

In dit artikel wordt de gemiddelde lichaamsgroei van Nederlandse jongens en meisjes beschreven met een logistisch en bilogistisch groeimodel. Hierbij gaat **André Heck** expliciet in op de wijze waarop de parameters in de wiskundige modellen kunnen worden geschat en hoe ICT dit werk voor vwo-leerlingen mogelijk maakt.

Een gewichtig model

Inleiding

In de CWI vakantiecursus *Wiskunde en Gezondheid* wordt in het artikel ‘Gewichtige wiskunde in de klas’ (Heck, 2002) uitgebreid ingegaan op modellen voor gemiddelde lengtegroei van jongens en meisjes van Nederlandse afkomst. We kijken nu voor de verandering eens naar de gemiddelde groei in gewicht van Nederlandse kinderen. De groeicijfers komen uit de meest recente Nederlandse groeistudie (Fredriks et al, 2002). Opnieuw wordt het groeimodel opgebouwd uit verschillende componenten, bijvoorbeeld:

$$\text{gewicht} = \text{ondergrens} + \text{gewicht}_{\text{kindertijd}} + \text{gewicht}_{\text{puberteit}}$$

waarbij $\text{gewicht}_{\text{kindertijd}}$ respectievelijk $\text{gewicht}_{\text{puberteit}}$ de bijdrage aan het gewicht is die kan worden toegeschreven aan de kindertijd, respectievelijk puberteit. De *ondergrens* is het gemiddelde gewicht dat aan het eind van de kleutertijd bereikt is. In het zogenaamde bilogistische groeimodel gebruiken we het logistische model voor de twee genoemde bijdragen. Voor de puberteit krijgen we dan de formule:

$$\text{gewicht}_{\text{puberteit}}(t) = \frac{c}{1 + ae^{-bt}}$$

Hierin duidt t de leeftijd aan, en zijn a , b en c parameters waarvan de meest geschikte waarden moeten worden geschat. Het rechterlid is de formule voor logistische groei zoals die in de meeste Nederlandse schoolboeken vermeld staat en in veel rekenmachines gebruikt wordt. Als we ons beperken tot het gewicht tijdens de adolescentie, dan blijkt de gemiddelde groei in gewicht wiskundig aardig te worden beschreven door een gemodificeerde logistische kromme, ook wel sigmoïde of S-kromme genoemd. Nu is de formule:

$$G(t) = \frac{c}{1 + ae^{-bt}} + d$$

met $G(t)$ het gewicht op leeftijd t . Het enige verschil met de eerdere formule is de parameter d , die een ondergrens

ongelijk 0 mogelijk maakt. Wij geven er de voorkeur aan de formule voor gewicht als functie van leeftijd in een andere vorm op te schrijven:

$$G(t) = \frac{c}{1 + e^{\frac{b}{a-t}}} + d$$

De parameter a heeft dan namelijk de betekenis van het tijdstip waarop de gewichtstoename maximaal is. Op dat moment is de groei halverwege en is de groeisnelheid gelijk aan $\frac{bc}{4}$. In een diagram waarin de groeisnelheid uitgezet is tegen de leeftijd is de waarde van a gemakkelijk af te lezen. In het diagram van gewicht naar leeftijd is het gewicht aan het einde van de adolescentie goed te schatten. Dit eindgewicht minus het gewicht op leeftijd a is gelijk aan $\frac{c}{2}$. De grafiek van de S-kromme is punt-symmetrisch in $(a, G(a))$. Hieruit volgt dat we parameter d kunnen schatten als $G(a) - \frac{c}{2}$.

Wat rest is het schatten van parameter b ; hiervoor kan het gewicht op een gegeven leeftijd of de groeisnelheid op leeftijd a worden gebruikt.

In dit artikel zullen we een robuustere wijze van parameterschatting in een sigmoïdaal regressiemodel bespreken, en ingaan op de rol die ICT hierin speelt.

Het logistische groeimodel in het schoolboek

Het is een goed idee om even stil te staan bij enkele zaken die een vwo-leerling over het logistische groeimodel in het schoolboek leert. Zij maken kennis met logistische groei bij het onderwerp differentievergelijkingen via de formule:

$$\Delta u(t) = c \cdot u(t) \cdot \left(\frac{M - u(t)}{M} \right)$$

met groeivoet c en maximum M (ook wel grenswaarde, verzadigingsniveau of capaciteit genoemd). Kenmerk van logistische groei is dat deze in het begin lijkt op exponentiële groei en aan het einde lijkt op geremde groei volgens een gemengd lineair-exponentieel groeimodel. Bij gegeven groeivoet c , het maximum M en een startwaarde $u(t_0)$ kan de formule worden gebruikt om alle volgende

waarden $u(t)$ uit te rekenen. In toepassingsvoorbeelden in het lesboek is het maximum M meestal goed herkenbaar in grafieken. Maar wat te doen als uit de gegevens niet onmiddellijk de grenswaarde M valt af te lezen? Hoe moeten de meest geschikte waarden voor de groeiparameters dan worden gevonden? Leerlingen wordt de indruk gewekt dat dit dan via een trial-and-error methode en enkel met het blote oog gebeurt. De methode om de toename van de grootheid grafisch uit te zetten tegen de grootheid zelf, en dan de best bijpassende parabool (met de grafische rekenmachine) op te sporen, blijft jammer genoeg onbesproken. Leerlingen kunnen logistische groei ook tegenkomen bij het onderwerp differentiaalvergelijkingen, of bij continue dynamische modellen. De formule:

$$y = \frac{c}{1 + ae^{-bt}}$$

is de oplossing van de differentiaalvergelijking:

$$\frac{dy}{dt} = b \cdot y \cdot \left(\frac{c-y}{c} \right)$$

met:

$$y(0) = \frac{c}{1+a}$$

Als belangrijkste kenmerk wordt vermeld dat de grafiek de lijn $y = c$ als horizontale asymptoot heeft. In rekenopdrachten wordt nog wel gevraagd om uit te rekenen op welk moment de grootste toename plaatsheeft, maar onvermeld blijft dat op dat moment de groei altijd halverwege is. De symmetrie van de logistische groeikromme blijft onbesproken. Ook zijn er geen opdrachten waarin nagegaan wordt in hoeverre de volgende uitspraak klopt: ‘Het kenmerk van logistische groei is dat deze aanvankelijk exponentieel toeneemt en aan het eind van het groei-proces exponentieel afneemt.’

Voor toepassingen van het logistische model is het vervelend dat er weinig of niets wordt gezegd over hoe geschikte waarden voor de parameters kunnen worden gevonden. Kennelijk gaat men uit van de rekenkracht van wiskundige software of van een grafische rekenmachine, zoals de TI-83, die logistische regressie mogelijk maakt. Maar succes van niet-lineaire regressie hangt vaak af van het al dan niet kunnen vinden van goede initiële benaderingen voor de parameters.

Automatische logistische regressie

In de meeste grafische rekenmachines is een functie voor logistische regressie beschikbaar die zonder veel hulp van de gebruiker kan worden toegepast. Laten we eens nagaan hoe dit gebeurt in de TI-83. In de handleiding (TI-83, 1996) is niet meer te lezen dan dat de functie LOGISTIC de regressievergelijking:

$$y = \frac{c}{1 + ae^{-bx}}$$

voor de gegevens berekent aan de hand van een iteratieve methode van kleinste kwadraten. In Scarborough (1996) wordt een tipje van de sluier gelicht. We bespreken het algoritme in het kort, omdat het een goed beeld geeft van de algemene aanpak van niet-lineaire regressie.

We passen het logistische regressiemodel toe op de gegevens $(x_1, y_1), \dots, (x_n, y_n)$ met $x_1 < x_2 < \dots < x_n$ via de kleinste-kwadraten methode. We minimaliseren dus de volgende functie in de parameters a, b en c :

$$f(a, b, c) = \sum_{i=1}^n \left(\frac{c}{1 + ae^{-bx_i}} - y_i \right)^2$$

Door de partiële afgeleiden naar a, b en c uit te rekenen (de gradiënt van f) en gelijk aan nul te stellen, krijgen we een stelsel van niet-lineaire vergelijkingen. Dit stelsel is niet exact oplosbaar. Daarom lineariseren we de gradiënt van f via een Taylorbenadering van de eerste orde. Dit leidt tot de volgende formule:

$$\nabla f(a, b, c) = \nabla f(a_0, b_0, c_0) + H(a_0, b_0, c_0) \cdot \begin{pmatrix} a - a_0 \\ b - b_0 \\ c - c_0 \end{pmatrix}$$

waarbij de gradiënt (vector) $\nabla f(a, b, c)$ en de Hessiaan (matrix) $H(a, b, c)$ gedefinieerd zijn door:

$$\nabla f(a, b, c) = \begin{pmatrix} \frac{\partial f}{\partial a} \\ \frac{\partial f}{\partial b} \\ \frac{\partial f}{\partial c} \end{pmatrix}, H(a, b, c) = \begin{pmatrix} \frac{\partial^2 f}{\partial a^2} & \frac{\partial^2 f}{\partial a \partial b} & \frac{\partial^2 f}{\partial a \partial c} \\ \frac{\partial^2 f}{\partial a \partial b} & \frac{\partial^2 f}{\partial b^2} & \frac{\partial^2 f}{\partial b \partial c} \\ \frac{\partial^2 f}{\partial a \partial c} & \frac{\partial^2 f}{\partial b \partial c} & \frac{\partial^2 f}{\partial c^2} \end{pmatrix}$$

Herschrijven leidt tot de iteratieformule van de driedimensionale vorm van het algoritme van Newton-Raphson:

$$\begin{pmatrix} a_{k+1} \\ b_{k+1} \\ c_{k+1} \end{pmatrix} = \begin{pmatrix} a_k \\ b_k \\ c_k \end{pmatrix} - H(a_k, b_k, c_k)^{-1} \cdot \nabla f(a_k, b_k, c_k)$$

Hiermee wordt op iteratieve wijze het stelsel vergelijkingen numeriek opgelost. Wat rest is een slimme keuze van startwaarden. Hierover zegt Scarborough (1996) weinig, alleen dat de randwaarden worden gebruikt voor een ruwe schatting van beginwaarden voor parameters. Vermoedelijk worden de volgende twee vergelijkingen bedoeld:

$$y_1 = y(0) = \frac{c}{1+a}, \quad y_n = \lim_{x \rightarrow \infty} y(x) = c$$

Voor de startwaarde van b worden wellicht de medianen \bar{x} en \bar{y} van de x en y data gebruikt via de formule:

$$\bar{y} = \frac{c}{1 + ae^{-b\bar{x}}}$$

Als alles goed gaat, itereert het algoritme in een eindig aantal stappen naar een oplossing. Het is echter denkbaar dat de startwaarden van de parameters dusdanig slecht gekozen zijn dat er geen convergentie optreedt of dat er onderweg een Hessiaan voorkomt die geen inverse heeft. De rekenmachine geeft in het laatste geval de boodschap dat een singulariteit optreedt.

Een veelgebruikt en meer verfijnd alternatief voor de Newton-Raphsonmethode is het Levenberg-Marquardt algoritme (Marquardt, 1963). Dit wordt bijvoorbeeld toegepast in de functie-fit-routine van Coach. Ook dan worden aan de hand van gekozen startwaarden van parameters iteratief verbeteringen aangebracht.

Een semi-automatische regressiemethode ligt nu voor de hand: eerst geschikte startwaarden bepalen en dan een automatische (iteratieve) verbetering van deze parameterwaarden. In Coach kunnen de startwaarden handmatig worden gekozen door met de computermuis een regressiekromme met het blote oog neer te leggen bij de data. Handig is ook dat het regressiemodel niet alleen op alle gegevens tegelijk hoeft worden toegepast, maar dat ook met slechts een deel van de gegevens kan worden gewerkt (door in het grafiekenvenster op de gewenste data in te zoomen). Het nut hiervan zal pas echt duidelijk worden in de paragraaf over het biologische groeimodel. De S-kromme in figuur 3 is op deze manier geconstrueerd.

S-kromme van gewicht naar leeftijd tijdens de adolescentie

We bespreken in deze paragraaf een systematische aanpak van het schatten van parameters in een sigmoïdaal regressiemodel. We doen dit aan de hand van de gemiddelde groei in gewicht van Nederlandse jongens tijdens de adolescentie. In figuur 1 staat een schermafdrruk van een Coach-activiteit over groei in gewicht. Linksboven is het gemiddelde gewicht van Nederlandse jongens (in kg) tegen de leeftijd (in jaren) uitgezet als punten in het diagram. In hetzelfde kwadrant is ook de grafiek van de groeisnelheid en die van een sigmoïdaal model getekend. Duidelijk is te zien dat de groeisnelheid in de eerste levensjaren snel afneemt, tijdens de kinderperiode tot de puberteit nagenoeg lineair toeneemt en dat er tijdens de puberteit een groeispuurt plaatsvindt met pieksnelheid rondom het veertiende levensjaar.

Het wiskundige model in dit diagram lijkt op het zicht heel goed te kloppen voor wat de adolescentie betreft. De sterkere toename in gewicht tijdens de puberteit dan in de jaren ervoor heeft overigens bij jongens vooral te maken met een toename in spier- en botweefsel, bij meisjes ook met een toename van vetmassa (Wit & Delemarre-van de Waal, 1998).

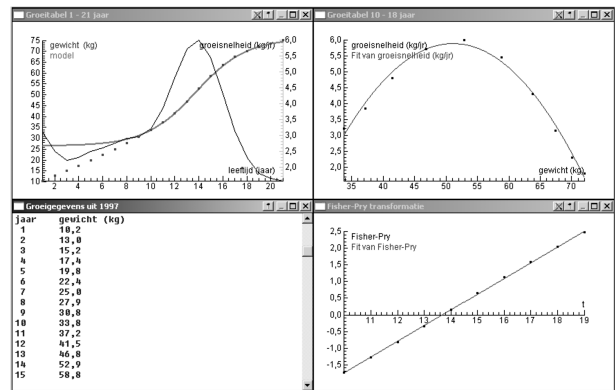


fig. 1 Schermafdrruk van een Coach-activiteit waarin een S-kromme wordt bepaald

We beperken ons voor jongens tot de periode van tien tot en met negentien jaar. We passen de formule van de S-kromme voor gewicht toe. Deze is gelijk aan:

$$G(t) = \frac{c}{1 + e^{b(a-t)}} + d$$

waarbij $G(t)$ het gewicht op leeftijd t is en a , b , c , d groeiparameters zijn. Deze formule past bij de volgende differentiaalvergelijking:

$$G' = \frac{b}{c}(G-d)(c+d-G)$$

Dit betekent dat, als de groeisnelheid uitgezet wordt tegen het gewicht, de meetpunten op een parabool moeten liggen. Voor de groeisnelheid op leeftijd t nemen we (bij gebrek aan beter) het centrale differentiequotient:

$$G'(t) = \frac{G(t+1) - G(t-1)}{2}$$

Rechtsboven in figuur 1 is de best bijpassende parabool te zien met als formule (bij benadering en in ontbonden vorm opgeschreven met behulp van de abc -formule):

$$G' \approx -0,00965(G - 26,4)(G - 75,8)$$

Kortom: $c \approx 49,4$ en $d \approx 26,4$. Leerlingen die de abc -formule niet meer paraat hebben of maar lastig vinden, kunnen ook de nulpunten van de parabool met hun rekenmachine bepalen en zo de ontbinding uitrekenen. Maar hoe nu verder met de bepaling van de groeiparameters a en b ? De zogenaamde Fisher-Pry transformatie biedt uitkomst. Deze is gedefinieerd als:

$$FP(t) = \ln\left(\frac{F(t)}{1 - F(t)}\right)$$

met: $F(t) = \frac{G(t) - d}{c}$

De reden waarom juist deze transformatie wordt geïntroduceerd, is pas duidelijk nadat we de formule voor de

S-kromme herschrijven. Start met:

$$F(t) = \frac{1}{1 + e^{b(a-t)}}$$

Neem aan weerszijden reciproque waarden en isoleer vervolgens de e-macht:

$$\frac{1}{F(t)} - 1 = e^{b(a-t)}$$

Breng het linkerlid onder één noemer, vermenigvuldig aan weerszijden met -1 en neem de logaritme. Dit levert het volgende lineaire verband op:

$$FP(t) = b(t - a)$$

Dit is precies de basisgedachte achter de Fisher-Pry transformatie, die vernoemd is naar de personen die deze methode populair hebben gemaakt (Fisher & Pry, 1971): als het sigmoïdale groeimodel goed functioneert, dan moet er een lineair verband herkenbaar zijn.

De parameters a en b zijn met lineaire regressie eenvoudig te bepalen. Rechtsonder in figuur 1 is het lineaire verband goed zichtbaar. Er geldt: $b \approx 0,474$ en $a \approx 13,7$. Dit impliceert een piek in de groeisprint bij de leeftijd van ongeveer dertien jaar en acht maanden. Dit komt overeen met het maximum in het diagram van groeisnelheid naar leeftijd. Voor meisjes in de leeftijd van negen tot achttien jaar kunnen we hetzelfde doen. We vinden dan:

$$G_{\text{meisjes}} = \frac{40,9}{1 + e^{0,468(11,9-t)}} + 22,7$$

Dit impliceert een piek in de groeisprint bij de leeftijd van ongeveer elf jaar en elf maanden. Dit is bijna twee jaar eerder dan bij jongens. Dit stemt overeen met het feit dat meisjes gemiddeld anderhalf jaar eerder in de puberteit geraken dan jongens (Wit & Delemarre-van de Waal, 1998; Mul et al, 2001).

ICT in eigen hand

In de vorige paragraaf hebben we een systematische aanpak voor het vinden van geschikte waarden van parameters in een S-kromme beschreven. Hoewel deze methode uitgelegd en geïllustreerd is aan de hand van een Coach-activiteit, heeft deze software niet het exclusieve recht hierop. Alle stappen kunnen ook met een spreadsheet-programma of een grafische rekenmachine worden gedaan.

In figuur 2 is een Excel rekenblad te zien waarin de groeisprint van Nederlandse jongens wat betreft lengte via het sigmoïdale model wordt onderzocht. De grafiek van de Fisher-Pry transformatie geeft een beeld van hoe goed het model werkt. De parameterwaarden kunnen via een schuifbalk worden veranderd. Dit geeft een idee van welke rol deze parameters spelen, hoe moeilijk het schatten

met het blote oog is, en hoe gevoelig het model is voor veranderingen in parameterwaarden.

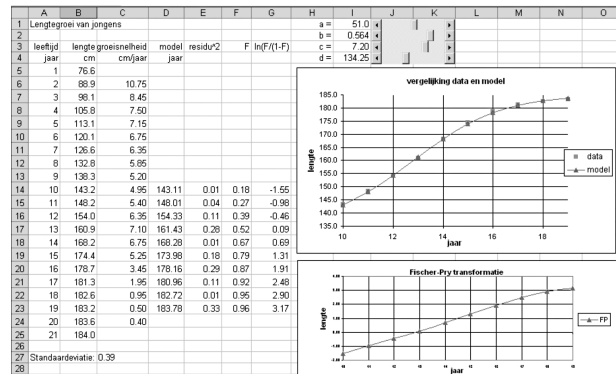


fig. 2 Excel rekenblad ter bepaling van een S-kromme

Laten we enkele details van het gebruik van een TI-83 bespreken. De eerste stap in bovenstaand algoritme, het schatten van de parameters c en d door in een grafiek de groeisnelheid tegen het gewicht uit te zetten en een kwadratisch model hierop toe te passen, is de lastigste. Het uitrekenen van het centrale differentiequotiënt van het gewicht kan op de volgende manier: stel dat de leeftijd in lijst L_1 opgeslagen is en dat het gemiddelde gewicht in lijst L_2 staat, voer dan het differentiequotiënt als nieuwe lijst, L_3 , in via het commando¹:

$$\text{seq}((L_2(X+1) - L_2(X-1)) / 2, X, 2, \dim(L_2) - 1)$$

Verwijder vervolgens het eerste en laatste element uit de lijsten L_1 en L_2 (met de DEL-toets) opdat de dimensie van elk van deze lijsten gelijk wordt aan die van L_3 en elk differentiequotiënt in de juiste rij staat.

Zet hierna de groeisnelheid (L_3) in een grafiek uit tegen het gewicht (L_2) en pas hierop een kwadratisch regressiemodel toe. Zorg ervoor dat de regressieformule opgeslagen wordt in een functievariable (Y_1). Het commando dat dit doet is QUADREG L_2, L_3, Y_1 met het resultaat dat hieronder staat.

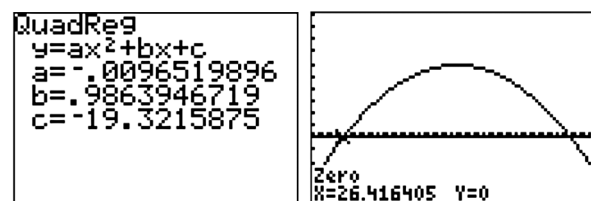


fig. 3 Het resultaat van QUADREG L_2, L_3, Y_1

De nulpunten van de kwadratische functie kunnen grafisch worden bepaald. In figuur 3 is het kleinste nulpunt in de grafiek af te lezen.

De tweede stap in het algoritme, de Fisher-Pry transformatie waarin een lineair verband wordt gecreëerd, heeft weinig uitleg: maak een nieuwe lijst, L_4 , met de formule:

$$L_4 = \ln\left(\frac{L_2 - d}{c} / \left(1 - \frac{L_2 - d}{c}\right)\right)$$

Zet deze lijst uit tegen L_1 en bepaal een lineaire regressie met een rechte lijn.

De rol van ICT in de wijze waarop we de S-kromme schatten vatten we samen als:

- *Visualisatie*: een tabel maken van groeigegevens, hiervan de grafiek tekenen, en de grafiek van de regressiekromme en de Fisher-Pry transformatie tekenen.
 - *Rekenen*: berekenen van numerieke afgeleide en vinden van nulpunten van een parabool (al dan niet grafisch).
 - *Analyse*: lineaire en kwadratische regressie toepassen.
 - *Controle*: nagaan hoe goed een regressiemodel werkt.
- Duidelijk is dat een zwaar beroep op de faciliteiten van software of rekenmachine wordt gedaan en dat het begrijpen en uitvoeren van het algoritme voor leerlingen een pittige opdracht is.

Dit roept de volgende vraag op: loont het wel de moeite om VWO-leerlingen met het algoritme te laten werken? Als een leerling een enkele keer een logistisch of sigmoïdaal regressiemodel moet toepassen op beschikbare gegevens, dan is het antwoord op deze vraag ontkenkend. In dat geval kun je bijvoorbeeld volstaan met het gebruiken van de in de TI-83 ingebouwde functie voor logistische regressie. Of nog handiger is het gebruik van de functie voor het schatten van S-krommen die in software zoals Coach en TI-Interactive, en in meer geavanceerde rekenmachines, zoals de TI-89 en TI-92, aanwezig is.

Als je daarentegen met leerlingen de algemene heuristiek van het omzetten van een niet-lineair regressiemodel in een lineair model wilt bespreken, dan is de beschreven aanpak van het schatten van een S-kromme een prachtig, niet-triviaal voorbeeld uit de wereld van wiskundige toepassingen.

Het biologische groeimodel

Met een gebruikersvriendelijke manier om op een zelfgekozen deel van gegevens een regressiemodel toe te passen, is het gemakkelijk om modellen te bestuderen die uit meerdere componenten bestaan.

We bekijken in deze paragraaf met behulp van Coach het biologische groeimodel voor gewicht naar leeftijd in de periode van drie tot negentien jaar:

$$\text{gewicht} = \text{ondergrens} + \text{gewicht}_{\text{kindertijd}} + \text{gewicht}_{\text{puberteit}}$$

waarbij voor $\text{gewicht}_{\text{kindertijd}}$ en $\text{gewicht}_{\text{puberteit}}$ elk een logistisch model wordt gebruikt. De schatting van de parameters gebeurt in twee stappen:

1. *S-kromme periode van negen tot en met negentien jaar*
In figuur 4 is te zien dat we in het grafiekenvenster ingezoomd hebben tot de periode negen tot en met negentien jaar. Met de menu-optie FUNCTIE-FIT krijg je

een apart venster waarin je het regressiemodel kunt kiezen, en geschikte startwaarden van de parameters handmatig kunt instellen.

In ons voorbeeld van een S-kromme kun je evenwel meteen op de knop voor automatische regressie klikken. Het resultaat past goed bij de gegevens.

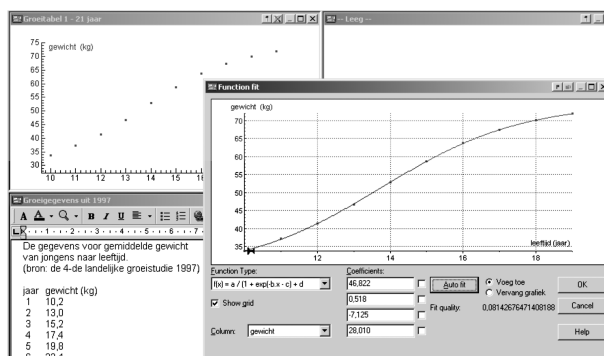


fig. 4 S-kromme voor de periode van 9 tot en met 19 jaar

2. *S-kromme voor het verschil van het model voor de adolescentie en het vastgestelde gemiddelde gewicht*
In figuur 5 is de grafiek van het model voor adolescentie getekend in het diagram van gewicht naar leeftijd. De verschilgrafiek van dit model en van het vastgestelde gemiddelde gewicht is hierin ook getekend. Het kwadrant rechtsboven is een ingezoomde versie van deze verschilgrafiek. Op het oog lijkt dit weer op een kromme die als S-kromme kan worden gemodelleerd. Het FUNCTIE-FIT venster illustreert dit.

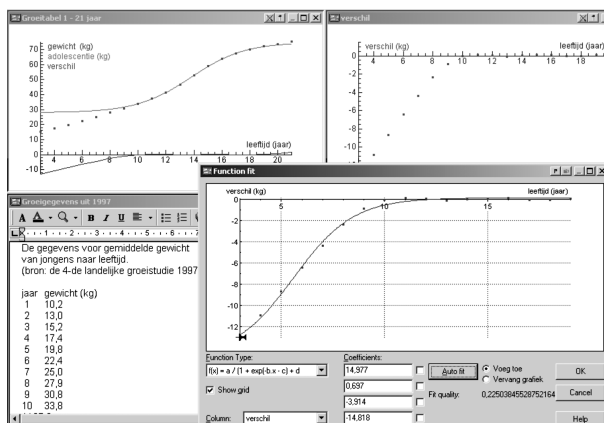


fig. 5 S-kromme voor het verschil van het adolescentiemodel en het vastgestelde gemiddelde gewicht

In figuur 6 staat de somgrafiek van de twee berekende S-krommen, samen met de verschilgrafiek van het samengestelde model en het vastgestelde gemiddelde gewicht. We lezen hier af dat de afwijking minder dan een halve kilogram is. Dit is geen slecht resultaat. Het laat zien dat het biologische model een goede kandidaat is om groei in gewicht voor de periode van drie tot en met negentien jaar te beschrijven.

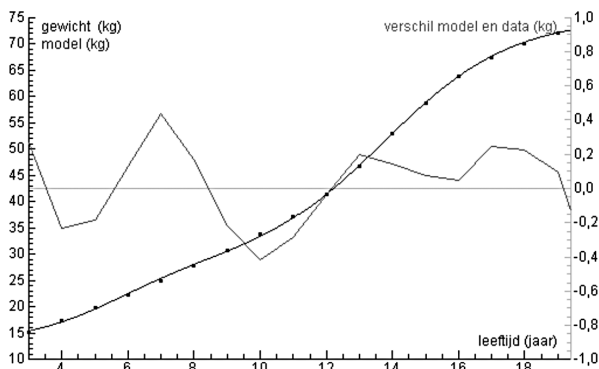


fig. 6 Grafische verificatie van het biologische model

De formules voor Nederlandse jongens en meisjes zijn in deze aanpak:

$$G_{\text{jongens}}(t) = 13,2 + \frac{15,0}{1 + e^{0,697(5,6-t)}} + \frac{46,8}{1 + e^{0,518(13,8-t)}}$$

$$G_{\text{meisjes}}(t) = 7,0 + \frac{18,1}{1 + e^{0,540(3,7-t)}} + \frac{39,1}{1 + e^{0,503(12,1-t)}}$$

met $G(t)$ het gewicht (in kg) op leeftijd t (in jaren). Deze formules leveren een goede start op voor een niet-lineaire regressie met behulp van het Levenberg-Marquardt algoritme. De verbeterde formules voor het biologische groeimodel die op deze manier worden gevonden⁴ zijn:

$$G_{\text{jongens}}(t) = 8,4 + \frac{26,5}{1 + e^{0,375(5,8-t)}} + \frac{39,5}{1 + e^{0,518(14,1-t)}}$$

$$G_{\text{meisjes}}(t) = 10,0 + \frac{19,1}{1 + e^{0,481(5,5-t)}} + \frac{33,9}{1 + e^{0,534(12,4-t)}}$$

Met deze nieuwe formules zijn de afwijkingen van het model met het vastgestelde gemiddelde gewicht minder dan tweehonderd gram over de hele leeftijdsperiode. Wat niet onvermeld mag blijven, is dat het biologische model ook een goede wiskundige beschrijving voor lengtegroei in de leeftijdsperiode van drie tot en met negentien jaar geeft. De formules zijn dan:

$$L_{\text{meisjes}}(t) = 59,3 + \frac{89,9}{1 + e^{0,330(4,0-t)}} + \frac{21,7}{1 + e^{0,753(11,7-t)}}$$

$$L_{\text{jongens}}(t) = 51,9 + \frac{106,7}{1 + e^{0,279(4,0-t)}} + \frac{26,6}{1 + e^{0,849(13,6-t)}}$$

met $L(t)$ de lengte (in cm) op leeftijd t (in jaren). De afwijkingen zijn minder dan 2,5 mm!

Waarvoor dienen de groeimodellen?

In de kindergeneeskunde hebben de groeimodellen geen klinisch nut; alleen de nationale referentiediagrammen voor gewicht en lengte spelen een rol bij de bepaling of

er sprake is van een groeistoornis. De wiskundige groeimodellen worden uitsluitend voor wetenschappelijk onderzoek gebruikt. We geven drie voorbeelden.

1. Kijkend naar bovenstaande formules voor gewicht en lengte naar leeftijd vallen twee dingen op:

- De leeftijden tijdens de puberteit waarop de groeisnelheden in lengte en gewicht hun piek bereiken zijn bij meisjes lager dan bij jongens. Dit heeft ook te maken met het feit dat meisjes eerder in de puberteit geraken dan jongens.

- Tijdens de puberteit treedt de piek in groeisnelheid bij lengte eerder op dan bij gewicht. Dit geldt zowel voor jongens als meisjes. Met andere woorden, tijdens de puberteit groeien kinderen eerst meer in de lengte dan in gewicht; ze krijgen dan een slanker postuur.

Uit de medische literatuur (Sinclair & Dangerfield, 1998; Roche & Sun, 2003) is bekend dat de piek bij groei in gewicht een paar maanden later optreedt dan bij lengtegroei, maar het verschil is bij meisjes niet zo groot als het biologisch model van gemiddeld gewicht naar leeftijd suggereert: slechts drie maanden in plaats van de negen maanden in ons model.

2. Volgens ons biologische model is de gewichtstoename van Nederlandse meisjes tijdens de puberteit het grootst op de leeftijd van twaalf jaar en vijf maanden. Dit is vóór de gemiddelde leeftijd van dertien jaar en twee maanden waarop zij voor de eerste keer menstrueren. Ook dit is bekend uit de medische literatuur. Er wordt gesuggereerd dat de eerste menstruatie (menarche) begint als een kritisch gewicht van pakweg 47 kg en een kritisch vetpercentage bereikt zijn.

In ons wiskundige model wordt het kritische gewicht bereikt bij een leeftijd van twaalf jaar en negen maanden. Overigens, de leeftijd van de menarche is ook gecorreleerd met lengte en Quetelet index (BMI): bij gegeven leeftijd hebben lange en zware meisjes een grotere kans op de menarche dan korte en lichte meisjes. Een lichaamsgewicht van meer dan 60 kg of een BMI van meer dan 20 kg/m² heeft geen effect meer op deze kans (Mul et al, 2001).

- Wiskundige modellen worden ook gebruikt om groei van kinderen van verschillende afkomst met elkaar te vergelijken. Denk bijvoorbeeld aan een vergelijking van allochtone en autochtone kinderen. Onder autochtone kinderen verstaan we dan kinderen met minstens één ouder van Nederlandse afkomst, terwijl de andere dan van West-Europese afkomst moet zijn.

- Het biologische model voor gemiddelde lengtegroei van Marokkaanse meisjes in Nederland levert de volgende formule op:

$$G_{\text{Marokkaanse}}(t) = 65,8 + \frac{71,5}{1 + e^{0,379(4,0-t)}} + \frac{24,2}{1 + e^{0,842(11,0-t)}}$$

Uit deze formule mogen we concluderen dat allochtone meisjes hun puberteitsspurten in lengte-groei gemiddeld eerder bereiken dan autochtone meisjes (pakweg acht maanden eerder). Dit komt omdat Marokkaanse meisjes in Nederland eerder in de puberteit geraken dan autochtone meisjes (Rikken, 1999; Fredriks et al, 2002).

Tot slot

Denk niet dat het biologische model het enige succesvolle wiskundige model voor groei in gewicht van jongens en meisjes is. Als alternatief voor logistische groei kan bijvoorbeeld de keuze vallen op Gompertz-groei. Ook dit levert een S-vormige groeikromme van begrensde groei op. Hierbij past de formule:

$$y(t) = e^{(b - e^{k-at})}$$

voor zekere parameters a , b en k . De bijpassende differentiaalvergelijking is:

$$y' = ay(b - \ln(y))$$

De gelijkens en het verschil met logistische groei komt door deze vergelijking goed tot uitdrukking. Het model van Gompertz vormt een onderdeel van het samengestelde groeimodel dat in Sumiya et al (2001) is gebruikt voor de wiskundige modellering van groei in gewicht van Japanse kinderen. Dit brengt ons bij een andere belangrijke rol die ICT speelt in het wiskundig modelleren: het biedt de mogelijkheid om te experimenteren. Zo hebben we al spelend met Coach gevonden dat voor gewichtsgroei van Nederlandse jongens het volgende model een maximale afwijking van driehonderd gram geeft:

$$G(t) = \begin{cases} G_1(t) & \text{als } t \in [3,10] \\ G_1(t) + G_2(t) & \text{als } t \in [11,19] \end{cases}$$

met:

$$G_1(t) = 0,07t^2 + 1,8t + 9,1$$

en:

$$G_2(t) = 4,0 \sin(0,6t - 1,9) + 3,8$$

Dit model is niet meer dan een speeltje, want het werkt niet zo goed bij groei in gewicht van meisjes en al helemaal niet voor lengtegroei van kinderen.

Het is eigenlijk meer een voorbeeld van betekenisloos modelleren. Het kan worden gebruikt om leerlingen te laten zien dat het altijd verstandig is om de zinvolheid en toepasbaarheid van een wiskundig model te onderzoeken, en dat het de moeite loont om een onderbouwing voor een wiskundig model te zoeken.

André Heck, Amstel Instituut, UvA, Amsterdam.

Noten

- [1] De operatoren SEQ en DIM haal je op door 2ND LIST in te toetsen, het menu OP te kiezen en dan de gewenste operator te selecteren.
- [2] Dit betreft Coach 6. Bij verschijning van dit nummer is dit nog een product in ontwikkeling.
- [3] De routine LogisticPlus komt niet in het statistiekmenu voor, maar is wel te gebruiken als wiskundige opdracht.
- [4] Hiervoor hebben we de NONLINEARREGRESS procedure uit het Statistics pakket van *Mathematica* gebruikt.