

# Testing different imputation methods on PUMS (MCAR) - RandomForest

---

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

# take 10,000 samples: df
set.seed(0)
n = 10000
sample <- sample(nrow(df), size = 10000)
df <- df[sample,]

# create MCAR scenario with 30% chance of missing: df_observed
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
}
```

## missForest

```
df.imp <- missForest(df_observed, verbose = FALSE)
d1 <- df.imp$ximpx
df.imp <- missForest(df_observed, verbose = FALSE)
d2 <- df.imp$ximpx
df.imp <- missForest(df_observed, verbose = FALSE)
d3 <- df.imp$ximpx
df.imp <- missForest(df_observed, verbose = FALSE)
d4 <- df.imp$ximpx
df.imp <- missForest(df_observed, verbose = FALSE)
d5 <- df.imp$ximpx
```

## Diagnostics

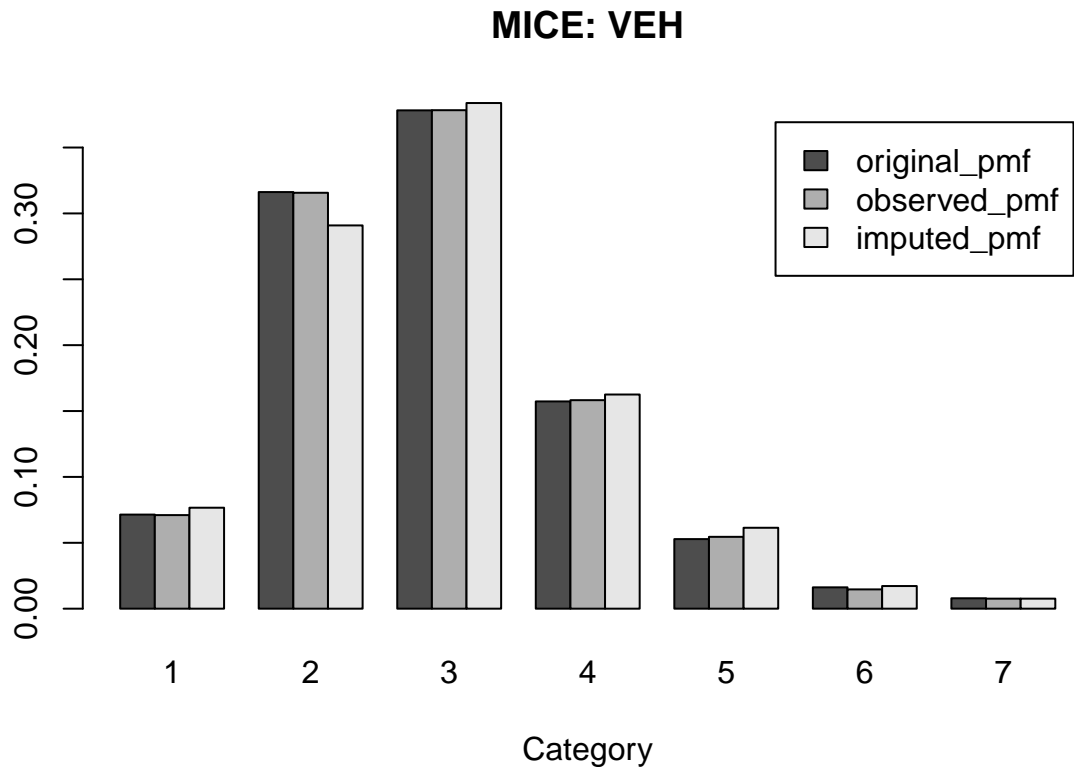
```
for (var_index in c(1,3,5,7,9,11)) {
  y_original = df[,var_index]
  original_pmf = table(y_original)/length(y_original)

  # Observed distribution
  missing_indicator = is.na(df_observed)[,var_index]
  y_observed = y_original[!missing_indicator]
  observed_pmf = table(y_observed)/length(y_observed)

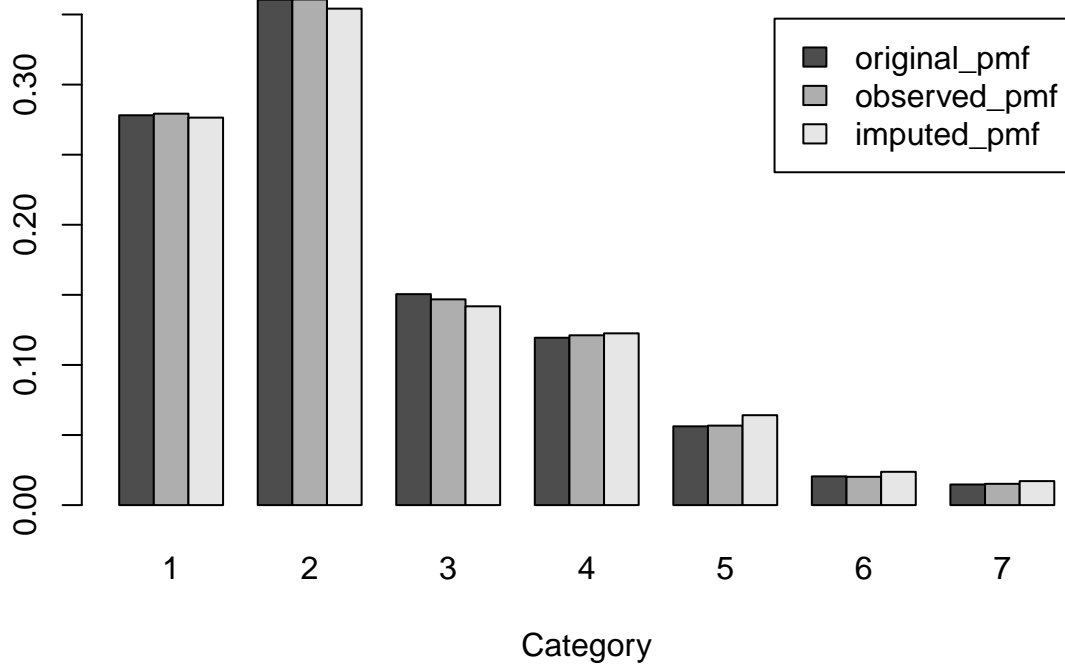
  # Extract variable from imputed data
  sample_estimate1 = table(d1[,var_index])/length(d1[,var_index])
  sample_estimate2 = table(d2[,var_index])/length(d2[,var_index])
  sample_estimate3 = table(d3[,var_index])/length(d3[,var_index])
  sample_estimate4 = table(d4[,var_index])/length(d4[,var_index])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

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

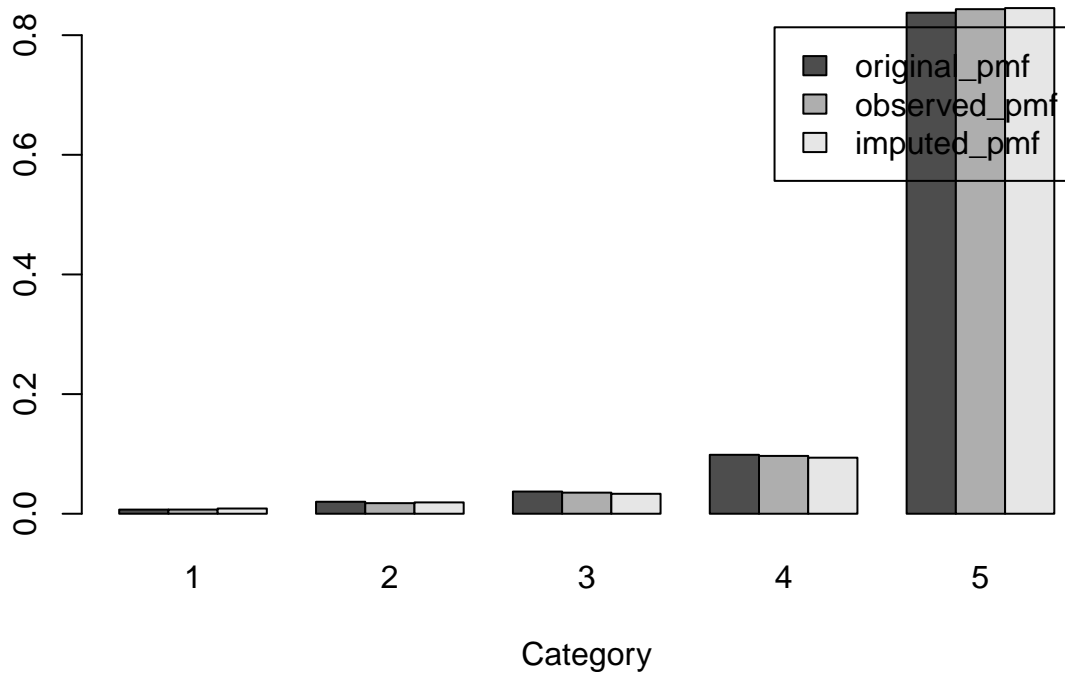
results = rbind(original_pmf,observed_pmf,imputed_pmf)
colnames(results)<- 1:dim(imputed_pmf)
barplot(results, xlab = 'Category', beside = TRUE,
        legend = TRUE,
        main = paste('MICE:', colnames(df)[var_index]))
}
```



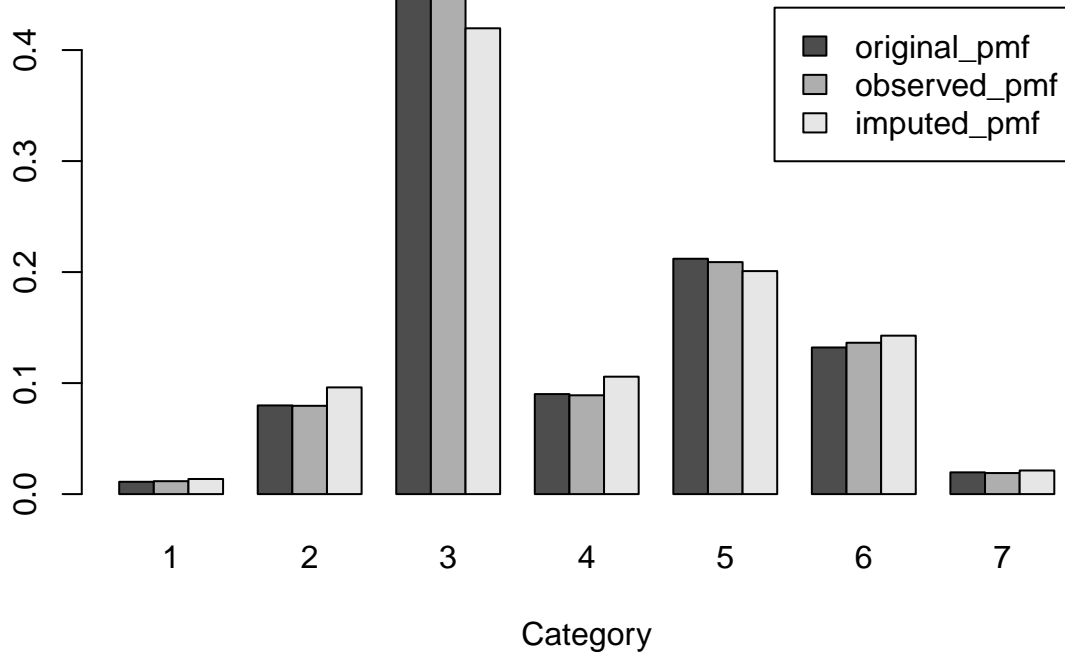
### MICE: NP



### MICE: ENG



### MICE: SCHL



### MICE: AGEP

