Testing different imputation methods on PUMS (MCAR) - MICE

```r
# load dataset: df
load('../Datasets/ordinalPUMS.Rdata')

# take 10,000 samples: df
n = 10000
sample <- sample(nrow(df), size = 10000)
df <- df[sample,]

# create MCAR scneario with 30% chance of missing: df_observed
set.seed(0)
missing_prob = 0.3
df_observed <- df
missing_col = colnames(df)[c(1,3,5,7,9,11)]
for (col in missing_col) {
  missing_ind <- rbernoulli(n,p = missing_prob)
  df_observed[missing_ind, col] <- NA
}
```

MICE

Create 5 imputed dataset

```r
library(mice)
```

```
##
## Attaching package: 'mice'
```

```
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```r
imputed_df <-  mice(df_observed,m=5,print=F)
```

```
## Warning: Number of logged events: 150
```

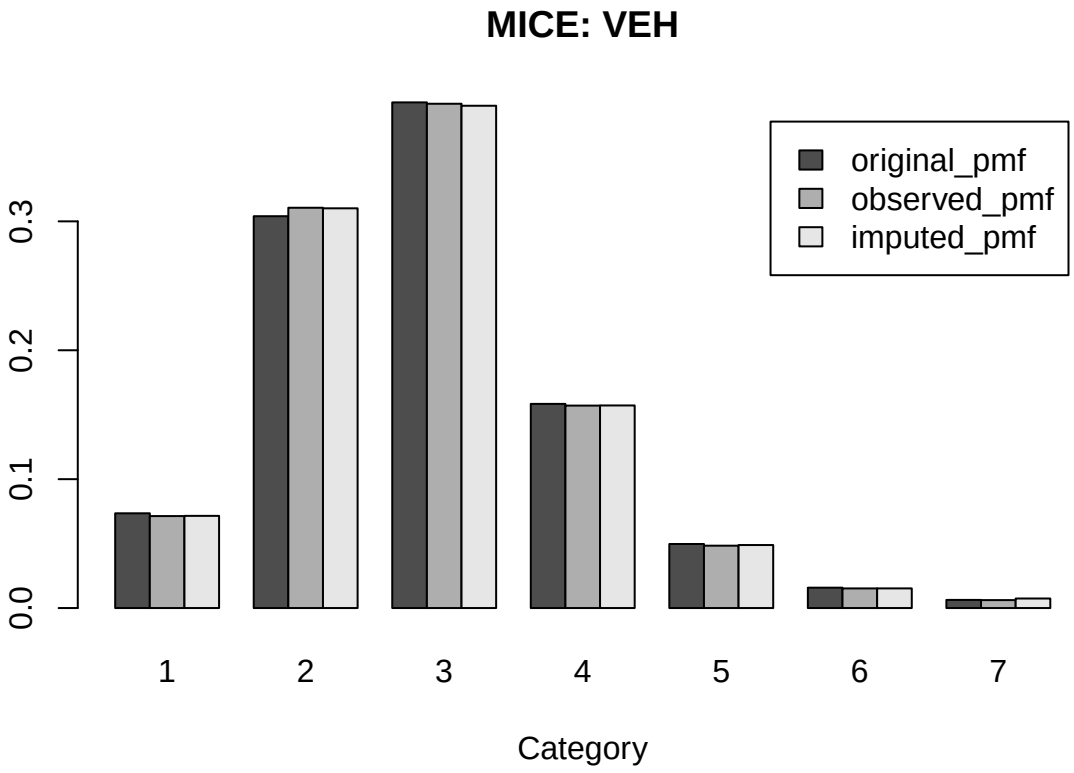
Extract the 5 imputed dataset

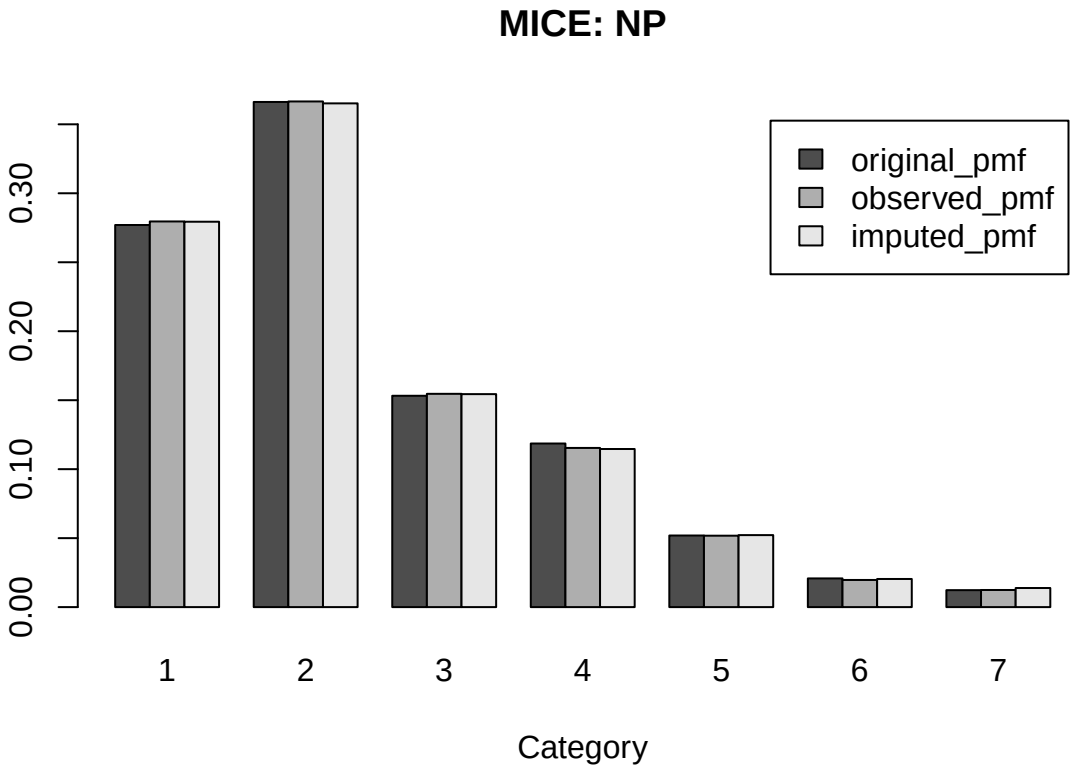
```r
d1 <-  complete(imputed_df, 1)
d2 <-  complete(imputed_df, 2)
d3 <-  complete(imputed_df, 3)
d4 <-  complete(imputed_df, 4)
d5 <-  complete(imputed_df, 5)
```

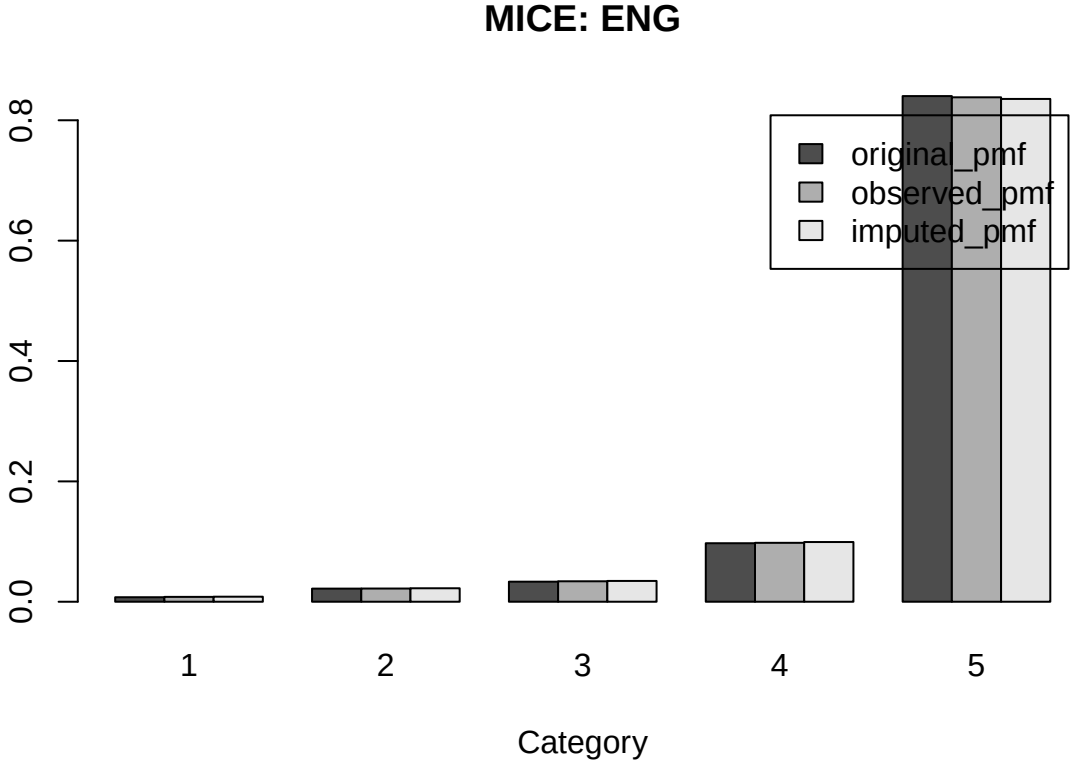
Diagnostics

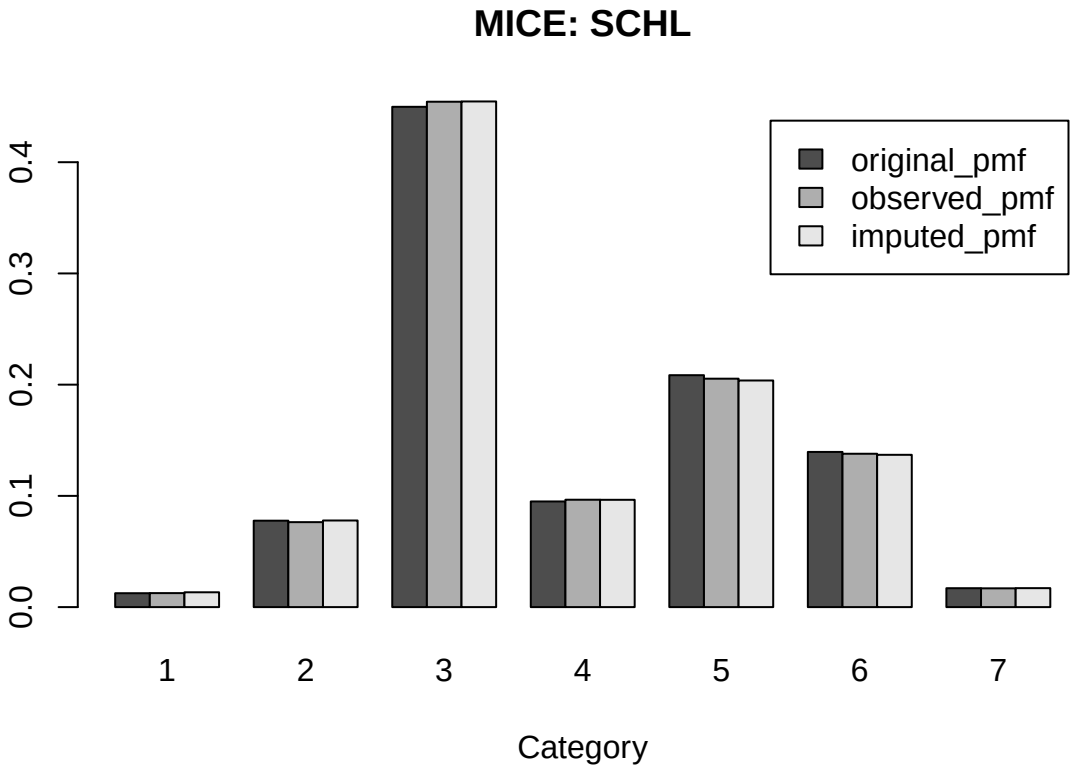
```r
for (var_index in c(1,3,5,7,9,11)) {
  y_original = df[,var_index]
  original_pmf = table(y_original)/length(y_original)

  # Observed distribution
  missing_indicator = is.na(df_observed)[,var_index]
```

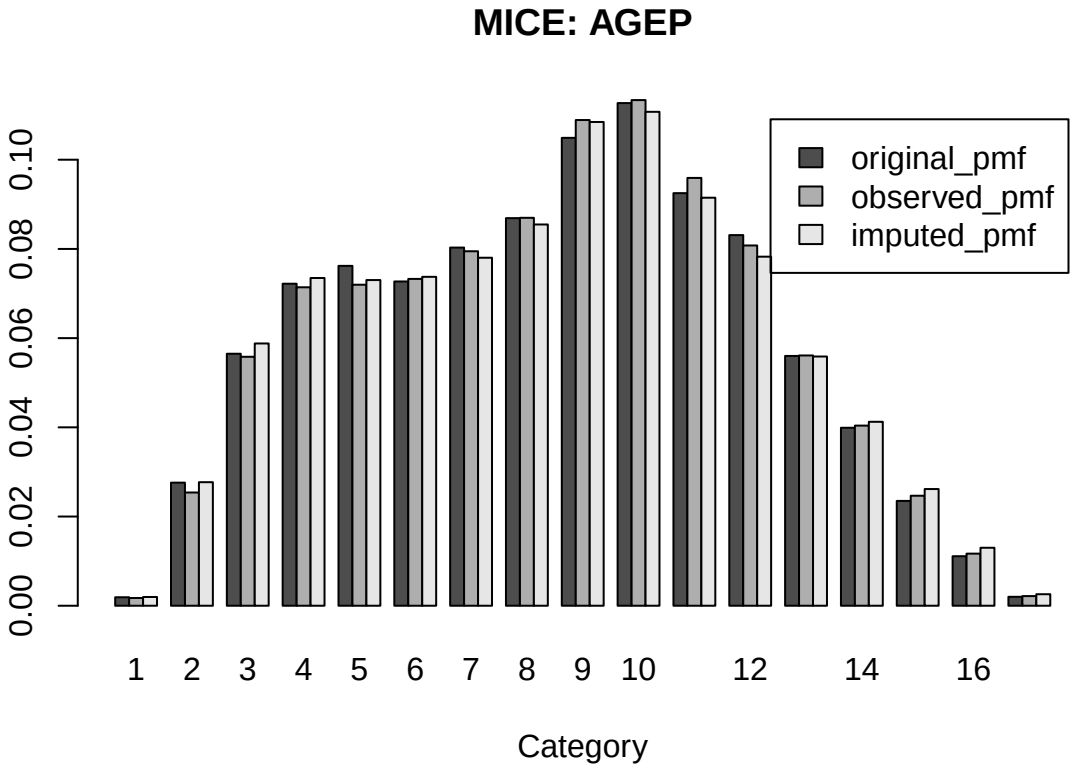
```
  y_observed = y_original[!missing_indicator]
  observed_pmf = table(y_observed)/length(y_observed)

  # Extract variable from imputed data
  sample_estimate1 = table(d1[,var_index])/length(d1[,var_index])
  sample_estimate2 = table(d2[,var_index])/length(d2[,var_index])
  sample_estimate3 = table(d3[,var_index])/length(d3[,var_index])
  sample_estimate4 = table(d4[,var_index])/length(d4[,var_index])
  sample_estimate5 = table(d5[,var_index])/length(d5[,var_index])

  imputed_pmf = (sample_estimate1 + sample_estimate2 + sample_estimate3 +
                   sample_estimate4 + sample_estimate5)/5

  results = rbind(original_pmf,observed_pmf,imputed_pmf)
  colnames(results)<- 1:dim(imputed_pmf)
  barplot(results, xlab = 'Category', beside = TRUE,
          legend = TRUE,
          main = paste('MICE:', colnames(df)[var_index]))
}
```

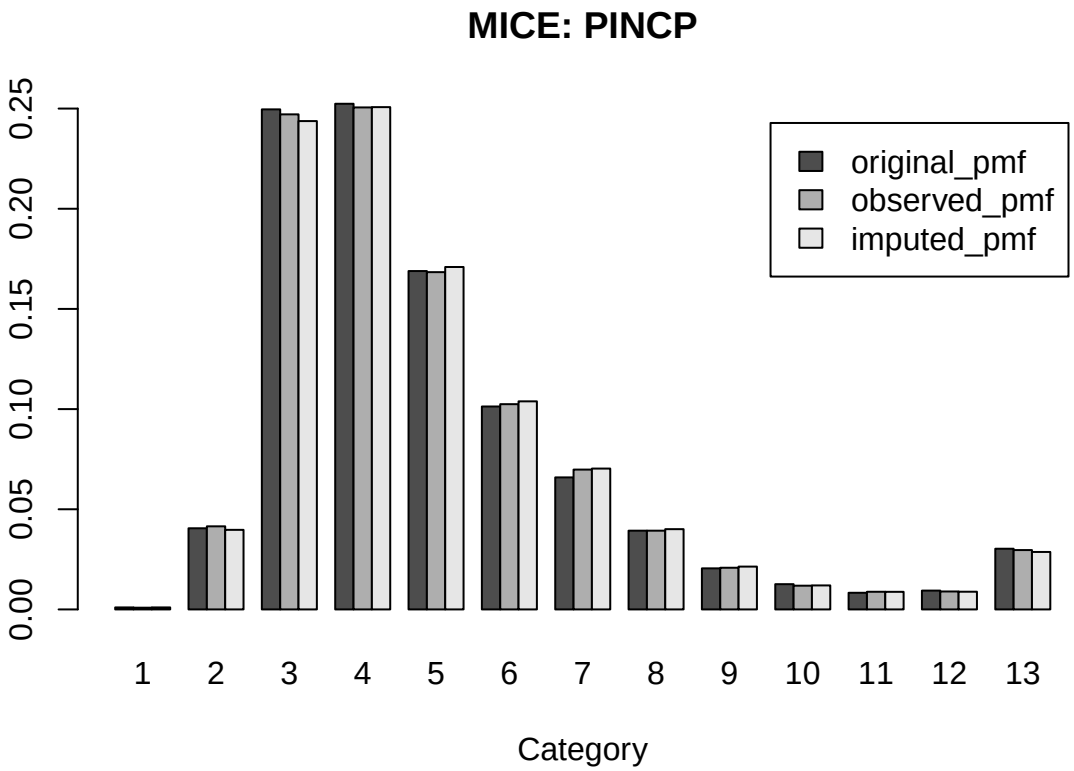



MICE: NP
original_pmf
observed_pmf
imputed_pmf
0.00
0.10
0.20
0.30
1
2
3
4
5
6
7
Category


MICE: ENG
original_pmf
observed_pmf
imputed_pmf
0.0
0.2
0.4
0.6
0.8
1
2
3
4
5
Category


MICE: 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


MICE: AGEP
original_pmf
observed_pmf
imputed_pmf
0.00
0.02
0.04
0.06
0.08
0.10
1
2
3
4
5
6
7
8
9
10
12
14
16
Category


MICE: 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