

faculteit

psychologie

Biometric decomposition of
phenotypic means in
human samples

Conor V. Dolan

Samenvatting.

Quantitatief genetisch onderzoek naar polygenetische menselijke eigenschappen heeft tot doel de bijdragen van genetische en omgevingsvariabelen tot individuele fenotypische verschillen, oftewel binnen groep variatie, te schatten. De genetische bijdrage is toe te schrijven aan verschillen tussen individuen in hun genetische constitutie of genotype. De omgevingsbijdrage is toe te schrijven aan verschillen in de mate waarin individuen blootstaan aan relevante omgevingsinvloeden. Met behulp van een aantal goed ontwikkelde onderzoeksopzetten is het mogelijk om de bijdragen van deze bronnen tot de totale variatie van een gegeven fenotype als variantie componenten te schatten. De identificatie van deze variantie componenten wordt ontleend aan theoretische kennis t.a.v. het percentage door familieleden gedeelde allelen en aan kennis omtrent (de geschiedenis van) hun cohabitatie.

Variatie tussen groepen wordt uitgedrukt als de variantie van de groepsgemiddelden en is evenals variatie binnen groepen toe te schrijven aan genetische en omgevingsinvloeden. In tegenstelling tot variatie binnen groepen, echter, kan de variantie van groepsgemiddelden niet zonder meer worden uiteengelegd in dergelijke componenten. Deze decompositie vereist specifieke kennis t.a.v. de verschillen tussen de onderscheiden groepen in de omgevingsinvloeden en kennis van frequenties van de relevante allelen in de groepen. Doorgaans ontbreekt deze kennis wanneer het continue verdeelde polygenetische eigenschappen betreft, zoals IQ of gewicht.

Gegeven de afwezigheid van dergelijke informatie, kan men door het maken van bepaalde aannamen toch komen tot een decompositie van de variantie van gemiddelden. Door aan te nemen dat de bronnen van variatie binnen groepen en tussen groepen dezelfde zijn, kan deze decompositie plaats vinden. Het doel van deze dissertatie is het onderzoeken van modellen die op basis van de genoemde aanname gebruikt kunnen worden om deze decompositie uit te voeren.

Uitgangspunt voor deze modellen is het idee dat de groepen tot stand komen door een vorm van selectie. Men denke, bijvoorbeeld, aan groepen die ontstaan zijn door selectie uit een oorspronkelijke populatie. Twee vormen van selectie worden onderscheiden: selectie op grond van de geobserveerde variabelen (de fenotypen) en selectie op grond van latente of niet-geobserveerde variabelen. Deze selectie processen leiden tot toetsbare modellen waarin de variantie tussen en binnen groepen gemodelleerd worden met eenzelfde set van parameters. De hierboven genoemde aanname komt zodoende in deze modellen op een exacte manier tot uitdrukking.

Zoals is uitgelegd in hoofdstuk 1, heeft het werk van Meredith (1964), Jöreskog (1971) en Sörbom (1974) op het gebied van latente selectie in het factor model geleid tot multi-groep covariante structuur modellen. Uit dit werk komt een belangrijk aspect van latente selectie naar voren: afgezien van het feit dat het selectie proces inderdaad op latent niveau plaats vindt, is verder kennis niet vereist. In hoofdstuk 3 is dit model uitgewerkt en toegepast op tweeling data. Deze onderzochte variabelen hebben betrekking op bloeddruk.

In hoofdstuk 4 wordt een op latente selectie gebaseerd longitudinaal model ontwikkeld en toegepast op herhaald gemeten gewicht in een steekproef van een- en tweeeïge tweelingen. Rekening houdend met de seriële afhankelijkheid van de data, wordt het ontwikkelingsproces dat ten grondslag ligt aan de stabiliteit van individuele verschillen en het verloop van het gemiddelde beschreven in termen van veranderende genetische en omgevingsinvloeden. Ten aanzien van selectie, worden de observaties op iedere tijdstip opgevat als voortvloeiend uit een proces van latente selectie van deze invloeden.

Selectie op grond van het fenotype leidt, evenals latente selectie tot een model waarin individuele verschillen en groepsverschillen uitgedrukt worden met dezelfde parameters. Zoals uitgelegd in hoofdstuk 1, is het probleem van fenotypische selectie in het factor model onderzocht door Thomson (1945) en recentelijk door Muthén (1989a). In hoofdstuk 5, is deze vorm van selectie toegepast op een multivariate dataset bestaande uit een aantal school variabelen gemeten in een steekproef tweelingen. In tegenstelling tot latente selectie, vereist phenotypische selectie specifieke kennis t.a.v. de selectie variabele.

De in deze dissertatie gekozen benadering heeft geleid tot toetsbare modellen voor de bronnen van variantie binnen en tussen groepen. De cruciale aanname die hierbij gemaakt wordt is moeilijk te bewijzen m.b.v. deze modellen. Echter aan de hand van verschillende passingsmaten, die aangeven in hoeverre deze modellen een adequate beschrijving geven van de data, is het mogelijk om een uitspraak te doen over de aannemelijkheid van de aanname. Afgezien van de specifieke modellen, die eruit voortvloeien, scheidt de notie van selectie een nuttig kader voor een geïntegreerde interpretatie van aan individuele en groepsverschillen gerelateerde toetsingsgrootheden.

Aangezien de modellen die hierin beschreven zijn covariantie structuur modellen zijn, is in het kader van dit onderzoeksproject, een computer programma geschreven om deze modellen te passen. Dit programma heeft twee voordelen boven commerciële programma's: het biedt de mogelijkheid om aan parameters een verscheiden aan restricties op te leggen en het kan problemen aan die te groot zijn voor deze programma's. De ontwikkeling van dit programma heeft geleid tot twee hoofdstukken die opgenomen zijn als de appendices in deze dissertatie. Hoofdstuk A1 bevat een beschrijving van het programma en enige toepassingen. In hoofdstuk A2 wordt een vergelijking gemaakt van een aantal methoden om standaard fouten van maximum likelihood schattingen te berekenen. De standaard methode is gebaseerd op Fischer's informatie matrix. In het genoemde programma zijn de standaard fouten gebaseerd op een benadering van de Hessiaan. Beide methoden geven bevredigende resultaten.

Tenslotte is hoofdstuk A3 gewijd aan een vergelijking van verschillende methoden om factor scores te berekenen in longitudinale covariantie structuur modellen. Hierbij is de binnen factor analyse ontwikkelde regressie methode vergeleken met de binnen 'state-space modelling' ontwikkelde Kalman filter (KF) en Fixed Interval Smoother (FIS). De aan de hand van een eerste orde autoregressief model uitgevoerde vergelijking wijst uit dat de FIS en de regressie methoden identiek zijn. De KF maakt gebruik van minder informatie en levert derhalve factorscores op met een grotere standaard fout.