

1 Setup

```
knitr::opts_chunk$set(echo = TRUE)
```

```
if (!require("pacman")) install.packages("pacman"); library(pacman)
```

```
## Loading required package: pacman
```

```
pacman::p_load(devtools)
```

```
if (!require("stdidx")) devtools::install_github("graemeblair/stdidx"); library(stdidx)
```

```
## Loading required package: stdidx
```

```
pacman::p_load(haven, readr, tidyverse, gtsummary, ggsci, tinytex)
```

```
here::i_am("project.Rmd")
```

```
## here() starts at /Users/laurakiemes/Library/Mobile Documents/com~apple~CloudDocs/LMU/Soziologie
```

```
theme_set(theme_light())
```

2 Datenbereinigung

```
allbus2006 <- read_dta(here::here("data", "ZA4500_v2-0-0.dta")) # gesis data
```

```
allbus2006 <- allbus2006 |>
```

```
  mutate(
```

```
    split = case_when(v3 == 2 ~ 0,  
                      TRUE ~ v3),
```

```
    split = factor(split,
```

```
                  levels = c(0, 1),
```

```
                  labels = c("Computer Assisted Self-Interviewing",
```

```

                                "Computer Assisted Personal Interviewing")),
age = case_when(v27 == 999 ~ NA,
                TRUE ~ v26),
age_cen = age - mean(age, na.rm = TRUE),
immigration_anpassen = case_when(v43 %in% c(1:7) ~ v43,
                                TRUE ~ NA),
immigration_heim = case_when(v44 %in% c(1:7) ~ v44,
                             TRUE ~ NA),
immigration_unpolit = case_when(v45 %in% c(1:7) ~ v45,
                                TRUE ~ NA),
immigration_heirat = case_when(v46 %in% c(1:7) ~ v46,
                               TRUE ~ NA),
immigration = idx_mean(immigration_anpassen,
                      immigration_heim,
                      immigration_unpolit,
                      immigration_heirat),
immigration_log = log(immigration),
split_selbst = factor(v47,
                      levels = c(0:4),
                      labels = c("Computer Assisted Personal Interviewing",
                                "ja, ohne Hilfe",
                                "ja, Hilfe vorher",
                                "ja, Hilfe dabei",
                                "nein")),
sex = factor(v174,
             levels = c(1, 2),
             labels = c("Männlich", "Weiblich")),
edu = case_when(v175 %in% c(1:5) ~ v175,
                TRUE ~ NA),
interviewer_sex = case_when(v702 == 2 ~ 0,
                             TRUE ~ v702),
interviewer_sex = factor(interviewer_sex,
                         levels = c(0, 1),

```

```

        labels = c("Männlich", "Weiblich")),
interviewer_age = v703,
interviewer_agediff = case_when(v703 == 999 ~ NA,
                                TRUE ~ abs(v703 - v27)),
interviewer_edu = factor(v705,
                        levels = c(1:4),
                        labels = c("Volks-/Hauptschulabschluss",
                                "Mittlere Reife",
                                "(Fach-)Hochschulreife",
                                "(Fach-)Hochschulabschluss"),
                        ordered = TRUE)
)

```

```
write_csv(allbus2006, here::here("data", "allbus2006.csv")) # save as csv
```

3 Results

3.1 Univariate Deskription

3.2 Zentraler Zusammenhang

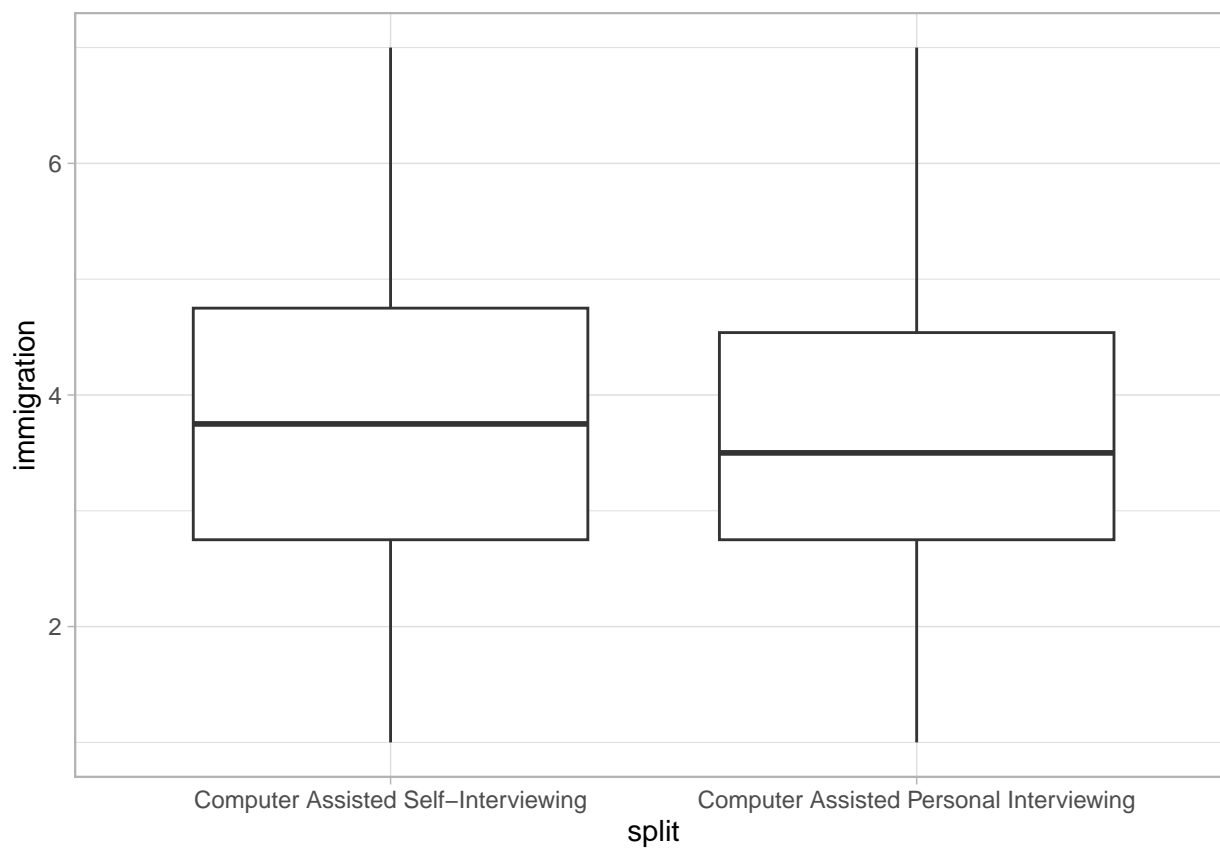
```
allbus2006 |>
  group_by(split_selbst) |>
  summarize(mean(immigration, na.rm = TRUE))
```



```
## # A tibble: 5 x 2
##   split_selbst      'mean(immigration, na.rm = TRUE)'
##   <fct>                                <dbl>
## 1 Computer Assisted Personal Interviewing      3.73
## 2 ja, ohne Hilfe                                3.55
## 3 ja, Hilfe vorher                              4.00
## 4 ja, Hilfe dabei                              4.10
## 5 nein                                          4.13
```

```
ggplot(data = allbus2006, mapping = aes(x = split, y = immigration)) +  
  geom_boxplot()
```

Warning: Removed 229 rows containing non-finite values ('stat_boxplot()').



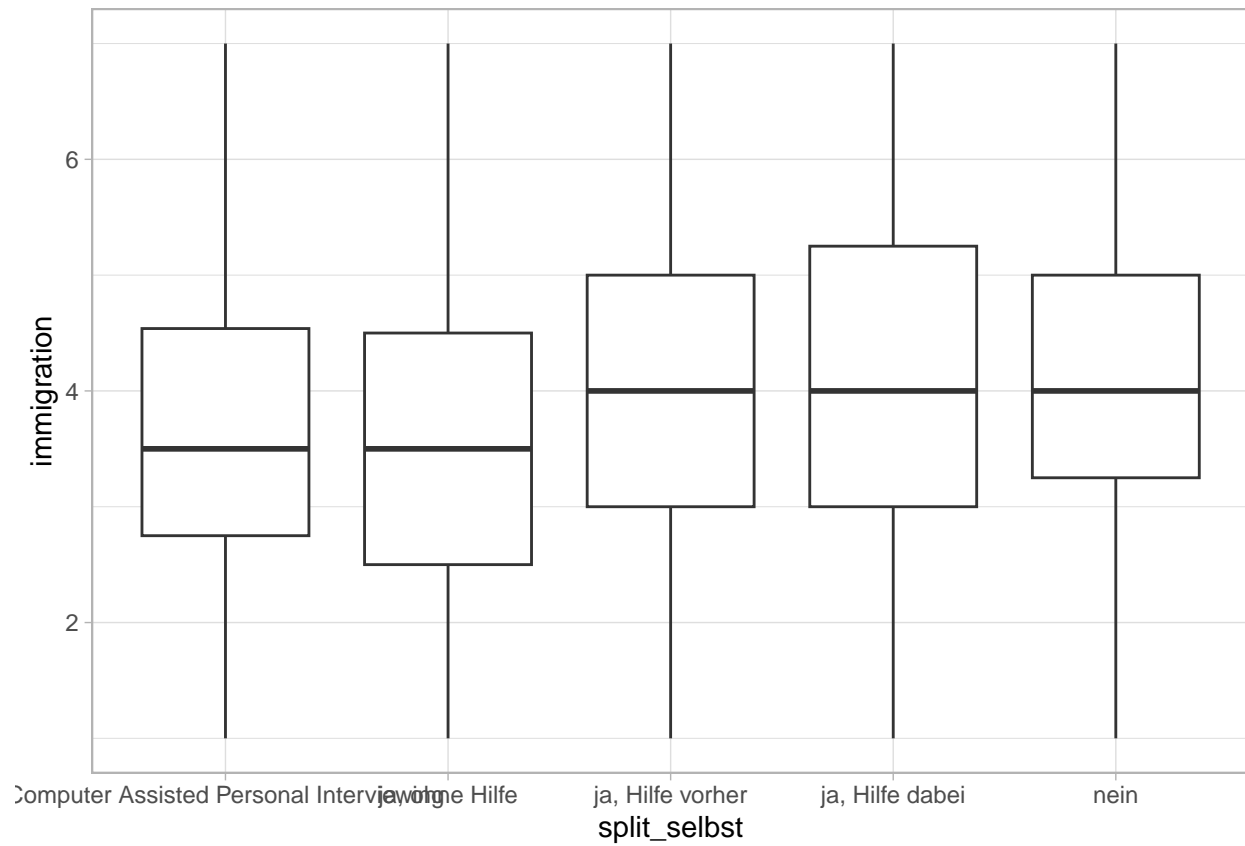
```
ggplot(data = allbus2006, mapping = aes(x = split, y = immigration)) +  
  geom_violin()
```

Warning: Removed 229 rows containing non-finite values ('stat_ydensity()').



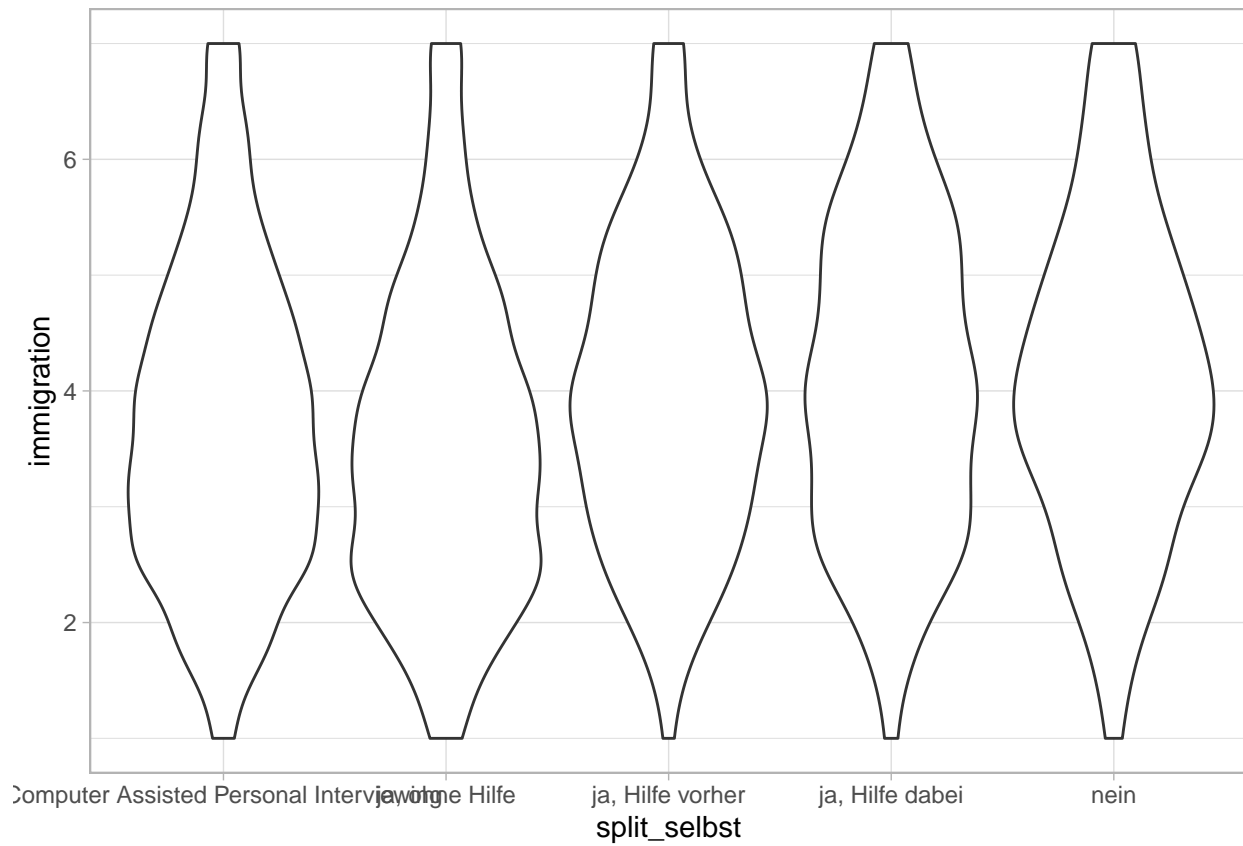
```
ggplot(data = allbus2006, mapping = aes(x = split_selbst, y = immigration)) +  
  geom_boxplot()
```

```
## Warning: Removed 229 rows containing non-finite values ('stat_boxplot()').
```



```
ggplot(data = allbus2006, mapping = aes(x = split_selbst, y = immigration)) +
  geom_violin()
```

```
## Warning: Removed 229 rows containing non-finite values ('stat_ydensity()').
```



4 Linear Regression

4.1 Unabhängigkeit

```
cor(allbus2006 |>
  select(
    interviewer_age,
    interviewer_agediff
  )) # all independant variables
```

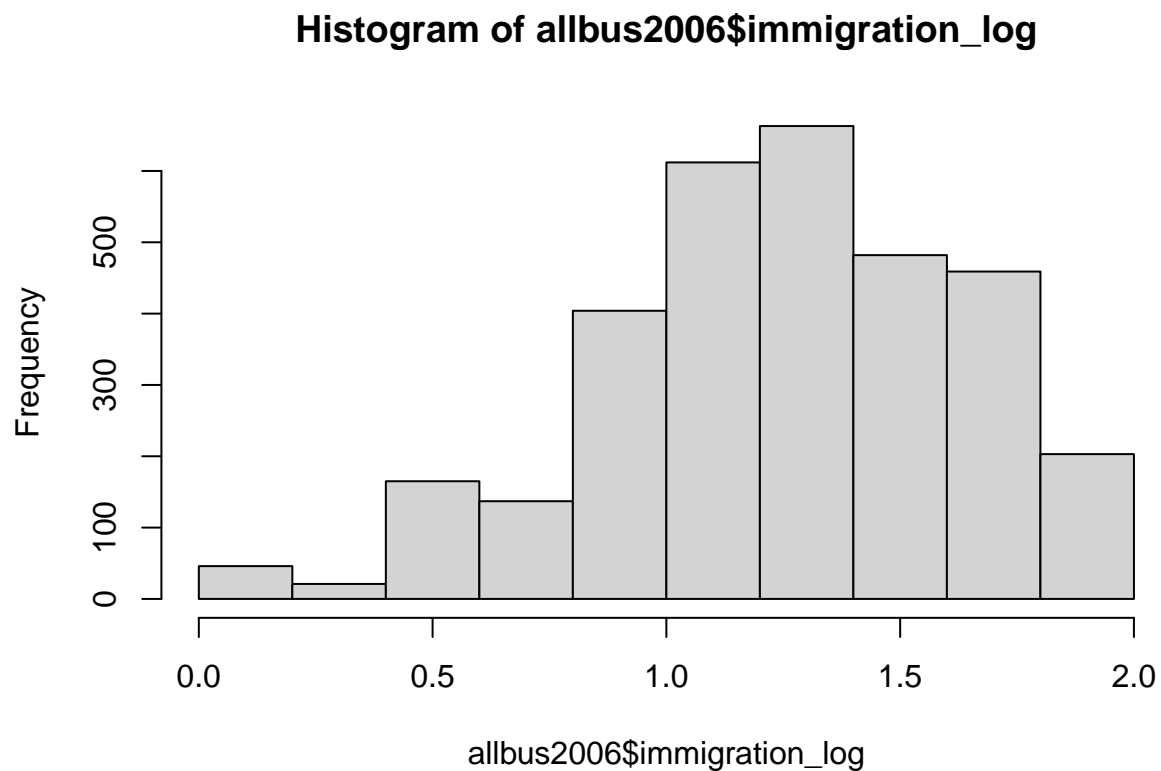
```
##               interviewer_age interviewer_agediff
## interviewer_age      1.00000000      0.01669369
## interviewer_agediff  0.01669369      1.00000000
```

```
cor(allbus2006 |>
  select(
    age,
    immigration
  ))
```

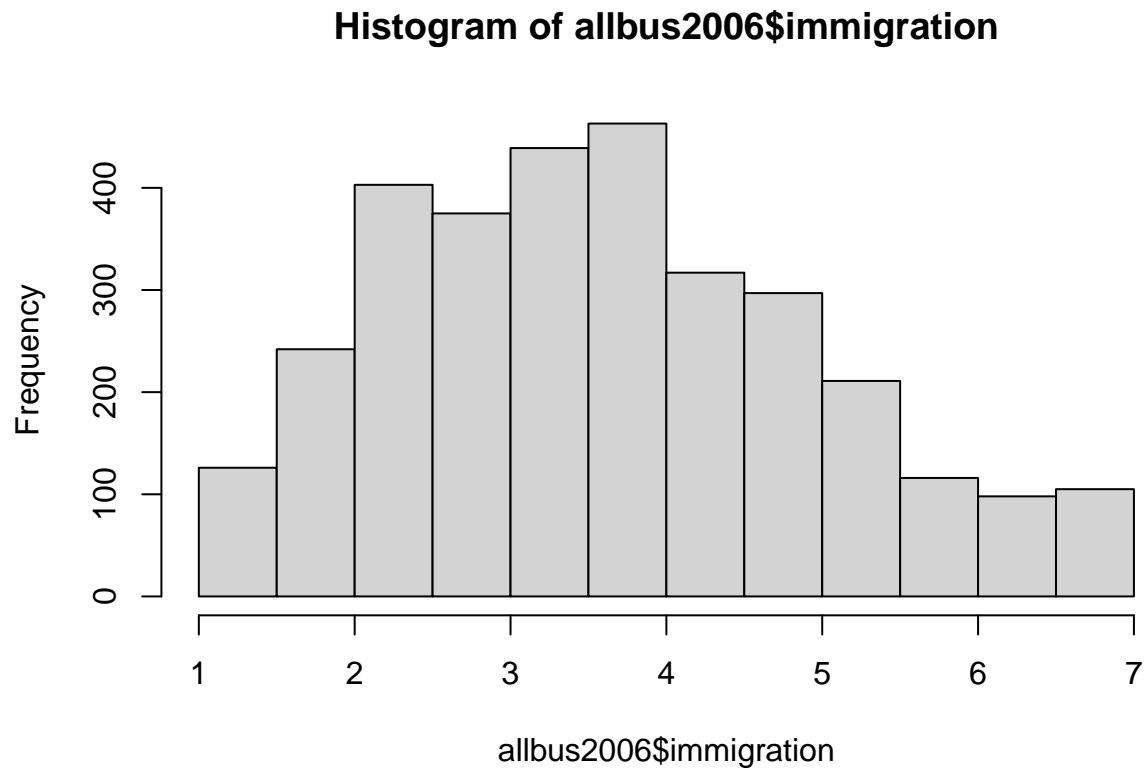
```
##           age immigration
## age           1         NA
## immigration NA           1
```

4.2 Normalverteilung

```
hist(allbus2006$immigration_log)
```



```
hist(allbus2006$immigration)
```

4.3 Einfache lineare Regression

```
split_immigration.lm <- lm(immigration ~ split_selbst, data = allbus2006)
```

```
summary(split_immigration.lm)
```

```
##
```

```
## Call:
```

```
## lm(formula = immigration ~ split_selbst, data = allbus2006)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -3.1293 -1.0509 -0.0970  0.9491  3.4491
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                3.73362    0.03414 109.352 < 2e-16 ***
## split_selbstja, ohne Hilfe  -0.18275    0.05597  -3.265 0.001106 **
## split_selbstja, Hilfe vorher 0.26220    0.09112   2.877 0.004036 **
## split_selbstja, Hilfe dabei  0.36337    0.09449   3.846 0.000123 ***
## split_selbstnein            0.39565    0.11801   3.353 0.000809 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.365 on 3187 degrees of freedom
## (229 observations deleted due to missingness)
## Multiple R-squared:  0.01715,    Adjusted R-squared:  0.01592
## F-statistic: 13.9 on 4 and 3187 DF,  p-value: 3.02e-11
```

4.4 Multiple lineare Regression

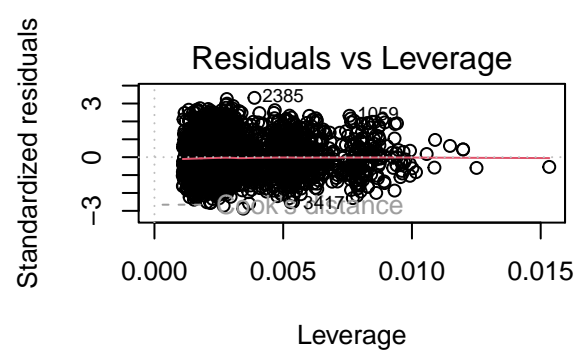
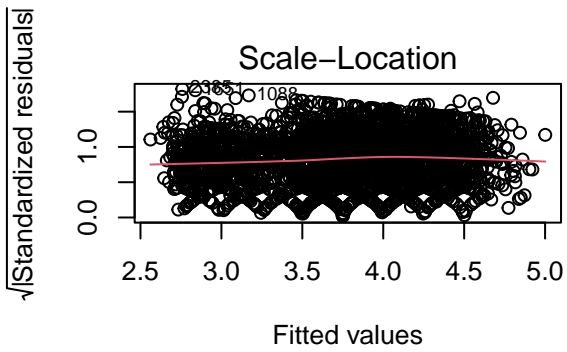
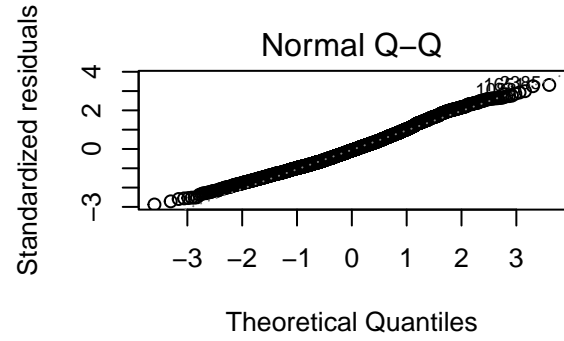
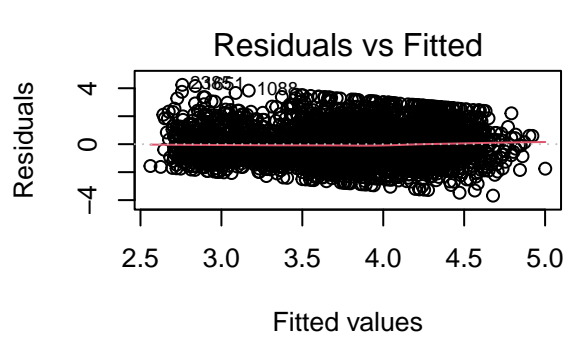
```
split_immigration_multiple.lm <- lm(immigration ~ split_selbst + age_cen + edu + sex + interviewer_sex + interviewer_agediff, data = allbus2006)
summary(split_immigration_multiple.lm)
```

```
##
## Call:
## lm(formula = immigration ~ split_selbst + age_cen + edu + sex +
##      interviewer_sex + interviewer_agediff, data = allbus2006)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6785 -0.9300 -0.0900  0.8111  4.2415
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.660188   0.092766  50.236 < 2e-16 ***
## split_selbstja, ohne Hilfe -0.044176   0.053628  -0.824  0.41015
## split_selbstja, Hilfe vorher  0.150288   0.086218   1.743  0.08141 .
```

```
## split_selbstja, Hilfe dabei    0.055588    0.090748    0.613    0.54022
## split_selbstnein              0.069251    0.112792    0.614    0.53928
## age_cen                      -0.012033    0.001595   -7.544  5.95e-14 ***
## edu                          -0.347240    0.021603  -16.074  < 2e-16 ***
## sexWeiblich                  0.120680    0.045853    2.632    0.00853 **
## interviewer_sexWeiblich      -0.004300    0.047464   -0.091    0.92782
## interviewer_agediff          0.004322    0.002184    1.979    0.04787 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.281 on 3136 degrees of freedom
## (275 observations deleted due to missingness)
## Multiple R-squared:  0.1341, Adjusted R-squared:  0.1316
## F-statistic: 53.96 on 9 and 3136 DF,  p-value: < 2.2e-16
```

4.5 Homoskedasität

```
par(mfrow=c(2,2))
plot(split_immigration_multiple.lm)
```



```
par(mfrow=c(1,1))
```