

Statistische software voor een industriële omgeving

Dee Denteneer*

Jan Engel†

2 juni 1995

Samenvatting

Het doel van ons onderzoek is het vinden van een geschikt statistisch pakket voor een industriële omgeving. Hiertoe zijn Minitab, SPSS, Stata, Statistica en Stagraphics *plus* onderling vergeleken op gebruiksgemak en functionaliteit, met name op het gebied van de proefopzetten en de quality control tools. De evaluatie leidde tot een aanbeveling van Minitab als het meest geschikte pakket voor de beoogde doelgroep.

*Adres: Philips Natuurkundig Laboratorium, Prof. Holstlaan 4, 5656 AA Eindhoven. Tel: 040-744767. E-mail: dentenee@prl.philips.nl

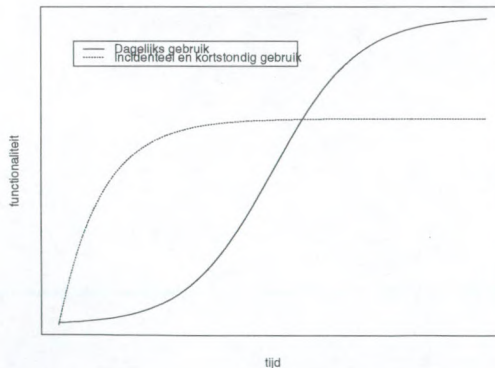
†Adres: CQM, Vonderweg 11, Postbus 414, 5600 AK Eindhoven. Tel: 040-758722. E-mail: eg1@cqm.iaf.nl

1 Inleiding

Het doel van ons onderzoek is het vinden van een geschikt statistisch pakket voor een industriële omgeving. Uitgangspunt voor dit onderzoek waren de wensen van de gebruiker, de 'industrial engineer'. Concreet betreft het R&D dan wel productie engineers die een inleidende cursus proefopzetten en proces beheersing hebben gevolgd en zelf proeven willen uitvoeren of hun proces karakteriseren. In tegenstelling tot de statisticus legt de engineer een grotere nadruk op gebruiksgemak dan op functionaliteit. Concreet betekent dit dat wij bij de beoordeling vooral gelet hebben op de user-interface van het programma en op het aanwezig zijn van een beperkt aantal zeer wel omschreven statistische hulpmiddelen.

De user-interface is van groot belang omdat de engineer slechts incidenteel en kortstondig gebruik zal maken van het pakket. Er is derhalve geen plaats voor een, steeds terugkerende, leerperiode waarin hij het pakket grondig leert kennen. Zijn prioriteiten liggen elders. Dit stelt heel andere eisen aan de leercurve voor het gewenste pakket dan voor een pakket dat dagelijks gebruikt wordt, zie bijvoorbeeld Figuur 1. Functionaliteit valt niet geheel los te zien van gebruiksgemak. Een pakket dat niet standaard de gewenste functionaliteit biedt, blijft onhandig in het gebruik, hoe fraai en overzichtelijk de menu's ook zijn. Onze eisen met betrekking tot de functionaliteit van de pakketten concentreren zich derhalve op die technieken die in de literatuur, zie bijvoorbeeld Banens *et al.* (1994), aanbevolen worden voor produktontwikkeling en -verbetering. De nadruk valt daarmee op proefopzetten, quality control tools en grafische hulpmiddelen, ten koste van algemene statistische technieken als bij Keller (1986, Tabel 3-5).

Na een korte vermelding in Sectie 2 van de verantwoording van het evaluatieonderzoek en de gevolgde procedure, volgt in Sectie 3 van dit rapport de eigenlijke evaluatie, dit aan de hand van een aantal evaluatietabellen en enkele korte opmerkingen per pakket. In Sectie 4 formuleren wij de belangrijkste conclusies.



Figuur 1: Gewenste leerkrommen voor een pakket voor dagelijks gebruik en een pakket voor incidenteel, kortstondig gebruik, aangepast overgenomen uit Nielsen (1993, Pagina 28). Bij het pakket voor dagelijks gebruik zien wij een langzame maar zekere accumulatie van expertise naar een hoge functionaliteit. Bij het pakket voor incidenteel gebruik is een snelle accumulatie van expertise gewenst; de functionaliteit die daarmee uiteindelijk wordt bereikt mag geringer zijn.

2 Verantwoording

In de evaluatie hebben wij ons beperkt tot menu-gestuurde pakketten waarmee in onze omgeving al enige ervaring bestond en die in de wandelgangen als 'prettig in het gebruik' gekarakteriseerd werden. Bekende afvallers zijn bijvoorbeeld Splus, SAS en RS/Discover (te ingewikkeld) en Jass (te beperkt). Minitab, SPSS, Statistica en StatGraphics *plus* bleven over. Daarnaast werd Stata meegenomen in de evaluatie omdat dit op het moment erg veel gebruikt wordt binnen Philips. Daarbij hebben wij CQM-Stata geëvalueerd: standaard Stata uitgebreid met een aantal hulpmiddelen ('kits') op het gebied van proefopzetten en quality control tools, zie Banens *et al.* (1994, Bijlage A).

Om onze wensen voor een pakket te objectiveren zijn voor het onderzoek enkele tabellen gemaakt, waarin onze eisen voor het pakket geconcretiseerd werden. Ons grote voorbeeld hierbij was Keller (1986), zij het dat wij de tabellen danig aangepast hebben aan wat wij relevant achten in een industriële context. Deze aanpassing is gebaseerd op uitvoerige consultancy-ervaring binnen de gestelde doelgroep.

De evaluatie is uitgevoerd door de twee auteurs binnen een zeer beperkt tijdsbestek. De evaluatieprocedure was als volgt. Elk pakket werd door de twee auteurs samen beoordeeld, waarbij per pakket zeker twee dagdelen werden gereserveerd. Daarvan diende het eerste dagdeel als een globale verkenning, terwijl in het tweede (en soms ook derde) dagdeel de eigenlijke evaluatie werd gedaan aan de hand van de tabellen. Het raadplegen van de handleidingen completeerde de informatie.

Wij zijn ons ervan bewust dat dit een grote concessie is aan de diepgang van het onderzoek. Een formele evaluatie van de waardering voor een user-interface impliceert immers een panel van proefpersonen (gebruikers, dus engineers) dat een aantal taken uitvoert volgens een van te voren vastgelegde opzet, terwijl de usability kwalificaties (tijd, aantal gemaakte fouten, ...) vastgelegd en geanalyseerd worden. Het geheel wordt bij voorkeur gecompleteerd door een fikse vragenlijst. Zie Nielsen (1993, Hoofdstuk 6).

Als verzachtende omstandigheid moge gelden dat de auteurs leek waren ten aanzien van via menu's te bedienen statistische programmatuur in het algemeen, en ten aanzien van de geëvalueerde pakketten in het bijzonder. Daarmee reflecteert de zeer korte evaluatieperiode juist de manier waarmee de onervaren gebruiker n onze doelgroep met het pakket om zal gaan. Bovendien is het zo dat het door onze evaluatie gepropageerde pakket een voorlopige keuze is. De definitieve pakketkeuze vindt pas plaats na acceptatie door de gebruikers en binnen die keuze fungeert dit overzicht niet als een diktaat, maar als een eerste ordening.

3 Evaluatie

In Subsectie 3.1 bespreken we kort de wensen ten aanzien van de pakketten zoals wij die voor het onderzoek vastgelegd hebben, en voor zover ze niet vanzelfsprekend zijn. In Subsectie 3.2 volgen dan de tabellen, de kern van ons onderzoek, en in Subsectie 3.3 volgen wat korte opmerkingen over de verschillende pakketten.

3.1 Wensen

Het software-wensenlijstje zoals geformuleerd door de door ons ondersteunde industriële omgeving, ziet er als volgt uit:

- Het pakket draait onder Unix op HP en op PC, ziet Tabel 1.

	Versie	DOS prijs	Windows prijs	Unix prijs	test machine	test OS
Minitab	10.2	£295	£395	£2990 ^{*)}	486	Windows
Statgraphics plus	1.1	Dfl 2095	Dfl 3395	nvt	486	Windows
Statistica	4.5	Dfl 3500	Dfl 4200	nvt	486	Windows
SPSS ^{**)}	5.0	Dfl 1295	Dfl 2590	Dfl 5180	HP-700	Unix/motif
CQM-Stata ^{***)}	3.1	Dfl 2395	nvt	nvt	486	Dos

Tabel 1: Er kunnen uiteraard significante verschillen bestaan in appreciatie van de versie van het pakket zoals dat door ons getest is (zie machine/OS-combinatie) en de andere verkrijgbare versies. ^{*)} Site licentie; geen individuele licenties beschikbaar. ^{**) Prijs van de geëvalueerde basis module, er geldt een verplicht renewal bedrag voor het tweede en derde jaar. ^{***)} Stata is wel beschikbaar voor Unix, de CQM uitbreiding echter niet.}

- Het pakket heeft een prettige user user-interface. Het is daardoor eenvoudig te leren, eenvoudig te herinneren en het is een genoegen om ermee te werken. Zie Tabel 2.
- Het pakket heeft minimaal de volgende functionaliteit:
 - Basis statistische analyses, zie Tabel 3.
 - Grafische analyses, zie Tabel 4.
 - De gangbare industriële proefopzetten: 2^n -proeven, Box-Behnken proeven, Central Composite Designs (CCD's), 3^n -proeven, zie bijvoorbeeld Box, Hunter en Hunter (1978). Het betreft hier zowel de opzet als de analyse van de proef. Zie Tabel 5.
 - Quality control tools. Zie Tabel 6.

De beoordeling is uitgevoerd door het toekennen van de scores + = goed, . = matig en - = slecht, dan wel afwezig.

3.2 Tabellen

De Tabellen 2 tot en met 4 mogen voor zichzelf spreken. In Tabel 5 worden de proefopzetten geëvalueerd. De categorie 'Randomisatie' onderzoekt of het mogelijke is om af te wijken van volledige randomisatie, bijvoorbeeld gerandomiseerde blokken en beperkte randomisatie. De categorie 'Centrumpunt' betreft de uitbreiding van 2^{n-k} -proeven met centrumpunten voor het onderzoeken van niet-lineariteiten en spreiding. De response surface designs worden afgedekt door Box (Box-Behnken designs) en CCD (Central Composite Designs).

Voor een uitleg van de quality control tools verwijzen wij naar Hoofdstuk 3 uit Banens *et al.* (1994). In Tabel 6 wordt de Control Chart alleen dan hoog gewaardeerd als het pakket naast de standaard Shewhart kaart ook de 'individuals control chart' bevat. Voorts betreft 'Pareto' de Paretoplot, 'Ishikawa' het Ishikawa, ook wel visgraatdiagram genoemd, en 'PCS' de process capability study.

3.3 Opmerkingen

In het voorafgaande hebben wij een systematische vergelijking gepresenteerd van de verschillende pakketten. Wij kunnen het niet nalaten deze vergelijking aan te vullen met enkele fragmentarische persoonlijke indrukken.

	Menu	Commando	Leren	Herinneren	Syntax	Macro	Help	Robuust	Leuk	Totaal
Minitab	+	+	+	+	.	+	+	+	+	+
Statgraphics <i>plus</i>	+	-	+	+	-	+	+	.	+	+
Statistica	+	+	.	.	.	+	+	.	.	.
SPSS	+	+	+	+	.	+	+	+	+	+
CQM-Stata	-	+	-	-	.	+	.	+	-	-

Tabel 2: Gebruiksvriendelijkheid

	Samenvatting	Conditioneren	Toetsen	Niet-parametrisch toetsen	ANOVA	Regressie	Totaal
Minitab	+	.	+	+	+	+	+
Statgraphics <i>plus</i>	+	-	.	-	+	+	.
Statistica	+	.	+	+	+	+	+
SPSS	+	.	+	+	.	+	+
CQM-Stata	+	+	+	+	+	+	+

Tabel 3: Statistiek

	Histogram	Scatter- plot	Box- plot	Regressie- plots	3d- plots	qq- plot	Visuele presentatie	Totaal
Minitab	+	+	+	+	+	+	+	+
Statgraphics <i>plus</i>	+	+	.	+	.	+	+	+
Statistica	+	+	+	+	+	+	+	+
SPSS	+	+	+	+	.	-	+	+
CQM-Stata	+	+	+	+	-	+	+	+

Tabel 4: Grafische mogelijkheden

	2^{n-k} Proeven	Randomisatie	Centrum punten	3^n	Box/ CCD	Mixture Designs	$2^m 3^n$ Proeven	Taguchi	Totaal
Minitab	+	+	+	-	+	+	-	-	+
Statgraphics <i>plus</i>	-	-	-	-	-	-	-	-	-
Statistica	+	.	-	-	.	-	-	+	.
SPSS	-	-	-	-	-	-	-	-	-
CQM-Stata	+	+	+	+	+	-	-	+	+

Tabel 5: Proefopzetten

	Control Chart	Trend Chart	Pareto	Ishikawa	PCS	Totaal
Minitab	+	+	+	+	+	+
Statgraphics <i>plus</i>	+	+	+	-	+	+
Statistica	.	+	+	-	.	.
SPSS	-	-	-	-	-	-
CQM-Stata	+	+	-	-	+	.

Tabel 6: Quality control tools

MiniTab was eigenlijk de verrassing van de geëvalueerde pakketten. Het is een overzichtelijk en gebruiksvriendelijk pakket; de oorsprong van Minitab als pakket voor lesdoeleinden, zie Lock (1993), zal daar niet vreemd aan zijn. Ook de samenhang tussen commando's en menubediening is overzichtelijk: de via het menu geactiveerde commando's worden getoond in het uitvoer-window. Op die manier leert men spelenderwijs de commando's kennen. De fraaie mogelijkheid bestaat om de commando's in het uitvoer-window te kopiëren naar een syntaxwindow, ze daar aan te passen en ze opnieuw uit te voeren. De commandostructuur zelf doet overigens wat ouderwets aan. Zo worden de variabelen standaard aangeduid met C1, C2, Het is mogelijk de variabelen van namen te voorzien, maar bij gebruik moeten dan altijd quotes om de namen worden geplaatst: met

```
name C1 = 'tijdsduur'
let C2 = C1           of      let C2 = 'tijdsduur'
```

krijgt variabele C1 de naam tijdsduur, en vervolgens wordt de waarde van variabele C1 aan variabele C2 toegekend.

Het gebruiksvriendelijke karakter van Minitab wordt in zekere zin mogelijk gemaakt door zijn beperking. Het aantal statistische technieken dat wordt aangereikt is beperkt, maar daardoor is het mogelijk het pakket op een consistente en overzichtelijke wijze via menu's aan te sturen. Uit de tabellen blijkt overigens weinig van een dergelijke beperking: door de technisch-wetenschappelijke oriëntatie van het pakket is er een grote overlap tussen hetgeen Minitab te bieden heeft en hetgeen door ons werd geëist.

Een laatste woord van lof betreft de handleidingen die met Minitab geleverd worden: ook de omvang hiervan heeft Minitab op zinvolle wijze weten te beperken. Als enige levert Minitab het moderne pendant van de 'quick reference card': het 28 pagina's tellende 'quick reference manual'.

StatGraphics plus Net als Minitab implementeert Statgraphics *plus* een beperkt aantal statistische technieken; echter door het ontbreken van proefopzetten is de overlap met het door ons gewenste niet erg groot. Overigens zijn de proefopzetten gepland voor een volgende release en waren zij ook aanwezig in de voorafgaande DOS-versie van Statgraphics. Door het snel verschijnen van opvolgende versies van de verschillende pakketten is ieder vergelijkend onderzoek aan de hand van de meest recente versies slechts een beperkte tijd relevant. Dit onderstreept ons pleidooi in de conclusies, waarin voor meer systematische en regelmatige pakketevaluatie wordt gepleit.

Ook bij StatGraphics *plus* leidt het beperkte aantal technieken tot een prettig en natuurlijk gebruik van de menu's. Het feit dat voor vrijwel iedere grootte die berekend wordt en voor iedere grafiek een apart window geopend wordt, maakt het gebruik van Statgraphics *plus* in eerste instantie vrij onoverzichtelijk.

Statgraphics *plus* biedt een keur aan grafieken, waarin op allerlei manieren de grootte, de kleur, het font, etc. te veranderen zijn. Toch scoort het pakket niet bijzonder hoog bij onze evaluatie ten aanzien van de grafische mogelijkheden, zie Tabel 4. Dit heeft zowel een subjectieve als een objectieve oorsprong. Wat betreft het eerste: één van de auteurs vond de gepresenteerde grafieken niet mooi: geen echt hoge resolutie en te veel niet-functioneel gebruik van kleuren ('te veel effectbejag, te weinig informatie, een typisch pc-pakketje'). Wat betreft het tweede: belangrijke grafieken als contour-plots ontbraken en bij de box-plots lukte het ons niet om meerdere variabelen tegelijk onder te brengen in de grafiek, hetgeen juist de essentie van de techniek is. (Dit laatste kan overigens te wijten

zijn aan een bug in een programma, waardoor het niet mogelijk was meerdere variabelen te selecteren bij de 'Multiple Comparison'.)

Statistica biedt een veel uitgebreider pakket statistische technieken dan de voorafgaande pakketten. Alle technieken en 'toeters en bellen' eraan zijn via het menu te bereiken en dit heeft zijn prijs. Ten eerste moeten er regelmatig nieuwe modules geladen worden wat een vervelende discontinuïteit in het werken met het pakket oplevert. Kan het eerste nog verholpen worden met een snellere computer en meer RAM-geheugen, het tweede nadeel lijkt onoverkomelijk. Doordat alle technieken en de opties daarin met de menu's te bereiken zijn, maakt het met menu's volgepropte scherm een onoverzichtelijke indruk. De gebruiker weet al snel niet meer welke van de menu's wel en welke niet actief zijn. Het ontbreken van een context-gevoelige helpfaciliteit completeert de problemen van de beginnende gebruiker.

Statistica is dan ook het enige pakket waarmee de auteurs na twee oefensessies niet goed uit de voeten konden bij een aantal eenvoudige analyses. Uiteraard hoeft dit geen diskwalificatie van Statistica in te houden. Door de hoeveelheid geboden technieken kan het een aantrekkelijk pakket zijn. Het pakket vereist echter, ondanks de menubediening, de nodige hoeveelheid computerexpertise en oefening. In hoeverre de handleidingen (2500 pagina's!) daarbij ondersteuning bieden is een open vraag. De auteurs troffen een afdeling met 8 Statistica licenties, intensief Statistica-gebruik en slechts één handleiding. Deze handleiding was overduidelijk nog nooit gebruikt.

SPSS was voor ons nauw verbonden met begrippen als 'mainframe', 'batch' en 'ponskaarten'. Wie dan de Unix-versie van SPSS draaiend op de HP-9000/700 onder motif opstart, gelooft zijn ogen niet: zeer prettige bediening met menu's en zeer overzichtelijke uitvoer. Alleen het wel erg trage scrollen door de tekstuitvoer verradt dat het pakket geen echte motif-achtergrond heeft.

SPSS behoeft zeker wat extra toelichting. Het bestaat uit de in de tabellen geëvalueerde basismodule, die naar behoefte uitgebreid kan worden met een aantal extra modules die meer geavanceerde statistische technieken bieden. Daardoor ontstaat een vrijwel onuitputtelijk reservoir aan technieken. De menubediening van dit geheel is op eenvoudige wijze overzichtelijk gehouden, namelijk door niet alle mogelijkheden via menu's aan te bieden. Het voordeel hiervan is dat de beginnende gebruiker een pakket voorgetoverd krijgt waarmee het een genoegen is om te werken. Om alle geboden technieken echter tot in detail te kunnen benutten, zal een gebruiker zich toch de weinig gebruiksvriendelijke, ouderwetse SPSS-programmeertaal eigen moeten maken. Dit laatste wordt overigens wel op fraaie wijze ondersteund door de mogelijkheid om via de menu's de basisversie van een commando in een syntaxwindow te plaatsen. Dit commando kan dan aangepast of uitgebreid, en vervolgens geëxecuteerd worden.

Door de modulaire structuur van SPSS hebben ook de plussen, punten en minnen in de tabellen extra toelichting nodig. Zoals aangegeven zijn deze toegekend op basis van de mogelijkheden van de basismodule. Bij de evaluatie hadden wij echter ook de beschikking over vijf extra modules: professional statistics, advanced statistics, tables, trends en categories. Indien wij ook de mogelijkheden van deze extra modules hadden verdisconteerd, zou er slechts sprake geweest zijn van een marginale verandering: alleen de '.' voor ANOVA zou een '+' geworden zijn. Onze conclusie is dat de kracht van SPSS in het verlengde van haar sociaal-wetenschappelijke achtergrond ligt bij discrete en multivariate data. De technieken voor de analyse van dergelijke data zijn echter anders dan die benodigd bij de analyse van de technisch-wetenschappelijke, vaak continue en univariate data van onze

doelgroep. Het ontbreken van de qqplot of quality control tools op een hoog nivo binnen de menubediening moge deze gedachte staven.

4 Conclusies

De evaluatie van een aantal met menu's te bedienen pakketten aan de hand van een aantal van tevoren opgestelde tabellen bracht een duidelijke winnaar naar voren: Minitab. Juist haar beperking tot technisch wetenschappelijke (continue en univariate) data maakt het mogelijk de gewenste technieken met de nodige diepgang te implementeren, zonder dat het gebruiksgemak moet worden opgeofferd. Daarmee is Minitab echter niet zaligmakend. Voor de analyse van andere type gegevens, levensduur of multivariate gegevens bijvoorbeeld, geeft Minitab niet thuis. SPSS is een mooi pakket dat een enorme hoeveelheid technieken omvat en complementair is aan Minitab. De modulaire structuur van SPSS maakt het mogelijk deze uitbreidingen gedoseerd in huis te halen. Bij SPSS komen de beperkingen van de menubediening naar voren: niet alle mogelijkheden zijn via de menu's te bereiken. Kleine aanpassingen van via de menu's te bereiken technieken worden handig ondersteund en vergen niet veel kennis van de SPSS-taal. Treedt men echter buiten de gebaande wegen en heeft men behoefte aan simulaties of cascades van elkaar aanvullende technieken, dan is een grondige kennis van de SPSS-syntax onontbeerlijk, en deze is, zoals vermeld, ouderwets en weinig flexibel. Voor zulke gevallen raden wij dan ook aan een pakket te nemen met een moderne programmeertaal, zoals Splus, en het menu-concept geheel te laten vallen.

Naast deze primaire conclusie betreffende de pakketkeuze is er een tweetal secundaire punten dat ons opviel gedurende de evaluaties. Deze hebben met name betrekking op de methodologie van een onderzoek als het voorliggende en wij geven deze kort weer:

- Er wordt in de literatuur opvallend weinig aandacht besteed aan het systematisch evalueren en vergelijken van programmatuur.
- De pakkevaluatie in de literatuur gebeurt weinig diepgravend. Als in ons relaas is er een overzicht van de benodigde hardware, een persoonlijke indruk van gebruiksgemak en ten slotte een selectie uit de geboden technieken. Te zelden leest men iets over bijvoorbeeld de nauwkeurigheid waarmee de technieken worden uitgevoerd, laat staan iets over gebruikerstesten.

Wij kunnen ons voorstellen dat de Sectie Statistische Programmatuur van de VVS op met name de laatste twee punten een ondersteunende rol wil gaan vervullen. Wij denken dan aan het concipiëren van een aantal standaardtabellen à la Keller, 1986. Deze zouden dan bijvoorbeeld bij programmatuurbesprekingen gehanteerd kunnen worden en de lezer een handvat bieden bij een systematisch vergelijking van de besproken programma's volgens een door hemzelf samen te stellen gebruikersprofiel. Ook zou KM een handzame verzameling taken ter beschikking kunnen stellen (het nare ASCII-bestand, de moeilijke regressie, het enge conditiegetal, ...) waarmee de pakketten wat zwaarder op de proef worden gesteld.

Referenties

- [1] Banens, P.J.A. *et al.* (1994), Industriële Statistiek en Kwaliteit. Kluwer.
- [2] Box, G.E.P., Hunter, W.G. en Hunter, J.S. (1978). Statistics for experimenters– An introduction to design, data analysis and model building. Wiley.
- [3] Keller, W.J. (1986). Statistical software for personal computers, Compstat-proceedings, 332-337.
- [4] Lock, R.H. (1993). A comparison of five student versions of statistics packages. The American Statistician 47, 136-145.
- [5] Nielsen, J. (1993). Usability engineering, Academic Press.

