



Multiple Imputation and subsequent calculations

Simon Ress | Ruhr-Universität Bochum

September 22, 2022

Workshop at hr&c, Bochum, 2021

1. Exemplary Data Set
2. Missing Patterns
3. Slide with R Output
4. Figures caption

Exemplary Data Set

Planing of the Data Set



Data Set Creation

```
#Create variables
```

```
set.seed(415)
```

```
Mann = ifelse(runif(5000,0,1) < 0.50, 1, 0)
```

```
Alter = as.numeric(cut(runif(5000,20,70), c(20,30,40,50,60,70)))
```

```
FK = ifelse((Mann*0.2 + Alter*0.1 + runif(5000,0,0.6)) > 0.5, 1, 0)
```

```
B2.1 = as.numeric(cut(FK*2 + rnorm(5000,2,0.35), c(0,1,2,3)))
```

```
set.seed(1015)
```

```
B5.1 = as.numeric(cut(FK*1.5 + B2.1*0.2 + Alter*(-0.30) + rnorm(5000,0,0.35), c(0,1,2,3)))
```

```
#Build data frame
```

```
df = data.frame(Mann, Alter, FK, B2.1, B5.1)
```

View Data Set

df	df	df	df	df	df
##	4189	0	1	0	2
##	4190	1	4	1	4
##	4191	0	4	0	5
##	4192	0	5	1	4
##	4193	1	4	0	3
##	4194	1	5	1	4
##	4195	0	3	0	3
##	4196	0	3	1	4
##	4197	1	4	0	3
##	4198	1	5	1	4
##	4199	0	4	0	3
##	4200	0	3	0	2
##	4201	1	4	0	2
##	4202	1	4	0	4
##	4203	0	2	0	3

Check

Missing Patterns

Missing Patterns: Missing completely at random (MCAR)

- Values are randomly missing in the dataset
 - Missing data values do not relate to any other data
 - There is no pattern to the actual values of the missing data themselves
- For instance, when smoking status is not recorded in a random subset of patients
- This is easy to handle, but unfortunately, data are almost never missing completely at random

MCAR: Inserting Missing values in data frame

Missing Patterns: Missing at random (MAR)

- Confusing and would be better stated as *missing conditionally at random*
- Missing data do have a relationship with other variables in the dataset
 - Whether a value is missing or not depends on other variables
- The actual values that are missing are random
- For example, smoking status is not documented in female patients because the doctor was too shy to ask

Missing Patterns: Missing not at random (MNAR)

- The pattern of missingness is related to other variables in the dataset
- In addition, the values of the missing data are not random
 - Whether a value is missing or not depends on other variables
- For example, when smoking status is not recorded in patients admitted as an emergency, who are also more likely to have worse outcomes from surgery

- Official GitHub Repo of Metropolis (formerly mtheme); older version in TeXLive
- My GitHub Repo for a local Ubuntu package of Metropolis – formerly mtheme
- Manuel

Slide with Bullets

- Bullet 1
- Bullet 2
- Bullet 3

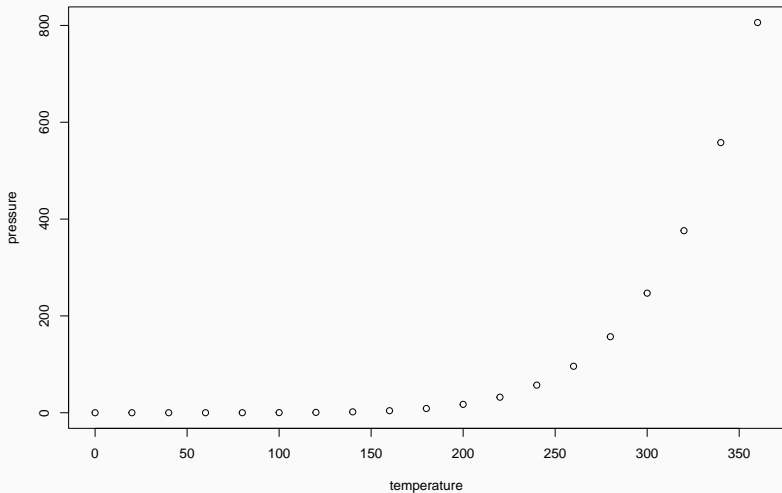
Slide with R Output

Slide with R Output

```
summary(cars)
```

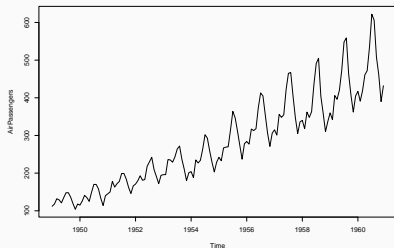
##	speed	dist
##	Min. : 4.0	Min. : 2.00
##	1st Qu.:12.0	1st Qu.: 26.00
##	Median :15.0	Median : 36.00
##	Mean :15.4	Mean : 42.98
##	3rd Qu.:19.0	3rd Qu.: 56.00
##	Max. :25.0	Max. :120.00

Slide with Plot



Two column layout

Here is some text above which goes over to whole slide



- Description of plot
- Second point

and here some text below which goes over to whole slide

Breakout page

Figures caption

Figures caption



Figure 1: Figure: Here is a really important caption.

Using LaTeX Parts: Blocks

As one example of falling back into \LaTeX , consider the example of three different block environments are pre-defined and may be styled with an optional background color.

Default

Block content.

Alert

Block content.

Example

Block content.