

Statistische Software voor Sociaal-Wetenschappelijk Onderzoek: Een evaluatie vanuit gebruikersperspectief.

Hubert Korzilius en Jan van Leeuwe*

Samenvatting

Evaluaties van statistische software beperken zich veelal tot beoordelingen ten aanzien van criteria die niet rechtstreeks verband houden met de werking van de statistische procedure of het gebruikte algoritme. Bovendien wordt vaak gebruik gemaakt van geconstrueerde datasets waardoor geen reëel beeld wordt verkregen van de werking van statistische software in een concrete onderzoekssituatie. In de onderhavige studie wordt daarom gebruik gemaakt van een reële onderzoeksvraagstelling en reële data die exemplarisch zijn voor sociaal wetenschappelijk onderzoek. Analyseresultaten worden gerapporteerd voor een zestal veelgebruikte statistische pakketten. De resultaten laten een hoge mate van overeenstemming zien. Niettemin vertonen de pakketten lacunes met betrekking tot hun statistische mogelijkheden en zijn op enkele deelgebieden opmerkelijke verschillen te constateren.

* Research Technische Dienstverlening, Pedagogische Wetenschappen en Onderwijskunde, Katholieke Universiteit Nijmegen, Postbus 9103, 6500 HD Nijmegen. Telefoon: 080-512559/512127.

De auteurs zijn dank verontschuldigd aan C.A.J. Aarnoutse van de vakgroep Interdisciplinaire Onderwijskunde voor het ter beschikking stellen van de data en voor zijn kritische commentaar.

Deze studie werd mede mogelijk gemaakt door het beschikbaar stellen van een IBM-PC/AT door IBM-Nederland N.V.

1. Inleiding

In evaluaties van statistische software worden vaak tientallen criteria gehanteerd voor de vergelijking van programma's en programmapakketten. Deze criteria kunnen veelal worden ondergebracht in een viertal globale categorieën: 1. benodigde hardware, 2. bedieningsgemak, 3. arsenaal aan statistische procedures en 4. rekensnelheid en nauwkeurigheid. De lezer wordt geconfronteerd met een veelheid aan tabellen waarin de onderzochte pakketten op deze criteria worden beoordeeld. In Keller (1986) wordt deze werkwijze ten aanzien van een elftal statistische pakketten voor IBM-compatibele personal computers in praktijk gebracht. Voor de Apple Macintosh werden door Lehman (1987) langs deze weg een twaalfal statistische pakketten vergeleken. Soortgelijke evaluaties zijn te vinden in Musegaas (1984), Gaaff en de Boer (1985), Bakker (1986), Fridlund (1986) en Lehman (1986).

Hoewel de gedetailleerde informatie voor de specialist interessant is, zal de onderzoeker, werkzaam in een toepassingsgebied, al gauw geneigd zijn deze informatie over te slaan en slechts de conclusies van de betreffende artikelen te betrekken bij de keuze van de statistische software. Daarin schuilt het gevaar dat de keuze van de statistische procedure voor een concreet onderzoek niet primair bepaald wordt vanuit de onderzoeksvraagstelling, maar vanuit de beschikbaarheid van een statistisch pakket in combinatie met de beschikbare hardware. Uiteraard stemmen de resultaten van dezelfde statistische procedures in de diverse statistische pakketten veelal, op rekenonnauwkeurigheid na, met elkaar overeen. De berekening van een rekenkundig gemiddelde zal voor alle pakketten waarschijnlijk hetzelfde resultaat opleveren. Toch kan men al snel verschillen constateren, bij de standaarddeviatie veroorzaakt deling door N of door $N - 1$ reeds afwijkingen. Als voorbeeld noemen wij tevens de behandeling van ontbrekende scores. Diverse pakketten hebben slechts de mogelijkheid van "listwise deletion" (verwijdering van onderzoekseenheden die op één of meer van de variabelen een ontbrekende score hebben). Met name in sociaal-wetenschappelijk onderzoek kan het vanuit de onderzoeksopzet noodzakelijk zijn in dit opzicht een andere procedure te hanteren. Dit aspect zou daarom van doorslaggevend betekenis kunnen zijn bij de keuze van een pakket. Zelfs de gedetailleerde artikelen van Keller en Lehman geven onvoldoende informatie over de werking van de pakketten in dit opzicht.

Een nog serieuzer gevaar schuilt in de uitwerking van de statistische procedures. Evaluatiestudies beperken zich veelal tot de constatering dat een bepaalde statistische procedure in een pakket is geïnstalleerd en geven daarbij soms een opsomming van het arsenaal aan geprogrammeerde statistische grootheden. De berekeningswijze van deze

statistische grootheden blijft vaak onbeschreven. Dit kan tot problemen aanleiding geven. Als voorbeeld beschrijven wij kort de resultaten van een tweevoudige variantieanalyse, zoals geprogrammeerd in een drietal pakketten. De data zijn ontleend aan Winer (1971, pag. 446) en betreffen een 2x4-design met ongelijke cel aantallen. Gebruikt werden de pakketten STATWORKS, STATVIEW en SYSTAT voor de Apple Macintosh. De resultaten van STATVIEW en SYSTAT kwamen geheel overeen. Voor het eerste hoofdeffect (A) werd in STATWORKS een F-waarde van 0.157 berekend, bij STATVIEW en SYSTAT was de F-waarde 0.022, terwijl de door Winer gegeven waarde 0.014 bedraagt. Voor het tweede hoofdeffect (B) berekent STATWORKS een waarde van 17.713, STATVIEW en SYSTAT komen uit op 18.696, terwijl Winer een waarde van 18.99 geeft. Nadere studie leert dat deze verschillen terug te voeren zijn tot verschillen in het algoritme dat gebruikt wordt voor de aanpassing van ongelijke cel aantallen. Het is hier niet de plaats om op dit probleem gedetailleerd in te gaan. Wel kan worden geconstateerd dat de uitwerking van de statistische procedures aanzienlijke verschillen mogelijk maakt. Grondige bestudering van de gebruikte procedures en hun onderliggende assumpties blijft geboden.

Een verdere beperking van het genoemde type van evaluatiestudies is dat de gebruikte data vaak geconstrueerd zijn of gekozen worden op grond van specifieke kenmerken. Zo worden de Longley (1967) data veelvuldig gebruikt voor het testen van programma's voor regressieanalyse omdat de correlatiematrix bijna singulier is. De werking van programma's op dergelijke datasets is evenwel niet representatief voor reëel onderzoek. In sociaal-wetenschappelijk onderzoek worden de resultaten van een eerdere analyse nogal eens gebruikt bij een volgende analysestap. Zo dienen factorscores vaak als inputvariabelen bij een volgende stap. Het genoemde type evaluatiestudies houdt met problemen die hierdoor kunnen ontstaan geen rekening.

De onderhavige evaluatiestudie heeft tot doel de resultaten van een reëel stuk onderzoek te vergelijken vanuit het perspectief van de onderzoeker. Eerst wordt nagegaan of een concreet pakket de statistische mogelijkheden bevat die uit de vraagstelling voortvloeien. Vervolgens wordt bekeken of de resultaten en daarmee de conclusies met elkaar in overeenstemming zijn. Dit gezichtspunt zou bepalend moeten zijn bij de keuze van statistische software. Aspecten als beschikbaarheid van hardware en gebruikersvriendelijkheid zijn belangrijk, maar hieraan ondergeschikt.

Bij de uitwerking van een reëel stuk onderzoek staat de methodologie en de daaruit voortvloeiende keuze van statistische technieken uiteraard centraal. Het is nodig om bij de uitwerking van de onderzoeksvraagstelling aan dit keuzeproces ruime aandacht te besteden. Omdat de doelstelling van deze studie echter is gelegen in de evaluatie van

statistische pakketten is een discussie over de in dit opzicht genomen beslissingen in dit artikel achterwege gelaten.

2. De beschikbare pakketten

De keuze van de pakketten.

Vanuit bovenstaand perspectief hebben we bij de keuze van de pakketten niet de beschikbaarheid van hardware laten prevaleren. De keuze van de pakketten is primair bepaald door de te verwachten kwaliteit voor sociaal-wetenschappelijk onderzoek en secundair door de relatieve gebruiksfrequentie van pakketten in onze onderzoeks-omgeving. Daarom beperken we ons tot zes pakketten die in dit opzicht in het recente verleden hun waarde reeds hebben bewezen. Deze pakketten zijn: SPSSx, SAS, SPSS-PC+, SAS-PC, SYSTAT en STATVIEW. Met deze keuze zijn drie verschillende typen computers vertegenwoordigd. Twee pakketten die op een IBM/VM-CMS mainframe draaien: SPSSx en SAS, twee bestemd voor een personal computer (afgekort: PC) van het soort IBM-compatibele: SPSS-PC+ en SAS-PC en twee voor de PC Apple Macintosh: SYSTAT en STATVIEW.

De pakketten.

Op IBM/VM-CMS mainframe:

SPSSx (versie 3.0).

SPSSx is een veelgebruikt standaardpakket voor de mainframeomgeving. Het bevat een omvangrijk arsenaal aan statistische procedures. De uitgebreide stuurtaal is evenwel in korte tijd aan te leren. In versie 3.0 zijn alle procedures met zogenaamde keywords aan te roepen. Programma's kunnen zowel interactief als batchgewijs worden aangeboden. SPSSx is voornamelijk gericht op data-analyse, minder op bijvoorbeeld data-invoer en databasetoepassingen.

SAS (versie 5.16).

SAS is ook een veelgebruikt standaardpakket voor statistische analyse op mainframes. Naast omvangrijke mogelijkheden voor statistische analyse bevat SAS ook faciliteiten voor databasemanagement, matrixbewerkingen, simulatie van data en grafische analyse. Het werken met zogenaamde datasteps maakt het aanleren van SAS nogal gecompliceerd.

Op IBM-compatibele PC:

SPSS-PC+ (versie 2.0).

SPSS-PC+ is de PC-uitvoering van SPSSx. SPSS-PC+ is niet zo uitgebreid als zijn grote broer. Voor PC-begrippen is het echter een fors statistisch pakket. Omdat het pakket ca. 7 Mb aan schijfruimte beslaat is een harde schijf op de PC noodzakelijk. In het pakket zijn de gangbare statistische technieken voorhanden. SPSS-PC+ werkt snel, nauwkeurig en prettig en geeft duidelijke foutmeldingen. Met versie 2.0 kan men volledig menugestuurd te werk gaan, terwijl batchgewijze aansturing ook mogelijk is. SPSSx en SPSS-PC+ zijn met enkele aanpassingen goed met elkaar uitwisselbaar.

Voor een bespreking van SPSS/PC+ (versie Update 1987), met daarin onder andere een vergelijking met SPSSx, raadplege men Bouts en Korzilius (1988).

SAS-PC (versie 6.03).

Wat voor SPSS-PC+ en SPSSx geldt, gaat mutatis mutandis ook op voor SAS-PC en SAS-mainframe. SAS-PC is een groot statistisch pakket. Een harde schijf is noodzakelijk. Het werken met het zogenaamde display manager system (dat is de werkomgeving van SAS-PC met daarin schermen voor de invoer en uitvoer van van SAS-programma's) draagt bij tot het gebruiksgemak. Uitwisseling tussen mainframe en PC gaat met behulp van een 'micro-to-mainframe link': programma's kunnen afhankelijk van de statistische procedure, de grootte van de datasets en benodigde rekentijd op mainframe danwel op PC worden uitgevoerd.

SAS-PC wordt besproken in Korzilius (1986).

Op Apple Macintosh:

In 1986 schrijft Keller nog: "De Apple Macintosh is een prachtige machine, doch de hoeveelheid statistische programmatuur is zeer beperkt." (Keller, 1986, p. 5). Deze visie konden we toen al en kunnen we zeker nu, enkele jaren later, weerleggen. Voor de Macintosh is een scala aan statistische software voorradig. Zie daartoe bijvoorbeeld Lehman (1986, 1987).

De Apple Macintosh kenmerkt zich door een andere 'userinterface' dan de IBM-compatibele PC. Dit komt tot uiting in het gebruik van pull-down menu's, vensters en muisbesturing.

SYSTAT (versie 3.0).

SYSTAT is een pakket met vergelijkbare mogelijkheden als SPSS-PC+ en SAS-PC. Het wordt gekenmerkt door een modulaire opbouw, hetgeen inhoudt dat de verschillende statistische procedures in aparte delen zijn ondergebracht. Een harde schijf is handig maar niet noodzakelijk. De aansturing gaat interactief maar het is ook mogelijk om programma's batchgewijs aan te bieden. Een nadeel van SYSTAT is dat het de Macintosh userinterface slechts beperkt ondersteunt.

SYSTAT voor de Apple Macintosh is beschreven in Korzilius (1988).

STATVIEW (versie 512+ 1.0).

STATVIEW is een statistisch pakket voor de Macintosh dat volledig gebruik maakt van de userinterface. De analyses worden uitgevoerd met behulp van menu's waarbij nauwelijks instructies behoeven te worden getypt. STATVIEW bevat minder statistische procedures dan de drie hierboven beschreven PC-pakketten. Een harde schijf is niet nodig. Een gebruiker die niet al te geavanceerde statistische technieken nodig heeft kan met dit pakket goed uit de voeten. Door de eenvoudige aansturing is STATVIEW met name bruikbaar voor onderwijs in de statistiek. Nadelig is het feit dat er geen mogelijkheid is om analyses via batch-verwerking aan te sturen. Meerdere gelijksoortige analyses moeten stuk voor stuk worden uitgevoerd.

Een handleiding van STATVIEW vinden we in Korzilius (1987).

3. De onderzoeksopzet

Voor de vergelijking van de resultaten van de onderzochte pakketten werd gebruik gemaakt van een reële onderzoeksvraagstelling. In het kader van het project "Leeractiviteiten en processen tijdens het onderwijs in begrijpend lezen" van de Facultaire vakgroep Onderwijskunde (Aarnoutse en Weterings, 1989) werden in 1987 op 21 basisscholen in de regio Nijmegen gegevens over begrijpend lezen verzameld. De totale steekproef omvatte 602 leerlingen uit groep 7 van het basisonderwijs. Het onderzoek omvatte naast observaties bij lezende kinderen ook de afname van een aantal tests en schalen met betrekking tot:

- a. De cognitieve aspecten van (begrijpend) lezen.
- b. Het oordeel van de leerkracht over de lees- en taalvaardigheid van de leerling.
- c. De attitude ten aanzien van lezen van de leerling.

De globale onderzoeksvraagstelling binnen het genoemde project is verder uitgewerkt in een vijftal onderzoeksvragen. Voor het onderhavige onderzoek beperken we ons tot een gedeelte van de binnen het project gedefinieerde vierde onderzoeksvraagstelling. Deze luidt:

4. Wat is bij begrijpende lezers de samenhang tussen cognitieve processen en leesprestaties en attitudes?

Daarbij zien wij af van de kwalitatieve analyse van de cognitieve processen. Slechts de neerslag ervan in de vorm van de resultaten op de tests voor begrijpend lezen is onderwerp van onderzoek. Verder richten wij ons met name op de verklaring van de leesattitude vanuit de prestaties op de tests voor begrijpend lezen en het oordeel van de leerkracht. Daarbij hebben we ook de verschillen tussen de scholen en de verschillen tussen allochtone en autochtone kinderen ten aanzien van de leesattitude betrokken.

Door deze inperkingen wordt slechts een bepaald aspect van de breder gedefinieerde onderzoeksvraagstelling in de beschouwing betrokken. Voor de vergelijking van statistische pakketten biedt deze toegespitste vraagstelling echter geschikte mogelijkheden. Dit type vraagstelling komt veel voor bij sociaal-wetenschappelijk onderzoek, met name in de pedagogische wetenschappen en de onderwijskunde. Er is sprake van een beperkt aantal afhankelijke variabelen, die vanuit een aantal onafhankelijke variabelen moeten worden verklaard. Daarbij worden twee soorten onafhankelijke variabelen onderscheiden: predictoren en omgevingsvariabelen. Predictoren worden verondersteld, al of niet in onderlinge samenhang, de mogelijke verklaringsgrond voor de afhankelijke variabele te leveren. Omgevingsvariabelen worden verondersteld een nadere specificatie van de onderzoeksgroep ten aanzien van de afhankelijke variabele of ten aanzien van de relatie tussen onafhankelijke en afhankelijke variabele op te leveren.

Afhankelijke variabele in het onderhavige onderzoek is de leesattitude. Voor de meting van de leesattitude werd gebruik gemaakt van de Leesattitudeschalen van Bisschop, Aarnoutse en Feenstra (1985). Deze meten de houdingen van de leerlingen ten aanzien van lezen en leesmateriaal. In concreto werd gekozen voor afname van de Leesattitudeschaal 2 (LAS2). Deze Likert-schaal omvat 18 items (5-puntsschalen).

De predictoren kunnen worden opgesplitst in cognitieve aspecten met betrekking tot (begrijpend) lezen en de lees- en taalvaardigheid van de leerling (geoperationaliseerd in het oordeel van de leerkracht). De cognitieve aspecten zijn gemeten door middel van de volgende meetinstrumenten:

- De toets Begrijpend Lezen van het CITO (1981), verder afgekort als CITO.
- De Synoniementest van Aarnoutse (1987), verder afgekort als SYNO.
- De Hoofdgedachtetest van Aarnoutse (1984), vorm A, verder afgekort als HGD.
- De Eén-Minuut-Test van Brus en Voeten (1972), vorm A, verder afgekort als EMT.

De toets Begrijpend Lezen, bestemd voor het midden van groep 7, meet het begrijpend lezen in het algemeen. De toets bestaat uit 25 items bij enkele informatieve en verhalende teksten.

De Synoniementest beoogt het begrijpen van de betekenis van geschreven woorden te meten. De test bestaat uit 30 items. Bij elk item moeten de leerlingen uit vier woorden dat woord kiezen dat het beste overeenkomt met de betekenis van het aangeboden woord.

De Hoofdgedachtetest is een taakgerichte test die een bepaald aspect van het begrijpend lezen meet, namelijk het afleiden van een hoofdgedachte uit een informatieve tekst. De test bestaat uit 29 items.

De Eén-Minuut-Test meet de vaardigheid in het ontsleutelen (decoderen) van gedrukte woorden. De test bestaat uit een kaart met 116 niet-samenhangende woorden. De ruwe score wordt gevormd door het aantal goed gelezen woorden in één minuut.

Het oordeel van de leerkracht over de taal- en leesvaardigheid van de leerlingen werd geregistreerd door een beoordeling van de leerkracht op een vijf-puntsschaal. Op deze wijze werd het oordeel van de leerkracht gevraagd ten aanzien van de vorderingen van de leerling op de volgende aspecten:

- Technisch Lezen, verder afgekort als TL.
- Begrijpend Lezen, verder afgekort als BL.
- De gemotiveerdheid van de leerling ten aanzien van lezen, verder afgekort als LM.
- De mondelinge taalvaardigheid, verder afgekort als MT.

De uitwerking van de onderzoeksvraagstelling verloopt in drie fasen. In de eerste fase wordt nagegaan of de leesattitudeschaal één homogene trek meet zoals bij de constructie van deze schaal de bedoeling was en voor de ijkingssteekproef in 1985 werd aangetoond. Omdat de leesattitude de afhankelijke variabele is in dit onderzoek is het belangrijk om te weten of gesproken kan worden van één algemene leesattitude of dat meerdere aspecten moeten worden onderscheiden. Mogelijke effecten betrekking hebbend op sociaal-culturele verschuivingen in de tijd danwel op specifieke kenmerken van de in dit onderzoek gebruikte steekproef mogen niet apriori worden uitgesloten. In de tweede fase wordt de leesattitude (afhankelijk van de resultaten van de eerste fase uitgedrukt in één danwel in meer scores) als afhankelijke variabele ingevoerd in regressieanalyses waarbij

de leestests en het oordeel van de leerkracht als onafhankelijke variabelen werden ingevoerd. Deze opzet is gekozen om de relatieve bijdrage van beide typen variabelen tot uiting te laten komen. In de derde fase tenslotte wordt nagegaan of de leesattitude verschilt per school en of er verschil bestaat tussen autochtone en allochtone leerlingen (herkomst). Voor de analyse van deze twee variabelen apart is gekozen voor enkelvoudige variantie-analyse.

Het bovenstaande leidt tot het volgende overzicht van de uit te voeren analysestappen:

A. Fase 1.

Factorstructuur Leesattitude-schaal (veronderstelling: eendimensionaal).

1. Iteratieve factoranalyse.
2. Principale Componentenanalyse.
3. Berekening factorscores en betrouwbaarheid.

B. Fase 2.

Invloed leesprestaties (tests + oordeel leerkracht) op leesattitude.

1. Normaliteitscontrole factorscore LAT.
2. Regressieanalyse met 8 onafhankelijke variabelen.
3. Bloksgewijze regressieanalyse.
4. Regressieanalyse op minimale verzameling predictoren.

C. Fase 3.

Verschillen tussen scholen en tussen autochtone en allochtone leerlingen t.a.v. LAT.

1. Enkelvoudige Variantieanalyses.
2. Kruistabel school - herkomst.
3. Gemiddelden per school uitgesplitst naar herkomst.

4. Resultaten per pakket

4.1. Inleiding.

In dit deel worden de resultaten van de verschillende analysestappen bij gebruik van de zes statistische pakketten naast elkaar gelegd. Vooral de verschillen tussen de pakketten worden uitgebreid besproken.

4.2. Resultaten per onderzoeksvraag.

A. Factorstructuur leesattitude-schaal.

De eerste stap in de data-analyse is de controle van de verdeling van de variabelen. Dit wordt gedaan middels frequentieverdelingen en de bepaling van enkele beschrijvende statistische maten zoals het minimum, maximum, gemiddelde en standaarddeviatie. Alle pakketten geven wat dit betreft gelijke resultaten.

A.1. Iteratieve factoranalyse.

Bij de uitvoering van factoranalyse hebben we allereerst gekeken naar de wijze waarop de correlatiematrix daarin wordt berekend. In STATVIEW is het niet mogelijk via pairwise deletion met ontbrekende scores rekening te houden, slechts listwise deletion is beschikbaar. In de andere pakketten zijn beide mogelijkheden geïmplementeerd. De genoemde correlatiematrix blijkt, gebruik makend van listwise deletion, voor alle pakketten identiek te zijn. Bij de bepaling van de factorstructuur van de variabelen, wordt het vergelijken van de pakketten wat problematischer. Met name SYSTAT en STATVIEW zijn in deze afwijkend.

Een veelgebruikte techniek om factoren te extraheren, is 'principal axis factoring' ook wel iteratieve factoranalyse genoemd. SYSTAT (versie 3.0) kent echter alleen 'principale componenten analyse'.

Met STATVIEW zijn er andere problemen. Het lijkt erop dat dit pakket nog enkele onvolkomenheden (zogenaamde bugs) in de procedure Factor Analyse heeft. Uit nader onderzoek blijkt dat STATVIEW in een aantal gevallen problemen kent bij het roteren van een factorstructuur naar een simpele structuur. In ons onderzoek werd de varimaxrotatie door STATVIEW niet uitgevoerd, als varimaxoplossing werd de ongeroteerde factor-matrix nogmaals vermeld.

Bij de iteratieve factoranalyse komen de andere pakketten voor en na iteratie tot een overeenkomstige factor(patroon)matrix, met gelijke communaliteiten, eigenwaarden en proporties verklaarde variantie. Zie tabel 1.

Tabel 1: Iteratieve factoranalyse: eigenwaarde en proportie verklaarde variantie.

verklaarde	beginwaarden		na iteratie	
	eigenwaarde	proportie verklaarde	eigenwaarde	proportie
		variantie		variantie
factor 1	5.40	30.0	4.76	26.5
factor 2	1.36	7.5	0.69	3.8
factor 3	1.05	5.8	0.33	1.8

De grafische weergave van de factorstructuur is in geen van de gebruikte pakketten direct interpreteerbaar. In STATVIEW worden de variabelen slechts weergegeven in de vorm van één gemeenschappelijk plotsymbool, de identificatie van de bijbehorende variabelen ontbreekt. In SAS, SAS-PC, SPSSx en SPSS-PC+ is identificatie slechts mogelijk via een bijgevoegde tabel. In geen van de pakketten is het mogelijk de namen of labels van de variabelen direct in de plot weer te geven. Indien deze mogelijkheid zou bestaan, hetgeen technisch vrij eenvoudig realiseerbaarheid lijkt, zou dit aanzienlijke tijdsbesparing opleveren, met name in situaties waarin verschillende factoroplossingen met elkaar worden vergeleken.

De vervaardiging van een grafische weergave in SYSTAT, tenslotte, is slechts mogelijk door het overbrengen van de factormatrix naar een separate plotroutine.

A.2. Principale Componenten Analyse.

Omdat in SYSTAT de mogelijkheid ontbreekt om iteratieve factoranalyse uit te voeren, verrichten we ook een principale componenten analyse om alle pakketten toch met elkaar te kunnen vergelijken.

De initiële schatting van de factormatrix is bij alle pakketten gelijk (eigenwaarden: 5.40, 1.36 en 1.05; proporties verklaarde variantie: 30.0, 7.5 en 5.8; factorpatroonmatrices komen overeen). De eigenwaarden na rotatie worden alleen door SAS, SAS-PC en SYSTAT gegeven: 3.30, 2.59 en 1.91. SYSTAT geeft als enige ook nog de proporties verklaarde variantie na rotatie: 18.35, 14.36 en 10.64. Bij STATVIEW wordt de varimaxrotatie ook bij principale componenten analyse niet uitgevoerd.

Uit het verloop van de eigenwaarden bij de initiële schatting (grafisch weergegeven bij SPSSx, SAS, SPSS-PC+ en SAS-PC door middel van het afzetten van de eigenwaarden tegen de factoren) wordt besloten tot een 1-component oplossing. Wij vonden geen aanleiding om de veronderstelling dat leesattitude één homogene trek is in twijfel te trekken. De factor leesattitude verklaart 30.0% van de variantie die ligt besloten in de 18

variabelen. Na deze vaststelling worden de factorscores voor de leesattitudeschaal berekend.

A.3. Berekening Factorscores en betrouwbaarheid.

De volgende stap is de bepaling van de betrouwbaarheid van de schaal. Alleen in SPSSx is de berekening van de betrouwbaarheid standaard aanwezig. Zelfs in een gerenomeerd pakket als SAS is dit niet standaard ingebouwd. Daarvoor moesten we onze toevlucht nemen tot een minder toegankelijke SAS-programmabibliotheek.

De leesattitudeschaal heeft overigens een betrouwbaarheidscoëfficiënt α van 0.85. Dat is redelijk tot goed te noemen.

Om een beter overzicht te krijgen op de scores van de leesattitudeschaal hebben we de factorscores bewerkt. Eerst zijn er positieve gehele getallen van gemaakt door de factorscores te vermenigvuldigen met 10 en er daarna 50 bij op te tellen. Het resultaat hiervan is vervolgens naar beneden afgerond. Het zou meer voor de hand liggend zijn geweest af te ronden naar het dichtstbijzijnde gehele getal. In SYSTAT bleek dit echter niet rechtstreeks mogelijk. Om de bewerkingen voor alle pakketten zoveel mogelijk uniform te laten verlopen, werd tot afronding naar beneden overgegaan. Deze transformaties resulteren in een variabele (LAT) die loopt van 1 tot en met 75. Na hercodering hiervan komen we tot een variabele bestaande uit 15 klassen met een klassebreedte van 5. Van deze variabele (LATREC) hebben we de diverse pakketten een aantal verdelingsmaten laten berekenen (zie tabel 2 en 3).

Tabel 2: Aantal cases, minimum, maximum, gemiddelde, standaarddeviatie van de bewerkte schaal.

	SPSSx SAS SPSS-PC+ SAS-PC	SYSTAT STATVIEW
aantal cases (N)	594	594
minimum	2	2
maximum	15	15
gemiddelde	10.308	10.315
standaarddeviatie	2.013	2.019

Uit tabel 2 blijkt dat er wat betreft het aantal cases, minimum en maximum geen verschillen zijn tussen de diverse pakketten. Het gemiddelde en de standaarddeviatie van SYSTAT en STATVIEW wijken iets af van de andere vier pakketten.

Tabel 3: Frequentieverdeling van de bewerkte schaal.

waarden:	SPSSx	SYSTAT
	SAS SPSS-PC+ SAS-PC	STATVIEW
	frequentie:	frequentie:
2	1	1
3	1	1
4	2	2
5	5	5
6	15	15
7	27	27
8	52	52
9	83	83
10	117	116
11	124	125
12	90	88
13	50	51
14	26	27
15	1	1
missing	8	8
N	602	602

Ook de frequentieverdelingen van LATREC in SYSTAT en STATVIEW wijken iets af van de frequentieverdelingen in de andere pakketten. Omdat ons niet is gebleken (zie ook paragraaf 4.2 onder A) dat er bij de vervaardiging van frequentieverdelingen en bij de berekening van bijbehorende statistische grootheden in enig pakket fouten optreden moet de oorzaak van de geconstateerde afwijkingen elders worden gezocht. Bij inspectie van de ruwe factorscores blijken die in SYSTAT en STATVIEW steeds verschillen van 4 à 5 honderdste te vertonen met de factorscores als berekend in de andere pakketten. Deze aanvankelijke geringe verschillen en enigszins andere implementatie van de daarop uitgevoerde transformaties leiden tot de gevonden verschillen in de frequentieverdelingen. Hoewel de hier gevonden verschillen gering zijn, wijzen we erop dat cumulatie van geringe afwijkingen in meerdere achtereenvolgende berekeningen tot grotere verschillen kan leiden.

B. Invloed leesprestaties (tests + oordeel leerkracht) op leesattitude.

B.1. Normaliteitscontrole factorscore LAT.

De volgende stap is controle of de variabele de normaalverdeling benadert. We gebruiken hiervoor de variabele LAT. Op dit punt komen we niet met alle pakketten tot gelijke conclusies. In tabel 4 staan de resultaten bij elkaar.

Opmerkelijk is het feit dat STATVIEW geen test op normaliteit bevat. SPSSx en SPSS-PC+ gebruiken beiden de 'Kolmogorov-Smirnov Goodness of fit test' en komen tot gelijke conclusies, namelijk dat er geen significante afwijking van de nulhypothese is: LAT mag normaal verdeeld verondersteld worden.

SAS hanteert Kolmogorov D voor steekproeven groter dan 50, bij kleinere steekproeven wordt de Shapiro-Wilk test gebruikt. SAS-PC maakt gebruik van Shapiro-Wilk als de steekproef kleiner of gelijk is dan 2000 en gebruikt Kolmogorov D bij grotere steekproeven. Met behulp van SAS komen we tot de conclusie dat de variabele een significante afwijking van de normaalverdeling te zien geeft. Bij SAS-PC heeft de grootheid W een waarde van 0.974. De bijbehorende linkeroverschrijdingskans is 0.001 waaruit eveneens een significante afwijking van de normaalverdeling blijkt.

SYSTAT heeft als toets op de normaalverdeling de 'Kolmogorov-Smirnov one sample test'. Ook toepassing van SYSTAT levert een significante afwijking van de normaalverdeling op.

Tabel 4: Normaliteitscontrole.

	SPSSx	SAS	SPSS-PC+	SAS-PC	SYSTAT	STATVIEW
test	KS ¹ -goodness of fit	Kolmogorov D	KS-goodness of fit	Shapiro-Wilk	KS-one sample test met stand. normaal verd.	niet aanwezig
waarde	KSZ 1.250	D: normal 0.051	KSZ 1.250	W: normal .974	(maxdif) 1.000	-
waarschijnlijkheid	p=0.088	Prob>D <.01	p=0.088	Prob<W .001	(2-tail) .000	-
significante afwijking op 5% nivo?	NEE	JA	NEE	JA	JA	-
¹ KS=Kolmogorov-Smirnov						

De verschillende toetsen leiden dus tot tegengestelde conclusies omtrent de normaliteit van de LAT-schaal.

Voorafgaande aan de regressieanalyses hebben we een correlatiematrix berekend van de acht onafhankelijke variabelen, CITO, SYNO, HGD, EMT, TL, BL, LM en MT (voor uitleg: zie paragraaf 3) en de afhankelijke variabele LAT. SPSSx, SAS en SAS-PC vinden gelijke correlaties. De drie andere pakketten, SPSS-PC+, SYSTAT en STATVIEW, vinden ook overeenkomstige correlaties. Het verschil wordt verklaard door

de behandeling van de ontbrekende scores. SPSSx, SAS en SAS-PC gebruiken pairwise deletion. Bij SPSS-PC+, SYSTAT en STATVIEW wordt listwise deletion gehanteerd.

B.2. Regressieanalyse met alle acht onafhankelijke variabelen.

Ter beantwoording van de onderzoeksvraagstelling op dit punt gebruiken we de methode waarbij alle onafhankelijke variabelen tegelijk in de analyse worden betrokken ('ENTER-methode'). Deze methode is in alle pakketten rechtstreeks uitvoerbaar.

Bij regressieanalyse van de acht onafhankelijke variabelen blijken de variabele CITO (de toets begrijpend lezen van het CITO) en LM (de motivatie van de leerling ten aanzien van lezen) de beste voorspellers van de score op de schaal Leesattitude (LAT). Alle pakketten komen op dit punt tot overeenkomstige resultaten. Soms wijkt de constante enkele tienden af. De niet gestandaardiseerde regressievergelijking is als volgt:

$$LAT = 27,97 + 0.53 \cdot CITO + 2.07 \cdot LM + \text{errorterm}$$

De gestandaardiseerde regressievergelijking is voor alle pakketten:

$$LAT = 0.21 \cdot CITO + 0.19 \cdot LM + \text{errorterm}$$

B.3. Bloksgewijze regressieanalyse.

We onderscheiden twee blokken van onafhankelijke variabelen: enerzijds variabelen die cognitieve aspecten meten (CITO, SYNO, HGD en EMT) en anderzijds variabelen waarin het oordeel van de leerkracht ligt besloten (TL, BL, LM en MT). Regressieanalyse van het blok met de cognitieve aspecten levert alleen de variabele CITO als significante regressor in de vergelijking op. De regressievergelijking luidt:

$$LAT = 29.64 + 0.60 \cdot CITO + \text{errorterm}$$

De gestandaardiseerde regressievergelijking:

$$LAT = 0.23 \cdot CITO + \text{errorterm}$$

Het blok variabelen met daarin het oordeel van de leerkracht heeft als significante voorspellers van leesattitude LM en MT (mondelinge taalvaardigheid). De regressievergelijking is:

$$LAT = 34.19 + 2.23 \cdot LM + 1.65 \cdot MT + \text{errorterm}$$

Alle pakketten komen ook hier met gelijklopende uitkomsten, al wijken de constante en de b-coëfficiënten bij SYSTAT en STATVIEW iets af (respectievelijk SYSTAT: c =

34.56, $b(\text{LM}) = 2.33$, $b(\text{MT}) = 1.69$ en STATVIEW: $c = 34.52$, $b(\text{LM}) = 2.32$, $b(\text{MT}) = 1.69$).

De gestandaardiseerde vergelijking is voor alle pakketten:

$$\text{LAT} = 0.21 * \text{LM} + 0.13 * \text{MT} + \text{errorterm}$$

B.4. Regressieanalyse op minimale verzameling predictoren.

Wanneer we alleen de variabelen CITO en LM in de regressieanalyse betrekken, krijgen we de volgende vergelijking:

$$\text{LAT} = 30.28 + 0.70 * \text{CITO} + 2.30 * \text{LM} + \text{errorterm}$$

De gestandaardiseerde vergelijking luidt:

$$\text{LAT} = 0.27 * \text{CITO} + 0.21 * \text{LM} + \text{errorterm}$$

De resultaten van de regressieanalyse wijzen voor alle pakketten in dezelfde richting: de variabelen LM en CITO zijn de beste voorspellers voor de afhankelijke variabele LAT.

C. Verschillen tussen scholen en tussen autochtone en allochtone leerlingen t.a.v. LAT.

In dit deel wordt nagegaan of de leesattitude verschilt per school en of er verschil bestaat tussen autochtone en allochtone leerlingen. Het was niet mogelijk om na te gaan of er interactie-effecten zijn omdat er op een aantal scholen te weinig allochtone kinderen waren.

C.1. Enkelvoudige Variantieanalyses.

Uit de enkelvoudige variantieanalyse komen we met alle pakketten tot dezelfde conclusie dat de leesattitude verschilt per school en dat er ook verschil tussen allochtone en autochtone kinderen (herkomst) bestaat.

De F-waarden van de modellen:

school - LAT: 4.17, $p = 0.0001$
 herkomst- LAT: 7.43, $p = 0.0066$

C.2. Kruistabel school-herkomst.

Alle pakketten geven hier gelijke resultaten.

C.3. Gemiddelden per school uitgesplitst naar herkomst.
Ook hier zijn de pakketten gelijk van oordeel.

5. Conclusies en discussie

De resultaten van de uitgevoerde analyses laten een hoge mate van overeenstemming zien. Dit geldt met name ook voor resultaten van analyses waarbij uitkomsten van vorige analyses als inputvariabelen werden gebruikt. De verschillen die dan optreden lijken terug te voeren tot verschillen in rekennauwkeurigheid en afrondtechniek.

Daarbij dient onmiddellijk te worden aangetekend dat in sommige pakketten een aantal mogelijkheden ontbreken die voor sociaal-wetenschappelijk onderzoek onontbeerlijk zijn. In SYSTAT is de factoranalyse slechts zeer beperkt.

Alleen in SPSSx is de berekening van de betrouwbaarheidscoëfficiënt standaard aanwezig.

In STATVIEW zitten (nog) aanwijsbare fouten met betrekking tot de rotatie van een factormatrix en is geen normaliteitstoets aanwezig.

In SPSSx wordt de verdeling van de verklaarde variantie over geroteerde factoren niet gegeven.

Daarnaast treden verschillen op die hun oorsprong vinden in de uitwerking van een statistische techniek of in een verschil in keuze van de statistische techniek. Juist op dit punt kan men tot verkeerde conclusies komen wanneer men zich verlaat op het toevallig beschikbare pakket. In de inleiding constateerden we reeds dat deze problematiek speelt met betrekking tot ongebalanceerde variantie analyse designs. In dit onderzoek is gebleken dat bij toetsing op normaliteit een grote verscheidenheid aan statistische toetsen in de verschillende pakketten is geprogrammeerd, die in ons geval bovendien tot tegengestelde conclusies leiden. De gevonden verschillen zijn een rechtstreeks gevolg van de verschillen in de statistische technieken. Het is een bekend feit dat resultaten van verschillende normaliteitstoetsen tot afwijkende conclusies kunnen leiden. De diverse toetsen gaan uit van verschillende assumpties, een uniform beste methode is niet aan te geven. Idealiter zou een onderzoek naar deze assumpties in het licht van de onderzoeksvraagstelling allereerst moeten leiden tot de keuze van de toetsingsmethode. Pas daarna kan een keuze worden gemaakt van het pakket waarin deze methode is geïmplementeerd. Deze werkwijze zou mutatis mutandis voor elke statistische analyse moeten gelden.

In de inleiding gaven we aan dat evaluatie van statistische software vaak plaatsvindt op grond van globale criteria die niet voortkomen uit een statistische operationalisering van de onderzoeksvraagstellingen maar ingegeven zijn door de beschikbaarheid en het bedieningsgemak van softwarepakketten. In het onderhavige onderzoek is aangetoond dat men daarmee op onderdelen tot onvolledige of verschillende resultaten en conclusies kan komen. Zonder te stellen dat iedere onderzoeker over een geldverslindend arsenaal aan hard- en software zou moeten beschikken, menen wij toch op een zorgvuldige keuze van statistische software te moeten aandringen. Dit kan betekenen dat voor de uitwerking van een concreet stuk onderzoek meerdere pakketten moeten worden gebruikt.

Keren wij nogmaals terug op het variantieanalysevoorbeeld dan mogen wij aannemen dat de gevonden verschillen in een tweevoudige variantieanalyse ook optreden in meer complexe designs en tevens ook tot verschillen aanleiding kunnen geven in multivariate technieken die op de verschillende benaderingen zijn gebaseerd. De literatuur op dit punt is in het algemeen niet toegankelijk voor onderzoekers in een toepassingsgebied. In dit opzicht is het takenpakket van de statistisch adviseur door de komst van een uitgebreid arsenaal aan software eerder in omvang toegenomen dan afgenomen.

Lehman (1987) besluit zijn bespreking van statistische software met:

"These are outstanding programs available for the Macintosh that play no second fiddle to none on any computer. For most problems, statistics users can cut the cord to the mainframe and do all their analysis right on their desk."

Hoewel de resultaten van onze studie een bevestiging te zien geven van de goede kwaliteiten van statistische pakketten voor de Macintosh kunnen wij bovenstaande conclusies niet onderschrijven. Het arsenaal aan statistische technieken is binnen elk pakket beperkt. Dit geldt zowel voor de mainframe- als voor de PC-pakketten. De gesuggereerde tegenstelling tussen mainframe- en PC-pakketten doet in dit opzicht niet ter zake. Veeleer kan worden vastgesteld dat verruiming van het aanbod aan statistische software leidt tot vergroting van de keuzemogelijkheden. Mits op de juiste manier gehanteerd, kan deze uitbreiding van het statistisch arsenaal bijdragen tot kwaliteitsverbetering van het empirisch onderzoek.

Literatuur

- Aarnoutse C.A.J. en A.C.E.M. Weterings (1989), Leeractiviteiten en processen tijdens het onderwijs in begrijpend lezen. Instituut voor Onderwijskunde. In voorbereiding.
- Bakker, J.G.M. (1986), SPSS9, SPSSX of SAS. Vergelijking van SPSS rel 9, SPSSX en SAS. CRI Handleiding 121, Uitgave van het Centraal RekenInstituut van de Rijks-universiteit Leiden.
- Bisschop, P., Aarnoutse, C. en H. Feenstra (1985), Leesattitudeschalen. Bestemd voor het derde, vierde en vijfde leerjaar van het basisonderwijs, Nijmegen, Berkhout.
- Bouts, L. en H. Korzilius (1988), SPSS/PC+. Serie Technical Notes van de RTD, PAW-KU-Nijmegen, no. 10.
- Fridlund, A.J. (1986), Statistics Software. InfoWorld, 31-39.
- Gaaff, A. en Th.W. de Boer (1985), Statistische programmatuur voor personal computers: geteld, gewogen en getest. In: Symposium Statistische Software 1985, Technisch Centrum FSW, Universiteit van Amsterdam.
- Keller, W.J. (1986), Statistische programmatuur voor de personal computer. Brochure bij de Teleaccursus Statistiek, Stichting Teleac, Utrecht.
- Korzilius, H. (1986), Inleiding SAS-PC. Serie Technical Notes van de RTD, PAW-KU-Nijmegen, no. 3.
- Korzilius, H. (1987), STATVIEW. Serie Technical Notes van de RTD, PAW-KU-Nijmegen, no. 9.
- Korzilius, H. (1988), SYSTAT (versie 3.0). Serie Technical Notes van de RTD, PAW-KU-Nijmegen, no. 11.
- Lehman, R.S. (1987), Statistics on the Macintosh. In-depth tests of 12 programs ranging from elementary to professional quality. BYTE, 12, 207-214.
- Lehman, R.S. (1986), Macintosh statistical packages. Behavior Research Methods, Instruments, & Computers, 18, 177-187.
- Longley, J.W. (1967), An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User. Journal of the American Statistical Association, 37, 83-86.
- Musegaas, B.E. (1984), Statistische programmatuur voor micro-computers, een overzicht van het aanbod (versie 2.1). Stichting Mens, Computer en Maatschappij, Utrecht.
- Winer, B.J. (1971), Statistical Principles In Experimental Design. McGraw-Hill, Inc., Kogakusha.