

ZA8775

**Monitoring Ländliche Räume - Befragung zur
Vereinbarkeit von Erwerbsarbeit und Familie
(Vertiefungsbefragung)**

**Hinweise zu den Originaldatensätzen
und zu Datenfehlern**

Hinweise zu den Originaldatensätzen Vertiefungsbefragung der Monitor-Bevölkerungsbefragungen

In Bezug auf einige wenige Variablen unterscheiden sich die von infas generierten Originaldatensätze von dem bei GESIS archivierten Datensatz. Diese Unterschiede werden nachfolgend erläutert.

A Datei „Monitor Vertiefungsbefragung“

(1) Nicht enthaltene Variablen

Zwei der mit dem „**Monitor Vertiefungsbefragung**“ gelieferten Variablen sind nicht im Datensatz enthalten, der bei GESIS archiviert ist.

- aus Datenschutzgründen Variable mit
 - Ortsangaben aus offener Frage im Fragebogenabschnitt „Voll- und Teilzeitarbeit: Arbeitsort und -weg, Flexibilität, Familienfreundlichkeit für Auskunftsperson“ zur Frage „Liegt Ihre Arbeitsstätte in Ihrer Gemeinde, wo Sie wohnen?“
sd12b_plz Postleitzahl
sd12b_02_o Gemeinde

(2) Variablen, die sich nicht aus dem Fragebogen ergeben

Der Datensatz enthält einige Variablen, die nicht im Fragebogen verwendet werden. Darunter sind

- fünf Variable, die im Originaldatensatz enthalten sind:
 - **thuenen_id**: Thünen ID [Reihenfolgennummer innerhalb des Datensatzes]
 - **kkz**: Kreiskennziffer
 - **raumtyp**: Thünen-Typologie ländlicher Räume [Angabe zur Lage der Gemeinde einer Auskunftsperson]
 - **intdat_m**: Interviewdatum: Monat
 - **intdat_j**: Interviewdatum: Jahr]
- drei Variable, die aus dem Gewichtungsdatsatz von infas zugespielt wurden:
 - **desgew**: Designgewicht
 - **perswich**: Personengewicht
 - **hh_wich**: Haushaltsgewicht

B Datei „Monitor Vertiefungsbefragung DJI Zuspil“

(1) Nicht enthaltene Variablen

Zwei der mit dem „**Monitor Vertiefungsbefragung DJI-Zuspil**“ gelieferten Variablen sind nicht im Datensatz enthalten, der bei GESIS archiviert ist.

- aus Datenschutzgründen Variable mit
 - offenen Angaben zur Beruflichen Stellung“
kfg23o2 Berufliche Stellung P (offenen Abfrage)
kfg23o Frühere berufl. Stellung P (offene Angabe)

C Zusammenfassung der beiden Datendateien

Die beiden Datensätze sind für die Analyse zusammenzufassen. Dies ist mit einigen Problemen, besonders in Hinblick auf Datenfehler bei der Variable „Ankerperson“ verbunden.

In der Textdatei „Vertiefungsbefragung Ankerperson Zuspil“ werden viele der entsprechenden Datenfehler beschrieben.

Joachim Kreis

Analyse von Datenfehlern zu „Ankerperson“ in der Analysedatei der Projektgruppe, in der die Befragungsdaten und die Zuspieldaten von infas zusammengeführt worden sind.

Auszug aus einer Textdatei der Projektgruppe

1. Nettostichprobe

Die Nettostichprobe fußt auf einer Befragung von infas für das Deutsche Jugendinstitut in Familien mit Kindern unter 15 Jahren im Haushalt. Sie wurde vom Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) beauftragt und steht in Zusammenhang mit dem Kinderförderungsgesetz (KiföG). Sie wird im Folgenden mit der Abkürzung „kfg“ bezeichnet. Aus dieser Befragung wurde für die Vertiefungsbefragung eine Stichprobe gezogen unter Familien mit mindestens einem Kind unter 13 Jahren im Haushalt. Der von infas gelieferte Datensatz umfasst 959 Fälle.

Aufgrund der Komplexität der exakten Altersberechnung für Kinder, die in der DJI-Befragung noch unter 13 Jahren alt waren und zum Befragungszeitpunkt der Vertiefungsbefragung gerade oder vor kurzem 13 Jahre alt geworden waren, wurde die Berechnung vereinfacht und auch Haushalte zugelassen, deren jüngste Kinder gerade oder vor kurzem 13 Jahre alt geworden sind.

Insgesamt sind im Datensatz 37 Fälle, in denen das jüngste Kind im Haushalt schon 13 Jahre alt ist.

Kinder im Haushalt

Unter den 959 Fällen sind zwei Fälle ohne Kinder im Haushalt. Sie sind für die Stichprobenziehung „not eligible“, d.h. sie dürften kein Bestandteil der Nettostichprobe sein, da sie nicht zur Grundgesamtheit der Vertiefungsbefragung gehören. Eventuell hat sich zwischen Stichprobenziehung und Befragungsbeginn in diesen Fällen die Familienkonstellation geändert. Sie sind aus dem Analysedatensatz zu entfernen. Es betrifft die Fälle mit den Identifikationsnummern (id) 125 und 565.

Die bereinigte Nettostichprobe umfasst 957 Fälle.

2. Analysedatensatz

Aus der bereinigten Nettostichprobe werden Fälle ausgeschlossen, die viele „weiß nicht“ und/oder „keine Angabe“ aufweisen und wenige gültige Werte. Es gibt nur einen Fall auf den dies zutrifft. Er hat die id 143 und weist nur 15 gültige Werte auf. Er wird aus dem Analysedatensatz gelöscht.

Die folgenden Angaben beziehen sich auf den Analysedatensatz mit 956 Fällen.

3. Datenfehler und fehlende Werte

Die Prüfung und Bereinigung von Datenfehlern lässt sich in der Vertiefungsbefragung gut anhand der Antworten zur Frage nach der Ankerperson darstellen. Hier wird gefragt, ob eine Auskunftsperson (AP) der Vertiefungsbefragung schon in der kfg-Befragung Auskunftsperson war. Falls ja, wird diese Person als Ankerperson eingeordnet. Die Antworten werden im Zusammenhang mit den Daten zu Geburtsmonat, -jahr und Geschlecht der Auskunftsperson (AP) und falls vorhanden mit den entsprechenden Daten von deren Partner*in analysiert.

Im Folgenden werden in diesem Zusammenhang 107 Datenfehler gezeigt.

Davon sind wahrscheinlich 32 Datenfehler wegen Vertippens und/oder Verhörens von Interviewer*innen. Unter den 32 sind 16 Fälle beim Geburtsmonat der AP. Ob sie in der kfg- oder Vertiefungsbefragung entstanden, ist unklar. Eine entsprechend berechnete Fehlerquote muss sich demnach auf die Anzahl der Angaben zum Geburtsmonat in beiden Befragungen beziehen, d.h. auf 956×2 Fälle, insgesamt 1912. Die 16 Fälle entsprechen damit einer Fehlerquote von 0,8 Prozent. Unter den 32 sind nur 5 Fälle, in denen bei der Geschlechtsangabe zur AP ein Eingabefehler vorliegt, was nur knapp 0,3 Prozent der 1912 Eingaben entspricht.

14 der 107 Datenfehler sind dem Datenmanagement zuzuordnen;

bei 61 der 107 ist unklar, ob sie durch Interviewer*innen oder durch Befragte verursacht sind. Unter den 61 unklaren Fällen sind 54 Fälle (5,6 Prozent der 956 Fälle), bei denen der Datenfehler entweder auftritt, weil Befragte eine falsche Angabe zu Ankerperson machen, oder weil sich Interviewer*innen bei der Antwort zur Ankerperson vertippt haben.

Die meisten der 107 Datenfehler sind im Analysedatensatz bereinigt. Nur wenige müssen als unklare Angaben ausgewiesen werden.

In der Vertiefungsbefragung hängen einige Filtersetzungen davon ab, ob die Auskunftsperson (AP) schon in der kfg-Befragung Auskunftsperson war. Bei Übereinstimmung wird sie als Ankerperson bezeichnet. In 76 Fällen, d.h. 8 Prozent der 956, wird für die AP angegeben, dass sie schon in der kfg-Befragung AP gewesen sei; aber ihre Geschlechtsangabe und/oder Geburtsjahrangabe stimmen gemäß der von infas generierten Variablen „ankerpersion“ nicht mit ihren Angaben aus der kfg-Befragung überein. Hier stellt sich die grundsätzliche Frage, inwieweit Angaben in der kfg- und Vertiefungsbefragung, die sich auf Fakten beziehen, in der gleichen Art und Weise von Befragten verstanden und beantwortet bzw. von den Interviewer*innen eingegeben wurden. Der Vergleich von Daten aus der kfg- und Vertiefungsbefragung gibt hierzu ein anschauliches Beispiel:

In 593 Fällen sind in beiden Befragungen gemäß der von infas generierten Variable „ankerpersion = 1“ die gleichen Personen Auskunftsperson, ihre Angaben zu Geschlecht und Geburtsjahr stimmen überein. Bei 17 unter ihnen besteht jedoch keine Übereinstimmung in den Daten zum Geburtsmonat.

Verhören (z.B. Juni statt Juli und deshalb die Dateneingabe „6“ statt „7“ für den Geburtsmonat) oder vertippen von Interviewer*innen (z.B. 8 statt 9) liegt sehr wahrscheinlich in 14 Fällen vor (Fehler Interviewer*in). Wie sich diese Fehlerquellen auf die kfg- und Vertiefungsbefragung verteilen, kann auf Grundlage der vorliegenden Daten nicht beurteilt werden. >

In 3 Fällen wurden eventuell von Befragten unterschiedliche Angaben zum Geburtsmonat gemacht, kfg/Vertief: 10/8, 6/10, 10/4. (Fehlerquelle unklar)

In 287 Fällen sind gemäß Variable „ankerpersion = 2“ die AP der Vertiefungsbefragung nicht Ankerperson. Bei sieben unter ihnen ist die Angabe aus der von infas generierten Variable „ankerpersion“ aber falsch (Fehler Datenmanagement): bei ihnen stimmen die Angaben zu Geschlecht und Geburtsjahr in beiden Befragungen überein. Sie werden deshalb im Analysedatensatz in der dort generierten Variablen „ankperik“ als Ankerpersonen kodiert.

In 76 Fällen haben die AP gemäß infas-Variable „ankerpersion = 3 bis 5“ in der Vertiefungsbefragung unzutreffend angegeben, auch in der kfg-Befragung AP gewesen zu sein.

Unter ihnen sind 12 Fälle, in denen die Angaben zum Geschlecht der AP in der Vertiefungsbefragung mit denen aus der kfg-Befragung übereinstimmen, aber nicht die zum Geburtsjahr („ankerpersion = 4“).

In sechs dieser 12 Fälle stimmen jedoch die Angaben zu den Geburtsmonaten überein und die Interviewer*innen haben sich wahrscheinlich beim Geburtsjahr vertippt: die entsprechenden Angaben sind kfg/Vertief: 1977/1975, 1985/1980, 1976/1966, 1971/1977, 1973/1979, 1972/1970. Die Angabe der AP schon in der kfg-Befragung AP gewesen zu sein ist demnach wahrscheinlich richtig (Fehler Interviewer*in).

In zwei weiteren Fällen stimmen Geburtsmonat und –jahr nicht überein; in einem Fall liegt beim Geburtsjahr wahrscheinlich ein Zahlendreher, d.h. ein weiteres Vertippen vor, kfg/Vertief: 1976/1967 (Fehler Interviewer*in), in dem anderen Fall bleibt unklar, wie die unterschiedlichen Angaben zustande gekommen sind. Hier bleibt unklar, ob die AP der Vertiefungsbefragung Ankerperson ist (Fehlerquelle unklar).

In zwei weiteren der 12 Fälle machen die AP in der Vertiefungsbefragung keine Angaben zu ihren Geburtsdaten. Sie werden gemäß ihrer eigenen Angabe im Analysedatensatz als Ankerperson kodiert (Fehler Datenmanagement).

In den zwei verbleibenden der 12 Fälle stimmen die Geburtsdaten von Partner*in der AP in der Vertiefungsbefragung mit denen der AP der kfg-Befragung überein, jedoch nicht mit deren Geschlecht. In den zwei Fällen sind wahrscheinlich die Geschlechtsangaben der Vertiefungs- und nicht der kfg-Befragung richtig, da die Geschlechtsangabe der AP der Vertiefungsbefragung mit der Angabe zur Stellung zum Kind übereinstimmt (Mutter bzw. Vater). In kfg-Befragung haben sich die Interviewer*innen bei der Eingabe zum Geschlecht der AP wahrscheinlich vertippt (Fehler Interviewer*in). Für die Angaben aus der Vertiefungsbefragung bleibt unklar, ob die AP fälschlicherweise angaben, Ankerperson gewesen zu sein, oder ob ein Dateneingabefehler seitens der Interviewer*innen vorliegt (Fehlerquelle unklar).

Letztere Frage stellt sich auch für die 53 Fälle, in denen die Angaben zu Geschlecht und Geburtsjahr der Auskunftsperson nicht mit denen aus der kfg-Befragung übereinstimmen („ankerson = 5“).

In 47 dieser 53 Fälle stimmen die Angaben zu den Partner*innen der Vertiefungsbefragung mit denen zur Auskunftsperson der kfg-Befragung überein, auch im Geburtsmonat, d.h. ihre Partner*innen waren damals Auskunftsperson (Fehlerquelle unklar).

In vier der verbleibenden sechs Fälle stimmen die Angaben der Auskunftsperson zu Geschlecht und Geburtsmonat mit den Angaben zu ihren Partner*innen überein, beim Geburtsjahr liegen dagegen sehr wahrscheinlich Fehler in der Dateneingabe vor: die entsprechenden Eingaben sind kfg/Vertief: 1971/1972, 1974/1973, 1975/1974, 1976/1975. Ob die Fehler bei der kfg- oder Vertiefungsbefragung gemacht wurden ist unklar (Fehler Interviewer*in).

In zwei der verbleibenden sechs Fälle stimmt nur der Geburtsmonat nicht mit den Angaben zu ihren Partner*innen überein (Fehler Interviewer*in). Auch hier ist unklar, ob der Fehler bei der Dateneingabe zu Ankerperson in der kfg- oder Vertiefungsbefragung liegt (Fehler Interviewer*in).

In 11 Fällen gibt gemäß Variable „ankerson = 3“ die Auskunftsperson der Vertiefungsbefragung an, auch in der kfg-Befragung Auskunftsperson gewesen zu sein. Die Geschlechtsangaben beider Befragungen stimmen jedoch gemäß der Variable „ankerson“ nicht überein, aber das Geburtsjahr.

In fünf der 11 Fälle stimmt das Geburtsjahr auch nicht überein, jedoch mit Geburtsjahr und monat von Partner*in der AP. Zwei verschiedene Datenfehler treten bei diesen fünf Fällen auf; zum einen hinsichtlich ihrer Kodierung in der Variable „ankerson“: sie werden mit „3“ statt mit „5“ kodiert (Fehler Datenmanagement); zum anderen hinsichtlich ihrer Kodierung zu ihrer eigenen Angabe als Ankerperson (Fehlerquelle unklar).

In weiteren drei der 11 Fälle haben AP und Partner*in verschiedene Geburtsdaten. Geburtsjahr und –monat von AP in kfg- und Vertiefungsbefragung stimmen überein. Dies lässt es wahrscheinlich erscheinen,

dass die Geschlechtsangabe aus der kfg-Befragung falsch ist, d.h. es liegt wahrscheinlich ein Fehler in der Dateneingabe vor (Fehler Interviewer*in). Die Angabe der AP schon in der kfg-Befragung AP gewesen zu sein ist demnach wahrscheinlich richtig.

In den letzten drei der 11 Fälle haben AP und Partner*in das gleiche Daten für Geburtsjahr und –monat. Hier bleibt unklar, ob es Eingabefehler hinsichtlich der Geschlechtsangabe in der kfg-Befragung oder bei der Ankerperson in der Vertiefungsbefragung gibt oder eine falsche Angabe der Befragten zur Ankerperson (Fehlerquelle unklar).

Staatsbürgerschaft der Auskunftsperson „sd32z“ und – falls vorhanden – von deren Partner*in mit den Variablen da05z, da18z und da21z wurden nur gestellt, falls im Vergleich zur kfg-Befragung Änderungen eingetreten waren. Falls die Auskunftspersonen angaben, es habe sich nichts geändert, wurden die entsprechenden Daten aus der kfg-Befragung dem Analysedatensatz zugespielt oder kfg-Daten konstruiert und dann zugespielt. Die entsprechenden Daten wurden vom DJI zur Verfügung gestellt.

Ob die Auskunftsperson und ihr*e Partner*in verheiratet sind „pa02z“, wurde in der Vertiefungsbefragung nur gefragt, wenn im Vergleich zur kfg-Befragung andere bzw. neue Partner vorhanden waren. Für den Vergleich werden aus der kfg-Befragung auch die Variablen kfg29 zur festen Partnerschaft und kfg31 zum Zusammenleben mit Partner*in verwendet.

Anmerkungen zu diesen Variablen

höchster Schulabschluss: in der kfg-Befragung entspricht dies den Variablen kfg28a, kfg28ap; im Analysedatensatz sind dies die Variablen sd05z, da05z. In der kfg-Befragung sind die beiden höchsten Abschlüsse „Fachhochschulreife, Abschluss einer Fachoberschule“ und „Allgemeine oder fachgebundene Hochschulreife/Abitur“ nicht trennscharf. In der Vertiefungsbefragung wurden sie trennscharf formuliert. Aufgrund des Zuspiels müssen jedoch die Antwortvorgaben aus der kfg-Befragung übernommen werden. Für Analysen bietet sich hier die Zusammenfassung der beiden höchsten Schulabschlüsse in „höherer Schulabschluss“ an.

Ausbildungsabschlüsse: in der kfg-Befragung entspricht dies den Variablen kfg28b, kfg28b_2 kfg28bp, kfg28bp_2. Die „_2“ steht für eine Variable, bei der eine zweite Nennung eines Ausbildungsabschlusses kodiert werden konnte. Es gab nur sehr wenige zweite Nennungen in der kfg-Befragung. In der Vertiefungsbefragung sind dies die Variablen sd06_01 bis sd06_13, die im Analysedatensatz zur Variable sd06z zusammengefasst werden und die Variablen da18_01 bis da18_13, zusammengefasst zur Variable da18z. In den Fällen mit zwei Nennungen von Abschlüssen wurde jeweils der höchste Abschluss in der zusammengefassten Variable kodiert. Insoweit geben die Variablen sd06z und sd18z über den höchsten Ausbildungsabschluss.

Es gab in den Fällen, bei denen eine Änderung des Ausbildungsabschlusses seit der kfg-Befragung erfolgte, wie in kfg-Befragung nur sehr wenige Mehrfachantworten. Die Antwortmöglichkeiten in der kfg- und Vertiefungsbefragung sind nicht völlig identisch. In der Vertiefungsbefragung gibt es zusätzlich die Abschlüsse „Beruflich-betriebliche Anlernzeit mit Abschlusszeugnis, aber keine Lehre“ und „Teilfacharbeiterabschluss (DDR)“ und die Differenzierung des Universitätsabschlusses in „Bachelor an Universität“ und „Universitätsabschluss (z. B. Master, Diplom, Magister, Staatsexamen)“. Aufgrund des Zuspiels werden die Antwortvorgaben aus der kfg-Befragung übernommen und die Differenzierungen aus der Vertiefungsbefragung den entsprechenden Antwortmöglichkeiten der kfg-Befragung zugewiesen.

Staatsbürgerschaft: in der kfg-Befragung wird die Staatsbürgerschaft mit Bezug auf die leiblichen Eltern des Zielkinds anhand der Variablen kfg34, kfg34c_1, kfg34c_2 erhoben. Ob Partner*in der Auskunftsperson leiblicher Elternteil ist, wird dort anhand der Variable kfg27a3 festgestellt. Aus diesen vier Variablen werden für die Vertiefungsbefragung die Zuspieldaten für die Variablen sd32z und da21z im Analysedatensatz konstruiert.

Partner verheiratet: in der kfg-Befragung wird die Frage nach der Heirat anhand der Variable kfg30 erhoben. Aus ihr werden die Zuspieldaten für die Variable pa02z konstruiert.

Die kfg-Variablen sind nicht Bestandteil des Analysedatensatzes. Die Merkmale höchster Schulabschluss höchster Ausbildungsabschluss Staatsbürgerschaft und mit Partner*in verheiratet werden durch die sieben Variable sd05z, da05z, sd06z, sd18z, sd32z, da21z und pa02z im Analysedatensatz erfasst. Aufgrund ihrer Sonderstellung im Hinblick auf fehlende Werte, die auch aus der kfg-Befragung resultieren können, sollten die sieben Variablen bei der Analyse fehlender Werte je Fall der Vertiefungsbefragung gesondert behandelt werden.

Joachim Kreis