

2.1 Hands-On: R Markdown

Marco Wähler

2022-03-25

Aufgabenstellung - Lösung

Hinweis: Sie können die Aufgaben in einem Skript oder in einem R Markdown File bearbeiten. Hinweise zur Lösung der Aufgaben finden Sie auch im R Einführungsskript Kapitel 5.6.

Aufgabe 1

Importieren Sie nun die Daten aus dem ALLBUS 2018.

```
library(rio)
```

```
df <- import("data/allbus_2018_gesamt.sav")
```

Aufgabe 2

Klassifizieren Sie das Alter der Befragten (Variable `age`) in Geburtskohorten. Die Ausprägungen der neuen Variable sollen ein Label haben: Befragte unter 20 Jahren erhalten den Wert “unter 20”; Befragte älter/gleich 20 Jahre und jünger als 30 Jahre erhalten den Wert “20 bis 29”; Befragte älter/gleich 30 Jahre und jünger als 40 Jahre erhalten den Wert “30 bis 39” usw. mit “40 bis 49”, “50 bis 59”, “60 bis 69”, Befragte älter/gleich 70 Jahre erhalten den Wert “über 70”.

```
df$age_class[df$age < 20] <- "unter 20"
df$age_class[df$age >= 20 & df$age < 30] <- "20 bis 29"
df$age_class[df$age >= 30 & df$age < 40] <- "30 bis 39"
df$age_class[df$age >= 40 & df$age < 50] <- "40 bis 49"
df$age_class[df$age >= 50 & df$age < 60] <- "50 bis 59"
df$age_class[df$age >= 60 & df$age < 70] <- "60 bis 69"
df$age_class[df$age >= 70] <- "über 70"
```

Aufgabe 3

Geben Sie nun die absolute Häufigkeit der neuen Variable zu den Geburtskohorten aus. Was fällt Ihnen an der Reihenfolge der Geburtskohorten auf?

```
table(df$age_class)
```

```
##
## 20 bis 29 30 bis 39 40 bis 49 50 bis 59 60 bis 69 über 70 unter 20
##      412      511      532      727      641      586      63
```

Aufgabe 4

Wenn Sie die Reihenfolge der Werte festlegen wollen, können Sie die Variable als **Factor** rekodieren. Nutzen Sie hierfür folgende Variable, um die Reihenfolge der Werte zu spezifizieren.

```
age_level <- c("unter 20", "20 bis 29", "30 bis 39",  
"40 bis 49", "50 bis 59", "60 bis 69", "über 70")
```

Nutzen Sie anschließend die **factor()**-Funktion um die Variable zu transformieren. Innerhalb der **factor**-Funktion können Sie ein Argument spezifizieren, um die Levels der Variable **age_level** zu übernehmen. Nutzen Sie ggf. die Help-Funktion. Geben Sie anschließend erneut die absolute Häufigkeit der neuen Variable aus, um die Reihenfolge zu überprüfen.

```
#spezifizierung über levels = age_level. Achtung: die Variable age_level muss zuvor erstellt werden  
df$age_class <- factor(df$age_class, levels = age_level)
```

```
#Häufigkeit  
table(df$age_class)
```

```
##  
##  unter 20 20 bis 29 30 bis 39 40 bis 49 50 bis 59 60 bis 69  über 70  
##      63      412      511      532      727      641      586
```

```
attributes(df$age_class)
```

```
## $levels  
## [1] "unter 20" "20 bis 29" "30 bis 39" "40 bis 49" "50 bis 59" "60 bis 69"  
## [7] "über 70"  
##  
## $class  
## [1] "factor"
```

Aufgabe 5

Geben Sie eine Tabelle mit den prozentualen Häufigkeiten der Kohorten aus. Die prozentualen Anteile sollen auf zwei Nachkommastellen gerundet werden.

```
#prozentuale Häufigkeit  
round(100 * prop.table(table(df$age_class)), 2)
```

```
##  
##  unter 20 20 bis 29 30 bis 39 40 bis 49 50 bis 59 60 bis 69  über 70  
##    1.81    11.87    14.72    15.32    20.94    18.46    16.88
```

Aufgabe 6

Erstellen Sie eine Tabelle - getrennt nach Ost und West - mit den prozentualen Häufigkeiten der Kohorten. Die prozentualen Anteile sollen auf zwei Nachkommastellen gerundet werden.

```
tapply(df$age_class, df$eastwest,
       function(x) round(100*prop.table(table(x)), 2))
```

```
## $`1`
## x
## unter 20 20 bis 29 30 bis 39 40 bis 49 50 bis 59 60 bis 69 über 70
##      2.14      13.55      14.73      15.78      20.73      16.95      16.11
##
## $`2`
## x
## unter 20 20 bis 29 30 bis 39 40 bis 49 50 bis 59 60 bis 69 über 70
##      1.10      8.17      14.69      14.33      21.40      21.76      18.55
```