ZA6265 / ZA6266

LEO 2018 - Leben mit geringer Literalität

Nutzungshinweise LEO 2018

1 Einleitung

Die Studie *LEO 2018 – Leben mit geringer Literalität* (kurz: LEO 2018) wurde vom 01. April 2017 bis 31. Mai 2020 an der Universität Hamburg, Fakultät für Erziehungswissenschaft unter der Leitung von Prof. Dr. Anke Grotlüschen und Dr. Klaus Buddeberg durchgeführt. Die Studie wurde vom Bundesministerium für Bildung und Forschung gefördert (Förderkennzeichen: W142900).

Ein Ziel von LEO 2018 war es, die Ergebnisse der leo. – Level-One Studie (kurz: LEO 2010; Grotlüschen & Riekmann (2012)) aus dem Jahr 2010 fortzuschreiben. LEO 2010 hat erstmals in einer bevölkerungsrepräsentativen Studie den Anteil von Erwachsenen mit geringen Lese- und Schreibkompetenzen an der Deutsch sprechenden Bevölkerung im Alter von 18-64 Jahren in Deutschland ermittelt. Das zweite Ziel bestand darin, den Alltag von gering literalisierten Erwachsenen genauer beschreiben zu können. Der Fokus lag hierbei auf den Bereichen Digitalisierung, Finanzen, Gesundheit, Politik, Familie, Arbeit, Migration und Weiterbildung. In der Phase der Haupterhebung wurden 7.192 Personen befragt und deren Lese- und Schreibkompetenz erhoben. In einer ersten Buchveröffentlichung wurden vorrangig deskriptive Analysen veröffentlicht (Grotlüschen & Buddeberg 2020). Der Datensatz von LEO 2018 wird zur weiteren Forschung der (Fach-)Öffentlichkeit zur Verfügung gestellt.

2 Datenzugang und Datenschutz

Die Bereitstellung der erhobenen Daten von LEO 2018 erfolgt in Form zweier Datensätze – dem Public-Use-File (PUF) und dem Scientific-Use-File (SUF). Beide Datensätze sind beim GESIS - Leibniz-Institut für Sozialwissenschaften (gesis.org) archiviert und über den GESIS Datenservice https://www.gesis.org/angebot/daten-finden-und-abrufen/datenservice erhältlich. Daten und Dokumente des Public-Use-Files sind nach Registrierung zugänglich und für die akademische Forschung und Lehre freigegeben. In diesem Datensatz fehlen eine Reihe an datenschutzsensiblen, vorrangig soziodemografischen Variablen. Der Scientific-Use-File ist umfangreicher, aber ebenfalls um einige datenschutzsensible Variablen bereinigt bzw. mit gruppierten Variablen versehen (siehe Kapitel 6). Der Zugang zum umfangreicheren Datensatz des Scientific-Use-Files ist ebenfalls über GESIS möglich, erfolgt jedoch nur bei einem berechtigten Forschungsinteresse und nach Prüfung und Freigabe durch die Projektleitung. Für die Nutzung des Scientific-Use-Files ist der Abschluss eines Datennutzungsvertrags erforderlich:

GESIS Datennutzungsvertrag – de:
http://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Datennutzungsvertrag_Off-Site.pdf

GESIS Datennutzungsvertrag – en:
http://www.gesis.org/fileadmin/upload/dienstleistung/daten/secure_data_center/GESIS_Data_Use_Agreement_Off-Site.pdf

3 Beschreibung der verfügbaren Dokumente

Neben den vorliegenden Nutzungshinweisen und dem Datensatz stehen noch eine Reihe von weiteren Dokumenten zum Download zur Verfügung.

Das Dokument **„ZA6265_fb.pdf“** enthält zwei Version des **LE0 2018 Fragebogens**: den Fragebogenentwurf und den von Kantar Public umgesetzten Fragebogen für die computerunterstützte persönliche Befragung. Im Fragebogenentwurf sind die im Interview gestellten Fragen ausführlich beschrieben. Hierzu gehören die knappe Dokumentation des Forschungsstandes, bzw. der Theorie, die Rationale hinter der Frage, welchem Konstrukt die Frage zugerechnet ist, und ob die Frage einer anderen Studie entlehnt ist. Der LEO 2018 CAPI-Fragebogen gibt die exakte Reihenfolge der Fragestellung im Interview wieder. Weiterhin sind in diesem Teil des Dokuments die Anweisungen für die Interviewer*innen, etwaige abweichende Fragestellungen, die Filterführung und im Verlauf des Interviews generierte Variablen dokumentiert.

Eine Beschreibung der Durchführung der Erhebung, der anschließenden Auswertung des Assessments und der Aufbereitung der Daten erfolgt im **LEO 2018 Methodenbericht (ZA6265_mb.pdf).** Die Autorinnen Frauke Bilger und Alexandra Strauß waren federführend für die Erhebung der LEO-Studie beim Umfrageinstitut Kantar verantwortlich.

Das **LEO 2018 Codebook (ZA6265_cod.xlsx)** gibt eine Übersicht aller im Public-Use-File und Scientific-Use-File vorhanden Variablen. Die Variablen werden zudem ausführlich beschrieben mit genauem Fragetext, den im Datensatz verwendeten Labeln, den verwendeten Skalen und Werten. Zudem gibt das Dokument Auskunft darüber, welche Variablen generiert sind.

4 Verwendung Plausibler Werte

Für LEO 2018 wurden auf Grundlage eines Rasch-Modells mit einer latenten Regression für jede getestete Person zehn sogenannte plausible Werte (PVs, plausible Values) gezogen, die als Schätzer für die Personenfähigkeiten (Lese- und Schreibkompetenz) dienen. Das genaue Vorgehen ist bei Dutz und Hartig (2020) beschrieben. Zu beachten ist, dass plausible Werte nicht zur individuellen Diagnostik geeignet sind, sondern nur aussagekräftig sind für die Population bzw. Sub-Populationen.

Um Berechnungen mit plausiblen Werten durchzuführen, müssen bestimmte Rechenregeln beachtet werden. Statistische Analysen müssen dabei jeweils entsprechend der Anzahl der PVs mehrmals durchgeführt werden und werden so bei statistischen Analysen wie multiple Imputationen behandelt. Für Punktschätzer (etwa Mittelwerte oder Regressionskoeffizienten) kann das arithmetische Mittel dieser mehrfachen Auswertungen als gepooltes Ergebnis

berechnet werden. Andere Schätzer, wie etwa Standardfehler, können nach anderen Formeln berechnet werden. Einführungen zu dieser Thematik finden sich bei Lüdtke & Robitzsch (2017) oder bei von Davier, Gonzalez & Mislevy (2009). Für solche Analysen werden üblicherweise Statistikprogramme genutzt. Erläuterungen zum Vorgehen bei R, SPSS und Stata finden sich weiter unten.

Im Datensatz zu LEO 2018 liegen mehrere verschiedene Sätze von plausiblen Werten vor. Diese sind:

- **`pv1`** bis **`pv10`**: Diese PVs repräsentieren die Lese- und Schreibkompetenz auf der kontinuierlichen LEO-Skala (0-100).
- **`alpha_pv1`** bis **`alpha_pv10`**: Diese PVs repräsentieren die Lese- und Schreibkompetenz in den Alpha-Leveln.

5 Analysen mit einschlägigen Statistikprogrammen

Komplexe Datensätze können mit einer Reihe von Statistikprogrammen bearbeitet und analysiert werden. Nachfolgend werden für drei Programme – R, SPSS und STATA – die Vorbereitung zu Analyse und einige Beispiele für Analysemethoden vorgestellt. Das LEO-Team der Universität Hamburg selbst hat für alle publizierten Ergebnisse das Programm R verwendet. Bei der Arbeit mit den unterschiedlichen Programmen können die Ergebnisse zwischen den Programmen in einigen Fällen um eine Nachkommastelle voneinander abweichen. Dies ist auf abweichende Rundungsvorgänge während der Ausführung der Berechnungen der Programme zurückzuführen.

5.1 Datenausfälle

Die Kodierung fehlender Antworten ist nicht über alle Variablen hinweg einheitlich. Im Falle von Item-Nonresponse orientiert sich der vergebene Wert z.B. für eine Antwortverweigerung am konkreten Wertebereich gültiger Antworten der jeweiligen Variable. Entsprechend variiert z. B. der vergebene Wert für "Keine Angabe" von Variable zu Variable zwischen 9, 99 oder 999. Fehlende Werte dieser Art sind im Datensatz grundsätzlich nicht als „fehlend“ vordefiniert und werden in Analysen miteingeschlossen. Fehlende Antworten durch missing by design, d.h. durch Filterführung bedingte Datenausfälle, wurden als „system missings“ gehandelt. System missings werden bei Analysen nicht berücksichtigt. Bevor Variablen für Analysen genutzt werden, sollte der Wertebereich gültiger Antworten kontrolliert und ggf. einzelne Werte von Berechnungen bewusst ein- bzw. ausgeschlossen werden.

5.2 R

R ist eine freie und kostenlose Statistiksoftware, die für Windows, Linux und Mac OS verfügbar ist[1]. Um die Arbeit mit R zu erleichtern empfiehlt sich die Verwendung von RStudio[2].

Auswertungen von gewichteten Querschnittsstudien mit plausiblen Werten (PVs) können in R mit den Paketen survey[3] und mitools[4] durchgeführt werden. Alle für die Auswertung notwendigen Pakete sind kostenfrei über das *Comprehensive R Archive Network* (CRAN) verfügbar[5] und können innerhalb der R-Umgebung heruntergeladen und installiert werden.

5.2.1 Vorbereitungen

Vor der Benutzung des Datensatzes in R müssen die benötigten Pakete installiert und geladen sowie der Datensatz geladen werden.

```
# Installation der benötigten Pakete.
# Sollte nur bei der erstmaligen Verwendung ausgeführt werden. Dazu # entfern
en
#install.packages(c("survey", "mitools", "haven", "tidyselect", "magrittr", "
dplyr", "purrr", "labelled"))

# Laden der benötigten Pakete.
library(haven)
library(survey)
library(mitools)
library(tidyselect)
library(magrittr)
library(dplyr)
library(purrr)
library(labelled)

# Einlesen der Daten mit Hilfe des haven-Pakets.
# Bitte beachten Sie die Pfadangabe anzupassen und dabei keine Backslashes zu
verwenden.
df <- haven::read_sav("C:/Pfad/und/Dateiname/anpassen/LEO2018-Use-File.sav")
```

Da im nachfolgenden Schritt die Daten entsprechend der Anzahl der PVs in LEO verzehnfacht wird, können für die vorgesehenen Analysen die nicht benötigten Variablen aussortiert werden.

---

[1] https://www.r-project.org

[2] https://rstudio.com

[3] https://cran.r-project.org/web/packages/survey/index.html

[4] https://cran.r-project.org/web/packages/mitools/index.html

[5] https://cran.r-project.org

Es sollte jedoch nicht vergessen werden, die PVs und die Gewichtungsvariable im Datensatz zu behalten.

Für Berechnungen mit PVs müssen anschließend das Survey-Design und die PVs spezifiziert werden. Faktorvariablen sollten mit `labelled::to_factor()` vor der Erstellung des survey-Objekts umgewandelt werden. Anschließend steht für die weiteren Analysen die PV-Variable pv für Berechnungen auf der kontinuierlichen LEO-Skala und alpha_pv für Berechnungen mit den Alpha-Levels zur Verfügung.

```
# Umwandlung von Faktorvariablen
df$fam002 <- labelled::to_factor(df$fam002)
df$f001 <- labelled::to_factor(df$f001)
df$altgr4 <- labelled::to_factor(df$altgr4)

# Spaltennamen der PVs mit Hilfe eines regulären Ausdrucks (regexp) finden
# kontinuierliche LEO-Skala
lit <- grep("^pv[0-9]+", colnames(df), value = TRUE)
# Alpha-Levels
alp <- grep("^alpha_pv[0-9]+", colnames(df), value = TRUE)

# Liste der Daten erstellen mit einer pv-Variable für LEO-Skala sowie Alpha-L
evels
all <- df %>% dplyr::select(-tidyselect::one_of(lit), -tidyselect::one_of(alp
))
all <- colnames(all)
dfList <- lit %>%
  map(function(x) {dplyr::select(df, tidyselect::all_of(all), paste0("alpha_"
, x), tidyselect::all_of(x)) %>%
    dplyr::rename(pv := !!x) %>%
    dplyr::rename(alpha_pv := !!paste0("alpha_", x)) %>%
    dplyr::mutate(alpha_pv = labelled::to_factor(alpha_pv))})

# Erstellen des survey-Objekts
leo.sd <- survey::svydesign(
  ids = ~0,
  weights = ~pgewges,
  data = mitools::imputationList(dfList))
```

5.2.2 Häufigkeitsverteilung der Alpha-Levels

```
table_alpha <- MIcombine(with(leo.sd, svymean(~alpha_pv)))
summary(table_alpha)
```

5.2.3 Kreuztabellen zwischen Alpha-Levels und einer weiteren Variable

```
cross_alpha <- MIcombine(with(leo.sd, svyby(~alpha_pv, ~fam002, svymean)))
summary(cross_alpha)
```

5.2.4 Regressionsanalyse

Regressionsanalyse mit den kontinuierlichen Alpha-Werten (pv) als abhängiger Variable und den Variablen Geschlecht (`f001`) und Alter (`altgr4`) als unabhängigen Variablen.

```
reg_pv <- MIcombine(with(leo.sd, svyglm(pv ~ f001 + altgr4)))
summary(reg_pv)
```

5.3 SPSS

Die Analysen mittels PVs mit der Statistiksoftware SPSS erfordern zunächst eine Bearbeitung des Datensatzes. Für die Analysen ohne Verwendung der Alpha-Levels ist eine Bearbeitung nicht erforderlich.

Die Bearbeitung des Datensatzes ist notwendig, da SPSS für jeden der zehn PVs einen eigenen Datensatz benötigt. Die Syntax mit Erläuterungen zur Bearbeitung befindet sich im Anhang dieses Dokuments. Im Rahmen der Bearbeitung werden drei neue Variablen gebildet:

- **Imputation_**: Weist die benötigten zehn Datensätze aus und nimmt daher Werte von 1 bis 10 an.
- **IW_62**: Gibt über die im Datensatz vorhandenen Variablen pv1 bis pv10 den zugehörigen PV an.
- **IW_Alpha_62**: Gibt über die Datensatz vorhandenen Variablen alpha_pv1 bis alpha_pv10 den zugehörigen Alpha-Level an. Daher sind hier die Werte von 1 bis 5, äquivalent zu den Alpha-Levels 1 bis 4 sowie 5 und höher. Anhand der Variable IW_Alpha_62 ist es möglich, in einer rekodierten Variable die Alpha-Levels zusammenzufassen, etwa indem die Werte 1 bis 3 zusammengefasst werden und somit den Bereich der geringen Literalität darstellen.

Die Analysen werden für jeden Datensatz einzeln bzw. jeden PV durchgeführt und anschließend der Durchschnittswert der zehn zuvor berechneten Werte ermittelt. D.h. für die Analysen sind nur die Fallzahlen, die in der Ausgabe-Datei von SPSS unter der Rubrik „Kombiniert“ dargestellt werden, relevant. Leider werden die Durchschnittswerte in der Ausgabe-Datei nicht mit Prozentwerten, sondern lediglich mit den gewichteten Fällen versehen. Dieser Umstand erfordert die selbstständige Berechnung der Prozentwerte auf Grundlage der kombinierten Fallzahlen, z.B. unter zur Hilfenahme von Excel.

5.3.1 Vorbereitungen

Vor der Analyse muss die Datendatei geöffnet werden und anhand der Variable Imputation_ sortiert und aufgeteilt werden. Dies ist bei jedem neuen öffnen der Datei nötig, um die zehn PVs für die weiteren Analysen zu berechnen. Die Syntax hierzu ist:

```
*/ 1. Öffnen der Arbeitsdatei und Datei anhand der Variable Imputation_ sorti
eren und Gruppen vergleichen aktivieren.
GET FILE= 'SPEICHERORT\LEO2018_Enddaten_mit_PVs.sav'.
SORT CASES BY Imputation_.
SPLIT FILE LAYERED BY Imputation_.
```

Anschließend wird festgelegt, welche Gewichtung die Berechnungsgrundlage bildet. Auch dieser Schritt ist bei jedem neuen Öffnen erforderlich. Zum einen kann die Variable pgewges gewählt werden. Diese Variable gibt den Personengewichtungsfaktor an und wird verwendet, um die gewichteten Fallzahlen zu bekommen. Zum anderen ermöglicht die Variable phochges die Gewichtung nach einem Hochrechnungsfaktor. Auf diese Weise wird der Bevölkerungsanteil in Tausend wiedergegeben. Die Syntax lautet:

```
*/ 2. Gewichtung nach Personengewichtungsfaktor Gesamtstichprobe.
WEIGHT BY pgewges.
```

bzw.

```
WEIGHT BY phochges.
```

5.3.2 Häufigkeitsverteilung der Alpha-Levels

```
*/ 1. Öffnen der Arbeitsdatei und Datei anhand der Variable Imputation_ sorti
eren und Gruppen vergleichen aktivieren.
*/ Syntax siehe oben

*/ 2. Gewichtung nach Personengewichtungsfaktor Gesamtstichprobe.
*/ Syntax siehe oben

*/ 3. Auswertung der Bevölkerung nach Alpha-Levels (IW_Alpha_62)
FREQUENCIES VARIABLES=IW_AlphaLevel_62
  /ORDER=ANALYSIS.
```

5.3.3 Kreuztabellen zwischen Alpha-Levels und einer weiteren Variable

```
*/ 1. Öffnen der Arbeitsdatei und Datei anhand der Variable Imputation_ sorti
eren und Gruppen vergleichen aktivieren.
*/ Syntax siehe oben

*/ 2. Gewichtung nach Personengewichtungsfaktor Gesamtstichprobe.
*/ Syntax siehe oben

*/ 3. Besitz einer Fahrerlaubnis (Variable fam002) gekreuzt mit den Alpha-Lev
els (IW_Alpha_62)
 CROSSTABS
  /TABLES=fam002 BY IW_AlphaLevel_62
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

5.3.4 Regressionsanalyse

Regressionsanalyse mit den kontinuierlichen Alpha-Werten (IW_62) als abhängiger Variable und den Variablen Geschlecht (f001) und Alter (altgr4) als unabhängigen Variablen.

```
*/ 1. Öffnen der Arbeitsdatei und Datei anhand der Variable Imputation_ sorti
eren und Gruppen vergleichen aktivieren.
```

```
*/ Syntax siehe oben

*/ 2. Gewichtung nach Personengewichtungsfaktor Gesamtstichprobe.
*/ Syntax siehe oben

*/ 3. Umkodieren und Erzeugen von Dummy-Variablen. Die Abkürzung dv steht hie
rbei für Dummy-Variable.
*/ 3.1. Umkodieren der Variable f001 (Geschlecht) in die beiden Ausprägungen
0 = männlich und 1 = weiblich (bisher: 1 = männlich und 2 = weiblich).
RECODE f001 (1=0) (2=1) INTO dv_f001.
VARIABLE LABELS dv_f001 'Geschlecht'.
VALUE LABELS dv_f001 0 'männlich' 1 'weiblich'.
FORMATS dv_f001 (F1.0).
EXECUTE.

*/ 3.2 Umkodieren der Variable altgr4 (Altersgruppen [4 Gruppen]) in mehrere
dichotome Variablen.
COMPUTE dv_alter_a = (altgr4=1).
VARIABLE LABELS dv_alter_a '18-29 Jahre'.

COMPUTE dv_alter_b = (altgr4=2).
VARIABLE LABELS dv_alter_b '30-39 Jahre'.

COMPUTE dv_alter_c = (altgr4=3).
VARIABLE LABELS dv_alter_c '40-49 Jahre'.

COMPUTE dv_alter_d = (altgr4=4).
VARIABLE LABELS dv_alter_d '50-64 Jahre'.

VALUE LABELS dv_alter_a dv_alter_b dv_alter_c dv_alter_d 0 'nicht ausgewählt'
1 'ausgewählt'.
FORMATS dv_alter_a dv_alter_b dv_alter_c dv_alter_d (F1.0).
EXECUTE.

*/ 4. Durchführen der Regressionsanalyse in zwei Modellen. Im ersten Modell w
ird nur das Geschlecht einbezogen. Im zweiten Modell zusätzlich die Altersgru
ppen.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT IW_62
/METHOD=ENTER dv_f001
/METHOD=ENTER dv_alter_b dv_alter_c dv_alter_d.
```

5.4 Stata

Um LEO-Daten in Stata auszuwerten, müssen zum einen die PVs als multiple Imputationen importiert werden und zum anderen die Gewichtung mittels des svy-Befehls eingestellt werden. Dies kann mit folgender Syntax erfolgen:

```
clear

use "N:\Erzwiss\Leo2018\Daten\2020-07-08-LEO2018-Scientific-Use-File.dta"

gen ALPHAPV=.
gen PV=.

capture mi unset

mi import wide, imputed(ALPHAPV=alpha_pv1-alpha_pv10 PV=pv1-pv10) clear drop
mi svyset[pw=phochges]
```

Anschließend können Auswertungen mit einer Kombination der Befehle mi und svy durchgeführt werden.

5.4.1 Häufigkeitsverteilung der Alpha-Levels

```
mi estimate: svy: proportion ALPHAPV
```

5.4.3 Regressionsanalyse mit PVs

```
mi estimate: svy: reg PV f001
```

6 Variablen

Der Datensatz der LEO-Studie 2018 umfasst neben den in der Umfrage direkt erhobenen Variablen aus dem Fragebogen zusätzlich generierte Variablen und aus Datenschutzgründen rekodierte Variablen. Einen Überblick mit Beschreibungen der generierten, rekodierten und standardisierten Variablen bietet dieses Kapitel.

6.1 Generierte Variablen

Eine Übersicht über die generierten Variablen ist mit dem **Codebook** möglich, wenn auf dem Blatt „Variablen“ in Spalte H unter der Überschrift „generiert“ die Auswahl entsprechend vorgenommen wird. Im **CAPI-Fragebogen** finden sich Erläuterungen zu Zweck und Ausgangsvariablen zu einer Vielzahl von generierten Variablen. Nachfolgend werden diejenigen generierten Variablen beschrieben, die nicht im CAPI-Fragebogen aufgeführt sind. Hierbei handelt es sich um generierte Variablen, die erst im Verlauf der Analyse des Datensatzes erstellt wurden und nicht bereits im Zuge der Datenerhebung und anschließenden Datenaufbereitung durch das Umfrageinstitut Kantar. Eine Ausnahme bilden die drei ersten Variablen über die Gemeindegröße. Diese drei Variablen werden im CAPI-Fragebogen zwar genannt, dennoch wird nachfolgend die unterschiedliche Bedeutung erläutert. Bei generierten Variablen, die lediglich

Ausprägungen der zu Grunde liegenden Variablen zusammenfassen, wurde auf eine Beschreibung der Kodierung verzichtet.

gkpol, bik, bik4: Die Größe der Gemeinde gibt gkpol wieder. Hier wird zur Einordnung die Einwohner*innenzahl benutzt. Mit der Variable bik wird das Einzugsgebiet, in dem die Gemeinde liegt, beschrieben (zur Definition des Bik-Gemeindetyps: Behrens & Wiese (2013)). Befindet sich etwa eine Gemeinde mit 9000 Einwohner*innen in einem Einzugsgebiet mit insgesamt 50 000 Einwohner*innen, wird für gkpol der Wert 3 kodiert und für bik der Wert 6. Die Variable bik4 ist eine Zusammenfassung auf Grundlage der Variable bik.

altgr4: Ist eine Zusammenfassung der Altersangaben in vier Gruppen; Grundlage sind die Variablen f002a1 und f002a2, die aufgrund des Datenschutzes nicht im Public-Use-File (PUF) enthalten sind.

wohnen_partnerschaft: Gibt Auskunft, ob die befragte Person in einer Partnerschaft lebt und wenn ja, ob sie in einem Haushalt mit der*dem Partner*in zusammenlebt. Grundlage sind die Variablen sdg001, sdg002, sdg003. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
1	wohnt mit Partner*in zusammen	(sdg002=1 und sdg003=1)[6]
2	wohnt nicht mit Partner*in zusammen	(sdg002=1 und sdg003=3) sowie (sdg002=1 und sdg001=1)
3	lebt nicht in einer Partnerschaft	sdg002=2
9	keine Angabe	sdg002=9 sowie sdg003=9

berabk: Ist eine Zusammenfassung des höchsten erreichten Berufsabschlusses der befragten Person. Grundlage sind die Variablen der Reihe f04401 bis f04499. Bei Mehrfachnennung sticht die höchste Nennung (f04401 bis f04420). Die Kodierung:

Wert	Wertbeschriftung	Kodierung
0	Keine Berufsausbildung	wenn nichts genannt von f04401 bis f04499 oder f04498
1	berufsvorbereitende Maßnahme	wenn f04401 bis f04404
2	betriebliche Lehre/Berufsfachschule	wenn f04405 bis f04408
3	Meister/Fachschule	wenn f04409 bis f04415
4	(Fach-)Hochschulabschluss, Promotion	wenn f04416 bis f04420

[6] In Klammern stehende Textteile bilden eine feste Verbindung. „Und“ ist in diesem Fall eine Bedingung und keine Aufzählung.

9	nicht zuordbar/keine Angabe	wenn f04499 oder nur f04421

schulab_eltern: Ist eine Zusammenfassung des höchsten erreichten Schulabschlusses der Eltern der befragten Person. Grundlage sind die Variablen schulab_vater und schulab_mutter. Der höhere Schulabschluss sticht den niedrigeren Schulabschluss.

text_erst_od: Gibt Auskunft darüber, in wie vielen Herkunftssprachen die befragte Person anspruchsvolle Texte lesen und schreiben kann, wenn Deutsch keine Herkunftssprache ist. Grundlage sind die Variablen mig005_erst1 und mig005_erst2. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
0	in keiner Sprache	(mig005_erst1≠4 und mig005_erst2≠4)
1	in einer Sprache	(mig005_erst1=4 oder mig005_erst2=4)
2	in zwei Sprachen	(mig005_erst1=4 und mig005_erst2=4)

text_fremd_od: Gibt Auskunft darüber, in wie vielen Fremdsprachen außer Deutsch die befragte Person anspruchsvolle Texte lesen und schreiben kann. Grundlage sind die Variablen mig005_eng und mig005_and. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
0	in keiner Sprache	(mig005_eng≠4 und mig005_and≠4)
1	in einer Sprache	(mig005_eng=4 oder mig005_and=4)
2	in zwei Sprachen	(mig005_eng=4 und mig005_and=4)

text_alle_od: Gibt Auskunft darüber, in wie vielen Herkunfts- oder Fremdsprachen außer Deutsch die befragte Person anspruchsvolle Texte lesen und schreiben kann. Grundlage sind die zuvor generierten Variablen text_erst_od und text_fremd_od, die zusammengefasst wurden.

erwk: Ist eine Zusammenfassung der Variable f006. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
1	erwerbstätig	(f006=1 oder f006=2)
2	arbeitslos	f006=3
3	erwerbsunfähig	f006=9
4	Hausfrau*mann, Elternzeit	(f006=4 oder f006=10)
5	Rentner*in, Pensionär*in	f006=8
6	in Ausbildung	(f006=5 oder f006=6)
7	Sonstiges	(f006=7 oder f006=11)
9	keine Angabe	f006=99

ges025_zus: Gibt Auskunft darüber, ob mindestens eine gesundheitliche Beeinträchtigung vorliegt. Grundlage sind die Variablen der Reihe ges025_1 bis ges025_6.

mehrspr: Gibt Auskunft, ob die befragte Person ein- oder mehrsprachig ist und ob sie Deutsch spricht. Grundlage sind die Variablen mig001a und erstspr. Letztere ist nicht im Public-Use-File (PUF) und Scientific-Use-File (SUF) enthalten. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
1	einsprachig, nur Deutsch	(mig001a=2 und erstspr=1)
2	mehrsprachig, mit Deutsch	(mig001a=1 und erstspr=1)
3	mehrsprachig, ohne Deutsch	(mig001a=1 und erstspr=2)
4	einsprachig, kein Deutsch	(mig001a=2 und erstspr=2)

mig011gr: Ist eine Zusammenfassung der Altersangaben zum Alter bei Zuzug in vier Gruppen. Grundlage ist die Variable mig011, die nicht im Public-Use-File (PUF) und Scientific-Use-File (SUF) vorhanden ist.

migragen2: Gibt Auskunft über den Migrationshintergrund der befragten Person bis zur Elterngeneration. Grundlage sind die Variablen mig009, mig018 und mig023. Die Kodierung:

Wert	Wertbeschriftung	Kodierung
0	kein Migrationshintergrund	mig009=1
1	Geburtsland ZP nicht Deutschland	mig009=2
2	Geburtsland Eltern nicht Deutschland	mig009=2 und (mig018=2 oder mig023=2)
9	keine Angabe	mig009=9 oder mig018=9 oder mig023=9

pv1 bis **pv10**, **logit_pv1** bis **logit_pv10** und **alpha_pv1** bis **alpha_pv10**: Sind die als Schätzer für die Personenfähigkeit durch eine latente Regression erzeugten plausiblen Werte.

6.2 Rekodierte Variablen

Die Bereitstellung des Datensatzes in Form eines Public-Use-Files und eines Scientific-Use-Files erfordert aus Gründen des Datenschutzes eine Anonymisierung von sensiblen Variablen. Die nachfolgenden rekodierten Variablen fassen die Ausgangsvariablen in übergeordneten Kategorien zusammen. Die rekodierten Variablen entstammen drei Themenbereichen. Hierbei handelt es sich um die Familien- bzw. Haushaltszusammensetzung, die Geburtsländer und Staatsangehörigkeit sowie die Sprachkenntnisse.

Bei der Rekodierung der Variablen, die Auskunft über die Familien- und Haushaltsstruktur geben, wurden Angaben ab einem bestimmten Wert zusammengefasst. Zum Beispiel gibt die Variable sdg005 Auskunft über die Anzahl der eigenen Kinder. Die Anzahl von drei Kindern wurde in der rekodierten Variable noch als eigenständige Kategorie belassen. Alle darüber hinaus gehenden Angaben wurden als vier oder mehr Kinder in einer Kategorie zusammengefasst. Nach diesem Schema wurden folgende Variablen zusammengefasst:

sdg001a_rec: Ist die Zusammenfassung der Anzahl der Personen, die mit der befragten Person in einem Haushalt leben.

sdg005_rec: Ist die Zusammenfassung der Anzahl der Kinder der befragten Person.

sdg006_1_rec, **sdg006_2_rec**, **sdg006_3_rec**, **sdg006_4_rec**, **sdg006_5_rec**: Bei diesen rekodierten Variablen handelt es sich jeweils um die Zusammenfassung der Anzahl der Personen mit einem gewissen Alter, die im selben Haushalt mit der befragten Person leben.

Maßgeblich für das Zusammenlegen von Ländern und Sprachen ist die Anzahl der in Deutschland lebenden Personen, die aus den jeweiligen Ländern stammen bzw. Sprachen der Herkunftsländer sprechen. Leben weniger als 50.000 Personen, die aus dem jeweiligen Land stammen, in Deutschland, werden Länder zusammengelegt. Bei diesem Vorgehen orientiert sich LEO an PIAAC, dass ebenfalls einen Grenzwert von 50.000 Personen benutzt. Weiterhin orientiert sich LEO bei der Bildung der zusammengelegten Länderkategorien an PIAAC (Perry, Helmschrott, Konradt & Maehler, 2017). Als Grundlage für die Ermittlung des Grenzwertes diente eine Veröffentlichung des Statistischen Bundesamtes über die Herkunftsländer der ausländischen Bevölkerung in Deutschland für das Erhebungsjahr der LEO-Studie (Destatis, 2019).

Bei den Sprachen wurde zunächst ermittelt, in welchem Land diese überwiegend gesprochen wird. Stammen aus den jeweiligen Ländern weniger als 50.000 Personen, wurde analog zum Vorgehen bei Staatsangehörigkeit und Geburtsland, Länder zusammengefasst. Sprachen, die in verschiedenen Ländern und Regionen gesprochen werden, wurden nicht zusammengefasst: hierzu zählen Englisch, Spanisch, Französisch, Russisch und Arabisch. Minderheitensprachen in Ländern wurden beispielsweise wie folgt kodiert: Sprachen in der Russischen Föderation. In diese Kategorie fällt nicht Russisch, das als Einzelkategorie aufgeführt ist. Zudem sind Sprachen in Sprachfamilien zusammengefasst, wie etwa Berbersprachen oder kurdische Sprachen, wenn diese über mehrere Länder verteilt gesprochen werden.

Die rekodierten Varablen bezogen auf das Geburtsland sind **gebland_rec**, **geblandv_rec** und **geblandm_rec**. Die rekodierten Variablen zum Bereich Staatsangehörigkeit sind **staat1_rec**, **staat2_rec**, **staatv_rec** und **staatm_rec**. Die neukodierten Werte und Wertbeschriftungen sowohl für die Geburtsländer als auch für die Staatsangehörigkeit sind in Tabelle 1 abgetragen.

Wert	Wertbeschriftung und ggf. Erläuterungen hierzu
-9	`gebland_rec`: Zusammenfassung von „Keine Angabe/weiß nicht“ und „Angabe nicht eindeutig“ zu „Keine Angabe/Angabe nicht eindeutig“, `geblandv_rec`: Zusammenfassung von „Keine Angabe/weiß nicht“, „Angabe nicht eindeutig“ und die zuvor herausgefilterten „Vater nicht bekannt“ zu „Keine Angabe/Angabe nicht eindeutig/Vater nicht bekannt“, `geblandm_rec`: Zusammenfassung von „Keine Angabe/weiß nicht“, „Angabe nicht eindeutig“ und die zuvor herausgefilterten „Mutter nicht bekannt“ zu „Keine Angabe/Angabe nicht eindeutig/Mutter nicht bekannt“,

	`staat1_rec`: Zusammenfassung von „Keine Angabe“, „Angabe nicht eindeutig“ und „staatenlos“ zu „Keine Angabe/Angabe nicht eindeutig/staatenlos“, `staat2_rec`: Zusammenfassung von „Keine Angabe“ und „Angabe nicht eindeutig“ zu „Keine Angabe/Angabe nicht eindeutig“, `staatv_rec`: Zusammenfassung von „Keine Angabe“, „Angabe nicht eindeutig“, „staatenlos“ und die zuvor herausgefilterten „Vater nicht bekannt“ zu „Keine Angabe/Angabe nicht eindeutig/Vater nicht bekannt/staatenlos“, `staatm_rec`: Zusammenfassung von „Keine Angabe“, „Angabe nicht eindeutig“, „staatenlos“ und die zuvor herausgefilterten „Mutter nicht bekannt“ zu „Keine Angabe/Angabe nicht eindeutig/Mutter nicht bekannt/staatenlos“.
40	Österreich; Schweiz
50	Bangladesch; Bhutan; Brunei Darussalam; Indonesien; Kambodscha; Demokratische Volksrepublik Laos; Malaysia; Malediven; Myanmar; Nepal; Pakistan; Philippinen; Singapur; Osttimor
76	Argentinien; Bolivien; Brasilien; Chile; Ecuador; Guyana; Kolumbien; Paraguay; Peru; Surinam; Uruguay; Venezuela (Bolivarische Republik); alle anderen südamerikanischen Länder
124	Kanada; Antigua und Barbuda; Bahamas; Barbados; Belize; Costa Rica; Dominica; Dominikanische Republik; El Salvador; Grenada; Guatemala; Haiti; Honduras; Jamaika; Kuba; Mexiko; Nicaragua; Panama; St. Kitts und Nevis; St. Lucia; St. Vincent und die Grenadinen; Trinidad und Tobago; alle anderen zentralamerikanischen Länder und die Karibik
208	Dänemark; Finnland; Schweden, Norwegen, Island
226	Äquatorialguinea; Gabun; Kamerun; Demokratische Republik Kongo; Kongo; São Tomé und Príncipe; Sudan; Tschad; Zentralafrikanische Republik; Äthiopien; Burundi; Dschibuti; Eritrea; Kamerun; Komoren; Kenia; Madagaskar; Mauritius; Ruanda; Seychellen; Somalia; Vereinigte Republik Tansania; Uganda; Angola; Botswana; Lesotho; Malawi; Mosambik; Namibia; Sambia; Simbabwe; Südafrika; Swasiland; Komoren; Kongo; Zentralafrikanische Republik
268	Armenien; Aserbaidschan; Georgien; Kirgisistan; Tadschikistan; Turkmenistan; Usbekistan; Kirgisistan; Tadschikistan; Turkmenistan; Usbekistan; Mongolei
288	Benin; Burkina Faso; Côte d’Ivoire; Gambia; Ghana; Guinea; Guinea-Bissau; Kap Verde; Liberia; Mali; Mauretanien; Niger; Nigeria; Senegal; Sierra Leone; Togo
356	Sri Lanka; Indien
410	Japan, Taiwan; Demokratische Volksrepublik Korea; Republik Korea
528	Belgien; Luxemburg; Niederlande
688	Montenegro; Serbien
705	Estland; Slowenien; Lettland; Belarus; Malta; Zypern

784	Jemen; Bahrain; Israel; Jordanien; Katar; Oman; Vereinigte Arabische Emirate; Saudi-Arabien; alle anderen Länder des Nahen Ostens
788	Algerien; Libyen; Tunesien; Ägypten
826	Irland; Vereinigtes Königreich

Tabelle 2 illustriert die Werte und Wertbeschriftungen der rekodierten Variablen aus dem Themenfeld Sprache. Hierbei handelt es sich um die Variablen **`f033ax_rec`**, **`mig001x_rec`**, **`mig002x_rec`**, **`mig004_1x_rec`**, **`mig004_2x_rec`**, **`mig004_3x_rec`**, **`mig004_4x_rec`**, **`mig004_5x_rec`** und **`mig004_6x_rec`**.

Wert	Wertbeschriftung und ggf. Erläuterungen hierzu
-2	Angabe nicht klassifizierbar, d.h. antike Sprachen (ausgenommen Latein), Minderheitensprachen in Deutschland oder in einem anderen Staat, wobei nicht feststeht, wie hoch die Anzahl der Sprecher*innen in Deutschland ist.
-1	Angabe nicht eindeutig, d.h. die Angabe konnte nicht als Sprache nach ISO 639-3 identifiziert wurden.
0	Latein
50	Sprachen in Bangladesch; Bhutan; Brunei Darussalam; Indonesien; Kambodscha; Demokratische Volksrepublik Laos; Malaysia; Malediven; Myanmar; Nepal; Pakistan; Philippinen; Singapur; Osttimor
76	Sprachen in Argentinien; Bolivien; Brasilien; Chile; Ecuador; Guyana; Kolumbien; Paraguay; Peru; Surinam; Uruguay; Venezuela (Bolivarische Republik); alle anderen südamerikanischen Länder
156	Sprachen in China
208	Sprachen in Dänemark; Finnland; Schweden, Norwegen, Island
226	Sprachen in Äquatorialguinea; Gabun; Kamerun; Demokratische Republik Kongo; Kongo; São Tomé und Príncipe; Sudan; Tschad; Zentralafrikanische Republik; Äthiopien; Burundi; Dschibuti; Eritrea; Kamerun; Komoren; Kenia; Madagaskar; Mauritius; Ruanda; Seychellen; Somalia; Vereinigte Republik Tansania; Uganda; Angola; Botswana; Lesotho; Malawi; Mosambik; Namibia; Sambia; Simbabwe; Südafrika; Swasiland; Komoren; Kongo; Zentralafrikanische Republik
232	Sprachen in Eritrea
268	Sprachen in Armenien; Aserbaidschan; Georgien; Kirgisistan; Tadschikistan; Turkmenistan; Usbekistan; Kirgisistan; Tadschikistan; Turkmenistan; Usbekistan; Mongolei
288	Sprachen in Benin; Burkina Faso; Côte d'Ivoire; Gambia; Ghana; Guinea; Guinea-Bissau; Kap Verde; Liberia; Mali; Mauretanien; Niger; Nigeria; Senegal; Sierra Leone; Togo
356	Sprachen in Sri Lanka; Indien
410	Sprachen in Japan, Taiwan; Demokratische Volksrepublik Korea; Republik Korea

504	Berbersprachen
528	Sprachen in Belgien; Luxemburg; Niederlande
586	Urdu, Panjabi
705	Sprachen in Estland; Slowenien; Lettland; Belarus; Malta; Zypern
784	Sprachen in Jemen; Bahrain; Israel; Jordanien; Katar; Oman; Vereinigte Arabische Emirate; Saudi-Arabien; alle anderen Länder des Nahen Ostens
810	Minderheitensprachen in der Russischen Föderation
840	nordamerikanische Sprachen

6.3 Angewandte Klassifikationsstandards

Der Rekodierung von Informationen zu Bildung und Beruf, Regionen und Wirtschaftszweige liegen folgende Standards zugrunde:

- **Klassifikationsystematik für Bildung** nach International Standard Classification of Education – Fields of education and training 2013 (UNESCO-UIS, 2014): ISCED-F-2013 (3-, 2-, 1-stellig), ISCED Fields 5-26 Basiskategorien, und ISCED Fields 5-26 Lernfelder.
- **Klassifikationssystematik für Gemeinden** nach BIK-Gemeindegrößenklassen (Behrens & Wiese, 2013): Politische Gemeindegrößenklasse, Gemeindegröße und Gemeindetyp.
- **Klassifkationssystematik für Beruf** nach International Standard Classification of Occupations ISCO-08 (ILO, 2012)
- **Klassifikationsystematik für Wirtschaftszweige**: Classification of Economic Activities, issue 2008 (Destatis, 2008).

7 Literatur

Behrens, K. & Wiese, K. (2013). Stadtregionen: Von Boustedt zu BIK (GESIS-Schriftenreihe). In Arbeitsgruppe Regionale Standards (Hrsg.), *Regionale Standards: Ausgabe 2013* (Band 12, S. 86–120). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.21241/ssoar.34820

Destatis. (2019). *Bevölkerung und Erwerbstätigkeit - Ausländische Bevölkerung: Ergebnisse des Ausländerzentralregisters 2018*. Wiesbaden: Statistisches Bundesamt (Destatis).

Destatis. (2008). *Klassifikation der Wirtschaftszweige*. Wiesbaden: Statistisches Bundesamt. Online: https://www.destatis.de/static/DE/dokumente/klassifikation-wz-2008-3100100089004.pdf

Grotlüschen, A. & Buddeberg, K. (Hrsg.). (2020). *LEO 2018: Leben mit geringer Literalität*. Bielefeld: wbv.

Grotlüschen, A. & Riekmann, W. (Hrsg.). (2012). *Funktionaler Analphabetismus in Deutschland: Ergebnisse der ersten leo. - Level-One Studie* (Alphabetisierung und Grundbildung). Münster New York München Berlin: Waxmann.

ILO. (2012). *International Standard Classification of Occupations: Structure, group definitions and correspondence tables* (Bd. 1). Genf: International Labour Organization. Online: https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/@publ/documents/publication/wcms_172572.pdf

Lüdtke, O. & Robitzsch, A. (2017). Eine Einführung in die Plausible-Values-Technik für die psychologische Forschung. *Diagnostica*, *63*(3), 193–205. Hogrefe Verlag. https://doi.org/10.1026/0012-1924/a000175

Perry, A., Helmschrott, S., Konradt, I. & Maehler, D. B. (2017). *User Guide for the German PIAAC Scientific Use File: Version II* (GESIS Papers) (Band 2017/23). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften.

UNESCO-UIS. (2014). *ISCED Fields of Education and Training 2013 (ISCED-F 2013)*. Montreal: UNESCO Institute for Statistics. Online: http://uis.unesco.org/sites/default/files/documents/isced-fields-of-education-and-training-2013-en.pdf.

von Davier, M., Gonzalez, E. & Mislevy, R. J. (2009). What Are Plausible Values and Why Are They Useful? (IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments). In IEA-ETS Research Institut (Hrsg.), (S. 9–36). Hamburg: IEA-ETS Research Institut.

8 Anhang

8.1 Anleitung zur Herstellung des SPSS-Datensatzes für imputierte Werte

```
*/ 1. Öffnen der Ausgangsdatei: ‚LEO2018-Scientific-Use-File.sav‘. Bei der Ve
rwendung der Ausgangsdatei ‚LEO2018-Public-Use-File.sav‘ muss in der nachfolg
enden Syntax der Dateiname angepasst werden.
*/ Hierzu bietet sich an über die Funktion "Suchen und Ersetzen" (STRG + F) d
en Begriff "Scientific" durch "Public" zu ersetzen.
*/ An Stelle von "SPEICHERORT" muss der Speicherort des Datensatzes angegeben
werden, z.B. C:\Users\LEO2018.
*/ Zu beachten ist, dass der Dateipfad in einfachen Anführungszeichen steht.
Auch hier ist das empfohlene Vorgehen mittels der Funktion "Suchen und Ersetz
en" (STRG + F)
*/ den Platzhalter "SPEICHERORT" mit dem Dateipfad zu den von Ihnen gewählten
Speicherort zu ersetzen.
GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.sav'.
```

```
DATASET NAME DataSet1 WINDOW=FRONT.
DATASET ACTIVATE DataSet1.

*/ 2. Anlegen einer neuen Variable „imputation_“.
COMPUTE Imputation_ = 0.
EXECUTE.
RECODE Imputation_ (0=SYSMIS).

*/ 3. Anlegen neuer Variablen für die imputierten Werte (IW_62 und IW_AlphaLe
vel_62).
COMPUTE IW_62 = 0.
COMPUTE IW_AlphaLevel_62 = 0.
EXECUTE.

*/ 4. Speichern des veränderten Ausgangsdatensatzes, um das Original unveränd
ert zu erhalten.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.1.sav'.

*/ 5. Neunmaliges Kopieren der Datendatei und Umbenennen der Kopie in ‘LEO201
8-Scientific-Use-File.2.sav‘ bis ‚LEO2018-Scientific-Use-File.10.sav‘.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.2.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.3.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.4.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.5.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.6.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.7.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.8.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.9.sav'.
SAVE OUTFILE = 'SPEICHERORT\LEO2018-Scientific-Use-File.10.sav'.

*/ 6. Schließen des nicht mehr benötigten DATASET1.
DATASET CLOSE DataSet1.

*/ 7. Nachfolgend werden die erzeugten Dateikopien jeweils geöffnet und dort
der Wert für die Variable ‚Imputation_‘ gesetzt.
*/ D.h. auf den Wert 1 in der ersten Kopie, auf 2 in der zweiten Kopie … auf
10 in der zehnten Kopie.
*/ Danach werden die Dateikopien unter dem Namen ‚LEO2018-Scientific-Use-File
.1neu.sav‘ etc. gespeichert.
GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.1.sav'.
RECODE Imputation_ (MISSING=1).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.1neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.2.sav'.
RECODE Imputation_ (MISSING=2).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.2neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.3.sav'.
RECODE Imputation_ (MISSING=3).
```

```
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.3neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.4.sav'.
RECODE Imputation_ (MISSING=4).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.4neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.5.sav'.
RECODE Imputation_ (MISSING=5).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.5neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.6.sav'.
RECODE Imputation_ (MISSING=6).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.6neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.7.sav'.
RECODE Imputation_ (MISSING=7).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.7neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.8.sav'.
RECODE Imputation_ (MISSING=8).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.8neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.9.sav'.
RECODE Imputation_ (MISSING=9).
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.9neu.sav'.

GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.10.sav'.
RECODE Imputation_ (MISSING=10).
EXECUTE.
SAVE OUTFILE='SPEICHERORT\LEO2018-Scientific-Use-File.10neu.sav'.

*/ 8. Verbinden des veränderten Ausgangsdatensatzes mit den zehn bearbeiteten
Dateikopien (Daten --> Dateien zusammenfügen).
GET FILE='SPEICHERORT\LEO2018-Scientific-Use-File.1neu.sav'.

ADD FILES /FILE=*
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.2neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.3neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.4neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.5neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.6neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.7neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.8neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.9neu.sav'
  /FILE='SPEICHERORT\LEO2018-Scientific-Use-File.10neu.sav'.
EXECUTE.

*/ 9. Füllen der neu angelegten Variablen „IW_62“ und „IW_AlphaLevel_62“ mit
den Werten der erzeugten Variablen „pv1“ bis „pv10“ bzw. „alpha_pv1“ bis „alp
```

```
ha_pv10“.
*/ Benennen des neu erstellten DATASET.
DATASET NAME DataSet2 WINDOW=FRONT.
DATASET ACTIVATE DataSet2.
IF (Imputation_ = 1) IW_62=pv1.
IF (Imputation_ = 2) IW_62=pv2.
IF (Imputation_ = 3) IW_62=pv3.
IF (Imputation_ = 4) IW_62=pv4.
IF (Imputation_ = 5) IW_62=pv5.
IF (Imputation_ = 6) IW_62=pv6.
IF (Imputation_ = 7) IW_62=pv7.
IF (Imputation_ = 8) IW_62=pv8.
IF (Imputation_ = 9) IW_62=pv9.
IF (Imputation_ = 10) IW_62=pv10.
IF (Imputation_ = 1) IW_AlphaLevel_62=alpha_pv1.
IF (Imputation_ = 2) IW_AlphaLevel_62=alpha_pv2.
IF (Imputation_ = 3) IW_AlphaLevel_62=alpha_pv3.
IF (Imputation_ = 4) IW_AlphaLevel_62=alpha_pv4.
IF (Imputation_ = 5) IW_AlphaLevel_62=alpha_pv5.
IF (Imputation_ = 6) IW_AlphaLevel_62=alpha_pv6.
IF (Imputation_ = 7) IW_AlphaLevel_62=alpha_pv7.
IF (Imputation_ = 8) IW_AlphaLevel_62=alpha_pv8.
IF (Imputation_ = 9) IW_AlphaLevel_62=alpha_pv9.
IF (Imputation_ = 10) IW_AlphaLevel_62=alpha_pv10.
EXECUTE.

*/ 10. Speichern des gesamten Datensatzes.
SAVE OUTFILE = 'SPEICHERORT\LEO2018_Enddaten_mit_PVs.sav'.

*/ 11. Datei anhand der Variable Imputation_ sortieren und Gruppen vergleiche
n aktivieren.
SORT CASES BY Imputation_.
SPLIT FILE LAYERED BY Imputation_.

*/ 12. Gewichtung nach Personengewichtungsfaktor Gesamtstichprobe.
WEIGHT BY pgewges.

*/ 13. Optional ist eine Gewichtung nach Hochrechnungsfaktor Gesamtstichprobe
(Bevölkerungsverteilung in Tsd.) möglich.
*/WEIGHT BY phochges.

*/ Jetzt können die Analysen starten.
```