Testing different imputation methods on PUMS (MCAR) - RandomForest

```r
# 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 scneario 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

```r
df.imp <- missForest(df_observed, verbose = FALSE)
d1 <-  df.imp$ximp
df.imp <- missForest(df_observed, verbose = FALSE)
d2 <-  df.imp$ximp
df.imp <- missForest(df_observed, verbose = FALSE)
d3 <-  df.imp$ximp
df.imp <- missForest(df_observed, verbose = FALSE)
d4 <-  df.imp$ximp
df.imp <- missForest(df_observed, verbose = FALSE)
d5 <-  df.imp$ximp
imputed_sets = rbind(d1, d2, d3, d4, d5)
```
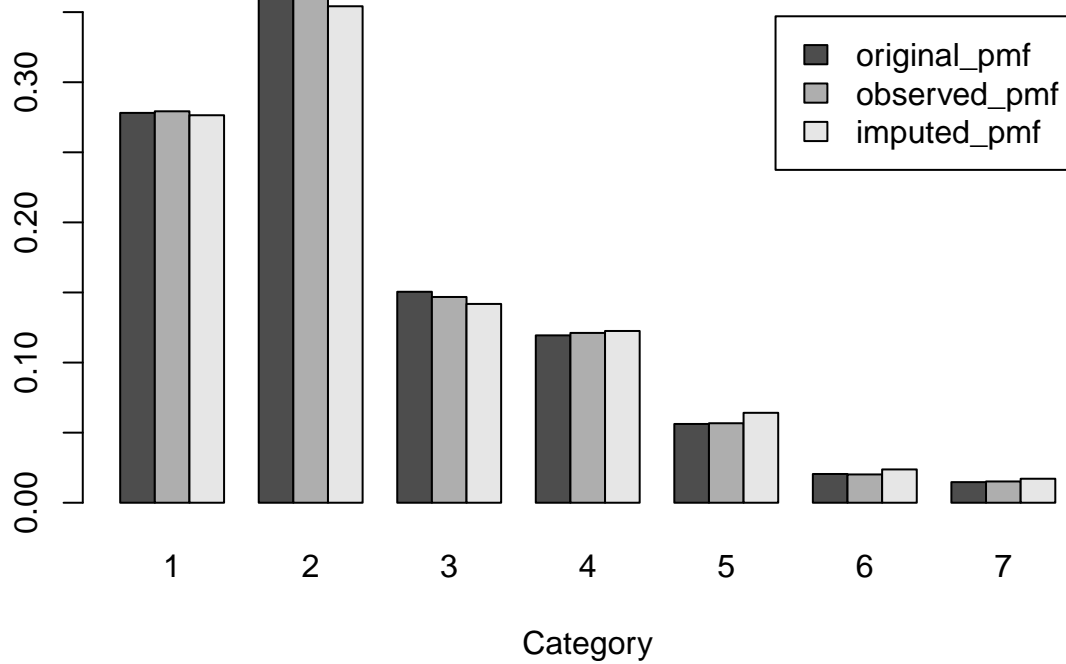
Diagnostics

Assess bivariate joint distribution
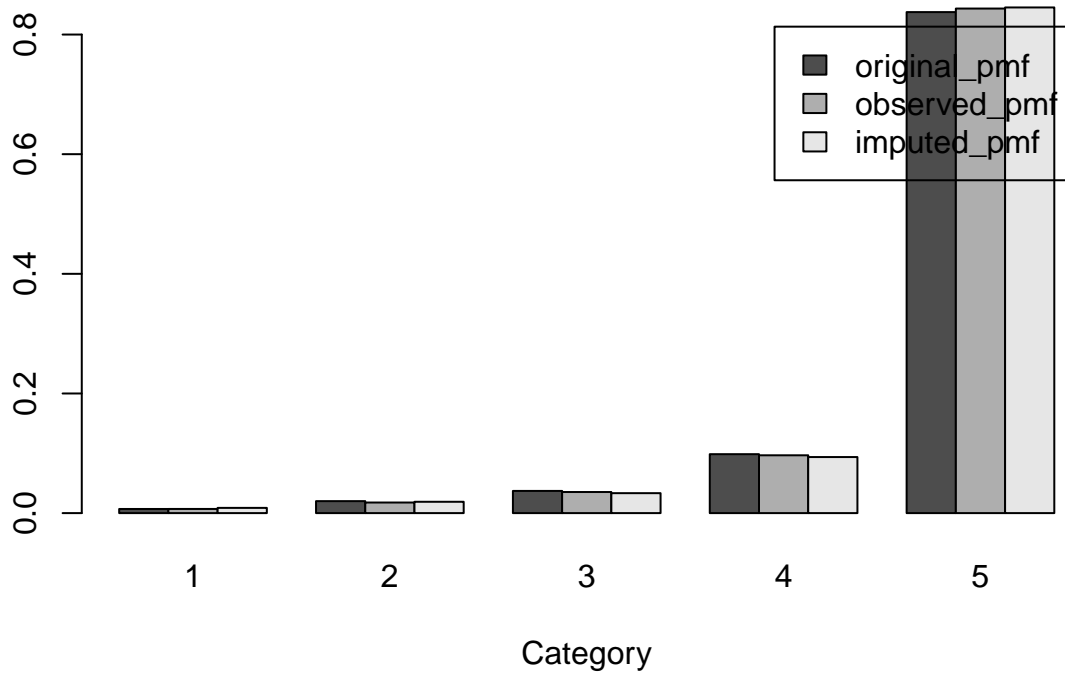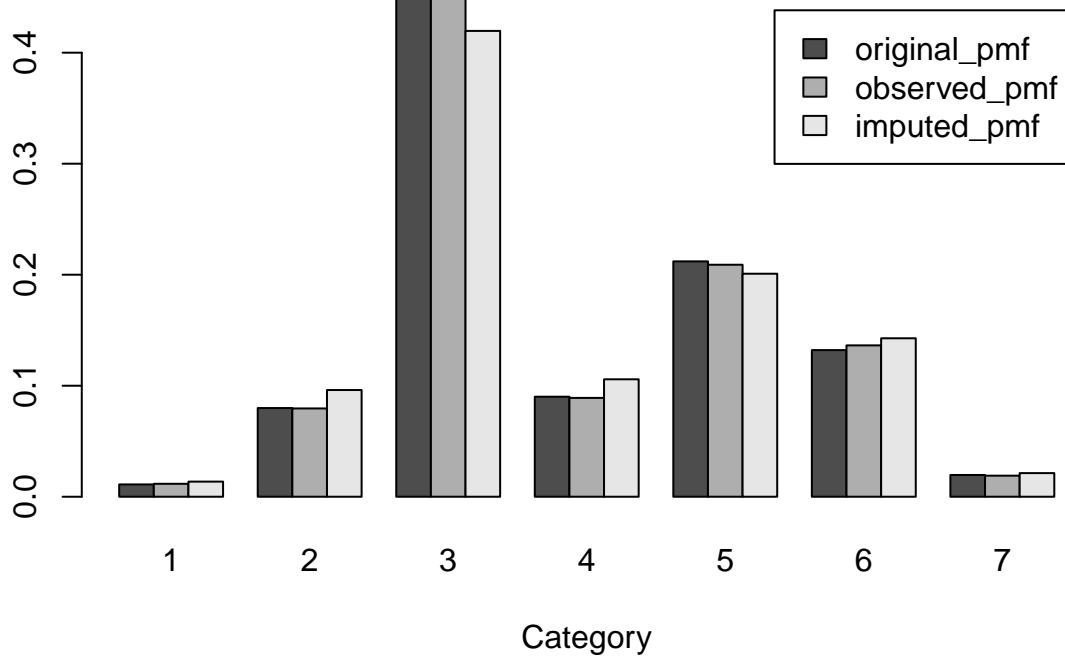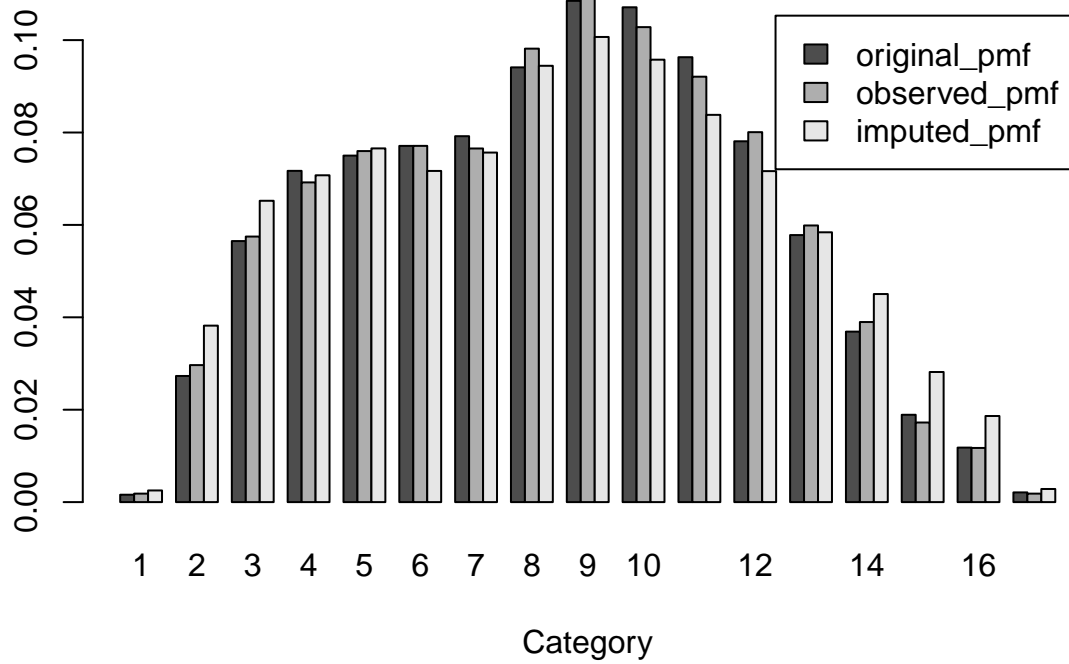
Assess trivariate joint distribution

MICE: VEH
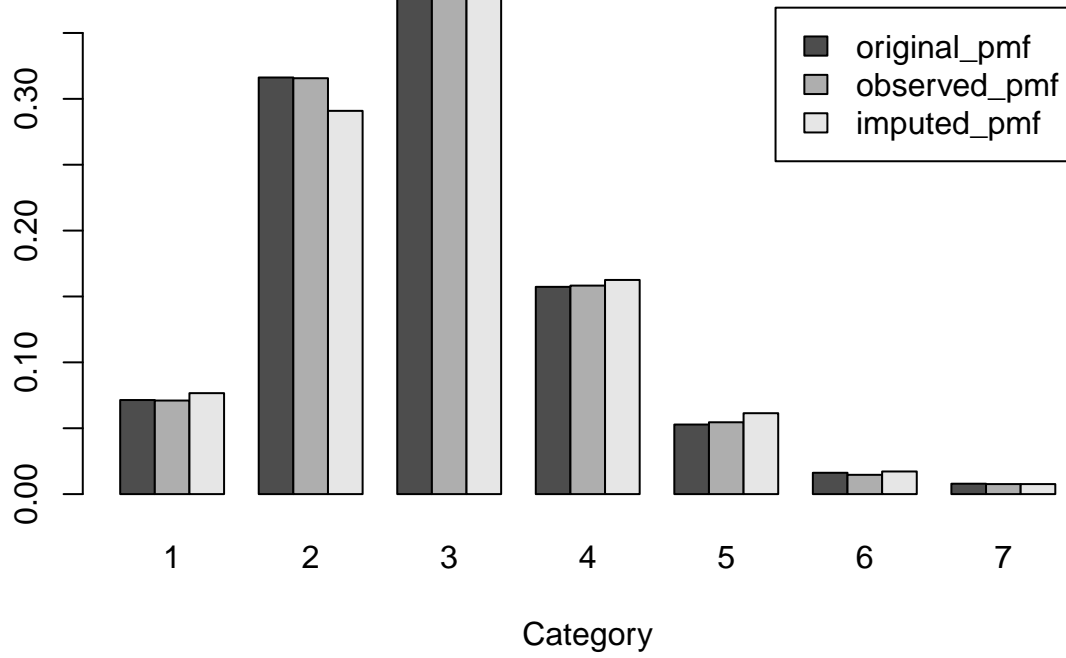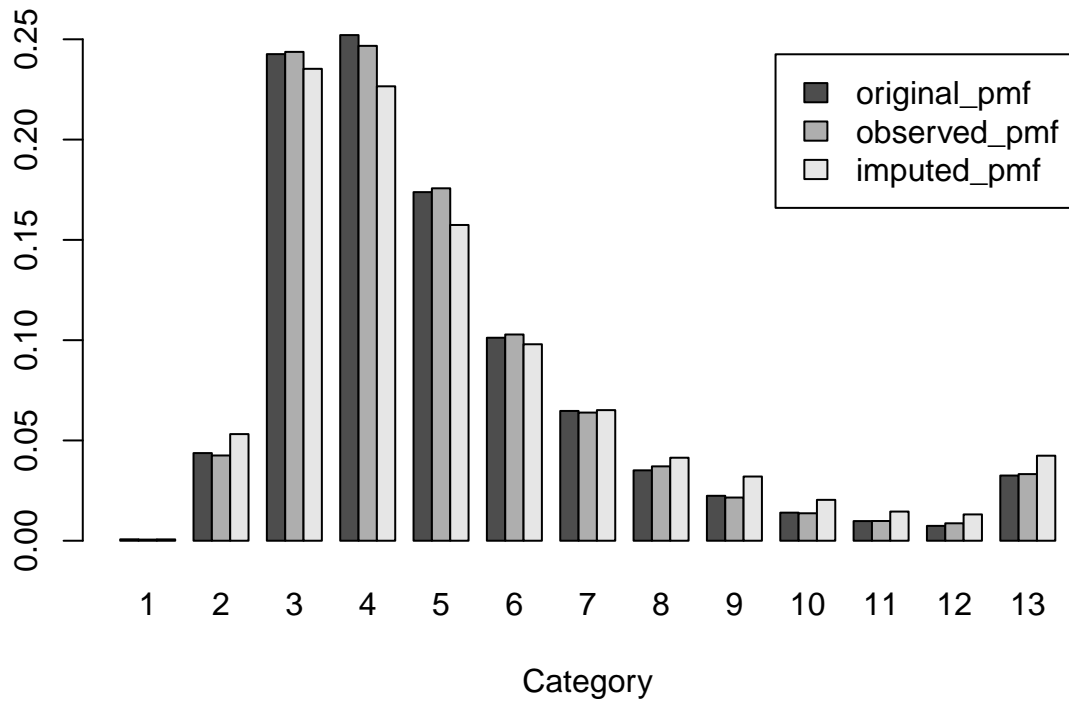
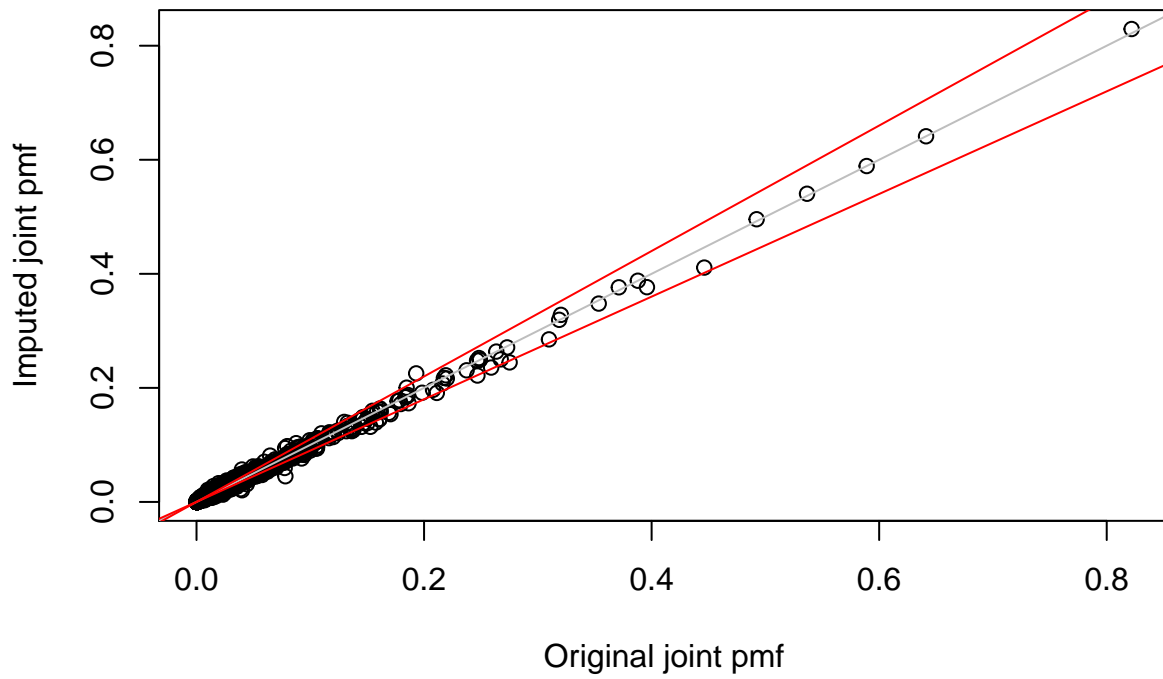**MICE: NP**

MICE: ENG

MICE: SCHL

MICE: AGEP

MICE: PINCP

Bivariate pmf , r square: 0.998

Trivariate pmf , r square: 0.995