Evaluation of Socio Demographic Results

Code

This is the evaluation of the experiment's socio demographic results. It expands the master thesis by data analysis and respective code as well as short explanations. This markdown does not replace the thesis, it serves solely as an appendix.

Library Downloads

Hide

```
library(tidyverse)
library(broom)
library(ggplot2)
library(dplyr)
```

Sample Description: Socio Demographics

First, the final sample is loaded:

Hide

```
getwd()
```

```
[1] "/Users/lara-aidajopp/Documents/Programming_Stuff/Universität Mannheim/Masterthesis/evaluation"
```

Hide

```
raw_data <- readxl::read_excel("data/raw_data.xlsx", sheet = 2)
view(raw_data)
glimpse(raw_data)
```

```
Rows: 131
Columns: 27
$ `NEW CASE NR` <db> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10…
$ SD01          <db> 1, 2, 2, 1, 1, 1, 2, 1, 1, 1,…
$ SD02          <db> 11, 2, 2, 5, 2, 8, 5, 2, 3, 4…
$ SD04          <db> 5, 7, 6, 11, 10, 6, 5, 10, 9,…
$ SD04_14       <chr> NA, NA, NA, NA, NA, NA, NA, N…
$ SD05          <db> 7, 2, 6, 5, 4, 4, 4, 4, 6, 4,…
$ TR01_CP       <db> 2, 2, 2, 3, 3, 4, 4, 5, 5, 5,…
$ TR01          <db> 3, 2, 1, 2, 2, 3, 4, 1, 3, 2,…
$ AC01_01       <db> 5, 4, 5, 5, 5, 5, 4, 5, 4, 5,…
$ AC01_02       <db> 5, 4, 5, 4, 5, 4, -1, 4, 4, 4…
$ AC01_03       <db> 5, 5, 4, 4, 2, 5, 4, 5, 4, 4,…
$ DI01_01       <db> 5, 2, 4, 4, 3, 5, 4, 4, -1, 5…
$ DI01_02       <db> 5, 1, 1, 1, 1, 1, 5, -1, 5, -…
$ DI01_03       <db> 5, 1, 1, 1, 1, 1, 5, -1, 5, -…
$ DI01_04       <db> 5, 2, 1, 1, 1, 1, -1, -1, 4, …
$ UI01_01       <db> 4, 5, 5, 5, 4, 5, 4, 5, 4, 5,…
$ UI01_03       <db> 5, 3, -1, -1, 4, -1, -1, 4, 3…
$ UI01_04       <db> 5, 2, 1, 1, 1, 1, -1, 2, 4, …
$ EX01_01       <db> 5, 4, 3, 3, -1, 2, -1, 5, 4, …
$ EX01_02       <db> 5, 2, 3, 1, 4, 2, -1, 2, 2, 2…
$ EX01_03       <db> 5, 1, 3, 1, -1, 2, -1, 2, 2, …
$ EX01_04       <db> 5, 1, 1, 1, 1, 2, 4, 1, 4, -1…
$ STARTED       <dttm> 2025-09-26 14:11:10, 2025-09…
$ LASTDATA      <dttm> 2025-09-26 14:16:16, 2025-09…
$ TIME_SUM      <db> 306, 430, 369, 555, 445, 309,…
$ FINISHED      <db> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,…
```

Out of the raw data, we can at least analyze the socio demographics of each participant to validate if the sample is representative.

Hide

```
socio_demographics <- raw_data[, c("SD01", "SD02", "SD04", "SD04_14", "SD05")]

socio_demographics <- rename(socio_demographics, sex = SD01, age_group_years = SD02, degree = SD04, degree_special = SD04_14, employment = SD05)

socio_demographics <- socio_demographics %>%
  mutate(
    sex = factor(sex,
                 labels = c("Female", "Male")),
  )

glimpse(socio_demographics)
```

```
Rows: 131
Columns: 5
$ sex              <fct> Female, Male, Male, Female,…
$ age_group_years  <db> 11, 2, 2, 5, 2, 8, 5, 2, 3, …
$ degree           <db> 5, 7, 6, 11, 10, 6, 5, 10, …
$ degree_special   <chr> NA, NA, NA, NA, NA, NA, NA,…
$ employment       <db> 7, 2, 6, 5, 4, 4, 4, 4, 6, …
```

Based on that the share of sex is evaluated:

Hide

```
sex <- socio_demographics %>%
  group_by(sex) %>%
  summarise(
    absolute_count = n()
  ) %>%
  mutate(relative_share_percent = round(100 * absolute_count / sum(absolute_count), 0)
  )

glimpse(sex)
```

```
Rows: 2
Columns: 3
$ sex                  <fct> Female, Male
$ absolute_count       <int> 56, 75
$ relative_share_percent <dbl> 43, 57
```

Hide

```

```

Now, the age is evaluated and plotted using a bar plot showing the absolute count:

Hide

```
age_group_labels <- c("<18", "19-24", "25-29", "30-34", "35-39", "40-44", "45-49", "50-54", "55-59", "60-64", ">65")

age <- socio_demographics %>%
  sort_by(socio_demographics$age_group_years) %>%
  mutate(age_group_years = factor(age_group_years, labels = age_group_labels))

ggplot(data = age, aes(x=age_group_years)) + geom_bar(fill=rgb(0.1,0.4,0.5,1))
```

Hide

```

glimpse(age)
```

```
Rows: 131
Columns: 5
$ sex              <fct> Female, Male, Male, Female,…
$ age_group_years  <fct> <18, 19-24, 19-24, 19-24, 1…
$ degree           <db> 1, 7, 6, 10, 10, 10, 9, 6, …
$ degree_special   <chr> NA, NA, NA, NA, NA, NA, NA,…
$ employment       <db> 1, 2, 6, 4, 4, 4, 4, 2, 3, …
```

The following shows the age group distribution in percentage:

Hide

```
age_group_labels <- c("<18", "19-24", "25-29", "30-34", "35-39", "40-44", "45-49", "50-54", "55-59", "60-64", ">65")

age <- socio_demographics %>%
    sort_by(socio_demographics$age_group_years) %>%
  mutate(age_group_years = factor(age_group_years, labels = age_group_labels)) %>%
  group_by(age_group_years) %>%
  summarise(absolute_count = n()
  ) %>%
  mutate(relative_share_percent= round(100 * absolute_count / sum(absolute_count), 2))

view(age)
```

Since the age group has been conducted, the degree is evaluated next:

Hide

```
degree_labels <- c("School-Student", "German: Hauptschule", "German: Mittlere Reife", "German: Ausbildungsabschluss", "Technical High School Diploma", "High School Diploma", "Bachelor Professional", "Master Craftsman", "Bachelor's", "Master's", "Diploma", "Doctorate", "Else")

degree <- socio_demographics %>%
  sort_by(socio_demographics$degree) %>%
  mutate(degree = factor(degree, labels = degree_labels))

ggplot(data = degree, aes(x=degree)) + geom_bar(fill=rgb(0.1,0.4,0.5,1)) +
  coord_flip()
```


Hide

```

glimpse(degree)
```

```
Rows: 131
Columns: 5
$ sex              <fct> Female, Female, Male, Femal…
$ age_group_years <db> 1, 9, 3, 8, 8, 9, 2, 11, 5,…
$ degree           <fct> School-Student, German: Hau…
$ degree_special   <chr> NA, NA, NA, NA, NA, NA, NA,…
$ employment       <db> 1, 4, 4, 4, 4, 4, 2, 7, 4, …
```

The following shows the degree distribution in percentage:

Hide

```
degree_labels <- c("School-Student", "German: Hauptschule", "German: Mittlere Reife", "German: Ausbildungsabschluss", "Technical High School Diploma", "High School Diploma", "Bachelor Professional", "Master Craftsman", "Bachelor's", "Master's", "Diploma", "Doctorate", "Else")

degree <- socio_demographics %>%
    sort_by(socio_demographics$degree) %>%
  mutate(degree = factor(degree, labels = degree_labels)) %>%
  group_by(degree) %>%
  summarise(
    absolute_count = n(),
    degree_special = if (first(degree) == "Else") {
      paste(sort(degree_special), collapse = ", ")
    } else {
      ""
    }
  ) %>%
  mutate(relative_share_percent = round(100 * absolute_count / sum(absolute_count), 2))

view(degree)
```

Finally the employment status is evaluated:

Hide

```
employment_labels <- c("Pupil", "In Training", "Student", "Employee", "Civil Servant", "Self-Employed", "Pensioner")

employment <- socio_demographics %>%
  sort_by(socio_demographics$employment) %>%
  mutate(employment = factor(employment, labels = employment_labels))

ggplot(data = employment, aes(x=employment)) + geom_bar(fill=rgb(0.1,0.4,0.5,1)) +
  coord_flip()
```


Hide

```

view(employment)
```

The following shows the employment distribution in percentage:

Hide

```
employment_labels <- c("Pupil", "In Training", "Student", "Employee", "Civil Servant", "Self-Employed", "Pensioner")

employment <- socio_demographics %>%
  sort_by(socio_demographics$employment) %>%
  mutate(employment = factor(employment, labels = employment_labels)) %>%
  group_by(employment) %>%
  summarise(absolute_count = n()
  ) %>%
  mutate(relative_share_percent= round(100 * absolute_count / sum(absolute_count), 2))

view(employment)
```