

Het leven van een statisticus

In 1989 startte ik met de studie Wiskunde. Natuur- en Scheikunde vond ik ook heel leuk, maar ik zag een beetje op tegen de practica. Ik kwam er echter al snel achter dat ik de studie Wiskunde te theoretisch vond. De bovenbouwstudie Statistiek was toen een goede uitkomst. Ik haal mijn motivatie namelijk uit de mogelijke toepassingen in de praktijk. Tijdens de studie Statistiek had ik niet te klagen over de begeleiding van de docenten. Een groot voordeel van een klein clubje.

Na vier jaar had ik genoeg van de collegebanken. Op het Centraal Bureau voor de Statistiek heb ik zeven maanden stage gelopen en een scriptie geschreven over statistische beveiliging van bestanden. Het doel van statistische beveiliging is om de kans op statistische onthulling klein te houden. Het kan voorkomen dat een persoon in een bestand wordt geïdentificeerd met behulp van een aantal variabelen in dit bestand. Een gebruiker van het bestand weet bijvoorbeeld dat er maar één vrouwelijke dierenarts in Lutjebroek is. Andere variabelen in het bestand worden dan bekend bij deze gebruiker en dan is er sprake van statistische onthulling.

Na mijn studie heb ik een aantal malen gesolliciteerd, maar de arbeidsmarkt was iets minder rooskleurig dan nu. Ik had tevens het gevoel dat ik met mijn 22 levensjaren nog niet toe was aan een vaste baan in het bedrijfsleven. Op een gegeven ogenblik werd ik gewezen op een vacature voor een onderzoeker in opleiding. Het onderzoeksvoorstel had betrekking op de robuustheid van schattingsmethoden voor LISREL-modellen. Robuustheid is een begrip dat mij tijdens de studie al aansprak (met dank aan Bert van Es). Verder moest er voor het bepalen van de robuustheid veel gesimuleerd worden. Ik had al tweemaal een uitgebreide simulatiestudie gedaan en vond dat zeer interessant.

Van LISREL-modellen had ik echter nog nooit gehoord. Met zulke modellen kunnen gerichte relaties tussen latente variabelen worden

onderzocht. Latente variabelen zijn bijvoorbeeld intelligentie en geluk. Deze variabelen kunnen alleen indirect gemeten worden met behulp van een vragenlijst. Met behulp van een LISREL-model kan dus worden onderzocht of intelligentie een positief effect heeft op de mate van geluk. Het onderzoek vond plaats op het Instituut voor Psychologisch Onderzoek in Groningen. Psychologie was een hobby van mij dus ik zag geen probleem.

De vrijheid die je als oio krijgt vond ik fantastisch net als de congressen in het buitenland. Na viereneenhalf jaar was mijn proefschrift klaar voor de drukker en besloot ik om bij het CBS te solliciteren. Sinds een paar jaar hebben ze daar een traineesysteem dat mij zeer aanspreekt. Tijdens een periode van twee jaar kunnen trainees verschillende projecten doen en zijn er tevens allerlei cursussen. Het is een gezellige club jongeren en buiten het werk doen we ook van alles.

Het eerste project wat ik gedaan heb op het CBS betrof een internationaal project van Eurostat, het Europese statistische bureau. Het doel van dit project was het ontwikkelen en evalueren van software voor imputatiedoeleinden. Als een persoon bepaalde vragen niet wil of kan beantwoorden ontbreken er waarden in het bestand. Het inschatten van deze non-response heet imputeren. Omdat het een internationaal project was en er daarom in vijf verschillende landen een pakket moest worden geëvalueerd danwel ontwikkeld was het zeer moeilijk om tot algemene conclusies te komen. Geheel tegen mijn zin in had elk land bijvoorbeeld een andere dataset gebruikt om mee te simuleren.

Momenteel ben ik bezig met het in kaart brengen van het statistische proces van de maandstatistiek detailhandel. Er moet tevens aangegeven worden op welke punten dit proces verbeterd kan worden. Het aandachtspunt van deze statistiek is de maandontwikkelingen van de omzet per bedrijfspgroep. Het proces is complex en slechts deels gedocumenteerd, waardoor ik vaak interviews moet afnemen. Een berichtgever hoeft bijvoorbeeld niet per maand de omzet van zijn bedrijf op te geven, maar kan dit ook per vier weken doen. Met behulp van een kooldagenpatroon moeten de 4-weekomzetten dan worden omgerekend naar maandomzetten. Verder zijn

er verschillende mogelijkheden voor het inschatten van de omzet van niet-responderende en/of niet-actieve bedrijven, bedrijven met een afwijkende omzet en bedrijven die in de verkeerde bedrijfs- of grootteklasse zitten. De uiteindelijk beschrijving van dit proces zal ongeveer 30 bladzijden bedragen. Er is veel belang bij een goede beschrijving omdat op korte termijn een aantal CBS-statistieken moet worden (her)ontworpen.

Op dit moment ben ik net terug van de Costa del Sol en dan valt het niet mee om de hele dag op kantoor te zitten. Het CBS bevalt echter prima dus ik zal hier de komende jaren zeker blijven werken.

Jeffrey Hoogland
jhgd@cbs.nl