

SORBet: Statistiek Op Relationele Bestanden

S. L. Kemme

Mathematisch Instituut, RU Groningen

Inleiding

In het kader van de integratie wiskunde-informatica in de bovenbouw van het vwo en de havo, is aan de Rijksuniversiteit Groningen het computerprogramma SORBet met bijbehorend lesmateriaal ontwikkeld voor het statistiekonderwijs in wiskunde A. In dit artikel wordt een beschrijving gegeven van het programma, het lesmateriaal en van ervaringen in de klas.

Een stukje geschiedenis

De invoering van wiskunde A en B in het vwo kwam tot stand op basis van het HEWET-rapport (1980). Daarin is in doelstellingen en globale leerstofomschrijvingen aangegeven welke kant het ongeveer uit moet. Voor wiskunde A wordt een omschrijving gegeven van het nieuwe vak Automatische Gegevens Verwerking:

'Als doeleinden van het informaticaonderwijs aan vwo-leerlingen wil de werkgroep onderscheiden:

- concrete problemen door analyse te leren formuleren en formaliseren;
- constructieve oplossingsmethoden te leren vinden voor dergelijke problemen;
- enigszins vertrouwd te raken met de werkwijze en toepassingen van de computer.'

In aansluiting op deze doelstellingen volgt een lijst van onderwerpen. De lijst is nogal uitgebreid en bevat onderdelen uit de (numerieke) wiskunde, maar ook typische onderdelen uit de informatica, zoals: leren programmeren, 'gegeven', gegevensverwerking, simulatie en tekstverwerking.

Bij de uiteindelijke invoering van Hewet is van al deze verwachtingen niet veel terechtgekomen. In één van de experimentele examens is een structuurdiagram opgenomen van een computersimulatie aan de hand waarvan een aantal vragen werden gesteld. Om de vragen te kunnen beantwoorden dienden de leerlingen een goed begrip te hebben van structuurdiagrammen. Sindsdien besteden leraren enkele lessen aan het kunnen lezen van structuurdiagrammen. De meeste leerlingen krijgen bij wiskunde A geen computer te zien, uitzonderingen daargelaten. Daarmee kan gerust gesteld worden dat aan de verwachtingen ten aanzien van het vak Automatische Gegevens Verwerking in wiskunde A niet is voldaan.

Intussen werd de druk steeds groter om meer aandacht te besteden aan het vak informatica in het voortgezet onderwijs. Op een gegeven ogenblik was zelfs de vraag aan de orde of informatica een zelfstandig vak dient te worden met een eigen bevoegdheidsregeling. Uiteindelijk leidde dit tot de uitspraak van het ministerie dat er in de onderbouw wat ruimte zal komen voor een zelfstandig vak informatica, maar dat informatica in de bovenbouw geïntegreerd dient te worden aangeboden binnen de andere vakken. Dus ook binnen wiskunde A in het vwo. Aan het Mathematisch Instituut van de Universiteit van Groningen werd in 1988 door PRINT de opdracht gegeven om hiervoor materiaal te ontwikkelen. Het lag voor de hand om een nieuwe invulling te geven aan AGV in wiskunde A. Daartoe werd geprobeerd informatica-aspecten uit de theorie van relationele gegevensbestanden te koppelen aan bestaande onderdelen uit de statistiek.

Halen meisjes betere resultaten dan jongens op het eindexamen?

Zo luidt vraag 10 van de leerlingentekst voor 5 vwo. Het aardige is dat dit een natuurlijke en zinvolle vraag is, die zich echter niet zo gemakkelijk laat beantwoorden. Er zullen eerst afspraken moeten worden gemaakt. Wat zijn 'betere' resultaten en naar welke vakken ga je kijken? Er zal een methode moeten worden gekozen waarmee de vraag statistisch en ondubbelzinnig kan worden beantwoord. Het leren nadenken over de methode van onderzoek bij statistische problemen, hoort essentieel bij het statistiekonderwijs. In de leerlingentekst worden de leerlingen bij deze vraag als volgt op weg geholpen:

'Om die vraag te kunnen beantwoorden moet je eerst afspreken wat je onder 'betere resultaten' verstaat. Laten we afspreken dat we zullen kijken naar de gemiddelde resultaten *per vak*. Het antwoord op de vraag zou dan kunnen zijn: meisjes zijn beter in wiskunde A en jongens zijn beter in Frans.'

Bij volgende vragen ontbreken deze aanwijzingen en zullen ze eerst zelf over de methode na moeten denken om tot een plan en tot opdrachten te komen waarmee met behulp van de computer een antwoord gegeven kan worden.

Zo word je door het gebruik van een computerpakket

op een vriendelijke doch dringende manier uitgedigd over de oplossingsstrategie van dergelijke vragen na te denken. Simpelweg omdat je een keuze moet maken uit een aantal mogelijke opdrachten die het programma biedt.

Hoe gaat het nu verder met de vraag of meisjes betere resultaten halen dan jongens op het eindexamen? De leerlingen hebben de beschikking over het computerprogramma SORBet waarin een grote tabel EXAMEN is opgenomen. In die tabel staan van drie scholen de cijfers van alle vakken van het examen vwo. Met het computerprogramma kunnen ze allerlei gegevens uit de tabel halen om die statistisch te verwerken in getallen of grafieken.

Ook kunnen verdelingen van een aantal variabelen (in dit geval: wiskunde A, Frans, geschiedenis, ...) vergeleken worden op hun karakteristieken, zoals gemiddelde, mediaan, modus, standaarddeviatie, kwartielafstand.

In dit geval zal gekeken worden naar de gemiddelden. De tabel moet worden gesplitst in twee tabellen: één voor jongens en één voor meisjes. Dat kan, want het programma vraagt of je wilt indelen in groepen. Door te kijken of de variabele 'seks' de waarde 'm' of 'v' heeft, worden twee nieuwe tabellen gemaakt en kunnen de karakteristieken van de aangegeven variabelen worden uitgerekend. Vanuit een statistisch overzichtelijke presentatie van de feiten in tabellen kunnen leerlingen zelf een verantwoorde conclusie trekken.

Plantlust

Tuincentrum 'Plantlust' weet de zegeningen van de natuur met die van de elektronika te verenigen door het automatiseren van de administratie. De bedrijfsgegevens over voorraad, inkoop en verkoop zijn ondergebracht in een verzameling tabellen, de database PLNTLUST. Eén van die tabellen, PLANTEN, bevat alle gegevens over het assortiment, gerangschikt in de kolommen: artcode, plntnaam, soort, kleur, hoogte, blb (begin bloeitijd), ble (einde bloeitijd), licht, bodem, prijs.

Meneer Rood wil alleen maar roodbloeiende planten in zijn tuin. Hij wil weleens weten wat Plantlust hem op dit punt te bieden heeft. De medewerker van Plantlust start het programma, kiest de goede database, gaat daarmee naar SORBet 0: **Database bewerken** en voert het volgende in (het vet gedrukte is als menu aanwezig):

Selecteer uit de tabellen: PLANTEN

De kolommen: plntnaam, soort, hoogte, blb, ble, licht, bodem, prijs.

Waarvoor geldt: kleur = 'ROOD'.

Het programma kiest nu de genoemde kolommen en het kiest de rijen die aan de voorwaarde voldoen. Dit levert een nieuwe (werk)tabel op waaruit de heer Rood een volgende keuze zou kunnen maken.

Het programma biedt dus de mogelijkheid om te selecteren uit een tabel. Er kunnen kolommen uit de tabel worden gehaald en er kunnen rijen worden geselecteerd op basis van een voorwaarde.

Ook de ontkenning van een voorwaarde is mogelijk. Mevrouw Rood, die een afschuw heeft van de kleur

rood, kan een overzicht van alle niet rood bloeiende planten krijgen door de voorwaarde te stellen: niet kleur = 'ROOD'.

De voorwaarden die aan de rijen gesteld kunnen worden, kunnen worden samengesteld met behulp van de logische voegwoorden 'en' en 'of'. Meneer Rood, die in tegenstelling tot zijn vrouw, wel van rood houdt maar niet van roodbloeiende bomen, kan rechtstreeks de voorwaarde stellen: kleur = 'ROOD' en niet soort = 'BOOM'.

Een opdracht om nagelbijter van te worden is de volgende:

Je wilt rode en blauwe planten. Wat moet je selecteren:
 kleur = 'ROOD' en kleur = 'BLAUW'
 of: kleur = 'ROOD' of kleur = 'BLAUW'?

Voor een visueel ingesteld iemand als meneer Rood, biedt het programma vele grafische mogelijkheden. Voor een snel grafisch overzicht kan hij in dit geval kiezen uit een histogram, een sectordiagram of een staafdiagram. In principe bevat het programma alle grafische mogelijkheden die in wiskunde A, zowel voor vwo als havo, aan de orde kunnen komen, inclusief correlatie en regressie.

In het volgende schema is de database PLNTLUST volledig samengevat:

PLNTLUST	
TABEL	KOLOMMEN
BESTLING:	bestelnr,levcode,bdatum,l datum,bedrag.
BESTLREG:	bestelnr,artcode,l aantal,bestprs.
LEVCIEERS:	levcode,levnaam,adres,woonplts.
OFFERTES:	levcode,artcode,artcode,levtijd,offprijs.
PLANTEN:	artcode,plntnaam,soort,kleur,hoogte, blb,ble,licht,bodem,prijs.
PLSCHEMA:	categori,hoogte1,hoogte2,afstand.

In de leerlingentekst wordt door middel van een denkbeeldige situatie uiteengezet wat die codes allemaal betekenen en waar je die voor nodig hebt.

Mevrouw Boom wil een ligusterheg bestellen bij PLANTLUST. Ze heeft uitgerekend dat ze 500 struiken nodig heeft. Dat aantal is niet in voorraad. Het tuincentrum probeert nu bij de gegeven prijs, zo voordelig mogelijk in te slaan. In de tabel PLANTEN wordt de artikelcode bij 'liguster' opgezocht (dat is: 210). Via de tabel OFFERTES wordt gekeken welke leverancierscodes (levcode) en welke leveranciers-artikelcodes (artcode) bij artcode 210 aanwezig zijn. Uit deze tabel blijkt dat er drie leveranciers zijn die ligusters leveren. De prijs is bij alledrie hetzelfde: f 0,20 per stuk, maar er zijn aanzienlijke verschillen in leveringstijd (21, 10, 14 dagen). Leverancier 021 heeft de snelste levertijd, dus wordt deze gekozen. In de tabel BESTLREG (bestelregels) wordt deze bestelling vastgelegd. Aan het eind van de dag worden alle bestellingen voor leverancier 021 klaargemaakt via de tabel BESTLING, het adres wordt opgespoord via LEVCIEERS en de bestellingen worden op de post gedaan. De hele procedure zie je hieronder nog eens in schema vastgelegd:

Klas	Resultaat
PLANTEN:	
plntnaam='LIGUSTER',	artcode = '210'

OFFERTES:

artcode='210';

levcode = '009, artcode =

'0,97', levtijd = 21

of: levcode = '021', artcode =

'LIG', levtijd = 10

of: levcode = '022', artcode =

'S · 015', levtijd = 14

LEVCIERS:

levcode='021';

levnaam, adres, woonplts.

De procedure is nogal ingewikkeld, maar bij een groot tuincentrum is dat onvermijdelijk. Bovendien valt het allemaal wel mee als de computer al het zoekwerk doet.

In principe zou je alle gegevens ook in één grote tabel kunnen stoppen, maar dan wordt die tabel zo groot dat je ieder overzicht verliest.

Bovendien stop je dan allemaal gegevens bij elkaar die eigenlijk niets met elkaar te maken hebben. Zo heeft de kleur van een plant natuurlijk niets te maken met het adres van de leverancier. Ook dat komt het overzicht niet ten goede. Het is de kunst om je gegevens zó in te delen dat alles wat bij elkaar hoort ook bij elkaar komt te staan in één tabel. De gegevens over de leveranciers vind je bijvoorbeeld terug in de tabel LEVCIERS. Als er een leverancier verhuist, of er een nieuwe leverancier bijkomt, hoef je alleen maar deze tabel te veranderen. Je weet meteen welke tabel je moet kiezen en wat je moet veranderen.

In deze tabellen wordt veel met 'codes' gewerkt. De kolom 'levcode' bijvoorbeeld bevat codenummers van leveranciers. Je had net zo goed de hele naam van de leverancier kunnen nemen, maar dit werkt gemakkelijker. De kolom 'artcode' bevat de codes die de leveranciers gebruiken voor hun artikelen, de kolom 'artcode' bevat de code die het tuincentrum zelf hanteert voor haar artikelen. In de tabel OFFERTES is vastgelegd hoe deze codes bij elkaar horen.

Via de codes kun je systematisch van tabel naar tabel gegevens gaan zoeken.

Vraag 36. Een boze mevrouw Boom belt op waar haar ligusters blijven die ze drie weken geleden besteld heeft. Hoe zoek je in het gegevensbestand van PLANTLUST wanneer die ligusters besteld zijn, bij welke leverancier en wanneer die worden geleverd? Maak een soortgelijk schema als hierboven. (N.B. Je hoeft niet het antwoord te vinden op de vragen, het gaat erom dat je aangeeft hoe je dat antwoord zou kunnen opzoeken.)

In het voorbeeld wordt het idee van een 'relationele database' geïntroduceerd. Tussen de tabellen bestaat een zekere relatie doordat ze een kolom gemeenschappelijk kunnen hebben. Door middel van deze relatie kunnen tabellen met elkaar gecombineerd worden.

Bovendien wordt in dit fragment een stukje didactiek zichtbaar gemaakt. Van de leerlingen wordt gevraagd of ze eerst iets met 'de hand' willen uitvoeren. In dit geval is dus bewust gekozen voor een theoretische oriëntatie vóór ze met de computer aan de gang kunnen.

Een kleine giro

Om duidelijk te maken wat er precies gebeurt als in een relationele database twee tabellen gekoppeld worden, is een kleine database GIRO bedacht. Deze bestaat uit twee tabellen:

GEMACHTN	
nr	naam
9999	G. Pas
12345	J. Rap
12345	J. Maat
20001	T.E. Bros

SALDI	
gironr	saldo
9999	2165
12345	-23
20001	885

Jan Rap en z'n maat werken onder hetzelfde nummer en via dat nummer kom je aan de weet dat ze een negatief saldo hebben. Om de naam met het bedrag te combineren moet je in de twee tabellen kijken via het gemeenschappelijke nummer. Het zou mooi zijn als de combinatie naam-bedrag geautomatiseerd in een nieuwe tabel werd gegeven. Dat betekent dat gegevens uit beide tabellen aan elkaar gekoppeld moeten worden. In het programma is het mogelijk uit twee (of drie) tabellen tegelijk te selecteren. Kies je 'Selecteer uit de tabellen: GEMACHTN, SALDI' zonder een extra voorwaarde te stellen of speciale kolommen te kiezen dan is het resultaat:

nr	naam	gironr	saldo
9999	G.Pas	9999	2165
9999	G.Pas	12345	-23
9999	G.Pas	20001	885
12345	J.Rap	9999	2165
12345	J.Rap	12345	-23
12345	J.Rap	20001	885
12345	J.Maate	9999	2165
12345	J.Maate	20001	885
12345	J.Maate	20001	885
20001	T.E.Bros	9999	2165
20001	T.E.Bros	20001	885
20001	T.E.Bros	20001	885

De domme computer heeft alles met alles gecombineerd! De eerste rij van de eerste tabel wordt met elke rij van de tweede tabel gecombineerd. Hetzelfde gebeurt met de andere rijen van de eerste tabel. Zo ontstaat een grote tabel van 4 (= 2 + 2) kolommen en 12 (= 4 × 3) rijen waarin gegevens met elkaar gecombineerd zijn die niet bij elkaar horen. Binnen de opzet van de relationele databases is deze wijze van koppelen van twee tabellen een standaardoperatie die op z'n Engels 'join' genoemd wordt. Het is een standaardbewerking waarbij twee of meer tabellen met elkaar gecombineerd worden, waarbij geen informatie verloren gaat maar waarbij tegelijkertijd een tabel verloren kan ontstaan.

Hoe krijg je hieruit de gewenste tabel? Het gaat om de namen en de saldi, dus selecteren we alleen deze twee kolommen. Vervolgens wil je alleen rijen met zinvolle combinaties hebben. Als je goed naar bovenstaande tabel kijkt, zie je dat dat precies de rijen zijn waarvoor geldt: 'nr = gironr'. Dit wordt dus als voorwaarde aan de rijen gesteld. Het resultaat is dan precies wat we willen hebben:

naam	saldo
G.Pas	2165
J.Rap	-23
J.Maat	-23
T.E.Bros	885

Door de beperkte afmetingen van deze kleine GIRO past de grote zinloze jointabel nog juist op het scherm en kan de leerling zien wat er precies is gebeurd en wat de betekenis is van de selectie-opdrachten voor deze tabel. Zo is het mogelijk het moeilijke begrip 'join' uit de informatica, voor leerlingen duidelijk te maken.

Wie leende 'Een liefde' en wanneer moet dat terug zijn?

BIBLIOTH is een database van een kleine bibliotheek, maar toch nog zo groot dat de tabellen niet meer overzichtelijk op het scherm of op het papier passen. De structuur is weergegeven in het volgende schema:

BIBLIOTH	
TABELNAAM	KOLOMMEN
BOEKEN	boeknr,titel,autnr,jaar,taal,rubriek.
LENERS	lenersnr,naam,adres,plaats.
AUTEURS	autnr,naam,geboorte,sterfte,natio.
UITLEEN	boeknr,lenersnr,terug.

Ook hier bestaan er weer allerlei relaties tussen de tabellen.

Het beantwoorden van de vraag in de aanhef van deze paragraaf vraagt enig nadenken vooraf. Zomaar iets intypen levert wel wat op, maar dat is zelden het goede resultaat. Je ziet leerlingen dan ook vaak druk overleg plegen over zo'n probleem. Het is weer het handigst eerst te bedenken wat je zou moeten doen als je het allemaal met de hand zou moeten uitzoeken. Je zoekt eerst de titel op, daarbij het boeknr en via dat boeknr zoek je het bijbehorende lenersnr en de datum van terugbrengen. Via het lenersnr vind je dan bij LENERS de bijbehorende naam. Dat betekent dat je uit de drie tabellen BOEKEN, LENERS, UITLEEN moet selecteren, daaruit de kolommen 'naam' (van de lener) en 'terug', met de voorwaarde dat het boeknr (van BOEKEN) gelijk is aan het boeknr (van UITLEEN) en het lenersnr (van LENERS) gelijk is aan het lenersnr (van UITLEEN). Hier doet zich een extra moeilijkheid voor dat er kolommen zijn in de verschillende tabellen die dezelfde naam hebben. Dat is natuurlijk opzet. Maar het programma weet geen raad met de voorwaarde 'boeknr = boeknr' omdat niet duidelijk is welk boeknr wordt bedoeld. Daarom hoort 'boeknr' voorzien te worden van de naam van de tabel waar de kolom bijhoort. Dat doe je door de tabelnaam ervoor te zetten met een punt als scheidingssteken: 'uitleen.boeknr' en 'boeken.boeknr'.

Helemaal volledig ziet de opdracht er dan als volgt uit:

Selecteer uit de tabellen: boeken, leners, uitleen

De kolommen: leners.naam, terug

Waarvoor geldt: boeken.boeknr = uitleen.boeknr en leners.lenersnr = uitleen.lenersnr.

In feite staat hier een fragment van de vraagtaal SQL. SQL is een programmeer-taal die de mogelijkheid biedt om op systematische wijze gegevens uit relationele databases te selecteren. Het hierboven beschreven gedeelte van het programma is een aanpassing van deze vraagtaal aan de Nederlandse situatie in het voortgezet onderwijs. Dat betekent dat we de opdrachten uit het Engels vertaald hebben, dat we een keuze hebben gemaakt uit de veelheid aan opdrachten in SQL en dat we een menu-structuur hebben aangebracht zodat leerlingen niet teveel hoeven op te zoeken en in te typen.

Proefwerk in havo 4

Met ingang van het schooljaar '89/'90 zijn experimenten gestart in drie havo-klassen, wiskunde A, op de twee Hawex-scholen in de omgeving van Groningen. In het eerste halfjaar hebben de leerlingen met het statistische gedeelte van het pakket leren werken in aansluiting op de statistiek van het Hawex-materiaal. Dit eerste blok is afgesloten met een proefwerk waarbij de leerlingen de computer konden gebruiken.

Proefwerk SORBet havo 4, wiskunde A

De spelers van de voetbalclub 'FC DIOW' volgen aan het begin van het seizoen een trainingsprogramma van een maand om de conditie te verbeteren. Om te kijken wat het effect is van dit programma, leggen de spelers aan het begin en aan het eind de zogenaamde 'Harvard stap'-test af. Ze moeten daarbij 30 keer per minuut een kist van 50 centimeter hoog op en af stappen en dat vijf minuten lang zien vol te houden. Aan het eind van de test wordt drie keer om de minuut de hartslag gemeten. Daarmee wordt de 'Harvard-index' uitgerekend:

$$\text{Harvard Index} = \frac{\text{tijdsduur van de test in seconden} \times 100}{2 \times \text{de som van de drie hartfrequenties}}$$

In de tabel TEST van de database CONDITIE staan de resultaten van de metingen voor de 31 spelers. In de kolom 'test 1' staan de resultaten van de test aan het begin van de training. In de kolom 'test 2' staan de resultaten van de test aan het eind van de training.

OPDRACHT 1:

- Stel dat Simon halverwege de eerste test afknapt en dat hij bij de tweede test wel de vijf minuten volmaakt en dat zijn hartslag bij beide testen hetzelfde was. Verklaar hiermee dat een hoge Harvard-index overeenkomt met een goede conditie.
- Een slechte conditie komt overeen met een hoge hartslag. Verklaar hiermee dat een hoge Harvard-index overeenkomt met een goede conditie.

OPDRACHT 2:

Bij het vergelijken van beide testen wil je een overzicht hebben van het verschil van de resultaten.

Selecteer de tabel TEST uit de database CONDITIE. Definieer een nieuwe vierde kolom 'verschil' met Maak/wijzig tabel definitie van tabel WERKTABEL, bereken dan de waarden van de nieuwe kolom met verAnder rijen van tabel (van de werktabel) en de formule: verschil = test 2 - test 1 (achter Maak:).

- Bij welke speler is het verschil het grootste, bij welke het kleinste?
- Wat is het gemiddelde verschil?

OPDRACHT 3:

Neem de volgende tabel over en vul die in:

	gemiddelde	spreiding	mediaan	kwartielen
test 1				
test 2				

OPDRACHT 4:

- ▶ Vergelijk de resultaten van test 1 en test 2 met behulp van repestelen (boxplot-grafieken).
 - a) Welke scores zijn de uitschieters bij test 2?
 - b) Is de conditie voor alle spelers in dezelfde mate veranderd? Waarom? (Geef een verklaring met behulp van de repestelen).

OPDRACHT 5:

Voor deelname aan de competitie geldt de eis dat een speler een score van 100 of meer moet hebben op test 2. Kan trainer Piet Fit een elftal voor de competitie samenstellen? Beschrijf hoe je deze vraag hebt beantwoord.

OPDRACHT 6 (extra):

Piet Fit presenteert vol trots de resultaten aan het bestuur om te laten zien hoe goed zijn training wel niet is geweest. Het bestuurslid S. Jagrijn heeft een bijzonder grote hekel aan trainer Fit. Hij beweert dat het lijstje helemaal niet aantoonde dat het positieve verschil tussen test 2 en test 1 een gevolg is van het trainingsprogramma, maar dat het betere resultaat een gevolg is van andere oorzaken. De overige bestuursleden snappen dat niet zo goed.

- ▶ Geef argumenten die de mening van S. Jagrijn ondersteunen.

In opdracht 1 hebben de leerlingen nog geen computer nodig. Het gaat er in eerste instantie om of ze die formule kunnen interpreteren en inzien dat door het feit dat de hartslag in de noemer staat, een kleinere hartslag een grotere index oplevert. Opdracht 2 is nog een hele gesloten opdracht. De leerlingen wordt precies verteld wat ze met de computer moeten doen en welk resultaat ze krijgen. De vragen zijn eigenlijk alleen maar controlevragen om na te gaan of ze er ook uit zijn gekomen.

Bij opdracht 3 zullen ze blij moeten geven van eigen initiatief in het omgaan met het programma. Ze moeten weten welke opdrachten van het programma de gevraagde resultaten opleveren.

Het begrip boxplot (repesteel) wordt in opdracht 4 getoetst. Leerlingen moeten een zinvolle interpretatie aan het plaatje kunnen geven om de vragen te kunnen beantwoorden.

Ten opzichte van opdracht 3 heeft opdracht 5 nog een grotere mate van openheid. Leerlingen zullen nu zelf moeten bepalen welke gegevens relevant zijn voor de oplossing van het probleem en hoe ze die gegevens met de computer boven tafel kunnen krijgen.

Met opdracht 6 kunnen leerlingen extra punten verdienen. De opdracht is bedoeld om na te gaan in hoeverre leerlingen in staat zijn op een kritische wijze te kijken naar het trekken van conclusies uit statistische gegevens.

In het proefwerk wordt ook getoetst of de leerlingen handig met de computer en met het programma kunnen omgaan. In de eerste plaats is er natuurlijk de beperkte tijd waarin de opdrachten moeten worden gemaakt. Onhandige leerlingen hebben daardoor

automatisch een lager cijfer. Daarnaast bestond de mogelijkheid om extra steun van de docent te 'kopen'. Hulp van de docent betekende een paraaf bij de opgave waardoor een punt (van de 100 punten) per paraaf werd afgetrokken.

De docent van de ene school reageerde enthousiaster dan de docent van de andere school. Het is niet goed duidelijk waardoor dat verschil wordt veroorzaakt. De cijfers vertonen het normale beeld van een proefwerk. Daar valt zo globaal niet veel uit af te leiden. Om conclusies over het leereffect te kunnen trekken is een grondige analyse nodig van de antwoorden en dat valt buiten het kader van dit artikel.

Het schoolonderzoek vwo

De experimenten in het vwo zijn gestart in 1988 en lopen dus een jaar vooruit op die van de havo. Dat betekent dat dit schooljaar een eerste schoolonderzoek wordt gehouden waarbij de computer gebruikt wordt. De inhoud van dit schoolonderzoek is uiteraard gekoppeld aan de statistiek, in het bijzonder correlatie en regressie. In opzet is het schoolonderzoek vergelijkbaar met het proefwerk van de havo. De leerlingen worden geconfronteerd met een aantal opdrachten, variërend in openheid en moeilijkheidsgraad, die met behulp van de computer kunnen worden aangepakt. De mogelijkheid om hulp bij de docent te kopen vervalt hier natuurlijk. De tijdsduur van het praktische gedeelte is 75 minuten, gevolgd door een half uur waarin de leerlingen de mogelijkheid wordt geboden de opdrachten in een nette vorm uit te werken. Alle opdrachten hebben betrekking op een relationele database die de resultaten van het examen bevat van twee scholen, inclusief gegevens over de docenten waarvan leerlingen les hebben gehad. Binnen deze context is het mogelijk om de beheersing van begrippen uit de relationele database-theorie en de statistiek in samenhang te toetsen. Een mooiere vorm van integratie tussen informatica en wiskunde is nauwelijks denkbaar!

Bij het schrijven van dit artikel waren de resultaten van één van de twee scholen bekend. Daar hebben 29 leerlingen aan het praktisch schoolonderzoek meegeëerd. Vanwege het beperkte aantal computers in een klaslokaal, zijn er drie zittingen georganiseerd. Door de tweede zitting direct achter de eerste te plannen kon volstaan worden met twee versies van het schoolonderzoek. Het behaalde resultaat gaf samen met een theoretisch schoolonderzoek één schoolonderzoekcijfer. In totaal kregen de leerlingen voor wiskunde A drie schoolonderzoekcijfers. Enkele karakteristieken van de behaalde resultaten zijn:

gemiddelde cijfer:	6.2
standaarddeviatie:	2.3
laagste/hogste cijfer:	2.5/10
aantal cijfers ≤ 5.5 :	10

Het experiment

In het voorgaande is tussen de regels door al iets geschreven over het uitproberen van het programma en bijbehorende lesmateriaal. In september 1988 is het project gestart. Het computerprogramma moest toen nog helemaal worden geschreven. Iedere keer als er

een stukje klaar was, werd dat uitgetoet in een viertal klassen van vwo-5, wiskunde A. Daardoor ontstond een situatie waarin de klasse-ervaringen direct verwerkt konden worden bij het programmeren. Dit had tot gevolg dat er in blokken gewerkt moest worden. In klas 5 is er zo gewerkt in drie blokken van vier aansluitende uren. Achteraf bleek dat een hele gelukkige keuze. Leerlingen hielden het na vier lessen wel weer voor gezien en gaan weer monter aan de gewone wiskunde. Bovendien is dit voor de scholen veel gemakkelijker te organiseren. Ook bij de tweede ronde in vwo-5 is dit schema van drie blokken van vier uur aangehouden. In het schooljaar '89/'90 is op twee experimentele Hawex-scholen met het experiment in klas 4, wiskunde A, begonnen. Het is de bedoeling ook hiermee tot een experimenteel schoolonderzoek in klas 5 te komen.

De lessen die aan het experiment worden besteed vallen gewoon binnen de wiskunde A-lessen. In feite worden er dus onderwerpen aan het examenprogramma toegevoegd. Het is echter de verwachting dat er door het gebruik van de computer minder lessen statistiek nodig zijn omdat de grafische mogelijkheden sneller en gevarieerder beschikbaar zijn. Zelf een histogram laten tekenen met een aantal verschillende klasse-indelingen is een kwestie van minuten. In het traditionele statistiekonderwijs is zo iets praktisch nauwelijks haalbaar. Hierin bewijst zich duidelijk de meerwaarde van de computer.

In het volgende overzicht wordt de situatie van het experiment nog eens samengevat:

- * *6 Atheneum:*
 - gestart in '88,
 - 4 klassen, 2 scholen, 4 docenten
 - eerste versie lesmateriaal
 - experimenteel schoolonderzoek in maart '90.
- * *5 Atheneum:*
 - gestart in '89,
 - 7 klassen, 4 scholen, 7 docenten,
 - tweede versie lesmateriaal.
- * *4 havo (Hawex):*
 - gestart in '89,
 - 3 klassen, 2 scholen, 3 docenten,
 - eerste versie lesmateriaal.

Het is nog onduidelijk hoe het verder gaat met het experiment. In ieder geval moeten de reeds gestarte experimenten tot een goed schoolonderzoek worden gebracht. Daarnaast is het wenselijk om ook in de havo een tweede ronde op te starten zodat het lesmateriaal nog eens in een tweede versie kan worden uitgetoet. De ervaring leert dat een eerste versie in de praktijk nog zeer primitief is, hoe goed die ook is doorgedacht, en dat je pas bij het ontwikkelen van een tweede versie succesvolle lijnen in het materiaal gaat zien.

Ook zijn er allerlei aanpassingen aan het programma nodig. De huidige versie is goed te hanteren in de school, maar de gebruikersvriendelijkheid en de snelheid kunnen nog aanzienlijk worden verbeterd.

Het zou jammer zijn als het daarmee is afgelopen. Er is veel tijd en geld gestoken in de ontwikkeling van materiaal dat een zinvolle bijdrage kan leveren aan het gebruik van de computer bij wiskunde A. Zolang die

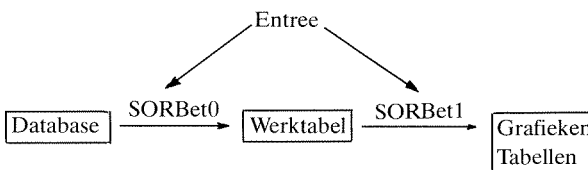
computer echter niet op een of andere manier deel uitmaakt van het eindexamen, is het gevaar groot dat al het werk en geld in een stoffige hoek blijft liggen.

Conclusies

- * Het blijkt goed mogelijk om in klas 5 van het vwo een geïntegreerde vorm van wiskunde en informatica te realiseren aan de hand van de combinatie van statistiek en relationele bestanden.
- * Het blijkt haalbaar een proefwerk te geven waarbij de computer gebruikt wordt, al vraagt dat wel enige organisatie.
- * Er zijn situaties waarin je beter eerst iets met je hoofd en handen kunt doen vóór je de computer gaat gebruiken. Er zijn ook situaties waarin je beter eerst wat met de computer gaat modderen vóór je begint te denken. Het is alleen nog niet duidelijk welke didactiek in welke situatie het beste werkt en waarom.
- * In het project is een ontwikkelmodel gebruikt waarin er sprake was van een directe interactie met de docenten en de klas. Dit levert in ieder geval enthousiaste docenten en materiaal op dat direct bruikbaar is in de klas.

Enkele feiten

- * Het citaat aan het begin van dit artikel is ontleend aan het HEWET-rapport, Den Haag, 1980.
- * Het programma SORBbet bestaat uit twee gedeeltes: SORBbet0, het databasepakket, en SORBbet1, het statistiek pakket. Deze beide programma's zijn in één gebruikersomgeving gezet waarin gemakkelijk van het ene naar het andere kan worden gesprongen, maar waarin het ook goed mogelijk is de onderdelen afzonderlijk te gebruiken. Wie bijvoorbeeld alleen geïnteresseerd is in statistische verwerking van gegevens, kan heel goed met alleen SORBbet1 uit de voeten. In feite ziet het pakket er als volgt uit:



- * Het programma SORBbet1 is een aangepaste versie van het pakket STatiON dat door Nijso B.V. is ontwikkeld bij de statistiekboeken van Nijdam en van Buuren.
- * De tabel EXAMEN is ontleend aan het pakket STatiON, zie: Nijdam, B.; Buuren, H. van: Statistiek, Samsom, 1988.
De database PLNTLUSt is ontleend aan: Eibers, H. B. e.a.: SQL in de praktijk, Academic Service, 1987.
- * Kees Visser en Jaap Helder waren de programmeurs. Ron Jansen, Frits Ferwerda, Ben Scholten, Harrie Oostijen, Paul Laan, Peter van Rijn, Carl Hegeman waren de docenten van de eerste ronde in vwo en havo.