Data Quality: Lecture 4

Philipp Drebes

21.03.2023

Task 01

Missing pattern / Test whether MCAR is given

A data set *riskfactors* is given.

How does the missing pattern look like?

Are the elements of the data set completely at random (MCAR)?

Apply a MCAR test.

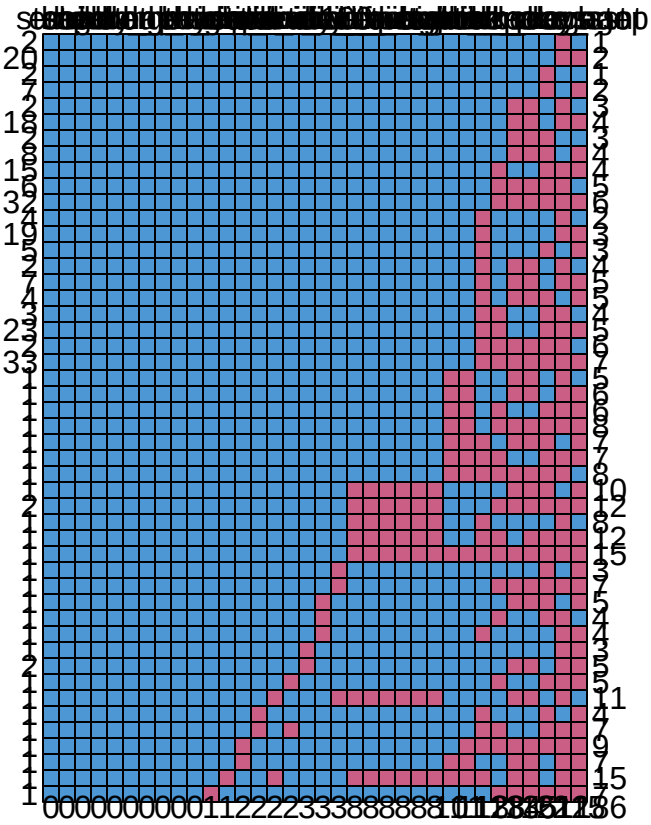
```
library(naniar)
library(mice)

head(riskfactors)
```

Missing pattern

```
## # A tibble: 6 x 34
##   state sex      age weight_lbs height_i~1   bmi marital pregn~2 child~3 educa~4
##   <fct> <fct>  <int>      <int>      <int> <dbl> <fct>   <fct>     <int> <fct>
## 1 26    Female    49        190         64  32.7 Married <NA>          0 6
## 2 40    Female    48        170         68  25.9 Divorc~ <NA>          0 5
## 3 72    Female    55        163         64  28.0 Married <NA>          0 4
## 4 42    Male      42        230         74  29.6 Married <NA>          1 6
## 5 32    Female    66        135         62  24.7 Widowed <NA>          0 5
## 6 19    Male      66        165         70  23.7 Married <NA>          0 5
## # ... with 24 more variables: employment <fct>, income <fct>, veteran <fct>,
## #   hispanic <fct>, health_general <fct>, health_physical <int>,
## #   health_mental <int>, health_poor <int>, health_cover <fct>,
## #   provide_care <fct>, activity_limited <fct>, drink_any <fct>,
## #   drink_days <int>, drink_average <int>, smoke_100 <fct>, smoke_days <fct>,
## #   smoke_stop <fct>, smoke_last <fct>, diet_fruit <int>, diet_salad <int>,
## #   diet_potato <int>, diet_carrot <int>, diet_vegetable <int>, ...
```

```
md.pattern(riskfactors)
```



```
##    state sex age children employment income health_general health_physical
## 2      1   1   1        1          1      1              1               1
## 20     1   1   1        1          1      1              1               1
## 2      1   1   1        1          1      1              1               1
## 7      1   1   1        1          1      1              1               1
## 2      1   1   1        1          1      1              1               1
## 18     1   1   1        1          1      1              1               1
## 2      1   1   1        1          1      1              1               1
## 8      1   1   1        1          1      1              1               1
## 15     1   1   1        1          1      1              1               1
## 6      1   1   1        1          1      1              1               1
## 32     1   1   1        1          1      1              1               1
## 4      1   1   1        1          1      1              1               1
## 19     1   1   1        1          1      1              1               1
## 5      1   1   1        1          1      1              1               1
## 2      1   1   1        1          1      1              1               1
## 7      1   1   1        1          1      1              1               1
## 4      1   1   1        1          1      1              1               1
## 3      1   1   1        1          1      1              1               1
## 23     1   1   1        1          1      1              1               1
## 2      1   1   1        1          1      1              1               1
## 33     1   1   1        1          1      1              1               1
## 1      1   1   1        1          1      1              1               1
## 1      1   1   1        1          1      1              1               1
## 1      1   1   1        1          1      1              1               1
## 1      1   1   1        1          1      1              1               1
```

```
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 2      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 2      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
## 1      1   1   1        1          1      1              1              1
##        0   0   0        0          0      0              0              0
##    health_mental health_cover marital education height_inch hispanic drink_any
## 2              1            1       1         1           1        1         1
## 20             1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 7              1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 18             1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 8              1            1       1         1           1        1         1
## 15             1            1       1         1           1        1         1
## 6              1            1       1         1           1        1         1
## 32             1            1       1         1           1        1         1
## 4              1            1       1         1           1        1         1
## 19             1            1       1         1           1        1         1
## 5              1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 7              1            1       1         1           1        1         1
## 4              1            1       1         1           1        1         1
## 3              1            1       1         1           1        1         1
## 23             1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 33             1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
```

```
## 2              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 2              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         1
## 1              1            1       1         1           1        1         0
## 1              1            1       1         1           1        0         1
## 1              1            1       1         1           1        0         1
## 1              1            1       1         1           0        1         1
## 1              1            1       1         1           0        1         1
## 1              1            1       1         0           1        1         0
## 1              1            1       0         1           1        1         1
##                0            0       1         1           2        2         2
##    smoke_100 veteran provide_care activity_limited diet_fruit diet_salad
## 2          1       1            1                 1          1          1
## 20         1       1            1                 1          1          1
## 2          1       1            1                 1          1          1
## 7          1       1            1                 1          1          1
## 2          1       1            1                 1          1          1
## 18         1       1            1                 1          1          1
## 2          1       1            1                 1          1          1
## 8          1       1            1                 1          1          1
## 15         1       1            1                 1          1          1
## 6          1       1            1                 1          1          1
## 32         1       1            1                 1          1          1
## 4          1       1            1                 1          1          1
## 19         1       1            1                 1          1          1
## 5          1       1            1                 1          1          1
## 2          1       1            1                 1          1          1
## 7          1       1            1                 1          1          1
## 4          1       1            1                 1          1          1
## 3          1       1            1                 1          1          1
## 23         1       1            1                 1          1          1
## 2          1       1            1                 1          1          1
## 33         1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          1          1
## 1          1       1            1                 1          0          0
## 2          1       1            1                 1          0          0
## 1          1       1            1                 1          0          0
## 1          1       1            1                 1          0          0
## 1          1       1            1                 1          0          0
```

```
## 1          1       1           1               0          1          1
## 1          1       1           1               0          1          1
## 1          1       1           0               1          1          1
## 1          1       1           0               1          1          1
## 1          1       1           0               1          1          1
## 1          1       0           1               1          1          1
## 2          1       0           1               1          1          1
## 1          0       1           1               1          1          1
## 1          1       1           1               0          0          0
## 1          1       1           1               1          1          1
## 1          0       1           1               1          1          1
## 1          1       1           1               1          1          1
## 1          1       1           1               1          1          1
## 1          1       1           1               1          0          0
## 1          1       1           1               1          1          1
##            2       3           3               3          8          8
##    diet_potato diet_carrot diet_vegetable diet_juice weight_lbs bmi health_poor
## 2            1           1              1          1          1   1           1
## 20           1           1              1          1          1   1           1
## 2            1           1              1          1          1   1           1
## 7            1           1              1          1          1   1           1
## 2            1           1              1          1          1   1           1
## 18           1           1              1          1          1   1           1
## 2            1           1              1          1          1   1           1
## 8            1           1              1          1          1   1           1
## 15           1           1              1          1          1   1           1
## 6            1           1              1          1          1   1           1
## 32           1           1              1          1          1   1           1
## 4            1           1              1          1          1   1           0
## 19           1           1              1          1          1   1           0
## 5            1           1              1          1          1   1           0
## 2            1           1              1          1          1   1           0
## 7            1           1              1          1          1   1           0
## 4            1           1              1          1          1   1           0
## 3            1           1              1          1          1   1           0
## 23           1           1              1          1          1   1           0
## 2            1           1              1          1          1   1           0
## 33           1           1              1          1          1   1           0
## 1            1           1              1          1          0   0           1
## 1            1           1              1          1          0   0           1
## 1            1           1              1          1          0   0           1
## 1            1           1              1          1          0   0           1
## 1            1           1              1          1          0   0           0
## 1            1           1              1          1          0   0           0
## 1            1           1              1          1          0   0           0
## 1            0           0              0          0          1   1           1
## 2            0           0              0          0          1   1           1
## 1            0           0              0          0          1   1           0
## 1            0           0              0          0          1   1           0
## 1            0           0              0          0          0   0           0
## 1            1           1              1          1          1   1           1
## 1            1           1              1          1          1   1           1
## 1            1           1              1          1          1   1           1
## 1            1           1              1          1          1   1           1
```

```
## 1            1           1              1          1          1   1           0
## 1            1           1              1          1          1   1           1
## 2            1           1              1          1          1   1           1
## 1            1           1              1          1          1   1           1
## 1            0           0              0          0          1   1           1
## 1            1           1              1          1          1   1           0
## 1            1           1              1          1          1   1           0
## 1            1           1              1          1          1   0           0
## 1            1           1              1          1          0   0           1
## 1            0           0              0          0          0   0           0
## 1            1           1              1          1          1   1           1
##              8           8              8          8         10  11         113
##    smoke_days drink_days drink_average smoke_last smoke_stop pregnant
## 2           1          1             1          1          0        1    1
## 20          1          1             1          1          0        0    2
## 2           1          1             1          0          1        1    1
## 7           1          1             1          0          1        0    2
## 2           1          0             0          1          0        1    3
## 18          1          0             0          1          0        0    4
## 2           1          0             0          0          1        1    3
## 8           1          0             0          0          1        0    4
## 15          0          1             1          0          0        0    4
## 6           0          0             0          0          0        1    5
## 32          0          0             0          0          0        0    6
## 4           1          1             1          1          0        1    2
## 19          1          1             1          1          0        0    3
## 5           1          1             1          0          1        0    3
## 2           1          0             0          1          0        1    4
## 7           1          0             0          1          0        0    5
## 4           1          0             0          0          1        0    5
## 3           0          1             1          0          0        1    4
## 23          0          1             1          0          0        0    5
## 2           0          0             0          0          0        1    6
## 33          0          0             0          0          0        0    7
## 1           1          0             0          1          0        1    5
## 1           1          0             0          1          0        0    6
## 1           0          1             1          0          0        0    6
## 1           0          0             0          0          0        0    8
## 1           1          0             0          0          1        0    7
## 1           0          1             1          0          0        0    7
## 1           0          0             0          0          0        1    8
## 1           1          0             0          0          1        0   10
## 2           0          0             0          0          0        0   12
## 1           1          1             1          1          0        1    8
## 1           0          1             0          0          0        0   12
## 1           0          0             0          0          0        0   15
## 1           1          1             1          0          1        0    3
## 1           0          0             0          0          0        0    7
## 1           1          0             0          0          1        0    5
## 1           0          1             1          0          0        1    4
## 1           1          1             1          1          0        0    4
## 1           1          1             1          1          0        0    3
## 2           1          0             0          1          0        0    5
## 1           0          1             1          0          0        0    5
```

```
## 1           1          0             0          1          0        1   11
## 1           1          1             1          0          1        0    4
## 1           0          1             1          0          0        0    7
## 1           0          0             0          0          0        0    9
## 1           1          0             0          1          0        0    7
## 1           1          0             0          1          0        0   15
## 1           0          0             0          0          0        0    7
##           128        134           135        161        212      215 1186
```

```
mcar_test(riskfactors)
```

MCAR Test

```
## Warning in norm::prelim.norm(data): NAs introduced by coercion to integer range
```

```
## # A tibble: 1 x 4
##   statistic    df  p.value missing.patterns
##       <dbl> <dbl>    <dbl>            <int>
## 1     1741.  1319 3.59e-14               48
```

Interpretation The null hypothesis in this test is that the data is MCAR. Given that the p-value of the MCAR test is below 0.05, we can conclude the *riskfactors* data is not missing completely at random.

Task 02

Missing analysis

A data set *grades* is given.

There are missings in the data set.

Using R, analyze the missing structure.

Apply single imputation for types “mean” and “sample” and compare them.

With this setting you will get to know the two simplest methods of MICE.

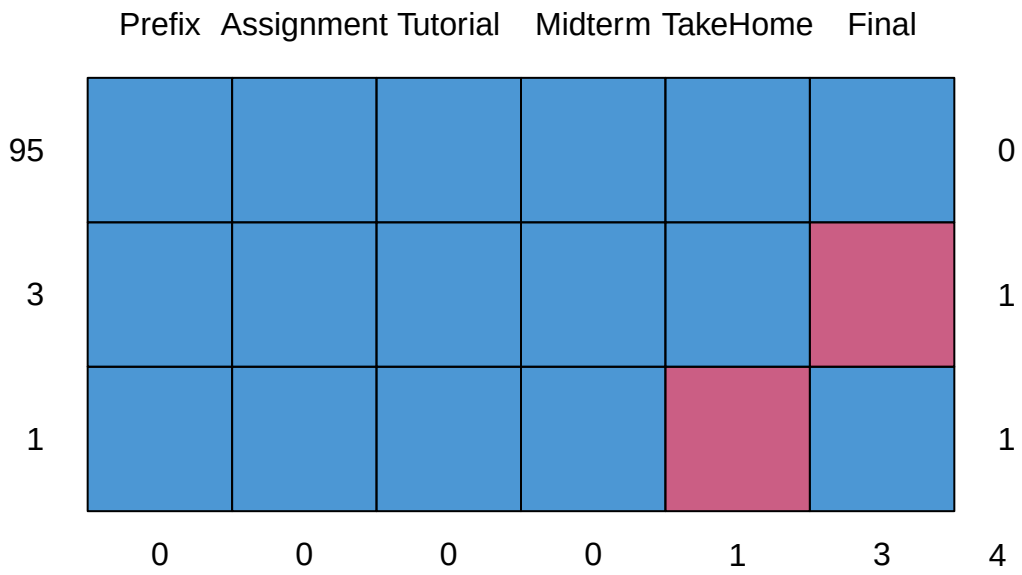
```
library(readxl)

grades <- read_excel('data/grades.xlsx')
head(grades)
```

Missing pattern

```
## # A tibble: 6 x 6
##   Prefix Assignment Tutorial Midterm TakeHome Final
##    <dbl>      <dbl>    <dbl>   <dbl>    <dbl> <dbl>
## 1      5       57.1     34.1    64.4     51.5  52.5
## 2      8       95.0    105.     67.5     99.1  68.3
## 3      8       83.7     83.2    30       63.2  48.9
## 4      7       81.2     96.1    49.4    106.   80.6
## 5      8       91.3     93.6    95      107.   73.9
## 6      7       95       92.6    93.1     97.8  68.1
```

```
md.pattern(grades)
```




```
##    Prefix Assignment Tutorial Midterm TakeHome Final
## 95      1          1        1       1        1     1 0
## 3       1          1        1       1        1     0 1
## 1       1          1        1       1        0     1 1
##         0          0        0       0        1     3 4
```

```
grades.imp.mean <- mice(grades, method="mean", m=1, maxit=1)
```

Imputation

```
##
##  iter imp variable
##   1   1  TakeHome  Final
```

```
grades.imp.sample <- mice(grades, method="sample", m=1, maxit=1)
```

```
##
##  iter imp variable
##   1   1  TakeHome  Final
```

```
grades.mean <- complete(grades.imp.mean)
grades.sample <- complete(grades.imp.sample)

cat('mean TakeHome before imputation: ', mean(grades$TakeHome, na.rm = TRUE))
```

```
## mean TakeHome before imputation:  80.82847
```

```
cat('mean TakeHome after imputation [mean]: ', mean(grades.mean$TakeHome, na.rm = TRUE))
```

```
## mean TakeHome after imputation [mean]:  80.82847
```

```
cat('mean TakeHome after imputation [sample]: ', mean(grades.sample$TakeHome, na.rm = TRUE))
```

```
## mean TakeHome after imputation [sample]:  81.07828
```

```
cat('mean Final before imputation:', mean(grades$Final, na.rm = TRUE))
```

```
## mean Final before imputation: 68.41438
```

```
cat('mean Final after imputation [mean]:', mean(grades.mean$Final, na.rm = TRUE))
```

```
## mean Final after imputation [mean]: 68.41438
```

```
cat('mean Final after imputation [sample]:', mean(grades.sample$Final, na.rm = TRUE))
```

```
## mean Final after imputation [sample]: 68.49606
```

```
head(grades.mean, n = 10)
```

```
##    Prefix Assignment Tutorial Midterm  TakeHome Final
## 1       5      57.14    34.09   64.38  51.48000 52.50
## 2       8      95.05   105.49   67.50  99.07000 68.33
## 3       8      83.70    83.17   30.00  63.15000 48.89
## 4       7      81.22    96.06   49.38 105.93000 80.56
## 5       8      91.32    93.64   95.00 107.41000 73.89
## 6       7      95.00    92.58   93.12  97.78000 68.06
## 7       8      95.05   102.99   56.25  99.07000 50.00
## 8       7      72.85    86.85   60.00  80.82847 56.11
## 9       8      84.26    93.10   47.50  18.52000 50.83
## 10      7      90.10    97.55   51.25  88.89000 63.61
```

```
head(grades.sample, n = 10)
```

```
##    Prefix Assignment Tutorial Midterm TakeHome Final
## 1       5      57.14    34.09   64.38    51.48 52.50
## 2       8      95.05   105.49   67.50    99.07 68.33
## 3       8      83.70    83.17   30.00    63.15 48.89
## 4       7      81.22    96.06   49.38   105.93 80.56
## 5       8      91.32    93.64   95.00   107.41 73.89
## 6       7      95.00    92.58   93.12    97.78 68.06
## 7       8      95.05   102.99   56.25    99.07 50.00
## 8       7      72.85    86.85   60.00   105.56 56.11
## 9       8      84.26    93.10   47.50    18.52 50.83
## 10      7      90.10    97.55   51.25    88.89 63.61
```

Task 03

MCAR, MAR or MNAR?

Develop your own examples of the emergence of the MCAR, MAR, and MNAR.

MCAR (Missing Completely At Random) For example, when data are missing for respondents because of a server error on the system that was running the online survey tool.

MAR (Missing At Random) For example, only younger people have missing values for IQ. In that case the probability of missing data on IQ is related to age.

MNAR (Missing Not At Random) For example, when data are missing on IQ and only the people with low IQ values have missing observations for this variable.