

Testing different imputation methods on PUMS (MCAR) - MICE-CART

```
# 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
}
```

MICE-CART

Create 5 imputed dataset

```
library(mice)

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

imputed_df <- mice(df_observed,m=5,method="cart",print=F)
```

Warning: Number of logged events: 150

Extract the 5 imputed dataset

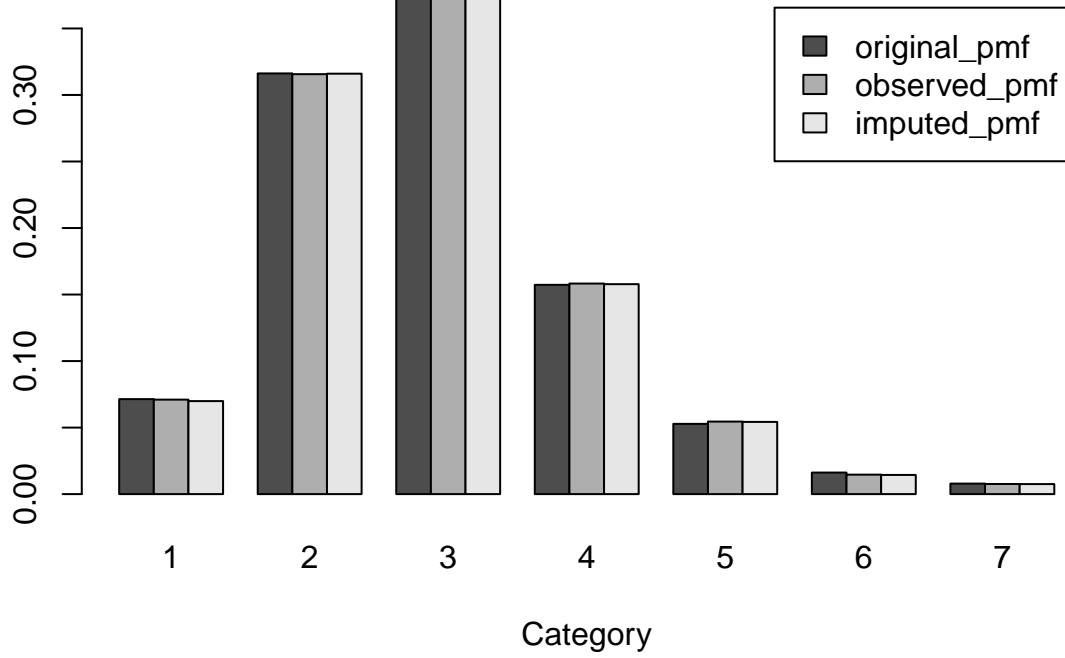
```
d1 <- complete(imputed_df, 1)
d2 <- complete(imputed_df, 2)
d3 <- complete(imputed_df, 3)
d4 <- complete(imputed_df, 4)
d5 <- complete(imputed_df, 5)
imputed_sets = rbind(d1, d2, d3, d4, d5)
```

Diagnostics

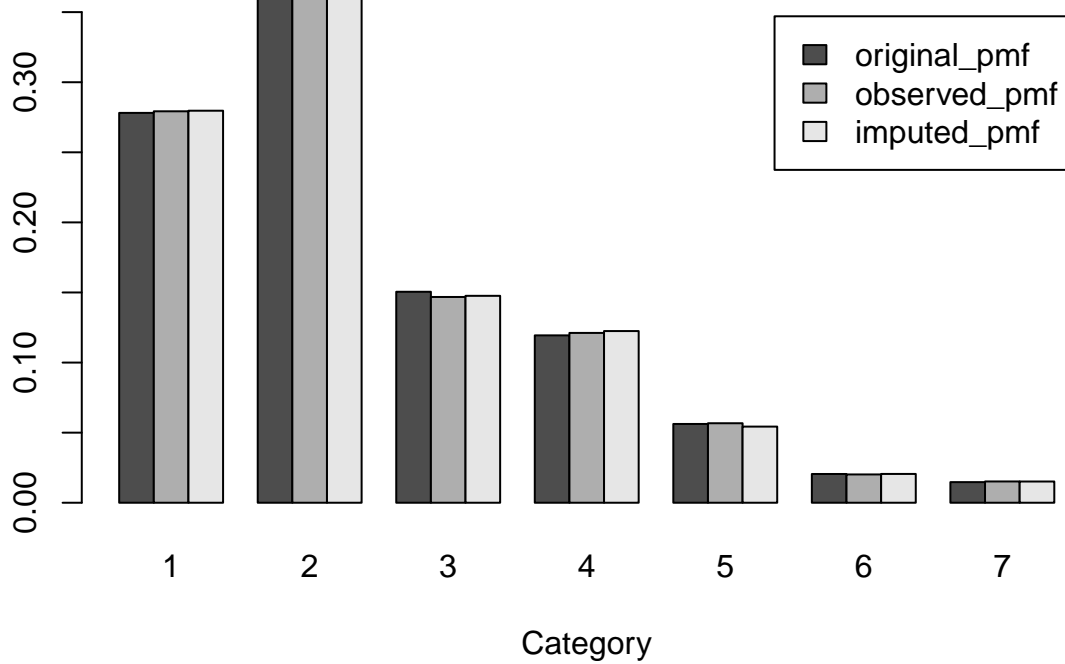
Assess bivariate joint distribution

Assess trivariate joint distribution

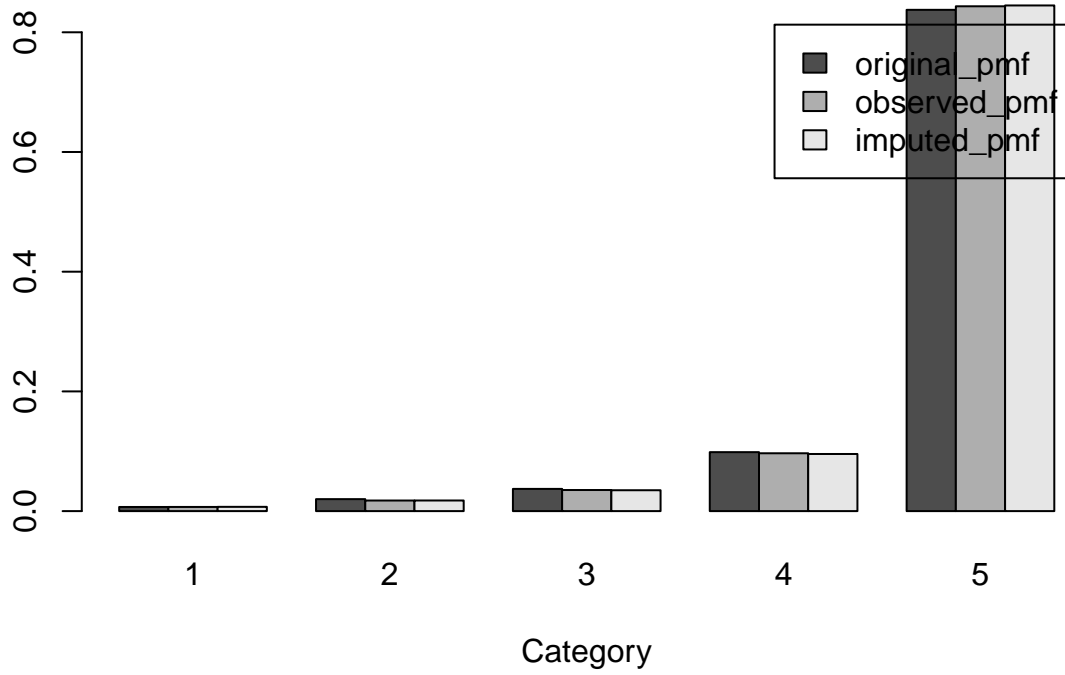
MICE-CART: VEH



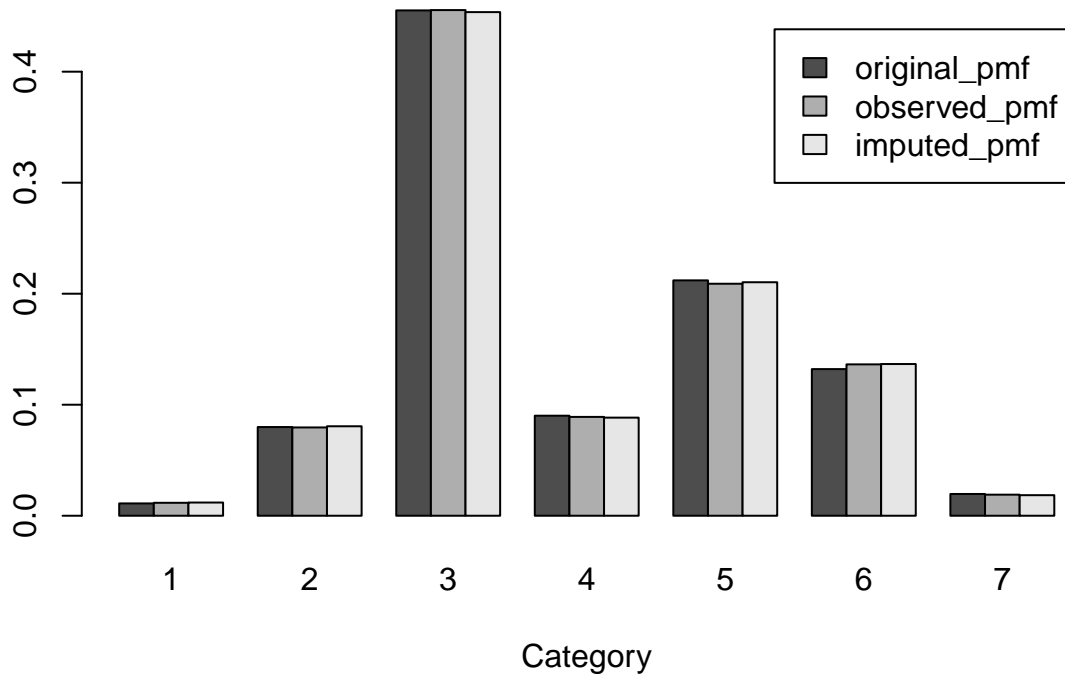
MICE-CART: NP



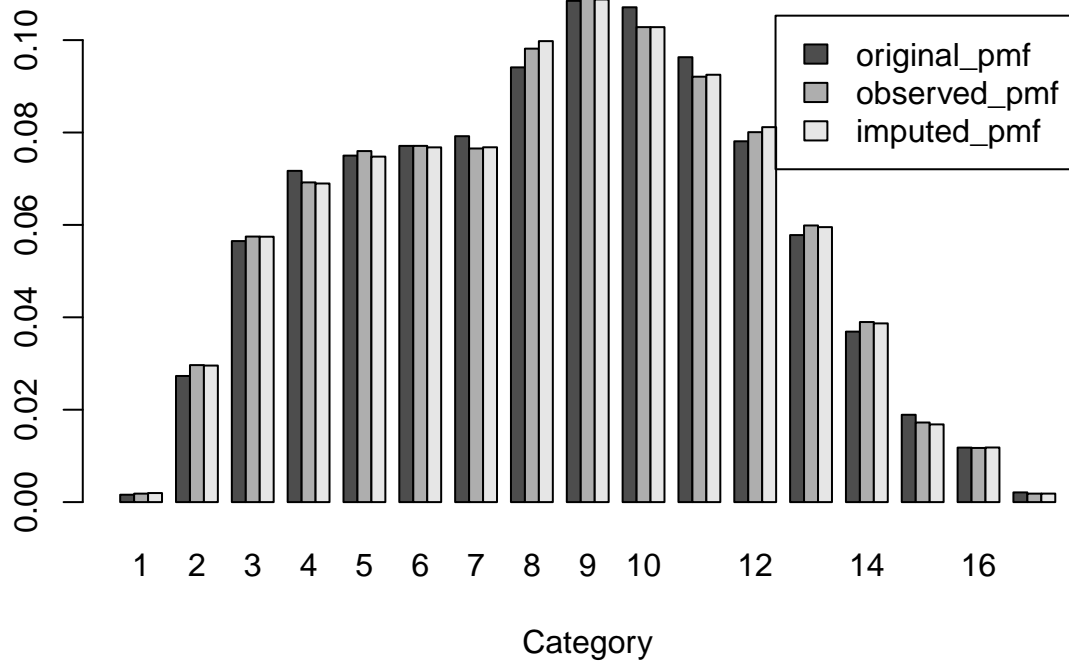
MICE-CART: ENG



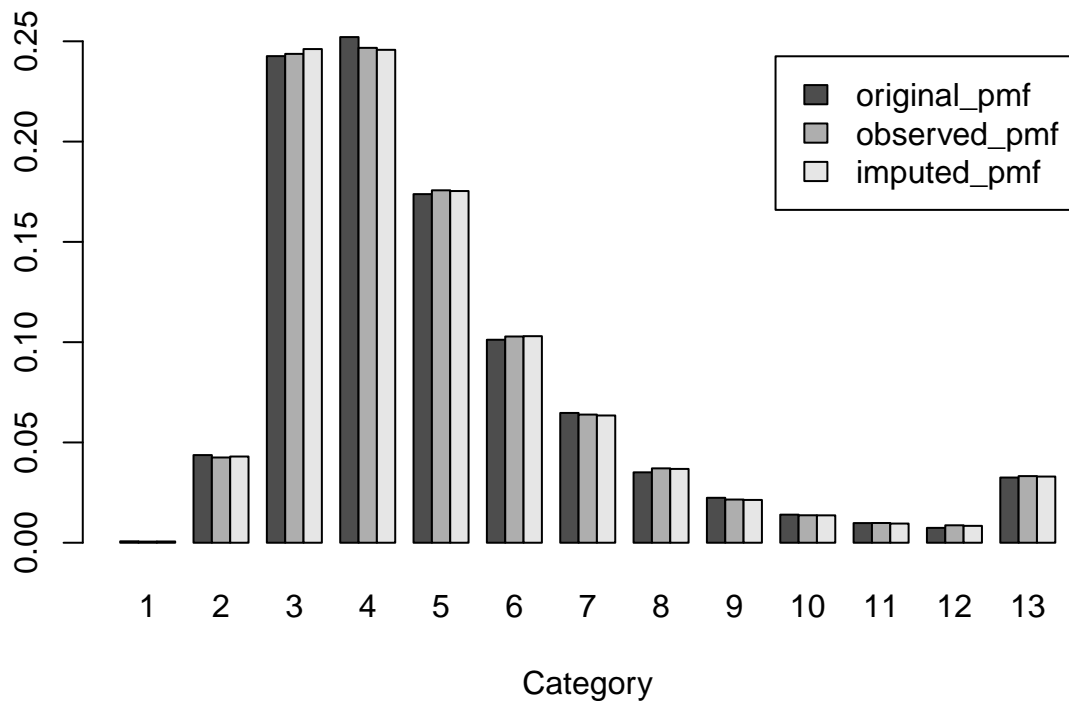
MICE-CART: SCHL



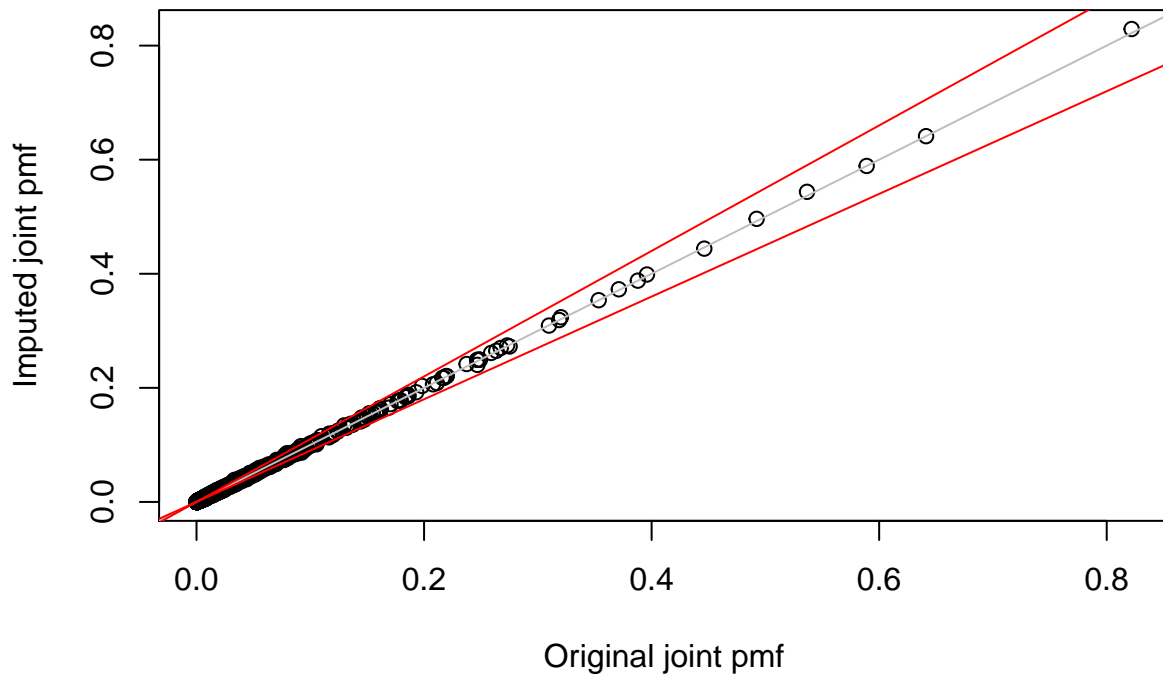
MICE-CART: AGEF



MICE-CART: PINCP



Bivariate pmf , r square: 1



Trivariate pmf , r square: 0.999

