

# 1. Abgabe (Paneldaten)

Bitte Namen eintragen

## 1 Organisation

### 1.1 Arbeitsverzeichnis festsetzen

```
# set working directory
```

### 1.2 Packages installieren und laden

```
# Packages
pkgs <- c(
  "tidyverse",
  "sjlabelled", # for variable labels
  "plm", # Panelregressionen
  "sjPlot" # um Regressionstabellen schön darzustellen
)

## Install uninstalled packages
lapply(pkgs[!(pkgs %in% installed.packages())], install.packages)

## Load all packages to library
lapply(pkgs, library, character.only = TRUE)
```

### 1.3 Daten einlesen

Wir arbeiten wieder mit dem Übungsdatensatz des SOEP (DOI:10.5684/soep.practice.v36). Es werden zwei Datensätze eingelesen. Der erste Datensatz enthält die tatsächlichen Umfragedaten, der zweite Datensatz enthält die Variablenlabels.

```

# read data
# df = read.csv("../daten/soep_uebung.csv")
# df_labels = read.csv("../daten/soep_labels.csv")
#
# # Add variable labels
# for (i in 1:nrow(df_labels)) {
#   variable_name <- df_labels$variable[i]
#   variable_label <- df_labels$variable_label[i]
#
#   # Apply the label to the corresponding column in df
#   df[[variable_name]] <- set_label(df[[variable_name]], variable_label)
# }

```

## 2 Kausalanalyse

Sie möchten herausfinden, welche Faktoren mit dem subjektiven Gesundheitszustand (gesund\_org) zusammenhängen.

### 2.1 Kausalmodell [2 Punkte]

Überlegen Sie sich ein Kausalmodell. Das heißt, überlegen Sie sich eine Variable, die einen kausalen Effekt auf den subjektiven Gesundheitszustand haben könnte und begründen Sie Ihre Wahl.

Antwort:

### 2.2 Auswahl des Regressionsmodells [4 Punkte]

Um den kausalen Effekt zu identifizieren, möchten Sie eine Regressionsanalyse durchführen und überlegen, ob Sie ein POLS, ein BE, ein RE oder ein FE Modell nutzen sollten. Beschreiben Sie Ihren Gedankenvorgang.

Antwort:

### 2.3 Kausalmodell [3 Punkte]

Bitte nennen Sie jeweils ein Beispiel für eine/n

- a) confounder

- b) collider
- c) Variable, die zu einem overcontrol bias führen könnte

und erklären Sie kurz, warum es sich um eine/n solche/n handelt.

Antwort:

### 3 Datenaufbereitung

Unabhängig vom tatsächlichen Kausalmodell soll nun eine Regressionsanalyse durchgeführt werden. Die abhängige Variable ist der subjektive Gesundheitszustand (als numerische Variable), als unabhängige Variablen sollen genutzt werden: Geschlecht (als numerische Variable mit dem Wert 1 für weiblich und 0 für männlich) und das log. Einkommen des Hauptberufs (als numerische Variable).

Bereiten Sie diese drei Variablen auf. Das heißt, definieren Sie die fehlende Werte und wandeln Sie bei Bedarf kategoriale Variablen in numerische Variablen um. Überprüfen Sie die Datenaufbereitung tabellarisch.

#### 3.1 Gesundheitszustand [1 Punkt]

#### 3.2 Geschlecht [1 Punkt]

#### 3.3 Einkommen (Hauptberuf) [1 Punkt]

Erstellen Sie eine neue Variable **ek**, die den log. Wert des Bruttoeinkommen/Jahr des Hauptberufs enthält. Nehmen Sie dazu die Variable *einkommenj1* (Bruttoeinkommen/Jahr Hauptberuf). Logarithmieren Sie diese. Da einige Beobachtungen 0 Euro Einkommen haben, addieren Sie zu jedem Wert 1 Euro **vor** der Log-Transformation.

#### 3.4 Analysesample bilden [3 Punkte]

- (a) Selektieren Sie nur die Variablen, die wir brauchen. Also die drei Variablen sowie die ID Variable (id) und die Zeit Variable (year)
- (b) Wir wenden "listwise deletion" an. Wie viele Beobachtungen gehen uns verloren?

Antwort:

## 4 Datenexploration

### 4.1 Wieviele Personenjahre sind im Datensatz? [1 Punkt]

Antwort:

### 4.2 Wie viele Personen sind im Datensatz? [1 Punkt]

Antwort:

### 4.3 Wieviele Personen nehmen pro Jahr teil? [2 Punkte]

Ist das Panel balanced?

Antwort:

### 4.4 Welche Variablen sind zeitkonstant und welche zeitveränderlich? [3 Punkte]

Antwort:

### 4.5 Teilnahme [2 Punkte]

Wie viele Personen nahmen an allen fünf Wellen teil? Und wie viele Jahre nahmen Personen durchschnittlich teil?

Antwort:

## 5 Regressionen

### 5.1 Schätzen Sie ein POLS Modell [3 Punkte]

... und interpretieren Sie die Koeffizienten.

Antwort:

### 5.2 Schätzen Sie ein BE Modell [3 Punkte]

... und interpretieren Sie die Koeffizienten.

Antwort:

### 5.3 Schätzen Sie ein FE Modell [3 Punkte]

... und interpretieren Sie die Koeffizienten.

Antwort:

### 5.4 Schätzen Sie ein RE Modell [3 Punkte]

... und interpretieren Sie die Koeffizienten.

Antwort:

### 5.5 Vergleich [5 Punkte]

Vergleichen Sie die vier Modelle. Lassen Sie sich 3 Nachkommestellen bei den Werten der Koeffizienten ausgeben. Erstellen Sie eine tabellarische Übersicht über alle Modelle.

a) Warum wird im FE Modell kein Koeffizient für *female* geschätzt?

Antwort:

b) Warum ist die Anzahl an Beobachtungen im BE Modell kleiner als die Anzahl an Beobachtungen in den anderen Modellen?

Antwort:

c) Interpretieren sie vergleichend die Koeffizienten des log. Einkommens.

Antwort:

### 5.6 Hausman Test [2 Punkte]

Sollen wir das RE oder das FE Modell nutzen? Führen Sie einen Hausman Test durch. Wie entscheiden Sie sich und warum?

Antwort:

## 6 Render [7 Punkte]

Wandeln Sie dieses Dokument in ein PDF und ein HTML Dokument um. Laden Sie bis zum 9.1.2025 zwei Dokumente hoch:

- das Quarto-File (1\_abgabe\_Nachname.qmd)
- entweder ein HTML ODER ein PDF (1\_abgabe\_Nachname.pdf oder 1\_abgabe\_Nachname.html)

Sie erhalten 7 Punkte, wenn die Prüferin lediglich das Arbeitsverzeichnis ändern muss und sonst das Quarto Dokument durchläuft.

## 7 Bewertung

Es gibt insgesamt 50 Punkte.

Punktebereich	Note
49 - 50	1,0
46 - 48	1,3
43 - 45	1,7
40 - 42	2,0
37 - 39	2,3
34 - 36	2,7
31 - 33	3,0
28 - 30	3,3
25 - 27	3,7
22 - 24	4,0