Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Statistik BFS

Haushaltstypologie in der STATPOP (HHTyp)

Produktionsprozess - Kurzbeschreibung

Neuchâtel, 2022

Herausgeber:	Bundesamt für Statistik (BFS)	**Layoutkonzept:**	Sektion PUB
Auskunft:	info.dem@bfs.admin.ch	**Download:**	www.experimental.bfs.admin.ch
Redaktion:	Sabrina Bravo, BFS; Estelle Crippa, BFS; Fabienne Helfer, BFS	**Copyright:**	BFS, Neuchâtel 2022 Wiedergabe unter Angabe der Quelle für nichtkommerzielle Nutzung gestattet
Inhalt:	Sabrina Bravo, BFS; Estelle Crippa, BFS		
Themenbereich:	00 Statistische Grundlagen		
Originaltext:	Französisch		
Übersetzung:	Sprachdienste BFS		

ESPACE DE L'EUROPE 10 CH-2010 NEUCHÂTEL WWW.STATISTIK.CH

Einführung

Für die Statistik der Bevölkerung und Haushalte (STATPOP) wurde eine Variable «Haushaltstyp» erstellt. Damit lässt sich jedem Privathaushalt in der ständigen Wohnbevölkerung am Hauptwohnsitz ein Haushaltstyp zuordnen. Der Haushaltstyp ist nicht nur im Bereich der Haushalte von Bedeutung, sondern auch in Zusammenhang mit anderen Thematiken. Bis 2000 wurde die Variable «Haushaltstyp» aus den Informationen der alle zehn Jahre durchgeführten Volkszählung abgeleitet. Seit der Einführung des neuen Volkszählungssystems 2010 wird diese Variable in der Strukturerhebung erfasst, die jährlich mit einer Stichprobe von rund 300 000 Personen durchgeführt wird. In der Strukturerhebung wird der «Haushaltstyp» anhand der Beziehungen zwischen den einzelnen Haushaltsmitgliedern ermittelt und ermöglicht für Gruppen von mindestens 15 000 Personen eine direkte und zuverlässige Schätzung. In den Einwohnerregistern, die der STATPOP zugrunde liegen, sind die Informationen zu diesen Beziehungen nicht verfügbar. Dank des Projekts «Haushaltstypologie» kann diese Variable nun in der STATPOP für nahezu alle Haushalte bereitgestellt werden. So lassen sich detailliertere Analysen durchführen als mit der Strukturerhebung; denkbar sind beispielsweise die Untersuchung der zeitlichen Entwicklung der Haushaltstypen auf Gemeindeebene oder ein Vergleich der Erwerbseinkommen nach Haushaltstyp mit einem detaillierten Regionalisierungsgrad. Der Haushaltstyp kann auch auf individueller Ebene als erklärende Variable in statistischen Modellen verwendet werden und so auch in Zusammenhang mit Problematiken in anderen Themenfeldern zu einem besseren Verständnis beitragen (z.B. Analysen zur Armut).

Die STATPOP-Variable «Haushaltstyp» wird in mehreren Schritten erstellt. Zunächst werden anhand verschiedener Datenquellen die Beziehungen zwischen Personen innerhalb eines Haushalts ermittelt. Auf Basis dieser Informationen ordnet ein Algorithmus jedem Haushalt, für den alle Beziehungen bekannt sind, einen Haushaltstyp zu. Ein weiterer, auf Machine Learning basierender Algorithmus imputiert schliesslich den Haushaltstyp von Haushalten, für die der Haushaltstyp fehlt (rund 13%).

Die Variable Haushaltstyp umfasst folgende Modalitäten[1]:

a. Einpersonenhaushalt
b. Ehepaar ohne Kinder
c. Konsensualpaar[2] ohne Kinder
d. Gleichgeschlechtliches Paar[3] ohne Kinder
e. Einelternhaushalt mit mindestens einem Kind[4] unter 25 Jahren
f. Ehepaar mit mindestens einem Kind unter 25 Jahren
g. Konsensualpaar mit mindestens einem Kind unter 25 Jahren
h. Gleichgeschlechtliches Paar mit mindestens einem Kind unter 25 Jahren
i. Anderer Haushalt mit mehreren Personen

[1] Nach dem gleichen Vorgehen wurde eine zweite Typologie berechnet, wobei das Höchstalter für Kinder auf 18 anstatt auf 25 Jahre festgelegt wurde.

[2] Als Konsensualpaare gelten ausschliesslich verschiedengeschlechtliche Paare.

[3] Als gleichgeschlechtliche Paare gelten Konsensualpaare und eingetragene Partnerschaften.

[4] Als Kind gilt die Tochter oder der Sohn eines Haushaltsmitglieds.

Quelle der Beziehungsinformationen

Die Informationen zu den Beziehungen zwischen den verschiedenen Personen in einem Haushalt stammen aus folgenden Quellen:

a. *Informatisiertes Zivilstandsregister* (**Infostar**): Die Infostar-Daten geben Auskunft über die Beziehungen «Vater / Kind», «Mutter / Kind», «Ehemann / Ehefrau» und «eingetragene Partner/innen» für alle Personen mit einem Zivilstandsereignis (z.B. Geburt, Heirat usw.) in der Schweiz, womit aktuell rund 85% der Wohnbevölkerung abgedeckt sind (99,8% der Schweizer/innen und 45,5% der Ausländer/innen).

b. *Register der Diplomatinnen und Diplomaten und der internationalen Funktionärinnen und Funktionäre* (**ORDIPRO**): Alle Personen mit einer Arbeitsbewilligung des Eidgenössischen Departements für auswärtige Angelegenheiten sowie deren Familienmitglieder und Personal sind im Register ORDIPRO erfasst. In diesem Register ist eine Variable zu den Beziehungen zwischen der Referenzperson (Diplomat/in oder internationale/r Funktionär/in) und bestimmten Familienmitgliedern wie der Ehefrau bzw. dem Ehemann, der Konsensualpartnerin bzw. dem Konsensualpartner oder den Kindern ersichtlich.

c. ***Strukturerhebung*** ab 2010 bis zum Jahr n–1: Sämtliche Beziehungen zwischen den einzelnen Haushaltsmitgliedern werden jährlich im Rahmen einer Stichprobenerhebung direkt bei der Bevölkerung ermittelt. Die in der Strukturerhebung angegebene Beziehung wird für alle «Personenpaare», die zum Zeitpunkt der Produktion der Typologie immer noch im gleichen Haushalt leben wie zum Erhebungszeitpunkt, übernommen, unter der Bedingung, dass es bei Konsensual- und Ehepaaren sowie Verwandten und Nichtverwandten keine Zivilstandsereignisse gab. So werden alle Beziehungstypen, die in der Strukturerhebung erfasst sind, berücksichtigt. Die Ergebnisse der Strukturerhebung werden erst nach der Produktion der Variable Haushaltstyp für die STATPOP veröffentlicht, weshalb jeweils lediglich die Daten des Jahres n–1 verfügbar sind.

d. ***Deterministische Algorithmen***: Auf Basis der demografischen Variablen in den Daten der STATPOP, die komplett auf Registerdaten beruht, lassen sich die Beziehungen zwischen zwei Personen mit relativ grosser Sicherheit bestimmen. Beispielsweise werden zwei verheiratete Personen mit demselben Heiratsdatum, die im gleichen Haushalt leben, als Ehefrau und Ehemann definiert. Hiermit lassen sich etwa verheiratete Paare ermitteln, die eingewandert sind und daher nicht im Infostar geführt werden, weil ihr Zivilstandsereignis nicht in der Schweiz stattgefunden hat.

Datenaufbereitung

Die aus den obengenannten Quellen ermittelten Beziehungen werden plausibilisiert (Beispiele: bei einer Beziehung Elternteil / Kind muss der Altersunterschied mindestens zwölf Jahre betragen; ein Kind kann nicht mehr als zwei Elternteile haben; eine Person kann lediglich mit einer Person verheiratet sein) und allfällige symmetrische Beziehungen werden ergänzt (z.B. wenn die Beziehung A/B beschrieben ist, aber nicht B/A).

Anschliessend werden **erweiterte Beziehungen** hinzugefügt. Dazu werden anhand der plausibilisierten Beziehungen die Verhältnisse abgeleitet, die in den Registern nicht enthalten sind (z.B. Grossvater, Schwager, Tante usw.). So sollen die Beziehungen bestmöglich vervollständigt werden, da der Haushaltstyp nur bestimmt werden kann, wenn alle Beziehungen im Haushalt bekannt sind.

Bei der Ermittlung der Beziehungen für das Jahr 2018 sind die Quellen wie folgt verteilt: Für die rund 16 Millionen haushaltsinternen «Personenpaare» konnten 84,04% der Beziehungen aus einer verfügbaren

Quelle entnommen oder direkt berechnet werden. Zur Bestimmung der Beziehungen werden die Quellen in folgender Reihenfolge herbeigezogen: Infostar (75,39%), ORDIPRO (+0,05%), Strukturerhebung (+6,12%), deterministischer Algorithmus (+1,78%), erweiterte Beziehung (+0,7%).

Tabelle 1: *Verfügbare Beziehungen nach Datenquelle*

	Infostar[5]	ORDIPRO	Strukturerhebung	Deterministischer Algorithmus	Entfernte Beziehungen
Ehefrau / Ehemann	X	X	X	X	
Eingetragene/r Partner/in	X	X	X	X	
Konsensualpartner/in		X	X	X[6]	
Mutter / Vater	X	X	X		
Tochter / Sohn	X	X	X		
Schwester oder Halbschwester / Bruder oder Halbbruder	X	X	X		
Stiefmutter / Stiefvater (Ehefrau / Ehemann eines Elternteils)	X		X		X
Stieftochter / Stiefsohn (Kind der Ehepartnerin / des Ehepartners)	X		X		X
Grossmutter / Grossvater	X	X	X		X
Enkel/in	X	X	X		X
Verwandt[7]		X	X		X
Nicht verwandt[8]	X		X		X

Anschliessend wird ein Algorithmus angewendet, der denjenigen Haushalten, für die alle Beziehungen bekannt sind, anhand der Beziehungen zwischen den einzelnen Haushaltsmitgliedern einen Haushaltstyp zuordnet. Für 2018 konnte 13,78% der Haushalte kein Haushaltstyp zugeordnet werden, d.h. mindestens eine der Beziehungen liess sich aus den verfügbaren Quellen nicht ableiten und erforderte eine Imputation.

Imputation

Es wurden mehrere Imputationsmethoden getestet, insbesondere ein «Random Forest» für die Beziehungen sowie für die Haushaltstypen, ein Verfahren zur deterministischen Imputation der Haushaltstypen, eine multinomiale Regression und ein Entscheidungsbaum. Die Wahl fiel schliesslich auf den Entscheidungsbaum. Mit einem «Random Forest» wurden zwar auf individueller Ebene leicht bessere Ergebnisse erzielt, aber mit einem Entscheidungsbaum lässt sich die Qualität der Imputationen besser einschätzen und das Verfahren flexibler anpassen.

Der Entscheidungsbaum wird anhand eines Lernensembles erstellt, das Haushalte mit bekanntem Haushaltstyp sowie ausgewählte Variablen umfasst, wobei diese Variablen sowohl im Lernensemble als auch in allen zu imputierenden Haushalten verfügbar sein müssen (z.B. Anzahl Personen im Haushalt, Durchschnittsalter der Haushaltsmitglieder, Anzahl verschiedener Nachnamen im Haushalt, Gemeindegrösse usw.). Mit statistischen Tests werden diejenigen Variablen bestimmt, mit denen sich die

[5] Lediglich die Beziehungen «Ehefrau/Ehemann», «eingetragene/r Partner/in», «Mutter/Vater» und «Tochter/Sohn» sind direkt verfügbar; alle anderen Verhältnisse aus dem Infostar werden daraus abgeleitet.

[6] bei nicht verheirateten Paaren mit gemeinsamen Kindern im gleichen Haushalt

[7] Zur Kategorie «Verwandt» gehören insbesondere Tanten, Onkel, Cousinen und Cousins.

[8] Zur Kategorie «Nicht verwandt» gehören insbesondere die Schwiegereltern (Eltern der Ehepartnerin/des Ehepartners), die Schwiegerkinder (Ehepartner/in der Kinder), die Kinder der Konsensualpartnerin/des Konsensualpartners und die Konsensaulpartnerin/der Konsensualpartner eines Elternteils.

Haushaltstypen am besten voneinander unterscheiden lassen. Diese Variablen bilden den Entscheidungsbaum, der anschliessend in einem zweiten Schritt auf die zu imputierenden Haushalte angewendet wird. So lässt sich anhand der Haushaltsmerkmale ein Typ imputieren.

Bei der Erstellung des Lernensembles für die Imputation spielt die Strukturerhebung eine grundlegende Rolle. Die Schnittmenge zwischen der nach dem Zufallsprinzip gezogenen Haushaltsstichprobe aus der Strukturerhebung und der Teilgesamtheit der zu imputierenden Haushalte des gleichen Jahres kann als Analysebereich innerhalb der Grundgesamtheit der Strukturerhebung betrachtet werden. Diese Schnittmenge wird für das Jahr n–1 berechnet und bildet das Lernensemble zur Imputation des Jahres n.

Ergebnisse

Vergleich der Verteilungen im Jahr 2018: Die aus der Teilgesamtheit der imputierten Haushalte ermittelte Verteilung wird mit der in der Strukturerhebung geschätzten Verteilung verglichen. Der euklidische Abstand zwischen diesen beiden Verteilungen beläuft sich auf 1,92.

Tabelle 2: *Verteilung der Haushaltstypen im Jahr 2018 gemäss Strukturerhebung und STATPOP*

Haushaltstyp	Schätzung aus der Strukturerhebung	Umfassende Typologie (STATPOP)
Einpersonenhaushalt	35,69	35,69
Einelternhaushalt mit mindestens einem Kind unter 25 Jahren	4,64	5,24
Anderer Haushalt mit mehreren Personen	7,59	8,01
Ehepaar ohne Kinder	20,11	19,45
Konsensualpaar ohne Kinder	6,53	7,51
Gleichgeschlechtliches Paar ohne Kinder	0.61	0,25
Ehepaar mit mindestens einem Kind unter 25 Jahren	22,12	20,87
Konsensualpaar mit mindestens einem Kind unter 25 Jahren	2,70	2.96
Gleichgeschlechtliches Paar mit mindestens einem Kind unter 25 Jahren	0,03	0,03

Schätzung der individuellen Fehler: Die Schnittmenge zwischen der Haushaltsstichprobe aus der Strukturerhebung und der Teilgesamtheit der imputierten Haushalte aus dem gleichen Jahr wird herangezogen, um den imputierten Typ auf individueller Ebene zu vergleichen. Dieser Vergleich ist möglich, weil die Daten der Strukturerhebung aus dem laufenden Jahr bei der Produktion des vollständigen «Haushaltstyps» nicht verwendet werden; sie dienen hier als Lernensemble für das Imputationsmodell. Die Gewichtungen der Strukturerhebung werden bei der Berechnung der geschätzten individuellen Fehler in der Haushaltstabelle berücksichtigt.

Tabelle 3: *Schätzung der individuellen Zuordnungsfehler nach Jahr*

Jahr	Geschätzter Anteil falscher Imputationen am Total der Haushalte
2018	1,45
2019	1,48
2020	1,47

Derzeit machen zwei Fehlertypen rund 76% aller Fehler aus (rund 1,1% der 1,45%). Es handelt sich um die Verwechslung der Haushaltstypen «Konsensualpaar ohne Kinder» und «Anderer Haushalt mit mehreren Personen» (rund 56,5%) sowie «Gleichgeschlechtliches Paar ohne Kinder» und «Anderer Haushalt mit mehreren Personen» (rund 19,5%). Diese Verwechslungen treten hauptsächlich bei Zweipersonenhaushalten auf. Personen unterschiedlichen Geschlechts werden vom Algorithmus als Konsensualpaar betrachtet, Personen gleichen Geschlechts werden hingegen als Wohngemeinschaft eingestuft. Dadurch wird die Kategorie «Gleichgeschlechtliches Paar ohne Kinder» unterschätzt. Dieser Punkt wird demnächst überarbeitet, sodass sich die Qualität der Imputationen deutlich verbessern dürfte.

Fazit

Zusammenfassend lässt sich sagen, dass die STATPOP-Variable «Haushaltstyp» mit relativ grosser Zuverlässigkeit verwendet werden kann. 87% der Haushaltstypen stammen direkt aus Erhebungsdaten und bei den 13% imputierten Haushalten ist der geschätzte Klassifikationsfehler gering (rund 1,45% falsche Zuordnungen beim Total aller Haushalte). Die ersten Rückmeldungen der Anwenderinnen und Anwender bestätigen den Nutzen dieser für nahezu alle Haushalte verfügbaren Variable. Bevor sie in die laufende STATPOP-Produktion integriert werden kann, gilt es allerdings noch einige Punkte zu verbessern. Insbesondere muss die Imputation bestimmter Kategorien verfeinert werden. Eine detaillierte Dokumentation der Methodik ist auf der Website des BFS verfügbar (Haushaltstypologie [STATPOP]).