MCAR 30% missing - Generative Adversarial Imputation Nets (GAIN)

```
# sample MCAR dataset from PUMS
source("../../utils/sampleMCAR.R")
n = 10000
missing_col = c(1,3,7,9,10,11)
missing_prob = 0.3
set.seed(0)

output_list <- sampleMCAR(n, missing_prob)
df <- output_list[['df']]
df_observed <- output_list[['df_observed']]
```

Generative Adversarial Imputation Nets (GAIN)

reference: https://arxiv.org/abs/1806.02920

```
# Load imputed dataset
d1 = read.csv('../../GAIN/imputed_dataset/MCAR_30percent_1.csv', header = FALSE, sep = ',')
d2 = read.csv('../../GAIN/imputed_dataset/MCAR_30percent_2.csv', header = FALSE, sep = ',')
d3 = read.csv('../../GAIN/imputed_dataset/MCAR_30percent_3.csv', header = FALSE, sep = ',')
d4 = read.csv('../../GAIN/imputed_dataset/MCAR_30percent_4.csv', header = FALSE, sep = ',')
d5 = read.csv('../../GAIN/imputed_dataset/MCAR_30percent_5.csv', header = FALSE, sep = ',')
```

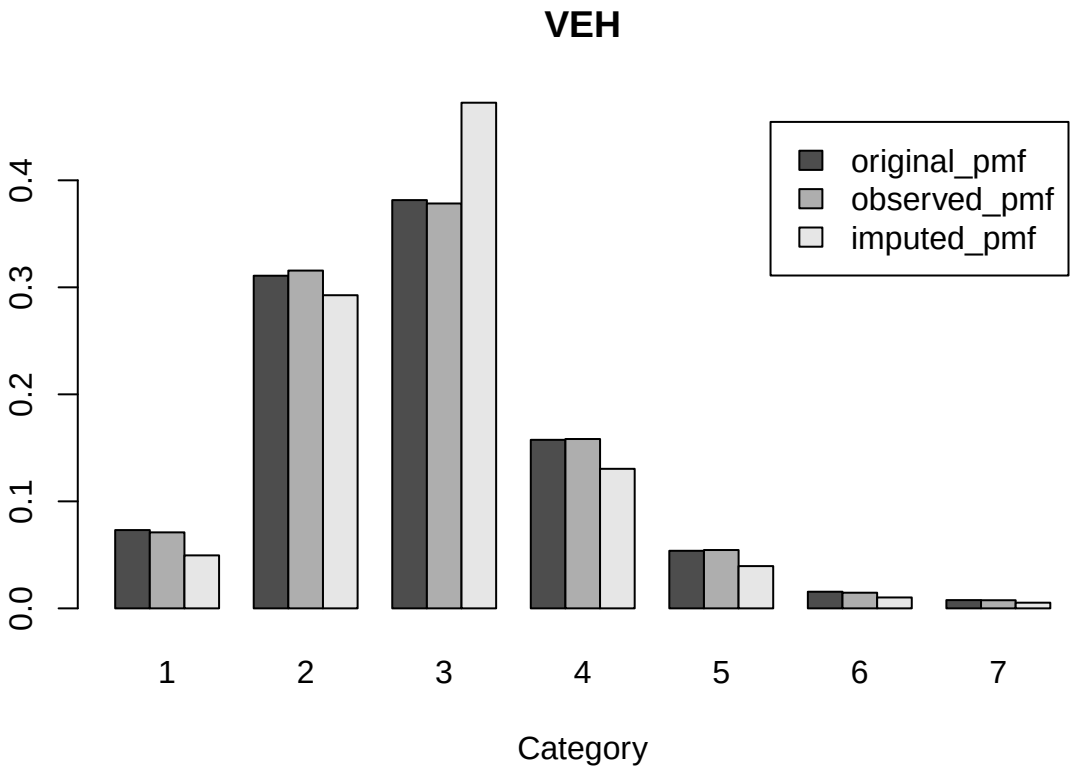
```
# Format imputed dataset into list
imputation_df = list(d1, d2, d3, d4, d5)
imputation_list = list()
levels = c(7,7,7,19,5,4,7,2,17,3,13)
for (i in 1:5) {
  # need to plus 1 here because the class index of DP function starts at 0
  d = imputation_df[[i]]
  colnames(d) = colnames(df_observed)
  # format columns of d
  for (col_index in 1:ncol(df_observed)) {
    d[,col_index] = factor(d[,col_index], levels = 1:levels[col_index], ordered = TRUE)
  }
  imputation_list[[i]] = d
}
```

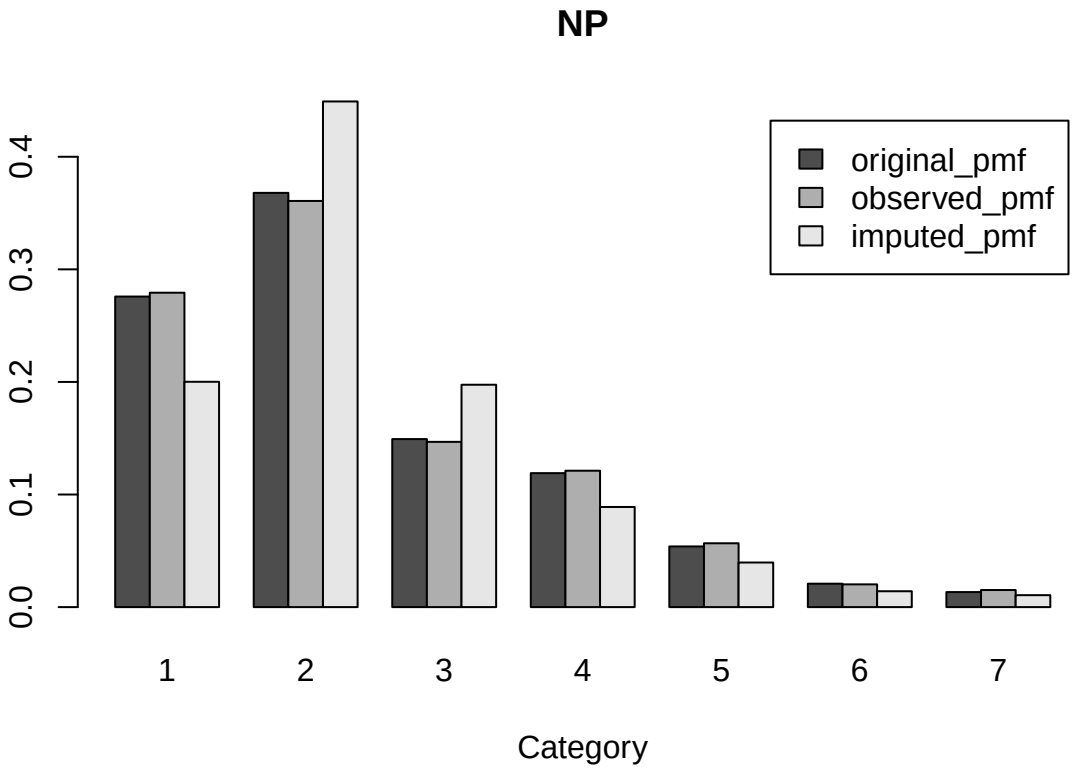
Diagnostics

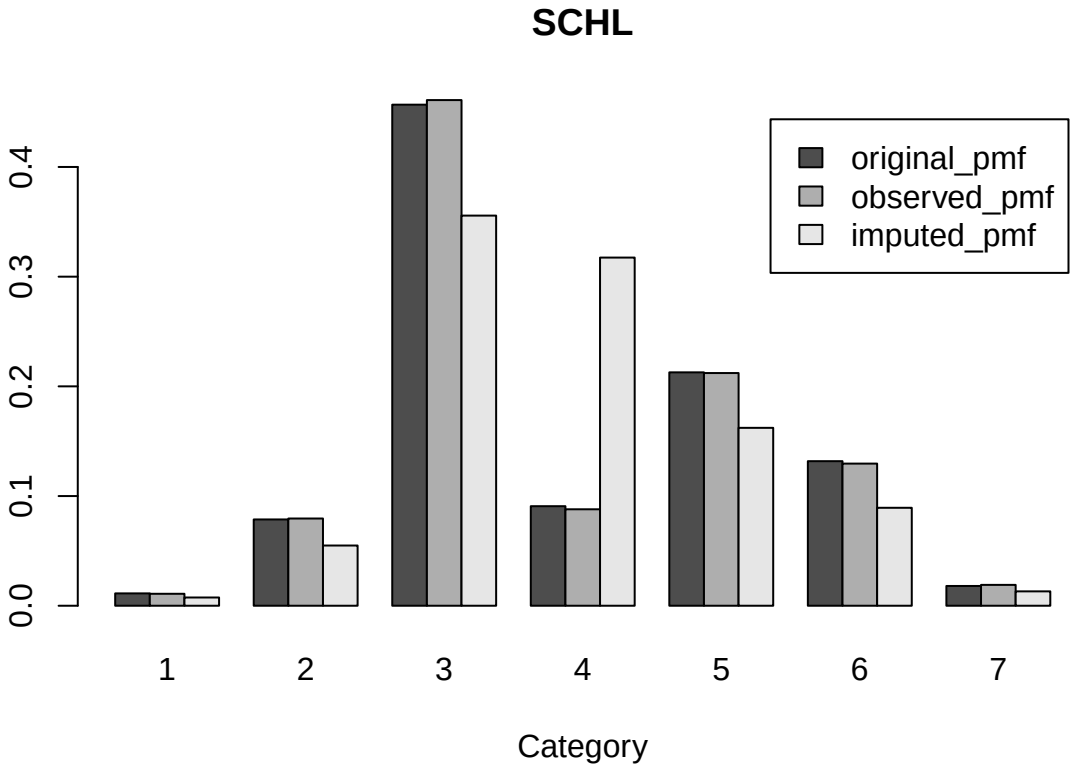
```
source("../../utils/create_report.R")
create_report(imputation_list, max_nway=4, missing_col, df_observed)
```

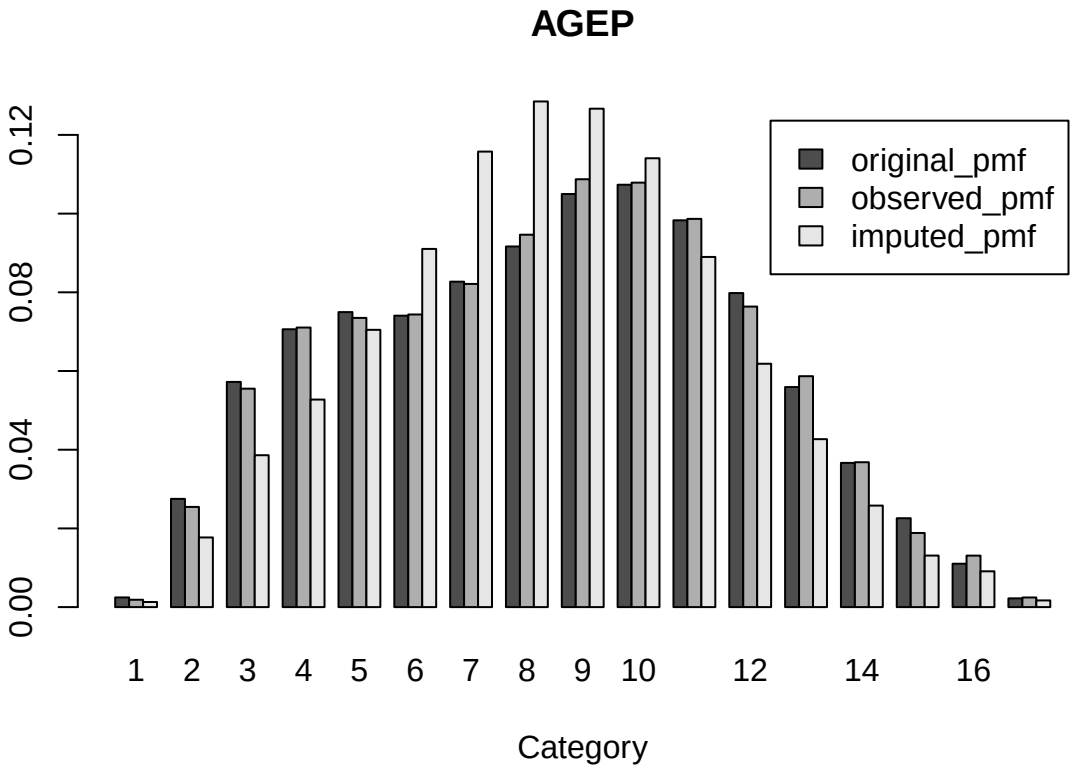
```
## ##### Coverage #####
## Coverage 1 way: 9.26 percent
## Coverage 2 way: 46.72 percent
## Coverage 3 way: 68.57 percent
## Coverage 4 way: 80.7 percent
##
## ##### RMSE #####
```

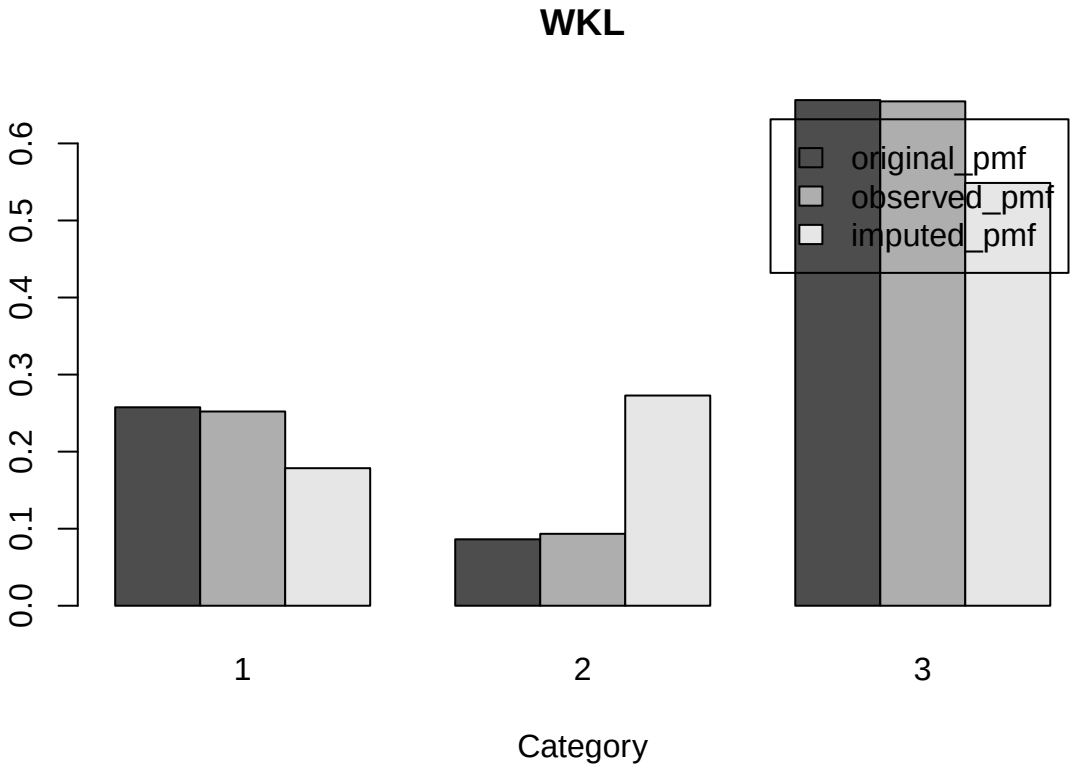
```
## RMSE 1 way: 0.054168
## RMSE 2 way: 0.013748
## RMSE 3 way: 0.003494
## RMSE 4 way: 0.000883
##
## ##### MAE #####
## MAE 1 way: 0.031345
## MAE 2 way: 0.004652
## MAE 3 way: 0.000766
## MAE 4 way: 0.000137
```

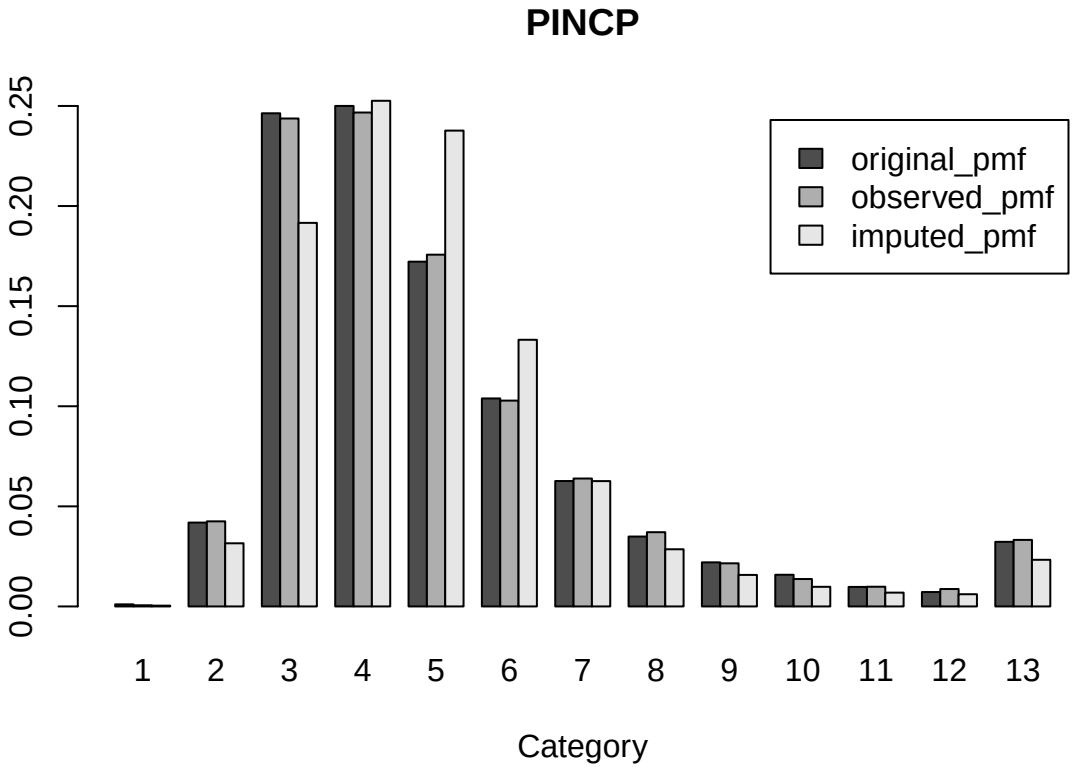


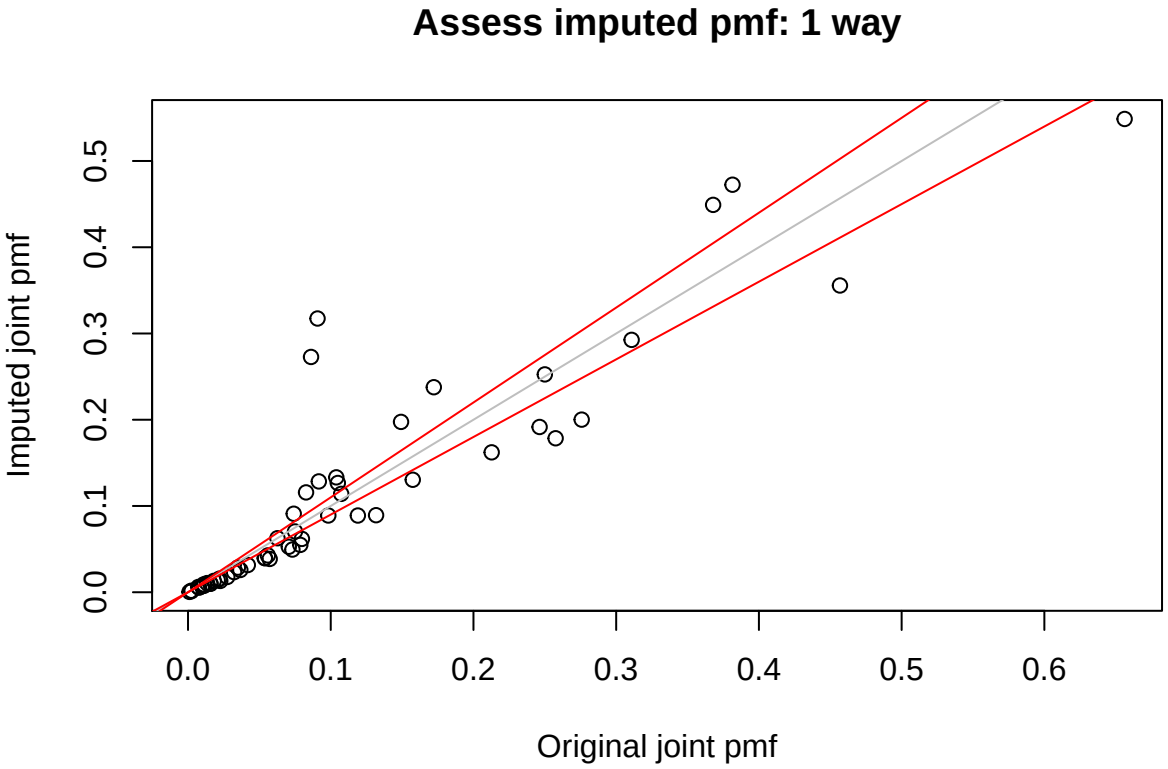

NP
original_pmf
observed_pmf
imputed_pmf
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
Category

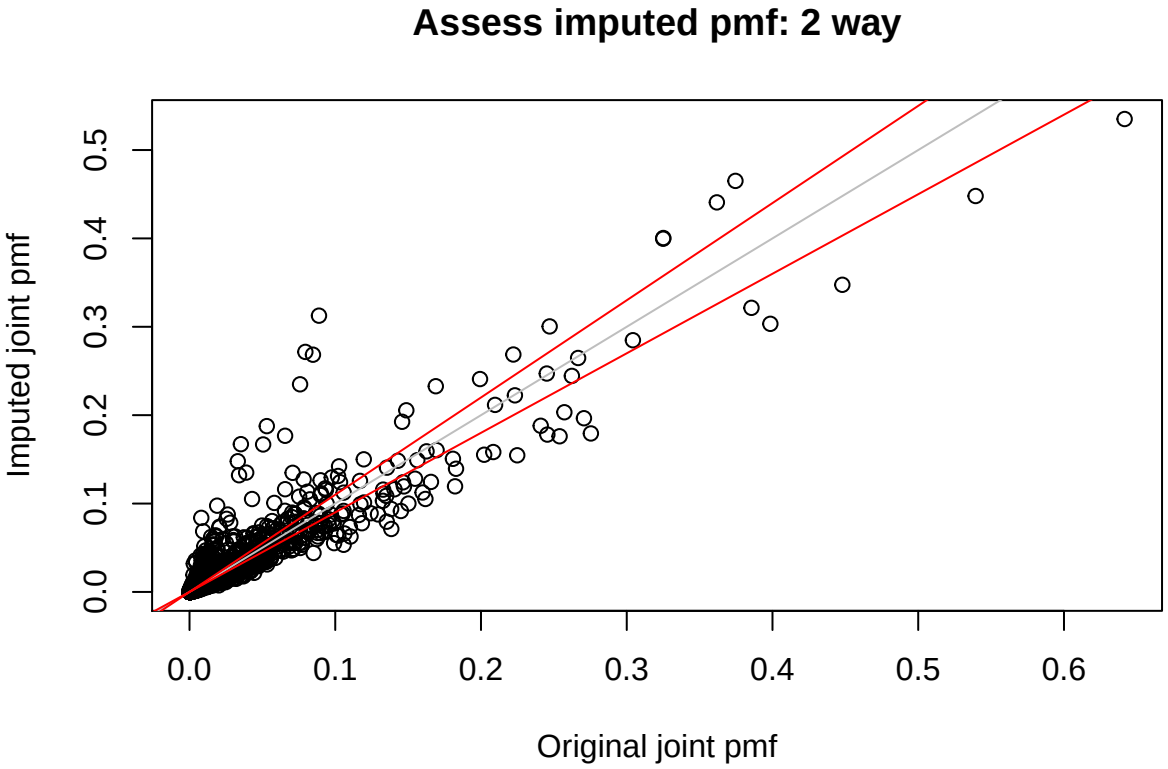

SCHL
original_pmf
observed_pmf
imputed_pmf
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
Category

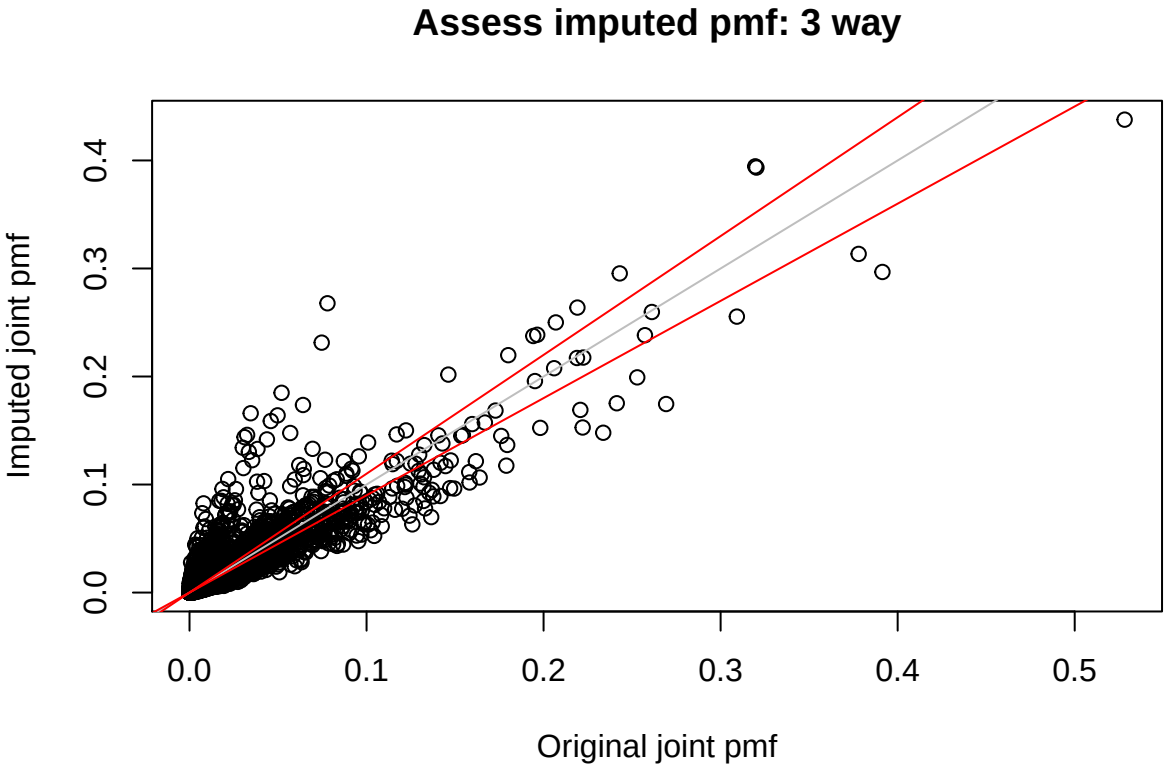

AGEP
0.00
0.04
0.08
0.12
1
2
3
4
5
6
7
8
9
10
12
14
16
Category
original_pmf
observed_pmf
imputed_pmf

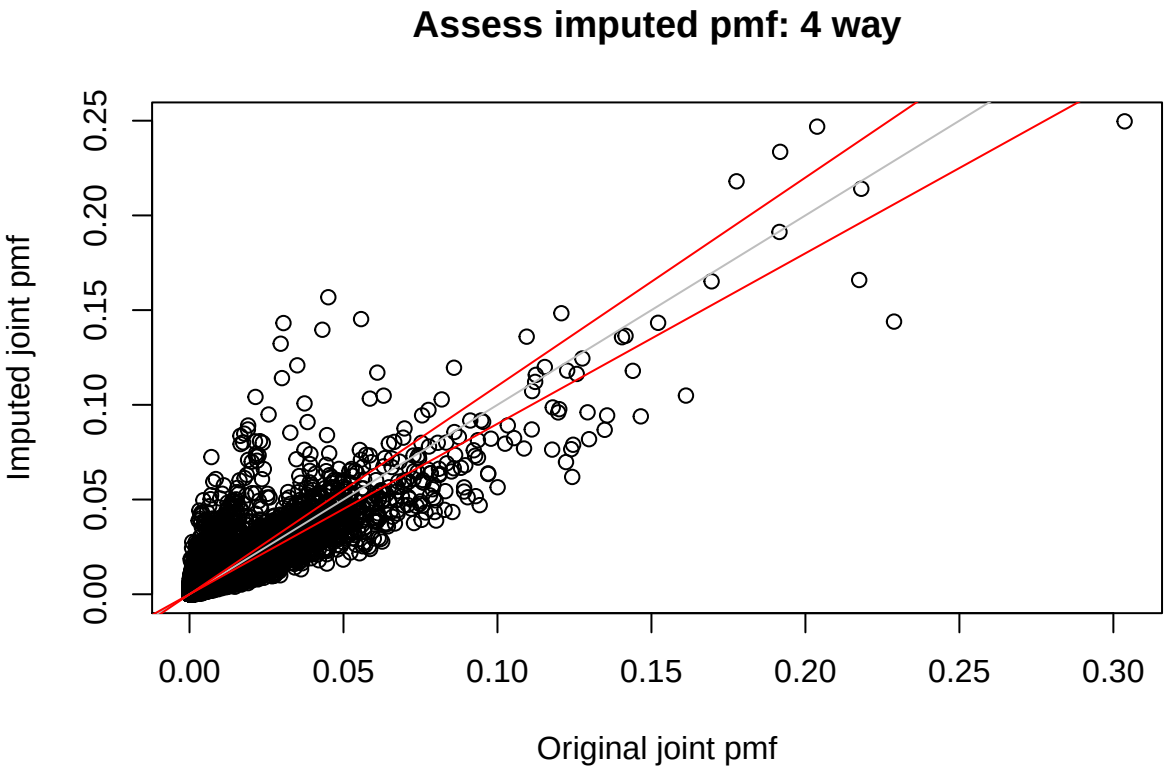

WKL
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
Category
original_pmf
observed_pmf
imputed_pmf


PINCP
original_pmf
observed_pmf
imputed_pmf
0.00
0.05
0.10
0.15
0.20
0.25
1
2
3
4
5
6
7
8
9
10
11
12
13
Category


Assess imputed pmf: 1 way
Imputed joint pmf
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Original joint pmf


Assess imputed pmf: 2 way
Imputed joint pmf
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Original joint pmf


Assess imputed pmf: 3 way
Imputed joint pmf
0.0
0.1
0.2
0.3
0.4
0.5
Original joint pmf


Assess imputed pmf: 4 way
Imputed joint pmf
0.00
0.05
0.10
0.15
0.20
0.25
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Original joint pmf