

STATISTICA

Jaap Dik
Arend Oosterhoorn

PRODUKTINFORMATIE

Besproken versie: STATISTICA/W+SIS (Release 5, 5.1)

Uitgever: StatSoft, Inc.
2325 East 13th Street, Tulsa OK 74104
e-mail: info@statsoft.com
WEB: <http://www.statsoft.com>

Importeur: Exstat & Exsyst
Postbus 6082
9702 HB Groningen
tel. 050-5267310
e-mail: exsyst@xs4all.nl
WEB: <http://www.xs4all.nl/~exsyst>

Handleiding: Quick Reference Guide (198 pp)
Uitgebreide handleiding in 5 delen (resp. 878, 958, 782, 478, 494 pp)
In totaal 3734 pp.

Hulpfunctie: Statistica kent een uitvoerige context-afhankelijke hulpfunctie.

Systeemvereisten: Windows (Windows 3.1 of Windows 95/NT) compatible computer.

Programma: Het programma wordt naar keuze geleverd op diskettes (niet aan te bevelen) of op CD-Rom. Installatie vanaf CD-Rom is snel en eenvoudig.
Netwerkversies zijn beschikbaar.

Er is tevens een Demo versie beschikbaar op CD-Rom.

Ontwikkeling: Statistica werd geprogrammeerd met Microsoft Assembler, C/C++ compilers, Microsoft Windows SDK en andere hulpmiddelen. Het programma sluit naadloos aan bij de programma's van Microsoft Office (Professional)

Toepassingen: (Full) Statistica is een algemeen programma voor de analyse van databestanden, met een nadruk op grafische weergave van resultaten. Quick Statistica is een samenhangende subset van Full Statistica. SIS (Statistica Industrial System) is een toegevoegde module die vooral bedoeld is voor toepassingen in de industrie (Quality Control, Process Analysis, Design of Experiments, Taguchi methoden). SIS draait zowel in combinatie met Quick Statistica als met Full Statistica.

Prijzen:	Quick Statistica/W	Dfl 1095	
	Quick Statistica/W+SIS	Dfl 2995	
	Full Statistica/W	Dfl 2195	
	Full Statistica/W+SIS	Dfl 3995	(besproken pakket)

Prijzen volgens prijslijst van 15 september 1997. Men betaalt *geen* abonnementskosten.

Upgrades worden tegen acceptabele prijzen ter beschikking gesteld. Kleine correcties en upgrades (in de vorm van 'patches') worden gratis ter beschikking gesteld en kunnen desgewenst van het Internet worden gedownload.

Voor universiteiten, hogescholen en academische instituten bestaan er voordelige licentieregelingen.

BESCHRIJVING ALGEMEEN

Statistica is een grafisch georiënteerd pakket voor het uitvoeren van statistische analyses van gegeven databestanden op een Personal Computer. Als zodanig is het een directe concurrent voor de PC versies van SAS, SPSS, Statgraphics, Systat, Stata en dergelijke.

Statistica is met name geschikt voor het interactief analyseren van data.

Het is duidelijk de bedoeling geweest van de ontwikkelaars om aansluiting te zoeken bij de mogelijkheden van Microsoft Windows en met name bij Microsoft Office (MS Excel, MS PowerPoint, MS Word). Dit blijkt onder andere uit het feit, dat je functies uit de Statistica Command Language vanuit een Excel werkblad rechtstreeks kan aanroepen. Dat maakt het bijvoorbeeld mogelijk overschrijdingskansen of kritieke waarden van continue (e.g. Logistic) en discrete verdelingen (e.g. Poisson) in een Excel werkblad op te nemen.

Statistica is modulair opgebouwd, waarbij de modules uit logisch samenhangende onderdelen bestaan. Bepaalde (eenvoudige) functies, zoals histogram, spreidingsdiagram, zijn vanuit elke module bereikbaar. De modules worden bereikt via de module-switcher die vanuit elk Statistica scherm bereikbaar is.

Modules

Full Statistica/W+SIS bestaat uit de volgende modules:

Basic Statistics	SEPATH (structural equation analysis)
Nonparametrics/Distrib.	Reliability/Item Analysis
ANOVA/MANOVA	Discriminant Analysis
Multiple regression	Log linear Analysis
Nonlinear Estimation	Survival Analysis
Time Series/Forecasting	Quality Control
Cluster Analysis	Process Analysis
Data Management/MFM	Variance Components
Factor Analysis	Experimental Design

Canonical Analysis	Command Language (SCL)
Multidimensional Scaling	STATISTICA File Server
Correspondence Analysis	Classification Trees (nog niet in 5.1)

Data

Data worden gerepresenteerd in een rechthoekige matrix, waarin de kolommen de variabelen bevatten en de rijen de cases (case = record). Op het scherm worden de data weergegeven in een spreadsheet, die sterk lijkt op een MS Excel werkblad:

	1	2	3	4	5
	CASENUM	DEPN_X	INDEP_A	INDEP_B	INDEP_C
1	1,000	1,100	10,000	125,000	2,000
2	2,000	2,200	20,000	275,000	4,000
3	3,000	2,800	30,000	325,000	6,000
4	4,000	3,900	40,000	500,000	8,000
5	5,000	5,400	50,000	675,000	10,000
6	6,000	5,500	60,000	725,000	12,000

Wanneer je de data voorbereidt met Excel en rekening houdt met de specifieke opbouw van een Statistica bestand, kun je de data rechtstreeks en moeiteloos uit het Excel werkblad overnemen. [Huidige gebruikers van Statistica die problemen hebben met het inlezen van Excel '97 werkbladen (Microsoft veranderde het format) kunnen een upgrade met Excel '97 support downloaden via de Web-site van Statsoft].

Bewerkingen in de Statistica spreadsheet zijn mogelijk, maar de manier waarop wijkt enigszins af van de manier waarop dat in Excel gebeurt. Zo voeg je cases of variabelen toe met behulp van de gelijknamige knoppen in de knoppenbalk en niet door het invoegen van rijen en/of kolommen zoals bij Excel het geval is d.m.v. de rechtermuisknop en submenu. Na afloop van de analyse kun je de data weer opslaan als een Excel bestand, of, uiteraard, als een Statistica Data File. [Een door een password beschermd Excel bestand kon door Statistica overigens niet rechtstreeks worden ingelezen].

Import en export van data is mogelijk, wanneer je gebruik maakt van de Data Management Module. Import en export is mogelijk naar SPSS, SAS, Excel, Lotus, Quattro Pro, Paradox, dBase, en ASCII bestanden. Voor extreem grote bestanden kan je gebruik maken van de mogelijkheden van de Mega File Manager (MFM). De Mega File Manager kan extreem grote records lezen (tot 32.000 kolommen), wat van pas kan komen wanneer geautomatiseerde kwaliteitscontrole systemen procesdata 'loggen'.

Het samenvoegen ('mergen') van databestanden, zowel door combinatie van cases, als door combinatie van variabelen is mogelijk. Data waarvan de records in twee verschillende bestanden zijn ondergebracht kunnen worden samengevoegd, mits beide bestanden een unieke 'key variable' bevatten die de cases identificeren.

Variabelen kunnen op eenvoudige wijze worden getransformeerd volgens de standaard wiskundige functies (log, e-macht, sin, cos, tg, functies hiervan, etc.) Ook het format en het type (numeriek of tekst) kunnen eenvoudig worden gewijzigd (je klikt op de naam van de variabele in de bovenste rij van het data scrollsheet).

Coderingen zoals "man en "vrouw" kunnen desgewenst als 0 en 1 worden ingevoerd. Wanneer je later "man " aan 0 koppelt, dan worden voor de betreffende variabele alle cellen waarin de 0 voorkomt automatisch vervangen door de tekst "man". Variabelen van het type tekst worden steeds dubbel gerepresenteerd: als tekst en als codenummer. Maak je bij het prepareren van het databestand hiermee vergissingen, dan kan dit overigens tot onvoorspelbare resultaten leiden.

Analyse

Voor het analyseren van de data wordt eerst de betreffende module geselecteerd met de module-switcher. Vervolgens kies je de opties "Analysis" en "Resume Analysis" uit het hoofdmenu. Hierdoor komt een dialoogschermb tevoorschijn. Je begint met het selecteren van de variabelen waarop de analyse betrekking moet hebben. Eventueel geef je aan welke records ('cases') geselecteerd moeten worden. Ook is het mogelijk om de cases van een gewicht te voorzien. In sommige gevallen kunnen nog enkele opties worden ge(de)selecteerd of aangekruist. Na deze keuzes druk je op de OK knop en de analyse wordt uitgevoerd. Hierna kan je eventuele verdere opties kiezen en/of wijzigen en de analyse opnieuw uitvoeren.

Resultaten

Voor het weergeven van de resultaten opent Statistica telkens een nieuw venster. In dit venster worden de voornaamste resultaten vermeld en tref je doorgaans knoppen aan voor het weergeven van gedetailleerdere resultaten, aankruisvakjes voor verdere opties en overige keuzemogelijkheden. Gezien het feit dat Statistica geheel op Windows is georiënteerd, en daarom batch-processing alleen als aanvullende mogelijkheid aanbiedt (zie ook Programmeren in Statistica), komen de keuzemogelijkheden voor de gebruiker soms wat massief over. Dit is echter een kwestie van wennen; in de praktijk werkt het goed en snel. Gekozen opties worden gehandhaafd totdat je een analyse van een nieuw databestand begint.

Gedetailleerdere resultaten (zoals een tabel met de effecten bij een regressie-analyse) worden opnieuw in een nieuw venster in de vorm van een spreadsheet (door Statistica Scrollsheet genoemd) weergegeven:

DEPENDENT_X	INDEP_A	INDEP_B	INDEP_C	
1,000	1,100	10,000	125,000	2,000
2,000	2,200	20,000	275,000	4,000

3. **Regression Summary for Dependent Variable: DEPENDENT_X**

4. Continue... R = ,99771319 R² = ,99543160 Adjusted R² = ,99238600
5. F(2,3)=326,84 p<,00031 Std.Error of estimate: ,15466

N=6	BETA	St. Err. of BETA	B	St. Err. of B	t(3)	p-level
Intercept			,213333	,143983	1,481652	,235034
INDEP_A	-,016586	,261046	-,001571	,024732	-,063538	,953335
INDEP_B	1,014110	,261046	,007600	,001956	3,884787	,030226

Deze resultaten kunnen in een eigen bestand (Scrollsheet File, *.scr) worden opgeslagen. Wil je de resultaten op hun beurt als een Statistica File analyseren, dan sla je de spreadsheet op met de menukeuze "Save As Data" uit het hoofdmenu (te vinden onder "File"). Dezelfde resultaten kunnen ook worden afgedrukt met de printfunctie uit het hoofdmenu. Al naar gelang de "Page/Output Setup" worden de resultaten van de analysesessie beperkt of juist uitvoerig gelogd.

Gezien de eerder genoemde compatibiliteit met Excel, kan je door selecteren van de spreadsheet en knippen, de gehele inhoud in een Excel werkblad plakken. Uiteraard is plakken in een Word document of een Power Point document eveneens mogelijk.

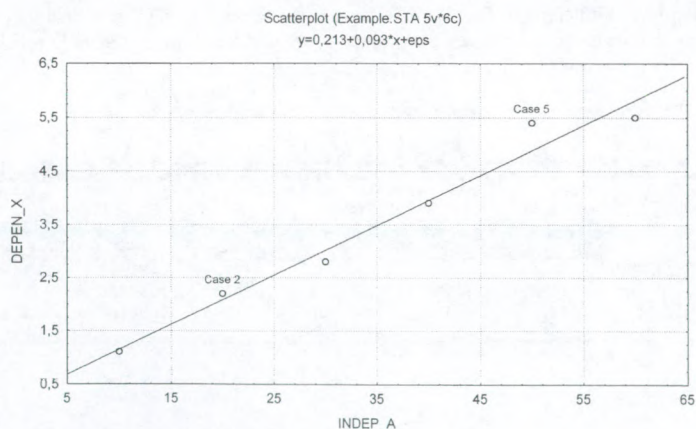
Grote scrollsheets (matrices) kunnen worden in- en uitgezoomd, zodat een goed overzicht behouden blijft. Door bijvoorbeeld in een correlatiematrix een cel aan te klikken, kan via de rechtermuisknop een spreidingsdiagram gemaakt worden van de betreffende variabelen.

Grafische mogelijkheden

Statistica kent een zeer groot aantal resultaten die in de vorm van een grafiek worden weergegeven (o.a. histogram, spreidingsdiagram, lijngrafieken, 3D weergaven, matrix graphs, probability plots etc. etc.). Elk grafisch resultaat wordt in z'n eigen venster weergegeven en kan op alle denkbare manieren aan de wensen van de gebruiker worden aangepast. Vrijwel alle aspecten van de grafiek (assen, eenheden, tickmarks, kleuren, lettergrootte, lettertype) kunnen worden gewijzigd. Net als bij de scrollsheets kunnen de grafieken apart worden opgeslagen en door knippen en plakken worden overgebracht naar MS Office toepassingen.

De kwaliteit van de afdrucken van de grafieken is uitstekend (bij afdrucken op een laserprinter). Tussen een direct vanuit Statistica afgedrukte grafiek en een die via MS Word of MS Power Point werd afgedrukt blijkt geen enkel verschil in kwaliteit te bestaan.

Een bijzonder nuttig hulpmiddel is het "brushing tool", een type cursor waarmee je individuele punten in een grafiek kan aanwijzen. Daarmee is het onder andere mogelijk te weten te komen bij welke case het desbetreffende punt behoort:



De betreffende cases zijn daarbij ook uit te schakelen, waarbij Statistica de grafiek weergeeft zonder dat die cases in de berekeningen worden betrokken.

Bovenstaande figuur (met uitsluitend 'default' ingestelde waarden) werd direct uit Statistica gekopieerd en in dit MS Word document geplakt.

Schermafhandeling

Het aantal vensters op het scherm is in te stellen (zoals bijna alles in dit programma). De ontwerper van de software moet wel een enorm geduld hebben gehad bij het ontwerpen van de iconen, want daarvan zijn er heel veel. In de handleiding staat een schermweergave waar meer dan 160 buttons zijn te zien. Dat hebben wij maar niet geprobeerd te reconstrueren op het scherm. De opmaak met al die vensters geeft je het gevoel dat er steeds dingen aan je gevraagd worden. De analyse wordt op de achtergrond uitgevoerd en je zegt maar welke resultaten je wilt zien en hoe je ze wilt zien.

Programmeren in Statistica

Het is mogelijk om steeds terugkerende taken binnen Statistica te automatiseren, door gebruik te maken van de Statistica languages (Basic en Statistica Command Language (SCL)). Ook kunnen macro's worden gebouwd, zelfs macro's waarin muisbewegingen zijn opgenomen. In de documentatie (deel V) tref je een uitgebreid overzicht van beschikbare commando's, uitleg en vele voorbeelden aan.

Macro's en in SCL of Basic geprogrammeerde onderdelen kunnen aan Windows-knoppen ('Autotask buttons') worden toegekend, zoals dat ook mogelijk is bij MS Word, MS Excel etc. Op die manier kan ook 'batch' verwerking binnen Statistica plaatsvinden. Zij kunnen ook als afzonderlijk programma worden gedraaid door een icoon op te nemen in een Windows programmagroep. Ook kan men deze 'programma's' opslaan en editen.

Met behulp van deze tools, kunnen resultaten worden verkregen die op zich in Statistica niet vertegenwoordigd zijn. Zo kunnen matrix berekeningen worden uitgevoerd (eigenwaarden, eigenvectoren, inversen etc.) en alle daarmee samenhangende grootheden.

Documentatie

Het voordeel van de boeken die worden meegeleverd is, dat er vooral wordt beschreven wat er met de betreffende termen en modules wordt bedoeld. Omdat Statistica zo menugestuurd is, bestaat er geen enkele noodzaak in de handleiding een behandeling van commando's op te nemen. De 3734 pagina's staan dan ook vol met beschrijvingen wat alle indicatoren en methoden weergeven en doen. Het boek over grafieken is het dikste (bijna 1000 pagina's) en heeft een hoofddeling naar 2D grafieken, 3D grafieken, 3D XYZ grafieken, matrixgrafieken, icon grafieken en andere. Dit onderstreept het karakter van Statistica, vooral gericht op grafische weergave van de resultaten. Aantal schermen, grafieken, buttons, het geheel geeft aan dat de gebruiker van Statistica een visueel ingesteld persoon zal zijn.

Er is een 10Mb grote Electronic manual met uitleg over de methoden op zich en een statistisch adviseur die help bij het kiezen van de juiste methodiek.

Statistica ondersteunt nog niet de intellimouse van Microsoft. Dat zou wel handig zijn gezien het vele aantal vensters dat op een gegeven moment aanwezig is.

In de helpfunctie is een 'animated overview' opgenomen. Daarbij wordt in multimedia show uitleg gegeven over het gekozen onderwerp. Je kunt merken dat er over is nagedacht, rechts onder in het scherm is een controlpanel opgenomen, waarmee je de show kunt stopzetten, laten pauzeren, opnieuw laten starten of versneld vooruit spoelen. Handig als je de zaken die worden uitgelegd nog een keer wilt horen of als je even rustig naar de resultaten wilt kijken.

De multimedia show is indrukwekkend.

De on-line documentatie en helpfunctie verwijst naar zo'n 150 data- en voorbeeld bestanden (afkomstig uit de statistische literatuur) die met het programma wordt meegeleverd. De gebruiker heeft daarmee steeds een adequaat voorbeeld bij de hand, waarmee hij de gekozen analysetechniek kan beproeven.

De geteste versie van Statistica bevatte nog niet de animated overview en multimedia show, maar deze werd aangeleverd via de demo CD. In de huidige CD versies zijn deze wel standaard opgenomen.

BESCHRIJVING SIS

Het Statistica Industrial System (SIS) bestaat uit de volgende modules:

Quality Control
Process Analysis
Experimental Design

SIS is een aanvulling op Statistica en is niet apart verkrijgbaar. Om met SIS te kunnen werken dient ofwel Quick Statistica of Full Statistica geïnstalleerd te zijn.

Quality Control

De Quality Control module bevat de meeste gangbare regelkaarten

- X-gemiddelde met R , s , s_2
- de attributieve kaarten c , u , np , p
- CUSUM, MA en EWMA kaarten
- Hotellings T_2 kaart
- regressie kaarten

en de mogelijkheid van een Pareto kaart. Volgens de documentatie is het mogelijk om automatische meetmiddelen aan Statistica te hangen (Statistica Connectivity Kit). Alle grafieken die gekoppeld zijn aan de dataset worden dan ook dynamisch (en in real time) aangepast. Dat kan mooie toepassingen geven, zoals regelkaarten van bepaalde grootheden bij een proces, waarbij ook kan worden aangegeven wat de oorzaak is van bijvoorbeeld waarnemingen buiten de regelgrens. Deze oorzaken kunnen dan in een bijbehorende Pareto grafiek worden bijgehouden. Met een beetje programmeren zal het ook wel mogelijk zijn om tijdens tussen meldingen van de regelkaart te registreren, waardoor ook een indicator voor de 'rust in het proces' weergegeven en ge-update kan worden.

Als je dus goed ingewerkt bent, dan is het mogelijk de SIS module in te zetten voor standaard kwaliteitszorg analyses door iedereen in de organisatie.

De patronen die worden herkend op de regelkaart als zijnde mogelijk afkomstig van afwijkend procesgedrag, kunnen worden ingesteld. Het aantal punten aan dezelfde kant van het gemiddelde, het aantal punten dat achtereenvolgens stijgend of dalend zijn op de regelkaart, de breedte van de zones (A, B en C) en de regels met betrekking tot die zones zijn allemaal in te stellen.

Tevens zijn waarschuwingslijnen in te stellen. Door de mogelijkheid iedere grafiek qua vormgeving aan te passen, zijn deze waarschuwingslijnen ook nog eens afwijkend van weergave te maken, zoals de gebruiker dat wenst (roze dikke stippellijnen zijn mogelijk).

Process Analysis

De Process analysis module bevat items voor het bepalen van de omvang van de steekproef bij ingangscntrole of partijkeuring, gegeven de gewenste risico's. Dit is niet het AQL systeem conform ISO 2859, er hoeft geen partijgrootte te worden ingegeven. De keuringskarakteristiek van bijvoorbeeld de steekproefbepaling voor een binomiaal verdeelde grootheid geeft op de x- as ook de term 'process shift in proportion defective'. Dus heeft deze module een ander doel dan beschreven is in de ISO 2859 norm over attributieve keuring van partijen.

De module bevat ook de procescapabiliteitsstudie en de analyse van de meetsystemen (measurement system analysis). De procescapabiliteitsstudie levert een veelheid aan capabiliteitsindices op.

Voor het uitvoeren van de proces analyse is het mogelijk de verdeling van de data te bekijken en te toetsen via het Kolmogorov-Smirnov procedure. Er zijn tien mogelijke verdelingen (normaal, niet-normaal (scheefheid, platheid), log-normaal, Rayleigh, exponentieel, beta, gamma, Weibull, extreme waarde verdeling). De resultaten worden weergegeven in een tabel, waarbij de verdelingen worden gerangschikt naar de waarde van de K-S grootheid. Ingeval er sprake is van niet-normaliteit, dan gebruikt Statistica de percentielpunten van de geselecteerde verdeling om de capabiliteitsindices uit te rekenen.

Ter toetsing van deze mogelijkheid om de verdelingen van de data eerst te beoordelen alvorens met een capabiliteitstest te beginnen, hebben we een standaard testset met data gebruikt van FORD. In deze datasets zijn gefingeerde waarnemingen uit bekende verdelingen gehaald en deze worden door FORD gebruikt om de juistheid van allerlei SPC software te testen. In de door ons uitgevoerde test bleek er toch niet een duidelijke overeenstemming. Data uit een Rayleigh verdeling werden terecht als non-normal aangeduid maar de Rayleigh verdeling stond pas op de derde plaats. Bij een testset met Weibull data kwam de Weibull verdeling pas op de zesde plaats als mogelijke verdelingen voor deze dataset. De waarde van deze uitkomst zou dieper moeten worden onderzocht (ligt het dan toch aan de data?), maar de waarde van deze uitgebreide mogelijkheid om de data te scannen met betrekking tot de meest waarschijnlijke verdeling moet toch ook met enige voorzichtigheid worden benaderd.

Experimental Design

De module over proefopzetten is zeer uitgebreid. Niet alleen de analyse is mogelijk, ook de proefschemata worden gegenereerd. Dit geldt voor

2^k , 3^k , en fractionering daarvan
Box-Behnken
CCD

mixture designs (2 en 3 niveaus, eventueel met constraints)
 D- en A-optimale schema's
 Taguchi methoden (met alias structuur)
 Latijnse- Griekse- en Hyper Grieks-Latijnse vierkanten

De opzet hier is steeds als bij de anderen. Je kiest een type proefschema, Statistica maakt het proefschema voor je, met de blokindeling die je zoal wilt. Je kunt na het opstarten van deze module ook kiezen voor 'Generate Design' en 'Analyze Results'. Kies je voor de eerste, dan kun je daarvoor weer een aantal wensen opgeven, bijvoorbeeld

maximum resolution (op to R specified below)
 maximum R plus maximum unconfounding
 maximum R plus minimum aberration

Na het genereren van het schema kun je weer kiezen uit een verzameling van buttons met resultaten als aliasstructuur, blokgenererende factor en dergelijke.

KRITISCHE OPMERKINGEN

Het is volstrekt onmogelijk om in het kader van deze softwarebespreking *alle* mogelijkheden van dit pakket op zijn merites te beoordelen. De aangegeven kritische punten zijn dus slechts wat wij in het gebruik hebben geconstateerd en heeft geenszins de pretentie om volledig te zijn.

Wij hebben de software vooral beoordeeld vanuit het gezichtspunt van de bruikbaarheid binnen een industriële onderneming en de datasets die daar gebruikelijk zijn. Deze recensie is *geen* poging tot vergelijking met andere pakketten. Statsoft publiceert zelf regelmatig overzichten van vergelijkende onderzoeken, waarbij Statistica wordt vergeleken met zijn concurrenten. Statistica blijkt steeds als eerste uit de bus te komen. Tevens publiceert Statsoft 'moeilijke' datasets, met de bijbehorende analyse door Statistica en door een aantal van de concurrerende pakketten. Dergelijke publicaties zijn op de WEB-site van Statsoft terug te vinden.

Beperkingen

Ook een pakket als Statistica kent uiteraard zijn beperkingen. De potentiële gebruiker beoordele zelf of deze beperkingen voor hem/haar een breekpunt vormen of niet.

Zo ontbreken mogelijkheden voor 'Bootstrap' analyses en zijn ongebalanceerde experimental designs niet ondersteund.

Ook factoriële schema's, bestaande uit zowel gekruiste als geneste classificaties, waarbij tevens zowel vaste als random factoren voorkomen, worden niet geheel correct ondersteund.

Voor industriële toepassingen zou het wenselijk zijn om naast de standaard SPC kaarten, waarvan de controlegrenzen zijn gebaseerd op

$$\bar{\bar{X}} \pm k \cdot \hat{\sigma} \text{ voor bepaalde } k,$$

te kunnen beschikken over controlegrenzen die van de kritieke waarden van de onderliggende verdeling gebruik maken (bij dergelijke grenzen is er ook een LCL voor de R- en s-kaart, ook wanneer de subgroepomvang n klein is).

Bugs

Tijdens het schrijven van deze bespreking hebben wij in de module "Multiple Regression" nog een bug aangetroffen. Na het installeren van een upgrade van versie 5.1 bleek deze bug opgelost te zijn.

Bestanden openen en sluiten

Voor het openen en sluiten van Statistica bestanden wordt helaas niet van de Microsoft conventies gebruik gemaakt. Zo is het bijvoorbeeld niet mogelijk om een Statistica sessie te beëindigen met het sluiten van het databestand. Open je een nieuw databestand dan wordt het reeds geanalyseerde bestand pas op dat moment verwijderd. Dit kan soms tot enige verwarring aanleiding geven.

Bij het opstarten van (een module van) Statistica wordt standaard het laatst geanalyseerde databestand geopend. Soms is dat erg handig, maar zeker niet in alle gevallen.

Bewerken meerdere kolommen tegelijk

Wanneer je een bepaalde bewerking op meerdere kolommen (variabelen) wilt uitvoeren door het selecteren van deze kolommen werkt dat ook niet altijd. Bij het opzetten van een 2^k experiment wordt het gegenereerde proefschema in de vorm van +1 en -1 weergegeven, maar met 5 nullen daarachter (dat laatste is wel te onderdrukken door dat in het submenu op te geven). Wil je dat wijzigen dan is er in de bovenste balk een button voor het reduceren van het aantal decimalen, maar dat werkt slechts per kolom.

Meerdere modules open

Diverse analysetechnieken zijn te kiezen via de module-switcher. Als je eenmaal voor zo'n module hebt gekozen, dan kun je alle technieken binnen de module toepassen, wil je een andere techniek toepassen (bijvoorbeeld multiple regressie analyse als je in de SIS module bezig bent), dan moet eerst de betreffende module worden opgestart. Het voordeel is wel dat je meerdere analyses (op hetzelfde databestand of op verschillende databestanden) gelijktijdig kan uitvoeren, hetgeen onderlinge vergelijking van de resultaten mogelijk maakt.

Het openen van meerdere modules kan dat leiden tot veel balkjes aan de onderkant van het scherm, waar alleen het woord STATISTICA te lezen is, de naam van de module staat daarachter en is door ruimtegebrek niet meer zichtbaar (is overigens te wijten aan Windows '95). Het ware beter de aanduiding in de balk te beperken tot het logo van Statistica en dan direct de aanduiding van de module. In de praktijk zal het wel meevallen, omdat de indeling van de modules behoorlijk afdoende is voor bepaalde typen van onderzoek. Maar toch, voor een beetje analyse van een proces heb je toch al snel de modules 'data management', 'quality control' en 'process analysis' open staan. Mogelijk voor specifieke toepassingen dan nog 'survival analysis' of 'experimental design' en dan is de onderste balk wel vol.

Dataopslag bij Meerdere modules open

Als je meerdere modules open hebt staan voor verschillende typen analyses (bijvoorbeeld quality control en process analysis) van hetzelfde databestand, dan moet je oppassen bij het aanpassen van die data tijdens de analyse. Veronderstel dat je de wijzigingen uitvoert in de module 'Data Management', dan worden die wijzigingen niet automatisch doorgegeven aan de andere modules. Je moet dan dus de dataset vanuit 'Data Management' eerst opslaan en daarna in de andere modules ophalen. Het is wel zo dat als je de dataset in 'Data Management'

wijzigt, je een melding krijgt als je dezelfde maar nog niet gewijzigde dataset in een andere module ophaalt.

Taguchi en DOE

De proefschemata's die je kan genereren in de module experimental design omvatten niet alle schema's die je bijvoorbeeld in de training-manuals van het ASI over Taguchi methoden aantreft. Met name de grotere schema's ontbreken.

Output Header

Wanneer je resultaten wil afdrukken is het verstandig om te beginnen met de pagina-indeling. Daartoe selecteer je "File" en daarna "Page/Output Setup". In het dialoogscherm kan je dan de gewenste opties selecteren en een 'output header' opgeven, een titel die dan bij elke pagina wordt afgedrukt. De laatst ingevoerde gegevens worden door Statistica onthouden, echter niet in connectie met het daarbij behorende databestand. Open je een nieuw bestand, dan heb je dus nog steeds de 'oude' paginainstellingen. Vergeet je deze instellingen te wijzigen, dan verschijnen dus de verkeerde titels en/of paginanummers op de afdrukken.

Module "Multiple Regression": Predicted Values dialog

Wanneer je een significant regressiemodel heeft gevonden voor een afhankelijke variabele en meerdere (zeg 8) onafhankelijke (verklarende) variabelen, dan wil je graag de voorspelde waarde berekenen bij bepaalde instellingen van die verklarende variabelen. Statistica biedt hiertoe de mogelijkheid via de "predicted values" dialog. Voer je de betreffende instellingen in, dan wordt de voorspelde waarde berekend, desgewenst met betrouwbaarheidsgrenzen. Statistica onthoudt de ingevoerde instellingen echter niet. Dat betekent dat je telkens opnieuw de instellingen moet invoeren, ook wanneer je slechts één instelling een beetje wil wijzigen. Dat is erg onhandig.

CONCLUSIES

Wij vinden Statistica een zeer aantrekkelijk en veelzijdig pakket, dat ook al door zijn gunstige prijsstelling met de alternatieve pakketten kan concurreren. De integratie in Windows 95 is een belangrijk pluspunt, aangezien dat het aanleren van het gebruik vergemakkelijkt. Het pakket is zeer gebruikersvriendelijk, zeker wanneer je eenmaal de structuur van het pakket doorziet. De grafische prestaties kunnen gerust verbluffend worden genoemd.

Het pakket is bovendien zeer geschikt voor de industrie, aangezien zij met aankoop van één of meerdere versies lange tijd voort kunnen. Het is voor de industrie immers minder van belang dat zij over de laatste stand van de statistische theorie kunnen beschikken, zodat de noodzaak voor updaten en upgraden minder groot is dan bij bijvoorbeeld Universitaire instellingen.

Het maken van presentaties en rapportages is met dit pakket aanzienlijk eenvoudiger geworden, ook al een voordeel voor industriële gebruikers.

Jaap Dik
OPQ Consultancy
Mierlo

Arend Oosterhoorn
Oosterhoorn Advies
Epe