

## University of Groningen

### Specification methods for the multilevel model

Berkhof, Johannes

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2000

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Berkhof, J. (2000). *Specification methods for the multilevel model*. [Thesis fully internal (DIV), University of Groningen]. DSWO Press.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

33, 84, 87

37, 39, 48, 55

Approximation, 32, 39-42

7, 36, 40

2, 56

del, 59-82

3-17, 42

ct, 39-40

0, 36, 44, 53

er, 64, 68, 74, 88, 95

ee cubic smoothing

7, 44, 45, 48, 61, 94

ct, 60, 72, 78, 86

1, 34, 42, 74

, 56, 61

, 1, 10, 11, 33, 77

testing, 31-58

ct, 8, 10, 12-17, 19-8

ponent test, see OLS

tion, 2

ctor, 54

## Samenvatting

In veel gegevens zit een hiërarchische of geneste structuur. Denk bijvoorbeeld aan gegevens van scholieren verzameld op meerdere scholen; hier kunnen we twee niveau's (student-school) onderscheiden of zelfs drie (student-school-regio). Vanuit inhoudelijk perspectief is deze hiërarchische structuur interessant omdat de prestatie van de student niet alleen afhangt van kenmerken van de student zelf, maar ook mede bepaald kan worden door de school waar hij of zij op zit. Voor de analyse van geneste gegevens ligt het dan ook voor de hand om een multiniveaumodel te gebruiken dat expliciet rekening houdt met de geneste structuur. Het specificeren van het multiniveaumodel wordt door gebruikers vaak als lastig ervaren. Om tot een zorgvuldig gespecificeerd model te komen is het daarom van belang om over ondersteunende specificatiemethoden te beschikken. Hiermee worden methoden of technieken bedoeld die misspecificaties detecteren en indicaties geven voor modelmodificaties. In dit proefschrift worden een aantal specificatiemethoden voorgesteld als aanvulling op reeds bestaande methoden.

In hoofdstuk 1 van het proefschrift wordt het multiniveaumodel beschreven door een vergelijking te trekken met klassieke binnengroeps- en tussengroepsregressiemodellen. De laatste modellen zijn nauw gerelateerd aan het multiniveaumodel en houden ook rekening met de multiniveauctuur van de gegevens. Een belangrijk verschil tussen het multiniveaumodel en de binnengroeps- en tussengroepsregressiemodellen is echter dat in het multiniveaumodel wordt verondersteld dat de groepen uit één onderliggende populatieverdeling komen. De verschillen tussen de groepen worden beschreven door een verzameling van toevalsvariabelen die meestal worden aangeduid als random effecten of random coëfficiënten. In hoofdstuk 1 worden verder argumenten genoemd om de keuze van een multiniveaumodel te motiveren en er wordt gewezen op de praktische moeilijkheid van het specificeren van zo'n model in de praktijk.

Een belangrijk onderdeel van het specificeren van een multiniveaumodel is het kiezen van de random coëfficiënten. Omdat de uiteindelijke keuze (sterk) van invloed kan zijn op de conclusies van een multiniveauanalyse is het verstandig om de schattingen voor verschillende multiniveaumodellen met elkaar te vergelijken. Voortbordurend op deze aanbeveling worden in hoofdstuk 2 wiskundige formules afgeleid die uitdrukken hoe de schatters van de parameters van het multiniveaumodel veranderen als een random coëfficiënt onterecht uit het model

wordt weggelaten. Ook worden formules afgeleid die uitdrukken hoe de geschatte standaardfouten van de vaste effecten dan veranderen. Deze formules kunnen van dienst zijn wanneer men een multiniveaumodel specificiert. Verder kunnen ze nuttig zijn bij het bestuderen van gerapporteerde resultaten van eerder gedaan onderzoek. Hoewel men bij het lezen van een rapport of artikel vaak niet direct kan beschikken over de ruwe gegevens, zou men toch graag direct inzicht willen krijgen in de mogelijke consequenties van het onterecht weglaten van een random coëfficiënt.

Een aantal situaties worden in hoofdstuk 2 in detail besproken, te weten (1) het weglaten van een random intercept in random-interceptmodel met twee niveau's, (2) het weglaten van een random helling in een tweenniveaumodel met random intercept en random helling, (3) het weglaten van een random intercept op niveau twee of drie in een drieniveaumodel, en (4) het weglaten van een random effect in een kruisklassificatiemodel. De formules in hoofdstuk 2 zijn gebaseerd op momentenschatters van de modelparameters. Omdat in de praktijk meestal de meest aannemelijke schatters worden gebruikt, wordt in hoofdstuk 2 ook nagegaan of de afgeleide formules een nauwkeurige benadering geven voor de verandering in de waarden van de meest aannemelijke schatters. Uit een analyse van examengegevens bleek dat (1) de met de afgeleide formules berekende veranderingen in de schattingen van modelparameters soms verschillen van de veranderingen in de meest aannemelijke schatters en (2) de met de afgeleide formules berekende veranderingen in de geschatte standaardfouten van de vaste effecten goed overeenkomen met de veranderingen in de geschatte standaardfouten van de meest aannemelijke schatters.

De beslissing om een random coëfficiënt aan het model toe te voegen wordt gewoonlijk bepaald door inhoudelijke argumenten en toetsingsresultaten. Met betrekking tot het laatste kan men kiezen uit een aantal bestaande statistische toetsen. De meest gangbare zijn de aannemelijkheidsratio-toets, de (asymptotische) Wald toets en een exacte toets gebaseerd op geschatte kleinste kwadratenhellingen per groep (Wald's variantiecomponententoets). De voor- en nadelen van deze toetsen worden uitvoerig besproken in hoofdstuk 3. Belangrijke conclusies zijn dat de asymptotische Wald toets (zeer) conservatief is en dat Wald's variantiecomponententoets beperkt toepasbaar is. Als aanvulling op de bestaande toetsen worden drie algemeen toepasbare scoretoetsen voorgesteld. Deze scoretoetsen zijn niet exact en verschillen met betrekking tot de manier waarop de nulverdeling wordt benaderd. Uit een Monte Carlo studie (waarin

de significantie van een random helling in een tweenniveau model met random intercept en random helling werd getoetst) bleek dat scoretoetsen minder conservatief zijn dan de aannemelijkheidsratioets. Verder bleken de scoretoetsen een minstens even hoog en dikwijls hoger onderscheidingsvermogen te hebben dan Wald's variantiecomponenttoets. Het onderscheidingsvermogen van de scoretoetsen en de aannemelijkheidsratioets was moeilijk te vergelijken aangezien het onderscheidingsvermogen van de aannemelijkheidsratioets (in tegenstelling tot dat van de scoretoetsen) sterk afhangt van de correlatie tussen het random intercept and de random helling.

In hoofdstuk 4 en 5 wordt een bepaald type multiniveau model onder de loep genomen, namelijk het herhaalde metingen- of longitudinale model. Dit model is ontworpen om longitudinale gegevens van een groep subjecten te analyseren. Drie doelstellingen van een herhaalde metingenanalyse zijn (1) het schatten van het algemeen gemiddelde van de responsevariabele over de tijd, (2) het schatten van de tijdsafhankelijke effecten van een verzameling van predictoren op de responsvariabele (net zoals het algemeen gemiddelde zijn deze tijdsafhankelijke effecten functies van de tijd), en (3) het schatten van de residuele individuele verschillen die door de in het model opgenomen covariaten worden beschreven (deze individuele verschillen zijn wederom functies van de tijd).

Het specificeren van het herhaalde metingenmodel begint vaak met het kiezen van een familie van functies waarvan de functies genoemd in (1), (2) en (3) speciale gevallen zijn. Soms is er theoretische kennis aanwezig over de familie van functies (bv. logaritmische groei). Indien dit niet het geval is kiest men meestal voor polynomen. Helaas kunnen sommige processen niet accuraat worden beschreven door polynomen van een lage graad (bv. een spurt in lichaamslengte tijdens de puberteit). Een mogelijkheid is dan om de graad van de polynoom op te voeren maar dit kan ernstig ten koste van de gladheid van de functies gaan.

In dit proefschrift wordt voorgesteld om in plaats van polynomen kubische "smoothing splines" te gebruiken. De vorm van deze spline functies wordt niet gereïcticeerd door te eisen dat ze tot een bepaalde parametrische familie van functies moeten behoren. Gladde functies worden verkregen door te middelen over (in de tijd) dichtbij elkaar gelegen waarnemingen. Spline functies zijn daarom soms beter dan polynomen in staat om bepaalde groeiprocessen accuraat te schatten. In hoofdstuk 4 worden alleen splines geformuleerd voor (1) het algemeen gemiddelde en (2) de tijdsafhankelijke effecten van predictoren. De

residuele individuele verschillen (3) worden nog steeds beschreven met polynomiale functies. Deze aanpak lijkt zinvol als alleen het algemeen gemiddelde en de tijdsafhankelijke effecten het onderwerp van studie zijn. In hoofdstuk 4 wordt voorgesteld om het model met "smoothing splines" te schatten met het iteratieve ECM-algoritme (Meng en Rubin, 1993). Dit levert gepenalizeerde meest aannemelijke schatters voor de spline functies. Een voordeel van het ECM-algoritme in vergelijking met andere algoritmes zoals Fisher scoring is dat het schatten van het model ook goed mogelijk als het aantal tijdspunten zeer hoog is. Voor het bepalen van de gladheid van de functies kunnen verschillende criteria worden gebruikt. Drie criteria worden vergeleken, te weten de kruisklassificatie aannemelijkheidsfunctie, Mallow's  $C_p$ , en Akaike's informatie criterium. Uit een van de voorbeelden blijkt dat ze ongeveer even gladde functies opleveren.

In hoofdstuk 5 wordt een model voorgesteld met "smoothing splines" voor het algemeen gemiddelde, de tijdsafhankelijke effecten van predictor en de residuele individuele verschillen. Met name in groeicurve studies zijn de laatstgenoemde verschillen vaak van groot inhoudelijk belang. Een exploratief model met "smoothing splines" in het random gedeelte van het model kan daarom zinvol zijn. Hoewel het model gebruikt kan worden voor voorspellingsdoeleinden, wordt in hoofdstuk 5 het exploratieve karakter van het model benadrukt. Zo wordt in een voorbeeld getoond dat het model kan worden gebruikt om uitbijters op te sporen. De spline functies worden geschat met een iteratief algoritme waarvan de convergentie wordt bewezen in het geval er geen missende gegevens zijn. Aan het eind van het hoofdstuk worden de resultaten van een Monte-Carlo simulatiestudie gepresenteerd waaruit blijkt dat het "smoothing splines" model goed in staat is onderliggende logaritmische en trigonometrische functies terug te schatten.