POK Memorandum 2009-1

Over het gebruik van continue normering

Timo Bechger
Bas Hemker
Gunter Maris

POK Memorandum 2009-1

Over het gebruik van continue normering

Timo Bechger
Bas Hemker
Gunter Maris

Cito
Postbus 1034
6801 MG Arnhem
Kenniscentrum

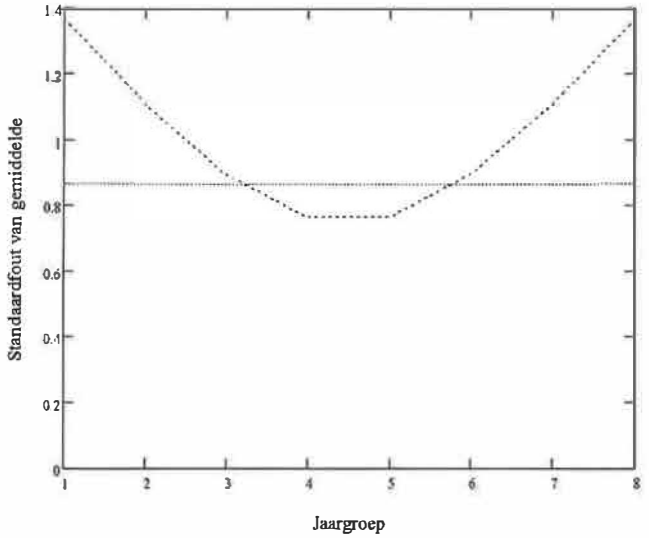
Cito
Arnhem, maart 2009

Introductie

Continue normering is een regressie methode die ontwikkeld is door Zachary and Gorsuch (1985) in verband met het normeren van de *Wechsler Adult Inteligence Scale R (WAIS-R)*. In de klassieke normering wordt de normverdeling in elke leeftijdsgroep apart geschat. In de praktijk is het aantal leeftijdsgroepen beperkt waardoor individuen die vrijwel even oud zijn toch in andere leeftijdsgroepen worden geplaatst.[1] Hierdoor vertonen verdelingen van IQ scores vaak "rare sprongen". Bij continue normering wordt de verdeling van de ruwe scores geschat als functie van leeftijd om te vermijden dat de vergelijking van iemand ruwe score met zijn of haar normgroep wordt beïnvloed wordt door het beperkte aantal leeftijdsgroepen. Daarnaast wordt bij continue normering gebruik gemaakt van de gegevens van meerdere leeftijdsgroepen tegelijkertijd waardoor de schattingen aan efficiëntie winnen.

Hoewel continue normering efficiëntere schatters oplevert dan klassieke normering rechtvaardigt continue normering niet het gebruik van zeer kleine steekproeven (zie ook Ravenzwaaij & Hamel, 2006). In de voorlopige versie van de nieuwe normen van de *COTAN* wordt gesteld dat de normen voor 8 opeenvolgende normgroepen met in iedere groep 50 waarnemingen in het geval van een continue normering in precisie equivalent is aan die van een klassieke normering met 300 waarnemingen per normgroep. Dit zou impliceren dat de efficiëntie van continue normering zes keer groter is dan die van klassieke normering en dat continue normering efficiëntere schatters oplevert voor alle leeftijdsgroepen. Dit is een opmerkelijke stelling waarvoor verder geen wetenschappelijke evidentie is gegeven. Verwacht mag worden dat op grond van deze stelling onderzoekers massaal zullen kiezen voor deze vorm van normering

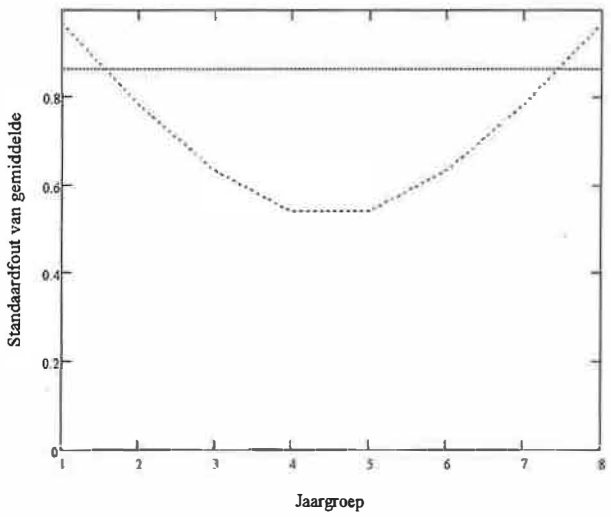
[1] In de engels-talige literatuur staat dit bekend als het *edge of cohort problem.*



Figuur 1. Standaardfout bij het schatten van het gemiddelde. 50 Observaties in elke groep voor continue normering en 300 voor klassieke normering.

aangezien dit de kosten van normering aanzienlijk verlaagt. Het gevaar is echter dat de gestelde efficiëntie niet gehaald wordt, hetgeen tot gevolg heeft dat de normen onnauwkeuriger worden en het percentage foute classificaties toeneemt.

In deze notitie zal middels een voorbeeld aangetoond worden dat zelfs onder de meest gunstige omstandigheden 50 waarnemingen niet voldoende is. De conclusie is dat verder onderzoek noodzakelijk is alvorens tot gefundeerde richtlijnen te komen voor het aantal benodigde observaties voor continue normering.



Figuur 2. Standaardfout bij het schatten van het gemiddelde. 100 Observaties in elke groep voor continue normering en 300 voor klassieke normering.

Een Rekenvoorbeeld

We concentreren ons op het schatten van de gemiddelden in de normgroepen die doorgaans beter geschat kunnen worden dan de percentielen of hogere momenten zoals de varianties. We vergelijken twee situaties:

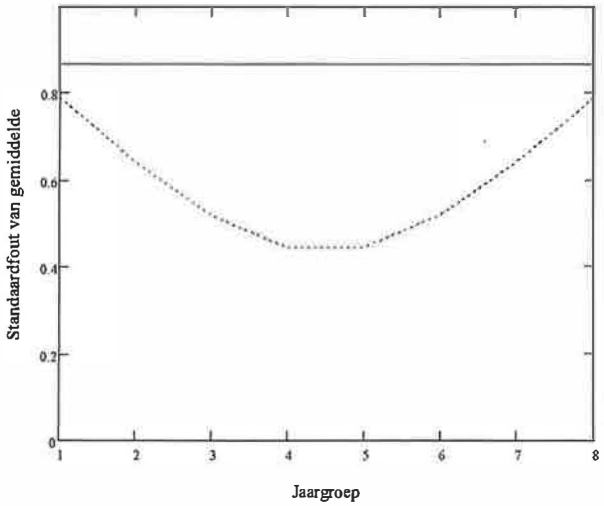
1. *Klassieke normering*: We schatten het gemiddelde in een steekproef van $n_j = 300$ waarnemingen uit een norm-populatie van kinderen in jaargroep j.

2. *Continue normering*: We schatten de lineaire regressie

$$\mu(j) = \alpha + \beta j \tag{1}$$

van score op leeftijd met n_j waarnemingen per jaargroep j.

We vergelijken beide procedures door te kijken naar de standaardfout van het geschatte gemiddelde in een groep. We beschouwen de standaardfout



Figuur 3. Standaardfout bij het schatten van het gemiddelde. 150 Observaties in elke groep voor continue normering en 300 voor klassieke normering.

van continue normering bij steekproeven van, respectievelijk $n_j = 50$, $n_j = 100$, en $n_j = 150$ in elke groep, in relatie tot de standaardfout bij klassieke normering bij een steekproef van 300. De resultaten staan samengevat in Figuur 1 en Figuur 3.

In het eerste scenario is de schattingsfout van het gemiddelde gegeven door de klassieke formule

$$\frac{\sigma^2}{n_j} \tag{2}$$

waarbij σ^2 het symbool is voor de variantie in de normpopulatie. We zullen daarbij, zoals in de praktijk gebruikelijk is, aannemen dat de variantie van de score in elke leeftijdsgroep gelijk is. Onder deze aanname hangt de *standaardfout* (de wortel van de variantie) alleen af van n_j. In Figuur 1 en Figuur 3 is de horizontale lijn de standaardfout bij klassieke normering. De lijn is horizontaal omdat de steekproeven even groot

zijn in alle jaargroepen. Om de getallen op de vertikale as te kunnen interpreteren krijgt de standaarddeviatie σ de waarde 15 zoals bij IQ testen gebruikelijk is. De standaardfout van het gemiddelde bij klassieke normering is dus

$$\sqrt{\frac{15^2}{300}} = 0.866 \tag{3}$$

In het tweede scenario is het geschatte gemiddelde de som van $\hat{\alpha}$ en $\hat{\beta}_j$, waarbij de hoedjes aangeven dat deze parameters geschat zijn. Gebruikmaken van welbekende regels voor varianties, volgt dat de schattingsfout van het gemiddelde gelijk is aan

$$Var(\hat{\alpha}) + Var(\hat{\beta})j^2 + 2jCov(\hat{\alpha}, \hat{\beta}) \tag{4}$$

Hierbij is:

$$\begin{aligned} Var(\hat{\alpha}) &= \frac{\sigma^2 \sum_j j^2 n_j}{SS \sum_j n_j} \\ Var(\hat{\beta}) &= \frac{\sigma^2}{SS} \\ Cov(\hat{\alpha}, \hat{\beta}) &= \frac{-\sigma^2 M}{SS} \end{aligned}$$

Hierbij is M gelijk aan de gemiddelde jaargroep en SS aan de *Sum of Squares*:

$$\sum_j (j - M)^2 n_j \tag{5}$$

Deze formules zijn in algemene statistiekboeken te vinden: zie bijvoorbeeld bladzijde 489 van Mood, Graybill, en Moes (1974). Voor het rekenvoorbeeld kiezen we voor een uniforme verdeling van leerlingen over de jaargroepen zodat $M = 4.5$. Merk op dat bij continue normering de standaardfout van het geschatte gemiddelde *niet* constant is over jaargroepen.

In Figuur 1 zien we dat, bij een steekproefgrootte van $n_j = 50$, de standaardfout bij continue normering niet uniform beter is dan die van

klassieke normering. Wanneer het aantal observaties groter wordt zal continue normering het winnen van gewone normering. Figuur 3 laat zien dat 150 observaties genoeg is onder de aannamen die we in dit voorbeeld hebben gemaakt.

Conclusie

Het rekenvoorbeeld laat zien dat, bij 50 waarnemingen per groep, de meetprecisie bij continue normering in meer dan de helft van de jaargroepen kleiner is dan die van klassieke normering met 300 waarnemingen per groep. De conclusie is dat 50 waarnemingen te weinig is. In het rekenvoorbeeld blijkt 150 observaties per groep genoeg te zijn. Dit geldt echter onder de aannamen die we in dit voorbeeld hebben gemaakt.

De efficiëntie bij continue normering wordt verkregen door aannamen te maken. We hebben informatie "gekocht"door middel van aannamen. In het rekenvoorbeeld zijn we uitgegaan van een ideale situatie. We hebben aangenomen dat de regressie van toetsscore op leeftijd lineair is. Dit is in veel gevallen onrealistisch en in de praktijk zal men doorgaans kiezen voor een hogere orde polynoom als benadering. De standaardfout van het gemiddelde zal hierdoor groter worden waardoor de situatie verslechtert ten opzichte van het rekenvoorbeeld. Verder hebben we aangenomen dat binnen iedere normgroep een normale verdeling geldt, waarbij de verdelingen over de normgroepen alleen van elkaar verschillen in gemiddelde en variantie. Merk op dat de steekproeven in elke leeftijdsgroep representatief moeten zijn.

De aannames die gemaakt worden dienen getoetst te worden. Indien namelijk niet aan de aannamen is voldaan kan continue normering vertekende resultaten geven. Dit betekent dat bij continue normering de steekproef niet alleen groot genoeg moet zijn de minimale meetpreci-

sie te halen zoals die voor klassieke normering is vereist. De steekproef moet tevens groot genoeg zijn om de aannamen te kunnen toetsen met voldoende statistische power. In hoeverre continue normering gevoelig is voor schending van welbepaalde aannamen zal met behulp van simulatie moeten worden onderzocht alvorens kan worden aangenomen dat continue normering onder alle omstandigheden te prefereren is boven klassieke normering.

Tenslotte willen we opmerken dat er verschillende varianten zijn van continue normering (zie bijvoorbeeld Breukelen & Vlaeyen, 2005) waarvan we er in dit stuk slechts één hebben beschouwd. Deze technieken zijn recent ontwikkeld en slechts op beperkte schaal gebruikt. Het verdient aanbeveling om terughoudendheid te betrachten bij het uitgeven van strikte richtlijnen met betrekking tot het gebruik van de verschillende varianten van continue normering totdat er meer onderzoek is gedaan.

References

Breukelen, G. J. P. van, & Vlaeyen, J. W. S. (2005). Norming clinical questionnaires with multiple regression: The pain recognition list. *Psychological Assessment*, *17*, 336-344.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (third ed.). Singapore: McGraw-Hill.

Ravenzwaaij, D. van, & Hamel, R. (2006). De Nederlandstalige WAIS-II na hernormering. *De Psycholoog*, 268-271.

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*, 86-94.