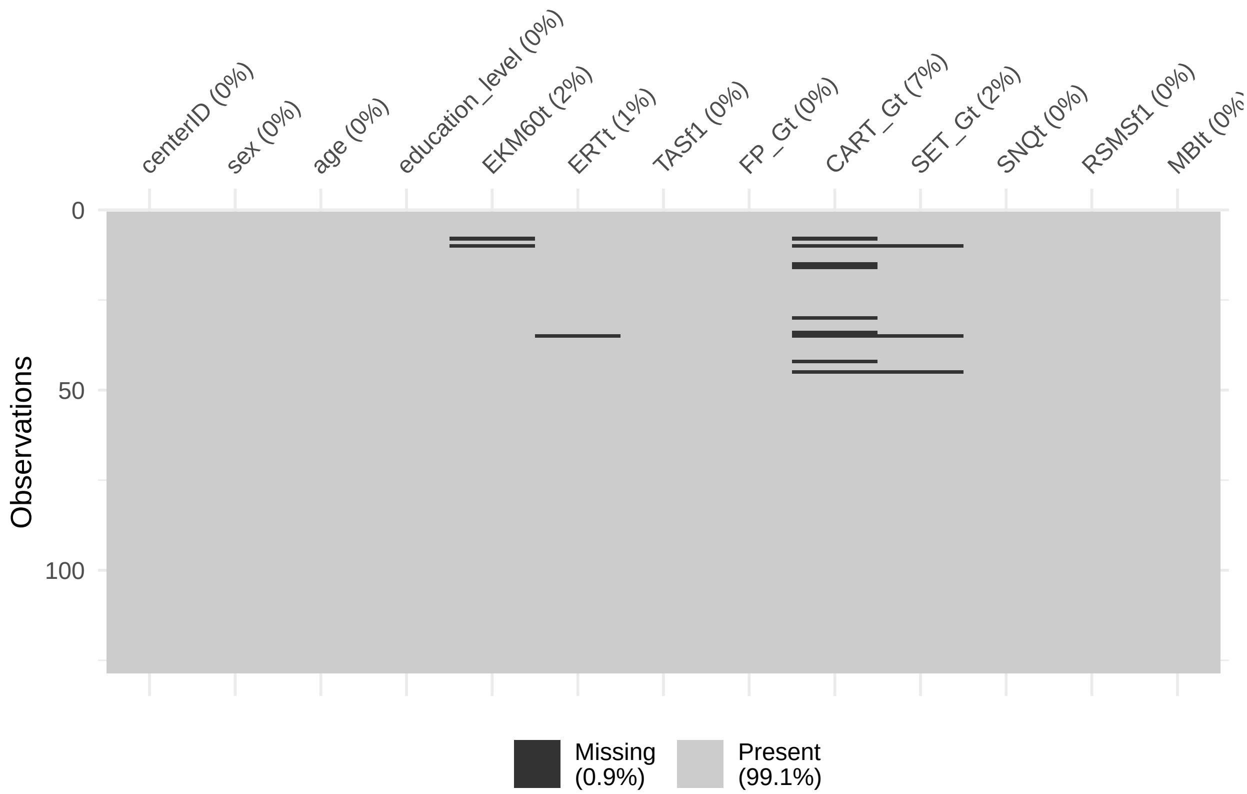
```
# Stochastic regression imputation plot for MBIt
plot(data_pooled_subset$age[!is.na(data_pooled_subset$MBIt)], data_pooled_subset_imp$MBIt[!is.na(data_po
     main = "Stochastic Regression",
     xlab = "age", ylab = "MBIt")
points(data_pooled_subset$age[is.na(data_pooled_subset$MBIt)], data_pooled_subset_imp$MBIt[is.na(data_po
       col = "red")
abline(lm(data_pooled_subset$MBIt ~ data_pooled_subset$age, data_pooled_subset_imp), col = "#1b98e0", lw
```



5.3 Mid term missing data evaluation

At this point, we have the following missing data:

```
vis_miss(data_pooled_subset_imp[,-1])
```



We will impute the remaining missing values, again with a stochastic approach based on significant correlations or significant t.test for sex.

5.4 Imputation predictor selection part 2

5.4.1 Predictor selection CART_Gt

Lets check the correlations of CART_Gt with other relevant variables.

```
matcor2 <- Hmisc:: rcorr(as.matrix(data_pooled_subset_imp[, c("age", "education_level",
                                        'EKM60t', 'ERTt', 'TASf1',
                                        'FP_Gt', 'CART_Gt', 'SET_Gt',
                                        'SNQt', 'RSMSf1', 'MBIt')]))

#Check whether sex is also a significant predictor
sexCART_Gt <- lm(CART_Gt ~ sex, data = data_pooled_subset_imp)
summary(sexCART_Gt)$coefficients["sex", "Pr(>|t|)"] #Significant
```

```
## [1] 0.0005648991
```

```
sigpred_CART_Gt <- matcor2$r["CART_Gt",matcor2$P["CART_Gt",]<.05]; sigpred_CART_Gt #significant correla
```

```
## education_level          EKM60t            ERTt           FP_Gt            <NA>
##       0.3372836       0.2394804       0.2402233       0.2187845              NA
##          SET_Gt          RSMSf1
##       0.1877775       0.2021467
```

```r
sigpred_CART_Gt$sex <- summary(sexCART_Gt)$coefficients["sex", "Pr(>|t|)"] #Significant
```

```
## Warning in sigpred_CART_Gt$sex <- summary(sexCART_Gt)$coefficients["sex", :
## Coercing LHS to a list
```

5.4.2 Predictor selection SET_Gt

Lets check the correlations of SET_Gt with other relevant variables.

```r
#Check whether sex is also a significant predictor
sexSET_Gt <- lm(SET_Gt ~ sex, data = data_pooled_subset_imp)
summary(sexSET_Gt)$coefficients["sex", "Pr(>|t|)"] #Not ignificant
```

```
## [1] 0.1354767
```

```r
sigpred_SET_Gt <- matcor2$r["SET_Gt",matcor2$P["SET_Gt",]<.05]; sigpred_SET_Gt #significant correlation
```

```
##        age       ERTt      FP_Gt    CART_Gt       <NA>
## -0.1980193  0.5010054  0.1821579  0.1877775         NA
```

5.4.3 Predictor selection EKM60t

Lets check the correlations of EKM60t with other relevant variables.

```r
#Check whether sex is also a significant predictor
sexEKM60t <- lm(EKM60t ~ sex, data = data_pooled_subset_imp)
summary(sexEKM60t)$coefficients["sex", "Pr(>|t|)"] #Not significant
```

```
## [1] 0.2449073
```

```r
sigpred_EKM60t <- matcor2$r["EKM60t",matcor2$P["EKM60t",]<.05]; sigpred_EKM60t #significant correlation
```

```
## education_level            <NA>            ERTt           FP_Gt         CART_Gt
##       0.1810381              NA       0.5057871       0.2988278       0.2394804
##          RSMSf1
##       0.1924306
```

5.4.4 Predictor selection ERTt (second round)

Lets check the correlations of ERTt with other relevant variables.

```r
#Check whether sex is also a significant predictor
sexERTt_2nd <- lm(ERTt ~ sex, data = data_pooled_subset_imp)
summary(sexERTt_2nd)$coefficients["sex", "Pr(>|t|)"] #Not significant
```

```
## [1] 0.7873681
```

```r
sigpred_ERTt_2nd <- matcor2$r["ERTt",matcor2$P["ERTt",]<.05]; sigpred_ERTt_2nd #significant correlation
```

```
##             age education_level          EKM60t            <NA>           FP_Gt 
##      -0.3202251       0.2671958       0.5057871              NA       0.4173242 
##         CART_Gt          SET_Gt 
##       0.2402233       0.5010054
```

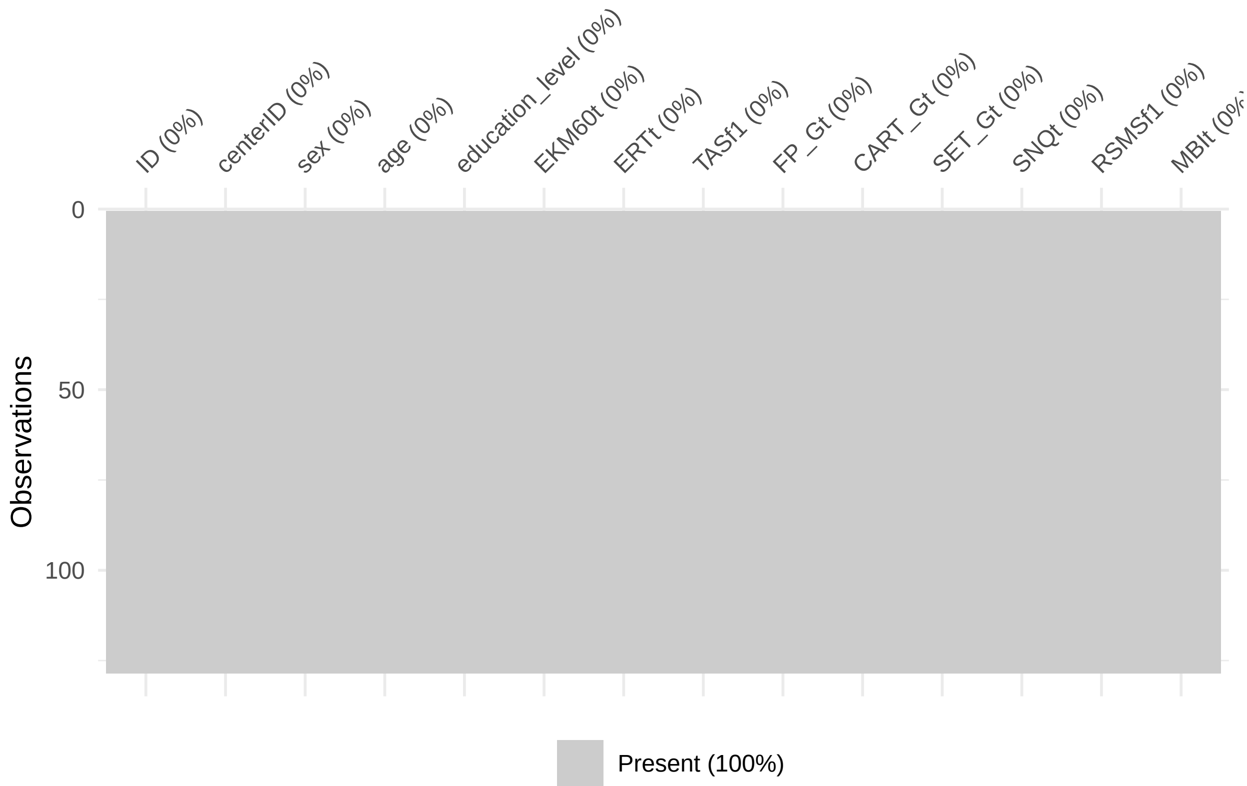
5.5 Imputation of CART_Gt, SET_Gt, EKM60t, ERTt (2nd imputation round)

```r
data_pooled_subset_imp2 <- data_pooled_subset_imp

predictor_matrix2 <- make.predictorMatrix(data_pooled_subset_imp)
predictor_matrix2[] <- 0

#assign the significant predictors to the prediction matrix
predictor_matrix2["CART_Gt", ] <-
  colnames(predictor_matrix2) %in% names(sigpred_CART_Gt)[!is.na(names(sigpred_CART_Gt))]
predictor_matrix2["SET_Gt", ] <-
  colnames(predictor_matrix2) %in% names(sigpred_SET_Gt)[!is.na(names(sigpred_SET_Gt))]
predictor_matrix2["EKM60t", ] <-
  colnames(predictor_matrix2) %in% names(sigpred_EKM60t)[!is.na(names(sigpred_EKM60t))]
predictor_matrix2["ERTt", ] <-
  colnames(predictor_matrix2) %in% names(sigpred_ERTt_2nd)[!is.na(names(sigpred_ERTt_2nd))]


#assign imputation medhods
imputation_methods2 <- make.method(data_pooled_subset_imp)
imputation_methods2[] <- ""  # Set all methods to "" initially
imputation_methods2[c("CART_Gt", "SET_Gt", "EKM60t", "ERTt")] <- "pmm"  # Use "pmm" for ERTt

#imputation
data_pooled_subset_imp2 <- complete(mice(data_pooled_subset_imp,
                               method = imputation_methods2,
                               predictorMatrix = predictor_matrix2,
                               m = 1,
                               seed = 42), 1)
```

```
## 
##  iter imp variable
##   1   1  EKM60t  ERTt  CART_Gt  SET_Gt
##   2   1  EKM60t  ERTt  CART_Gt  SET_Gt
##   3   1  EKM60t  ERTt  CART_Gt  SET_Gt
##   4   1  EKM60t  ERTt  CART_Gt  SET_Gt
##   5   1  EKM60t  ERTt  CART_Gt  SET_Gt
```

5.5.1 Check after imputation part 2

```r
vis_miss(data_pooled_subset_imp2)
```



No missing values anymore

6 Final dataset

6.1 Characteristics

```r
data_final <- data_pooled_subset_imp2

descriptives_data_final <- psych::describe(data_final[,-1]) %>%
  select(n, min, max, mean, median, sd, skew, kurtosis); descriptives_data_final
```

```
##                   n  min max  mean median    sd  skew kurtosis
## centerID*       128  1.0   2  1.63      2  0.48 -0.54    -1.72
## sex             128  0.0   1  0.64      1  0.48 -0.58    -1.68
## age             128 18.0  79 35.66     24 18.61  0.75    -1.00
## education_level 128  2.0   7  5.91      6  0.83 -1.57     4.69
## EKM60t          128 34.0  58 47.16     48  4.66 -0.35    -0.27
## ERTt            128 34.0  76 59.30     60 10.03 -0.34    -0.68
## TASf1           128  5.0  31 20.77     21  4.92 -0.71     0.49
## FP_Gt           128 10.5  24 20.31     21  2.42 -0.95     1.32
```

```
## CART_Gt          128  0.0  12  8.39      9  2.57 -0.41    -0.41
## SET_Gt           128  8.0  12 11.36     12  0.85 -1.52     2.03
## SNQt             128 16.0  22 19.35     19  1.21 -0.43    -0.07
## RSMSf1           128 11.0  32 22.63     23  4.57 -0.39    -0.37
## MBIt             128 42.0  70 52.94     53  6.39  0.03    -0.31
```

```
corr_data_final <- Hmisc:: rcorr(as.matrix(data_final[,-c(1:2)])); corr_data_final
```

```
##                   sex   age education_level EKM60t  ERTt TASf1 FP_Gt CART_Gt
## sex              1.00 -0.02           -0.14   0.10 -0.03 -0.03 -0.11   -0.35
## age             -0.02  1.00           -0.28  -0.06 -0.32  0.04 -0.10   -0.04
## education_level -0.14 -0.28            1.00   0.18  0.27  0.09  0.14    0.38
## EKM60t           0.10 -0.06            0.18   1.00  0.50  0.16  0.30    0.28
## ERTt            -0.03 -0.32            0.27   0.50  1.00  0.10  0.42    0.25
## TASf1           -0.03  0.04            0.09   0.16  0.10  1.00  0.11    0.01
## FP_Gt           -0.11 -0.10            0.14   0.30  0.42  0.11  1.00    0.26
## CART_Gt         -0.35 -0.04            0.38   0.28  0.25  0.01  0.26    1.00
## SET_Gt          -0.13 -0.18            0.15   0.11  0.50  0.05  0.17    0.19
## SNQt             0.02 -0.09            0.26   0.09  0.09  0.06  0.23    0.12
## RSMSf1          -0.22  0.05           -0.01   0.19  0.01  0.25  0.01    0.21
## MBIt             0.15  0.31            0.05   0.06 -0.07  0.12 -0.07    0.08
##                 SET_Gt  SNQt RSMSf1  MBIt
## sex              -0.13  0.02  -0.22  0.15
## age              -0.18 -0.09   0.05  0.31
## education_level   0.15  0.26  -0.01  0.05
## EKM60t            0.11  0.09   0.19  0.06
## ERTt              0.50  0.09   0.01 -0.07
## TASf1             0.05  0.06   0.25  0.12
## FP_Gt             0.17  0.23   0.01 -0.07
## CART_Gt           0.19  0.12   0.21  0.08
## SET_Gt            1.00  0.10   0.06 -0.05
## SNQt              0.10  1.00   0.00  0.04
## RSMSf1            0.06  0.00   1.00  0.00
## MBIt             -0.05  0.04   0.00  1.00
##
## n= 128
##
##
## P
##                 sex    age    education_level EKM60t ERTt   TASf1  FP_Gt
## sex                    0.7994 0.1036          0.2440 0.7703 0.6977 0.2066
## age             0.7994        0.0015          0.5290 0.0002 0.6536 0.2670
## education_level 0.1036 0.0015                 0.0399 0.0023 0.2938 0.1080
## EKM60t          0.2440 0.5290 0.0399                 0.0000 0.0765 0.0006
## ERTt            0.7703 0.0002 0.0023          0.0000        0.2653 0.0000
## TASf1           0.6977 0.6536 0.2938          0.0765 0.2653        0.2054
## FP_Gt           0.2066 0.2670 0.1080          0.0006 0.0000 0.2054
## CART_Gt         0.0000 0.6772 0.0000          0.0015 0.0052 0.8905 0.0031
## SET_Gt          0.1567 0.0399 0.0819          0.2271 0.0000 0.6087 0.0497
## SNQt            0.8589 0.2871 0.0028          0.3173 0.3092 0.4734 0.0107
## RSMSf1          0.0121 0.5607 0.9181          0.0304 0.8924 0.0046 0.8941
## MBIt            0.0884 0.0004 0.5449          0.4963 0.4276 0.1680 0.4558
##                 CART_Gt SET_Gt SNQt   RSMSf1 MBIt
## sex             0.0000  0.1567 0.8589 0.0121 0.0884
```

```
## age             0.6772  0.0399 0.2871 0.5607 0.0004
## education_level 0.0000  0.0819 0.0028 0.9181 0.5449
## EKM60t          0.0015  0.2271 0.3173 0.0304 0.4963
## ERTt            0.0052  0.0000 0.3092 0.8924 0.4276
## TASf1           0.8905  0.6087 0.4734 0.0046 0.1680
## FP_Gt           0.0031  0.0497 0.0107 0.8941 0.4558
## CART_Gt                 0.0279 0.1654 0.0158 0.3729
## SET_Gt          0.0279         0.2702 0.5180 0.5542
## SNQt            0.1654  0.2702        0.9812 0.6244
## RSMSf1          0.0158  0.5180 0.9812        0.9830
## MBIt            0.3729  0.5542 0.6244 0.9830
```

```r
corr_data_final_r <- round(as.data.frame(corr_data_final$r), 3); corr_data_final_r
```

```
##                    sex    age education_level EKM60t   ERTt  TASf1  FP_Gt
## sex              1.000 -0.023          -0.145  0.104 -0.026 -0.035 -0.112
## age             -0.023  1.000          -0.277 -0.056 -0.320  0.040 -0.099
## education_level -0.145 -0.277           1.000  0.182  0.267  0.094  0.143
## EKM60t           0.104 -0.056           0.182  1.000  0.504  0.157  0.299
## ERTt            -0.026 -0.320           0.267  0.504  1.000  0.099  0.418
## TASf1           -0.035  0.040           0.094  0.157  0.099  1.000  0.113
## FP_Gt           -0.112 -0.099           0.143  0.299  0.418  0.113  1.000
## CART_Gt         -0.354 -0.037           0.379  0.278  0.245  0.012  0.260
## SET_Gt          -0.126 -0.182           0.154  0.108  0.500  0.046  0.174
## SNQt             0.016 -0.095           0.262  0.089  0.091  0.064  0.225
## RSMSf1          -0.221  0.052          -0.009  0.191  0.012  0.249  0.012
## MBIt             0.151  0.306           0.054  0.061 -0.071  0.123 -0.066
##                 CART_Gt SET_Gt   SNQt RSMSf1   MBIt
## sex              -0.354 -0.126  0.016 -0.221  0.151
## age              -0.037 -0.182 -0.095  0.052  0.306
## education_level   0.379  0.154  0.262 -0.009  0.054
## EKM60t            0.278  0.108  0.089  0.191  0.061
## ERTt              0.245  0.500  0.091  0.012 -0.071
## TASf1             0.012  0.046  0.064  0.249  0.123
## FP_Gt             0.260  0.174  0.225  0.012 -0.066
## CART_Gt           1.000  0.194  0.123  0.213  0.079
## SET_Gt            0.194  1.000  0.098  0.058 -0.053
## SNQt              0.123  0.098  1.000 -0.002  0.044
## RSMSf1            0.213  0.058 -0.002  1.000  0.002
## MBIt              0.079 -0.053  0.044  0.002  1.000
```

```r
corr_data_final_P <- round(as.data.frame(corr_data_final$P), 3); corr_data_final_P
```

```
##                   sex   age education_level EKM60t  ERTt TASf1 FP_Gt CART_Gt
## sex                NA 0.799           0.104  0.244 0.770 0.698 0.207   0.000
## age             0.799    NA           0.002  0.529 0.000 0.654 0.267   0.677
## education_level 0.104 0.002              NA  0.040 0.002 0.294 0.108   0.000
## EKM60t          0.244 0.529           0.040     NA 0.000 0.076 0.001   0.001
## ERTt            0.770 0.000           0.002  0.000    NA 0.265 0.000   0.005
## TASf1           0.698 0.654           0.294  0.076 0.265    NA 0.205   0.890
## FP_Gt           0.207 0.267           0.108  0.001 0.000 0.205    NA   0.003
## CART_Gt         0.000 0.677           0.000  0.001 0.005 0.890 0.003      NA
## SET_Gt          0.157 0.040           0.082  0.227 0.000 0.609 0.050   0.028
```

```
## SNQt            0.859 0.287           0.003  0.317 0.309 0.473 0.011   0.165
## RSMSf1          0.012 0.561           0.918  0.030 0.892 0.005 0.894   0.016
## MBIt            0.088 0.000           0.545  0.496 0.428 0.168 0.456   0.373
##                 SET_Gt  SNQt RSMSf1  MBIt
## sex              0.157 0.859  0.012 0.088
## age              0.040 0.287  0.561 0.000
## education_level  0.082 0.003  0.918 0.545
## EKM60t           0.227 0.317  0.030 0.496
## ERTt             0.000 0.309  0.892 0.428
## TASf1            0.609 0.473  0.005 0.168
## FP_Gt            0.050 0.011  0.894 0.456
## CART_Gt          0.028 0.165  0.016 0.373
## SET_Gt              NA 0.270  0.518 0.554
## SNQt             0.270    NA  0.981 0.624
## RSMSf1           0.518 0.981     NA 0.983
## MBIt             0.554 0.624  0.983    NA
```

```
#knitr::kable(corr_data_final_r, format = "latex", booktabs = TRUE, caption = "Correlation matrix final
#knitr::kable(corr_data_final_P, format = "latex", booktabs = TRUE, caption = "P-values corresponding t
```

6.2 Z-scores

```
names_items_analysis <- c("EKM60t", "ERTt", "TASf1", "CART_Gt", "FP_Gt", "SET_Gt", "SNQt", "MBIt", "RSMS
data_final_Z <- cbind(data_final[, c("ID", "centerID", names_covariates)], round(scale(data_final[, name
```

7 Analyses

7.1 CFA models

```
#full 3 factor model
m1_f3fm <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Allow factors to be correlated
f1_P ~~ f2_U
f1_P ~~ f3_BR
f2_U ~~ f3_BR
'

#two factor model: f1 = f2
m2_2fm_1eq2 <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt
```

```
# Allow factors to be correlated
f1_P ~~ 1*f2_U  # Fix correlation between f1_P and f2_U at 1
f1_P ~~ f3_BR
f2_U ~~ f3_BR
'

#model 3: two factor model: f1 = f3 (CART)
m3_2fm_1eq3 <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Allow factors to be correlated
f1_P ~~ f2_U
f1_P ~~ 1*f3_BR # Fix correlation between f1_P and f3_BR at 1
f2_U ~~ f3_BR
'

#model 4: two factor model: f2 = f3 (CART)
m4_2fm_2eq3 <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Allow factors to be correlated
f1_P ~~ f2_U
f1_P ~~ f3_BR
f2_U ~~ 1*f3_BR # Fix correlation between f2_U and f3_BR at 1
'

#model 5: one factor model: f1 = f2 = f3 (CART)
m5_1fm <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Fixing all interfactor correlations at 1
f1_P ~~ 1*f2_U
f1_P ~~ 1*f3_BR
f2_U ~~ 1*f3_BR
'

#model 6: 3 one factor model: independent factors (CART)
m6_3x1fm <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Fixing all interfactor correlations at 0
```

```
f1_P ~~ 0*f2_U
f1_P ~~ 0*f3_BR
f2_U ~~ 0*f3_BR
'

 #model 6.3alt: 3 factor model; BR independent
m6_3fm_BRindep <- '
# Defining the factors (latent variables)
f1_P =~ EKM60t + ERTt + TASf1
f2_U =~ SET_Gt + FP_Gt + CART_Gt
f3_BR =~ RSMSf1 + SNQt + MBIt

# Fixing all interfactor correlations at 0
f1_P ~~ f2_U
f1_P ~~ 0*f3_BR
f2_U ~~ 0*f3_BR
'
```

7.2 CFA EMC data (N=47)

```
#model 1; EMC data
fit_m1_EMC <- cfa(model = m1_f3fm, data = data_final_Z[data_final_Z$centerID == "EMC",])
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m1_EMC <- summary(fit_m1_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m1_EMC <- fitMeasures(fit_m1_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 2; EMC data
fit_m2_EMC <- cfa(model = m2_2fm_1eq2, data = data_final_Z[data_final_Z$centerID == "EMC",])
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f2_U
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m2_EMC <- summary(fit_m2_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m2_EMC <- fitMeasures(fit_m2_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 3; EMC data
fit_m3_EMC <- cfa(model = m3_2fm_1eq3, data = data_final_Z[data_final_Z$centerID == "EMC",])
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f3_BR
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m3_EMC <- summary(fit_m3_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m3_EMC <- fitMeasures(fit_m3_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 4; EMC data
fit_m4_EMC <- cfa(model = m4_2fm_2eq3, data = data_final_Z[data_final_Z$centerID == "EMC",])
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f2_U f3_BR
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m4_EMC <- summary(fit_m4_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m4_EMC <- fitMeasures(fit_m4_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 5; EMC data
fit_m5_EMC <- cfa(model = m5_1fm, data = data_final_Z[data_final_Z$centerID == "EMC",])
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f2_U
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f3_BR
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m5_EMC <- summary(fit_m5_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m5_EMC <- fitMeasures(fit_m5_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
# #model 6; EMC data
# fit_m6_EMC <- cfa(model = m6_3x1fm, data = data_final_Z[data_final_Z$centerID == "EMC",])
# summ_m6_EMC <- summary(fit_m6_EMC, standardized = TRUE, fit.measures = TRUE)
# fm_m6_EMC <- fitMeasures(fit_m6_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi",
```

```
#model 6.3; EMC data
fit_m6.3alt_EMC <- cfa(model = m6_3fm_BRindep, data = data_final_Z[data_final_Z$centerID == "EMC",])
summ_m6.3alt_EMC <- summary(fit_m6.3alt_EMC, standardized = TRUE, fit.measures = TRUE)
fm_m6.3alt_EMC <- fitMeasures(fit_m6.3alt_EMC, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi",
```

7.3 CFA pooled data (N=128)

```r
#model 1; pooled data
fit_m1_pld <- cfa(model = m1_f3fm, data = data_final_Z)
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```r
summ_m1_pld <- summary(fit_m1_pld, standardized = TRUE, fit.measures = TRUE)
fm_m1_pld <- fitMeasures(fit_m1_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```r
#model 2; pooled data
fit_m2_pld <- cfa(model = m2_2fm_1eq2, data = data_final_Z)
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f2_U
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```r
summ_m2_pld <- summary(fit_m2_pld, standardized = TRUE, fit.measures = TRUE)
fm_m2_pld <- fitMeasures(fit_m2_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```r
#model 3; pooled data
fit_m3_pld <- cfa(model = m3_2fm_1eq3, data = data_final_Z)
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f3_BR
```

```
## Warning: lavaan->lav_object_post_check():
##    covariance matrix of latent variables is not positive definite ; use
##    lavInspect(fit, "cov.lv") to investigate.
```

```r
summ_m3_pld <- summary(fit_m3_pld, standardized = TRUE, fit.measures = TRUE)
fm_m3_pld <- fitMeasures(fit_m3_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```r
#model 4; pooled data
fit_m4_pld <- cfa(model = m4_2fm_2eq3, data = data_final_Z)
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f2_U f3_BR
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated ov variances are negative
```

```
## Warning: lavaan->lav_object_post_check():
##    covariance matrix of latent variables is not positive definite ; use
##    lavInspect(fit, "cov.lv") to investigate.
```

```
summ_m4_pld <- summary(fit_m4_pld, standardized = TRUE, fit.measures = TRUE)
fm_m4_pld <- fitMeasures(fit_m4_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 5; pooled data
fit_m5_pld <- cfa(model = m5_1fm, data = data_final_Z)
```

```
## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f2_U

## Warning: lavaan->lav_start_check_cov():
##    starting values imply a correlation larger than 1; variables involved are:
##    f1_P f3_BR

## Warning: lavaan->lav_object_post_check():
##    covariance matrix of latent variables is not positive definite ; use
##    lavInspect(fit, "cov.lv") to investigate.
```

```
summ_m5_pld <- summary(fit_m5_pld, standardized = TRUE, fit.measures = TRUE)
fm_m5_pld <- fitMeasures(fit_m5_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

```
#model 6; pooled data
fit_m6_pld <- cfa(model = m6_3x1fm, data = data_final_Z)
```

```
## Warning: lavaan->lav_object_post_check():
##    some estimated lv variances are negative
```

```
summ_m6_pld <- summary(fit_m6_pld, standardized = TRUE, fit.measures = TRUE)
fm_m6_pld <- fitMeasures(fit_m6_pld, c("npar", "chisq", "df", "pvalue", "aic", "srmr", "cfi", "ifi", "n
```

7.4 SEM models

7.5 SEM EMC data (N=47)

8 Results

8.1 CFA EMC data

8.1.1 Fit indices

```
results_CFA_EMC <- data.frame(round(rbind(fm_m1_EMC, fm_m2_EMC, fm_m3_EMC, fm_m4_EMC, fm_m5_EMC
                                          #, fm_m6_EMC
                                          ), 3))
modelNames_CFA <- c("Full three-factor", "Two-factor: P = U", "Two-factor: P = BR", "Two-factor: U = BR
                    #, "Independent three factor"
                    )
rownames(results_CFA_EMC) <- NULL
results_CFA_EMC <- cbind(Model = modelNames_CFA, results_CFA_EMC)
results_CFA_EMC
```

```
##                    Model npar  chisq df pvalue      aic  srmr   cfi   ifi   nfi
## 1      Full three-factor   21 25.952 24  0.356 1176.076 0.084 0.962 0.969 0.705
## 2      Two-factor: P = U   20 43.386 25  0.013 1191.510 0.374 0.646 0.708 0.506
## 3     Two-factor: P = BR   20 35.310 25  0.083 1183.434 0.248 0.801 0.836 0.598
## 4     Two-factor: U = BR   20 49.263 25  0.003 1197.387 0.432 0.532 0.614 0.439
## 5 One-factor: P = U = BR   18 50.079 27  0.004 1194.204 0.464 0.555 0.621 0.430
##   rmsea      bic
## 1 0.042 1214.929
## 2 0.125 1228.513
## 3 0.094 1220.437
## 4 0.144 1234.390
## 5 0.135 1227.506
```

```
results_CFA_EMC[,c("Model", "npar", "df", "chisq", "pvalue", "aic", "bic", "srmr", "cfi", "ifi")]
```

```
##                    Model npar df  chisq pvalue      aic      bic  srmr   cfi
## 1      Full three-factor   21 24 25.952  0.356 1176.076 1214.929 0.084 0.962
## 2      Two-factor: P = U   20 25 43.386  0.013 1191.510 1228.513 0.374 0.646
## 3     Two-factor: P = BR   20 25 35.310  0.083 1183.434 1220.437 0.248 0.801
## 4     Two-factor: U = BR   20 25 49.263  0.003 1197.387 1234.390 0.432 0.532
## 5 One-factor: P = U = BR   18 27 50.079  0.004 1194.204 1227.506 0.464 0.555
##     ifi
## 1 0.969
## 2 0.708
## 3 0.836
## 4 0.614
## 5 0.621
```

8.1.2 Model comparisson

```
anova(fit_m1_EMC, fit_m2_EMC)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_EMC 24 1176.1 1214.9 25.952
## fit_m2_EMC 25 1191.5 1228.5 43.386     17.434 0.59133       1  2.974e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit_m1_EMC, fit_m3_EMC)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_EMC 24 1176.1 1214.9 25.952
## fit_m3_EMC 25 1183.4 1220.4 35.310     9.3578 0.42169       1    0.00222 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_EMC, fit_m4_EMC)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_EMC 24 1176.1 1214.9 25.952
## fit_m4_EMC 25 1197.4 1234.4 49.263     23.311 0.68898       1  1.378e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_EMC, fit_m5_EMC)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_EMC 24 1176.1 1214.9 25.952
## fit_m5_EMC 27 1194.2 1227.5 50.079     24.128 0.38709       3  2.349e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#anova(fit_m1_EMC, fit_m6_EMC)
```

8.1.3 Favoured model

```r
est_par_CFA_EMC <- parameterEstimates(fit_m1_EMC, standardized = TRUE)
rownames(est_par_CFA_EMC) <- NULL
est_par_CFA_EMC
```

```
##        lhs op     rhs   est    se     z pvalue ci.lower ci.upper std.lv
## 1     f1_P =~  EKM60t 1.000 0.000    NA     NA    1.000    1.000  0.951
## 2     f1_P =~    ERTt 0.671 0.185 3.634  0.000    0.309    1.033  0.638
## 3     f1_P =~   TASf1 0.308 0.186 1.655  0.098   -0.057    0.674  0.293
## 4     f2_U =~  SET_Gt 1.000 0.000    NA     NA    1.000    1.000  0.194
## 5     f2_U =~   FP_Gt 3.354 2.444 1.372  0.170   -1.436    8.144  0.652
## 6     f2_U =~ CART_Gt 3.571 2.593 1.377  0.168   -1.512    8.653  0.694
## 7    f3_BR =~  RSMSf1 1.000 0.000    NA     NA    1.000    1.000     NA
## 8    f3_BR =~    SNQt 0.334 0.357 0.933  0.351   -0.367    1.034     NA
## 9    f3_BR =~    MBIt 0.302 0.358 0.843  0.399   -0.400    1.003     NA
## 10    f1_P ~~    f2_U 0.133 0.101 1.327  0.184   -0.064    0.330  0.722
## 11    f1_P ~~   f3_BR 0.293 0.150 1.953  0.051   -0.001    0.587  0.549
## 12    f2_U ~~   f3_BR 0.085 0.069 1.230  0.219   -0.051    0.222  0.783
## 13  EKM60t ~~  EKM60t 0.206 0.210 0.980  0.327   -0.206    0.618  0.206
## 14    ERTt ~~    ERTt 0.457 0.133 3.439  0.001    0.196    0.717  0.457
## 15   TASf1 ~~   TASf1 1.129 0.236 4.780  0.000    0.666    1.592  1.129
## 16  SET_Gt ~~  SET_Gt 0.614 0.130 4.737  0.000    0.360    0.868  0.614
## 17   FP_Gt ~~   FP_Gt 0.602 0.174 3.464  0.001    0.261    0.943  0.602
## 18 CART_Gt ~~ CART_Gt 0.596 0.184 3.237  0.001    0.235    0.958  0.596
```

```
## 19  RSMSf1 ~~  RSMSf1  1.338 0.605  2.211  0.027    0.152    2.524  1.338
## 20    SNQt ~~    SNQt  0.895 0.198  4.517  0.000    0.507    1.283  0.895
## 21    MBIt ~~    MBIt  1.260 0.270  4.658  0.000    0.730    1.790  1.260
## 22    f1_P ~~    f1_P  0.904 0.305  2.965  0.003    0.306    1.501  1.000
## 23    f2_U ~~    f2_U  0.038 0.053  0.715  0.474   -0.066    0.141  1.000
## 24   f3_BR ~~   f3_BR -0.315 0.508 -0.620  0.535   -1.311    0.681     NA
##    std.all
## 1    0.902
## 2    0.686
## 3    0.266
## 4    0.241
## 5    0.643
## 6    0.668
## 7       NA
## 8       NA
## 9       NA
## 10   0.722
## 11   0.549
## 12   0.783
## 13   0.186
## 14   0.529
## 15   0.929
## 16   0.942
## 17   0.586
## 18   0.553
## 19   1.308
## 20   1.041
## 21   1.023
## 22   1.000
## 23   1.000
## 24      NA
```

8.2 CFA pooled data

8.2.1 Fit indices

```r
results_CFA_pld <- data.frame(round(rbind(fm_m1_pld, fm_m2_pld, fm_m3_pld, fm_m4_pld, fm_m5_pld, fm_m6_
modelNames_CFA <- c("Full three-factor", "Two-factor: P = U", "Two-factor: P = BR", "Two-factor: U = BR
rownames(results_CFA_pld) <- NULL
results_CFA_pld <- cbind(Model = modelNames_CFA, results_CFA_pld)
results_CFA_pld
```

```
##                       Model npar   chisq df pvalue      aic  srmr   cfi   ifi
## 1        Full three-factor   21  45.363 24  0.005 3185.383 0.075 0.831 0.845
## 2        Two-factor: P = U   20  80.179 25  0.000 3218.199 0.298 0.563 0.598
## 3       Two-factor: P = BR   20 106.314 25  0.000 3244.334 0.317 0.356 0.407
## 4       Two-factor: U = BR   20  95.357 25  0.000 3233.377 0.294 0.442 0.487
## 5   One-factor: P = U = BR   18 116.076 27  0.000 3250.096 0.457 0.294 0.341
## 6 Independent three factor   18 105.401 27  0.000 3239.420 0.146 0.379 0.420
##     nfi rmsea      bic
## 1 0.720 0.083 3245.275
## 2 0.506 0.131 3275.239
```

```
## 3 0.344 0.159 3301.374
## 4 0.412 0.148 3290.417
## 5 0.284 0.161 3301.432
## 6 0.350 0.151 3290.757
```

8.2.2 Model comparisson

```r
anova(fit_m1_pld, fit_m2_pld)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_pld 24 3185.4 3245.3 45.363
## fit_m2_pld 25 3218.2 3275.2 80.179     34.816 0.51399       1  3.624e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_pld, fit_m3_pld)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC   Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_pld 24 3185.4 3245.3  45.363
## fit_m3_pld 25 3244.3 3301.4 106.314     60.951 0.68438       1   5.85e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_pld, fit_m4_pld)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC  Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_pld 24 3185.4 3245.3 45.363
## fit_m4_pld 25 3233.4 3290.4 95.357     49.994 0.61868       1  1.542e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_pld, fit_m5_pld)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC   Chisq Chisq diff   RMSEA Df diff Pr(>Chisq)
## fit_m1_pld 24 3185.4 3245.3  45.363
## fit_m5_pld 27 3250.1 3301.4 116.076     70.713 0.41992       3  3.003e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit_m1_pld, fit_m6_pld)
```

```
##
## Chi-Squared Difference Test
##
##            Df    AIC    BIC   Chisq Chisq diff  RMSEA Df diff Pr(>Chisq)
## fit_m1_pld 24 3185.4 3245.3  45.363
## fit_m6_pld 27 3239.4 3290.8 105.401     60.038 0.3854       3   5.77e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# anova(fit_m2_pld, fit_m5_pld)
# anova(fit_m3_pld, fit_m5_pld)
# anova(fit_m4_pld, fit_m5_pld)
```

8.2.3 Favoured model

```r
est_par_CFA_pld <- parameterEstimates(fit_m1_pld, standardized = TRUE)
rownames(est_par_CFA_pld) <- NULL
est_par_CFA_pld
```

```
##        lhs op     rhs    est    se      z pvalue ci.lower ci.upper std.lv
## 1     f1_P =~  EKM60t  1.000 0.000     NA     NA    1.000    1.000  0.511
## 2     f1_P =~    ERTt  1.916 0.461  4.154  0.000    1.012    2.820  0.978
## 3     f1_P =~   TASf1  0.203 0.177  1.147  0.251   -0.144    0.550  0.104
## 4     f2_U =~  SET_Gt  1.000 0.000     NA     NA    1.000    1.000  0.516
## 5     f2_U =~   FP_Gt  0.919 0.212  4.337  0.000    0.504    1.334  0.474
## 6     f2_U =~ CART_Gt  0.603 0.194  3.110  0.002    0.223    0.983  0.311
## 7    f3_BR =~  RSMSf1  1.000 0.000     NA     NA    1.000    1.000     NA
## 8    f3_BR =~    SNQt  2.018 1.823  1.107  0.268   -1.556    5.592     NA
## 9    f3_BR =~    MBIt -0.403 0.808 -0.498  0.618   -1.987    1.182     NA
## 10    f1_P ~~    f2_U  0.243 0.076  3.186  0.001    0.094    0.393  0.923
## 11    f1_P ~~   f3_BR  0.023 0.026  0.862  0.389   -0.029    0.075  0.491
## 12    f2_U ~~   f3_BR  0.083 0.068  1.220  0.222   -0.050    0.216  1.769
## 13  EKM60t ~~  EKM60t  0.731 0.104  7.023  0.000    0.527    0.936  0.731
## 14    ERTt ~~    ERTt  0.035 0.183  0.190  0.849   -0.324    0.393  0.035
## 15   TASf1 ~~   TASf1  0.981 0.123  7.996  0.000    0.741    1.222  0.981
## 16  SET_Gt ~~  SET_Gt  0.726 0.108  6.693  0.000    0.513    0.939  0.726
## 17   FP_Gt ~~   FP_Gt  0.767 0.108  7.075  0.000    0.555    0.980  0.767
## 18 CART_Gt ~~ CART_Gt  0.895 0.115  7.809  0.000    0.671    1.120  0.895
## 19  RSMSf1 ~~  RSMSf1  1.000 0.132  7.572  0.000    0.742    1.259  1.000
## 20    SNQt ~~    SNQt  1.026 0.214  4.798  0.000    0.607    1.445  1.026
## 21    MBIt ~~    MBIt  0.994 0.125  7.974  0.000    0.749    1.238  0.994
## 22    f1_P ~~    f1_P  0.261 0.098  2.674  0.008    0.070    0.452  1.000
## 23    f2_U ~~    f2_U  0.266 0.103  2.575  0.010    0.064    0.469  1.000
## 24   f3_BR ~~   f3_BR -0.008 0.040 -0.209  0.835   -0.086    0.069     NA
##    std.all
## 1    0.513
## 2    0.982
## 3    0.104
```

```
## 4    0.518
## 5    0.476
## 6    0.312
## 7       NA
## 8       NA
## 9       NA
## 10   0.923
## 11   0.491
## 12   1.769
## 13   0.737
## 14   0.035
## 15   0.989
## 16   0.732
## 17   0.773
## 18   0.902
## 19   1.008
## 20   1.034
## 21   1.001
## 22   1.000
## 23   1.000
## 24      NA
```