

# MCAR 30% missing - Probit

```
# 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']]
```

## Ordinal bayesian nonparametric model

```
source("../probitBayes_by_cluster.R")
N = 40
Mon = 10
B = 10
thin.int = 2
start = Sys.time()
# function(y, N = 40, Mon = 2000, B = 300, thin.int = 5, seed = 0)
output_list <- probitBayesImputation(df_observed, N, Mon, B, thin.int)
Sys.time() - start

sampled_y <- output_list[['sampled_y']]
sampled_z <- output_list[['sampled_z']]

# extract 5 imputed dataset from probit model
imputation_index = as.integer(seq(1, dim(sampled_y)[1], length.out = 5))
imputation_list = list()

levels = c(7,7,7,19,5,4,7,2,17,3,13)
for (i in 1:length(imputation_index)) {
  index = imputation_index[i]
  # need to plus 1 here because the class index of DP function starts at 0
  d = sampled_y[index,,]
  d = data.frame(d)
  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

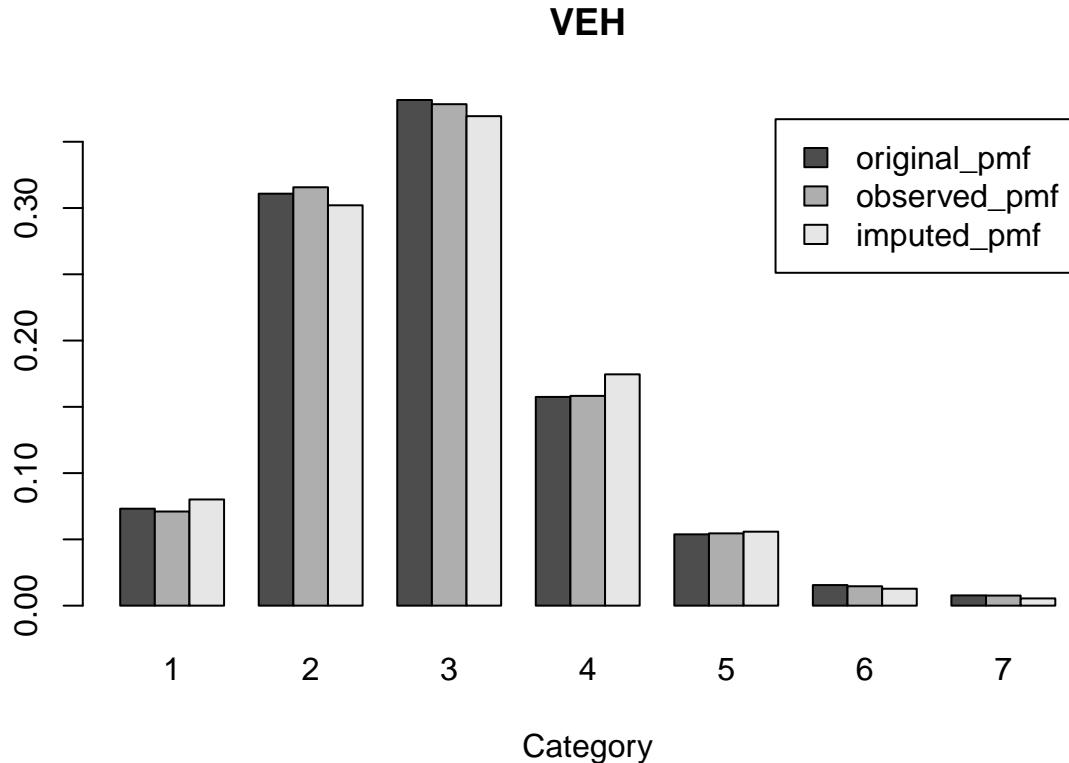
```
source("../utils/create_report.R")
create_report(imputation_list, max_nway=4, missing_col, df_observed)
```

```
## ##### Coverage #####
```

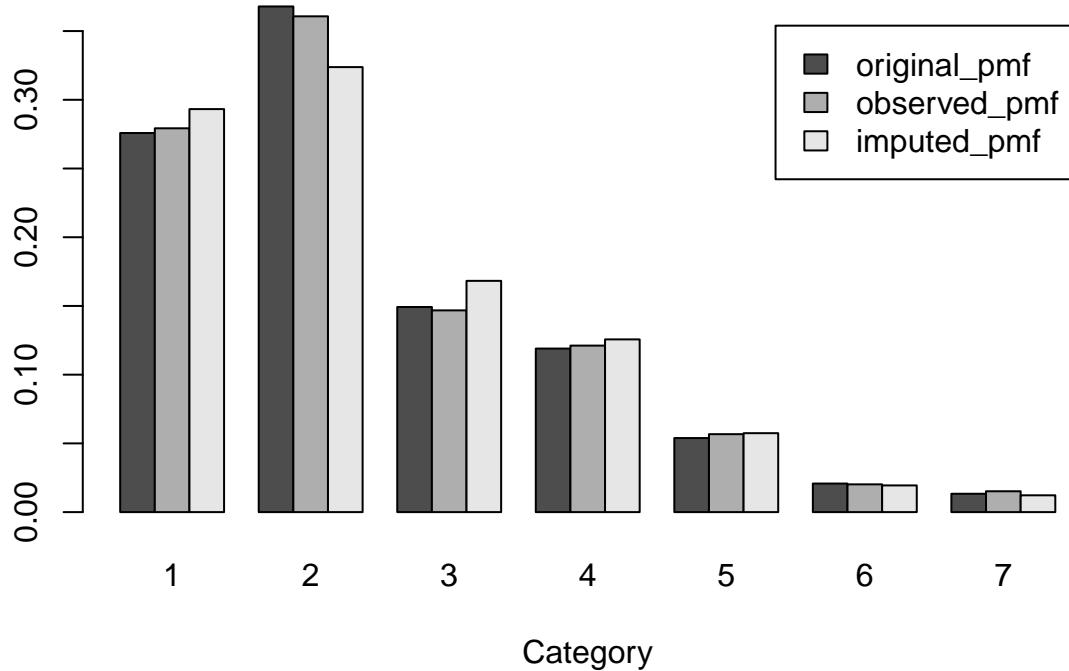
```

## Coverage 1 way: 37.04 percent
## Coverage 2 way: 78.43 percent
## Coverage 3 way: 92.81 percent
## Coverage 4 way: 97.26 percent
##
## ##### RMSE #####
## RMSE 1 way: 0.018005
## RMSE 2 way: 0.004875
## RMSE 3 way: 0.001333
## RMSE 4 way: 0.000365
##
## ##### MAE #####
## MAE 1 way: 0.011146
## MAE 2 way: 0.0019
## MAE 3 way: 0.000363
## MAE 4 way: 7.5e-05

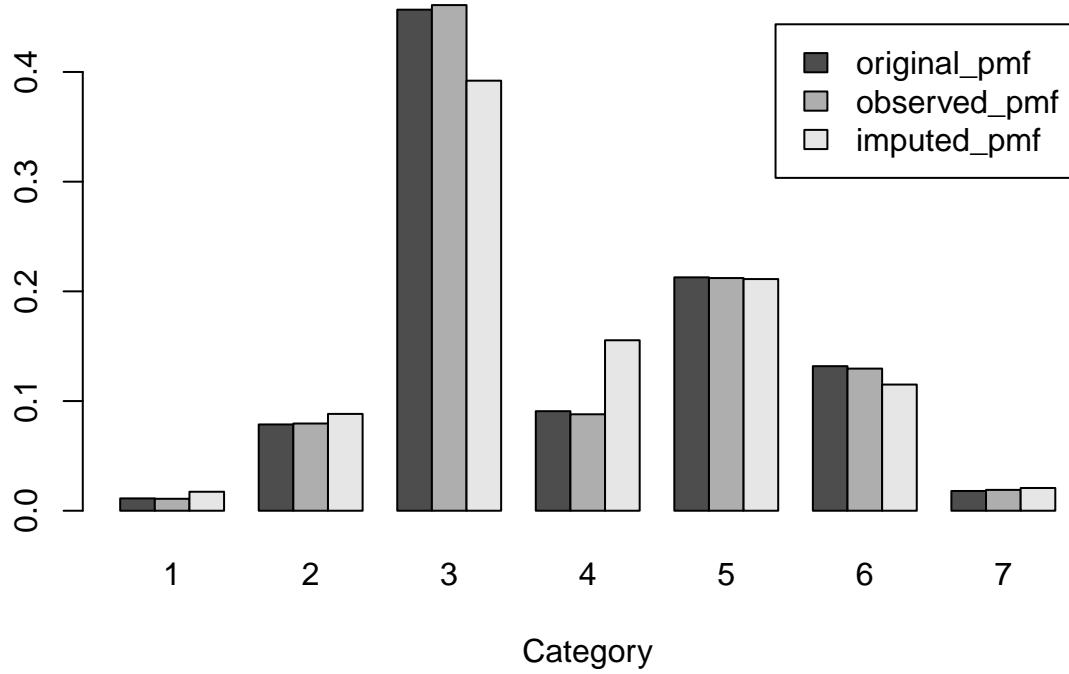
```



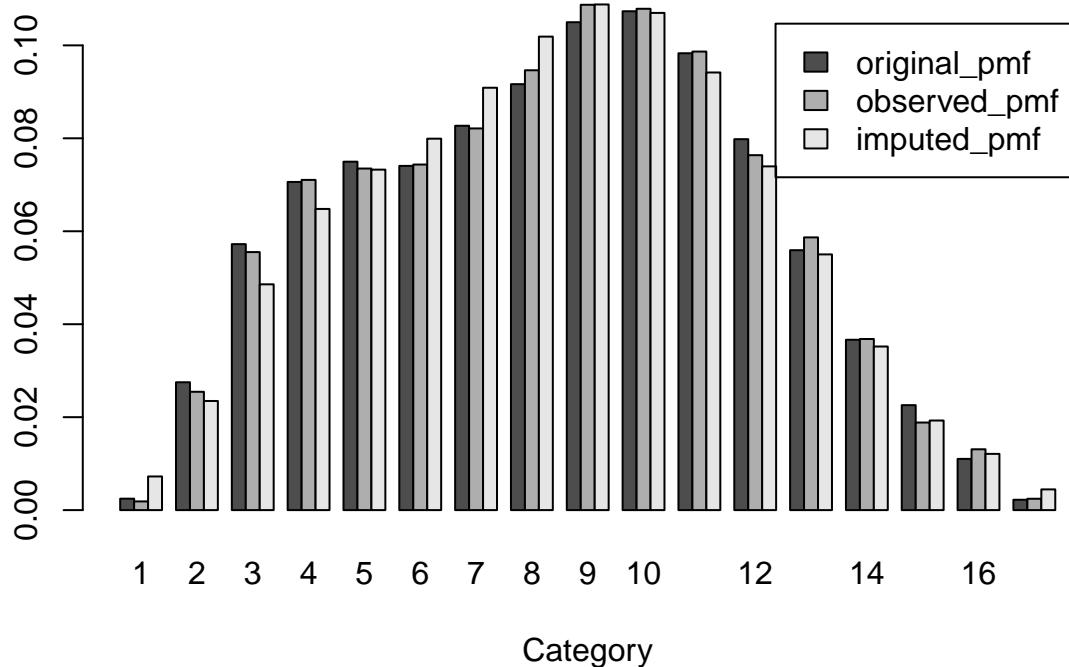
## NP



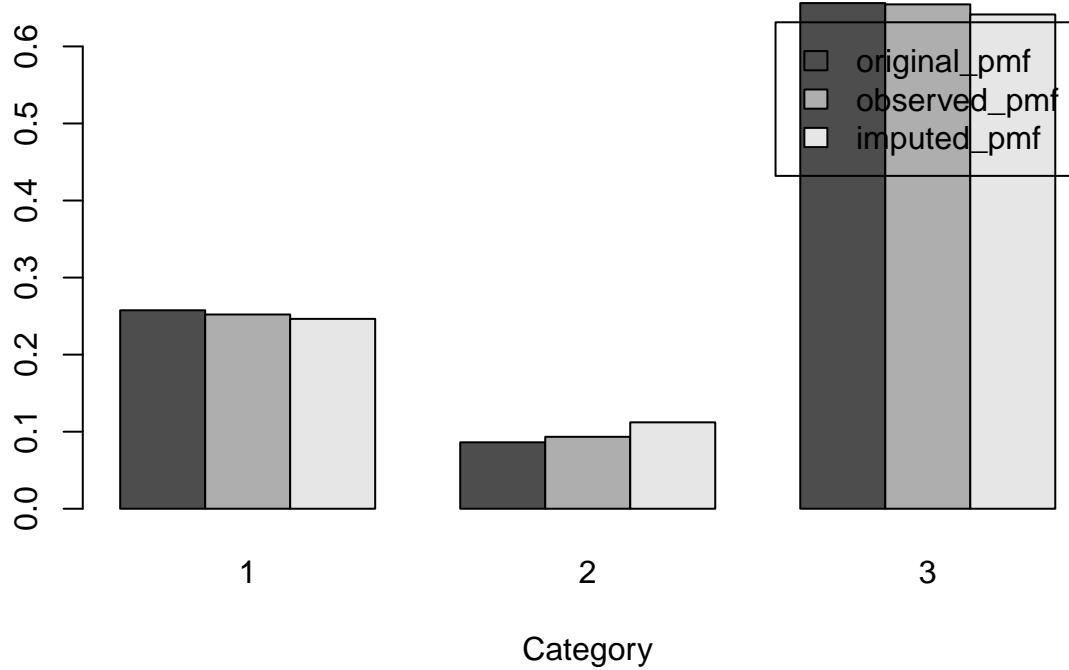
## SCHL



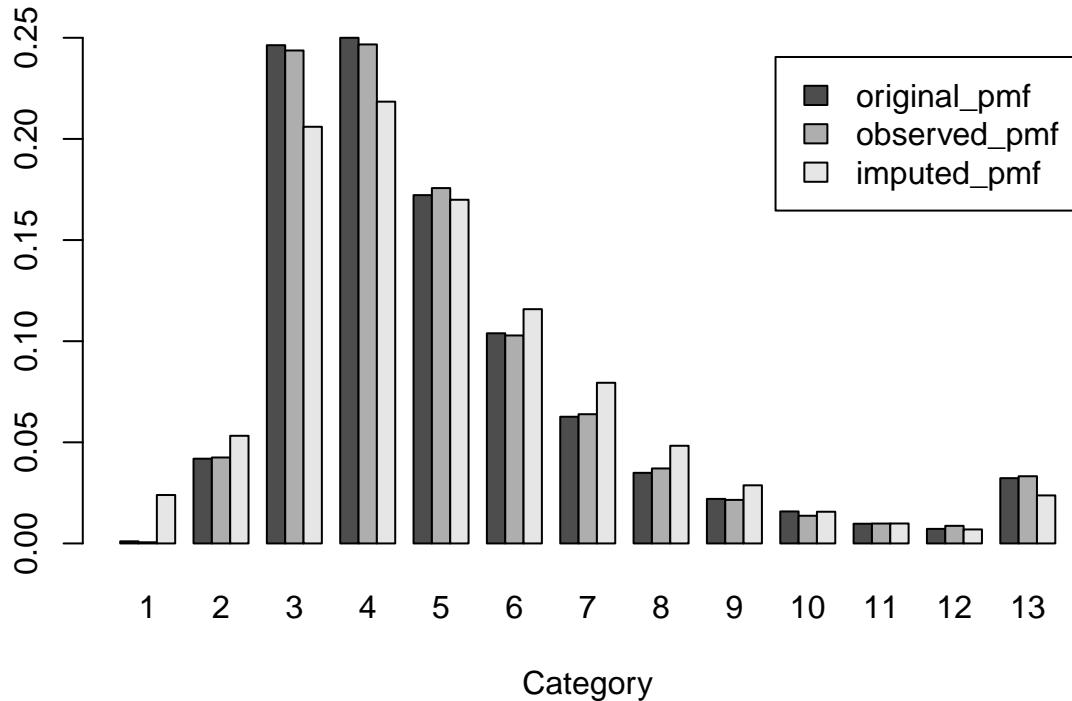
## AGEP



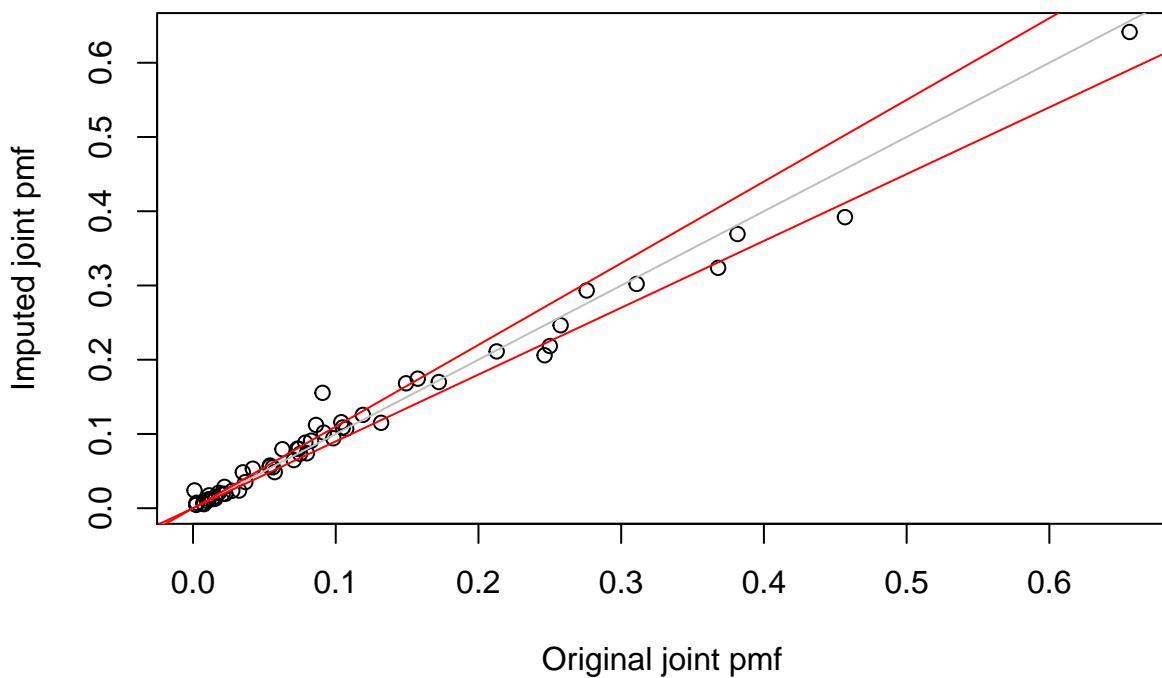
## WKL



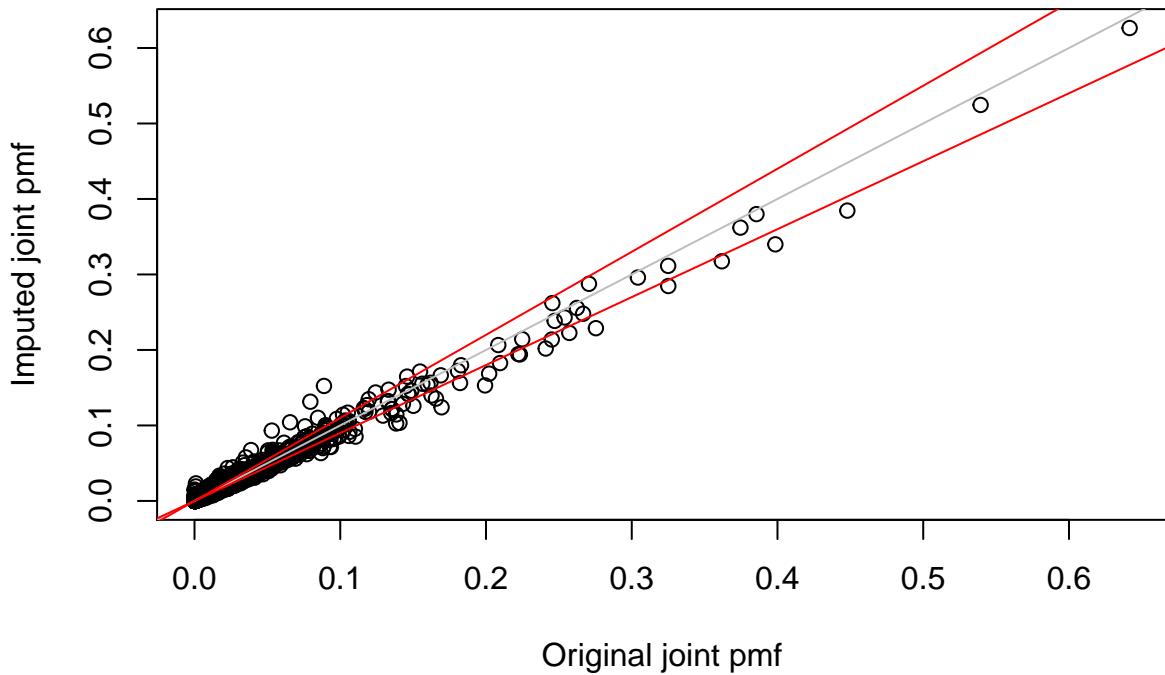
## PINCP



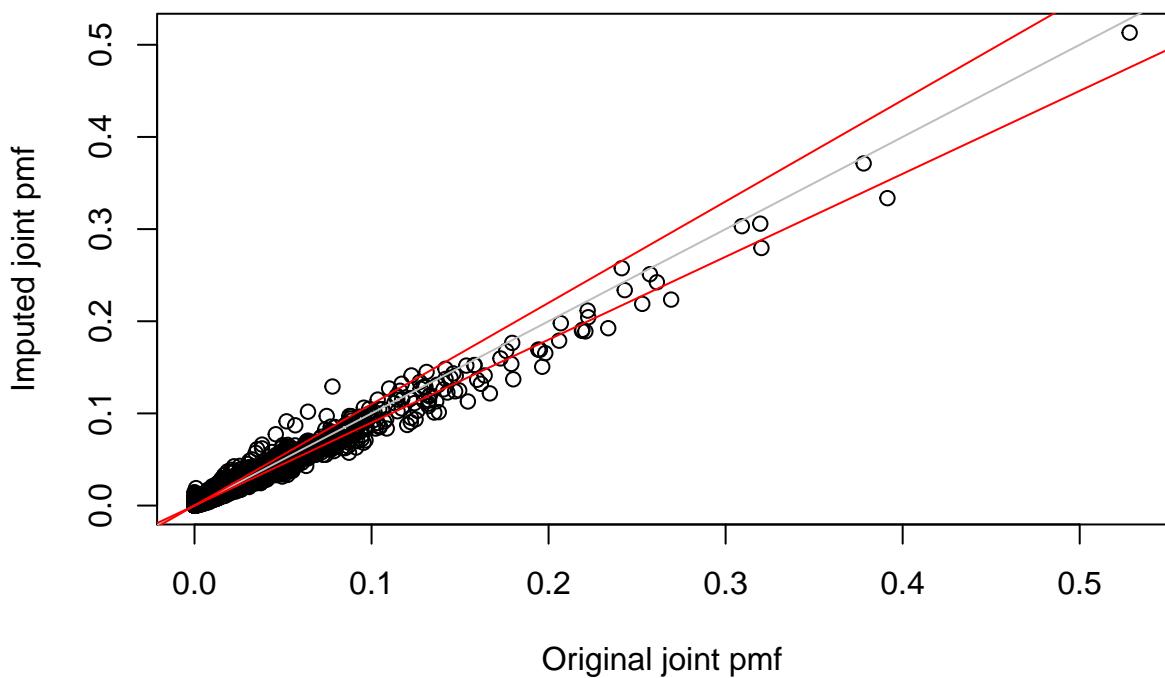
## Assess imputed pmf: 1 way



**Assess imputed pmf: 2 way**



**Assess imputed pmf: 3 way**



### Assess imputed pmf: 4 way

