

# MAR 45% missing - RandomForest

```
# sample MCAR dataset from PUMS
source("../utils/sampleMAR45.R")
n = 10000
missing_col = c(1,3,7,9,10,11)
set.seed(3)

output_list <- sampleMAR45(n)
df <- output_list[['df']]
df_observed <- output_list[['df_observed']]

apply(is.na(df_observed), MARGIN = 2, mean)

##      VEH      MV      NP    RMSP      ENG    MARHT    SCHL RACNUM     AGEP      WKL    PINCP
## 0.4456 0.0000 0.3998 0.0000 0.0000 0.0000 0.4842 0.0000 0.4670 0.4478 0.4384

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 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 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

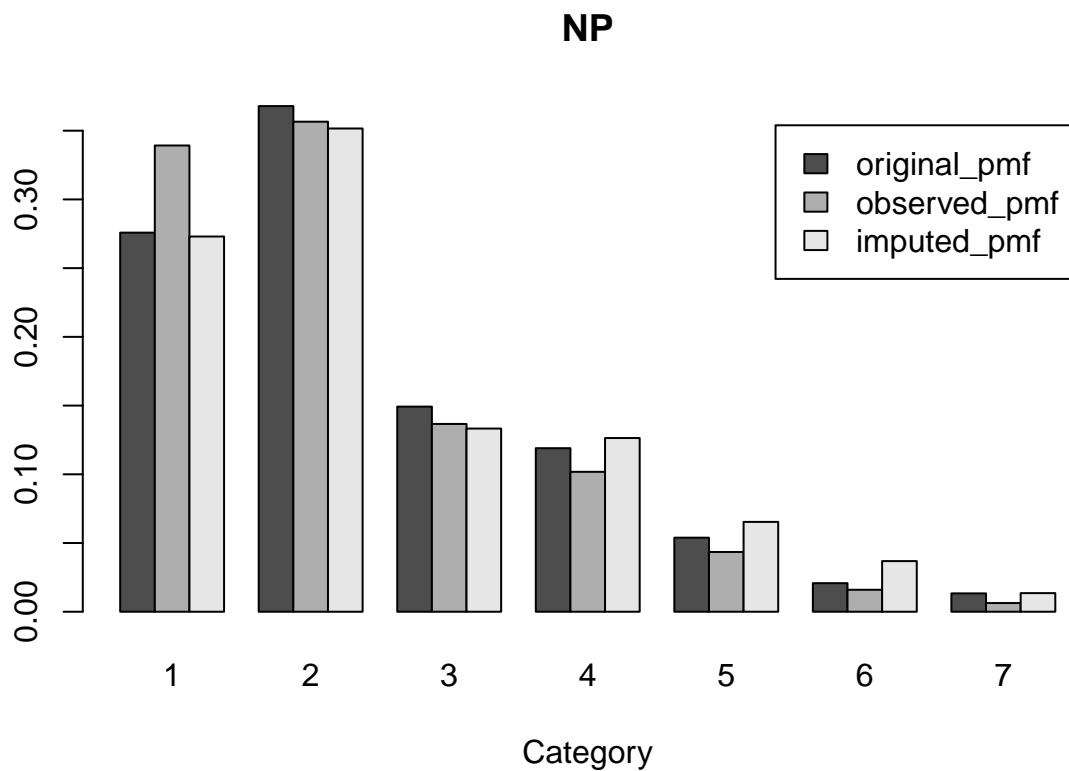
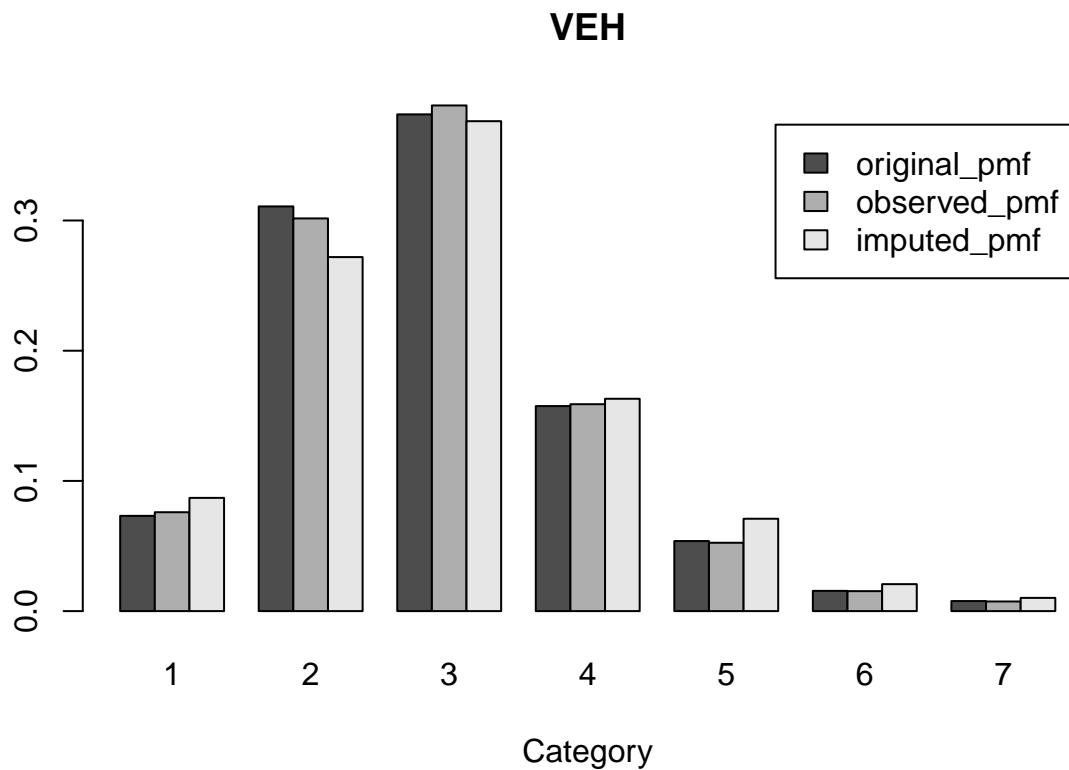
```

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

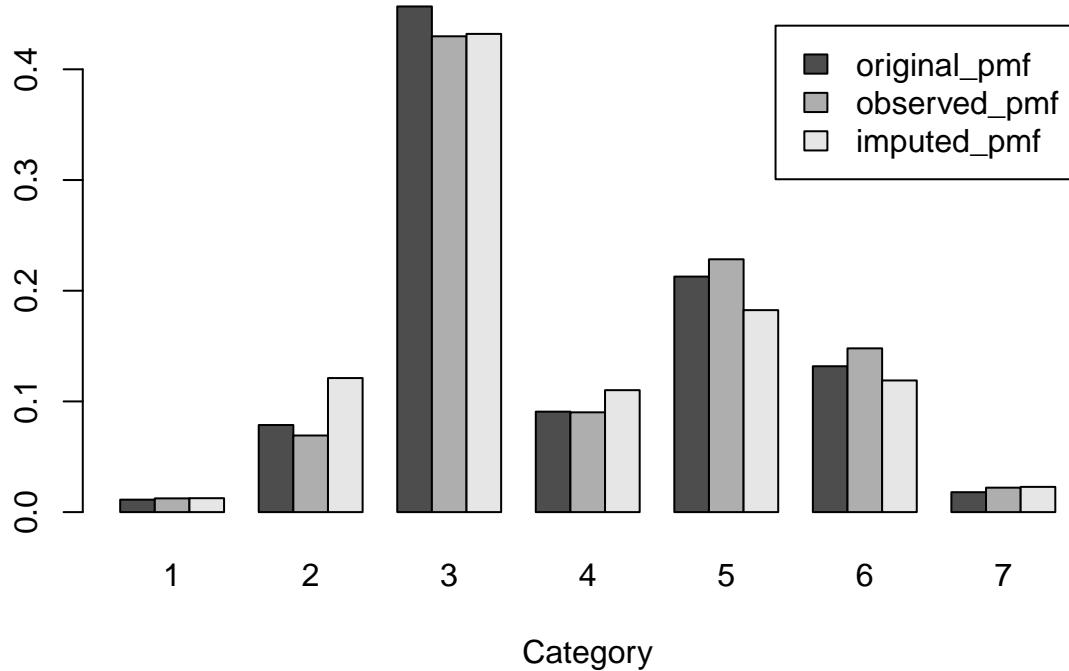
## ##### Coverage #####
## Coverage 1 way: 38.89 percent
## Coverage 2 way: 74.22 percent
## Coverage 3 way: 84.76 percent
## Coverage 4 way: 90.23 percent
##
## ##### RMSE #####
## RMSE 1 way: 0.017125
## RMSE 2 way: 0.005291
## RMSE 3 way: 0.00165
## RMSE 4 way: 0.000518
##
## ##### MAE #####
## MAE 1 way: 0.012956
## MAE 2 way: 0.00251
## MAE 3 way: 0.000528
## MAE 4 way: 0.000114

```

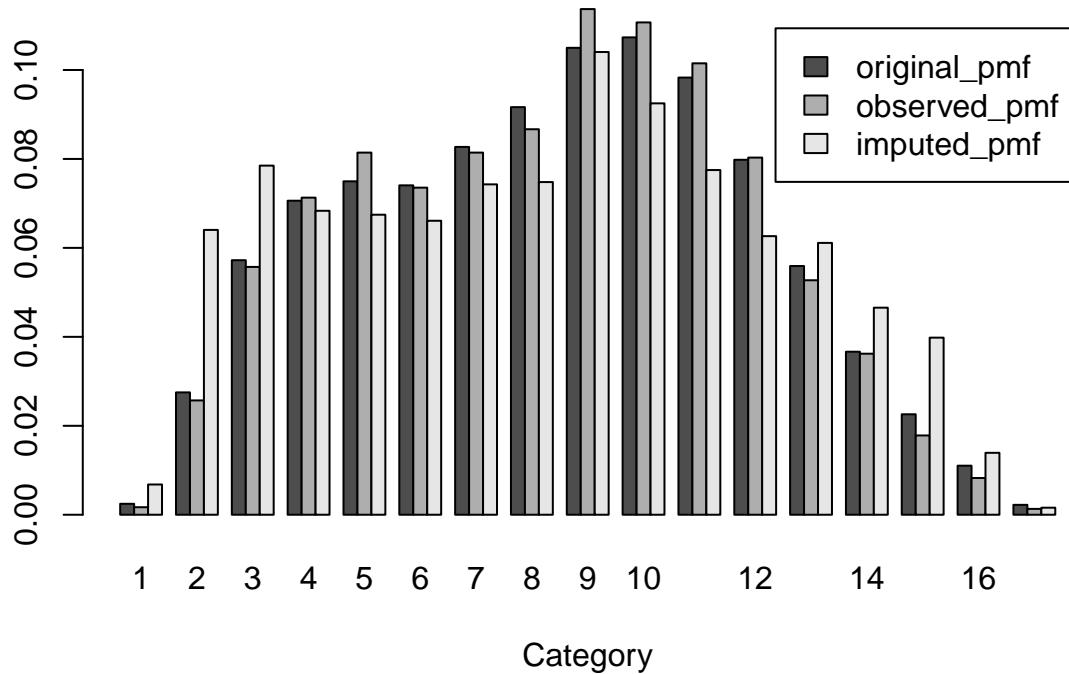
---



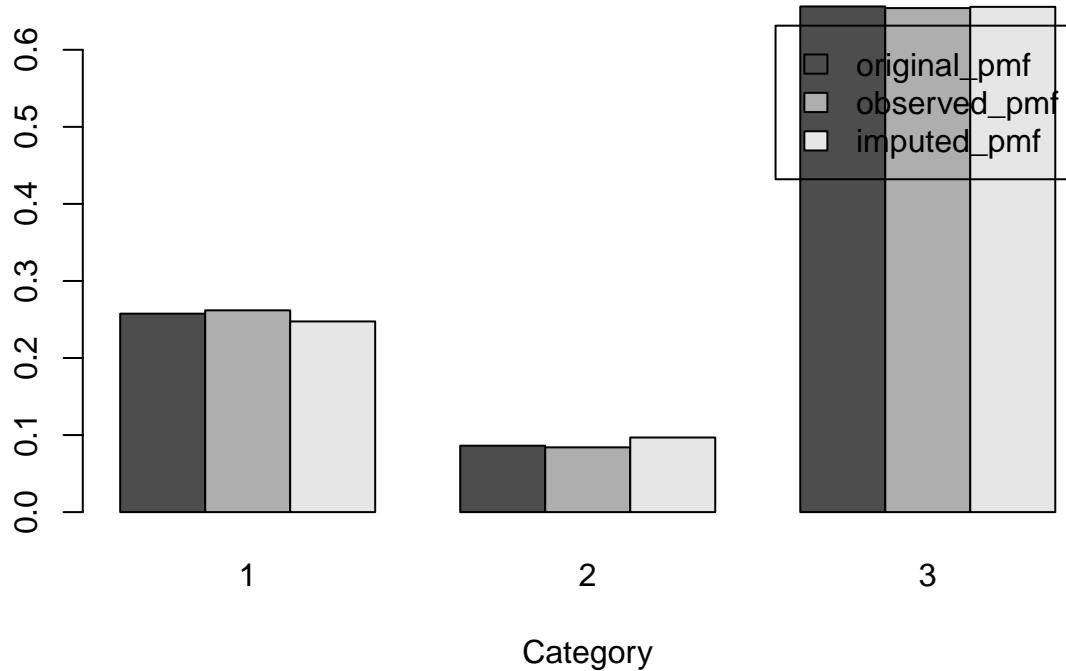
## SCHL



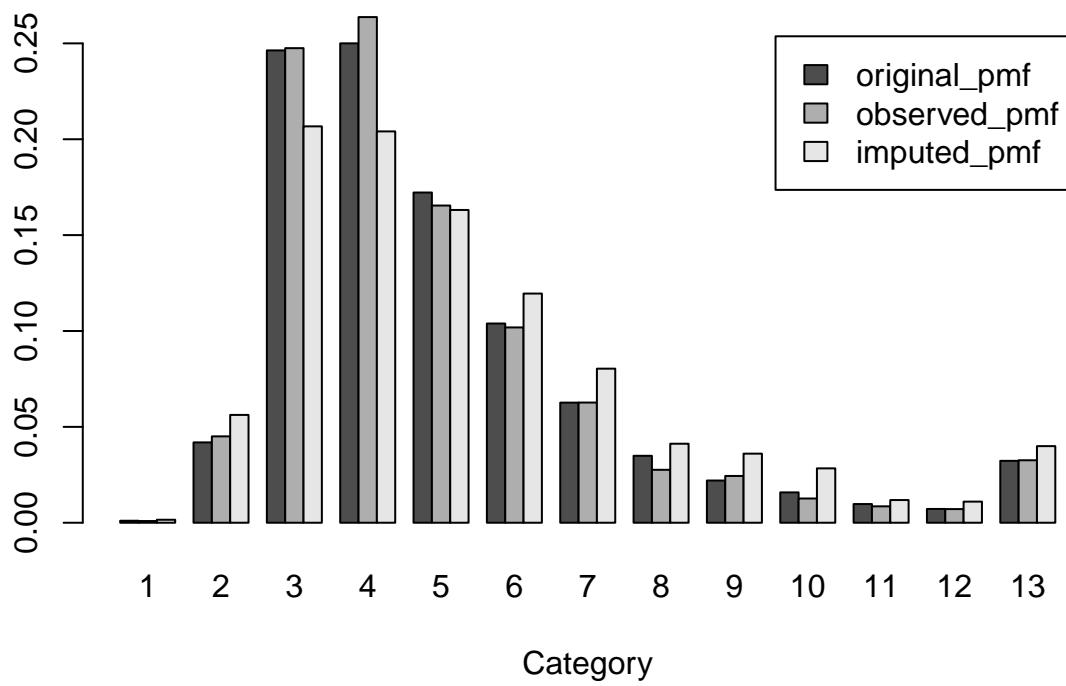
## AGEP



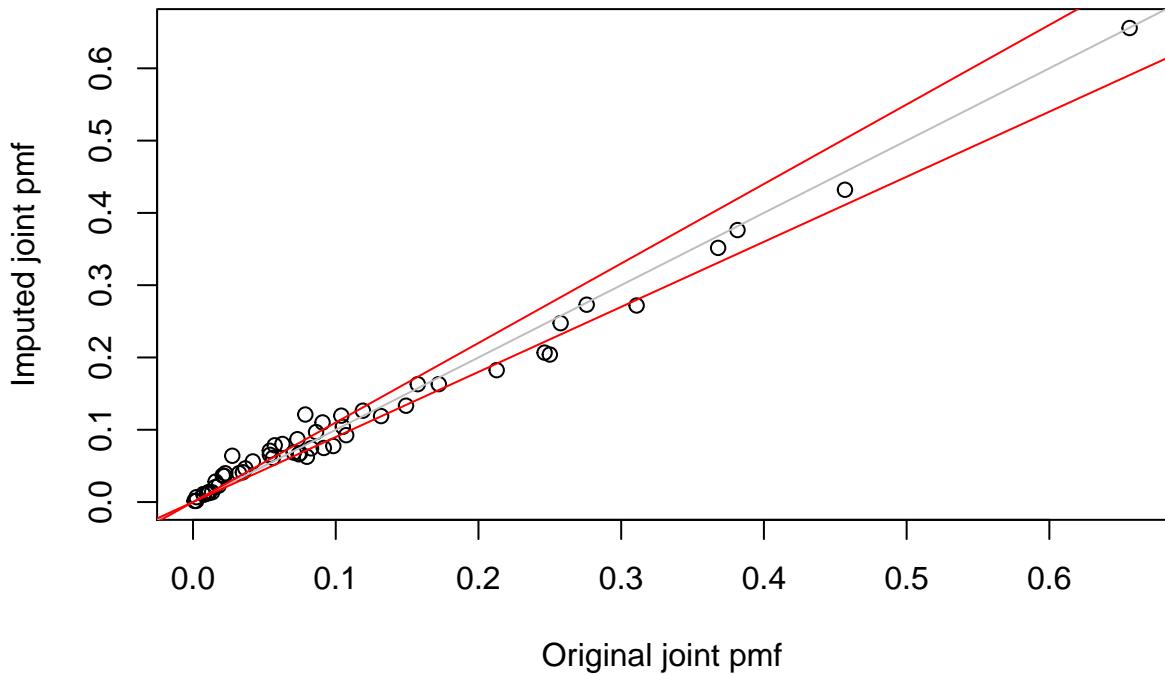
## WKL



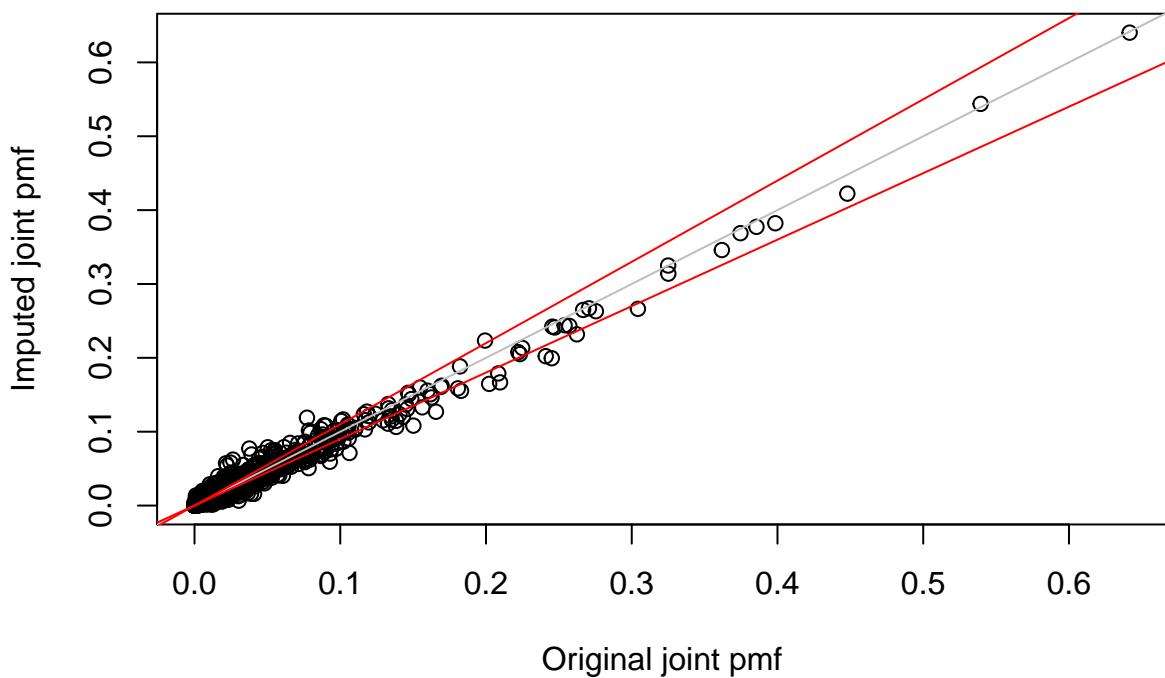
## PINCP



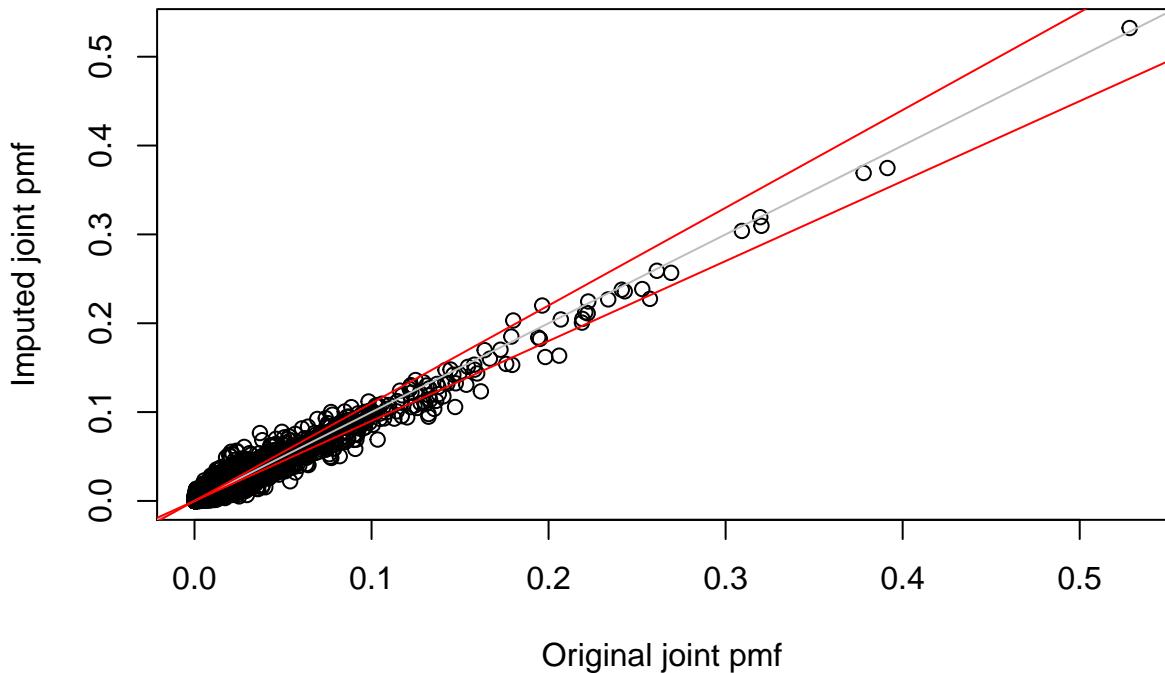
### Assess imputed pmf: 1 way



### Assess imputed pmf: 2 way



**Assess imputed pmf: 3 way**



**Assess imputed pmf: 4 way**

