

TFS-Info-13 Informationen zu Korrekturgewichten

Version: 1 vom 9.4.2024

Autor*innen: Anke Macdonald

Einleitung

Seit März 2024 sind die Korrekturgewichte im NAKO TransferHub beantragbar.

Ein besonderer Dank gebührt Dr. Matthias Sand (Leibniz-Institut für Sozialwissenschaften GESIS, Abteilung 'Survey Design and Methodology') für seine fachliche Anleitung und tatkräftige Unterstützung bei der Ermittlung der Korrekturgewichte.

Nachfolgend finden Sie einige Hinweise zur Erstellung und Verwendung der Korrekturgewichte. Kontaktperson für Fragen zu diesem Thema ist Stefan Rach (rach@leibniz-bips.de).

1. Gewichtung

Die Korrekturgewichte werden als Gesamtgewichte (total weights) bereitgestellt, die sowohl den Einfluss unterschiedlicher Einschlusswahrscheinlichkeiten aufgrund des Stichprobendesigns (design weights), als auch den Einfluss von Non-response und die Verteilung bestimmter Merkmale in der Bevölkerung (calibration weights) berücksichtigen.

Die Korrekturgewichte dürfen **nur mit den Daten der Basiserhebung verwendet werden**; für spätere Untersuchungswellen und/oder die GEFU-Befragungen müssen ggf. jeweils gesonderte Korrekturgewichte bestimmt werden.

Im NAKO-Datenverzeichnis finden Sie die Gesamtgewichte unter dem Punkt Basisdaten->Korrekturgewichte->Gesamtgewichte. (Beachten: Aktuell noch unter Basisdaten-> Demographie -> Für den NAKO-Datensatz... -> Korrekturgewichte -> Gesamtgewichte. Wird demnächst korrigiert.)

Im Regelfall beantragen Sie die Variablenpaare

- "Gesamtgewicht normiert" (wgt_total_18sc) und "Geografische Zentrumszuordnung normiert" (wgt_geoSC_18sc)

und/oder

- "Gesamtgewicht normiert" (wgt_total_16sc) und "Geografische Zentrumszuordnung normiert" (wgt_geoSC_16sc)

(Zum Unterschied zwischen **_18sc** und **_16sc** siehe Abschnitt Normierung)

Sofern für bestimmte Analysen nur Designgewichte benötigt werden, können diese auch bereitgestellt werden, wenn dies bei der Beantragung der Daten entsprechend begründet wird. Die Designgewichte finden Sie im NAKO-Datenverzeichnis unter dem Punkt Basisdaten->Korrekturgewichte->Qualitätssicherung.

2. Repräsentativität

Die Gewichtung erfolgte nicht anhand der Bevölkerungsverteilung Gesamtdeutschlands, sondern nur anhand der **Bevölkerungsverteilung in den tatsächlich durch die Studienzentren der NAKO abgedeckten Studienregionen**. Damit sind Aussagen zur Repräsentativität von gewichteten Analyseergebnissen (sofern man diese überhaupt treffen kann/möchte) genau auf eben diese Studienregionen eingeschränkt.

3. Normierung / Renormierung

Die Gesamtgewichte wurden auf die Größe der Unterstichproben der Studienzentren normiert, d.h. für jedes Studienzentrum gleicht die Summe der Gesamtgewichte genau der Anzahl der aus diesem Studienzentrum enthaltenen Fälle im Datensatz. Hierbei ist zu beachten, dass hierfür die **geografische** Studienzentrumzugehörigkeit laut Adresse **zum Einladungszeitpunkt** (bzw. Zeitpunkt der Stichprobenziehung) betrachtet wird. Für eine geringe Anzahl von Fällen (Zentrumswechsler/Umzügler) weicht diese vom Studienzentrum ab, in dem die Teilnehmenden letztlich untersucht wurden. Die geografische Studienzentrumzugehörigkeit wird in den Variablen wgt_geoSC_18sc bzw. wgt_geoSC_16sc bereitgestellt.

Die Korrekturgewichte werden in zwei Normierungen bereitgestellt:

- Normierung auf 18 Studienzentren: Die Gesamtgewichte (total weights) sind in der Variable wgt_total_18sc enthalten; die Zuordnung zum Studienzentrum ist in der Variable wgt_geoSC_18sc enthalten.

- Normierung unter Zusammenlegung der drei Berliner Studienzentren (51, 52, 53) zu einer Studienregion (50): Die Gesamtgewichte (total weights) sind in der Variable wgt_total_16sc enthalten; die Zuordnung zum Studienzentrum ist in der Variable wgt_geoSC_16sc enthalten.

(Die Unterschiede zwischen den Variablen wgt_total_18sc und wgt_total_16sc beschränken sich auf die 3 Berliner Studienzentren; die restlichen Werte sind identisch.)

4. Renormierung der Gewichte vor der Verwendung

Der Umfang des NAKO Samples verringert sich durch eingehende Widersprüche laufend. Zum Zeitpunkt der Erstellung der Korrekturgewichte befanden sich 204.754 Datensätze im Sample. Datennutzende, deren Daten zu einem früheren Zeitpunkt geliefert wurden, könnten daher Fälle in ihren Datensätzen haben, für die keine Korrekturgewichte vorhanden sind. Für Analysen unter Verwendung der Korrekturgewichte wären diese dann auszuschließen.

Datennutzende, deren Daten zu einem späteren Zeitpunkt geliefert werden, werden mglw. weniger als diese 204.754 Datensätze erhalten. Des Weiteren werden auch im Rahmen der

Analysen regelmäßig Fälle ausgeschlossen, so dass sich auch hier die Anzahl der Fälle verringert.

Immer wenn aus o.g. Gründen im Analysedatensatz nicht alle 204.754 Fällen mit Korrekturgewicht enthalten sind, wird Datennutzenden empfohlen, vor der Verwendung der Gewichte eine Renormierung durchzuführen, die dafür sorgt, dass wieder für jedes Studienzentrum die Summe der Gesamtgewichte genau der Anzahl der aus diesem Studienzentrum enthaltenen Fälle im Datensatz gleicht. Für diese Renormierung werden für jedes Studienzentrum gesondert die individuellen Gewichte durch die Summe der Gewichte geteilt und mit der Anzahl der Fälle multipliziert.

5. Nachträgliche Beantragung von Korrekturgewichten

Es besteht für alle Nutzungsprojekte die Möglichkeit, die Korrekturgewichte nachträglich zu beantragen, auch dann, wenn die vollständige Datenübergabe bereits erfolgt ist.

Wenn nur die **Gesamtgewichte** benötigt werden, genügt ein formloser Antrag per E-Mail an transferstelle@nako.de mit folgendem Wortlaut:

Hiermit beantrage ich im Rahmen des Nutzungsprojekts NAKO-xxx nach bereits erfolgter Datenübergabe die Gesamtgewichte bestehend aus den Variablenpaaren

- "Gesamtgewicht normiert" (wgt_total_18sc) und "Geografische Zentrumszuordnung normiert" (wgt_geoSC_18sc)

und/oder

- "Gesamtgewicht normiert" (wgt_total_16sc) und "Geografische Zentrumszuordnung normiert" (wgt_geoSC_16sc)

Die Variablen werden dann vom Datenmanagement mit gleicher Pseudonymisierung wie der ursprüngliche Variablensatz zum Nutzungsantrag hochgeladen.

Wenn die **Designgewichte** benötigt werden, ist ein ebenfalls formloser Antrag per E-Mail mit folgendem Wortlaut an transferstelle@nako.de zu richten:

Hiermit beantrage ich im Rahmen des Nutzungsprojekts NAKO-xxx nach bereits erfolgter Datenübergabe die Designgewichte.

Begründung: xxx

Das UAC prüft die Begründung und bei entsprechender Empfehlung zur Freigabe werden die Variablen vom Datenmanagement mit gleicher Pseudonymisierung wie der ursprüngliche Variablensatz zum Nutzungsantrag hochgeladen.