



Centraal Bureau voor de Statistiek

Centrum voor Beleidsstatistiek

Documentatierapport Geboorteplaats 2004V1

VIDIV-nummer: 2008-060-mcb

Datum: 20 februari 2008

Opgesteld door: Jan van der Worp

Bronvermelding

Publicatie van uitkomsten geschiedt door het onderzoeksbureau of de opdrachtgever op eigen titel. Verwijzing naar het CBS betreft uitsluitend het gebruik van de microdatabestanden. Dat wordt als volgt geformuleerd:

“Eigen berekening [*naam onderzoeksbureau, c.q. opdrachtgever*] op basis van bij het CBS beschikbaar gestelde microdatabestanden betreffende geboorteplaatsen, welke afkomstig zijn van het CBS.”

Versiegeschiedenis

Versie 1

080220 Geboorteplaats 2004V1

080220 GBA-selectiebestand Geboorteplaats 2004V1

De gebruiker dient rekening te houden met de volgende *do's & don'ts*:

- Uitsluitend de combinatie van RIN met SRTNUM 'R' of 'S' identificeert een persoon. Zie blz. 10 voor een beschrijving.
 - De variabelen in het bestand betreffen tekstvelden die nog veel onderzoek vergen voordat deze informatie in standaardpublicaties kan worden gebruikt. De geboorteplaats van in Nederland geboren personen is gevuld met in het geboortjaar van de persoon geldige gemeentecodes. Dat is niet structureel gecheckt.
-

Inhoudsopgave

Bronvermelding	2
Versiegeschiedenis.....	3
1. Inleiding	7
2. Inhoud van de bestanden.....	8
2.1 Structuur van de bestanden.....	8
2.2 Overzicht van variabelen.....	8
2.3 Algemene toelichting.....	9
3. Beschrijving van de variabelen.....	10

1. Inleiding

Bij het Centrum voor Beleidsstatistiek (CvB) worden microdatabestanden van het Centraal Bureau voor de Statistiek (CBS) beschikbaar gemaakt voor gebruik door externe onderzoekers. Van alle personen die in het hier beschreven microdatabestand voorkomen, zijn gegevens uit de Gemeentelijke basisadministratie (GBA) geselecteerd. Deze selectie wordt het GBA-selectiebestand genoemd. Dit documentatierapport beschrijft de inhoud en structuur van het microdatabestand '080220 Geboorteplaats 2004V1'. Het complete GBA-bestand wordt in een apart documentatierapport beschreven.

Hoofdstuk 2 beschrijft de structuur van het microdatabestand, geeft een opsomming van de variabelen en bevat een toelichting op het bestand met daarbij een indicatie van de kwaliteit. In hoofdstuk 3 wordt per variabele meer informatie gegeven. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.

Indien gewenst, kan een bijlage geleverd worden met daarin de frequentieverdelingen van de categoriale variabelen: '080220 Bijlage 1 Frequentietellingen Geboorteplaats 2004V1'. In deze bijlage worden van de (semi-) continue variabelen de eerste en laatste vijf waarden gegeven.

2. Inhoud van de bestanden

2.1 Structuur van de bestanden

Microdatabestand

Een record in het microdatabestand bevat informatie van de geboorteplaats van een persoon. Een persoon kan dus niet meerdere records in het microdatabestand hebben. Het bestand bevat SRTNUM en RIN als sleutel(s) waarmee gekoppeld kan worden met andere bestanden.

GBA-selectiebestand

Het bijbehorende GBA-selectiebestand bevat persoons- en huishoudensgegevens van alle personen die in het microdatabestand voorkomen en ook geregistreerd staan in de GBA. De persoonsgegevens bestrijken een periode van 1 januari 1995 tot 1 januari 2006. De huishoudensgegevens worden door het CBS berekend voor de stand op 1 januari. In het GBA-selectiebestand zijn huishoudensgegevens beschikbaar van 1998 tot en met 2006.

Zowel het microdatabestand als het GBA-selectiebestand is gesorteerd op respectievelijk SRTNUM en RIN. Met behulp van deze twee variabelen als koppelsleutel kan de gebruiker persoons- en huishoudensgegevens uit het GBA-selectiebestand toevoegen aan het microdatabestand, zodat deze gegevens in het onderzoek meegenomen kunnen worden.

Om RIN-nummers toe te kennen aan het microdatabestand is gebruik gemaakt van de meest recente versie van de GBA. Hierin is slechts een beperkt aantal variabelen aanwezig, daarom wordt dit bestand niet gebruikt voor het GBA-selectiebestand. Het GBA-selectiebestand maakt gebruik van gegevens tot 1 januari 2006. Hierdoor kunnen personen die na 1 januari 2006 naar Nederland zijn geïmmigreerd, niet worden teruggevonden in het GBA-selectiebestand.

2.2 Overzicht van variabelen

Onderstaand volgt een lijst van alle variabelen in het microdatabestand. Zie hoofdstuk 3 voor een nadere toelichting op de variabelen in het microdatabestand.

<i>Nummer</i>	<i>Variabele & Label</i>	<i>Format</i>	<i>Lengte</i>	<i>Decimalen</i>
1	SRTNUM Soort identificatiecode	Tekst	1	0
2	RIN Identificatienummer	Getal	9	0
3	GEBJAAR Geboortejaar	Getal	4	0
4	GEBOORTEGEMEENTE Geboortegemeente	Tekst	4	0

2.3 Algemene toelichting

Populatieafbakening

De massa bestaat uit de bevolking van Nederland per $\pm$ 1 januari 2004. De geboorteplaats van personen die in het buitenland zijn geboren is leeggemaakt.

Kwaliteit

Het microdatabestand is geen publicatiebestand. Dit betekent dat voor veel variabelen door het CBS weinig of geen controles en correcties hebben plaatsgevonden. Op de variabelen die wel door het CBS bekeken zijn, zijn op grond van de beschikbare informatie correcties uitgevoerd. Deze correcties zijn als zodanig in het microdatabestand overgenomen.

Omdat het geen officieel vastgesteld publicatiebestand betreft, heeft dit tot gevolg dat de bestanden niet noodzakelijk exact overeenkomen met Statline, de elektronische databank van het CBS (statline.cbs.nl).

Een goede stelregel om in eerste instantie te kunnen inschatten of een variabele in het microdatabestand van goede kwaliteit is, is te kijken naar Statline. Als een kenmerk is opgenomen in Statline, is dat over het algemeen van goede kwaliteit. Als een kenmerk niet is opgenomen, dan is de kans groot dat deze variabele van matige of slechte kwaliteit is.

De gegevens op Statline kunt u vinden onder Nederland regionaal > Indelingen > Plaatsnamen in Nederland.

3. Beschrijving van de variabelen

In dit hoofdstuk wordt per variabele meer informatie gegeven. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.

1	SRTNUM
	Soort identificatiecode
R	RIN-persoon (persoon is gevonden in GBA) Bronbestanden die aan het CBS geleverd worden, bevatten doorgaans als persoonsidentificator het sofinummer/ burgerservicenummer. Het sofinummer/ burgerservicenummer wordt als zeer identificerend beschouwd en daarom vervangt het CBS dit nummer door een intern persoonsidentificatienummer, het betekenisloze RIN-nummer. Met dit RIN-nummer kan het bestand worden gekoppeld aan andere bestanden bij het CBS. Het bronbestand van dit microdatabestand bevatte geen sofinummer/ burgerservicenummer. Via andere identificerende gegevens dan het sofinummer/ burgerservicenummer kon het RIN-nummer ook aan dit bestand worden toegevoegd. Dit is gebeurd door van ieder record in het bronbestand de combinatie van GBAA-nummer op te zoeken in de GBA. Het CBS heeft namelijk iedere persoon in de GBA voorzien van een RIN-nummer. Als de specifieke combinatie voorkomt in de GBA, dan is het RIN-nummer van die persoon toegevoegd aan het record in het bronbestand. Op deze manier kon het RIN-nummer van personen in het bronbestand worden achterhaald. Het microdatabestand is voorzien van de door het CBS toegevoegde variabelen soort identificatiecode (SRTNUM) en identificatienummer (RIN). Records in het microdatabestand die konden worden voorzien van een RIN-nummer hebben de waarde 'R' op SRTNUM gekregen. Uitsluitend de combinatie van SRTNUM 'R' en RIN identificeert een uniek persoon. Koppelen aan andere bestanden of aggregeren op persoonsniveau moet daarom alleen worden uitgevoerd op records met SRTNUM 'R'. Met behulp van SRTNUM en RIN kan de gebruiker - gegevens uit het GBA-selectiebestand toevoegen aan de records in het microdatabestand met SRTNUM 'R'.
2	RIN
	Identificatienummer
	Persoonsidentificatienummer, ter vervanging van het sofinummer/ burgerservicenummer. Zie toelichting bij SRTNUM.
3	GEBJAAR
	Geboortejaar
4	GEBOORTEGEMEENTE
	Geboortegemeente
	Lege waarde betekent 'Geboren in het buitenland' Zie: mspv1f1cvb3\AP_Data\8_Uilities\Code_Listings\Gemeentecodes