

## 2.2 Hands-On: Gewichtung

Marco Wähler

2022-03-25

### Aufgabenstellung - Lösung

In der GLES sind Personen, die in Ostdeutschland leben, im Verhältnis zu den Bevölkerungsanteilen überdurchschnittlich repräsentiert (Oversampling). Um die Disproportionalität aufzuheben, beinhaltet der Datensatz ein Designgewicht. Das Designgewicht finden Sie in der Variablen „w\_ow“. Nutzen Sie für die folgenden Aufgaben das R-Package „survey“

#### Aufgabe 1

Importieren Sie die Daten der GLES zur Vor- und Nachwahlbefragung 2021. Zur Erinnerung: Im Datensatz sind sowohl die Vorwahl- als auch die Nachwahl-Daten. In der Variable “sample” ist die jeweilige Befragung gespeichert. Bitte bearbeiten Sie die Aufgaben nur mit Befragten aus der Nachwahl-Befragung. Filtern Sie demnach nach der Variable “sample” und speichern Sie die Befragten aus der Nachwahl-Befragung im neuen Dataframe “df\_post” ab.

```
library(rio)
library(survey)
```

```
df <- import("data/gles_2021_gesamt.sav")
```

```
attributes(df$sample)
```

```
## $label
## [1] "Stichprobe"
##
## $format.spss
## [1] "F32.0"
##
## $labels
## GLES Querschnitt 2021, Vorwahl GLES Querschnitt 2021, Nachwahl
##                                7                                8
```

```
#option 1
df_post <- df[df$sample == 8, ]
```

```
#option 2
df_post <- subset(df, sample == 8)
```

## Aufgabe 2:

Am Beispiel von zwei Variablen untersuchen Sie nun, ob sich die Mittelwerte zwischen gewichteten und ungewichteten Daten unterscheiden. Die erste Variable misst die subjektive Beurteilung der aktuellen wirtschaftlichen Lage (q23). Rekodieren Sie die Variable so um, dass hohe Werte eine gute Beurteilung signalisieren. Achten Sie auf die fehlenden Werte.

```
#recode q23
df_post$wrt_akt[df_post$q23 == 1] <- 5
df_post$wrt_akt[df_post$q23 == 2] <- 4
df_post$wrt_akt[df_post$q23 == 3] <- 3
df_post$wrt_akt[df_post$q23 == 4] <- 2
df_post$wrt_akt[df_post$q23 == 5] <- 1
```

Die zweite Variable fragt nach der Verantwortung der Bundesregierung für die wirtschaftliche Entwicklung. Rekodieren Sie auch diese Variable um, sodass hohe Werte eine starke Verantwortung signalisieren.

```
#recode q25
df_post$wrt_reg[df_post$q25 == 1] <- 5
df_post$wrt_reg[df_post$q25 == 2] <- 4
df_post$wrt_reg[df_post$q25 == 3] <- 3
df_post$wrt_reg[df_post$q25 == 4] <- 2
df_post$wrt_reg[df_post$q25 == 5] <- 1
```

## Aufgabe 2: Gewichteten Datensatz erstellen

Auf der Grundlage der Nachwahl-Befragung und der Rekodierung, erstellen Sie nun mittels der Gewichtung (Variable: "w\_ow") einen gewichteten Datensatz.

```
attributes(df$w_ow)
```

```
## $label
## [1] "Gewicht: Ost/West (MZ2019)"
##
## $format.spss
## [1] "F9.2"
```

```
table(df$w_ow)
```

```
##
## 0.58712500333786 1.20658254623413
##           2848           5692
```

```
#gewichteter Datensatz
df.w <- svydesign(ids =~ 1, data = df_post, weights =~ w_ow)
```

## Aufgabe 3:

Berechnen Sie nun das arithmetische Mittel mit den ungewichteten und den gewichteten Daten für die rekodierten Variablen zur aktuellen wirtschaftlichen Lage und zur Verantwortung der Bundesregierung. Unterscheiden sich die Mittelwerte zwischen gewichteten und ungewichteten Daten?

```
#mittelwert gewichtete Daten: wirtschaftliche Lage, aktuell  
svymean(~ wrt_akt, df.w, na.rm = TRUE)
```

```
##           mean      SE  
## wrt_akt 3.4335 0.0136
```

```
#mittelwert ungewichtete Daten  
mean(df_post$wrt_akt, na.rm=TRUE)
```

```
## [1] 3.420083
```

```
#mittelwert gewichtete Daten: Verantwortung wi  
svymean(~ wrt_reg, df.w, na.rm = TRUE)
```

```
##           mean      SE  
## wrt_reg 3.4816 0.0157
```

```
#mittelwert ungewichtete Daten  
mean(df_post$wrt_reg, na.rm=TRUE)
```

```
## [1] 3.502709
```