

# Nieuwe data, 'nieuwe' methode? Een sociologisch historische toepassing van multiniveau-analyse\*

---

R.L. Zijdemans

*Diverse technische en methodische ontwikkelingen hebben grootschalig vergelijkend onderzoek naar historische sociale stratificatie mogelijk gemaakt. Met de kwantiteit van de onderzoekspopulatie is echter ook de diversiteit van de onderzoekspopulatie toegenomen. Aan de ene kant biedt deze variatie mogelijkheden voor onderzoek naar de invloed van contextuele kenmerken op het proces van statusverwerving. Aan de andere kant betekent het vaak dat de onderzoekseenheden uit één en dezelfde context niet onafhankelijk zijn, een belangrijke aanname bij veel gebruikte regressietechnieken. In een dergelijk geval raden statistici het gebruik van multiniveau-analyse aan. Na een beknopte uitleg van deze techniek volgt een toepassing ervan. Hierbij wordt onderzocht of het verband tussen beroepsstatus van vaders en zonen in Zeeland eind negentiende en begin twintigste eeuw verschilde tussen gemeenten en over de tijd. Tevens wordt getoetst of dergelijke verschillen samenhangen met urbanisering, migratie en religie.*

## Introductie

In 1983 wees Ultee op het gat tussen negentiende-eeuws historisch stratificatie-onderzoek en naoorlogs empirisch onderzoek.<sup>1</sup> In de jaren daarna verschenen enkele studies over dit onderwerp, maar Boonstra en Mandemakers concludeerden een decennium later dat het gebrek aan kennis over de stratificatie van de Nederlandse samenleving door deze studies slechts in beperkte mate was verminderd.<sup>2</sup> Een oorzaak hiervan zou kunnen liggen in de vergelijkbaarheid van de studies, met betrekking tot zowel de onderzochte populatie als de gebruikte meetinstrumenten.

Onderzoekers op het gebied van historische stratificatie hebben veelal handmatig data over individuen uit archieven moeten verzamelen, wat tamelijk arbeidsintensief is. Dit heeft noodzakelijkerwijs geleid tot het beperken van de

---

\* Graag bedank ik Vincent Buskens, Ineke Maas, Tom Snijders en Jeroen Weesie voor hun hulp bij het begrijpen van teksten en toepassingen van multiniveau-analyse. Eventuele fouten in de uitleg van de multiniveau-analyse komen geheel voor mijn rekening. De commentaren van twee anonieme referenten hebben me erg geholpen bij het schrijven van de uiteindelijke versie. Dit stuk is mede mogelijk gemaakt door NWO MAGW-Open subsidie 400-05-054 voor het project 'Status attainment during industrialization, Life courses in context'. De laatste versie is geschreven tijdens een verblijf bij de afdeling Toegepaste Sociale Wetenschappen aan de Universiteit van Stirling, Schotland.

1 Ultee, 'Het aanzien van beroepen', 28-48.

2 Boonstra en Mandemakers, 'Ieder is het kind', 125-141.

omvang van de onderzochte populatie. Studies hebben zich dan ook vaak op een bepaalde populatie en/of geografisch gebied gericht, zoals bijvoorbeeld: de elite in Amsterdam<sup>3</sup>. De diversiteit in onderzochte populaties en regio's maakt het lastig om onderzoeksresultaten te vergelijken en conclusies te trekken over het stratificatieproces in Nederland in de negentiende en begin twintigste eeuw.

Naast de beperkte omvang van onderzoekspopulaties heeft onenigheid over een juist meetinstrument van 'status' eenduidigheid in sociaal-historisch onderzoek bemoeilijkt. Status is in het verleden op diverse wijzen gemeten, onder andere op basis van de hoogte van betaalde belastingen<sup>4</sup>, of de positionering van iemands zitplaats in de kerk.<sup>5</sup> Het gebruik van beroep als indicator van status (beroepsprestige) zoals dat gebruikelijk is in hedendaags stratificatieonderzoek, is in het verleden om ten minste twee redenen bekritiseerd. Ten eerste zou een beroepstitel te weinig informatie verschaffen over iemands sociaal-economische positie.<sup>6</sup> Ten tweede zouden (historische) beroepstitels in betekenis verschillen van hedendaagse beroepstitels en daarmee niet geschikt zijn voor het doen van historisch onderzoek.<sup>7</sup>

### Nieuwe mogelijkheden voor sociaal-historisch onderzoek

Recente ontwikkelingen stellen sociaalhistorische onderzoekers echter nu *wel* in staat op grote schaal en op vergelijkbare wijze onderzoek te doen naar sociale stratificatie in de negentiende en begin twintigste eeuw.<sup>8</sup> Naast hedendaags onderzoek laat ook historisch onderzoek zien dat er een sterke overeenkomst is tussen het gemiddelde inkomen van een beroepsgroep en de sociaaleconomische positie die aan die beroepsgroep wordt toegekend door middel van een beroepsprestigeschaal, zoals die van Van Heek.<sup>9</sup> Het gebruik van beroepstitels lijkt daarmee ook in sociaalhistorisch onderzoek gerechtvaardigd,<sup>10</sup> des te meer omdat het probleem van verschil in de betekenis van beroepstitels tussen regio's, talen en over de tijd bestudeerd en vertaald is naar een historische internationale standaard voor classificatie van beroepen: HISCO.<sup>11</sup> Tevens zijn er verschillende schema's verschenen om de HISCO-gecodeerde beroepen om te zetten naar speciaal daarvoor gemaakte

---

3 De Vries, *Electoraat en elite*.

4 De Vries, *Electoraat en elite*.

5 Lucassen en Trienekens, 'Om de plaats', 239-304.

6 De Vries, *Electoraat en elite*, 12.

7 Van Leeuwen, Maas en Miles, *HISCO*, 9-10.

8 Voorbeelden van dergelijke studies zijn te vinden in: Mandemakers en Boonstra (ed.), *De levensloop*, en in Van Leeuwen, Maas en Miles, *Marriage choices*.

9 Van Heek, *Sociale stijging*. Zie bijvoorbeeld: Mandemakers, *Aanzet tot een beroepsstratificatie*, 198-222.

10 Boonstra en Mandemakers, 'Ieder is het kind', 125-141.

11 Van Leeuwen, Maas en Miles, *HISCO*.

historische klassenschema's, bijvoorbeeld HISCLASS<sup>12</sup> en SOCPO<sup>13</sup> of beroeps-prestigescalen, zoals HIS-CAM.<sup>14</sup>

Behalve methodische ontwikkelingen dragen ook ontwikkelingen op het gebied van dataverzameling bij aan de mogelijkheden om historische sociale stratificatie te onderzoeken. Digitalisering van persoonsgegevens leidt tot datasets die representatief zijn voor grotere regio's, langere perioden en vrijwel alle lagen van de bevolking. Voorbeelden van dergelijke digitaliseringprojecten zijn de Historische Steekproef Nederland (HSN)<sup>15</sup> en Genlias<sup>16</sup>. Maar ook diverse meer regionale digitaliseringsprojecten, zoals de Groene Hart-archieven<sup>17</sup> en het Noord-Hollands Archief<sup>18</sup> dragen bij aan de kennis over personen in de negentiende en begin twintigste eeuw.

### Individen in contexten

Om veranderingen in het stratificatieproces te verklaren gebruiken sociologen en historici echter niet alleen data over individuen, maar ook over de contexten waarin deze individuen zich bevinden. Ook op dit vlak vinden grootschalige ontwikkelingen plaats. Met behulp van de gedigitaliseerde volkstellingen en de Historische Database Nederlandse Gemeenten (HDNG) kunnen veranderingen in het stratificatieproces op individueel niveau gerelateerd worden aan de gemeentelijke context.<sup>19</sup> Zo kan de invloed van onder andere urbanisatiegraad, migratiepatronen, en dominante godsdienst op het stratificatieproces worden bepaald. Met de komst van de zogenaamde 'Amsterdamse code' worden deze analyses steeds minder afhankelijk van mutaties in namen en grondgebied van gemeenten.<sup>20</sup> Het Nederlands Historisch Geografisch Informatiesysteem (NLGIS) brengt contextuele informatie op gemeenteniveau en zelfs wijkniveau in kaart.<sup>21</sup> Bovengenoemde ontwikkelingen bieden sociaalwetenschappers en historici de mogelijkheid om onderzoek te doen naar grotere aantallen individuen uit meerdere contexten.

Bij het analyseren van deze grootschalige historische datasets, dient men in acht nemen dat ze zelden uit onafhankelijke onderzoekseenheden bestaan. Zo is er bij datasets die alle gedigitaliseerde aktes uit een bepaalde regio of periode omvatten geen sprake van een random sample. Maar zelfs wanneer de onderzoekseenheden via een aselechte steekproef verkregen zijn, kunnen onderzoekseenheden

---

12 Maas en Van Leeuwen, 'Total and relative endogamy', 275-295.

13 Van de Putte en Miles, 'A class scheme', 61-92.

14 Maas et al., HIS-CAM. Lambert et al., 'Testing the universality'.

15 Internetbron: HSN. <http://www.iisg.nl/~hsn/indexnl.html>

16 Internetbron: Genlias. <http://www.genlias.nl/nl/page0.jsp>

17 Internetbron: Groenehart Archieven. <http://www.groenehartarchieven.nl/>

18 Internetbron: Noord-Hollands Archief, <http://www.noord-hollandsarchief.nl/>

19 Beekink et al., *Nederland in verandering*.

20 Van der Meer en Boonstra, *Repertorium*.

21 Internetbron: NLGIS. <http://ip042.niwi.knaw.nl:9099/hgin/app>

systematische onderlinge samenhang vertonen. Wanneer onderzoekseenheden uit verschillen contexten komen, is het aannemelijk dat er tussen eenheden uit één en dezelfde context meer onderlinge samenhang is, dan tussen eenheden uit verschillende contexten. Zo kan bijvoorbeeld het verband tussen beroepsstatus van vaders en zonen verschillen tussen plaatsen en over de tijd.

Om rekening te houden met dergelijke regionale en temporele verschillen in de afhankelijke variabele wordt in regressieanalyse vaak gecontroleerd voor contextuele verschillen door dummyvariabelen op te nemen voor plaatsen en tijdstippen. Door het opnemen van interacties tussen variabelen op het individuele niveau en deze dummy's kan zelfs achterhaald worden of de invloed van individuele kenmerken verschilt tussen contexten. In het voorbeeld van intergenerationale statusverwerving kan men zodoende achterhalen of de invloed van vaders beroepsstatus op dat van zijn zoon verschilt tussen plaatsen en perioden.

Een andere manier om data van individuen uit verschillende contexten te analyseren is met behulp van multiniveau-analyse.<sup>22</sup> Deze analysetechniek is ontworpen om rekening te houden met het feit dat onderzoekseenheden op het zogenaamde eerste (lagere) niveau niet onafhankelijk zijn van elkaar, maar samenhangen in een bepaalde context (het tweede (hogere) niveau). In bovengenoemde voorbeelden zijn individuen bijvoorbeeld gegroepeerd (genest) binnen plaatsen en tijdstippen. Andere voorbeelden van data met een dergelijke multiniveaustructuur zijn: werknemers in bedrijven, kinderen in gezinnen, maar ook meetmomenten in proefpersonen.

Hoewel men met OLS regressie (op één niveau) de mogelijkheid heeft om te controleren voor een geneste datastructuur, is er een aantal redenen om multiniveaumodellen te gebruiken. Eén reden is dat het controleren voor context in een OLS regressiemodel ten koste gaat van de spaarzaamheid van het model. Wanneer men denkt aan een datastructuur waarbij individuen genest zijn in steden, zou men, om te controleren voor contexteffecten, voor alle steden dummyvariabelen moeten opnemen. Dit gaat ten koste van het aantal vrijheidsgraden. In een multiniveaumodel is een groot aantal eenheden op een hoger niveau echter geen probleem. Het specificeren van een hoger niveau gaat ten koste van slechts één vrijheidsgraad ongeacht het aantal eenheden op dat hogere niveau. Multiniveau-analyse is dus veel zuiniger in het gebruik van de data. Een tweede reden is dat men met multiniveau-analyse macro-micro hypothesen kan toetsen, *terwijl* men controleert voor de geneste data-structuur. Zodoende kan men vragen beantwoorden als: Is de invloed van beroepsstatus op kindersterfte groter in meer landelijke of in meer stedelijke gebieden? Neemt de invloed van vaders beroepsstatus op dat van zijn zoon af met industrialisering? Dergelijke vragen zijn in het verleden ook wel onderzocht met behulp van OLS regressie-analyse, maar met

---

22 Hiërarchisch lineair model, randomcoëfficiëntmodel of variantiecomponentmodel zijn andere aanduidingen voor het multiniveaumodel.

OLS regressie kan men niet tegelijkertijd contexthypothese toetsen en controleren voor de geneste datastructuur. Immers wanneer men naast dummyvariabelen voor gemeenten ook indicatoren voor gemeentekennmerken in het model opneemt, zal een deel van de effecten van de gemeentekennmerken al worden opgevangen door de dummy variabele van de gemeente zelf. Effecten van gemeentekennmerken worden dan onderschat. Daarnaast is het nagenoeg onmogelijk om te controleren voor de geneste structuur van de data door enkel en alleen te controleren voor gemeentekennmerken. Daarvoor zou men immers moeten beschikken over alle relevante gemeentekennmerken. Bij multiniveau-analyse wordt gecontroleerd voor de geneste datastructuur door regressiecoëfficiënten van variabelen op het eerste niveau per groep te laten verschillen. Door vervolgens interactie-effecten tussen variabelen op het eerste en tweede niveau in het model op te nemen, kunnen macro-micro hypothesen worden getoetst. De volgende sectie beschrijft in meer detail hoe in multiniveau-modellen voor één variabele meerdere regressiecoëfficiënten worden geschat. Daarna volgt een toepassing van multiniveau-analyse op een dataset van Zeeuwse huwelijksakten uit de negentiende en begin twintigste eeuw.

### Multiniveau-modellen<sup>23</sup>

Multiniveau-modellen typeren zich door de variatie in de afhankelijke variabele op te splitsen tussen verschillende niveaus. Om dit inzichtelijk te maken geeft vergelijking 1 een OLS regressie model weer met een individuele variabele  $X$ .  $\beta_0$  is de constante (intercept) en de residuen  $R_{ij}$  zijn de afwijkingen van het gemiddelde van  $Y$  die niet verklaard zijn door  $X$ .

$$[1] \quad Y_{ij} = \beta_0 + \beta_1 x_{ij} + R_{ij}$$

De subscript  $i$  voor individu en  $j$  voor groep geven aan op welk niveau een variabele betrekking heeft. De waarden van de variabele  $X$  verschillen voor alle individuen  $i$  uit alle groepen  $j$ . (Een variabele  $Z$  op het groepsniveau zou dus alleen het subscript  $j$  krijgen, omdat de waarden voor die variabele dan alleen van groepen afhangen.) In vergelijking (1) hebben geen van beide schatters  $\beta_0$  en  $\beta_1$  een subscript. Ze hebben dezelfde *vaste* waarde voor alle individuen uit alle groepen. Om de waarde van de schatters te laten variëren voor groepen, worden ze gesplitst in een gemiddeld effect (voor alle groepen hetzelfde) en een groep-specifieke afwijking van dit gemiddelde effect (verschillend per groep). De schatters krijgen dan het subscript  $j$ :

$$[2] \quad \begin{aligned} \beta_{0j} &= \gamma_0 + U_{0j} \\ \beta_{1j} &= \gamma_1 + U_{1j} \end{aligned}$$

---

23 De uitleg van het multiniveau-model is gebaseerd op hoofdstuk 4 en 5 van Snijders en Bosker, *Multilevel analysis*.

In vergelijking (2) zijn  $\gamma_0$  en  $\gamma_1$  de *vaste* (fixed) effecten van intercept en helling.  $U_{0j}$  en  $U_{1j}$  zijn de *random*-effecten, de groepsspecifieke afwijkingen van respectievelijk de intercept en helling. Wanneer deze groepsspecifieke schatters worden ingevuld in vergelijking (1) ontstaat het volgende multiniveaumodel:

$$[3] \quad Y_{ij} = \gamma_0 + \gamma_1 x_{ij} + U_{0j} + U_{1j} x_{ij} + R_{ij}$$

Het eerste deel van vergelijking (3) bestaat uit vaste effecten, het tweede deel uit random-effecten. De random effecten  $R_{ij}$  en  $U_{0j}$  en  $U_{1j}$  geven de (door X) onverklaarde afwijkingen van het gemiddelde van Y weer. De onverklaarde variantie is dus nu gesplitst naar het individuele niveau:  $\text{var}(R_{ij}) = \sigma^2$  en het groepsniveau:  $\text{var}(U_{0j}) = \tau_1^2$  en  $\text{var}(U_{1j}) = \tau_2^2$ . Om de onverklaarde variantie te verminderen kunnen aan het model extra variabelen worden toegevoegd. Variabelen op het individuele niveau kunnen variantie op het individuele niveau verklaren, maar ook op het groepsniveau. Dit is het geval wanneer de samenstelling van groepen in individuele kenmerken tussen groepen varieert. Het toevoegen van contextvariabelen verkleint ook de onverklaarde variantie op het groepsniveau. Zo kan aan vergelijking (3) een hoofd- en interactie-effect van groepsvariabele Z worden toegevoegd:

$$[4] \quad Y_{ij} = \gamma_0 + \gamma_1 x_{ij} + \gamma_2 x_{ij} z_j + \gamma_3 z_j + U_{0j} + U_{1j} x_{ij} + R_{ij}$$

De random component, het tweede deel uit vergelijking (4), is dus dezelfde als in vergelijking (3), maar de residuen hebben nu betrekking op de onverklaarde afwijkingen van het gemiddelde van Y na controle voor X, Z en het interactie-effect. Aan de hand van dit model kan een macro-micro hypothese worden getoetst, waarbij tegelijkertijd gecontroleerd wordt voor de geneste datastructuur. In de volgende paragrafen volgt een toepassing van zo een multiniveau-model, waarbij de invloed van gemeentekenmerken op het proces van intergenerationale statusverwerving wordt onderzocht.

## Probleemstelling

Om enkele van de mogelijkheden van multiniveau-analyse te laten zien, stel ik twee vragen aan de orde. Ten eerste, in hoeverre verschilt de invloed van vaders beroepsstatus op dat van zijn zoon tussen plaatsen en over de tijd in Zeeland tussen 1851 en 1915? Ten tweede, in hoeverre kunnen contextuele kenmerken verschillen in de invloed van vaders beroepsstatus op dat van zijn zoon tussen plaatsen en over de tijd verklaren? Daarbij toets ik drie macro-micro hypothesen, waarbij gesteld wordt dat de associatie tussen beroepsstatus van vader en zoon (individuele niveau) beïnvloed wordt door contextuele kenmerken (groepsniveau).

De eerste twee hypothesen hebben betrekking op verschillen in de mate van modernisering van gemeenten. Volgens de moderniseringstheorie is het effect van vaders beroepsstatus kleiner in meer geurbaniseerde gebieden en in gebieden met

meer migratie.<sup>24</sup> De redenering is dat in grotere steden niet iedereen elkaar kent en men dus minder voordeel (of last) heeft van zijn achtergrond. Daarnaast is de diversiteit in achtergrond groter in grotere steden en zou een anders dan gemiddelde afkomst van minder belang zijn. Meer migratie zou een gelijk effect te weeg brengen en eveneens het verband tussen beroepsstatus van vader en zoon verkleinen.

De derde hypothese heeft betrekking op culturele verschillen in context. Volgens Webers wereldbeeldentheorie hebben mensen een bepaald wereldbeeld dat meer of minder activistisch is.<sup>25</sup> Naarmate mensen een meer activistisch wereldbeeld hebben, beschouwen ze de wereld als veranderbaar en zou afkomst een minder grote rol spelen. In het algemeen wordt gesteld dat katholieken een minder activistisch wereldbeeld hebben dan protestanten, waardoor de invloed van afkomst voor katholieken groter zou zijn dan voor protestanten. De protestanten in Zeeland in de negentiende eeuw waren echter veelal Nederlands gereformeerd, maar ook orthodox calvinistisch.<sup>26</sup> Ik verwacht daarom dat deze meer conservatieve protestanten vasthielden aan meer traditionele waarden en een minder activistisch wereldbeeld hadden dan de katholieken. Helaas omvatten de data alleen godsdienstgegevens op gemeente niveau en niet op individueel niveau. Daarom neem ik aan dat de invloed van beroepsstatus van de vader op dat van de zoon groter is in meer conservatief protestantse gebieden.

### Data en specificatie van de multiniveauctuur

Om deze hypothesen te toetsen maak ik gebruik van de data op het individuele en groepsniveau. Data over individuen zijn afkomstig uit het databestand van huwelijksakten van het Zeeuws Archief te Middelburg. Uit dit bestand gebruik ik voor de periode 1851-1915 alleen de huwelijksakten van bruidegommen die voor het eerst huwen. Bruidegommen die voor een tweede of op eenvolgende keer huwen zijn verder in hun carrière, waardoor voor hen het verband tussen beroepsstatus van vader en zoon verschilt van nog niet eerder gehuwde bruidegommen. Helaas wordt niet altijd het beroep van de vader en zoon (bruidegom) vermeld op de huwelijksakte. Soms ontbreekt een beroepstitel, omdat er sprake was van een andere bron van inkomsten, maar meestal omdat de persoon in kwestie is overleden. In dat geval werd er geen beroep genoteerd. Het is dus vrijwel altijd het beroep van de vader dat ontbreekt. In totaal kunnen er van de 85.982 huwelijksakten van zonen die voor het eerst huwen, 41.544 (48,3%) gebruikt worden in de analyse. De keuze voor de periode 1851-1915 heeft te maken met de data die voor handen zijn op het contextuele niveau. De contextuele data zijn afkomstig uit datasets van de Historisch Ecologische Database en de Historische Databank Nederlandse Gemeenten die gebaseerd zijn op de volkstellingen. De data

---

24 Treiman, 'Industrialization', 207-234.

25 Weber, *The sociology*. Lenski, Lenski en Nolan, *Human societies*. Ultee, Arts en Flap. *Sociologie*.

26 Bras en Kok, 'They live', 247-274.

zijn beschikbaar voor elke 10 jaar. De tussenliggende jaren zijn geschat met een gewogen gemiddelde van de jaren waar wel informatie voor bekend was.

De afhankelijke variabele in de analyses is de beroepsstatus van de zoon (bruidgom). Beroepstitels zijn eerst gecodeerd volgens HISCO en vervolgens omgezet naar een continue historische beroepsprestigeschaal: HIS-CAM(v.0.1). HIS-CAM is een historische versie van de CAMSIS-schalen die worden gebaseerd op patronen van sociale interactie (bijvoorbeeld huwelijken).<sup>27</sup> De veronderstelling daarbij is dat naarmate mensen van verschillende beroepsstrata vaker met elkaar omgaan, de sociale afstand tussen de mensen uit deze strata kleiner is. De HIS-CAM schaal is een geschatte historische beroepsstratificatie op basis van anderhalf miljoen huwelijksakten uit zes verschillende landen (Canada, Duitsland, Frankrijk, Nederland, het Verenigd Koninkrijk en Zweden) voor de periode 1800-1938. De HIS-CAM schaal loopt van 1 tot en met 99.

De onafhankelijke variabelen zijn:

*Status vader:* evenals de beroepstitels van de zonen, zijn de beroepstitels van de vaders in HISCO gecodeerd en is er vervolgens een HIS-CAM score aan toegekend. Vaders beroepsstatus is gecentreerd over het grootgemiddelde (grand mean) over de periode 1851-1915.

*Urbanisering:* dit is bevolkingsomvang per 1000 inwoners in de huwelijksgemeente en het huwelijksjaar.

*In-migratie:* is gemeten als de proportie migranten ten opzichte van de bevolkingsomvang in de gemeente en het jaar van huwelijk.

*Conservatief protestantisme:* is gemeten als de proportie Nederlands gereformeerden en orthodox calvinisten ten opzichte van de bevolkingsomvang. Aangezien er nauwelijks niet-religieuzen waren, neem ik aan dat het hier om de verhouding tussen conservatieve protestanten en katholieken gaat.

Naast de onafhankelijke variabelen zijn er in de modellen controlevariabelen opgenomen:

*Leeftijd:* de leeftijd van de zoon gecentreerd over het grootgemiddelde.<sup>28</sup>

*Migrant:* geeft aan of een zoon in de gemeente huwde, waar hij geboren was of niet.

*Moeder overleden:* deze dummy variabele geeft aan of de moeder van de zoon reeds was overleden voor de huwelijksvoltrekking.

*Decade:* is een continue variabele voor tijd. Decade is 1 in 1851 en neemt ieder jaar toe met 0,1.

Alvorens de data te analyseren dienen de niveaus in het multiniveaumodel gespecificeerd te worden. In de Zeeland-data zijn individuen waargenomen tijdens het

---

27 Prandy, 'The social interaction', 215-249. Prandy en Lambert, 'Marriage', 397-411.

28 Het grootgemiddelde heeft betrekking op de periode 1811-1915.



huwelijk dat in een bepaalde gemeente en in een bepaald jaar gesloten is. Enerzijds zijn individuen dus gegroepeerd binnen plaatsen (bijv. alle huwelijken in Middelburg), anderzijds binnen tijd (bijv. alle huwelijken in 1880). Nu vereist een hiërarchisch lineair model (het type multi-niveau model dat hierboven besproken is) dat de verschillende niveaus binnen elkaar vallen. Bijvoorbeeld: leerlingen binnen klassen, of burgers binnen gemeenten. Tijd en plaats zijn echter niet hiërarchisch gerangschikt en zou men normaliter gebruik maken van een gemengd-klasse-multiniveaumodel (cross-classified model).<sup>29</sup> In zo'n model zijn individuen dan enerzijds genest binnen gemeenten en anderzijds genest binnen tijd. Een nadeel van deze aanpak is echter dat alleen gemeente-specifieke variabelen gebruikt kunnen worden die niet veranderen over de tijd (bijv. ligging aan zee) of tijds-specifieke variabelen die niet verschillen per gemeente (bijv. verandering in nationale wetgeving). Daarom wordt in de literatuur voorgesteld om elke combinatie van gemeente en tijd als groep te zien.<sup>30</sup> Dus de huwelijken in Middelburg 1880 vormen een groep, de huwelijken in Middelburg 1881 vormen een groep en zo ook de huwelijken in Vlissingen 1881. Het aantal groepen is dus gelijk aan het product van tijdstippen en gemeenten. Op deze manier kunnen wel tijd- en regio specifieke kenmerken, zoals bevolkingsomvang, worden opgenomen.

## Resultaten

Tabel 1 geeft de resultaten van de multiniveau regressie van beroepsstatus van zonen op individuele en contextuele indicatoren weer. Model 0 is het eerste model in de tabel en bevat geen verklarende variabelen. In dit *lege model* wordt de totale variantie in Y opgesplitst tussen variantie op het individuele ( $\sigma^2$ ) en groepsniveau ( $\tau^2$ ). Om een indruk te krijgen van hoeveel van de variantie in Y zich op het groepsniveau bevindt, deel ik de variantie op het groepsniveau ( $\tau^2$ ) door de totale variantie ( $\sigma^2 + \tau^2$ ). Dit quotiënt, de *intraklasse correlatiecoëfficiënt*, bedraagt 0,118 ( $19,489 / (19,489 + 145,481)$ ). Met andere woorden: 11,8 procent van de variantie in Y wordt op het groepsniveau bepaald. Tussen de beroepsstatus van zonen uit dezelfde gemeente en jaar van huwelijk bestaat dus meer samenhang dan op basis van toeval verwacht zou mogen worden. De gemiddelde HIS-CAM score van de beroepsstatus van de zoon is 42,047, maar dit gemiddelde verschilt dus tussen groepen. Om te zien hoe groot die verschillen zijn, vergelijk ik de groepen met de laagste en hoogste score met elkaar. In de 2,5 procent groepen met de laagste score, is de beroepsstatus van de zoon 33,329 ( $42,047 - 2\sqrt{19,489}$ ), terwijl in de 2,5 procent 'hoogste' groepen de beroepsstatus anderhalf keer zo groot is: 50,765 ( $42,047 + 2\sqrt{19,489}$ ).

In model 1 is er een vast en een random effect voor vaders beroepsstatus aan het model toegevoegd. Naast een gemiddelde invloed van vaders beroepsstatus

---

29 Voor toepassing van gemengd-klasse-niveaumodellen zie: Tubergen, *The integration*.

30 Duncan, Jones en Moon, 'Health related behaviour', 817-830.

Tabel 1. – Multiniveau regressie van beroepsstatus zoon (HIS-CAM) op individuele en contextuele variabelen voor Zeeland 1851-1915 (N=41.544).

	Model 0		Model 1		Model 2		Model 3	
	Coëff.	Err.	Coëff.	Err.	Coëff.	Err.	Coëff.	Err.
<i>Vaste effecten</i>								
Constante	42,047	0,088	40,907	0,169	39,262	0,245	39,257	0,237
Status vader			0,571	0,012	0,713	0,021	0,713	0,019
Decade			0,550	0,037	0,449	0,034	0,442	0,030
* status vader			-0,002	0,003	0,001	0,003		
Urbanisatie					0,518	0,016	0,529	0,012
* status vader					-0,001	0,001		
In-migratie					10,607	2,311	10,662	2,308
* status vader					-0,425	0,206	-0,424	0,204
Cons. Protestantisme					0,114	0,196	0,116	0,196
* status vader					-0,150	0,018	-0,150	0,018
Leeftijd			0,222	0,012	0,238	0,012	0,238	0,012
Migrant			1,462	0,098	1,454	0,097	1,455	0,097
Moeder overleden			-0,241	0,112	-0,199	0,109	-0,199	0,109
<i>Level 2 random effecten</i>								
Intercept ( $U_{0j}$ )	19,489	0,832	10,007	0,483	5,385	0,385	5,378	0,358
Status vader ( $U_{1j}$ )			0,048	0,003	0,048	0,003	0,048	0,003
<i>Level 1 variantie</i>								
Intercept ( $R_{ij}$ )	145,481	1,081	88,799	0,687	89,137	0,680	89,150	0,680
Deviantie	328.282,500		308.816,200		307.467,700		307.469,200	

wordt er dus een groepsspecifiek effect geschat, zoals in vergelijking (3) het geval was. Eveneens wordt er gecontroleerd voor individuele variabelen, tijd en een interactie van tijd met vaders beroepsstatus. De resultaten laten zien dat gemiddeld de beroepsstatus van zonen toeneemt (afneemt) met 0,571 voor ieder punt dat de beroepsstatus van vader toeneemt (afneemt). Het random effect van vaders beroepsstatus is echter ook significant, wat betekent dat de invloed van vaders beroepsstatus inderdaad tussen groepen verschilt. Hoewel deze verschillen ook door niet-lineaire tijdseffecten worden veroorzaakt, hebben plaatselijke verschillen waarschijnlijk het grootste aandeel in de variërende invloed van vaders beroepsstatus. Op de grootte van de verschillen te laten zien, vergelijk ik weer de groepen met de laagste en hoogste score van vaders beroepsstatus. In de 2,5 procent groepen met de laagste score, is de invloed van vaders beroepsstatus 0,133 ( $0,571 - 2\sqrt{0,048}$ ) terwijl in de 2,5 procent 'hoogste' groepen de invloed van vaders beroepsstatus bijna tien keer zo groot is: 1,009 ( $0,571 + 2\sqrt{0,048}$ ). Het interactie effect van vaders beroepsstatus en decade is niet significant, wat veronderstelt dat er geen sprake was van een lineaire afname in de invloed van vaders beroepsstatus. De controlevariabelen in dit model en volgende modellen laten zien dat de beroepsstatus van zonen die ouder zijn, en zonen die in een andere plaats dan hun geboorteplaats huwen een hogere beroepsstatus hebben. Zonen waarvan de moeder voor het huwelijk is overleden hebben in dit model een lagere status, maar dit effect is niet langer significant wanneer gecontroleerd wordt voor contextuele kenmerken (model 2 en 3).

In Model 2 zijn de contextuele kenmerken voor urbanisatie, in-migratie en de mate van conservatief protestantisme opgenomen. Daarnaast zijn interacties van deze kenmerken met beroepsstatus van de vader opgenomen. Model 3 is verkregen door alle niet-significante interactie-effecten uit Model 2 weg te laten. De invloed van vaders beroepsstatus op dat van de zoon neemt niet af met urbanisatie, zoals verondersteld werd door de moderniseringstheorie. De invloed van vaders beroepsstatus neemt wel af met de proportie in-migranten in een gemeente. In een gemiddelde gemeente met vijf procent in-migranten, is het effect van vaders beroepsstatus bijna drie procent kleiner dan in een gemeente zonder in-migranten ( $(-0,424 * 0,05) / 0,713$ ). De derde hypothese over de invloed van de proportie conservatieve protestanten in een gemeente wordt weerlegd. De invloed van vaders beroepsstatus neemt af met de proportie conservatieve protestanten. Per gemeente zijn er gemiddeld 79,0 procent conservatieve protestanten. De invloed van vaders beroepsstatus is in een dergelijke gemeente 16,6 procent kleiner dan de invloed van vaders beroepsstatus in een volledig katholieke gemeente ( $(-0,150 * 0,79) / 0,713$ ).

## Conclusies en discussie

Deze studie heeft beargumenteerd dat bij de grootschalige historische datasets het probleem op kan treden, dat de onderzoekseenheden niet onafhankelijk zijn, maar onderlinge samenhang vertonen. Dit is geïllustreerd aan de hand van een

dataset gebaseerd op Zeeuwse huwelijksakten uit de periode 1851-1915, waarbij bleek dat de beroepsstatus van zonen niet alleen afhing van individuele kenmerken, maar ook verschilde tussen huwelijksplaats en jaar. Met behulp van multiniveau-analyse kunnen wetenschappers echter rekening houden met een dergelijke samenhang en gelijktijdig de invloed van contextuele kenmerken op individuele processen schatten. De resultaten wijzen uit dat invloed van vaders beroepsstatus op dat van de zoon verschilde tussen gemeenten. In gemeenten met meer in-migratie en een hogere proportie conservatieve protestanten bleek de invloed van vaders beroepsstatus kleiner. Er is geen bewijs gevonden dat het proces van intergenerationale statusverwerving samenhangt met urbanisatie of lineair afnam over de tijd. Dit laatste sluit aan bij eerdere bevindingen over het proces van intergenerationale statusverwerving in Nederland in de negentiende en begin twintigste eeuw.<sup>31</sup>

Hoewel voor de geneste structuur in de gebruikte data is gecontroleerd, zijn er mogelijk andere moeilijkheden bij het analyseren van (deze) historische data. In eerder onderzoek is gewezen op problemen bij gebruik van huwelijksakten voor onderzoek naar sociale stratificatie.<sup>32</sup> Zo bevatten de huwelijksakten geen informatie over niet-gehuwden. Hoewel dit aantal misschien is toegenomen over de onderzochte tijdsperiode, was het aantal huwelijken in de negentiende en begin twintigste eeuw relatief hoog. Een groter selectieprobleem lijkt het ontbreken van de beroepstitel van de vader op de huwelijksakte door vroegtijdig overlijden. Dit hangt mogelijk samen met slechte leef- en arbeidsomstandigheden. Zonen van vaders uit lagere sociaal-economische strata zullen daarom vaker in de data ontbreken dan zonen van vaders uit hogere strata. Dit heeft tot gevolg dat het verband tussen vaders en zonen wordt onderschat. Door verbeterende leef- en arbeidsomstandigheden is dit selectieprobleem aan het eind van de onderzochte periode minder groot dan aan het begin van de periode.

Samengevat, met behulp van multiniveau-analyse is het mogelijk om de invloed van context op individuele processen aan de hand van een dataset met een geneste structuur te schatten. De invloed van vaders beroepsstatus op dat van de zoon werd niet minder over de tijd, maar was wel kleiner in gemeenten met hogere percentages conservatieve protestanten en nam af naarmate er in een gemeente meer in-migratie was. Dit zijn echter slechts enkele kenmerken waarop gemeenten onderling en over de tijd kunnen verschillen. Naarmate er meer data op contextueel niveau digitaal beschikbaar komen, zal het begrip van regionale en temporale verschillen in het statusverwervingsproces toenemen.

---

31 Van Dijk, Visser en Wolst, 'Regional differences', 435-452. Van Leeuwen en Maas, 'Social mobility', 619-664.

32 Delger en Kok, 'Bridegrooms and biases', 113-121.

## Literatuurlijst

- Beekink, E., Boonstra, O., Engelen, T. et al., *Nederland in verandering. Maatschappelijke ontwikkelingen in kaart gebracht 1800-2000* (Amsterdam 2003).
- Boonstra, O.W.A. en K. Mandemakers, “Ieder is het kind zijner eigene werken”. Sociale stratificatie en mobiliteit in Nederland in de achttiende en negentiende eeuw’, in: Dronkers, J. en W.C. Ultee (eds.), *Verschuivende ongelijkheid in Nederland* (Assen 1995), 125-141.
- Bras, H. en J. Kok, “‘They live in indifference together’, Marriage mobility in Zeeland, the Netherlands’, 1796-1922, in Leeuwen, M.H.D. van, Maas, I. en A. Miles (eds.), *Marriage choices and class boundaries: social endogamy in history* (Cambridge 2005), 247-274.
- Delger, H. en J. Kok, ‘Bridegrooms and biases’, *Historical Methods*, 31 (1998), 113-121.
- Dijk, H. van, J. Visser en E. Wolst, ‘Regional differences in social mobility patterns in the Netherlands between 1830 and 1940’, *Journal of Social History*, vol. 17,3 (1984), 435-452.
- Duncan, C., Jones, K. en G. Moon, ‘Health related behaviour in context: a multilevel modeling approach’, *Social Science Medicine*, vol.42, 6 (1996), 817-830.
- Genlias, <http://www.genlias.nl/nl/page0.jsp>, (6 september 2007).
- Groenehart Archieven, <http://www.groenehartarchieven.nl/>, (6 september 2007).
- Heek, F. van, *Sociale stijging en daling in Nederland. I*, (Leiden, Stenfert Kroese, 1958).
- Historische Steekproef Nederland (HSN), <http://www.iisg.nl/~hsn/indexnl.html>, (6 september 2007).
- Lambert, P.S., Zijdemans, R.L., Maas, I., et al., ‘Testing the universality of historical occupational stratification structures’, paper gepresenteerd op *de ISA RC28 Social Stratification and Mobility spring meeting*, Nijmegen, 2006.
- Leeuwen, M.H.D. van en I. Maas, ‘Social mobility in a Dutch province, Utrecht 1850-1940’, *Journal of Social History*, 30 (1997), 619-664.
- Leeuwen, M.H.D. van, Maas, I. en A. Miles, *HISCO. Historical international standard classification of occupations*, (Leuven, 2002).
- Leeuwen, M.H.D. van, Maas, I. en A. Miles, *Marriage choices and class boundaries: social endogamy in history*, IRSH Supplement 13, (Cambridge 2006).
- Lenski, G., Lenski, J. en P. Nolan, *Human societies: an introduction to macrosociology*, (New York, 1991).
- Lucassen, J. en G. Trienekens, ‘Om de plaats in de kerk’, *Tijdschrift voor Sociale Geschiedenis*, 4 (1978), 239-304.
- Maas, I. en M.H.D. van Leeuwen, ‘Total and relative endogamy by social origin: a first international comparison of changes in marriage choices during the nineteenth century’, in: Leeuwen, M.H.D. van, Maas, I. en A. Miles, *Marriage choices and class boundaries: social endogamy in history* IRSH Supplement 13, (Cambridge 2005), 275-295.

- Maas, I., Lambert, P.S., Zijdeman, R.L. et al., 'HIS-CAM – The derivation then implementation of an historical occupational stratification scale', Paper gepresenteerd op de *Sixth European Social Science History Conference*, Amsterdam, 22-25 March (2006).
- Mandemakers, K., 'Aanzet tot een beroepsstratificatie voor Nederland omstreeks 1900. Tiel 1884 en 1918', in: *Tijdschrift voor Sociale Geschiedenis*, vol. 13 (1987), 198-222.
- Mandemakers, K., en O. Boonstra (eds.), *De levensloop van de Utrechtse bevolking in de 19e eeuw* (Assen 1995).
- Meer, A. van der en O. Boonstra, *Repertorium van Nederlandse gemeenten 1812-2006*, (Den Haag 2006).
- Nederlands Historisch Geografisch Informatiesysteem (NLGIS), <http://ip042.niwi.knaw.nl:9099/hgin/app>, (6 september 2007).
- Noord-Hollands Archief, <http://www.noord-hollandsarchief.nl/>, (6 september 2007).
- Prandy, K., 'The social interaction approach to the measurement and analysis of social stratification', *International Journal of Sociology and Social Policy*, 19 (2000a), 215-249.
- Prandy, K. en P.S. Lambert, 'Marriage, social distance and the social space: an alternative derivation and validation of the Cambridge scale', *Sociology*, 37 (2003), 397-411.
- Putte, B. van de en A. Miles, 'A class scheme for historical occupational data. The analysis of marital mobility in industrial cities in 19<sup>th</sup> century Flanders and England', *Historical Methods*, vol. 38,2 (2005), 61-92.
- Snijders, T.A.B. en R.J. Bosker, *Multilevel analysis. An introduction to basic and advanced multilevel modeling* (Londen 1999).
- Treiman, D.J., 'Industrialization and social stratification', *Social inquiry*, 40 (1970), 207-234.
- Tubergen, F. van, *The integration of immigrants in cross-national perspective. Origin, destination, and community effects* (Utrecht, 2005).
- Ultee, W.C., 'Het aanzien van beroepen, op andere plaatsen en vooral in andere tijden. Een analyse van een aantal recent historische studies', *Tijdschrift voor Sociale Geschiedenis*, 9 (1983), 28-48.
- Ultee, W.C., Arts, W.A. en H.D. Flap, *Sociologie. Vragen, uitspraken, bevindingen*, 3<sup>e</sup> druk (Groningen, Martinus Nijhoff, 2003).
- Vries, B. de, *Electoraat en elite. Sociale structuur en sociale mobiliteit in Amsterdam 1850-1895* (Amsterdam 1986).
- Weber, M. *The sociology of religion*, (Boston [1922] 1993).