

Naar een Sectie Statistische Programmatuur
de geschiedenis van de WSC en de CSP

Ton van Meurs ¹⁾

Leo van der Weele ²⁾

Samenvatting

De voorgeschiedenis van de Sectie Statistische Programmatuur van de VVS begon reeds in 1966 met de oprichting van de Werkgroep Statistiek en Computer (WSC). De WSC heeft een vijftal jaren met wisselende intensiteit en succes bestaan. Na de opheffing van de WSC was er op het gebied van het georganiseerd overleg op het raakvlak van de statistiek en de computer een vacuum van enige jaren dat werd afgesloten met de oprichting van de Contactgroep Statistische Programmatuur (CSP) in 1976. De geschiedenis van de WSC en de CSP tot haar formele opheffing per 1 januari 1982 wordt in dit artikel in enige feiten en beschouwende opmerkingen verhaald.

De Werkgroep Statistiek en Computer

De oprichting

Tijdens de Statistische Dag in 1965 besprak Van Reeken (werkzaam op de Katholieke Hogeschool in Tilburg) met Wiggers (AKZO-medewerker en bestuurslid van de VVS) een plan om te komen tot een sectie "Statistiek en Computer". De aanleiding hiertoe was een idee van Van Reeken dat een bundeling van mankracht voor de programmering van computerprogramma's van statistische technieken noodzakelijk was. Besloten werd om een oproep te plaatsen in "Statistisch Nieuws", de voorloper van het VVS-bulletin. Hierop reageerden vijf VVS-leden.

Een jaar later, op de Statistische Dag van 1966, deed Van Reeken opnieuw een poging, en ziedaar: het aantal reacties was nu zo groot dat besloten werd een oprichtingbijeenkomst te beleggen. In deze bijeenkomst, die in april 1966 plaatsvond, zette Van Reeken zijn ideeën uiteen, en zij vonden in grote trekken bijval.

1) N.V. Philips Eindhoven

2) Rekencentrum Rijksuniversiteit Groningen

Men werd het eens over de oprichting van een Werkgroep Statistiek en Computer (WSC) die zich voorlopig ten doel stelde :

- . de vervaardiging van een overzicht van statistische computerprogramma's in Nederland;
- . de bespreking van deze programma's;
- . de vervaardiging van nieuwe programma's.

In februari 1967 werd door het VVS-bestuur de status van de Werkgroep binnen de VVS aan de orde gesteld. Naar aanleiding hiervan werd een bestuur geformeerd en werden de definitieve doelstellingen vastgelegd. In ruil daarvoor kon het VVS-bestuur helaas alleen maar morele steun aanbieden, want geld had de VVS niet.

Hoe moesten de doelstellingen van de Werkgroep luiden? Om daar achter te komen werd een enquête onder de leden gehouden, met als resultaat:

- . Het vergemakkelijken van de programmering van statistische analysetechnieken alsmede het gebruik van statistische standaardprogramma's, ten dienste van statistici en rekencentra.

Hieruit werd toen een viertal taken afgeleid:

- . De oprichting van Werkcommissies die zich bezig moesten houden met de vastlegging van de eisen waaraan een standaardprogramma moet voldoen, de vervaardiging van dit programma (inclusief de wiskundige hulproutines), de vervaardiging van een programmabeschrijving, en de opsomming van de toepassingsvoorwaarden.
- . De organisatie van een centrale programmatheek in de vorm van een inventarislijst, om de uitwisseling van programma's te bevorderen.
- . Het beleggen van discussie-bijeenkomsten voor de uitwisseling van ideeën en ervaringen.
- . Het organiseren van voordrachten, bedoeld als voorlichting aan de leden van de Werkgroep en andere belangstellenden.

Het zal de lezer opvallen dat het begrip "Statistisch Pakket" in de doel- en taakstellingen nog geheel ontbreekt. Dit is niet zo verwonderlijk gezien het feit dat dit soort statistische software nog maar langzaam begon te ontstaan en zo er al iets van bekend was, het nog lang niet die mate van gebruikersvriendelijkheid bleek te bezitten die men hedentendage kan waarnemen.

Het functioneren

Hoe productief was de Werkgroep? Daarover verschilden de meningen. In ieder

geval waren er tot januari 1969 zestien bijeenkomsten geweest, waarin de leden over bepaalde problemen discussieerden, of waarin voordrachten werden gehouden. Het accent lag vooral op de voordrachten.

Daarnaast waren overeenkomstig de afgesproken taakstelling een aantal werkcommissies ingesteld die zich bezig hielden met specifieke onderwerpen, te weten: regressieanalyse, variantieanalyse, verdelingsfuncties en multivariate analyse. Bovendien bestond een latente behoefte aan commissies voor factoranalyse en covariantieanalyse.

De laatste twee zijn nooit van de grond gekomen, de voorlaatste twee hebben nauwelijks gefunctioneerd en de commissie variantieanalyse werd wegens een onvoldoende aantal leden op non-actief gesteld voor dat een tastbaar resultaat tot stand kon komen. De commissie regressieanalyse bracht het er nog het best van af: er kwam een ontwerp-rapport tot stand (6). Helaas is er nooit een definitieve versie van verschenen.

Overigens mag wel vermeld worden dat de Stichting het Nederlandse Studiecentrum voor Administratieve Automatisering in samenwerking met de Werkgroep een publikatie van 210 programma's aan het voorbereiden was.

Was de toestand inderdaad zo somber als uit het voorgaande zou kunnen worden afgeleid? Misschien wel, want de gewenste produktiviteit liet in vele opzichten te wensen over.

De al dan niet geplande discussies leverden geen concrete resultaten op. Er bestond zo'n grote verscheidenheid van computers en programmeertalen, en zo'n grote verscheidenheid in belangstelling van de leden (statistici enerzijds, computerdeskundigen anderzijds), dat de discussies veelal ontaardden in monologen zonder onderling verband.

Op een enkele uitzondering na waren de werkcommissies niet actief, of lieten de resultaten te lang op zich wachten. Een belangrijke oorzaak hiervan was het tijdgebrek bij de deelnemers. Bovendien werden de irrelevante discussies, die het kenmerk waren van de Werkgroep, binnen sommige werkcommissies voortgezet. En de regressiecommissie stagneerde bovendien door wiskundige/numerieke problemen waarvoor op korte termijn geen oplossing voor handen bleek te zijn. Het was dus niet verwonderlijk dat een aantal leden hun enthousiasme verloren, en dat was zowel een oorzaak als een gevolg van de zojuist geschetste gang van zaken.

De opheffing

Toen het Werkgroepbestuur zelf ook begon te lijden aan tijdgebrek, brak een tijdperk van bezinning aan. Het bestuur erkende dat het niet in staat was de

doelstellingen van de Werkgroep te bereiken, zeker niet als de leden om welke reden dan ook niet daadwerkelijk actief waren.

Op de vraag wat er nu moest gebeuren, werd tijdens een gesprek met Wiggers in augustus 1969 getracht een antwoord te vinden. Het Werkgroepbestuur kwam met twee alternatieven aandragen: ontbinding van de Werkgroep of een nieuwe taakstelling formuleren.

Het eerste alternatief - de ontbinding van de Werkgroep - werd niet aantrekkelijk geacht. Het nut van het bestaan van een werkgroep in een of andere vorm werd door niemand ontkend. Opvallend was bijvoorbeeld de wonderbaarlijk grote belangstelling voor de bijdrage van de Werkgroep op de Statistische Dag van 1968. Daarnaast ontstonden in andere landen (in Engeland gesteund door de Royal Statistical Society en de redactie van Applied Statistics, en in de Verenigde Staten gesteund door de American Statistical Association) soortgelijke groepen, die zeer actief waren en een samenwerking met een Nederlandse groep op prijs stelden.

Met betrekking tot het tweede alternatief - nieuwe taakstelling - zou een betere basis voor de Werkgroep gevonden moeten worden in een minder ambitieuze en een meer concrete taakstelling. De aandacht zou in de eerste plaats gericht moeten zijn op het inventariseren en uitwisselen van programma's, omdat dit de leden van de Werkgroep een direct voordeel opleverde. Dit hoofddoel was tot dan toe niet snel genoeg ontwikkeld. De inventarisatie kon dan een uitgangspunt zijn voor het instellen van werkcommissies, die de essentie van de Werkgroep zouden zijn en blijven.

Aan het uitwisselen van ideeën en ervaringen zou meer kunnen worden gedaan in de vorm van korte artikelen in Statistica of in het VVS-bulletin (dat op het punt stond opgericht te worden).

Het geven van voorlichting door het organiseren van lezingen diende buiten de directe activiteiten van de Werkgroep te vallen. Tot dusver was weliswaar een aantal voordrachten gehouden die door een redelijk aantal toehoorders werden bezocht (gemiddeld 17), doch het effect en nut ervan kon betwijfeld worden.

Op 4 juni 1970 werd deze boodschap aan het VVS-bestuur overhandigd, met de vraag om er iets aan te doen. Daarop belegde dit bestuur een bijeenkomst, waarop een aantal aanwezigen beloofden pogingen in het werk te stellen de Werkgroep een nieuw leven in te blazen. Achteraf gezien hebben wij het vermoeden dat we de bedoelde leden toen verkeerd hebben begrepen, en dat zij hadden voorgesteld behulpzaam te zijn om de Werkgroep de laatste adem te laten uitblazen. Hoe dan ook, het werd in Nederland erg stil op het raakvlak van statistiek en computer.

De Contactgroep Statistische Programmatuur

De oprichting

Althans ... het werd stil op het georganiseerde vlak. De ontwikkelingen op het raakvlak van de statistiek en de computer stonden allerminst stil, vooral doordat de beoefenaars van de sociale wetenschappen de statistische programmatuur ontdekten, niet in de laatste plaats door het pakket SPSS. De aandacht voor statistische programmatuur verschoof dan ook van de ontwikkeling van programma's naar het gebruik ervan. Wiggers, inmiddels voorzitter van de VVS, haakte hierop in door in december 1975 een oproep voor belangstellenden in een contactgroep statistische programmatuur in het elfde nummer van het VVS-bulletin te plaatsen. Het aantal reacties was groot. De geschiedenis herhaalt zich: op de voorlopige oprichtingsvergadering op 8 april 1976 waren 36 belangstellenden verzameld om te komen tot een "door de VVS ondersteunde" contactgroep statistische programmatuur, de CSP. Een voorbereidingscommissie kreeg de opdracht een nadere uitwerking te geven aan de levende ideeën. De commissie hield een enquête waarop een vijftigtal reacties kwam, stelde een concept-reglement op en organiseerde een oprichtingsbijeenkomst welke op 1 september 1976 te Utrecht werd gehouden. Op deze bijeenkomst werd de Contactgroep Statistische Programmatuur definitief opgericht met als status: werkgroep van de VVS en als doelstelling: "het creëren van een forum voor allen die geïnteresseerd zijn in statistische programmatuur", een wat vager en minder ambitieus doel dan de voormalige Werkgroep zich had gesteld. De doelstelling van de CSP werd als volgt nader gespecificeerd:

- het leggen van contacten, zowel nationaal als internationaal;
- het inventariseren van beschikbare programmatuur;
- het tot stand brengen van een uitwisseling van statistische programmatuur;
- het opstellen van kwaliteitsrapporten betreffende statistische programmatuur;
- het uitwisselen van ervaringen bij de invoering, het onderwijs en het gebruik van statistische programmatuur.

Een opmerkelijk verschil met de doelstelling van de WSC wordt gevormd door het ontbreken van het vervaardigen van programmatuur.

Bij de voorstellen om tot realisatie van het doel te komen werd de nadruk gelegd op zogenaamde subgroepen, werkgroepen die zich bezig zouden houden met bepaalde samenhangende facetten van statistische programmatuur. Daarnaast werden het houden van voordrachten, het publiceren van bereikte resultaten en het leveren van bijdragen aan congressen en de Statistische Dag, als activiteiten opgevoerd.

De 59 ter vergadering aanwezigen kozen een Dagelijks Bestuur kwamen tot instelling van een aantal subgroepen, en woonden een forum bij met als thema "statistische pakketten, een zegen of een ramp?", geleid door de voormalige voorzitter van de WSC, Van Reeken.

Conform de bij de oprichtingsvergadering gelanceerde strategie, richtten de activiteiten van de CSP zich op jaarlijkse vergaderingen, congressen, publikaties en vooral: subgroepen.

Op de jaarvergadering van 1977 kwam de vraag "een zegen of een ramp?" terug in het thema "interactieve statistische programmatuur". De vraag werd behandeld door W. Crawford (SPSS, Inc.), die tevens een demonstratie gaf van SCSS.

Een jaar later kwamen de subgroepen uitvoerig aan bod met de door hen behandelde onderwerpen. Een gastspreker was uitgenodigd om te komen vertellen over EASIT (European Association for Software access and Information Transfer). De jaarvergadering van 1979 had de dienstverlening bij "software-centra" als thema, en in 1980 werd aandacht besteed aan Logit-, Logistische, Log-lineaire en LISREL-modellen en de bijbehorende programmatuur. De belangrijkste activiteiten op het gebied van congressen werden gevormd door bijdragen aan de Statistische Dag. Daarnaast waren de CSP-leden zeer actief op COMPSTAT-congressen; niet alleen als sprekers maar ook als organisatoren. Wat de publikaties betreft wordt de lezer verwezen naar de referenties aan het eind van dit artikel, waaraan kunnen worden toegevoegd de vele bijdragen van CSP leden aan het voorliggende nummer van Kwantitatieve Methoden.

De subgroepen

Zoals bij de oprichtingsvergadering van de CSP was voorgesteld, is het inderdaad uitgekomen dat de subgroepen het belangrijkste forum voor de activiteiten van de CSP vormen. Elk der subgroepen houdt zich bezig met een bepaald samenhangend onderwerp, behalve de "pakkettenclub" maar daarover straks meer. Over zo'n onderwerp worden voordrachten gehouden, wordt onderzoek gedaan zowel naar statistische methoden als naar beschikbare programmatuur en worden ervaringen en programma's uitgewisseld. Van de leden van een subgroep wordt een actieve participatie in het werk van de groep verondersteld, waaraan in het algemeen ook zeker wordt voldaan. De subgroepen bestaan uit 10 tot 35 leden die gemiddeld eens per twee maanden bijeenkomen. Tussen de bijeenkomsten wordt door enige leden gewerkt aan vooral het evalueren van methoden en programmatuur. We zullen kort de belangrijkste activiteiten der subgroepen bespreken.

Regressieanalyse

Deze subgroep heeft slechts ruim een jaar gefunctioneerd, maar kwam daarbij wel tot een rapportage over een classificatie van eigenschappen van regressieprogramma's (2).

Clusteranalyse

De clusterclub hield zich bezig met het evalueren van clustertechnieken in het bijzonder die technieken die in het CLUSTAN pakket waren opgenomen. De resultaten van de studie zijn vastgelegd in een eindrapport (7), waarna de subgroep werd opgeheven.

Tijdreeksenanalyse

De tijdreeksen-subgroep heeft vanaf de oprichting van de CSP een actief bestaan geleid, ondanks het probleem dat zowel uitgesproken theoretici als typische toepassers aan de werkzaamheden deelnemen. Bovendien kampt de groep met een gebrek aan goede en betaalbare programma's, waardoor het uitvoeren van experimenten wordt bemoeilijkt. Programma's worden geëvalueerd door bepaalde tijdreeksen met verschillende programma's op uiteenlopende computers te verwerken, waarbij opmerkelijke verschillen in uitkomsten zijnesignaleerd (5).

SPSS

De subgroep SPSS stelt zich vooral ten doel het inventariseren van fouten in de diverse SPSS versies, het bestuderen van methoden en het evalueren van methodieken om SPSS te leren gebruiken. In dat kader mag de bijdrage aan het PLATO-SPSS-project van de Universiteit van Amsterdam in samenwerking met Control Data worden genoemd. De subgroep heeft de ambitie als gebruikersgroep te functioneren, maar stuit daarbij op aanzienlijke verschillen in de voor de diverse computers beschikbare SPSS versies.

Opzet van statistische pakketten in het algemeen

Deze "pakkettenclub", opgezet om het ontwerp van pakketten en pakketten te bestuderen, heeft zich ontwikkeld tot de meest algemene subgroep, in bepaalde opzichten als "praatgroep". Als belangrijkste taken worden gezien: het uitwisselen van ervaringen met het gebruik van statistische programmatuur, het opstellen van criteria voor de evaluatie van programmatuur, de uitvoering van evaluaties en het voorkomen van het misbruik van pakketten. Om de gevarieerdheid van de in de pakkettenclub behandelde onderwerpen aan te geven, volgen enige onderwerpen van gegeven voordrachten: "het ontwikkelen van een bibliotheek van distributiefuncties", "defaults" (1), "portabiliteit van programmatuur", "het testen van programma's", "het schatten van varianties", terwijl ook enige pakketten, zoals P-STAT, STAP en WESP, en programmabibliotheken als KUNST en LISTOR werden besproken.

MANOVA

De subgroep MANOVA bestudeert de theorie van multivariate variantieanalyse en voert experimenten uit met diverse variantieanalyse programma's, terwijl ook enige praktijkproblemen onderwerp van bespreking zijn geweest.

MDS

De subgroep voor multidimensionale schaalanalyse heeft zich beziggehouden met de evaluatie van MDS programmatuur door toepassing van diverse programma's op een aantal data-verzamelingen (4). De resultaten van de evaluatiestudies zijn vastgelegd in een rapport dat goede diensten kan bewijzen bij de keuze van een methode en een programma voor een gegeven schaalanalyse probleem (3).

Terugblik

Terugblikkend op vijf jaar CSP moet men zich onherroepelijk de vraag stellen: "CSP, een zegen of een ramp?", ofwel: is de CSP erin geslaagd aan het gestelde doel te beantwoorden. Het belangrijkste doel, het creëren van een forum voor allen die geïnteresseerd zijn in statistische programmatuur, is ongetwijfeld bereikt: een tachtigtal leden heeft zich actief met de CSP beziggehouden, vooral door participatie in de subgroepen, maar ook op de jaarvergaderingen, terwijl leden van de CSP in groten getale acte de presence hebben gegeven op bijeenkomsten als de Statistische Dag en de COMPSTAT congressen. Hebben deze activiteiten nu ook geleid tot concrete resultaten? Voor een aantal deelnemers aan de CSP ongetwijfeld wel: er is een hoop geleerd van de voordrachten en van enige onderzoekingen en verder zijn er zodanige contacten tot stand gebracht dat uitwisseling van statistische programmatuur en ervaringen met "computational statistics" zeker is vergemakkelijkt. Hiermee is ook de gebruiker van statistische programma's ontegenzeggelijk geholpen. Op het gebied van het vooruitbrengen van het vakgebied "computational statistics" zijn de resultaten echter enigszins teleurstellend; er zijn zeker wel resultaten geboekt, zoals de evaluatie-rapporten van de subgroepen Clusteranalyse en MDS, er is meer inzicht gegeven in het defaultprobleem (1) en in missing data (9), (10), terwijl ook van enige studies die nog aan de gang zijn concrete resultaten mogen worden verwacht. Echter een meer structurele aanpak van bijvoorbeeld evaluatie van beschikbare programmatuur is te weinig van de grond gekomen. De oorzaak hiervan is duidelijk: "computational statistics" wordt in Nederland nog niet erkend als (toegepaste) wetenschap, waardoor de werkzaamheden op dit terrein zich te veel hebben moeten afspelen in de sfeer van liefhebberij en moeten worden uitgevoerd in de vrije tijd van de leden van de CSP. Wel zijn de omstandigheden zeer veel gunstiger dan in de tijd van de WSC: veel meer mensen

houden zich bezig met statistische programmatuur en er zijn veel meer programma's beschikbaar, waardoor de behoefte aan in CSP-verband het zelf ontwikkelen van programma's op een laag pitje kon worden gezet. Er is een verschuiving van programmering naar toegepaste statistische methoden en vooral naar het gebruik van programmatuur en methoden voor onderwijs te constateren. De ontwikkeling dat steeds meer "leken" programma's gebruiken, waardoor de kans op misbruik sterk wordt verhoogd, maakt het belang van kritische studies naar de kwaliteit van statistische programmatuur echter alleen maar groter. De tijd is voor de CSP wat dat betreft echter gunstig. De VVS is geïnteresseerd in het opnemen van statistische programmatuur in de opleidingen voor statistisch assistent en analyst en statisticus VVS. Ook de pogingen aan verschillende universiteiten om te komen tot een studierichting Sociaal Wetenschappelijke Informatica betekenen een belangrijke stap naar volwassenheid van de studie van statistische programmatuur in al zijn facetten. Het kanaliseren van onderzoek op dit gebied door de NOSMO (= Nederlandse Organisatie voor Sociaal-Wetenschappelijk Methodologisch Onderzoek) (12) kan een belangrijke stimulans voor werkers op het terrein van "computational statistics" betekenen.

Sectie Statistische Programmatuur van de VVS

De overgang van de CSP naar de status van sectie van de VVS heeft enerzijds een formele reden, namelijk het wegnemen van een onduidelijke en niet in de structuur van de VVS passende status, maar anderzijds moet dit initiatief van het VVS-bestuur worden opgevat als een erkenning van het belang van statistische programmatuur als wezenlijk onderdeel van de (toegepaste) statistiek. Deze formele wijziging in de status kan een belangrijke stap zijn naar een bloeiende contactgroep danwel sectie statistische programmatuur, waarvan dit artikel de voorgeschiedenis in enige feiten en beschouwende opmerkingen heeft verhaald.

Referenties

- (1) N. Bendermacher, A. Cappel, F. van Nes (red.), D. van der Sluis, G. van der Vorst: "Defaults, een zegen of een ramp?", MDN (1978)3.
- (2) J. Bethlehem: "Classificatie van de eigenschappen van eenheden regressie-analyse-programmatuur", VVS Bulletin 11 (1978)1.
- (3) U. Brouwer, P. Debets, W. van Schuur: "Multidimensional Scaling: some evaluations concerning the distance model in MDS programs, using symmetric input matrices", (1980).
- (4) U. Brouwer, P. Debets: "Vergelijken van MDS programmatuur", VVS Bulletin 14 (1981)5.
- (5) M.C. Gels: "Voorlopige resultaten van de subgroep tijdreeksenanalyse", VVS Bulletin 12 (1979)6.
- (6) A. van Meurs, A.J. van Reeken et al: "Regressie-analyse" (ontwerp-rapport, 1970).
- (7) A. van Meurs (red.): "Clusteranalyse", (1978).
- (8) A.J. van Reeken: "Report on the Work of the Dutch Working Party on Statistical Computing", Statistical Computing (1970).
- (9) D.M. van der Sluis: "Statistical programs and missing data", Kwantitatieve Methoden 2 (1981)2.
- (10) G.P. van der Vorst, D.M. van der Sluis: "Statistische programmatuur en ontbrekende gegevens", VVS Bulletin 12 (1979)6.
- (11) L.Th. van der Wee: "Statistische programmatuur, een zegen of een ramp?", VVS Bulletin 12 (1979) 4.
- (12) "Oprichting Stichting NOSMO", VVS Bulletin 14 (1981)7/8.