

Op het Stedelijk Gymnasium te Leiden is de module Levende Statistiek uitgetoetst, Een verslag van **Jacob van Eeghen** en **Liesbeth de Wreede**.

Levende Statistiek, een module voor vwo wiskunde D

Statistiek is typisch wiskunde A en dus eigenlijk geen wiskunde.

Statistiek is saai.

Je hebt niets aan statistiek, de voorbeelden op school hebben niets met de praktijk te maken.

Veel wiskundeleraren zullen deze opvattingen – van leerlingen of collega's – herkennen en het er soms mee eens zijn. Toen onze inzending voor een workshop Levende Statistiek tot de prijswinnaars voor de Nationale Wiskundedagen 2010 bleek te behoren, waren wij niet alleen vereerd, maar zagen we hierin ook een kans om tegen deze opvattingen ten strijde te trekken. Met enige uitbreiding van de gebruikelijke stof is statistiek namelijk om te vormen tot één van de mooiste voorbeelden van toegepaste wiskunde op de middelbare school. De toepassingen worden vooral door leerlingen die met hun profielwerkstuk bezig zijn, uiterst relevant gevonden. Onze workshop was gebaseerd op de door ons geschreven module Levende Statistiek voor wiskunde D op het vwo. Destijds was deze module nog in ontwikkeling en waren we delen ervan aan het uittesten met leerlingen. Inmiddels is de module af en beschikbaar op internet.

Het belang van statistiek, ook op school

Statistiek is één van de meest toegepaste onderdelen van de wiskunde. Statistiek is een verplicht en dikwijls omvangrijk onderdeel bij talloze universitaire studies: bij vrijwel alle gammavakken, bij studies in de biomedische sfeer en soms ook, maar minder frequent, bij de 'harde' bètavakken. Voor veel leerlingen is statistiek het enige onderdeel van de wiskunde dat expliciet aan de orde komt in hun vervolgstudie.

Statistiek vinden we ook terug in ons dagelijks leven en via de media: Is Lucia de B. terecht veroordeeld? Is een mediterraan dieet gezonder dan Hollandse kost? Is het een goed idee om alle meisjes te vaccineren tegen baarmoederhalskanker? Om berichten over dergelijke problemen op waarde te kunnen schatten, is enige kennis over de opzet van onderzoek en de rol van statistiek hierin onontbeerlijk.

Ook op school worden leerlingen vaak rechtstreeks geconfronteerd met het belang van statistiek als ze voor hun profielwerkstuk aan het experimenteren of enquêteren slaan en ze vervolgens de data proberen te analyseren.

Kortom: het belang van statistiek is overduidelijk en daarmee is het onderwerp een uitstekend voorbeeld voor 'contextrijke' wiskunde.

Toch maakt statistiek, net als kansrekening, inmiddels geen deel meer uit van het wiskunde-B-curriculum. Bij wiskunde A en D zijn kansrekening en statistiek wél verplicht; de inhoud van deze onderdelen is nagenoeg ongewijzigd ten opzichte van die van de 'oude' wiskunde A12, B1 en B12. De stof gaat echter net niet ver genoeg om de geleerde statistiek in de praktijk toepasbaar te maken. Dit wordt ieder jaar weer duidelijk als leerlingen voor hun profielwerkstuk de verzamelde gegevens uit hun experiment of enquête moeten verwerken. In het beste geval beseffen ze dat ze statistiek nodig hebben om tot een goede conclusie te komen, maar het komt maar zelden voor dat ze in staat zijn om de noodzakelijke analyse zelf uit te voeren. Dat is jammer. Wiskundeleraren steken veel energie in het motiveren van leerlingen door te wijzen op de relevantie van hun vak. Bij veel profielwerkstukken dient zich een concrete vraag aan die met wiskunde moet worden opgelost en wordt de leerling met een kluitje in het riet gestuurd. ("Dat zul je later nog wel leren.") Als de leerling geluk heeft, schrijft de wiskundeleraar de statistische paragraaf.

Dit kan ook anders. In de module Levende Statistiek wordt het aantal statistische technieken dat de leerling tot zijn beschikking heeft, zodanig uitgebreid dat op veel vragen uit de praktijk direct antwoord kan worden gegeven. Op het Stedelijk Gymnasium Leiden krijgen leerlingen die wiskunde D in hun pakket hebben, de mogelijkheid om in plaats van het schrijven van een eigen profielwerkstuk de statistiek te verzor-

gen van een aantal profielwerkstukken van anderen en zo als consultant op te treden. Tot dusver konden alle aangedragen praktijkgevallen adequaat door de leerling worden behandeld. Verderop worden twee eenvoudige voorbeelden gepresenteerd.

Inhoud van de module Levende Statistiek

Zoals gezegd behandelt de module de meest gangbare statistische onderwerpen, waarbij de verplichte onderdelen van het wiskunde D-curriculum bekend worden verondersteld. Deze onderwerpen zijn, na enige herhaling en uitbreiding van de principes van de kansrekening: de chi-kwadraat- en de t -verdeling, betrouwbaarheidsintervallen, t -toetsen (enkelvoudig, gepaard en ongepaard), verdelingsvrije toetsen, het vergelijken van populatiepercentages, enkelvoudige lineaire regressie, correlatie en meervoudige lineaire regressie.

De in de module aangeboden statistische technieken komen niet uit de lucht vallen, maar worden alle wiskundig afgeleid. Over het algemeen geldt: geen bewering zonder bewijs, waarbij een uitzondering wordt gemaakt voor onderdelen waarin met kansdichtheden moet worden gemanipuleerd. Voor deze onderdelen weten de leerlingen nog te weinig van integraalrekening. De wiskundige fundering zorgt ervoor dat de leerlingen statistiek niet als trukendoos zien. Zo heeft de module belangrijke neveneffecten: de leerlingen oefenen hun algebraïsche vaardigheden, het leveren van bewijzen en het toepassen van de principes uit de kansrekening.

De meeste berekeningen kunnen op de GR worden uitgevoerd. Voor grotere gegevensverzamelingen en voor lineaire regressie met meer dan één verklarende variabele is gebruik van statistische software op een computer onmisbaar. Wij kozen voor het gebruik van de software *R*. *R* is gratis en relatief eenvoudig te downloaden, zodat leerlingen ook thuis kunnen oefenen. *R* wordt veel gebruikt door professionele statistici omdat er inmiddels talloze gespecialiseerde functies in geschreven zijn. *R* is een commandotaal, maar er bestaat een menu-gestuurde ‘schil’, *RCommander*, waarmee de meeste standaardanalyses (waaronder alles wat in de module Levende Statistiek voorkomt) eenvoudig kunnen worden uitgevoerd. Onze ervaring is dat leerlingen er makkelijk mee overweg kunnen.

De meeste praktijkvoorbeelden uit de module zijn afkomstig van het Leids Universitair Medisch Centrum en zijn gebaseerd op medisch onderzoek. Het is onze ervaring dat dit toepassingsgebied de meeste leerlingen aanspreekt, zeker omdat de gekozen voorbeelden dicht bij echte onderzoeksvragen liggen.

Leerlingen raken door deze voorbeelden vertrouwd met de principes van het gebruik van wiskundige modellen. De praktische situatie wordt “vertaald” in een model waar je statistische technieken op kunt loslaten, maar je moet je steeds blijven afvragen of het model de essentie van de werkelijkheid adequaat weergeeft. Het woord ‘levende’ uit Levende Statistiek krijgt zo een dubbele betekenis: de statistiek is levend omdat je er problemen uit het ‘echte leven’ mee aankunt en de meeste voorbeelden komen uit de bio- oftewel de levenswetenschappen.

Twee voorbeelden

YOGA

De onderzoeksvraag uit dit profielwerkstuk luidde: “Vermindert yoga stress?”. Stress werd gemeten door de proefpersoon een formulier te laten invullen, nadat hem of haar een aantal stressverhogende vragen was gesteld. De antwoorden op het formulier worden omgezet in een score. Hoe hoger de score, hoe meer stress.

De proefpersonen werden aselekt onderverdeeld in twee groepen. Bij beide groepen werd de stress gemeten. Vervolgens kregen de leden van één van de groepen de opdracht om gedurende één week dagelijks een aantal voorgeschreven yoga-oefeningen uit te voeren. De andere groep kreeg geen opdracht. Na deze week werden beide groepen weer blootgesteld aan stressverhogende vragen en moesten ze opnieuw het formulier invullen. Zijn de gevonden scoreverschillen significant?

ESPERANTO

De onderzoeksvraag uit dit profielwerkstuk luidde: “Is Esperanto makkelijker te leren dan Spaans?”. De proefpersonen, die bij aanvang vertrouwd waren met Spaans noch Esperanto, kregen lessen in Spaans gevolgd door een toets en evenveel lessen Esperanto, gevolgd door een qua moeilijkheidsgraad vergelijkbare toets. Een deel van de proefpersonen werd eerst aan Spaans en dan aan Esperanto blootgesteld, bij de rest was de volgorde omgekeerd. Zijn de gevonden scoreverschillen significant?

In de kaders hierboven is het onderzoek van twee profielwerkstukken kort beschreven. Uiteraard roepen deze korte beschrijvingen allerlei vragen op: over vergelijkbaarheid van scores, samenstelling van groepen van proefpersonen etcetera. Deze vragen zijn gerechtvaardigd en maken zeker deel uit van een training in statistiek, maar in dit artikel gaan we er korthedshalve aan voorbij. We concentreren ons op de vraag of de gevonden verschillen in scores significant zijn. We

gaan ervan uit dat de waargenomen scores afkomstig zijn uit een normale verdeling. In de praktijk levert deze aanname doorgaans geen problemen op. Als dit wel het geval is, dan kan gebruik worden gemaakt van een zogenaamde verdelingsvrije methode.

In de eerste plaats constateren we dat de statistiek zoals geleerd bij wiskunde A onvoldoende is om deze situaties te analyseren. Het eerste probleem is dat op school altijd wordt verondersteld dat σ , de standaardafwijking van de betreffende verdeling, bekend is. In deze voorbeelden is σ , zoals gebruikelijk in de praktijk, onbekend. Het tweede probleem is dat op school meestal alleen de toets wordt behandeld waarin μ , het gemiddelde van de betreffende verdeling, wordt vergeleken met een van buitenaf gegeven waarde. In deze voorbeelden moeten de μ 's uit twee populaties worden vergeleken, hetgeen eveneens een situatie is die in de praktijk veelvuldig voorkomt. In de volgende paragraaf wordt in het kort de theorie behandeld, waarmee zowel YOGA als ESPERANTO succesvol kunnen worden behandeld.

Het principe van de t-toets

We concentreren ons eerst op het probleem van de onbekende σ .

Gegeven is de steekproef X_1, X_2, \dots, X_n uit een populatie die normaal verdeeld is met gemiddelde μ en variantie σ^2 . Het steekproefgemiddelde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

is normaal verdeeld met gemiddelde μ en variantie

$$\frac{\sigma^2}{n}.$$

Hieruit volgt dat de variabele $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ standaardnormaal verdeeld is.

Als we, bij gegeven μ_0 , de hypothese $H_0: \mu = \mu_0$ toetsen tegen een een- of tweezijdig alternatief, kunnen we, als σ bekend is, Z gebruiken als toetsingsgrootheid. Onder H_0 kunnen we Z immers berekenen.

Als σ onbekend is, ligt het voor de hand dat we σ (of σ^2) uit de gegevens proberen te schatten. σ^2 is gedefinieerd als $E(X - \mu)^2$, dus een logische keuze voor de schatter van σ^2 is

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

Deze schatter heeft als prettige eigenschap dat hij *zuiver* is, dat wil zeggen dat zijn verwachtingswaarde gelijk is aan de parameter die hij schat. Immers, $E(S^2) = \frac{1}{n} \cdot n\sigma^2 = \sigma^2$.

σ^2 heeft ook een onprettige eigenschap: hij is in de praktijk onbruikbaar omdat μ onbekend is. Het is ver-

leidelijk om nu in de formule voor S^2 μ te vervangen door zijn schatter \bar{X} . We zouden dan S^2 vervangen door

$$\frac{1}{n} \sum (X_i - \bar{X})^2.$$

Dit is echter niet onproblematisch, zoals duidelijk moge worden uit de volgende berekening.

$$\begin{aligned} \sum (X_i - \mu)^2 &= \sum ((X_i - \bar{X}) + (\bar{X} - \mu))^2 = \\ \sum (X_i - \bar{X})^2 + \sum 2(X_i - \bar{X})(\bar{X} - \mu) + \sum (\bar{X} - \mu)^2 &= \\ \sum (X_i - \bar{X})^2 + 2(\bar{X} - \mu) \sum (X_i - \bar{X}) + n(\bar{X} - \mu)^2 &= \\ \sum (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \end{aligned}$$

Uit deze gelijkheid volgt dat $\frac{1}{n} \sum (X_i - \bar{X})^2$ altijd kleiner of gelijk is aan de zuivere schatter S^2 en dus zelf niet zuiver is. Eenvoudig valt echter aan te tonen (ook door leerlingen) dat

$$s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

wél zuiver is en dit is dan ook de grootheid die we als schatter van σ^2 zullen gebruiken.

We kunnen nu de toetsingsgrootheid $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ vervangen door $T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$.

Om T te kunnen gebruiken voor het toetsen van $H_0: \mu = \mu_0$ is het nodig om de verdeling van T te bepalen.

Bovenstaande afleiding is, met een beetje extra uitleg, goed te volgen voor wiskunde D-leerlingen. De afleiding van de verdeling van T vergt echter te veel kennis van integraalrekening. We volstaan hier dan ook met de mededeling dat de verdeling van T alleen afhangt van de parameter n ; deze verdeling wordt de 't-verdeling met $n-1$ vrijheidsgraden' genoemd. De t -verdeling lijkt sterk op de standaardnormale verdeling, maar de 'staarten' zijn dikker, vooral bij een klein aantal vrijheidsgraden, zie figuur 1 op de volgende bladzijde. Op de GR kunnen berekeningen met de t -verdeling eenvoudig worden uitgevoerd via de functies tcdf en invT, te vergelijken met normalcdf en invNorm.

Gepaarde en ongepaarde t-toetsen

Voor het vergelijken van de gemiddelden van twee populaties, zoals in het voorbeeld YOGA, gebruiken we de volgende notatie.

X_{1i} ($i = 1, \dots, n_1$) is de stressvermindering (stressscore in de eerste meting minus stressscore in de tweede meting) van elk van de n_1 proefpersonen uit de groep die aan yoga deed.

$X_{2i} (i = 1, \dots, n_2)$ is de stressvermindering voor de proefpersonen uit de andere groep.

We veronderstellen dat alle waarnemingen onafhankelijk van elkaar zijn en normaal verdeeld.

μ_1 is het gemiddelde van de X_{1i} 's en μ_2 het gemiddelde van de X_{2i} 's. Verder veronderstellen we dat de X_{1i} 's en de X_{2i} 's dezelfde variantie hebben: σ^2 .

We toetsen $H_0 : \mu_1 = \mu_2$ tegen $H_1 : \mu_1 > \mu_2$, omdat we proberen aan te tonen dat yoga leidt tot meer stressvermindering. Het ligt voor de hand om deze hypothese te verwerpen als $\bar{X}_1 - \bar{X}_2$ 'groot' is. Je kunt aantonen (en leerlingen doen dat via een opgave) dat

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s \sqrt{1/n_1 + 1/n_2}},$$

waarin

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

het gewogen gemiddelde is van de geschatte varianties uit beide steekproeven, een t -verdeling heeft met $n_1 + n_2 - 2$ vrijheidsgraden en daarmee is het probleem opgelost.

Het voorbeeld ESPERANTO lijkt weliswaar op YOGA, maar kan niet op precies dezelfde manier worden aangepakt. In dezelfde notatie als bij YOGA gebruiken we het subscript 1 voor de score behaald bij de toets over Esperanto en het subscript 2 voor de score bij Spaans. Ook hier toetsen we $H_0 : \mu_1 = \mu_2$ tegen $H_1 : \mu_1 > \mu_2$. De veronderstellingen die we maakten bij YOGA kunnen alle worden overgenomen op één na: de X_{1i} 's en de X_{2i} 's zijn *niet* onafhankelijk, want het zijn de scores van dezelfde proefpersoon. We zeggen daarom dat de steekproef bij ESPERANTO uit *gepaarde* waarnemingen bestaat en die bij YOGA uit *ongepaarde* waarnemingen.

Voor een correcte aanpak van ESPERANTO definiëren we voor proefpersoon i : $X_i = X_{1i} - X_{2i}$ als het verschil tussen de score voor Esperanto minus de score voor Spaans. De X_i 's zijn onafhankelijk en normaal verdeeld met gemiddelde $\mu = \mu_1 - \mu_2$. De te toetsen hypothese is te herschrijven als $H_0 : \mu = 0$ tegen $H_1 : \mu > 0$.

Hiermee is de vergelijking van twee populaties teruggebracht tot een enkelvoudige t -toets.

In beide voorbeelden zijn we ervan uitgegaan dat de geobserveerde scores afkomstig zijn uit een normale verdeling. Als deze veronderstelling niet kan worden gerechtvaardigd, bestaat voor elk van de gebruikte t -toetsen een zogenaamde verdelingsvrije variant, die ook in de module Levende Statistiek wordt behandeld. De

uit wiskunde A bekende tekentoets is een voorbeeld van een verdelingsvrije toets. Deze zou in principe gebruikt kunnen worden voor ESPERANTO, maar het onderscheidingsvermogen van deze toets is klein. In deze situatie is Wilcoxon's rangtekentoets beter geschikt.

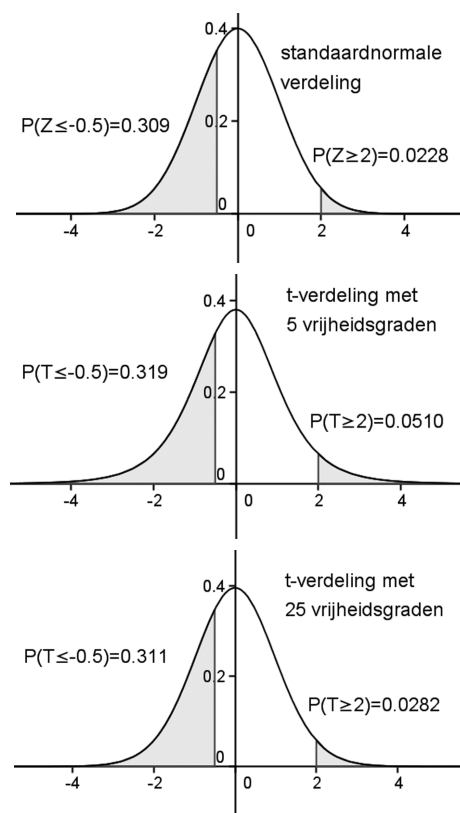


fig. 1 Verschillende verdelingen.

Ervaringen in de klas

De module Levende Statistiek is door drie klassen van het Stedelijk Gymnasium Leiden afgerond. De module is gesplitst in twee delen aangeboden, het eerste deel aan het eind van klas 5, het tweede aan het begin van klas 6. Per saldo nam de stof van Levende Statistiek, inclusief het R-computerpracticum, dertig lessen van 50 minuten in beslag.

De leerlingen waren vooral enthousiast over het realiteitsgehalte van de stof ("we leren iets echt"). Aan het begin van klas 6 begint het profielwerkstuk van de leerlingen vorm te krijgen. Voor veel leerlingen betekent dit onder andere dat een eigen onderzoek moet worden verricht: data worden verzameld en deze moeten worden geanalyseerd. Dit leidde tot veel gesprekjes na de les of tijdens pauzes. "Meneer, wilt u even kijken? Hier moeten we toch de xyz-toets toepassen?" Leerlingen lijken verbaasd als blijkt dat het wiskundig instrumentarium dat nodig is voor de oplossing van een praktisch probleem binnen handbereik ligt. Vervolgens is de voldoening groot als ze zelf de oplossing kunnen vinden.

Voor velen is het de eerste keer dat ze wiskunde toepassen in een situatie die niet is verzonnen door de leraar ter illustratie van de leerstof.

De wiskunde achter de statistiek wordt door de meeste leerlingen als lastig ervaren. Dit geldt met name voor de theoretische afleidingen; het toepassen van resultaten in een concreet voorbeeld wordt makkelijker gevonden. Voor wiskundeleraren is dit een bekend fenomeen. Er zijn maar weinig leerlingen die de productregel bij differentiëren kunnen afleiden, ook nadat deze afleiding een keer is voorgedaan. Er zijn er velen die de regel kunnen toepassen. Bij statistiek komt daar de complicatie van de notatie bij. Bij berekeningen over steekproeven van lengte n moet veelvuldig gebruik worden gemaakt van het sommatie-symbool \sum en de bijbehorende algebra. Ook moet er steeds een duidelijk onderscheid worden gemaakt tussen (populatie)parameters, waarvan de waarde in de praktijk onbekend is, maar waar we kansuitspraken over willen kunnen doen, en hun schatters, die kansvariabelen zijn omdat ze gebaseerd zijn op de steekproef.

In Levende Statistiek hebben wij deze moeilijkheden niet omzeild. Door middel van opgaven waarin ze zelf (kleine) theoretische resultaten moeten afleiden, worden leerlingen getraind om tot op zekere hoogte abstract te denken. Deze training verloopt niet zonder dat dikwijls de vraag wordt gesteld: “Moeten we dit weten voor de toets?”. Het antwoord “Ja” houdt ze in ieder geval aan het werk. Het is onze ervaring dat leerlingen juist bij dit onderdeel (formeel wiskunde en kansrekening toepassen om statistische resultaten af te leiden) het meest de hulp van de docent nodig hebben. Die hulp bestaat dan vooral uit het ‘vertalen’ van abstracte begrippen en symbolen in concrete voorbeelden. Opmerkelijk is dat het nut van de theoretische achtergrond op een gegeven moment wel wordt ingezien. Zo vertelde een leerling naar aanleiding van het bezoek aan een open dag van een universiteit dat zij de enige was die begreep wat er met de gepresenteerde formules (met veel \sum 's) bedoeld werd. Wij willen hier wel benadrukken dat onze doelgroep bestond uit leerlingen met wiskunde D, een groep die zich niet snel door wiskundige moeilijkheden uit het veld laat slaan. Bij een andere doelgroep is het raadzaam om wat minder de nadruk te leggen op de wiskundige achtergrond van de statistiek.

Het computerpracticum werd gewaardeerd en de leerlingen hadden weinig moeite met de software. Dat is

prettig, want zo kon de aandacht vooral uitgaan naar de interpretatie van de resultaten. Deze interpretatie is niet iets vanzelfsprekends. Leerlingen moeten er af en toe aan worden herinnerd dat ze hun gezonde verstand niet moeten uitschakelen. Als binnen een populatie van studenten het kenmerk ‘lichaamslengte’ sterk (en significant) gecorreleerd blijkt te zijn met ‘gemiddeld aantal alcoholische consumpties per week’ moeten ze zelf op het idee komen dat het kenmerk ‘geslacht’ wel eens van belang zou kunnen zijn. Bij de praktische opdracht, een analyse van een gegevensverzameling van ruim 1200 patiënten met nierproblemen, speelde de combinatie van gezond verstand en kennis van statistische technieken een grote rol.

Uiteindelijk wordt de tevredenheid van leerlingen voor een belangrijk deel bepaald door de behaalde cijfers. Deze waren bevredigend.

Conclusie

Aan het begin van dit artikel hebben wij enkele opvattingen geciteerd die illustratief zijn voor het negatieve imago dat statistiek bij sommigen heeft. Wij hopen dat de lezer met ons inziet dat, zeker als je iets verder kijkt dan op school meestal gedaan wordt, het vak statistiek dit imago niet verdient en dat de volgende opvattingen eerder van toepassing zijn:

Statistiek is niet het makkelijkste maar wel het mooiste voorbeeld van toegepaste wiskunde waar een middelbare scholier mee in aanraking kan komen.

Statistiek is leuk.

De meeste onderzoekers kunnen niet zonder statistiek en dat begint al op school.

Meer weten?

- De module Levende Statistiek, waaronder de instructies voor het downloaden van het software pakket R, zijn te vinden op de site van cTWO (www.fi.uu.nl/ctwo/WiskundeD/MateriaalDomeinenWiskundeD/welcome.html) onder VWO, Domein B.
- Voor verdere vragen kunt u contact opnemen met Liesbeth de Wreede, l.c.de_wreede@lumc.nl (zie ook <http://www.lumc.nl/con/3020/38285/901050350082510?setlanguage=Dutch&setcountry=nl>).

*Jacob van Eeghen,
docent wiskunde Stedelijke Gymnasium Leiden
Liesbeth de Wreede,*

biostatistisch onderzoeker Leids Universitair Medisch Centrum