

MCAR 45% missing - RandomForest

```
# sample MCAR dataset from PUMS
source("../utils/sampleMCAR.R")
n = 10000
missing_col = c(1,3,7,9,10,11)
missing_prob = 0.45
set.seed(1)

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

missForest

```
# Perform missing data imputation and format output into list
imputation_list = list()
for (i in 1:5) {
  df.imp <- missForest(df_observed, verbose = FALSE)
  imputation_list[[i]] = df.imp$ximp
}
```

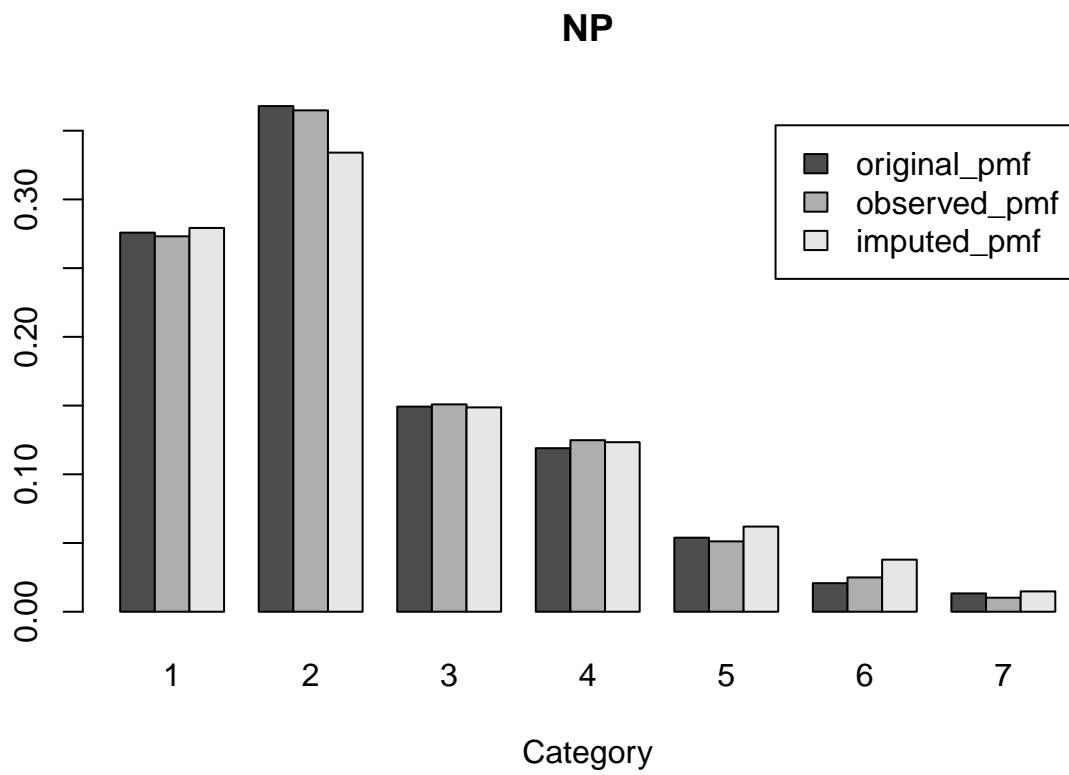
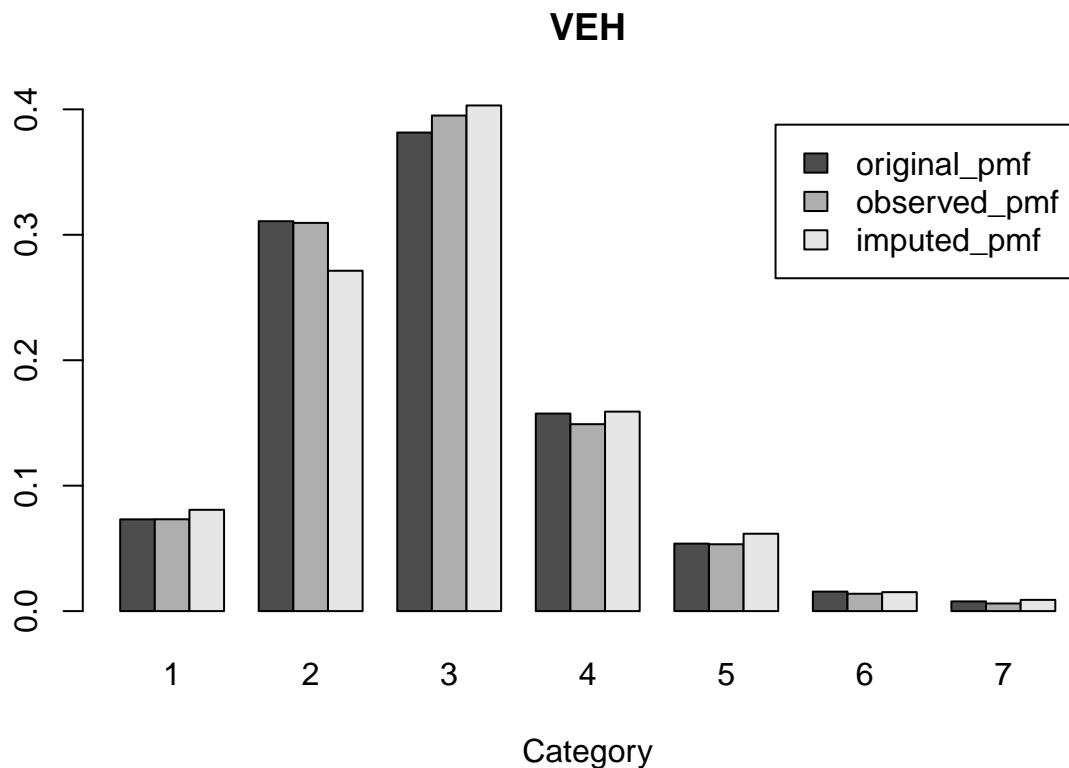
```
## missForest iteration 1 in progress...done!
## missForest iteration 2 in progress...done!
## missForest iteration 3 in progress...done!
## missForest iteration 4 in progress...done!
## missForest iteration 5 in progress...done!
## missForest iteration 6 in progress...done!
## missForest iteration 7 in progress...done!
## missForest iteration 8 in progress...done!
## missForest iteration 9 in progress...done!
## missForest iteration 10 in progress...done!
## missForest iteration 1 in progress...done!
## missForest iteration 2 in progress...done!
## missForest iteration 3 in progress...done!
## missForest iteration 4 in progress...done!
## missForest iteration 5 in progress...done!
## missForest iteration 6 in progress...done!
## missForest iteration 7 in progress...done!
## missForest iteration 1 in progress...done!
## missForest iteration 2 in progress...done!
## missForest iteration 3 in progress...done!
## missForest iteration 4 in progress...done!
## missForest iteration 5 in progress...done!
## missForest iteration 6 in progress...done!
## missForest iteration 7 in progress...done!
## missForest iteration 8 in progress...done!
## missForest iteration 9 in progress...done!
## missForest iteration 10 in progress...done!
## missForest iteration 1 in progress...done!
## missForest iteration 2 in progress...done!
```

```
## missForest iteration 3 in progress...done!
## missForest iteration 4 in progress...done!
## missForest iteration 5 in progress...done!
## missForest iteration 6 in progress...done!
## missForest iteration 7 in progress...done!
## missForest iteration 8 in progress...done!
## missForest iteration 9 in progress...done!
## missForest iteration 10 in progress...done!
## missForest iteration 1 in progress...done!
## missForest iteration 2 in progress...done!
## missForest iteration 3 in progress...done!
## missForest iteration 4 in progress...done!
## missForest iteration 5 in progress...done!
## missForest iteration 6 in progress...done!
## missForest iteration 7 in progress...done!
## missForest iteration 8 in progress...done!
## missForest iteration 9 in progress...done!
## missForest iteration 10 in progress...done!
```

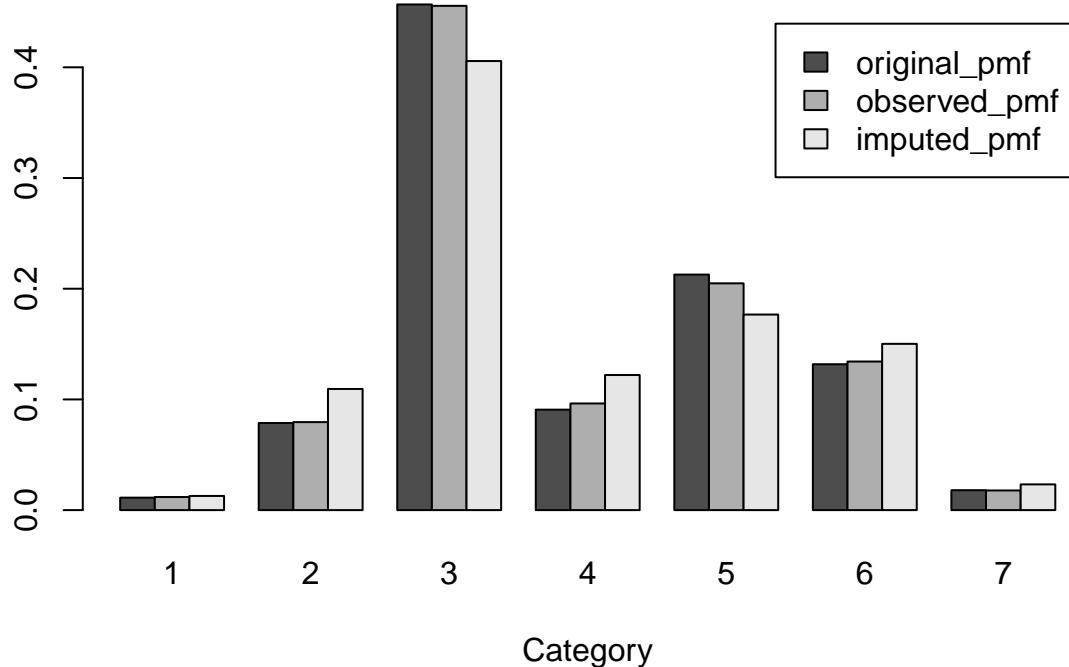
Diagnostics

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

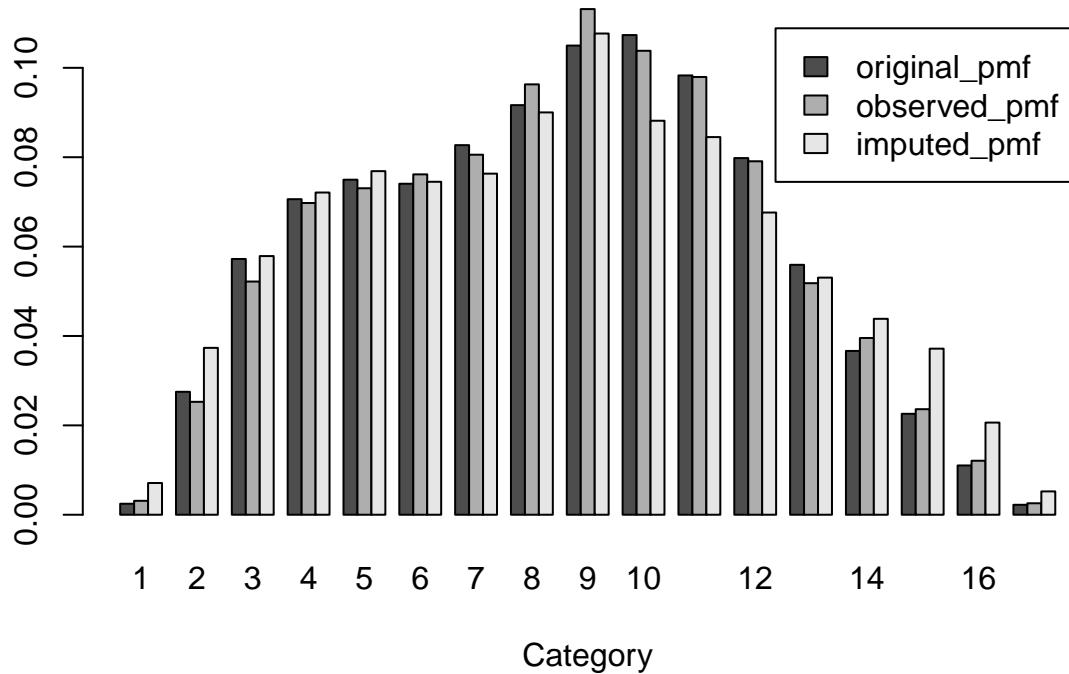
```
## ##### Coverage #####
## Coverage 1 way: 37.04 percent
## Coverage 2 way: 75.15 percent
## Coverage 3 way: 86.35 percent
## Coverage 4 way: 91.17 percent
##
## ##### RMSE #####
## RMSE 1 way: 0.016277
## RMSE 2 way: 0.004982
## RMSE 3 way: 0.001523
## RMSE 4 way: 0.000466
##
## ##### MAE #####
## MAE 1 way: 0.011725
## MAE 2 way: 0.002261
## MAE 3 way: 0.000475
## MAE 4 way: 0.000104
```



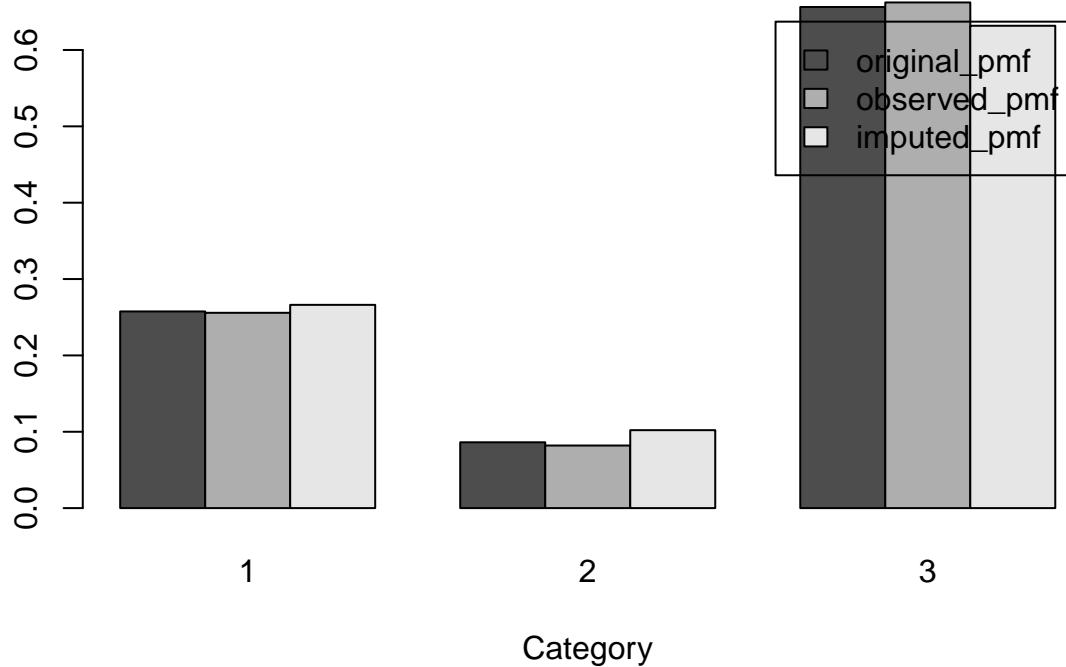
SCHL



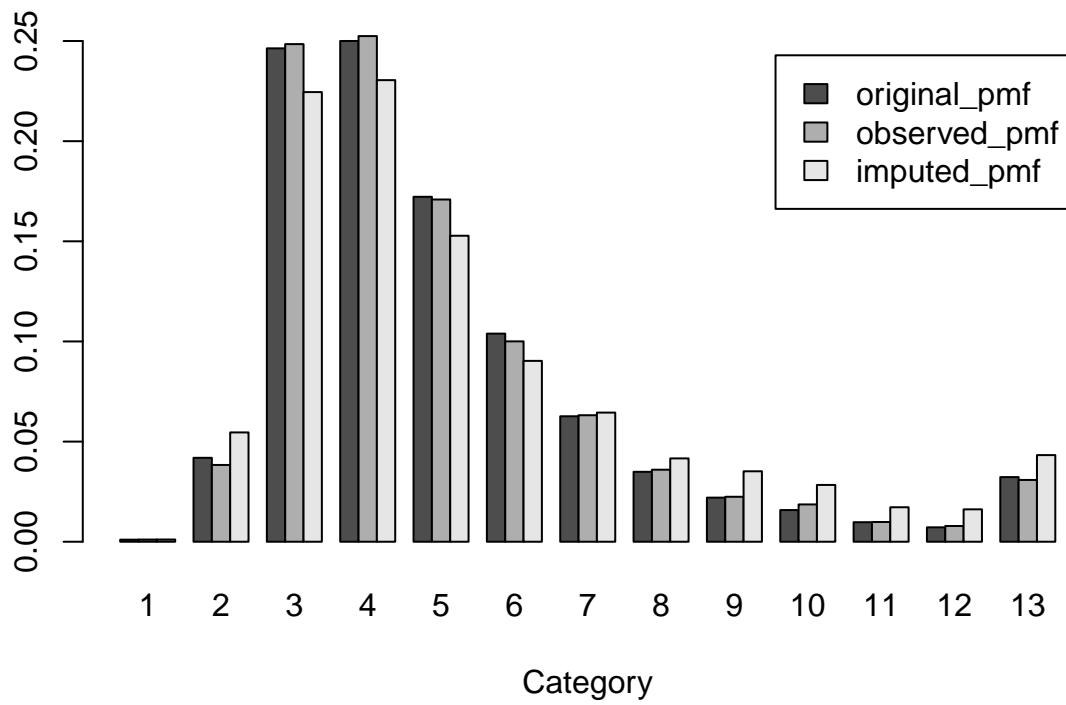
AGEP



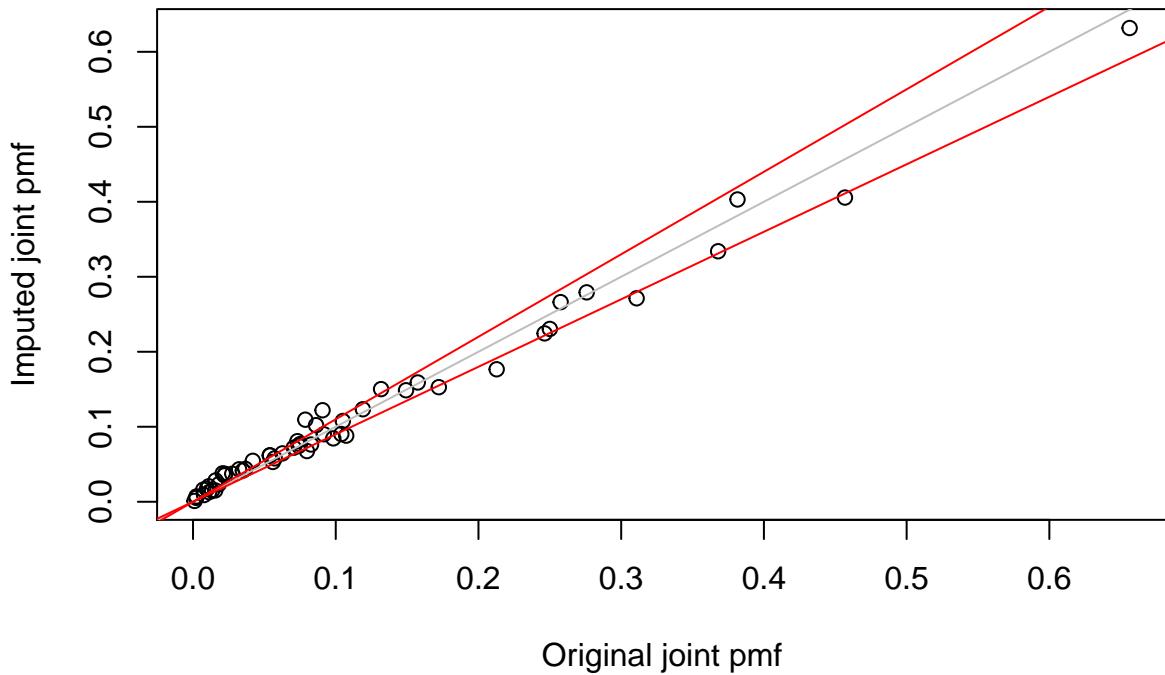
WKL



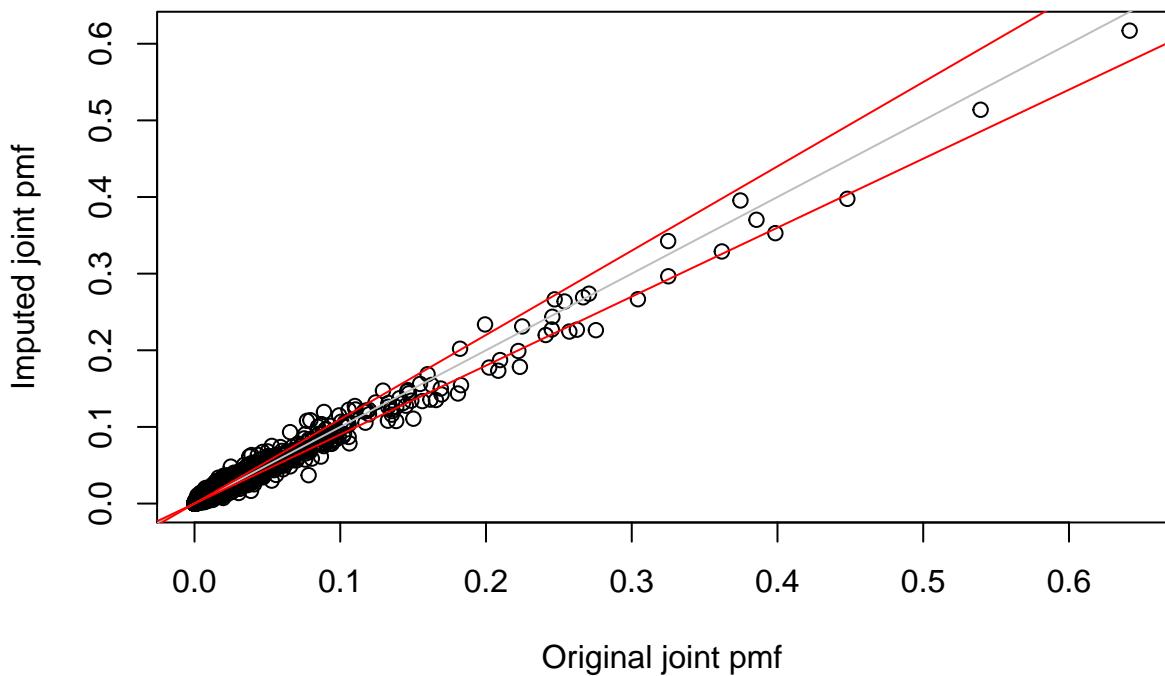
PINCP



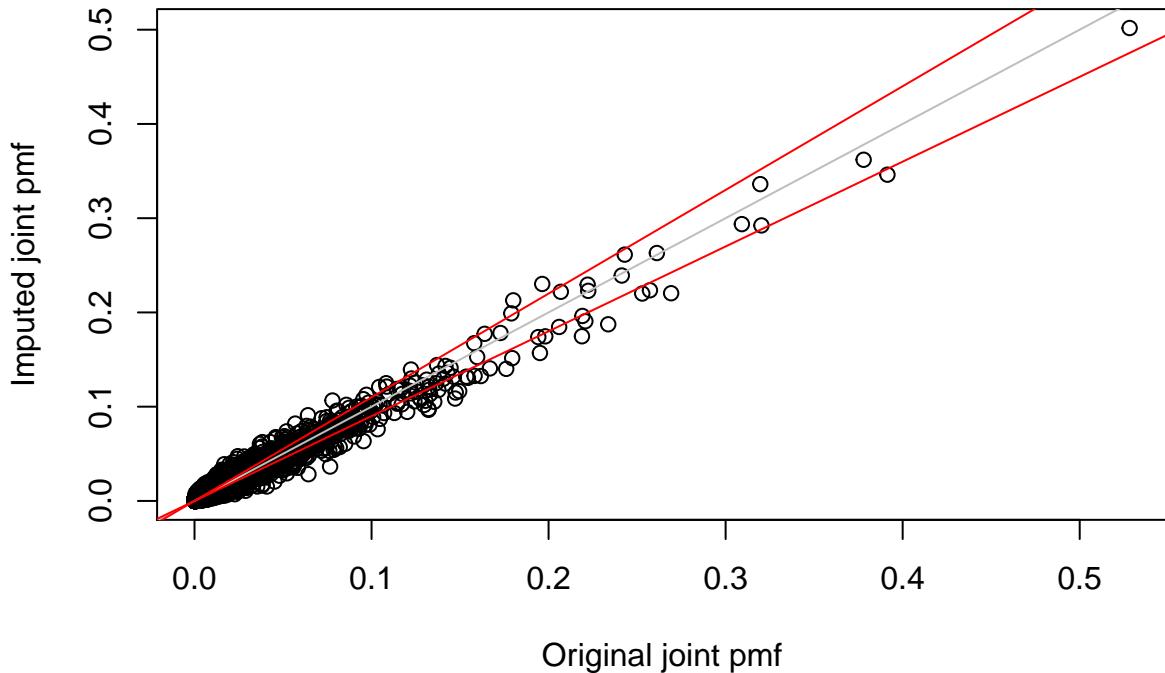
Assess imputed pmf: 1 way



Assess imputed pmf: 2 way



Assess imputed pmf: 3 way



Assess imputed pmf: 4 way

