MCAR 30% missing - MICE

```
# 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']]
```

MICE

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,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)
```

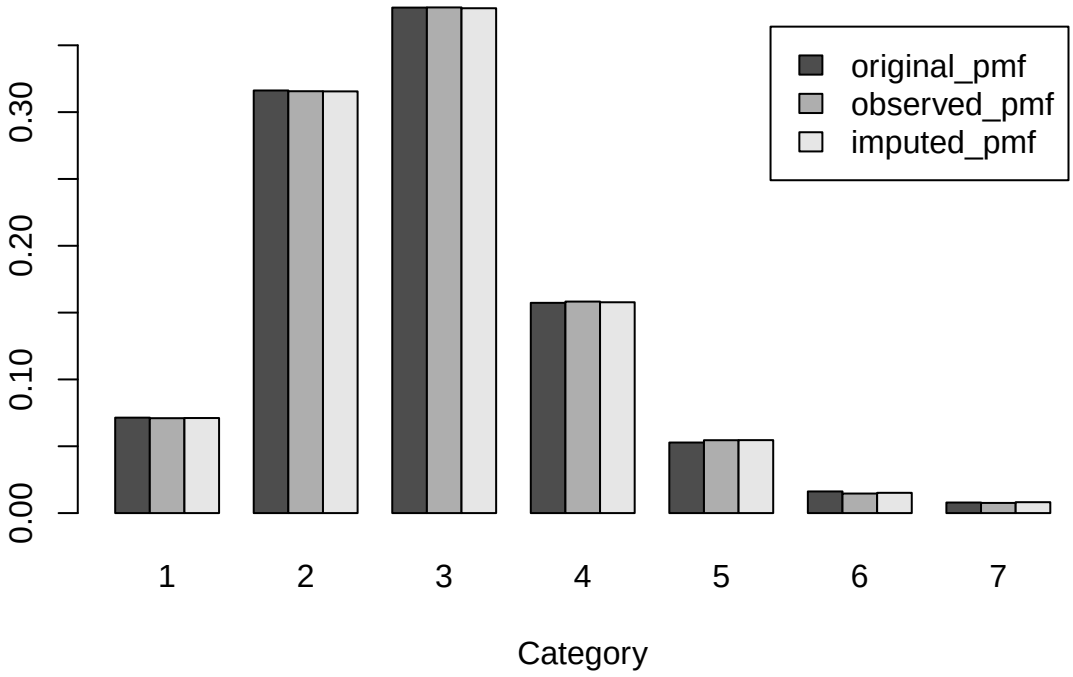
Diagnostics

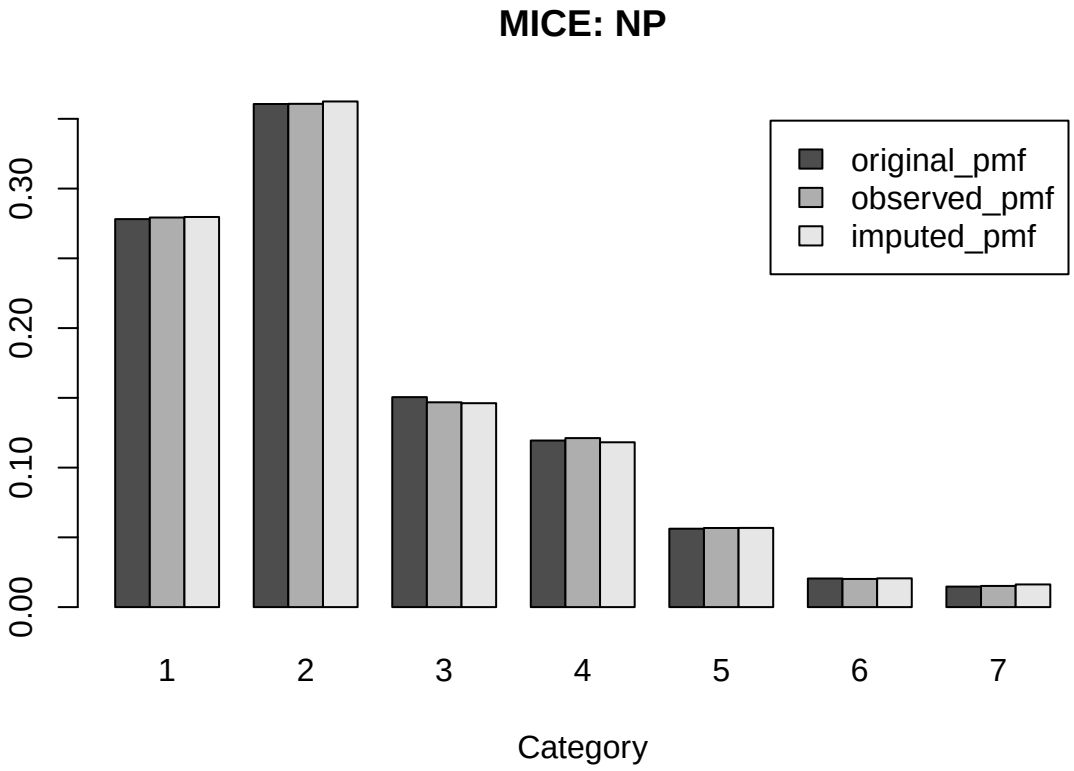
Assess bivariate joint distribution

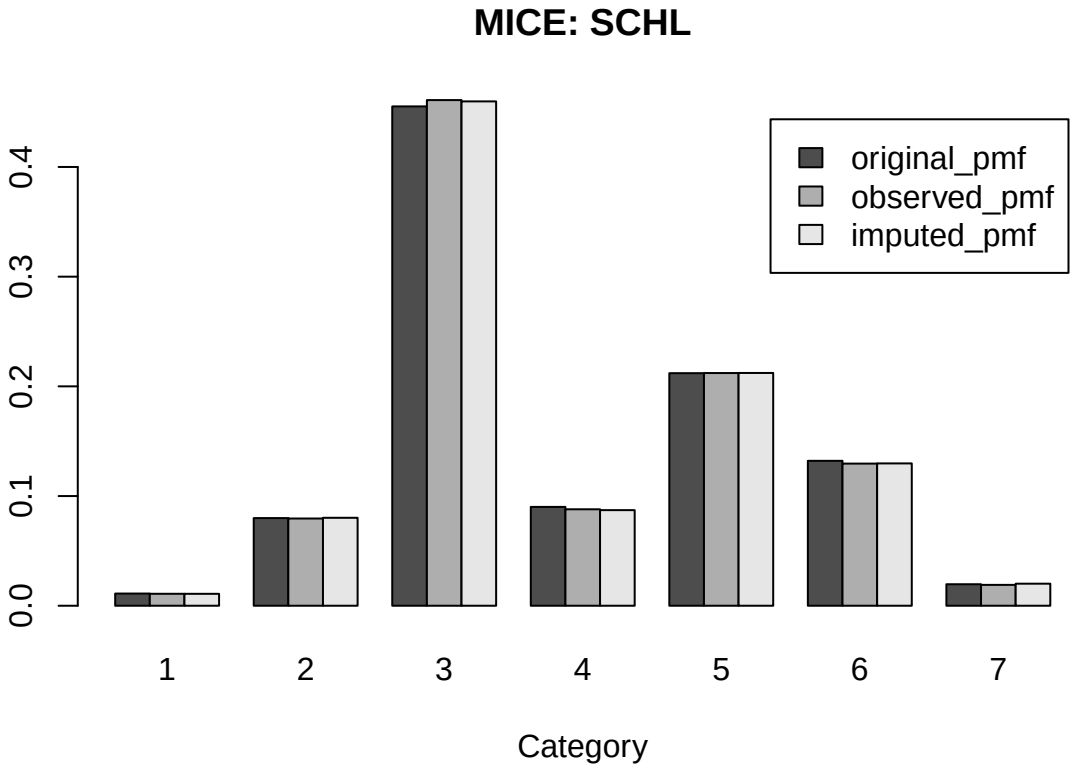
Assess trivariate joint distribution

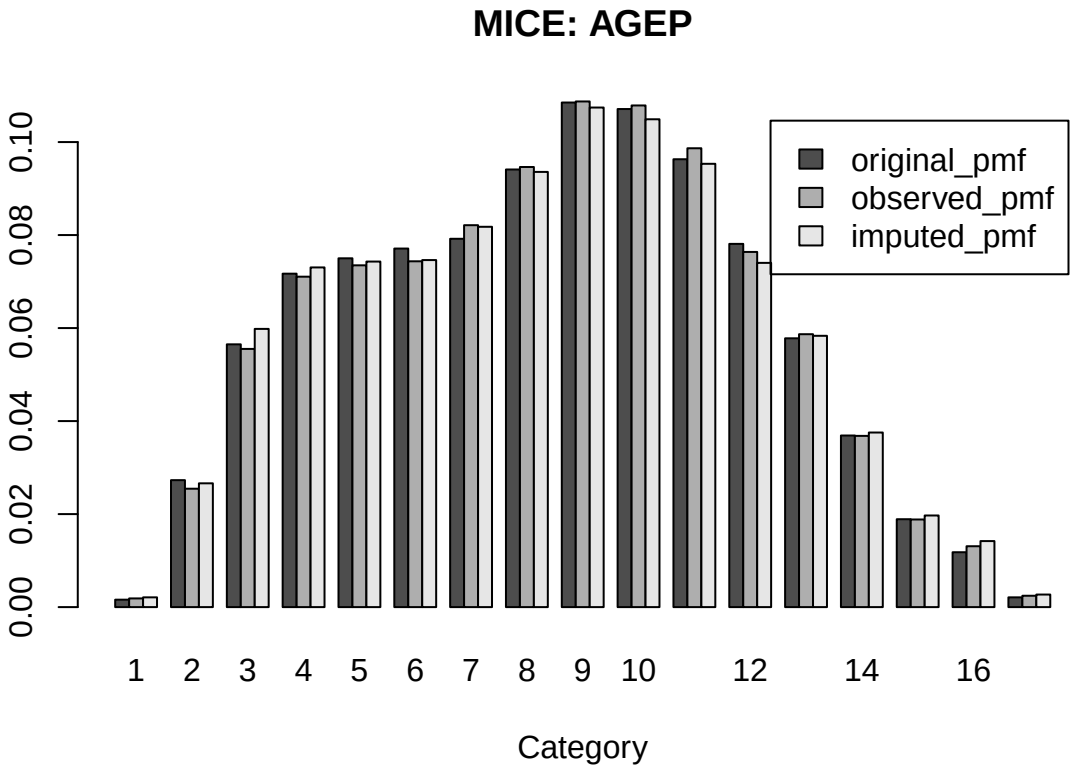
```
# calculate rmse
numeric_df = sapply(df, as.numeric)
normalized_df = t(t(numeric_df-1)/(apply(numeric_df, MARGIN = 2, FUN = max)-1))
numeric_impute = sapply(d1, as.numeric)
normalized_impute = t(t(numeric_impute-1)/(apply(numeric_df, MARGIN = 2, FUN = max)-1))

missing_matrix = is.na(df_observed)

rmse = sqrt(sum((normalized_df[missing_matrix] - normalized_impute[missing_matrix])^2)/sum(missing_matri
rmse
```

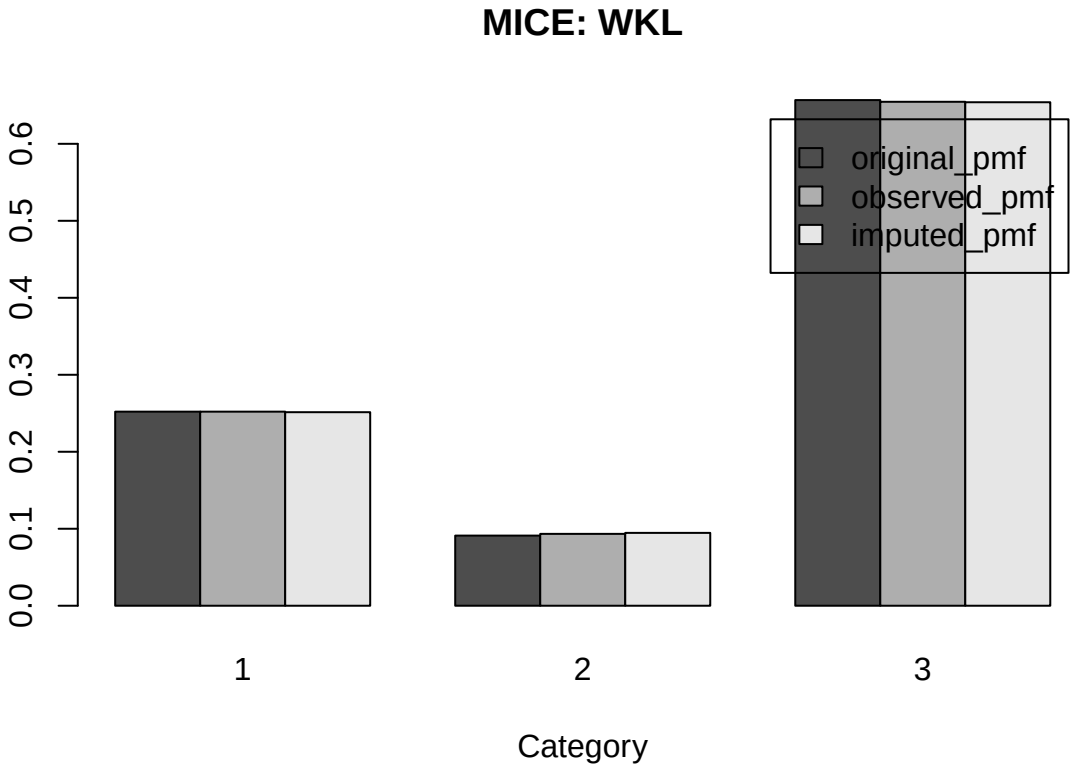
```
## [1] 0.3145641
```

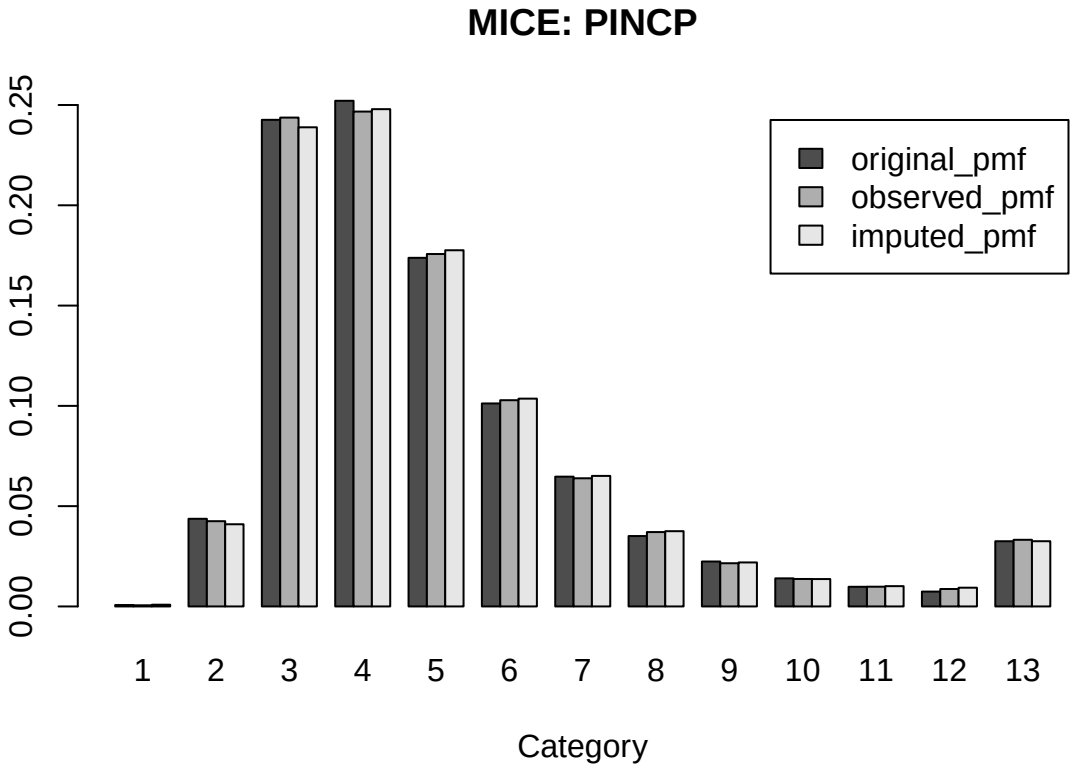

MICE: VEH
original_pmf
observed_pmf
imputed_pmf
0.00
0.10
0.20
0.30
1
2
3
4
5
6
7
Category

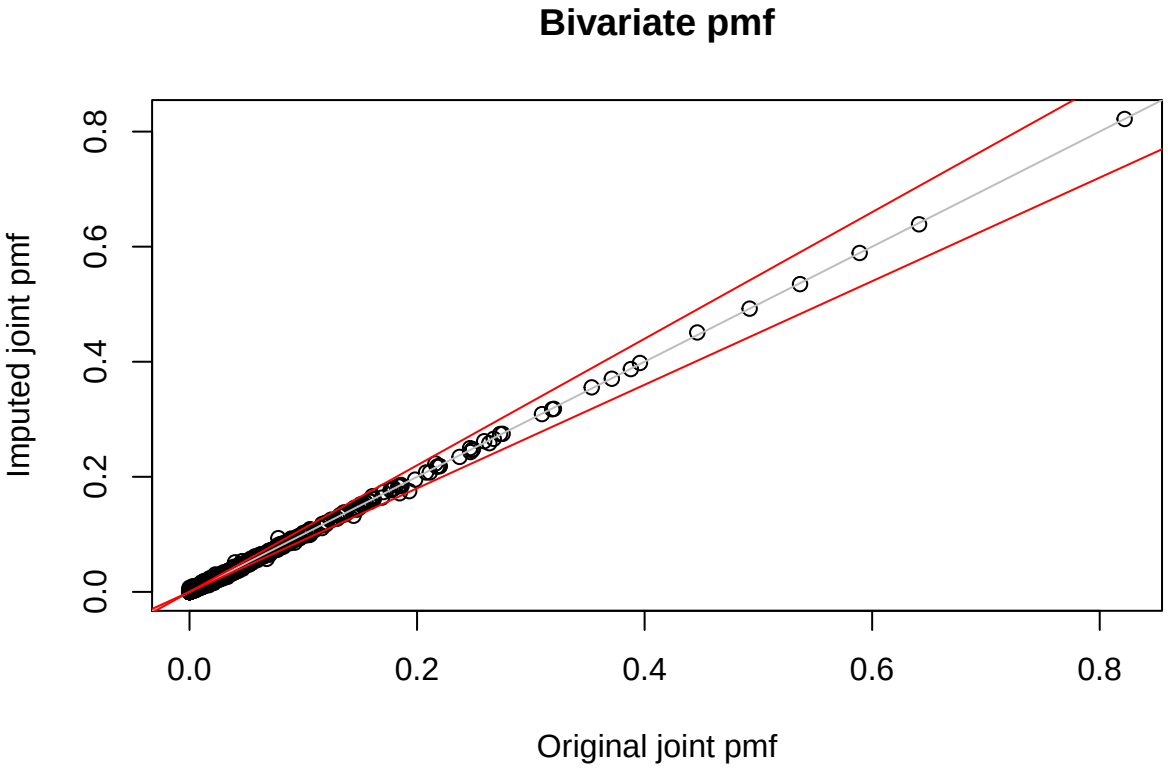

MICE: NP
original_pmf
observed_pmf
imputed_pmf
0.00
0.10
0.20
0.30
1
2
3
4
5
6
7
Category

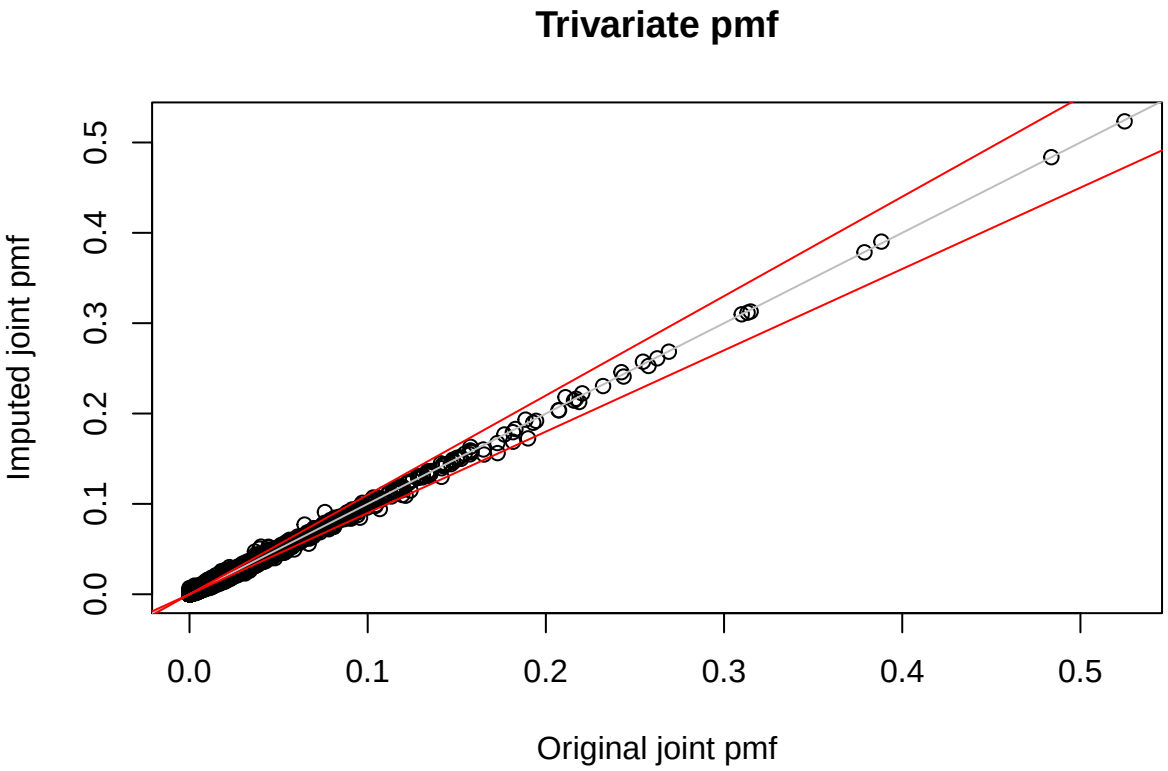

MICE: SCHL
original_pmf
observed_pmf
imputed_pmf
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5
6
7
Category


MICE: AGEP
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
original_pmf
observed_pmf
imputed_pmf


MICE: WKL
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
Category
original_pmf
observed_pmf
imputed_pmf


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


Bivariate pmf
Imputed joint pmf
0.0
0.2
0.4
0.6
0.8
Original joint pmf


Trivariate pmf
Imputed joint pmf
Original joint pmf
0.0
0.1
0.2
0.3
0.4
0.5