

# Sensitivity analysis on imputed dataset

Andreas Gammelgaard Damsbo

Knitted: 10 February, 2022

## Import

```
dta_all<-read.csv("/Volumes/Data/depression/dep_dataset.csv",na.strings = c("NA","unknown")) ## Extending
```

## Defining patients to include for analysis

Only including cases with complete pase\_0 and MDI at 1 & 6 months

```
dta<-dta_all[!is.na(dta_all$pase_0),]  
# &!is.na(dta$mdi_1)&!is.na(dta$mdi_6)
```

## Formatting

```
# Backup  
dta_b<-dta
```

## Libraries

```
library(daDoctor)  
library(mice)
```

## Variables to include in imputation

```
# Possible variables to include  
coval<-c("pase_0_bin","rtreat","age","sex","smoke_ever","civil","bmi","diabetes","hypertension","afli
```

## Imputation

```

# Output variables added to include in model. Excluded from predicting.
outc<-c("mdi_1_enr", "mdi_6_newobs_enr")
# Adding all
covar<-c(coval, outc)
# Selecting dataset
r<-dta[,c("rnumb", covar)]

```

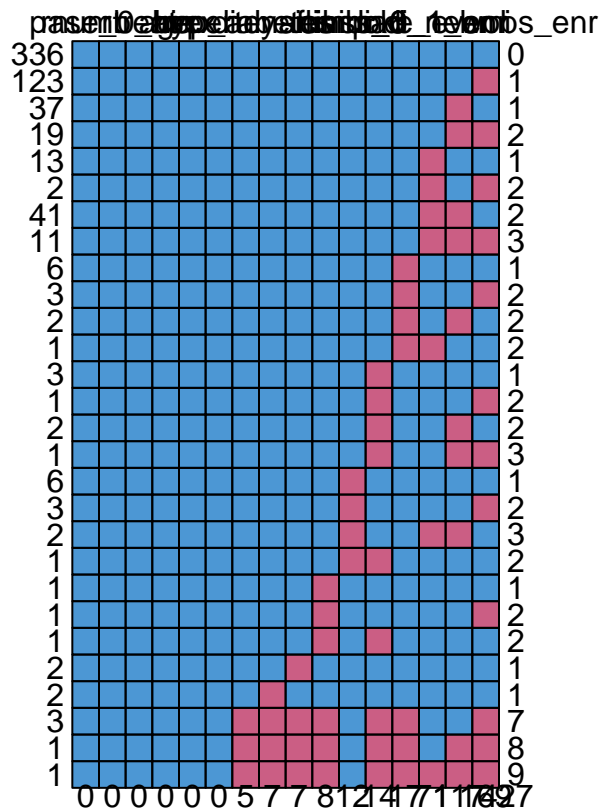
```

# Iterations
mxt=20
# Imputations
mis=5

```

<https://datascienceplus.com/imputing-missing-data-with-r-mice-package/>

```
md.pattern(r) # Missing pattern
```



```

##      rnumb pase_0_bin rtreat age sex rep_any hypertension diabetes afli civil
## 336      1           1      1  1  1      1           1           1  1     1
## 123      1           1      1  1  1      1           1           1  1     1
## 37       1           1      1  1  1      1           1           1  1     1
## 19       1           1      1  1  1      1           1           1  1     1
## 13       1           1      1  1  1      1           1           1  1     1
## 2        1           1      1  1  1      1           1           1  1     1
## 41       1           1      1  1  1      1           1           1  1     1
## 11       1           1      1  1  1      1           1           1  1     1
## 6        1           1      1  1  1      1           1           1  1     1

```

```

## 3      1      1      1  1  1      1      1      1  1  1
## 2      1      1      1  1  1      1      1      1  1  1
## 1      1      1      1  1  1      1      1      1  1  1
## 3      1      1      1  1  1      1      1      1  1  1
## 1      1      1      1  1  1      1      1      1  1  1
## 2      1      1      1  1  1      1      1      1  1  1
## 1      1      1      1  1  1      1      1      1  1  1
## 6      1      1      1  1  1      1      1      1  1  1
## 3      1      1      1  1  1      1      1      1  1  1
## 2      1      1      1  1  1      1      1      1  1  1
## 1      1      1      1  1  1      1      1      1  1  1
## 1      1      1      1  1  1      1      1      1  1  0
## 1      1      1      1  1  1      1      1      1  1  0
## 1      1      1      1  1  1      1      1      1  1  0
## 2      1      1      1  1  1      1      1      1  0  1
## 2      1      1      1  1  1      1      1      0  1  1
## 3      1      1      1  1  1      1      0      0  0  0
## 1      1      1      1  1  1      1      0      0  0  0
## 1      1      1      1  1  1      1      0      0  0  0
##          0      0      0  0  0      0      5      7  7  8
##      nihss_0 pad smoke_ever mdi_1_enr mdi_6_newobs_enr bmi
## 336      1  1      1      1      1  1  0
## 123      1  1      1      1      1  0  1
## 37       1  1      1      1      0  1  1
## 19       1  1      1      1      0  0  2
## 13       1  1      1      0      1  1  1
## 2        1  1      1      0      1  0  2
## 41       1  1      1      0      0  1  2
## 11       1  1      1      0      0  0  3
## 6        1  1      0      1      1  1  1
## 3        1  1      0      1      1  0  2
## 2        1  1      0      1      0  1  2
## 1        1  1      0      0      1  1  2
## 3        1  0      1      1      1  1  1
## 1        1  0      1      1      1  0  2
## 2        1  0      1      1      0  1  2
## 1        1  0      1      1      0  0  3
## 6        0  1      1      1      1  1  1
## 3        0  1      1      1      1  0  2
## 2        0  1      1      0      0  1  3
## 1        0  0      1      1      1  1  2
## 1        1  1      1      1      1  1  1
## 1        1  1      1      1      1  0  2
## 1        1  0      1      1      1  1  2
## 2        1  1      1      1      1  1  1
## 2        1  1      1      1      1  1  1
## 3        1  0      0      1      1  0  7
## 1        1  0      0      1      0  0  8
## 1        1  0      0      0      0  0  9
##          12 14      17      71      117 169 427

```

```

# library(VIM)
# aggr_plot <- aggr(r, col=c('navyblue','red'), numbers=TRUE, sortVars=TRUE, labels=names(data), cex.a

```

```

init <- mice(r, maxit=0) # Creating initial imputation list to assess methods
meth <- init$method
meth

```

```

##          rnumb      pase_0_bin      rtreat      age
##          ""          ""          ""          ""
##          sex      smoke_ever      civil      bmi
##          ""          "logreg"      "logreg"      "pmm"
##          diabetes hypertension      afli      pad
##          "logreg"      "logreg"      "logreg"      "logreg"
##          nihss_0      rep_any      mdi_1_enr mdi_6_newobs_enr
##          "pmm"          ""          "pmm"          "pmm"

```

```

predM <- init$predictorMatrix

```

```

predM[, c("rnumb")] <- 0 # Defining variables not to be used for predicting imputed values
# meth[outc]="
# Defining variables not to be imputed.
# Commented out as all included variables will be imputed

```

```

# summary(imputed)

```

```

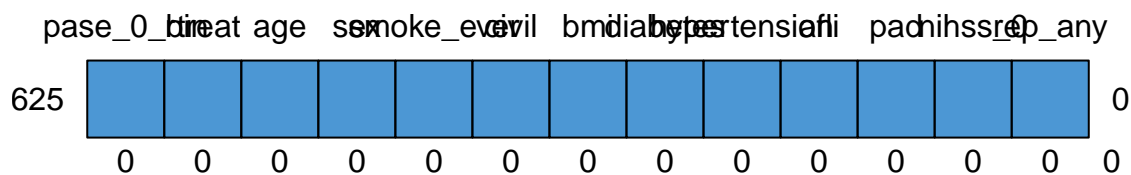
library(dplyr)
export<-dta %>%
  select(-all_of(coval)) %>% # Leaving out imputed variables from original dataset
  left_join(mice::complete(imputed,1),.,by="rnumb") # Join with the first imputed dataset for a full da
md.pattern(export[coval]) # Ensuring complete data

```

```

## /\      /\
## { '---' }
## { 0 0 }
## ==> V <== No need for mice. This data set is completely observed.
## \ \|/ /
## '-----'

```



```

##      pase_0_bin rtreat age sex smoke_ever civil bmi diabetes hypertension afli
## 625          1      1  1  1          1      1  1          1          1  1
##          0      0  0  0          0      0  0          0          0  0
##      pad nihss_0 rep_any
## 625  1          1      1  0
##          0          0      0  0

```

```
write.csv(export, "/Volumes/Data/depression/imputed.csv", row.names = FALSE) # Export
```

## Regression analyses

```
print(adjs_10<-rep_lm(meas = "mdi_6", string=c("pase_0_bin", "rtreat", "coval"), data=dta, cut.p = .1)[[2]])
```

```
## [1] "pase_0_bin" "sex" "civil" "hypertension" "nihss_0"
```

## Bivariabel

Function to format collected data from pool function

```
pool_table<-function(clls){  
  ## Variables needed: estimate, p.value, term  
  
  coll$lo<-round(coll$estimate-coll$std.error*1.96,2)  
  coll$hi<-round(coll$estimate+coll$std.error*1.96,2)  
  
  pa<-coll$p.value  
  pa <- ifelse(pa < 0.001, "<0.001", round(pa, 3))  
  pa <- ifelse(pa <= 0.05 | pa == "<0.001", paste0("*,",pa), ifelse(pa > 0.05 & pa <= 0.1, paste0(".", pa),  
  cl<-data.frame(id=coll$term,diff=paste0(round(coll$estimate,2)," (",coll$lo," to ",coll$hi,")"),p=pa,sta  
  return(cl)  
}  
  
keeps<-c("term","estimate","std.error","p.value")
```

## Repeated bivariabel analysis

Not necessary for this, but an interesting addition

```
coll<-c()  
for (i in c("rtreat",adjs_10)){  
  ## Bivariable linear regression analysis of all  
  coeffs<-summary(pool(  
    with(imputed,lm(as.formula(paste0("mdi_1_enr~",i))))  
  ))[-1,c("term","estimate","std.error","p.value")]  
  
  coll<-rbind(coll,coeffs)  
  
  ## Inspiration: https://stackoverflow.com/questions/40132829/r-for-loop-in-a-formula  
  ## Also: https://gist.github.com/AaronGullickson/3ccb3fdd1778b32fc46df40d78faf5d3  
}  
  
## Collecting  
  
coll$lo<-round(coll$estimate-coll$std.error*1.96,2)
```

```

coll$hi<-round(coll$estimate+coll$std.error*1.96,2)

pa<-coll$p.value
pa <- ifelse(pa < 0.001, "<0.001", round(pa, 3))
pa <- ifelse(pa <= 0.05 | pa == "<0.001", paste0("*",pa), ifelse(pa > 0.05 & pa <= 0.1, paste0(".", pa)

coll_bi<-data.frame(diff=paste0(round(exp(coll$estimate),2)," (" ,coll$lo," to " ,coll$hi,")"),p=pa,id=col

```

## Unadjusted analyses

```

adjs_10m<-adjs_10[adjs_10!="pase_0_bin"]

adj_m<-c("rtreat", "pase_0_bin")

coll<-c()
nms<-c()
for (l in outc){

for (i in adj_m){
  coeffs<-summary(pool(
    with(imputed,lm(as.formula(paste0(l,"~",i))))
  ))[-1,keeps]
  coll<-rbind(coll,coeffs)

  nms<-c(nms,paste(l,i,sep = "_"))

  d.long <- mice::complete(imputed,"long",include = T)

  # Inspiration: https://stackoverflow.com/questions/53014141/mice-splitting-imputed-data-for-further-a

  for (j in levels(d.long[[i]])){
    k<-length(adj_m)-grep(i,adj_m)+1 ## This only works to select the "opposite" of i for length(adj_m)

    s.imp<-mice::as.mids(d.long[which(d.long[[i]] == j),]) # Subsetting long and convert to "mids" form

    coeffs<-summary(pool(
      with(s.imp,lm(as.formula(paste0(l,"~",adj_m[k]))))
    ))[-1,keeps]

    coll<-rbind(coll,coeffs)

    nms<-c(nms,paste(l,j,sep = "_"))
  }

  ## Inspiration: https://stackoverflow.com/questions/40132829/r-for-loop-in-a-formula
  ## Also: https://gist.github.com/AaronGullickson/3ccb3fdd1778b32fc46df40d78faf5d3
}

}

```

```
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
```

```
coll$term<-nms
```

## Collecting

```
biv_coll<-pool_table(coll)
```

## Adjusted analyses

```
adjs_10m<-adjs_10[adjs_10!="pase_0_bin"]

adj_m<-c("rtreat", "pase_0_bin")

coll<-c()
nms<-c()
for (l in outc){
  # l="mdi_1_enr"
  for (i in adj_m){
    coeffs<-summary(pool(
      with(imputed,lm(as.formula(paste0(l,"~",paste(i,paste(adjs_10m,collapse="+"),sep="+")))))
    ))[2,keeps]
    coll<-rbind(coll,coeffs)

    nms<-c(nms,paste(l,i,sep = "_"))

    d.long <- mice::complete(imputed,"long",include = T)

    # Inspiration: https://stackoverflow.com/questions/53014141/mice-splitting-imputed-data-for-further-a

    for (j in levels(d.long[[i]])){
      k<-length(adj_m)-grep(i,adj_m)+1 ## This only works to select the "opposite" of i for length(adj_m)

      s.imp<-mice::as.mids(d.long[which(d.long[[i]] == j),]) # Subsetting long and convert to "mids" form

      coeffs<-summary(pool(
```

```

with(s.imp,lm(as.formula(paste0(1,"~",paste(adj_m[k],paste(adjs_10m,collapse="+"),sep="+")))))
)) [2,keeps]

coll<-rbind(coll,coeffs)

nms<-c(nms,paste(1,j,sep = "_"))
}

## Inspiration: https://stackoverflow.com/questions/40132829/r-for-loop-in-a-formula
## Also: https://gist.github.com/AaronGullickson/3ccb3fdd1778b32fc46df40d78faf5d3
}

}

```

```

## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1
## Warning: Number of logged events: 1

```

```
coll$term<-nms
```

## Collecting

```

mul_coll<-pool_table(coll)
colnames(mul_coll)[-1]<-paste0("adj_",colnames(mul_coll)[-1])

```

```

library(lubridate)
write.csv(full_join(biv_coll,mul_coll),paste0("/Volumes/Data/depression/imp_regression_",today(),".csv"))

```