

Inventarisatie van Statistische Programmatuur voor Microcomputers

J.Th. Geilenkirchen. (1)

Samenvatting.

In dit artikel wordt een 'checklist' gepresenteerd welke is opgesteld ten behoeve van een systematische en volledige inventarisatie van statistische programmatuur voor microcomputers.

De lijst is opgesteld in het kader van de activiteiten van de subgroep 'microcomputers en statistische programmatuur' van de VVS.

Inleiding.

In het najaar van 1983 is door een aantal belangstellenden een studiegroep gevormd als subgroep van de Sectie Statistische Programmatuur der Vereniging Voor Statistiek. Deze subgroep richt zich met name op microcomputers en statistische programma-tuur (2).

Deze subgroep zag als een van haar eerste doelstellingen het maken van een inventarisatie en evaluatie van beschikbare statistische programma's en programmapakketten voor microcomputers.

Om deze inventarisatie en evaluatie systematisch te laten gebeuren, is door een apart werkgroepje een lijst van onderwerpen opgesteld, welke in geregeld overleg met de gehele subgroep nu haar definitieve vorm heeft gekregen.

In dit artikel willen wij deze lijst presenteren met als doel enerzijds om mogelijke opmerkingen, aanvullingen enz. te ontvangen en anderzijds om aan deze lijst en aan het bestaan van de 'microclub' meer bekendheid te geven. Zodoende hopen wij bij onze inventarisatie ook gebruik te kunnen maken van de ervaringen van lezers van dit artikel.

De 'checklist'.

De lijst zoals deze nu is opgesteld, is opgenomen aan het einde van dit artikel. In de hier gepresenteerde versie is zij enkel inventariserend. Criteria ter waardering van een programma of pakket naar verschillende gezichtspunten, resulterend in een eindoordeel van het betreffende produkt, zijn nog niet in deze lijst opgenomen.

1): Erasmus Universiteit, Postbus 1738, 3000 DR Rotterdam,
tel. 010-525511, tst. 3275.

2): Verdere inlichtingen omtrent deze subgroep kunnen ingewonnen worden bij Th. W. de Boer, Economische Faculteit, Postbus 800,
9700 AV Groningen.

De lijst is ingedeeld in 5 hoofdonderdelen te weten:

- A: een globale omschrijving en identificatie van het programma
- B: het gegevensbeheer
- C: de statistische procedures incl. de uitvoer
- D: de communicatie tussen gebruiker en programma
- E: documentatie en ondersteuning.

De totale lijst is erg uitgebreid. De subgroep heeft zich hierbij duidelijk gerealiseerd dat het lang niet altijd mogelijk is om de lijst volledig in te vullen voor een bepaald programma; soms zullen zelfs gedeelten van de lijst helemaal niet van toepassing zijn. Van de andere kant echter kunnen bij de keuze van een programma of pakket kleine detailpunten vaak van doorslaggevende betekenis zijn.

Een mogelijk nadeel van deze volledigheid is dat de lijst enkel goed kan worden ingevuld door iemand die duidelijk al enige tijd ervaring heeft met de betreffende programmatuur; advertenties en folders geven te weinig informatie hiervoor.

In schril contrast met de uitgebreidheid van de onderdelen A, B, D en E staat onderdeel C: dit deel, dat de kern vormt van de lijst, namelijk een overzicht van de in het programma opgenomen statistische procedures, bestaat in de huidige versie slechts uit 1 regel. Reden hiervoor is dat er tientallen verschillende procedures bestaan met binnen elke procedure weer de nodige varianten en opties: het leek ons momenteel nog niet zinvol om een volledige opsomming van al deze mogelijkheden in de lijst op te nemen, waarbij dan per programma de aanwezigheid aangekruist zou moeten worden. Gekozen is nu voor de open vorm, waarbij een beschrijving gegeven kan worden van de aanwezige procedures.

Deel A: Globale omschrijving.

Dit eerste deel is bedoeld om in één oogopslag een indruk te krijgen of het betreffende programma biedt wat u zoekt of niet: vooral de korte omschrijving van de inhoud, de computersystemen waarvoor het beschikbaar is en de namen van mogelijke concurrerende programma's kunnen al meteen aangeven of het de moeite loont om de lijst verder door te nemen.

Hetzelfde kan gelden voor de prijs en voor de beoogde gebruikersgroep: is het een programma met eenvoudige statistische mogelijkheden, bedoeld voor een leek op statistisch gebied of is het juist vooral bedoeld voor de geavanceerde statisticus.

Deel B: Gegevensbeheer.

Statistische procedures bestaan natuurlijk al heel lang. Tegenwoordig, en vooral ook bij microcomputers, wordt de organisatie rond de procedures echter steeds belangrijker. Een echt statistisch pakket moet dan ook goede mogelijkheden bieden om gegevens in te voeren, te veranderen, te bewaren, aan te vullen enz., waarbij dit alles op een 'gebruikersvriendelijke' manier moet kunnen gebeuren.

In deel B wordt nagegaan welke mogelijkheden het produkt biedt op dit terrein, terwijl in deel D wordt nagegaan hoe 'gebruikers-vriendelijk' dit gebeurt. Deze veelgebruikte, vage term wordt in deel D nader geconcretiseerd.

Deel C: Procedures.

Het eerste onderdeel, het overzicht van de geïmplementeerde procedures, is om al genoemde redenen volledig open gelaten: hier kan bijv. een kopie van de inhoudsopgave van de handleiding worden bijgevoegd.

Het tweede deel, de betrouwbaarheid van de procedures, is ook nog open gelaten. Dit onderdeel, dat al meer beoordelend dan inventariserend is, moet nog verder uitgewerkt worden door de microclub. Momenteel wordt in dit verband al gewerkt aan een inventarisatie van bestaande testpakketten om de betrouwbaarheid en nauwkeurigheid te meten.

Het derde deel betreft de uitvoer van de resultaten. De hierin opgenomen punten spreken verder voor zich.

Deel D: Communicatie.

Zoals al onder B gezegd, geeft dit onderdeel een uitwerking van het begrip 'gebruikersvriendelijkheid': is het programma star, traag en moet de gebruiker continu de handleiding raadplegen om te zien wat er moet gebeuren. Wordt de gebruiker in onzekerheid gelaten of het systeem voor hem aan het werk is of dat het 'hangt'; moet hij na een kleine fout weer helemaal opnieuw beginnen enz.

Deel E: Documentatie en ondersteuning.

Dit laatste deel betreft niet het programma zelf, maar de ondersteuning welke de (potentiele) gebruiker kan krijgen.

Hierbij valt aan een groot aantal mogelijkheden te denken: cursussen voor beginners en gevorderden, een goede, nederlandse, handleiding, een goed, liefst ook nederlands, 'reference-manual', literatuurverwijzingen enz., maar ook de in deel A al gestelde vragen naar een nederlandse dealer en andere gebruikers in Nederland (of zelfs een gebruikersclub) horen hierbij.

Tot Slot.

In de inleiding is al gezegd dat deze lijst opgesteld is als basis voor een uniforme inventarisatie van statistische programmatuur en programmapakketten voor microcomputers.

De onderwerpen welke erin zijn opgenomen, zijn zeker niet willekeurig gekozen maar zijn vaak, zij het impliciet, een uitdrukking van de wensen van de subgroep m.b.t. wat minimaal in een 'ideaal' statistisch pakket aanwezig behoort te zijn.

Dit blijkt al duidelijk uit vragen welke de woorden 'gemakkelijk'

of 'logisch en duidelijk' bevatten (zie. bijv. B.3., D.1.e., D.2.b. enz.), terwijl sommige vragen misschien zelfs enigszins tendentieus gesteld zijn zoals bijv. vraag D.3.c.: "is het mogelijk dat een bestand ongewild wordt vernietigd".

Deze aanpak bergt het gevaar in zich dat in de lijst bepaalde hobby's van de samenstellers te veel benadrukt worden, terwijl anderzins hobby's juist weggelaten of te summier behandeld zijn. Om deze reden is de subgroep dan ook zeer geïnteresseerd in commentaar van lezers: schrijver dezes zal hierbij als postbus voor de microclub fungeren.

Daarnaast is de microclub met betrekking tot de inventarisatiefase natuurlijk eveneens zeer geïnteresseerd in ervaringen van lezers met verschillende programma's. Kontaktpersoon hierbij is B. Musegaas (Centrum voor Data-analyse, Varkensmarkt 2, 3511 BZ Utrecht), zowel voor informatie m.b.t. reeds geïnventariseerde software als voor inlichtingen over andere software, het beoordelen ervan enz. Een rapport van de hand van B. Musegaas met een voorlopig nog globaal overzicht van in totaal 77 programma's is al verschenen bij het Centrum voor Data-analyse.

Uiteindelijk stelt de subgroep zich hierbij tot doel om te komen tot een database waarin een Kritisch overzicht van de bestaande statistische programmatuur voor microcomputers is opgeslagen, waarna de club een voorlichtende functie kan vervullen t.b.v. (potentiele) gebruikers.

Tenslotte, hoewel de hier gepresenteerde lijst duidelijk bedoeld is voor statistische programmatuur, menen wij dat grote delen ervan (misschien met enige wijzigingen) ook toepasbaar zijn op andere gebieden. Mocht dit inderdaad het geval zijn en mocht deze lijst meer toepassing vinden, dan is hiermee een goede stap gezet in de richting van een gesystematiseerde, volledige en uniforme, (en daardoor gemakkelijk te raadplegen), documentatie van software.

De lijst ter inventarisatie van statistische programmatuur.

A. Globale omschrijving.

1. naam van het programma
2. a. korte omschrijving van de inhoud
b. beoogde gebruikersgroep
3. a. naam en adres leverancier
b. naam auteur(s)
c. naam en adres van importeur/dealer(s) in Nederland
4. a. prijs
b. is de handleiding in deze prijs inbegrepen
c. is de handleiding ook apart verkrijgbaar
d. prijs van de handleiding
5. a. recht van verspreiding (bijv. binnen onderneming)
b. recht van gebruik
c. recht en mogelijkheden om het te copieren
d. overige leverings- en gebruiksvoorwaarden
6. a. op welke computersystemen is het beschikbaar
b. voor elk van deze systemen:
wat is de minimaal vereiste configuratie m.b.t.:
1. hardware
2. operating system
3. overige software
c. wat is de gewenste/aanbevolen configuratie m.b.t.:
1. hardware
2. operating system
3. overige software
d. kan de gebruiker het op een ander systeem overzetten
zo ja, hoe gemakkelijk (zie ook de punten 7 en 8)
7. a. welke programmeertaal (-talen)
b. hoeveel in hogere programmeertaal en hoeveel in assembler
c. 1. is source listing beschikbaar
zo ja, is hij goed leesbaar, goed gedocumenteerd
2. in hoeverre is gebruik gemaakt van machine-afhankelijke taalvarianten
8. omvang van het programma in Kbytes/aantal schijven
9. a. jaar van eerste verschijnen
b. welke versie wordt nu geleverd
c. jaar van verschijnen van huidige versie
d. in hoeverre is het in de tussentijd uitgebreid, aangepast enz.
e. mate van verspreiding
f. zijn er gebruikers in Nederland
g. is demonstratie en/of bruikleen bij een dealer mogelijk

10. a. zijn er concurrerende produkten (met ongeveer dezelfde inhoud en geschikt voor een soortgelijk systeem)
zo ja, welke
 - b. wat zijn de specifieke voordelen van dit produkt t.o.v. deze concurrenten
11. a. op basis van welke informatie is deze lijst ingevuld
 - b. door wie is deze lijst ingevuld

B. Gegevensbeheer.

1. de datamatrix
 - a. maximaal aantal rijen/cases
 - b. maximaal aantal kolommen/variabelen
 - c. maximaal aantal elementen
 - d. per variabele:
 1. zijn alfanumerieke variabelen mogelijk
 2. is het formaat door de gebruiker te specificeren
 3. zijn boven- en ondergrens door de gebruiker te specificeren
 4. zijn er ook andere ingebouwde controles
 5. zijn 'missing values' te specificeren
zo ja, hoeveel per variabele
 6. zijn variabele-namen mogelijk
zo ja, hoe lang mogen deze maximaal zijn
 7. zijn 'value-labels' mogelijk
zo ja, hoe lang mogen deze maximaal zijn
 - e. per case:
 1. zijn case-namen mogelijk
zo ja, hoe lang mogen deze maximaal zijn
 - f. is ook gegevensbeheer anders dan in de vorm van een rechthoekige datamatrix mogelijk:
 1. als database-systeem
 2. in de vorm van tabellen (bijv. frequentieverdelingen)
 3. andere matrixvormen (bijv. correlatiematrix)
 4. tijdreeksen
 5. overige (bijv. variabele case-lengte)
 - g. overige opmerkingen
2. invoer
 - a. invoer via het toetsenbord
 1. bedieningsgemak (wordt bijv. aangegeven welke variabele je nu moet intypen enz.)
 2. kan invoer naar keuze per case of per variabele
 3. indien per case, moeten dan alle variabelen ingetypt worden
 4. indien per variabele, moeten dan alle cases ingetypt worden
 5. kunnen later nieuwe cases toegevoegd worden
 6. kunnen later nieuwe variabelen toegevoegd worden
zo ja, wat gebeurt met deze nieuwe variabelen in de al bestaande cases
 - b. invoer van schijf of cassette
 1. zijn bestanden van andere programma's geschikt (te ma-

- ken) als invoerbestand voor dit programma
 - zo ja, hoe gemakkelijk
- 2. is het mogelijk om andere bestanden aan de datamatrix te 'mergen' en/of te 'appenden' (resp. als nieuwe kolommen en nieuwe rijen toe te voegen)
- 3. correcties
 - a. is het mogelijk om correcties aan te brengen in de ingevoerde gegevens
 - 1. tijdens het invoeren
 - zo ja, hoe gemakkelijk
 - 2. achteraf
 - zo ja, hoe gemakkelijk
- 4. berekeningen/selecties/transformaties/sorteren
 - a. zijn deze binnen het programma gedefinieerd of door de gebruiker op te geven
 - b. indien vooraf gedefinieerd:
 - 1. hoeveel functies zijn ingebouwd en welke
 - c. indien door de gebruiker op te geven:
 - 1. in welke taal
 - 2. welke mogelijkheden bestaan hierbij
 - 3. hoe gemakkelijk werkt dit
 - d. wordt het resultaat van een transformatie/berekening naar keuze aan een bestaande of nieuwe variabele toegekend
 - e. zijn ook tijdelijke berekeningen/transformaties enz. mogelijk (d.w.z. zonder het bestand permanent te veranderen)
 - f. in hoeverre is sorteren van (sub-)bestanden mogelijk naar een of meer variabelen en/of case-namen
- 5. verwijderen van rijen/kolommen
 - a. blijven verwijderde rijen of kolommen nog in het bestand
 - zo ja,
 - 1. bestaat de mogelijkheid om de verwijdering weer ongedaan te maken
 - zo ja, hoe gemakkelijk
 - 2. bestaat de mogelijkheid om de verwijdering definitief te maken
 - zo ja, hoe gemakkelijk
- 6. uitvoer van het gegevensbestand
 - a. is uitvoer naar het scherm mogelijk
 - zo ja, is het aantal tekens per regel te specificeren
 - b. is uitvoer naar de printer mogelijk
 - c. is het mogelijk te switchen van scherm naar printer en vv.
 - d. kan de gebruiker de volgorde van uitvoer zelf specificeren (bijv. sorteren van cases en andere volgorde van variabelen)
 - e. zijn er 'report'-opties beschikbaar (bijv. paginering, titels, labels)
 - f. is het mogelijk om werkbestanden te maken en deze apart te bewaren
 - g. bij het opslaan van het bestand op schijf
 - 1. welke informatiecode wordt hierbij gebruikt
 - 2. is het mogelijk hierbij een format op te geven

- h. kan het gegevensbestand als invoer voor andere programma's dienen
zo ja, hoe gemakkelijk

C. Procedures.

1. overzicht van de beschikbare procedures
2. betrouwbaarheid van de algoritmes
3. uitvoer van de procedures
 - a. zijn de resultaten (geheel of gedeeltelijk) als bestand op schijf weg te schrijven (bijv. de residuen bij regressie)
 - b. kan de uitvoer aan het gegevensbestand toegevoegd worden
 - c. is uitvoer naar de printer mogelijk
zo ja, welke opties bestaan hierbij (titels, paginering, labels, schaalindeling enz.)
 - d. is grafische uitvoer mogelijk
zo ja, welke opties bestaan hierbij (kleuren, tekst enz.)
kan een tekening ook op de printer worden afgedrukt
kan een tekening ook op schijf worden opgeslagen
 - e. is gebruik van de plotter mogelijk
zo ja, welke opties bestaan hierbij
 - f. is het mogelijk te switchen van scherm naar printer en vv.
4. overig
 - a. hoe gemakkelijk kan de gebruiker van de ene procedure naar de andere overstappen
 - b. hoe gemakkelijk kan de gebruiker van gegevensbeheer naar procedures overstappen en terug
 - c. hoe vaak wordt hierbij van schijf gelezen of naar schijf geschreven en hoe lang duurt dit
 - d. is het mogelijk om eigen procedures in het programma op te nemen

D. Communicatie met het programma.

1. algemeen
 - a. is het programma zelfverklarend
 - b. is de communicatie af te stemmen op de deskundigheid van de gebruiker
 - c. zijn er help-opties beschikbaar
zo ja, zijn deze qua niveau af te stemmen op de deskundigheid van de gebruiker
 - d. is de schermindeling duidelijk
 - e. wordt een logisch en duidelijk gebruik gemaakt van verstekwaarden
 - f. is het 'command-' of 'menudriven'
 - g. wordt er steeds een overzicht gegeven van de mogelijke commando's en van de status van het systeem
 - h. wordt een logboek van alle opdrachten bijgehouden
 - i. is het mogelijk om eigen definities te maken
 - j. is het mogelijk om eigen macro's te maken

2. taal/nomenclatuur

- a. welke taal (nederlands/engels/overig)
- b. wordt er een logisch en duidelijk gebruik gemaakt van funktietoetsen
- c. wordt er een logisch gebruik gemaakt van geluid
- d. zie B.1.d. en B.1.e. voor gebruik van namen en labels

3. fouten

- a. mate van beveiliging tegen invoerfouten (bijv. een letter intypen i.pl.v. een cijfer)
- b. mate van beveiliging tegen systeemfouten (bijv. delen door nul)
- c. is het mogelijk dat het bestand ongewild wordt vernietigd
- d. kan het voorkomen dat veel werk (bijv. veel invoer) voor niets is geweest door een 'system crash'
- e. zijn er, binnen het programma blijvend, doorstartmogelijkheden na een fout
- f. zijn er correctiemogelijkheden na een foutief commando

4. tijd

- a. 1. wordt bij het starten het programma in zijn geheel ingelezen
- 2. wordt bij het starten het bestand in zijn geheel ingelezen
- b. snelheid van de verschillende onderdelen
- c. bij langdurige jobs:
 - 1. is er een waarschuwing vooraf dat het lang gaat duren
 - 2. wordt een schatting gegeven van de benodigde tijd
 - 3. wordt steeds aangegeven dat hij bezig is
 - 4. wordt steeds aangegeven hoever hij is
 - 5. bestaat de mogelijkheid om het programma te onderbreken
- d. kan met 'command-files' gewerkt worden
- e. is inspectie vooraf van de commando's mogelijk (droogzwemmen)

5. overig

- a. behoeft redundante informatie niet ingetypt te worden
- b. is altijd een afdruk op papier mogelijk
- c. hoe kritisch is het systeem t.o.v. de ingevoerde commando's (mag je bijv. afkorten tot 3 letters, zijn typefouten toegestaan)

E. Documentatie en ondersteuning.

1. worden voorlichtingssessies, cursussen enz. gehouden

2. de handleiding

- a. welke taal (nederlands/engels/overig)
- b. duidelijkheid (voorbeelden enz.)
- c. wordt ook de theorie behandeld
- zo ja, op welk niveau

- d. zijn de gebruikte algoritmes beschreven
- e. worden literatuurverwijzingen gegeven
- f. de kwaliteit van de handleiding als introductie
- g. de kwaliteit van de handleiding als naslagwerk
- h. overige opmerkingen

- 3. wordt er een 'reference-manual' bij geleverd
- 4. wordt er een demonstratiebestand bij geleverd met voorbeelden
- 5. is er een programmeurshandleiding bij
- 6. is er een lijst van gedocumenteerde fouten bij en wordt deze lijst bijgehouden
- 7. overige ondersteuning door leverancier: bulletin enz.
- 8. overige opmerkingen.

Ontvangen: 28-12-1983