

RIVM rapport 422 516 001

## **S-PLUS voor het RIVM**

*Krachtig statistisch software gereedschap*

A.L.M. Dekkers

mei 2001



Dit onderzoek werd verricht in het kader van project nr. 422516.



## **SAMENVATTING**

Statistische analyses worden op het RIVM veelvuldig uitgevoerd en gerapporteerd. Een grote hoeveelheid goede software is er op de markt om deze analyses betrouwbaar uit te voeren. Voor een goede communicatie tussen statistici en inhoudelijke onderzoekers, die statistiek binnen hun onderzoek toepassen, is het van belang dat het RIVM slechts enkele standaard pakketten heeft. Bovendien levert dit financiële voordelen op. Zo is binnen sector 5 sinds 1992 S-PLUS het standaard statistisch pakket en binnen sector 2 al sinds jaren SAS met S-PLUS als snelle ontwikkelsoftware.

Vanaf 1 januari van dit jaar bezit het RIVM een geheel nieuwe licentie van S-PLUS, waarbij het pakket op alle platforms vrij beschikbaar is voor de sectoren 2, 3/4 en 5 inclusief het gebruik van alle modules én thuisgebruik.

Deze nieuwe situatie is de aanleiding geweest tot het schrijven van dit rapport.

De eerste drie hoofdstukken bevatten algemene informatie over S-PLUS en zijn geschreven voor potentiële gebruikers én voor het management. Hierbij wordt gedacht aan labhoofden, RIShoofden en projectleiders van laboverstijgende projecten zoals de Milieubalans en de Milieuverkenningen. In deze hoofdstukken komt de brede inzetbaarheid van S-PLUS aan de orde als statistisch pakket, grafisch pakket en 4e generatietaal. Het overzicht van enkele S-PLUS toepassingen binnen RIVM projecten spreekt voor zich.

De overige hoofdstukken zijn speciaal voor gebruikers geschreven en bevatten een schat aan inhoudelijke informatie. Het betreft o.a. de installatievoorschriften van S-PLUS voor thuisgebruik en gebruik op het RIVM. Het rapport wordt afgesloten met 6 uitvoerige bijlagen waarin twee inleidende cursussen over de S-PLUS commando taal en over de S-PLUS grafische user interface zijn opgenomen, alsmede een bijlage voor SPSS gebruikers die met S-PLUS gaan beginnen.

Dit rapport beoogt een standaard gebruik van S-PLUS te bevorderen voor goede, reproduceerbare statistische analyses en voor optimale communicatie tussen onderzoekers.



## SUMMARY

### *The Statistical Analysis package S-PLUS*

Statistical analyses are carried out and reported on in abundance at the RIVM. There is a lot of good software on the market for executing these analyses reliably. For good communication between statisticians and scientific researchers, who use statistical applications in their research, RIVM needs only a few standard packages. This also has financial advantages. For example, S-PLUS has been used since 1992 as the standard package in Division 5 and moreover in Division 2, as rapid development software combined with SAS. On 1 January of this year RIVM took out a completely new licence for S-PLUS, making the package available for Divisions 2, 3/4 and 5 at no extra costs, including the use of all the modules and the use of S-PLUS at home.

This new situation was one of the reasons for preparing this document. The first three chapters, written for potential users and management, e.g. laboratory and department heads, IT heads in the Division 5 departments, and managers of interdepartmental projects like the Environmental Balance and Outlooks. These chapters deal with the broad application of S-PLUS as statistical and graphics package, and as a fourth-generation language. The overview of several S-PLUS applications within RIVM speaks for itself.

The other chapters, providing a wealth of substantial information, are written especially for users. Included here are the installation instructions of S-PLUS for both home use and use at the RIVM. The report closes with four elaborate appendixes incorporating two introductory courses on the S-PLUS command language and the S-PLUS graphics user interface.

This report is thus intended to promote the standard use of S-PLUS for good-quality, reproducible statistical analyses and optimal communication between researchers.



## INHOUD

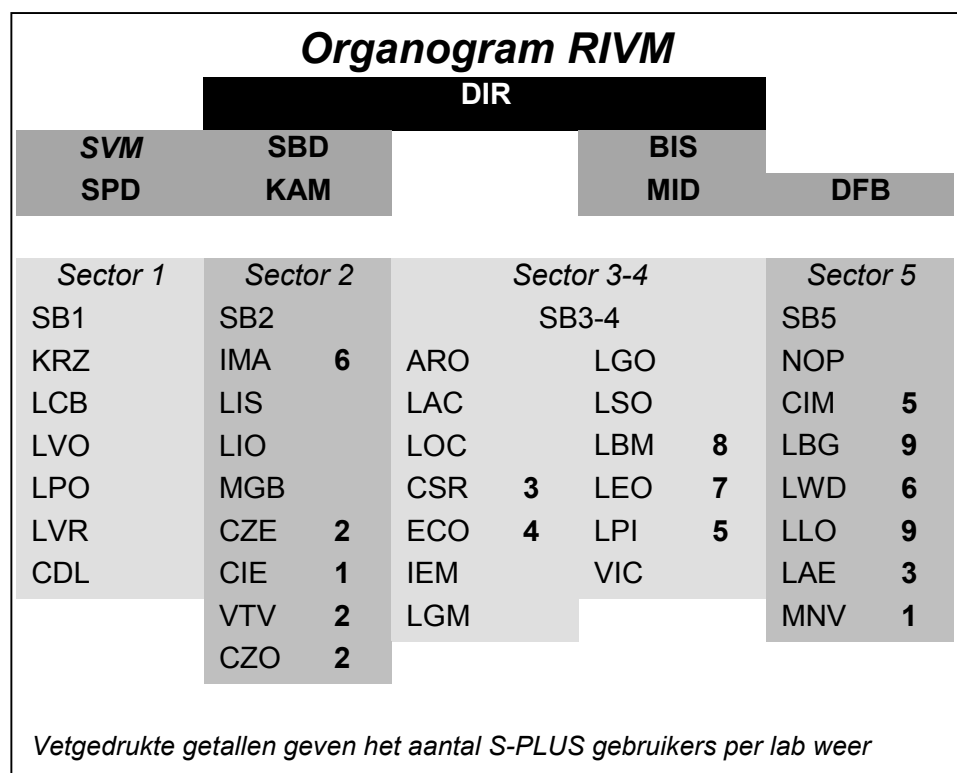
<b>1. INLEIDING</b>	<b>9</b>
<b>2. S-PLUS 2000</b>	<b>11</b>
2.1 S-PLUS 2000 eigenschappen	11
2.2 S-PLUS toepassingen op het RIVM	12
<b>3. MODULES VAN S-PLUS</b>	<b>15</b>
3.1 S+Dox	15
3.2 EnvironmentalStats	15
3.3 S+Garch	15
3.4 S+Nuoqt	15
3.5 S+SeqTrial	15
3.6 S+SpatialStats	15
3.7 S+Wavelets	16
3.8 S+ArcViewGis	16
<b>4. S-PLUS 2000 OP HET RIVM</b>	<b>17</b>
4.1 Back-up van de eigen S-PLUS gegevens en verwijdering oude versie S-PLUS	17
4.2 Installatie S-PLUS 2000 release 3	17
4.3 Installatie van de modules	18
4.3.1 <i>Probleem: er gebeurt niets na het aanklikken van de NAL S-PLUS Modules</i>	19
4.3.2 <i>Opstarten modules / libraries binnen S-PLUS</i>	19
4.3.3 <i>Beschikbaarheid helpbestanden binnen S-PLUS</i>	20
4.4 Installatie van de S+ArcViewGis	20
4.5 De RIVM S-PLUS bibliotheek RIVMfun	20
4.6 RIVM ondersteuning bij het gebruik van S-PLUS	21
4.7 Beginnen met S-PLUS voor SPSS gebruikers	21
<b>5. S-PLUS 2000 THUIS</b>	<b>23</b>
5.1 Regeling thuisgebruik	23
5.2 Installatie stand-alone versie onder Windows 9x	23
5.2.1 <i>Openen bestand met installatiecodes</i>	23
5.2.2 <i>Installatie S-PLUS</i>	23
5.2.3 <i>Installatie modules</i>	25
5.2.4 <i>Afzonderlijke installatie Acrobat Reader</i>	25
5.2.5 <i>Demo S-PLUS</i>	25
5.2.6 <i>Inhoud installatie CD benaderen via Explorer</i>	25
<b>6. VERWIJDEREN S-PLUS 2000</b>	<b>27</b>
6.1 Verwijdering van modules	27
6.2 Verwijdering van S-PLUS 2000	27
<b>7. S-PLUS INSTELLINGEN EN BESTANDSORGANISATIE</b>	<b>29</b>
7.1 Algemene S-PLUS instellingen	29
7.2 Gebruik en organisatie van S-PLUS via de commandotaal	31
7.3 Gebruik en organisatie van S-PLUS via de GUI	32
7.4 Het schrijven van functies via GUI en de history	32
7.5 S-PLUS in BATCH mode	33
7.6 Wetenswaardigheden en tips	34
7.6.1 <i>Naamgeving Windows bestanden en paden vanuit S-PLUS</i>	34
7.6.2 <i>Naamgeving van S-PLUS objecten buiten S-PLUS</i>	34
7.6.3 <i>Kopiëren en/of verplaatsen van S-PLUS objecten tussen mappen</i>	34
7.6.4 <i>Systematische naamgeving S-PLUS objecten</i>	35
7.6.5 <i>Figuren vanuit S-PLUS in WORD documenten plaatsen</i>	35
7.6.6 <i>De S-PLUS functies rm en remove</i>	36
7.6.7 <i>Bewaren van uitgevoerde commando's voor een volgende S-PLUS sessie</i>	36
7.6.8 <i>Het uitvoeren en bewaren van analyses in S-PLUS</i>	37
7.6.9 <i>Het maken van back-ups en het terugzetten van back-ups in S-PLUS</i>	38
7.6.10 <i>Versiebeheer binnen S-PLUS</i>	38

7.6.11	<i>Back-up van S_PROJ op een CD en het teruglezen</i>	38
<b>8.</b>	<b>S-PLUS DOCUMENTATIE EN INFORMATIE</b>	<b>41</b>
8.1	Boeken te bestellen bij Candiensten	41
8.2	Literatuur over gebruik van en programmeren in S-PLUS	42
8.3	Literatuur over data analyse met S-PLUS	42
8.4	Literatuur over grafische weergaven in S-PLUS	43
8.5	S-PLUS via het WEB	44
<b>9.</b>	<b>DBMS/COPY</b>	<b>45</b>
	<b>VERZENDLIJST</b>	<b>49</b>
	<b>APPENDIX A UITVOERIGE BESCHRIJVINGEN VAN DE MODULES</b>	<b>53</b>
	<b>APPENDIX B INLEIDING IN DE S-PLUS COMMANDO TAAL</b>	<b>61</b>
B.1	Het commando window	61
B.2	S-PLUS als rekenapparaat	62
B.3	De functie "options" en het gebruik van een editor	65
B.4	Gebruik grafisch scherm	66
B.5	Verschillende DATASTRUCTUREN in S-PLUS	68
B.5.1	<i>Vectoren en factors</i>	68
B.5.2	<i>Matrices en arrays</i>	69
B.5.3	<i>Lists</i>	69
B.5.4	<i>Data.frames</i>	70
B.5.5	<i>Tijdreeksen</i>	71
B.6	Functies in S-PLUS	72
B.6.1	<i>Het maken van functies</i>	72
B.6.2	<i>Menu's</i>	74
	<b>APPENDIX C INLEIDING IN HET GEBRUIK VAN DE GUI</b>	<b>75</b>
	<b>APPENDIX D BEGINNEN MET S-PLUS VOOR SPSS GEBRUIKERS</b>	<b>77</b>
D.1	Inleiding	77
D.2	Bestanden en variabelen	77
D.3	Naar een data-frame	78
D.4	get file	78
D.5	save file	78
D.6	display labels	79
D.7	compute	79
D.8	(add) value labels	79
D.9	variable labels	80
D.10	rename	80
D.11	recode	80
D.12	if en do if	81
D.12	rank ntiles	81
D.14	count	81
D.15	frequencies en descriptives	82
D.16	crosstabs	82
D.17	add files	83
D.18	match file	83
D.19	missing values	83
	<b>APPENDIX E BOEKBESPREKINGEN VAN CANDIENSTEN</b>	<b>85</b>
	<b>APPENDIX F AANTEKENINGEN</b>	<b>89</b>



## 1. INLEIDING

S-PLUS 2000 is de nieuwste versie van S-PLUS, statistische software die al sinds 1992 hét standaard pakket van de sector Milieu van het RIVM is. Sinds een aantal jaren wordt het ook door de Volksgezondheidssector gebruikt als snelle ontwikkelsoftware naast hun standaard software, SAS. Tevens wordt S-PLUS intensief gebruikt door sector 3/4. Op dit moment bestaat de S-PLUS gebruikersgroep op het RIVM uit ruim 70 personen, verdeeld over 16 labs zoals uit onderstaand schema blijkt.



Sinds 1 januari 2001 heeft het RIVM een nieuwe S-PLUS licentie, bestaande uit

- onbeperkt gebruik van S-PLUS Professional onder Windows NT, UNIX en LINUX voor sector 2, 3/4 en 5
- onbeperkt gebruik van alle S-PLUS modules,
- thuisgebruik voor RIVM-ers van S-PLUS 2000 Professional en de modules.

Vanwege deze nieuwe situatie wordt in dit rapport actuele informatie gegeven over de mogelijkheden en het gebruik van S-PLUS en de bijbehorende modules ter bevordering van een adequate verspreiding en toepassing van S-PLUS binnen het RIVM. Het rapport is daarom geschreven voor een zeer uiteenlopende groep RIVM-ers die vanuit verschillende invalshoeken naar statistische software kijken en het voor verschillende doelen toepassen.

De volgende twee hoofdstukken zijn met name geschreven voor labhoofden, RISHoofden en projectleiders van laboverstijgende projecten zoals MB/MV, maar natuurlijk zijn ze ook interessant voor S-PLUS gebruikers of potentiële gebruikers. Deze hoofdstukken bevatten respectievelijk een summiere beschrijving van S-PLUS en de bestaande modules, alsmede voorbeelden over de inzet ervan bij een groot aantal uiteenlopende projecten.

De volgende hoofdstukken zijn geschreven voor de S-PLUS gebruikers, waarbij weer een scala aan type gebruikers mogelijk is: van incidenteel gebruik tot dagelijks gebruik, van mensen die veelal de commandotaal gebruiken tot degenen die vooral de menu's gebruiken. In hoofdstuk 4 komen de installatievoorschriften van S-PLUS 2000 (netwerk versie) via de

NAL onder Windows NT aan de orde, waarbij een aantal tips worden aangereikt. Vervolgens wordt beschreven hoe men de modules kan installeren via een tweede NAL en treft men een beschrijving van de RIVM S-PLUS bibliotheek, RIVMfun genaamd aan alsmede de ondersteuning die er bij het S-PLUS gebruik op het RIVM gegeven wordt. In hoofdstuk 5 staat de regeling voor het thuisgebruik van S-PLUS beschreven en in hoofdstuk 6 komt het verwijderen van afzonderlijke modules en van S-PLUS 2000 zelf aan de orde. Hoofdstuk 7 bevat veel nuttige informatie over algemene S-PLUS instellingen, het organiseren van verschillende projecten binnen S-PLUS en het gebruik van S-PLUS resp. via de commandotaal en de grafische user interface (GUI). Tevens treft men in hoofdstuk 7 aan hoe men via de GUI zelf scripts kan maken, bewaren en uitvoeren, alsmede het gebruik van S-PLUS in de BATCH mode, waarmee veel geheugenruimte bespaard wordt. Het hoofdstuk wordt afgesloten met een aantal wetenswaardigheden en tips over het S-PLUS gebruik in de dagelijkse praktijk. In hoofdstuk 8 staat allerlei achtergrond informatie over S-PLUS bijeen gebracht waaronder boeken over nieuwe statistische methoden die men via S-PLUS kan uitvoeren. Het hoofdstuk wordt afgesloten met verwijzingen naar belangrijke WEB-sites met informatie over S-PLUS. In hoofdstuk 9 wordt het dataconversieprogramma DBMS/COPY besproken dat een goede aanvulling op S-PLUS vormt voor wat betreft data uitwisseling met andere software.

Een aantal appendices voorziet in verdere achtergrondinformatie. Zo bevat Appendix A uitvoerige beschrijvingen van de extra functionaliteiten van de modules t.o.v. S-PLUS 2000 Professional. Appendix B bevat een degelijke minicursus in het gebruik van de S-PLUS commando taal terwijl Appendix C een minicursus in het gebruik van de GUI bevat. In Appendix D staat extra informatie voor SPSS gebruikers die met S-PLUS gaan beginnen en in Appendix E zijn een aantal boekbesprekingen van Candiensten opgenomen. Tenslotte bevat Appendix F ruimte voor aantekeningen van de gebruiker zelf.

## 2. S-PLUS 2000

Deze algemene beschrijving van S-PLUS 2000 is op de eerste plaats bestemd voor lab- en RISHoofden, alsmede voor projectleiders van lab-overstijgende projecten, die mogelijk uitzien naar breed inzetbare software op statistisch en grafisch gebied voor hun lab, afdeling of project. Speciaal voor hen is vervolgens een beknopt, zeker niet volledig overzicht opgenomen met een aantal RIVM onderzoeken van de afgelopen 10 jaar waarin intensief gebruik gemaakt is of wordt van S-PLUS. Verder is dit hoofdstuk van belang voor RIVM-ers die op zoek zijn naar goede statistische software en ondersteuning bij het gebruik ervan.

### 2.1 S-PLUS 2000 eigenschappen

Het statistisch softwarepakket S-PLUS 2000 Professional<sup>1</sup> combineert de kracht en functionaliteit van het S-PLUS data analyse systeem met een flexibele grafische user interface (GUI). Hierdoor kan een brede groep mensen gebruik maken van geavanceerde statistische methoden en onthullende data-visualisatie technieken. S-PLUS bevat meer dan 2000 statistische functies, waarvan vele via de GUI beschikbaar zijn. Tevens staan er een zeer groot aantal grafische functionaliteiten ter beschikking om hoogwaardige (en indien gewenst interactief) grafieken te produceren. Daarnaast zijn er, al dan niet via modules, rechtstreekse koppelingen (Add-ins) beschikbaar voor SPSS, Excel, PowerPoint, Ingres (via ODBC) en ArcView, die in een complete, geïntegreerde werkomgeving voorzien.

S-PLUS 2000 bevat de interessante mogelijkheid om interactief via de GUI uitgevoerde analyses als S-PLUS programma's op te slaan. Zo'n programma kan dan vervolgens omgezet worden in een op maat gesneden programma dat aan derden kan worden doorgegeven, mits deze persoon ook S-PLUS bezit. Desgewenst kan men alleen gebruik maken van de commando-taal (aanvankelijk een wat steile leercurve, maar de geïnvesteerde moeite verdient zich ruimschoots terug) en daarin zelf programmeren, gebruikmakend van S-PLUS 2000 als 4<sup>e</sup> generatie taal. Gerenommeerde S-PLUS gebruikers stellen geheel kosteloos via het internet de nieuwste ontwikkelingen in de statistiek wereldwijd beschikbaar, vastgelegd in zelf geprogrammeerde S-PLUS functies en/of bibliotheken. Verder zijn er een aantal specialistische modules die in hoofdstuk 3 beknopt en in Appendix A uitvoerig beschreven staan.

In S-PLUS kunnen gegevens worden ingelezen vanuit bijna elke bron. Deze gegevens kunnen verder gemanipuleerd en bewerkt worden in een data window. Door middel van "point-and-click" en "drag-and-drop" kunnen van de gegevens gemakkelijk kwaliteitsgrafieken geproduceerd worden die geschikt zijn voor publicaties. Deze grafieken zijn ook eenvoudig in andere applicaties te gebruiken zoals bijvoorbeeld Word of Excel. Het aanroepen van de geavanceerde statistische routines gaat eenvoudig door middel van menu's en toolbars.

Tevens kunnen allerlei resultaten worden opgeslagen, zoals grafieken en rapporten met analyse gegevens in wat genoemd wordt de "workspace". Dit houdt in dat alle openstaande windows binnen S-PLUS in één keer kunnen worden opgeslagen met de aanwezige inhoud. Deze workspace kan in een volgende S-PLUS sessie weer worden geopend en de gebruiker kan dan gewoon verder gaan daar waar hij of zij gebleven was.

Samenvattend zijn een aantal kenmerken van S-PLUS 2000 Professional

1. Import van allerlei databestanden o.a. SPSS, SAS, Excel en via ODBC.

---

<sup>1</sup> S-PLUS 2000 Professional bevat de mogelijkheid om zelf in S-PLUS te programmeren en de nieuwe functies aan het S-PLUS systeem toe te voegen, dit in tegenstelling tot S-PLUS 2000. In het vervolg wordt korthedshalve als werknaam S-PLUS 2000 gebruikt, ook daar waar S-PLUS 2000 Professional bedoeld wordt.

2. Exploratieve data analyse en ruime grafische mogelijkheden.
3. Algemene statistische methodes.
4. Statistische Modelbouw.
5. 4<sup>e</sup> generatie programmeertaal.
6. Bouw gebruikersvriendelijke interactieve tools.
7. Bouw van op maat gesneden modellen.
8. Bouw van eigen functionaliteit voor standaard verwerking van gegevens.
9. Add-ins voor SPSS<sup>2</sup>, Excel, PowerPoint en ArcView.

S-PLUS 2000 Release 3 wordt via de NAL beschikbaar gesteld aan iedereen die behoort tot de RIVM gebruikersgroep van S-PLUS. Via een e-mail<sup>3</sup> aan [Arnold.Dekkers@rivm.nl](mailto:Arnold.Dekkers@rivm.nl), de beheerder van S-PLUS, kan men hier lid van worden. Via deze gebruikersgroep wordt men regelmatig op de hoogte gehouden van o.a. het S-PLUS cursus aanbod en andere belangrijke wetenswaardigheden.

## **2.2 S-PLUS toepassingen op het RIVM**

S-PLUS is sinds 1992 op het RIVM in gebruik. Het eerste jaar is het op proef door het toenmalige CWM (Centrum Wiskundige Methodes) gebruikt en voor het eerst ingezet om 640 tijdreeksen met sterftcijfers bijna volledig automatisch te analyseren t.b.v. de eerste Volksgezondheid Toekomst Verkenning. Na evaluatie van de bruikbaarheid is de statistische afdeling van het CWM overgegaan op S-PLUS en is het o.a. gebruikt bij de AIDS modellering via Bayesiaanse statistische methodes. Dit heeft geleid tot de dissertatie van S.H. Heisterkamp, *Quantitative analysis of AIDS/HIV: development of methods to support policy making for infectious disease control*, 1995.

S-PLUS is sindsdien intensief gebruikt bij:

### **Volksgezondheidsanalyses**

- De Zutphen studie. Deze studie omvat gegevens van het Zutphen-cohort bestaande uit 878 mannen, aanvankelijk met leeftijd 40-59 jaar, die gevolgd zijn vanaf 1960. Hiervoor is o.a. het model van Manton&Stallard in S-PLUS geprogrammeerd. Met dit model zijn positieve veranderingen met de leeftijd gevonden m.b.t. de verdeling van de risicofactorniveaus in tegenstelling tot de uitkomsten van eenvoudige regressie analyses.
- Schiphol studie. In dit onderzoek zijn ruimtelijke patronen van ziekenhuisontslagen voor o.a. een aantal hartvaatziekten in de regio Schiphol bestudeerd aan de hand van kaarten. Het onderzoek is een uitbreiding van eerdere analyses die zijn uitgevoerd in het kader van de integrale Milieu Effectrapportage Schiphol en Omgeving (iMER) en is uitgevoerd in het kader van het Evaluatie en Monitoringsprogramma Schiphol (EMSO).
- De Tweede Nationale Studie. Met deze studie worden relaties onderzocht tussen gezondheidsklachten en het zorggebruik via huisartsen. Het aantal huisartsen en de duur van hun deelname aan het onderzoek bepalen de betrouwbaarheid van de schattingen. Door een goed design van de steekproefopzet zijn er aanzienlijke kosten bespaard.
- Het doorrekenen in combinatie met het gebruik van Mathematica van verschillende mogelijke interventies (al of niet medicatie bijv.) in het kader van de recent verschenen cholesterol en bloeddruk consensus. S-PLUS is hierbij gebruikt voor het fitten van diverse regressiemodellen.

### **Risico-analyses**

- De implementatie van verschillende klassieke overlevingsmodellen en van het "two-stage clonal expansion" carcino-genese-model, alsmede bij het fitten van deze modellen op gegevens van stralingsproeven op muizen.

---

<sup>2</sup> Appendix D bevat extra informatie voor SPSS gebruikers die met S-PLUS gaan beginnen.

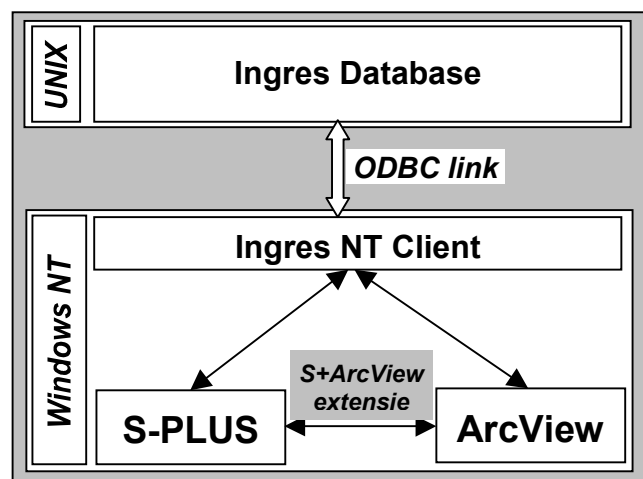
<sup>3</sup> In principe kan men op dinsdag-woensdag-donderdag worden toegevoegd.

- De ontwikkeling en toepassing van PROAST (possible risk obtained from animal studies) voor de modellering van doses-effect relaties op basis van waarnemingen in dierstudies, en de vertaling (extrapolatie) van deze relaties naar humane effecten. Met behulp van dit RIVM S-PLUS programma kan het effect bij een bepaalde blootstelling worden geschat. De effectschatting in het proefdier wordt met behulp van een waarschijnlijkheidsbenadering voor onzekerheden rond inter- en intraspeciesvariatie geëxtrapoleerd naar een effectschatting in de mens. Deze methodiek kan een onderdeel vormen van de "effectschattingsmodule" in de leidraad "Handelen bij normoverschrijding".
- De ontwikkeling en toepassing van STEM (Statistical Exposure Model) voor het schatten van de inname van stoffen (contaminanten, natuurlijke toxinen, additieven) via de voeding. Het afgelopen jaar is STEM o.a. gebruikt voor blootstellingbeoordelingen van o.a. dioxinen, PCBs en nitraat.

### Milieu-analyses

- De toepassing van ruimtelijke statistiek, met name Universal Kriging, voor de optimalisatie van de meetnetten LML (Landelijk Meetnet Luchtkwaliteit) en LMR (Landelijk Meetnet Regenwater). Deze methode levert naast de geïnterpoleerde meetwaarden ook de bijbehorende onbetrouwbaarheden.
- De ontwikkeling en toepassing van MOVE (ecologische modellering) voor het schatten van kansen op voorkomen van vegetatie op basis van abiotische factoren. Dit heeft geleid tot het proefschrift van A. Bio, *Does vegetation models suit our models? Data and model assumptions and the assessment of species distribution in space*, 2000.
- De (ruimtelijke) modellering via Gegeneraliseerde Additieve Modellen van zware metalen en pesticiden in de bodem. Zie proefschrift A. Tiktak, *Modeling Non-Point Source Pollutant in Soils. Applications to the Leaching and Accumulation of Pesticides and Cadmium*, 1999.
- De modellering van de daling van eenheidskosten van nieuwe technologieën in de tijd, vastgelegd in het rapport *TECHNO 2000*. De resultaten van deze modellering geven een indruk van de toekomstige milieukosten. Voorbeelden betreffen o.a. investeringen in Hoogrendementsketels, windturbines en LOWNO<sub>x</sub>-branders.
- De aansturing van inrichting en beheer van beken. Aan de hand van modelresultaten worden beleidsbeslissingen genomen over de inrichting (zoals uitgraven van een beek) en beheer (schoonhouden oevers). De modellering betreft het voorspellen van het voorkomen van soorten vissen en/of waterplanten (responsies) aan de hand van de gemeten milieufactoren zoals nutriëntenconcentraties en meandering van een beek (doses). Hierbij wordt gebruik gemaakt van stapsgewijze logistische regressie.
- De correctie van uurgemiddelde ozon concentraties op basis van uurgemiddelde meteorologische gegevens met het specifiek hiervoor ontwikkelde programma RBA-MC. Hiermee kunnen trends in jaargemiddelde concentraties ontdaan worden van jaar-specifieke meteorologische invloeden om zo beleidsinvloeden beter zichtbaar te maken.

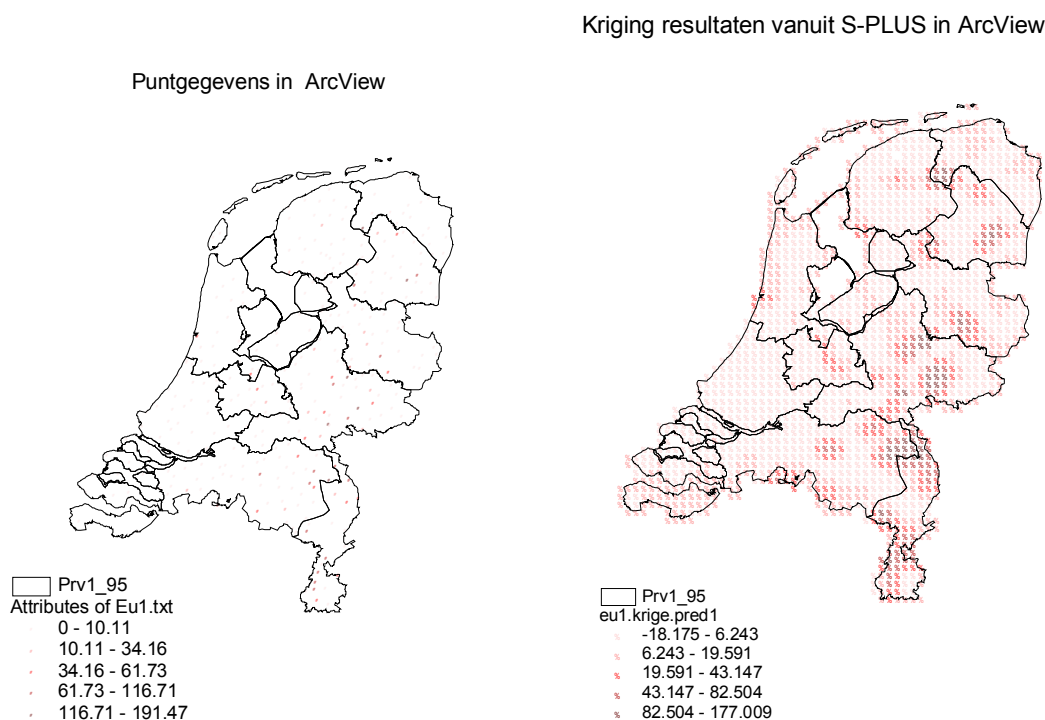
Onlangs is een gezamenlijke CIM / LBG pilotstudy gestart om tot een integraal instrumentarium voor GIS en statistiek te komen. Met de module S+ArcViewGis kan men data vanuit Arcview binnen S-PLUS halen en daar (geo)statistisch analyseren. Vervolgens kunnen de resultaten intern weer naar ArcView overgebracht worden om daar geografisch weer te geven. Hiermee wordt een optimale combinatie van geavanceerde geografische weergave met



ArcView en van moderne statistiek in S-PLUS bereikt. Via een ODBC link is het daarbij mogelijk om gegevens rechtstreeks uit de LMG Ingres database op een UNIX platform te halen. Zie schema op de vorige pagina.

Een eenvoudig voorbeeld, dat voor het begrip van deze aanpak inhoudelijk niet helemaal reëel is uitgewerkt maar dat wel de kracht van het instrumentarium weergeeft, is het maken van landsdekkende kaarten van nitraat waarnemingen in het diepe grondwater. In figuur 1 staan ze uitgezet m.b.v. ArcView. De database staat in ArcView waarbij elk record een unieke sleutel bevat. De *S+ArcViewGis* extensie geeft twee extra menu's in ArcView. Een is de link naar S-PLUS waarmee data en figuren intern kunnen worden uitgewisseld of waarmee een lineaire regressie kan worden uitgevoerd. Het andere menu biedt de mogelijkheid om rechtstreeks met S-PLUS geostatistische analyses uit te voeren op ArcView gegevens.

De onderliggende gegevenstabel met nitraat concentraties in ArcView (uitgezet in figuur 1) is intern overgebracht naar S-PLUS. Binnen S-PLUS heeft de geostatistische analyse plaatsgevonden waarbij, op basis van de oorspronkelijke data en het semivariogram model, de lineaire interpolatie via ordinary kriging is uitgevoerd op een 5x5 km grid. Dit levert een nieuwe tabel op met x- en y-coördinaten en de geïnterpoleerde nitraat concentraties. Deze tabel is vervolgens intern naar ArcView gezet waar de resultaten in een kaart geografisch zijn uitgezet met een legenda die in figuur 2 te zien is.



*Figuur 1 : De ruim 350 meetgegevens (1995) van nitraat uitgezet in Nederland.*

*Figuur 2 : De geïnterpoleerde kaart van nitraat in Nederland 1995 op basis van een sferisch model met range = 15.375 km, sill = 925 en de nugget gelijk aan nul.*

### 3. MODULES VAN S-PLUS

Nu volgen korte beschrijvingen van de S-PLUS modules. Uitvoerige beschrijvingen van de modules staan in Appendix A. De volgende modules van S-PLUS zijn beschikbaar:

#### 3.1 S+Dox

Module voor het ontwerpen en analyseren van industriële experimenten, bevat o.a.: fractional factorial, mixed factorial, irregular fractions, Plackett-Burman, central composite, face centered cube, Tachuchi orthogonal arrays, Box-Behken, robust designs.

#### 3.2 EnvironmentalStats

Module voor milieu statistiek en probabilistische risk analysis. EnvironmentalStats is een zeer goede aanvulling op de statistische functionaliteit van S-PLUS, en is toegespitst op milieu-statistiek met o.a. monitoringsstatistiek en risk assessment. In het handboek *Environmental Statistics with S-PLUS (2001, S.P. Millard and N.K. Neerchal, CRC)* staat de statistische achtergrond zeer goed uitgelegd. Het boek bevat maar liefst 25 pagina's aan referenties.

#### 3.3 S+Garch

Deze module is voor geavanceerde tijdreeks-analyse via GARCH modellen en is geschikt voor univariate en multivariate modellen, voor het modelleren en voorspellen van tijdreeksen, bevat standaard deviaties en correlaties. Tevens bevat het modellen met heteroscedastische varianties in de tijd, d.w.z modellen waarbij de variantie over de tijd varieert. S+Garch is o.a. toepasbaar in financiële modellering, kwaliteitscontrole en milieu-monitoring.

#### 3.4 S+NuoOpt

Module voor het oplossen van (zeer grote) optimalisatie problemen. Deze module bevat functies voor: lineaire programmeren, gemengd integer programmeren, kwadratische programmeren, transport problemen, unconstrained en constrained niet-lineair programmeren en multi-objective programmeren.

#### 3.5 S+SeqTrial

SeqTrial is een module voor monitoring, ontwerp en analyse van clinical trials en bevat geavanceerde routines voor berekening van o.a. power curven, ASN curven en stop kansen en voor groep sequentiële methoden zoals Lan en DeMets, Wang en Tsiatis, en Whitehead methode.

#### 3.6 S+SpatialStats

Module voor statistische analyse van ruimtelijke data: SpatialStats bevat routines voor exploratieve data analyse, modellering van puntpatronen, geostatistische data en gegevens op een rooster, geschikt voor problemen in milieu-monitoring, mijnbouw en oliewinning, epidemiologie, demografie etc.

### **3.7 S+Wavelets**

Module voor visualisatie en signaal analyse, bevat routines voor data analyse en visualisatie, niet-parametrische schattingen, signaal en beeld compressie via een volledige rekenomgeving voor wavelets analyse.

### **3.8 S+ArcViewGis**

S+ArcViewGis combineert de statistische technieken en grafische mogelijkheden van S-PLUS met de GIS technieken van ArcView. Deze module kan men alleen installeren indien men een ArcView licentie heeft (zie 4.4).



## 4. S-PLUS 2000 OP HET RIVM

S-PLUS 2000 wordt op het RIVM beschikbaar gesteld via verschillende NAL's. NAL staat voor "Network Application Launcher" en geeft de mogelijkheid om installatie of verwijdering van software centraal te verspreiden op een uniforme wijze. S-PLUS wordt via twee NAL's beschikbaar gesteld:

- een NAL voor de installatie van S-PLUS 2000 Professional, Release 3
- een NAL voor de installatie van de modules (optioneel).

Indien men S-PLUS 2000 voor het eerst installeert kan men 4.1 overslaan. Ieder die S-PLUS 2000 al gebruikt, dient voor de zekerheid eerst 4.1 te doorlopen (een back-up van de eigen gegevens maken).

Het is raadzaam eerst dit hele hoofdstuk door te lezen voordat met de installatieprocedure begonnen wordt! De commando's die men dient op te geven worden in ASCII weergegeven.

Met de installatie NAL wordt het S-PLUS programma in `C:\Program Files\sp2000` geplaatst en de S-PLUS projectdirectory `S_PROJ` wordt `C:\splusdata`. Hierin worden alle eigen S-PLUS gegevens opgeslagen. Deze laatste instelling kan men wijzigen, zie hiervoor 4.2 onder nummer 6.

### 4.1 Back-up van de eigen S-PLUS gegevens en verwijdering oude versie S-PLUS

Maak een volledige back-up van de `S_PROJ` directory. Dit is de directory `C:\Splusdata` of `N:\Splusdata`. Het is voldoende om deze map inclusief alle sub-mappen naar een andere map te kopiëren, bijv. `d:\temp\splusdatabackup`. Op deze wijze kan er niets verloren gaan van de eigen gegevens, mocht er onverhoopt iets mis gaan met het verwijderen en/of installeren van S-PLUS.

### 4.2 Installatie S-PLUS 2000 release 3

Na toevoeging aan de S-PLUS gebruikersgroep door de S-PLUS beheerder is het installatieprogramma S-PLUS 2000 R3 zichtbaar in `Start / Sector5 Applicaties / Overige Software` mits men minimaal zo'n 120 Mb vrij heeft op de C-schijf<sup>4</sup>.

#### Probleem: De NAL S-PLUS 2000 (R3) is niet zichtbaar.

Klik met de rechter muisknop op het icoontje  rechtsbeneden in de werkbalk en selecteer *Refresh*. Er gaat nu een soort klokje draaien. Als de icon  weer verschijnt en er voldoende schijfruimte vrij is op de C-schijf, verschijnt er een nieuwe NAL met de naam "S-PLUS 2000 R3" in `Start / Sector5 Applicaties / Overige Software`. Hiermee wordt S-PLUS geïnstalleerd.

Nu volgt puntsgewijs wat er allemaal gebeurt, zodat waar nodig achteraf veranderingen kunnen worden aangebracht.

1. Start de NAL door een dubbelklik: S-PLUS 2000 Release 3 wordt nu geïnstalleerd in de directory: `c:\Program Files\sp2000`  
met als standaard project directory `c:\splusdata`

---

<sup>4</sup> Indien men geen 130 Mb op de C-schijf beschikbaar heeft, is de NAL niet zichtbaar. Dit ter beveiliging van de schijf! Vanwege de NAL is het alleen mogelijk om het S-PLUS programma zelf op de C-schijf te zetten. Wel kan men de gegevensdirectory SPLUSDATA op de D- of N-schijf zetten.

2. Na het kopiëren van alle bestanden wordt gevraagd of de computer opnieuw opgestart kan worden. Dit is noodzakelijk om S-PLUS vervolgens te kunnen gebruiken.
3. Onder *Start/Programs* is een Personal Map aangemaakt, genaamd S-PLUS 2000 (R3), waarin een snelkoppeling naar S-PLUS staat, genaamd S-PLUS 2000.
4. Klik deze snelkoppeling met de rechtermuisknop aan en kies *Properties*.
5. Ga naar 2e tabblad: *Shortcut of Snelkoppeling*.
6. Onder *Target* staat het IP adres van de Windows NT server (*Milntg28*) al aangegeven, *S\_ELMHOST=131.224.235.177*. Hierop draait de license manager van S-PLUS en de modules.

#### **Veranderen van S\_PROJ map**

Achter *S\_ELMHOST* staat de S-PLUS project map *S\_PROJ=c:\splusdata* opgegeven en in het vakje *Start* staat nog eens de opstart directory opgegeven: *c:\splusdata*.

*Indien men op de N-schijf of D-schijf werkt qua data*, kan men hier de S-PLUS omgevingsvariabele *S\_PROJ* wijzigen, bijv op *n:\splusdata*. Men moet dan op beide plaatsen *c:\splusdata* vervangen door *n:\splusdata*.

#### **Gewijzigde snelkoppeling en werking NAL**

Als *S\_PROJ* gewijzigd wordt, is S-PLUS *niet* meer via de NAL op te starten, omdat deze NAL naar *c:\splusdata* verwijst. Men dient dan steeds de veranderde snelkoppeling te gebruiken die onder *Start/Programs/S-PLUS 2000 (R3)* staat.

#### **Instellen S-PLUS sneltoets**

Men kan het *opstarten* aanmerkelijk *versnellen* door in deze snelkoppeling onder het tweede tab-blad als code in de sneltoets op te geven *Ctrl-Alt-S*. Men kan dan in het vervolg S-PLUS opstarten met deze sneltoets combinatie zonder de muis te hoeven gebruiken.

7. Als er geen complicaties zijn opgetreden, is S-PLUS met de eigen gegevens weer volledig beschikbaar. Als dit zo is, kan men, indien gewenst, de back-up weer verwijderen.  
Indien de eigen gegevens niet meer onder de vroegere *S\_PROJ* staan, dan kan men alles terugzetten vanuit de map met de back-up *d:\temp\splusdatabackup* (zie 4.1)

### **4.3 Installatie van de modules**

Nadat S-PLUS 2000 geïnstalleerd is, kan men alle modules (exclusief S+ArcViewGis) in één keer met de NAL "S-PLUS modules" installeren. De niet gewenste modules kan men vervolgens eenvoudig verwijderen (zie 6.1).

Op het RIVM is de S-PLUS 2000 Professional netwerk versie beschikbaar. Helaas zijn twee alle modules nog niet als netwerk module beschikbaar. Dit compliceert wel de installatie maar niet het gebruik. De modules waarvan geen netwerkversie beschikbaar is, kunnen als library binnen S-PLUS aangekoppeld worden met behoud van hun functionaliteit.

In onderstaand schema staan enkele technische gegevens over de modules weergegeven. Slechts van enkele modules is een online handleiding beschikbaar, in PDF formaat, die te lezen is met de Acrobat Reader.

Tabel 1: De beschikbare S-PLUS modules met bijbehorende technische gegevens.


Module	Versie	GUI	Online Handleiding	Beschikbaar op RIVM als	Opgenomen in registry
S+DOX	1.1.2	nee	nee	module	ja
EnvStats	1.1.0	nee	nee, goede help	library	
S+Garch	1.3.1	nee	ja	module	ja
S+Nuoapt	1.0.1	nee	nee	library	ja
S+SeqTrial	1.0.1	ja	ja	module	ja
S+SpatialStats	1.5.1	ja	supplement	module	ja
S+Wavelets	1.1.2	nee	nee	module	ja
S+ArcviewGis <sup>5</sup>	1.1	ja	nee	module	ja

"GUI = nee" betekent dat de module alleen vanuit het commando window te gebruiken is. "GUI = ja" betekent dat bij het laden van de module een nieuw menu aan de menubalk wordt toegevoegd. Online handleidingen zijn er voor S-PLUS 2000 en de aangegeven modules. Indien een module geen netwerk-versie heeft, is deze module als library aan te koppelen binnen S-PLUS.

Na installatie van S-PLUS 2000 (R3) is nu de installatie NAL, SPLUS 2000 Modules (R3), zichtbaar in Start/Sector5 Applicaties/Overige/Software. Hiermee worden in één keer alle S-PLUS modules (exclusief S+ArcViewGis) geïnstalleerd.

#### 4.3.1 **Probleem: er gebeurt niets na het aanklikken van de NAL S-PLUS Modules**

Als deze NAL al een keer gebruikt is, kan het voorkomen dat het aanklikken ervan geen enkele reactie tot gevolg heeft, omdat de NAL "denkt" dat alles al geïnstalleerd is, terwijl de modules misschien handmatig verwijderd zijn. Dit is als volgt op te lossen:

1. Klik *rechts* op de startknop  en selecteer Explore.
2. Klik vervolgens in het linker scherm van Explore op Sector5 applicaties en vervolgens op Overige software.
3. Klik tenslotte in rechterscherm met de *rechtermuisknop* op SPLUS Modules en klik de optie Verify aan. Nu worden de modules wel geïnstalleerd!

#### 4.3.2 **Opstarten modules / libraries binnen S-PLUS**

Nadat men S-PLUS heeft opgestart kan men handmatig de gewenste module of library aankoppelen. Dit kan op twee manieren:

1. Via de GUI: ga naar File/Load Module en selecteer de gewenste module. Voor een library gaat het analoog.
2. Via het commando window: type `module("Spatial")` bijv. of `library("Envstats")`.

Men kan ook automatisch een veel gebruikte module of library laden bij het opstarten. Men dient dan de gewenste commando's op te nemen in de functie `.First` (zie 7.1).

Als men al een `.First` heeft met daarin de regel `library("SpatialStats")`, verander deze dan **voor** de installatie van de S-PLUS modules in: `module("spatial")`. Dit voorkomt foutmeldingen bij het opstarten van de nieuwe S-PLUS versie!

<sup>5</sup> De module S+ArcViewGis dient men op het RIVM handmatig te installeren, aangezien ArcView op de betreffende PC al geïnstalleerd dient te zijn. Wijze van installeren: zie 4.4.

### 4.3.3 Beschikbaarheid helpbestanden binnen S-PLUS

Na installatie, het opnieuw opstarten van S-PLUS en het aankoppelen van de gewenste module of library zijn de betreffende helpbestanden en de vermelde manuals online beschikbaar via het Help menu binnen S-PLUS.

Voor het verwijderen van modules of van S-PLUS 2000: zie hoofdstuk 6.

### 4.4 Installatie van de S+ArcViewGis

De installatie van de module S+ArcViewGis verloopt op een geheel eigen wijze vanwege dat men ook een Arcview licentie dient te hebben. Indien men van deze modules gebruik wil maken, dient men een aanvraag per e-mail te richten aan [Arnold.Dekkers@rivm.nl](mailto:Arnold.Dekkers@rivm.nl).

### 4.5 De RIVM S-PLUS bibliotheek RIVMfun

Via de O-schijf is er een eigen RIVM bibliotheek beschikbaar met een aantal handige functies. Deze functies zijn ontwikkeld om het gebruik van S-PLUS te vereenvoudigen en enigszins te standaardiseren uitgaande van de bestandsorganisatie beschreven in 7.2 en van het gebruik van S-PLUS via de commandotaal (zie Appendix B). De bibliotheek kan worden aangekoppeld op het RIVM via de S-PLUS opdracht

```
library("RIVMfun", first=T, lib.loc="O:/milieu/applics/splus 2000/RIVM")
```

Door de bibliotheek via de O-schijf aan te koppelen, heeft men altijd de meest bijgewerkte versie. Indien het netwerk langzaam is kan men de bibliotheek op de eigen C-schijf zetten. Hiertoe dient men de map RIVMfun te kopiëren naar `c:\Program Files\sp2000\library`. Het aankoppelen gebeurt dan met de opdracht `library("RIVMfun", first=T)`.

Via `help(rivmfun)` start men de helpfunctie op en kan men de verschillende gegevens en functies in de bibliotheek bekijken.

De bibliotheek bevat op dit moment 4 onderwerpen zoals uit onderstaande figuur blijkt:



De handige functies worden in 7.2 en 7.6 verder besproken. De *Spatial Plot Functions* zijn `NL` en `f.image.NL`. Beide functies zijn alleen vanuit het commando scherm te gebruiken. Met `NL` kan men de kaart van Nederland tekenen met optioneel de provinciegrenzen erbij. Via de gebruikelijke `help(f.image.NL)` kan de online informatie opvragen. De *Spatial Modelling*

*Functions* heeft als basis de functie `f.variogram.fit`, waarmee men een theoretisch variogram kan fitten aan een van te voren berekend empirisch semivariogram. Het laatste item *Tips and Background information* bevat een drietal handige wetenswaardigheden.

#### **4.6 RIVM ondersteuning bij het gebruik van S-PLUS**

De ondersteuning bij het gebruik van S-PLUS op het RIVM wordt momenteel grotendeels door de S-PLUS beheerder (de auteur) gegeven en omvat o.a.

- concrete hulp bij het gebruik van S-PLUS
- advisering en concrete inhoudelijke hulp qua statistiek
- advisering en eventueel ontwikkelen en schrijven van specifieke S-PLUS programma's voor een optimaal gebruikersgemak binnen een project. Een voorbeeld hiervan is het meteorocorrectie programma RBA-MC (zie 2.2).
- Ruim aanbod van cursussen bij Candiensten of intern op het RIVM.

Voor concrete hulpvragen is er in principe een S-PLUS mailing groep op het RIVM waar men vragen aan zou kunnen stellen. Doordat men meestal na enige tijd zelfstandig met S-PLUS gaat werken of het gemakkelijker is om rechtsreeks RIVM-ers om hulp te vragen, wordt daar geen gebruik van gemaakt.

Verder worden alle S-PLUS gebruikers regelmatig per e-mail op de hoogte gehouden van het cursusaanbod van de leverancier, Candiensten uit Amsterdam (zie ook 8.5).

Daarnaast zijn er de laatste 10 jaar een aantal S-PLUS cursussen op het RIVM gegeven door zowel Candiensten als door de auteur (Inleidende cursus commandotaal en Ruimtelijke Statistiek). Bij voldoende belangstelling kunnen deze opnieuw gegeven worden, maar ook cursussen over specifieke onderwerpen, zoals bijvoorbeeld lineaire regressie, kunnen worden georganiseerd.

#### **4.7 Beginnen met S-PLUS voor SPSS gebruikers**

RIVM-ers die gewend zijn om SPSS te gebruiken, vinden de overgang van SPSS, waarvan ze veel zaken uit hun hoofd weten, naar S-PLUS soms een grote stap. Op korte termijn lijkt dan de aanschaf van SPSS voor een individuele gebruiker een snelle, pragmatische oplossing, maar op de lange duur komt deze aanpak de communicatie binnen het RIVM niet ten goede. Om de overgang van SPSS naar S-PLUS te vergemakkelijken is in Appendix D een document van Hans de Wit (voorheen Sociaal Cultureel Planbureau, nu Stichting Farmaceutische Kengetallen) opgenomen getiteld "Beginnen met S-PLUS voor SPSS gebruikers".



## 5. S-PLUS 2000 THUIS

RIVM-ers die S-PLUS op het RIVM gebruiken, mogen S-PLUS 2000 en de bijbehorende modules ook thuis gebruiken. Duidelijk is dat het pakket niet doorgegeven mag worden aan derden.

### 5.1 Regeling thuisgebruik

De distributie van de stand-alone versie van S-PLUS voor thuisgebruik verloopt via de bibliotheek. Daar zijn twee installatiepakketten bestaande uit een CD en dit rapport beschikbaar. Leden van de S-PLUS gebruikersgroep kunnen daar een pakket lenen om S-PLUS thuis te installeren. De bibliotheek houdt de registratie bij.

### 5.2 Installatie stand-alone versie onder Windows 9x

De installatie CD bevat een autorun programma dat automatisch het installatiemenu met zes buttons opstart. Mocht dit niet gebeuren dan kan men via de Explorer naar de CD gaan en dubbelklikken op `Setup.exe`. Het is raadzaam eerst dit hele hoofdstuk door te lezen, voordat met de installatie wordt begonnen.

#### 5.2.1 Openen bestand met installatiecodes

Met de eerste knop wordt (alleen onder Windows 9x) automatisch via het Windows programma `notepad` een informatiebestand opgestart waarin de installatiecodes voor S-PLUS 2000 Release 3 en de modules staan. Men kan de benodigde code uit dit bestand met Ctrl-c kopiëren, starten met de installatie van S-PLUS of een module en de code met Ctrl-v in het registratieveld plakken.

Mocht het opstarten van Notepad niet lukken (onder bijv. Windows NT) dan kan men het bestand `licentie.txt` via knop 6 (Opstarten Explorer) handmatig aanklikken en openen.

#### 5.2.2 Installatie S-PLUS

Met de tweede knop wordt S-PLUS 2000, Release 3 geïnstalleerd. Er volgt nu een serie installatiestappen waarin men zelf S-PLUS kan configureren. Dit staat uitvoerig in het nu volgende beschreven. Het is aan te bevelen deze beschrijving in zijn geheel te lezen *voordat* met de installatie begonnen wordt, zodat men niet voor verrassingen komt te staan.

1. Het openingsscherm raadt aan alle andere programma's te sluiten en bevat de gebruikelijke copyright verklaringen. [Klik next].
2. Het tweede scherm bevat de naam [rivm-er], compagny [rivm] en vraagt om de installatiecode in te typen. Als deze al gekopieerd is (zie eerste knop), kan men deze nu met Ctrl-v kopiëren. [Klik next].
3. Dan wordt gevraagd of de opgegeven informatie correct is. [Klik next].
4. In de volgende stap worden er twee dingen gevraagd: het type installatie én de plaats waar S-PLUS geïnstalleerd dient te worden.

De plaats is standaard `C:\Program Files\sp2000`. Dit kan, indien gewenst, veranderd worden.

Indien men zelf volledig wil bepalen hoe S-PLUS geïnstalleerd wordt, dient men `custom` te kiezen. Indien men in het hierop volgende menu niets verandert, krijgt men de installatie volgens de keuze `Typical installation`.

In de nu volgende tabel staan kort alle opties beschreven en aanbevelingen voor de instellingen.

<b>Optie</b>	<b>Uitleg optie</b>	<b>Aanbeveling</b>
Readme Files	Bevat laatste informatie over S-PLUS	handig / optioneel
Program Files	Programma en ondersteunende bestanden voor S-PLUS	<b>noodzakelijk</b>
User-Interface and Preference files	Standaard menu/voorkeurs- en andere bestanden	handig / alleen niet-nodig als er voorkeursbestanden zijn van een eerdere installatie
Adobe Acrobat Reader	Installatie Acrobat Reader	<b>noodzakelijk</b> voor online manuals, kan ook met de 4e knop worden geïnstalleerd
Microsoft ODBC files	Indien ODBC import/export gewenst is	optioneel
Windows Files for program	Gemeenschappelijke Windows Files	<b>noodzakelijk</b>
PowerPointPresentation Wizard Files	Bestanden voor het maken van PowerPoint presentaties	optioneel
Visual Demo files	Online demo: visueel en auditief	optioneel
Development System files	Bestanden nodig om functies in andere talen geschreven, statisch aan S-PLUS te kunnen koppelen.	i.h.a. niet nodig; noodzakelijk om zelf gemaakte Fortran programma's te laden.
Excel Add-In files	S-PLUS is dan rechtsreeks vanuit Excel aan te roepen om bijv. grafieken te maken	optioneel, niet nodig als Excel niet beschikbaar is
SPSS Add-In files	idem voor SPSS	optioneel, niet nodig als SPSS niet beschikbaar is
Object System HTML help files	??	??
Setup Disk Image	Maakt een Disk image om S-PLUS vanaf een ander systeem te kunnen installeren	niet nodig

5. Nu wordt gevraagd waar de S-PLUS working directory dient te komen. Standaard is dit `c:\Program Files\sp2000\users\usernaam`. Indien men over meerdere partities op de harde schijf beschikt of meerdere harde schijven heeft, dan is het aan te raden om de S-PLUS working directory op de gegevenspartitie te zetten. Bijv. op `d:\splusdata`. Zie voor dit onderdeel ook punt 6 onder 4.2.
6. Vervolgens wordt gevraagd of in `Start/Program` de map S-PLUS 2000 kan worden aangemaakt [klik Next].
7. En of dit een Common group (voor alle gebruikers zichtbaar) of een Personal Group dient te zijn [maak keuze en klik Next].
8. Dan komt de vraag welke schermen er standaard opgestart dienen te worden: het commando scherm en/of de object explorer [maak keuze en klik Next].
9. Tenslotte komt een volledig overzicht van alle gemaakte instellingen op het scherm. Klik Next als alles naar wens is en, de installatie van S-PLUS 2000 wordt gestart.



### **5.2.3 *Installatie modules***

Met de derde knop wordt een nieuw menu opgestart voor de installatie van de modules. Elke module dient apart geïnstalleerd te worden, maar men kan wel via het indrukken van de Shift-toets in één keer alle gewenste modules selecteren.

Het is tijdens de installatie van de modules handig om het bestand `licentie.txt` open te hebben staan in Notepad. Men kan dan rechtstreeks de installatiecode van ieder module van Notepad in het installatieprogramma kopiëren via resp. Ctrl-c (copy) en Ctrl-v (paste).

### **5.2.4 *Afzonderlijke installatie Acrobat Reader***

Met de vierde knop kan men Acrobat Reader 4 apart installeren. Dit programma is nodig om de gebruikershandleidingen op het scherm te kunnen lezen. Deze handleidingen zijn te vinden onder het S-PLUS menu (dus na het opstarten van S-PLUS) Help. De beschikbare PDF handleidingen voor de modules (zie tabel 1) zijn pas zichtbaar na het aankoppelen van de betreffende module.

### **5.2.5 *Demo S-PLUS***

Met de vijfde knop kan men een demo over S-PLUS opstarten. Installatie van S-PLUS is voor het draaien van de demo niet nodig. De demo is te verlaten via de Esc-toets.

### **5.2.6 *Inhoud installatie CD benaderen via Explorer***

Met de zesde knop wordt de Explorer opgestart waarmee men de bestanden op de installatie CD kan bekijken. Deze knop dient men ook te gebruiken om de RIVM S-PLUS bibliotheek RIVMfun te installeren. Daarvoor dient men de map RIVMfun eigenhandig te kopiëren naar `c:\Program Files\sp2000\library`. Het aankoppelen van de bibliotheek gebeurt via de S-PLUS opdracht `library("RIVMfun", first=T)` (zie 4.5).



## 6. VERWIJDEREN S-PLUS 2000

Het verwijderen van S-PLUS en de geïnstalleerde modules op PC's dient in een bepaalde volgorde te gebeuren met het verwijderingsprogramma dat beschikbaar is via het besturingssysteem. Dit komt omdat de meeste modules en natuurlijk S-PLUS 2000 zelf in de registry genoteerd staan. Om deze registry niet te vervuilen én om *alle* relevante bestanden te verwijderen dient men *eerst de modules* te verwijderen en dan pas S-PLUS zelf.

### 6.1 Verwijdering van modules

Ga via Start/Settings/Control Panel naar Add/Remove Programs en start dit onderdeel op. In het venster zijn nu alle programma's te zien die in de registry staan. Het betreft voor S-PLUS (indien alle modules geïnstalleerd zijn):

S+Nuoopt Module	
S-PLUS 2000 ## in "C:\Program Files\sp2000	<b>[hoofdprogramma S-PLUS]</b>
S-PLUS DOX Module ## in "C:\Program Files\sp2000	
S-PLUS Garch Module ## in "C:\Program Files\sp2000	
S-PLUS SeqTrial Module ## in "C:\Program Files\sp2000	
S-PLUS SpatialStats 1.5 Module ## in "C:\Program Files\sp2000	
S-PLUS Wavelets Module ## in "C:\Program Files\sp2000	

Hierbij bestaat ## uit # met een nummer erachter. Uit dit schema volgt dat S-PLUS in "C:\Program Files\sp2000" geïnstalleerd staat. De modules staan dan vervolgens in de directory C:\Program Files\sp2000\modules\modulenaam.

Voor het **thuisgebruik** is het voldoende om een module te selecteren (bijv. S+Nuoopt Module) en op de Add/Remove knop te drukken. Indien er een melding komt dat niet alles verwijderd is, kan men de directory C:\Program Files\sp2000\modules\modulenaam vervolgens handmatig verwijderen.

Op het **RIVM** dient men naast de aanwijzingen voor het thuisgebruik tevens handmatig de directory C:\Program Files\sp2000\library\modulenaam te verwijderen. Hiermee zijn de bestanden van de betreffende module verwijderd. Deze extra handeling is nodig omdat op het RIVM niet-netwerk modules als library binnen de netwerkversie van S-PLUS 2000 zijn geïnstalleerd.

### 6.2 Verwijdering van S-PLUS 2000

Indien *alle* modules verwijderd zijn, kan men S-PLUS 2000 verwijderen. Hierbij dient men goed op de vraag te letten of men de eigen gegevens wil verwijderen of niet. Als men de gegevens wil bewaren, is het veilig om eerst een back-up te maken van de betreffende gegevens in de map S\_PROJ met alle sub-mappen (bijv. c:\splusdata, zie 4.2 en 5.2).

Voor het verwijderen van S-PLUS 2000 kan gebruik gemaakt worden van het verwijderingsprogramma Start / Programs / S-PLUS 2000 / Uninstall #1 of van de methode zoals voor de modules beschreven staat. Ook hier kan het zijn dat het verwijderingsprogramma niet alle onderdelen kan verwijderen, bijv. zelf geschreven libraries zoals RIVMfun. Handmatig kan men dan de directory C:\Program Files\sp2000 verwijderen, nadat men wel de betreffende gegevens veilig gesteld heeft.



## 7. S-PLUS INSTELLINGEN EN BESTANDSORGANISATIE

In dit hoofdstuk worden enkele tips gegeven om S-PLUS zo in te stellen dat men er prettig mee kan werken en dat men dezelfde instellingen elke keer weer heeft. In 7.1 komen algemene instellingen aan de orde. In 7.2 en 7.3 wordt onderscheid gemaakt tussen mensen die S-PLUS heel vaak zullen gaan gebruiken en mensen die S-PLUS incidenteel gaan gebruiken. Eerstgenoemden zullen namelijk S-PLUS vooral aansturen via het commando window, terwijl incidentele gebruikers S-PLUS vanwege de lage insteekdrempel het liefst via de GUI (Grafische User Interface) zullen gebruiken. RIVM-ers die PROAST gebruiken, dienen allereerst de aanwijzingen in de PROAST handleiding te volgen. In 7.4 wordt beschreven hoe men handelingen, met de GUI verricht, kan vastleggen in een zogenaamd scriptbestand. Verder komt in 7.5 het gebruik van de BATCH mode van S-PLUS aan de orde en worden in 7.6. een aantal tips over S-PLUS aangereikt.

In dit hoofdstuk wordt er van uitgegaan dat men al enige kennis van de achtergronden van de opzet van S-PLUS heeft. Om te starten is het goed om de online manual "Getting started Guide" te lezen, die te vinden is onder `Help/Online Manuals` en Appendix B.

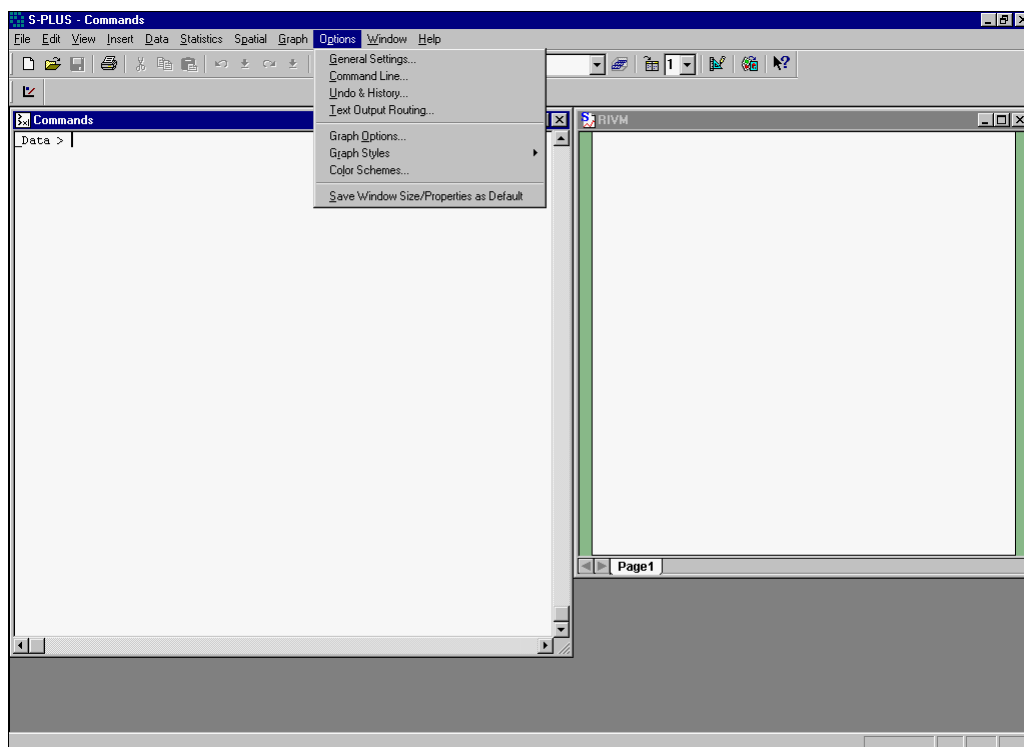
### 7.1 Algemene S-PLUS instellingen

Bij het opstarten van S-PLUS kan men een aantal instellingen bepalen zoals de grootte van het commandoscherm, het te gebruiken lettertype in het commandoscherm etc.

Als men S-PLUS voor het eerst opstart, zal het commandoscherm i.h.a. niet de gewenste grootte en positie op de monitor hebben. Men kan dit op de gebruikelijke wijze via "drag and drop" tot stand brengen. Vervolgens kan men de ingestelde posities bewaren via

`Options/Save Window Size/Properties as default.`

Onder `options` zijn een heleboel andere instellingen van S-PLUS vast te leggen, zoals bijv. het wel of niet laten vragen of men een graph wil bewaren als men een grafisch window afsluit.



Als men onder `Options` een van de menu's opent en dan vervolgens op `help` klikt krijgt men een uitvoerige beschrijving van alle opties te zien. Vandaar dat de vele mogelijkheden hier niet besproken worden. Met de `.First` functie kan men een aantal zaken automatisch opstarten, zoals het aankoppelen van modules en libraries die men vaak gebruikt. Onderstaande `.First` functie kan men zelf maken via het S-PLUS commando `fix(.First)`. Er wordt dan een notepad window geopend waarin een lege S-PLUS functie staat, die men kan wijzigen en opslaan:

```
function()
{
}
```

Indien men een syntax fout heeft gemaakt, wordt dit bij het opslaan vermeld. Met de opdracht `fix()` kan men dan de functie wijzigen en de fout proberen eruit halen.

Onderstaande `.First` functie opent automatisch een grafisch window, stelt de regelbreedte en paginalengte van het commando window in en laadt automatisch de `SpatialStats` module, `Spatial` genaamd, en de RIVM S-PLUS bibliotheek `RIVMfun` automatisch.

```
function()
{
# Tekst na een # is commentaar. Op het RIVM staat deze functie
# in O:/milieu/applics/Splus 2000/RIVM/TXT/first.txt
  module("spatial") #
# Laadt de SpatialStats module
  library("RIVMfun,first=T,lib.loc="o:/milieu/applics/Splus 2000/RIVM")
# Koppelt deze RIVM bibliotheek aan, zie 4.5
  if(interactive()) {
    options(length = 24, width = 60, object.size = Inf)
    guiSetOption(option.name = "ImagecolorTableStd", value.string
      = get("ImageColorTable.15")) # laadt kleurentabel
    graphsheet(18, 18, "normal", 0.45, 0.75, 0.55, 0, 12,
      "portrait", 2.54, pages = "within expression",
      color.scheme = "Standard", image.color.scheme =
      "Standard", num.image.colors = 16, num.image.shades = 0,
      Name = "RIVMStandard") # Open speciaal grafisch scherm
    f.pwd() # functie in RIVMfun die prompt aanpast
  }
  else print(search()[1])
}
```

Men komt bij het opstarten van S-PLUS terecht in `S_PROJ\Data`, waarbij `S_PROJ` de directory is die in de snelkoppeling staat. Indien men steeds in deze directory opstart wordt ook steeds `.First` aangeroepen. Om bovenstaande functie niet te hoeven intypen, kan men op het RIVM deze functie inlezen vanuit het ASCII bestand `first.txt`. Hiertoe dient men de volgende opdracht in S-PLUS te geven:

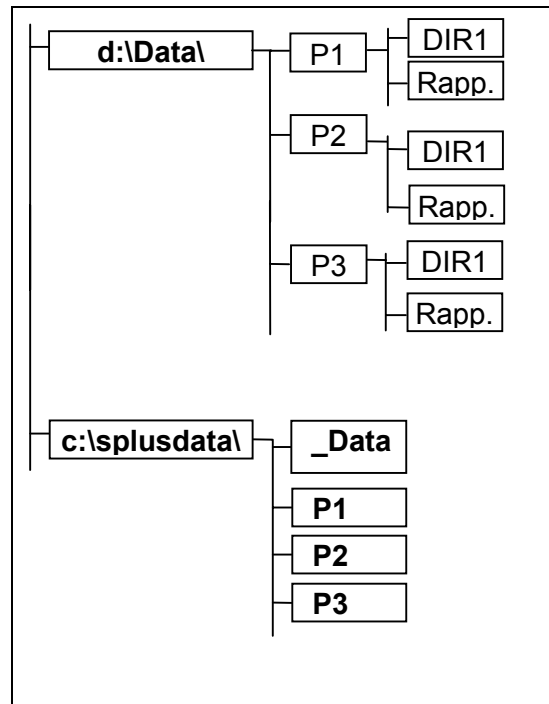
```
source("o:/milieu/applics/SPLUS 2000/RIVM/TXT/first.txt")
```

Typt men vervolgens `ls()` dan krijgt men een lijstje te zien met beschikbare S-PLUS objecten in `S_PROJ\Data`. Een ervan is `.First`.

## 7.2 Gebruik en organisatie van S-PLUS via de commandotaal

Indien men S-PLUS vaak gebruikt, dan zal dit gebeuren in meerdere, verschillende projecten. Hierbij is het natuurlijk wenselijk dat elk project een eigen gegevensdirectory krijgt met daarin de aangemaakte S-PLUS objecten. Binnen S-PLUS wordt in de eerst aangekoppelde map [werkmap genaamd en via het commando `search()` te zien] alles weggeschreven wat er aan gegevens of uitkomsten van berekeningen aangemaakt wordt. Bij het opstarten van S-PLUS is dit de `_Data` directory in `S_PROJ`. Daarom is het handig om per project een specifieke werkmap te maken. Dit kan men het beste doen onder de `S_PROJ` directory, dus in `splusdata`, op de C- of D-schijf (of als men op het RIVM dagelijkse backups wil in `n:\splusdata`), naargelang men voor één van deze drie directories heeft gekozen.

In nevenstaande figuur is te zien dat de gegevens (bijv. ASCII of Excel bestanden) van project P1 in de `d:\data\P1` directory staan. Binnen S-PLUS wordt parallel aan `d:\Data` in `c:\Splusdata` een werkmap, genaamd P1, voor de bijbehorende S-PLUS objecten van dit project gemaakt.



Om deze verschillende gegevensmappen snel binnen een S-PLUS sessie te kunnen aan- en afkoppelen, zijn in RIVMfun de twee functies `f.attach` en `f.detach` opgenomen. De eerste functie kijkt naar alle beschikbare mappen in `S_PROJ`, geeft hiervan een menu-lijstje en koppelt de geselecteerde map op de opgegeven positie aan. Wil men `RBA-MC` als werkmap gebruiken dan dient `RBA-MC` op positie 1 aangekoppeld te worden via de interactieve opdracht `f.attach(,1)`. Twee neven effecten zijn: dat de gebruikelijke werkmap `_Data` op de tweede positie komt te staan en dat de S-PLUS prompt via `f.pwd` wordt gewijzigd in de huidige werkmap naam. Het afkoppelen van de werkmap en aanpassen van de S-PLUS prompt gebeurt via de RIVMfun opdracht `f.detach(1)`.

```


_Data > f.attach(,1)
Which directory do you want to
attach ? (0 = STOP)
1: Functions
2: LAE_BetrEmissie
3: LLO Historisch
4: LLO NH3
5: oo_lmr2001
6: prefbkup
7: RBA-MC
8: RIVMfun
9: Temp
10: _Data
11: _Prefs
Selection: 7
RBA-MC >
  
```

Voordelen van deze aanpak zijn:

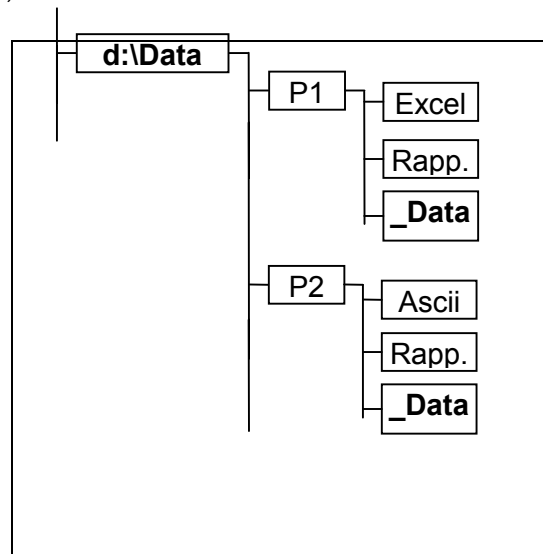
- men kan snel de gewenste werkmap aankoppelen
- bij het opstarten van S-PLUS wordt steeds dezelfde `.First` functie gebruikt, die in `c:\splusdata\_Data` staat en krijgt men dus altijd dezelfde S-PLUS functionaliteiten ter beschikking
- zelf gemaakte functies voor een project zijn snel terug te vinden en in te zetten binnen een ander project.

In Appendix B is een korte inleiding op de S-PLUS commando taal opgenomen.

### 7.3 Gebruik en organisatie van S-PLUS via de GUI

Wanneer men S-PLUS incidenteel gaat gebruiken, kan dit zonder de commandotaal te hoeven leren. Via de GUI kan men snel en intuïtief S-PLUS gebruiken. Deze aanpak vereist wel een andere bestandsorganisatie van S-PLUS dan in 7.2 beschreven. In plaats van het commando scherm zal men van de S-PLUS object browser gebruik maken, op te starten met de volgende knop .

Met deze object browser kan men, net als met de Filemanager voor Windows, de S-PLUS objecten binnen een project beheren: openstaande grafieken, dataframes, functies etc. Men kan per project een eigen object browser instellen. Hiervoor is het noodzakelijk om onder elke projectmap in d:\Data (P1 en P2 in nevenstaand figuur) een map `_Data` te maken, de S-PLUS werkmap van het project.



#### Snelkoppeling per project

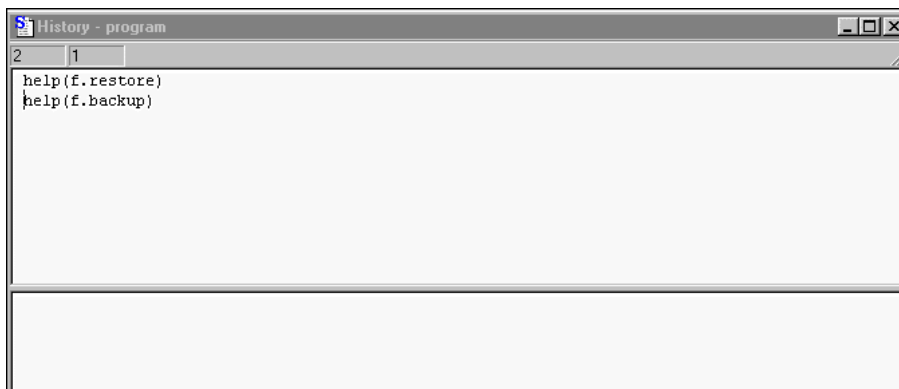
Men kan S-PLUS opstarten en meteen de juiste werkmap met object browser voor een project (bijv. P1 of P2) krijgen, door een snelkoppeling van S-PLUS met dat project te maken. Dit is eenvoudig te doen door een kopie van de bestaande snelkoppeling te maken en die kopie de naam te geven `S-PLUS P1`. Vervolgens wijzigt men in de zojuist gemaakte snelkoppeling `S_PROJ=c:\splusdata` in `S_PROJ=d:\data\P1` op de twee bekende plaatsen `Target` en `Start` (zie 4.2 punt 6). Door nu dubbel te klikken op deze snelkoppeling start met S-PLUS op met als werkmap `d:\data\P1\_Data`.

Deze projectgerichte snelkoppelingen kan men op de desktop zetten of bijvoorbeeld opslaan onder `Start/Programs/Splus 2000 (R3)`. Wil men ook nog gebruik maken van de instellingen van `.First` dan dient men deze functie te kopiëren naar `d:\Data\P1\_Data`. Dit kan het beste *binnen* S-PLUS gebeuren, zie 7.6.3.

In Appendix C is een korte inleiding in het gebruik van de GUI van S-PLUS opgenomen. Het boekje *An introduction to S-PLUS for Windows* van Longhow Lam (Candiensten, 1999) is een goede inleiding op het gebruik van de GUI (zie 8.1 voor bestel mogelijkheid).

### 7.4 Het schrijven van functies via GUI en de history

Binnen S-PLUS worden de opgegeven commando's van zowel de commando taal als de GUI bewaard in een history. De commando's zijn terug te halen via het commando `Ctrl-I` en als men de history van de vorige sessies heeft bewaard (zie 7.6.6), dan staan de commando's van vorige sessies er ook in.






De commando's opgegeven via de GUI zijn te vinden onder `Window/History/Display` (zie bovenstaande figuur). Via deze opdracht wordt de History script file geopend. Bovenin staan de commando's. Deze zijn via `copy-paste` in het onderste gedeelte te kopiëren. Met toets `F10` is de opdracht dan nogmaals uit te voeren. Wil men het geheel opslaan in een nieuw script bestand dan kan men de gewenste regels uit de history kopiëren naar bijvoorbeeld `mijnlescript.ssc`. Dit bestand is aan te maken door onder `File/New/Script file` een nieuwe script file te openen en deze op te slaan als `mijnlescript.ssc`.

Men kan commentaar toevoegen door eerst een hekje (`#`) te typen en daarachter het commentaar te geven.

Voor het gebruik van script bestanden zie de User's Guide, pagina 460 en verder.

## 7.5 S-PLUS in BATCH mode

S-PLUS programma's die veel geheugen vergen kunnen S-PLUS laten vastlopen. Simulatie programma's zijn hier een voorbeeld van. Om geheugen te besparen, bijv. de ruim 32Mb voor de grafische user interface, kan S-PLUS in batch mode gebruikt worden. Daartoe dient het volgende te gebeuren:

1. Maak in een directory drie ASCII bestanden voor de invoer, uitvoer en de foutenafhandeling:  
c:\temp\file\_in.txt  
c:\temp\file\_out.txt  
c:\temp\file\_err.txt  
Deze bestanden hoeven alleen maar lege bestanden te zijn. Ze zijn te maken door in notepad in een nieuw bestand één spatie te typen en het bestand vervolgens op de gewenste plaats met de juiste naam op te slaan.  
Ze staan ook in de map `o:/milieu/applics/splus 2000/RIVM/TXT/Batch` op het RIVM en op de "thuis-CD" onder `RIVM/TXT/Batch` en kunnen rechte lijnen naar `c:\temp` gekopieerd worden.
2. Het bestand `file_in.txt` dient het S-PLUS programma of de functie aanroep te bevatten. Er mogen overigens geen grafieken worden aangemaakt.  
Merk op dat als werkdirectory gebruikt wordt: `S_PROJ_Data`.  
Met `source("o:/milieu/applics/splus 2000/RIVM/TXT/Batch/fbatch.txt")` op het RIVM en thuis met de opdracht `source("e:/RIVM/TXT/Batch/fbatch.txt")` kan men een testprogramma in S-PLUS laden in `S_PROJ_Data`.
3. Vervolgens dient er een nieuwe snelkoppeling gemaakt te worden (S-PLUS BATCH, bijv.) naar S-PLUS waar onder `Properties / shortcut / Target` aan de opstartregel dient te worden toegevoegd:  
`[BATCH c:\temp\file_in.txt c:\temp\file_out.txt c:\temp\file_err.txt]`.  
Indien een andere werkmap gewenst is, dient dit op de gebruikelijke wijze worden aangegeven door het instellen van de commando line parameter `S_PROJ`.  
De snelkoppeling is te vinden in `o:/milieu/applics/splus 2000/RIVM/TXT/Batch` op het RIVM en op de "thuis-CD" onder `RIVM/TXT` en heet S-PLUS BATCH.  
De snelkoppeling is ook te maken door met de muis rechts te klikken op  `Start` en dan explorer te kiezen, vervolgens gaat men in het rechterscherf van de explorer naar `Programs/S-PLUS 2000 (R3)`. Kopieer nu de snelkoppeling naar S-PLUS noem deze S-PLUS BATCH en wijzig de Target regel.
4. Om de batch verwerking te starten dient S-PLUS niet opgestart te zijn. De batch wordt eenvoudig gestart door dubbel te klikken op de nieuwe snelkoppeling S-PLUS BATCH.

## 7.6 Wetenswaardigheden en tips

Nu volgen een aantal meer technische wetenswaardigheden van S-PLUS die handig zijn bij het geavanceerd werken met S-PLUS. Het is een verzameling van antwoorden op vragen die de laatste jaren binnen het RIVM zijn gesteld. De onderdelen 7.6.7 t/m 7.6.9 zijn vooral gericht om de reproduceerbaarheid van het gedane onderzoek te garanderen.

Voor een goed begrip van sommige zaken is kennis van Appendix B vereist.

### 7.6.1 Naamgeving Windows bestanden en paden vanuit S-PLUS

Bestanden en paden onder Windows zijn rechtstreeks te benaderen vanuit S-PLUS. Zelfs complete DOS opdrachten kan men vanuit S-PLUS uitvoeren. Alle paden en bestandsnamen dienen omgeven te worden met dubbele quotes. Omdat de S-PLUS commandotaal een interpreter is, wordt elke "\\" (twee back-slashes) vervangen door één back-slash "\". Het bestand invoer.txt in de map temp op de D-schijf dient men dus aan te geven als

```
"D:\\temp\\invoer.txt"
```

Gelukkig is dit te vereenvoudigen omdat binnen S-PLUS 2000 met ook met slashes kan werken! Daardoor kan men genoemd bestand eenvoudiger weergeven met

```
"D:/temp/invoer.txt"
```

Het gebruik van DOS commando's gaat via de S-PLUS opdrachten `dos` en `win3`. Voor verdere informatie type binnen S-PLUS `help(dos)`.

### 7.6.2 Naamgeving van S-PLUS objecten buiten S-PLUS

S-PLUS is oorspronkelijk onder UNIX ontwikkeld. Daardoor konden objecten willekeurig lange namen hebben. Bij de overgang naar Windows 3.1 heeft men dit opgevangen door binnen S-PLUS een eigen naamgeving bij te houden in de bestanden `__nonfi` en `__sum4.txt`. Het eerstgenoemde bestand kan men met bijv. `notepad` openen. Het bestand bevat één kolom met achtereenvolgens een bestandsnaam (tussen dubbele quotes) en daaronder de bijbehorende DOS/Windows naamgeving bestaande uit twee liggende streepjes en een getal (zie figuur).

In feite zou dit systeem onder Windows 9x en hoger niet meer nodig zijn omdat daarin ook lange namen toegestaan zijn. Helaas is dit niet zo en dat kan soms tot problemen leiden. Objectnamen die men met de Windows Explorer buiten S-PLUS verwacht te zien, zijn er niet. Dikwijls komt dit omdat de S-PLUS objectnamen niet aan de DOS conventie voldoen en dus via bovenstaand systeem onder Windows staan.

```
".last.value"
"__1"
"last.dump"
"__2"
"f.image.da"
"__4"
"no3.akker"
"__5"
"no3.natuur"
"__6"
"no3.gras"
"__7"
"no3.pred.total"
"__8"
```

Gevolg is dat het verstandig is om nooit met de Windows Explorer S-PLUS objecten van een map naar een andere te kopiëren. Dit kan men veel beter binnen S-PLUS doen, zie 7.6.3.

### 7.6.3 Kopiëren en/of verplaatsen van S-PLUS objecten tussen mappen

Het kopiëren van S-PLUS objecten kan op verschillende manieren gebeuren. Stel men heeft in de huidige werkmap `_Data` object `handig` staan en dit object wil men voortaan ook in project `ProRIVM` gebruiken. Alle S-PLUS mappen zijn volgens 7.2 georganiseerd en staan dus in `c:\\splusdata`.

#### Methode 1

1. `f.attach(,1)` en selecteer `ProRIVM` (waarbij `f.attach` alleen beschikbaar is als men de S-PLUS bibliotheek `RIVMfun` heeft aangekoppeld, zie 4.5).
2. `handig <- handig` (omdat alle toekenningen in de huidige werkmap gebeuren, wordt object `handig` uit de map op positie 2 gekopieerd naar de map op positie 1).
3. `f.detach(1)` als men weer verder wil in `_Data`.

4. Indien `handig` niet meer nodig is in `_Data` type dan nog `rm(handig)`.

### Methode 2

Ken object `handig` rechtstreeks toe aan S-PLUS map `ProRIVM`. Indien deze map al aangekoppeld is, zeg op positie 2, dan gebeurt de toekenning als volgt:

```
assign("handig",handig,where=2)
```

hierbij is `"handig"` de naam van het nieuwe object en mag dus ook een andere naam zijn, `handig` is de inhoud van het object `handig` en `where = 2` geeft de positie op de search-list aan, de lijst met alle aangekoppelde S-PLUS mappen, die binnen S-PLUS is op te vragen met de opdracht `search()`. Is `ProRIVM` echter niet aangekoppeld, dan is de volgende `assign` opdracht voldoende:

```
assign("handig",handig,where="c:/splusdata/ProRIVM")
```

### 7.6.4 Systematische naamgeving S-PLUS objecten

De ervaring leert dat een systematische naamgeving van objecten binnen een groot project veel tijdswinst oplevert. Door een goede en consequente naamgeving kan men ook na jaren nog belangrijke objecten zoals gegevens en functies binnen S-PLUS terugvinden.

Stel men heeft de gegevens in een S-PLUS object, genaamd `no3.gras` en gaat daarmee een empirisch variogram berekenen. Het is dan handig om het nieuwe, berekende object `no3.gras.vario` te noemen. Ten eerste blijven zo alle objecten m.b.t. `no3.gras` bij elkaar staan en ziet men aan de extensie `.vario` dat het om een resultaat van de S-PLUS functie `variogram` gaat. Met de opdracht `f.variogram.fit(no3.gras.vario)` kan men vervolgens een theoretisch variogram fitten aan het berekende empirische variogram. Met het resultaat (dat niet in een S-PLUS object kan worden opgeslagen) kan men dan een kriging model opzetten via de S-PLUS opdracht `krige`. Het resultaat kent men toe aan `no3.gras.krige`. Met dit object kan men tenslotte kriging predicties op het 5x5 km grid van Nederland maken via

```
no3.gras.krige.pred <- predict(no3.gras.krige,newdata=nl.grid)
```


Deze analyse levert dus de 4 objecten:

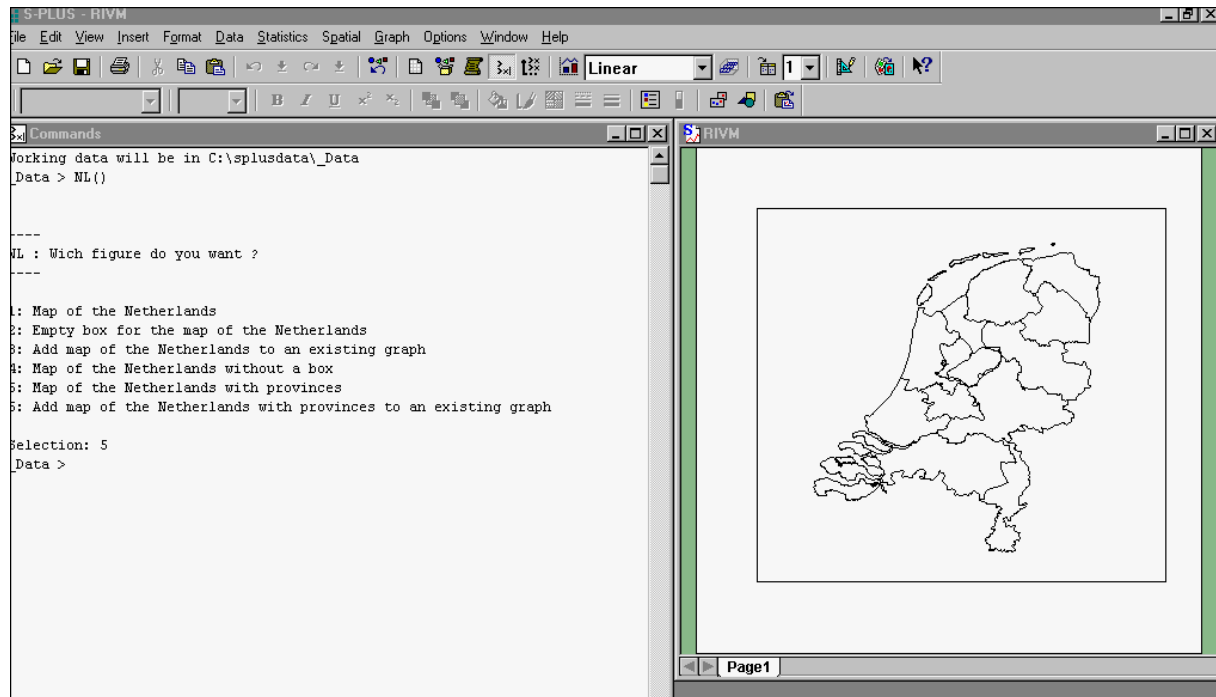
```
no3.gras      no3.gras.vario      no3.gras.krige      no3.gras.krige.pred
```


Wanneer men eigen functies maakt binnen een project is het handig om de naam met `f.` te laten beginnen zoals in `f.variogram.fit`. Dit komt omdat in RIVMfun een oude S-PLUS functie is opgenomen, `ls` genaamd, die de inhoud van een aangekoppelde map laat zien, zoals dit onder UNIX gebruikelijk is. De opdracht `ls("f.*",2)` laat dan alle functies die met `f.` beginnen zien in de map die is aangekoppeld op positie 2. Hiervoor is het wel nodig dat de bibliotheek RIVMfun op positie twee of drie is aangekoppeld (zie 4.5).

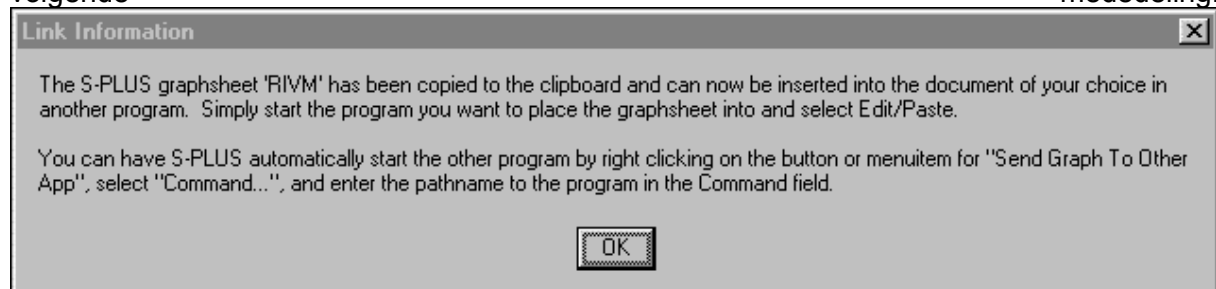
### 7.6.5 Figuren vanuit S-PLUS in WORD documenten plaatsen

S-PLUS bevat een knop om rechtstreeks een S-PLUS figuur in een *openstaand* WORD document te plaatsen. In WORD kan men de figuur desgewenst groter of kleiner maken en kiezen hoe men de tekst om de figuur laat lopen.

De werkwijze is als volgt: open zowel S-PLUS als WORD. Maak vervolgens in S-PLUS een figuur aan, dit kan via het commando scherm of met een menu. Klik dan het grafisch scherm aan. Hierdoor wordt een derde menubalk boven het grafische en commando scherm geopend, waarin de icon om te kopiëren staat, , zoals in de figuur op de volgende pagina te zien is.



Door nu op  te klikken wordt de figuur op het klembord geplaatst en verschijnt de volgende mededeling:



Ga nu naar WORD, plaats daar de cursor op de gewenste plaats en type Ctrl-v. Door rechts op de figuur te klikken in WORD kan men de eigenschappen zoals grootte en tekstomloop wijzigen.

### 7.6.6 De S-PLUS functies *rm* en *remove*

Met de functie `rm` kan men, zoals onder UNIX met de gelijknamige functie, S-PLUS objecten uit de werkmap verwijderen. bijv. door `rm(no3.gras)` in te typen. De functie `remove` kan werken op één van de aangekoppelde mappen en heeft als argument de namen van de objecten als string geschreven. Zo kan men de 4 bovenstaande `no3.*` objecten verwijderen via `remove(ls("no3.*"), where=1)`. Als de map met deze objecten op de tweede positie was aangekoppeld dan volstaat de opdracht `remove(ls("no3.*", 2), where=2)`.

### 7.6.7 Bewaren van uitgevoerde commando's voor een volgende S-PLUS sessie

Bij het afsluiten van een S-PLUS sessie gaan standaard alle gebruikte commando's van de sessie verloren. Om bij een volgende sessie de oude commando's nog ter beschikking te hebben, dient men `S CMDFILE=+history.q` toe te voegen aan de opstart commando regel in de S-PLUS snelkoppeling.

Hiermee wordt in S\_PROJ het bestand `history.q` geopend waarin de S-PLUS opdrachten die interactief via het commando window worden opgegeven, aan het bestaande bestand `history.q` worden toegevoegd. Een meer volledige beschrijving is te vinden in de online Programmer's Guide, in het onderdeel "Command line options". Men kan in Acrobat Reader

ook zoeken naar S\_CMDFILE door naar Help/Online Manual/Programmer's Guide te gaan, de manual te openen en dan Ctrl-f te typen.

### 7.6.8 Het uitvoeren en bewaren van analyses in S-PLUS

Het uitvoeren en bewaren van analyses binnen S-PLUS gebeurt het gemakkelijkst via het opslaan in bijv. een WORD document van de handelingen. De opdrachten die men via het commando scherm opgeeft en de uitvoer die in het commando scherm verschijnt kan men selecteren en vervolgens kopiëren via Ctrl-c. Met Alt-Tab kan men dan naar het openstaande WORD document gaan en met Ctrl-v de geselecteerde tekst in het "logboek" plakken. De figuren kan men vanuit S-PLUS in het WORD document zetten volgens de methode beschreven in 7.6.5.

Hieronder staat een eenvoudig voorbeeld in en twee figuren uitgezet. De eerste is het commando scherm in S-PLUS met de geselecteerde tekst. De volgende is het WORD document met een deel van de gekopieerde tekst in het lettertype COURIER new.

```

815.735 -0.0006629414 0.0003071783

Degrees of freedom: 320 total; 317 residual
Dropped 24 cases due to missing values
Residual standard error: 73.48516
Data > knmi.lm<-lm(jr92~x+y, knmi,na.action=na.omit)
Data > summary(knmi.lm)

Call: lm(formula = jr92 ~ x + y, data = knmi, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-165.6  -46.18  -2.508   47.92  250.5

Coefficients:
            Value Std. Error  t value Pr(>|t|)
(Intercept)  815.7350    26.8970   30.3281  0.0000
            x   -0.0007     0.0001   -8.9251  0.0000
            y    0.0003     0.0001    4.9297  0.0000

Residual standard error: 73.49 on 317 degrees of freedom
Multiple R-Squared:  0.2039
F-statistic: 40.6 on 2 and 317 degrees of freedom, the p-value is 2.22e-016

24 observations deleted due to missing values

Correlation of Coefficients:
(Intercept)  x
x  0.0291
y -0.9061  -0.4261
Data > plot(knmi.lm)
Data >

```

```

Microsoft Word - logboek.doc
File Edit View Insert Format Tools Table Window Help
ASCII Courier New 10 B I U
knmi.lm<-lm(jr92~x+y, knmi,na.action=na.omit)
_Data > summary(knmi.lm)

Call: lm(formula = jr92 ~ x + y, data = knmi, na.action = na.omit)
Residuals:
    Min       1Q   Median       3Q      Max
-165.6  -46.18  -2.508   47.92  250.5

Coefficients:
              Value Std. Error  t value Pr(>|t|)
(Intercept)  815.7350   26.8970   30.3281  0.0000
            x   -0.0007    0.0001   -8.9251  0.0000
            y    0.0003    0.0001    4.9297  0.0000

Residual standard error: 73.49 on 317 degrees of freedom
Multiple R-Squared:  0.2039
F-statistic: 40.6 on 2 and 317 degrees of freedom, the p-value is 2.22e-016

24 observations deleted due to missing values

Correlation of Coefficients:
(Intercept)      x
x  0.0291
y -0.9061      -0.4261

```

### 7.6.9 **Het maken van back-ups en het terugzetten van back-ups in S-PLUS**

Vooral bij het ontwikkelen van S-PLUS functies is het handig om regelmatig een back-up te maken van de in ontwikkeling zijnde functie. Dit kan men doen met de S-PLUS functie `dump`, waarmee men een S-PLUS object wegschrijft als een ASCII bestand. De functie is weer terug te zetten in S-PLUS via de opdracht `source`.

Om het e.e.a. te vergemakkelijken zijn er twee eenvoudige functies ontwikkeld en in RIVMfun opgenomen: `f.backup` en `f.restore`, waarmee men eenvoudig en grotendeels geautomatiseerd back-ups kan maken en weer kan terug zetten. Een volledige beschrijving van de functies vindt men door de RIVMfun bibliotheek te laden (zie 4.5) en dan in te tikken: `help(f.backup)` resp. `help(f.restore)`. Merk op dat met deze functies de oude back-up wordt overschreven.

### 7.6.10 **Versiebeheer binnen S-PLUS**

Het versiebeheer binnen S-PLUS is geen eenvoudige zaak, omdat elke toekenning meteen wordt vastgelegd op de harde schijf, waarbij het oude object wordt overschreven. Het kan dus gemakkelijk voorkomen dat S-PLUS objecten ongewild worden overschreven. Het eenvoudigst is het versiebeheer uit te voeren via de bestandsorganisatie zoals beschreven in 7.2 en daarbij voor elke versie een nieuwe werkmapp te maken.

### 7.6.11 **Back-up van S\_PROJ op een CD en het teruglezen**

Om alle projecten binnen S-PLUS van een gegeven jaar te bewaren is het voldoende om S\_PROJ en alle onderliggende mappen op een CD-ROM te branden. Daarbij dient men op te letten dat ook alle *lege* bestanden zoals `__sum4.txt` worden mee genomen.

Via `attach` kan men de S-PLUS werkmappen op de CD-ROM aan een S-PLUS sessie koppelen, waarbij men dit *niet* op de eerste positie dient te doen, daar S-PLUS dan niets kan wegschrijven. Op deze wijze blijft men oude S-PLUS objecten ter beschikking houden.

Indien men verder wil werken in een oude S-PLUS werkmap op de CD-ROM, dan dient men de betreffende werkmap in zijn geheel terug te kopiëren naar S\_PROJ via Windows Explorer. Na het kopiëren dient men wel voor de hele werkmap het attribute *Read-only* uit te zetten. Dit doet men door in Windows Explorer rechts te klikken op de teruggezette werkmap en onder Properties het vakje *Read-only* aan te klikken. Vervolgens kan men in de betreffende werkmap aan de slag.





## 8. S-PLUS DOCUMENTATIE EN INFORMATIE

De S-PLUS manuals zijn online te vinden onder het menu `Help/Online Manuals`. Het zijn PDF bestanden die met Acrobat Reader op het scherm te lezen zijn. Velen vinden het lezen van manuals op een scherm niet zo prettig. Een voordeel van lezen met Acrobat Reader is de mogelijkheid om met de gebruikelijke `Ctrl-f` gemakkelijk naar de tekst kan worden gezocht. Verder hebben de modules S+SeqTrial en S+GARCH online manuals.

### 8.1 Boeken te bestellen bij Candiensten

Wil men een gedrukte versie van de manuals van S-PLUS of van modules dan dient men zich te wenden tot CANDIENSTEN in Amsterdam, de leverancier van S-PLUS.

In onderstaand schema staan de bij hen te bestellen handleidingen en boeken weergegeven. Tevens staat aangegeven hoe men de bestellingen kan verrichten. De levertijd bedraagt ongeveer een week.

Uitgebreide boekbesprekingen van een aantal boeken staan in Appendix E .

Titel	Auteurs	Prijs	Verzendkosten per bestelling
Manual set S-PLUS (Windows/ UNIX/LINUX#) CD (ja/nee#) Bevat : Users Manual Programmers Manual Guide to Statistics I Guide to Statistics II	Insightful Corporation	NLG 295.-  <b>excl. BTW</b>	NLG 35.-
Manual set S-PLUS + Modules (Windows/ UNIX/LINUX#) CD (ja / nee #)	Insightful Corporation	NLG 350.-  <b>excl. BTW</b>	NLG 35.-
An introduction to S-PLUS for Windows	Longhow Lam Candiensten, 1999	NLG 35.- incl. BTW	---
Modern and applied Statistics with S-PLUS (3e editie)	Venables and Ripley Springer, 1999	NLG 145.- incl. BTW	---
Modern and applied Statistics with S-PLUS (2e editie)	Venables and Ripley	NLG 95.- incl. BTW	---
Modelling Survival Data, Extending the COX Model	Therneau and Grambsch Springer, 2000	NLG 125.- incl. BTW	---
Mixed Effects models in S and S-PLUS	Pinheiro and Bates Springer, 2000	NLG 165.- incl. BTW	---

De bestellingen zijn te plaatsen per fax / email / telefoon en post:  
 CANDIENSTEN tel.: 020 - 560 8430  
 Nieuwpoortkade 23-25 fax: 020 - 560 8448  
 NL-1055 RX AMSTERDAM email : [fischer@candiensten.nl](mailto:fischer@candiensten.nl)

# Doorhalen wat niet van toepassing is. De drie besturingssystemen Windows / UNIX / LINUX hebben verschillende manuals.

## **8.2 Literatuur over gebruik van en programmeren in S-PLUS**

### THE NEW S LANGUAGE

Richard A. Becker, John M. Chambers, and Allan R. Wilks  
Wadsworth & Brooks/Cole, Pacific Grove, CA (1988)

### PROGRAMMING WITH DATA

John M. Chambers  
Springer-Verlag, New York, NY (1998), ISBN 0-387-98503-4

### THE BASICS OF S AND S-PLUS (*second edition*)

Andreas Krause and Melvin Olson  
Springer-Verlag, New York (2000), ISBN 0-387-98961-7

### AN INTRODUCTION TO S AND S-PLUS

P. Spector  
Duxbury Press, Belmont, CA (1994), ISBN 0-534-19866-X

### S PROGRAMMING

W. N. Venables and B. D. Ripley  
Springer, New York (2000), ISBN 0-387-98966-8

## **8.3 Literatuur over data analyse met S-PLUS**

### AN INTRODUCTION TO S-PLUS AND THE HMISC AND DESIGN LIBRARIES

C.F. Alzola and F.E. Harrell  
vrij verkrijgbaar document op <http://hesweb1.med.virginia.edu/biostat/s/splus.html>

### ROBUST DIAGNOSTIC REGRESSION ANALYSIS

Anthony Atkinson and Marco Riani  
Springer (2000), ISBN 0-387-95017-6

### APPLIED SMOOTHING TECHNIQUES FOR DATA ANALYSIS: THE KERNEL APPROACH WITH S-PLUS ILLUSTRATIONS

Bowman & Azzalini  
Claredon Press (1997), ISBN 0-19-852396-3

### APPLIED WAVELET ANALYSIS WITH S-PLUS

Andrew Bruce and Hong-Ye Gao  
Springer-Verlag, New York, NY (1996), ISBN 0-387-94714-0

### STATISTICAL MODELS IN S

John Chambers and Trevor Hastie  
Chapman and Hall (1992), ISBN 0-412-05291-1 (hb) , ISBN 0-412-05301-2 (pb)

### BOOTSTRAP METHODS AND THEIR APPLICATION

A.C. Davison and D.V. Hinkley  
Cambridge University Press (1997), ISBN 0-521-57391-2

### AN INTRODUCTION TO THE BOOTSTRAP

Bradley Efron and Robert Tibshirani  
Chapman and Hall (1994), ISBN 0-412-04231-2

### A HANDBOOK OF STATISTICAL ANALYSES USING S-PLUS

B. Everitt  
Chapman and Hall, London (1994)

### SMOOTHING TECHNIQUES WITH IMPLEMENTATION IN S

W. Haerdle  
Springer-Verlag, New York (1991), ISBN 0-387-97367-2

## GENERALIZED ADDITIVE MODELS

T. Hastie and R. Tibshirani  
Chapman and Hall (1990)

REGRESSION MODELING STRATEGIES WITH APPLICATIONS TO LINEAR MODELS, LOGISTIC  
REGRESSION, AND SURVIVAL ANALYSIS

F.E. Harrell  
Springer (2001)

STATISTICAL TOOLS FOR NONLINEAR REGRESSION: A PRACTICAL GUIDE WITH S-PLUS  
EXAMPLES

S. Huet, A. Bouvier, M.-A. Gruet, and E. Joliet  
Springer-Verlag, New York, NY (1996), ISBN 0-387-94727-2

## S+SPATIALSTATS USER'S MANUAL

S.P. Kaluzny, S.C. Vega, T.P. Cardoso, and A.A. Shelly  
Springer-Verlag, New York, NY (1997)

## ALGORITHMS, ROUTINES AND S FUNCTIONS FOR ROBUST STATISTICS

A. Marazzi  
Wadsworth and Brooks/Cole, Pacific Grove, CA (1993)

## STATISTICAL METHODS FOR RELIABILITY DATA

W. Meeker and L. Escobar  
John Wiley & Sons, Inc. (1998), ISBN 0-471-14328-6

## ENVIRONMENTAL STATISTICS WITH S-PLUS

Steven P. Millard and Nagaraj K. Neerchal  
CRC Press (2000), ISBN 0-8493-7168-6

## STAT LABS: MATHEMATICAL STATISTICS THROUGH APPLICATIONS

D. Nolan & T.P. Speed  
Springer (2000), ISBN 0-387-98974-0

## MODERN APPLIED BIOSTATISTICAL METHODS USING S-PLUS

S. Selvin  
Oxford University Press (1998), ISBN 0-19-512025-6

## INTRODUCTION TO ROBUST ESTIMATION AND HYPOTHESIS TESTING

R. Wilcox  
Academic Press (1997), ISBN 0-12-751545-3

**8.4 Literatuur over grafische weergaven in S-PLUS**

## GRAPHICAL METHODS FOR DATA ANALYSIS

J.M. Chambers, W.S. Cleveland, B. Kleiner, and P.A. Tukey  
Duxbury Press, Belmont, CA (1983)

## VISUALIZING DATA

William S. Cleveland  
Hobart Press, Summit, NJ (1993), ISBN 0-9634884-0-6

## THE ELEMENTS OF GRAPHING DATA

William S. Cleveland  
Hobart Press, Summit, NJ (1994)

### **8.5 S-PLUS via het WEB**

S-PLUS staat al jaren bekend om een zeer actieve gebruikersgroep die dagelijks met elkaar communiceert. De internationale groep is zo actief dat de Nederlandse email groep, die op de eerste Nederlandse gebruikersdag is opgericht, amper emails bevat.

Deelname aan de internationale S-PLUS nieuwsgroep start met door een email te sturen naar

[s-news-request@lists.biostat.wustl.edu](mailto:s-news-request@lists.biostat.wustl.edu)

met als onderwerp `subscribe`. De rest wijst zich vanzelf. Het archief van deze nieuwsgroep is bereikbaar via

[www.biostat.wustl.edu/s-news](http://www.biostat.wustl.edu/s-news)

De Nederlandse S-PLUS email groep is te bereiken via

<http://www.candiensten.nl/s+gebruikersgroep/>

Informatie over S-PLUS en alle achtergronden zijn te vinden bij de Nederlandse leverancier CANDIENSTEN:

<http://www.candiensten.nl/>

en bij de producent van S-PLUS:

<http://www.insightful.com/>

waar ook bovenstaande literatuurlijsten in 8.2 t/m 8.4 vandaan komen. Op het volgende adres zijn de lijsten te bekijken en kan men meer informatie over de betreffende boeken vinden:

<http://www.insightful.com/resources/biblio.html>

Verder is er een bibliotheek, Statlib genaamd, waar men allerlei gratis beschikbaar gestelde S-PLUS programma's kan vinden en downloaden

<http://lib.stat.cmu.edu/S/>

Men kan er ook bibliotheken voor andere pakketten vinden via

<http://lib.stat.cmu.edu/>

## 9. DBMS/COPY

Ondanks dat DBMS/COPY geen onderdeel van S-PLUS is, maar een zelfstandig software pakket om de uitwisseling van gegevens tussen software pakketten te vereenvoudigen, is het de moeite waard om een korte beschrijving van DBMS/COPY in dit rapport op te nemen.

DBMS/COPY versie 7 is voor de gehele Milieusector beschikbaar en stelt de gebruikers in staat om bijvoorbeeld Excel worksheets of SAS bestanden in S-PLUS dataframes om te zetten buiten Excel, SAS en S-PLUS om. In S-PLUS zelf kan dit interactief ook. Het voordeel van DBMS/COPY is echter dat men op het einde van het interactieve proces in de stap vóór de omzetting, de voorbereide opdracht in een BATCH programma kan opslaan. Heeft men meerdere Excel worksheets in één bestand staan met dezelfde opbouw, dan kan men met eenvoudig programmeren het BATCH programma uitbreiden en vervolgens alle gegevens in één keer omzetten naar S-PLUS. Dit is binnen S-PLUS *niet* mogelijk. Het zal duidelijk zijn dat vanuit kwaliteitsoogmerk deze werkwijze de herhaalbaarheid van de actie garandeert.

Dit werd onlangs bevestigd toen een rapport enkele jaren na het onderzoek uitkwam. Toen iemand de gepubliceerde gegevens opvroeg, bleek dat de gegevens binnen S-PLUS door een back-up fout niet meer beschikbaar waren. Met een DBMS/COPY BATCH programma konden de oorspronkelijke Excel gegevens opnieuw in S-PLUS worden ingelezen. Daar is met een gelukkig nog wel beschikbaar S-PLUS programma de gepubliceerde tabel gereproduceerd.

De ervaring leert dat DBMS/COPY in het algemeen sneller inspeelt op nieuwe data formaten zoals SAS versie 8, Excel 2000 etc. Daarbij kan DBMS/COPY ook vele vormen van ASCII bestanden aan met verschillende delimiters of met een vaste kolommen-indeling.

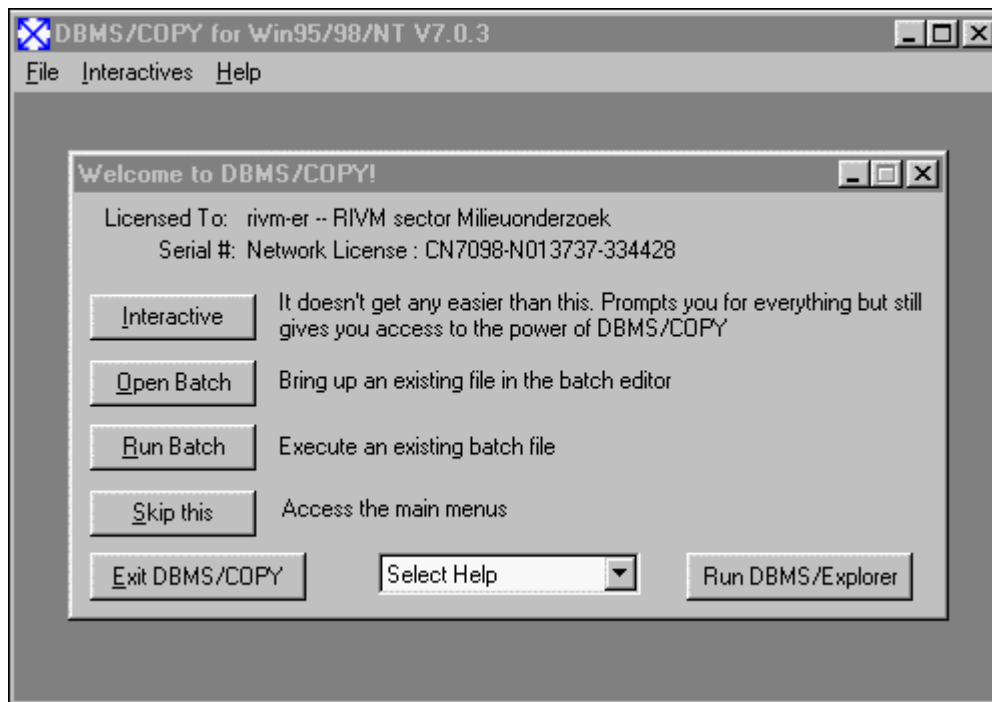
DBMS/COPY is ook erg handig als men vaak veel bestanden met dezelfde opbouw in S-PLUS dient te importeren. Dit kan dan snel geprogrammeerd worden via het kopiëren van commando regels en vervolgens in één keer worden uitgevoerd.

In de nu volgende figuren wordt kort uiteen gezet hoe men met DBMS/COPY een worksheet van een Excel bestand (met meerdere worksheets) in S-PLUS kan importeren.

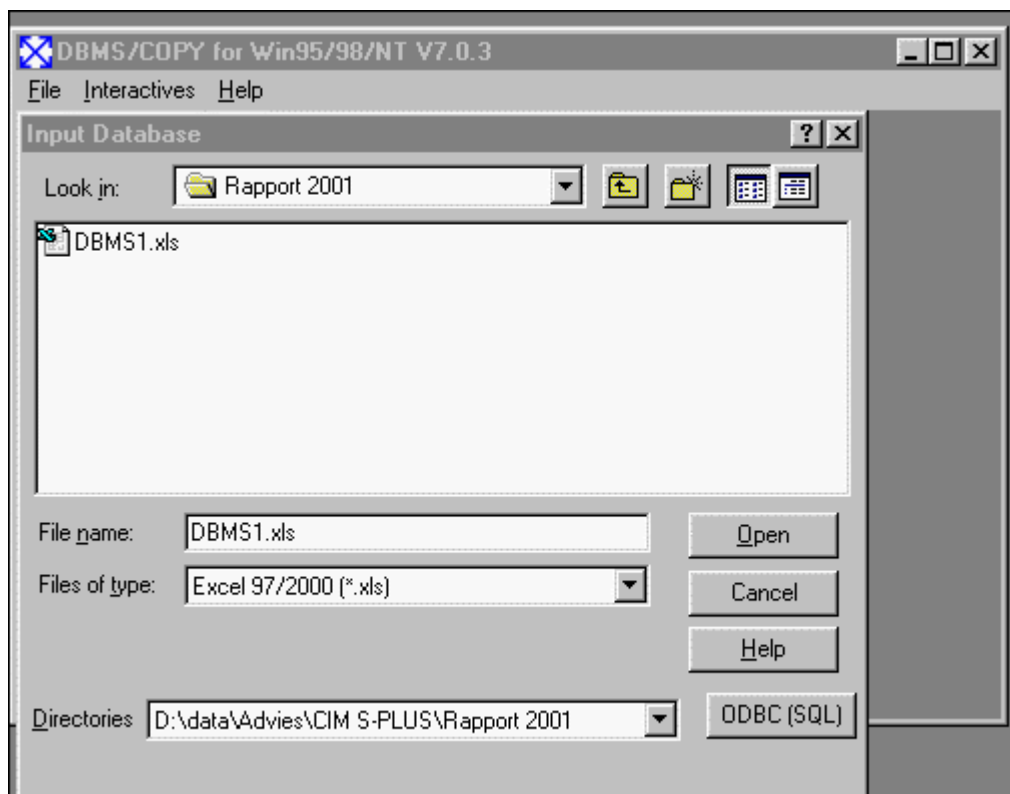
De Excel spreadsheets bevatten voor de jaren 1993, 1994 en 1995 gefingeerde ammonium, sulfaat en nitraat gegevens voor steeds dezelfde stations en dezelfde maanden.

	A	B	C	D	E	F	G	H	I	J
1	start	st1	st2	st3	st4	st5	st6	st7	st8	
2	30-12-1992	36	37	26	29	36	37	26	29	
3	27-1-1993	48	54	28	61	48	54	28	61	
4	24-2-1993	71	92	66	94	71	92	66	94	
5	24-3-1993	50	41	40	47	50	41	40	47	
6	21-4-1993	100	79	52	63	100	79	52	63	
7	19-5-1993	47	36	31	30	47	36	31	30	
8	16-6-1993	27	27	22	66	27	27	22	66	
9	14-7-1993	33	27	18	23	33	27	18	23	
10	11-8-1993	37	28	22	18	37	28	22	18	
11	8-9-1993	21	20	19	21	21	20	19	21	
12	6-10-1993	17	14	13	17	17	14	13	17	
13	3-11-1993	22	23	23	21	22	23	23	21	
14	1-12-1993	27	27	17	21	27	27	17	21	

Bij het opstarten van DBMS/COPY komt men in het volgende scherm:



Kies Interactive en selecteer het juiste datatype en database:



Door nu naar de juiste pagina te klikken en aan te geven waar de in te lezen gegevens staan, kan men het gewenste gegevensgebied selecteren. Linksboven wordt dan automatisch aangegeven welk gebied men geselecteerd heeft. Automatisch wordt aangegeven dat op de eerste rij de kolomnamen staan. Dit kan men ook door de juiste rij aan te wijzen, veranderen. Dit alles is in de volgende figuur te zien.

Spreadsheet Grabber

Data Area: A2 To I14 Next Page Ok  
 Variable Name Rows: 1 To 1 Prev Page Cancel  
 (Use 0 for no names)

	A	B	C	D	E	F	G	H	I
1	start	st1	st2	st3	st4	st5	st6	st7	st8
2	12/30/92	36	37	25	28	36	37	25	28
3	01/27/93	47	53	27	61	47	53	27	61
4	02/24/93	70	91	66	94	70	91	66	94
5	03/24/93	49	41	40	46	49	41	40	46
6	04/21/93	100	79	51	63	100	79	51	63
7	05/19/93	46	36	31	30	46	36	31	30
8	06/16/93	27	27	22	66	27	27	22	66
9	07/14/93	33	26	18	23	33	26	18	23
10	08/11/93	36	28	21	17	36	28	21	17
11	09/08/93	20	20	18	21	20	20	18	21
12	10/06/93	16	14	12	17	16	14	12	17
13	11/03/93	21	23	23	21	21	23	23	21
14	12/01/93	27	26	17	21	27	26	17	21

Door nu op Ok te klikken, gaat men naar de volgende fase waarin staat samengevat wat men heeft geselecteerd. Men kan hier ook nog bepaalde kolommen weglaten etc.

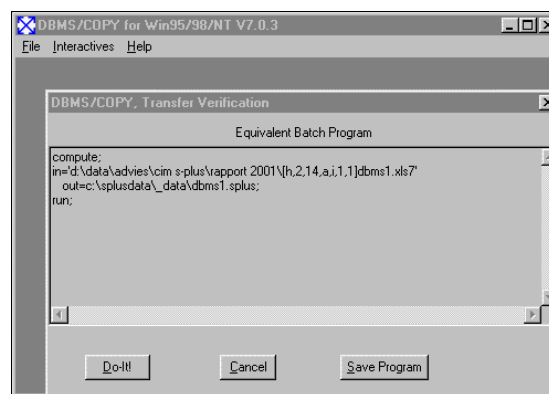
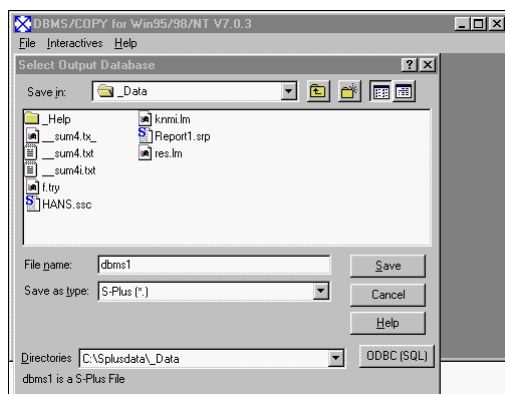
DBMS/COPY Spreadsheet Variable Type Override

Here are variables DBMS/COPY wants to create. Change the types and names if you want.

	Name	Drop	Type	Length	# Char	# Num	# Date	# Time	#Date
1	start	<input type="checkbox"/>	Date	0			13		
2	st1	<input type="checkbox"/>	Number	12		13			
3	st2	<input type="checkbox"/>	Number	12		13			
4	st3	<input type="checkbox"/>	Number	12		13			
5	st4	<input type="checkbox"/>	Number	12		13			
6	st5	<input type="checkbox"/>	Number	12		13			
7	st6	<input type="checkbox"/>	Number	12		13			
8	st7	<input type="checkbox"/>	Number	12		13			
9	st8	<input type="checkbox"/>	Number	12		13			
10		<input type="checkbox"/>							

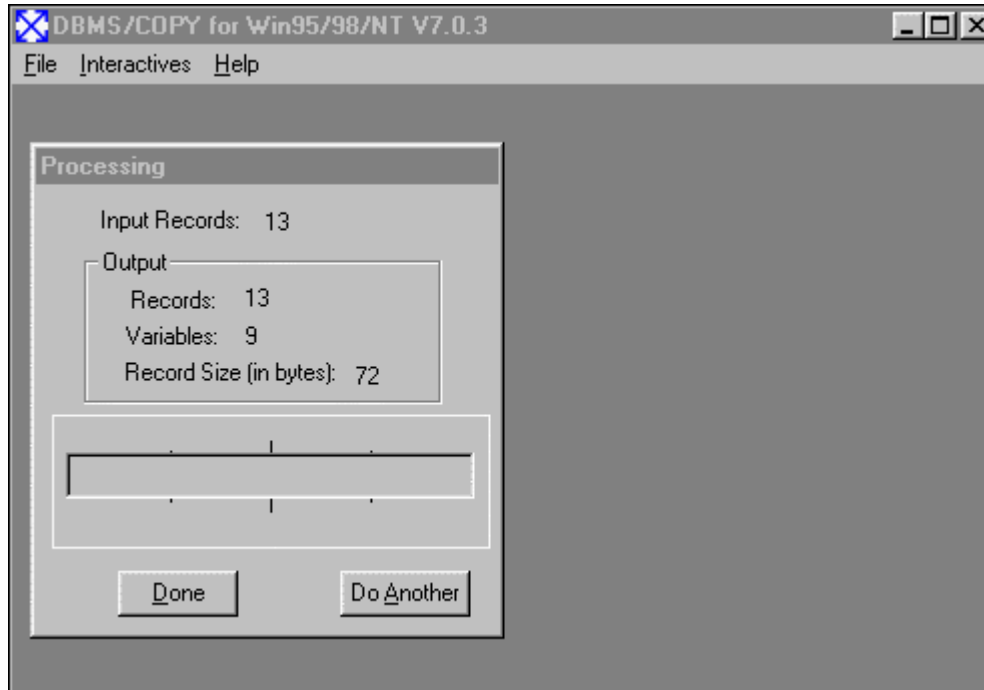
Indien alles goed is, klikt men weer op ok.

Nu kan men zelf bepalen in welke S-PLUS map men het nieuwe S-PLUS object wil hebben. In onderstaande figuur is het in c:\splusdata\\_Data.



Wanneer men tenslotte op Save klikt, krijgt men de gelegenheid om de interactief opgestelde opdracht als een DBMS/COPY BATCH programma in een \*.prg bestand op te slaan, in dezelfde map als het originele Excelbestand.

Klik op save en vervolgens op Do-it!





## VERZENDLIJST

1. Directie Rijksinstituut voor Volksgezondheid en Milieu
2. Ir. F. Langeweg
3. G. Fischer, Candiensten, Amsterdam
4. Dr. R. Koning, RijksUniversiteit Groningen
5. Dr. M. Schipper, Universiteit Utrecht
6. D. Verkerk, Candiensten, Amsterdam
7. Dr. ir. W. de Winter, Candiensten, Amsterdam
8. H. de Wit, Stichting Farmaceutische Kengetallen
9. Depot van Nederlandse publicaties en Nederlandse bibliografie
10. Drs. J.J.M. Aben
11. Drs. T. Aldenberg
12. Dr. ir. J.R.M. Alkemade
13. Mevr. C.B. Ameling
14. Drs. A.H. Bakema
15. Ing. M. Bakkenes
16. Mevr. Dr. M.I. Bakker
17. Dr. R.B. Beems
18. Ir. R. v.d. Berg
19. Ing. W.G.M. Bergkamp
20. Drs. A.U.C.J. van Beurden
21. Drs. A.H.W. Beusen
22. Ing. J.D. te Biesenbeek
23. Dr. ir. B.P.M. Bloemberg
24. Mevr. drs. J.A.M. Borsboom-van Beurden
25. Mevr. prof. dr. G.A.M v.d. Bos
26. Mevr. Dr. H.C. Boshuizen
27. Ir. L.J.M. Boumans
28. Dr. ir. A.F. Bouwman
29. Drs. A.A. Bouwman
30. Dr. L.C. Braat
31. Ir. A.H.M. Bresser
32. Dr. ir. J.J.B. Bronswijk
33. Drs. E. Buijsman
34. Ir. K. Buurman
35. Drs. J.H. Canton
36. L.E.M. Crommentuijn
37. Ir. J.D. van Dam
38. Ir. A.G.M. Dassen
39. Mevr. ir. G. Doornbos
40. Mevr. ir. A. Dusseldorp
41. Mevr. ir. P.M. van Egmond
42. Drs. J.C.H. van Eijkeren
43. H.W. Evers
44. Mevr. P.F.L. Feimann
45. Ir. B. Fraters
46. Dr. J.I. Freijer
47. Dr. J. Garssen
48. Mevr. ir. M.L.L. van Genugten
49. Drs. A. v.d. Giessen
50. Ir. B.J. deHaan
51. Ir. P. Hammingh

52. N. ter Harkel
53. M.J. 't Hart
54. Mevr. C.F. den Hartog Jager
55. Mevr. ir. M. de Heer
56. Dr. S.H. Heisterkamp
57. Dr. P.S.C. Heuberger
58. Mevr. dr. B. Hoebee
59. J.H.G. Hoekman
60. Mevr. dr. J.A. Hoekstra
61. Ir. R.T. Hoogenveen
62. Dr. R.Hoogerbrugge
63. Ir. D.J.M. Houthuijs
64. Ir. J. Jabben
65. Drs. J.H. Janse
66. Dr. ir. P.H.M. Janssen
67. Dr. L.H.J.M. Janssen
68. Mevr. ing. G.A. Janssen
69. Ir. G.J. Kommer
70. Dr. W.H. Könemann
71. Drs. J.L. Kool
72. Drs. F.J.Kragt
73. Dr. ing. H. van Kranen
74. Mevr. Dr. M.E.E. Kretzschmar
75. Mevr. dr. M.A.J. Kuypers-Linde
76. Drs. W. de Lange
77. Dr. ir. E. Lebret
78. Drs. C. de Lezenne-Coulander
79. Drs. F. Lips
80. Dr. ir. D. van Lith
81. Ir. A.H.P. Luijben
82. Drs. R.J.M.Maas
83. Ing. C.W.M. v.d. Maas
84. Ing. M. Marra
85. Mevr. E.A.M. Mathijssen
86. Dr. ir. D. van de Meent
87. Mevr. dr. M. Moerbeek
88. Ir. W.J.A.Mol
89. Dr. N.J.D. Nagelkerke
90. Dr. M.J.Nauta
91. Mevr. D.S. de Nie
92. Drs. A.C.M. de Nijs
93. Drs. H. Noordijk
94. Drs. F.M.M. van Och
95. Dr. M.M.P. van Oorschot
96. Dr. A. Opperhuizen
97. P.F. Otte
98. Mevr. dr. ir. M. Pieters
99. Drs. M.J.J.C. Poos
100. Dr. H.F.R. Reijnders
101. L. Renaud
102. Mevr. E.C.M. Rentinck
103. Ing. A.J. Rietveld
104. Ir. F.J.A. van Rijn
105. Drs. M.G.J. Rikken
106. Dr. D. Ruwaard

107. Mevr. Ing. E. Schenk
108. Dr. ir. J.C. Seidell
109. Dr. H. Slaper
110. Dr. W. Slob
111. Ing. J. Slootweg
112. Ir. P.A.M. de Smet
113. Dr. R.C.G.M. Smetsers
114. Dr. J.P.J.M. Smits
115. Drs. I. Soenario
116. Drs. J.A. Suurmond
117. Ing. W.J.R. Swart
118. Mevr. dr. ir. C. Tabak
119. Dr. P.F.M. Teunis
120. Dr. A. Tiktak
121. Dr. M.P. van Veen
122. Dr. ir. G.J.M. Velders
123. Drs. K. v. Velze
124. A. Verhoef
125. Dr. ir. C.W. Versluijs
126. Dr. ir. H. Visser
127. Prof. dr. J.G.Vos
128. Ir. K. Wieringa
129. Drs. J. Wiertz
130. Mevr. M.A.H. Wolters-Balk
131. Drs. F.G. Wortelboer
132. Dr. ir. M.J. Zeilmaker
133. Drs. D. de Zwart
134. Auteur
135. SBD/Voorlichting & Public Relations
136. Bureau Rapportenregistratie
137. Bibliotheek RIVM
- 138-147. Bureau Rapportenbeheer
- 148-149. Bibliotheek RIVM Uitleenbalie (voor thuisgebruik)
- 150-175. Reserve exemplaren auteur



## APPENDIX A UITVOERIGE BESCHRIJVINGEN VAN DE MODULES

---

### S+DOX (1998) Version 1.1 Release 2

---

S+DOX is an add-on module for the design and analysis of fractional factorial, response surface and robust (Taguchi) designs.

The module strongly emphasizes graphical displays for exploring and analysing experimental data. It uses new methodology for estimating standard errors in saturated models, enabling better assessment of important experimental factors.

The response surface functions provide an easy analysis and display of response surfaces with contour, image and surface plots.

The robust design functions provide designs for control and noise factors and allow calculation and analysis of signal to noise ratios and other responses. The analysis identifies factors that enable

the process to be optimized with minimal sensitivity to noise.

Some primary functions of the module are:

design.digest	fac.aov	rsm.design	rsm.lm	robust.design	robust.sn
pareto	tfiplot	acplot	ebplot		

---

### S+GARCH (2000)

---

S+GARCH : An S-PLUS module for Generalized AutoRegressive Conditional Heteroscedastic models.

S+GARCH is an object-oriented toolkit for modeling volatility of univariate and multivariate time series. S+GARCH offers the following tools:

- Univariate GARCH models
- Multivariate GARCH models
- Fractionally Integrated GARCH models
- A Variety of Visual Data Analysis Displays

---

## **NUOPT for S-PLUS (1998)**

---

NUOPT for S-PLUS : An S-PLUS module for defining and optimizing various numerical models.

NUOPT for S-PLUS is a combination of S-PLUS programming environment, the modeling language SIMPLE and the optimizer NUOPT. NUOPT for S-PLUS offers a wide collection of functions for descriptions of various optimization problems as well as tools to solve these problems. NUOPT for S-PLUS offers the following tools:

- Primal-dual interior point method with higher order correction for Linear Programming (LP) models.
- Simplex method for Linear Programming (LP) and Mixed Integer Programming (MIP) models.
- Primal-dual interior point method based on line search for general Convex Programming (CP) models.
- Primal-dual interior point method based on trust region method for general Non-Linear Programming (NLP) models.
- Primal-dual interior point method based on quasi-Newton method for general Non-Linear Programming (NLP) models.

---

## **S+SeqTrial (1999) Version 1.0 Release 1**

---

S+SeqTrial is the leading tool for the design and analysis of clinical trials and research studies, including both fixed sample designs and sequential designs. S+SeqTrial includes:

- A Microsoft Office-compatible graphical user interface designed for the clinical trialist;
- A comprehensive set of specialized plots for designing studies, including power curves, ASN curves, and stopping probabilities;
- Trellis graphics for powerful and effective comparison of designs;
- A complete set of sequential designs, including all those commonly used or offered in other software packages (for example, Lan & DeMets, Wang & Tsiatis, and Whitehead);
- Comprehensive evaluation tools, including conditional power, inference at the boundaries, and Bayesian analyses;
- Adjusted point estimates, p-values and confidence intervals based on sequential stopping rules;
- Full integration into S-PLUS, and extensible to fit your needs.

---

## **S+SpatialStats (2000) Version 1.5 Release 1**

---

S+SpatialStats is an object-oriented toolkit for analysis of spatial data.

S+SpatialStats offers a wide collection of tools, including:

- Variogram and correlogram calculation
- Fitting of theoretical variogram models
- Graphical exploration for geometric anisotropy
- Variogram clouds with interactive point identification
- Side by side boxplots of a variogram cloud
- Ordinary and Universal Kriging with prediction errors
- Point and block kriging predictions
- Finding of nearest neighbors of a point using a quad-tree
- Measures of Spatial Correlation such as Moran, Geary, as well as Monte Carlo estimates of their permutation distribution
- Determination of spatial neighbors of several orders and computation of corresponding spatial weight matrix
- Fitting of spatial linear (Generalized Least Squares) regression models for covariance families including: Conditional Spatial Autoregression (CAR), Simultaneous Spatial Autoregression (SAR), and Moving Averages (MA)
- Likelihood ratio test estimation for spatial linear models
- Sparse Matrix operations, including determinants
- Hexagonal binning for large 2-dimensional data
- Rayplots with confidence arcs
- Generation and analysis of spatial point pattern objects
- Empirical distribution functions of origin-to-point and point-to-point nearest neighbor distances
- Ripley's K-Function for spatial point patterns
- Kernel estimation of intensity for spatial point patterns
- Simulation of a spatial point process for distributions including: Poisson (homogeneous and non-homogeneous), Matern's Cluster, Strauss, and Strauss Sequential Inhibition
- Functions to compute Polygonal areas and membership
- Delaunay's Triangulations

---

## EnvironmentalStats for S-PLUS

---

What software do you use for your analyses of environmental data? Chances are, whatever it is, it doesn't have everything you want. Many people who deal with environmental data use a variety of software packages, including databases, statistical software, GIS software, spread sheets, and specialty packages such as groundwater or air plume modeling software.

Environmental data often present special problems such as data reported as "less than detection limit". Sometimes regulations require certain types of analyses which may or may not be appropriate.

ENVIRONMENTALSTATS for S-PLUS, created by Dr. Steven P. Millard of Probability, Statistics & Information (PSI), is a new *comprehensive* software package for Environmental Statistics. An add-on module to the statistical software package S-PLUS (sold by MathSoft, Inc.), it provides a set of powerful yet simple-to-use functions for performing graphical and statistical analyses of environmental data, bringing the major environmental statistical methods found in the literature and regulatory guidance documents into one statistical package, along with an extensive hypertext help system that explains what these methods do, how to use these methods, and where to find them in the environmental statistics literature. Also included are numerous built-in data sets from regulatory guidance documents and the environmental statistics literature.

ENVIRONMENTALSTATS for S-PLUS, combined with S+SPATIALSTATS and S-PLUS for ArcViewGis, provide the environmental scientist and researcher with a wide range of tools to "get the job done"!

### **What Can I Do With ENVIRONMENTALSTATS for S-PLUS?**

- Plot probability distributions so you can see how they change with the value of the distribution parameter(s).
- Estimates distribution parameters and compute confidence intervals for commonly used probability distributions.
- Plot the results of a Goodness-of-Fit test:
  - Observed and Fitted Distributions
  - Quantile-Quantile Plots
  - Results of Shaprio-Wilks, Kolmogorov-Smirnov, etc.
- Compute parametric and non-parametric prediction and tolerance intervals.
- Perform nonparametric estimation and tests for seasonal trend.
- Create power and sample size plots.
- Handle singly and multiply censored (less-than-detection-limit) data:
  - Quantile-Quantile Plots
  - Parameter Estimation
  - Confidence Intervals
  - Two-Sample Rank Tests



- Perform probabilistic risk assessment.
- Look up statistical methods in the environmental literature in a hypertext help system that explains the equations, links the equations to the original reference, includes abstracts of selected references, and contains a glossary of statistical and environmental terms.
- Reproduce specific examples in EPA guidance documents by using built-in data sets from these documents.

### ***Who Uses ENVIRONMENTALSTATS for S-PLUS?***

Anyone who has to make sense of environmental data, including hydrologists, soil scientists, atmospheric scientists, statisticians, geochemists, environmental engineers and consultants, hazardous and solid waste site managers, and regulatory agency analysts and enforcement officers. Parts of ENVIRONMENTALSTATS for S-PLUS incorporate statistical methods that have appeared in the environmental literature but are not commonly found in any statistical software package. Parts of ENVIRONMENTALSTATS for S-PLUS are specifically aimed at users who are

required to collect and analyze environmental monitoring data in order to comply with federal and state Superfund, RCRA, CERCLA, and Subtitle D regulations. All of the functions in ENVIRONMENTALSTATS for S-PLUS, however, are useful to anyone who needs to analyze environmental data.

### ***Would You Recommend ENVIRONMENTALSTATS for S-PLUS to Anyone Else?***

Emphatically YES! Anyone who uses S-PLUS will find all sorts of useful functions in ENVIRONMENTALSTATS for S-PLUS, including functions to plot probability and cumulative distribution functions, a standardized set of functions for estimating distribution parameters and quantiles and computing confidence intervals, new functions and methods for Goodness-of-Fit tests (including Shapiro-Wilk), functions to compute prediction and tolerance intervals, functions to estimate and test for trend, functions to easily create power and sample size plots, functions to handle Type I censored data, and functions to perform probabilistic risk assessment. Furthermore, the Glossary of Statistical Terms is a valuable instructional tool to anyone who uses S-PLUS to teach statistics.

### ***What's in ENVIRONMENTALSTATS for S-PLUS?***

Because ENVIRONMENTALSTATS for S-PLUS is an S-PLUS module, you automatically have access to all the features and functions of S-PLUS, including powerful graphics, standard hypothesis tests, and the flexibility of a programming language. In addition, specific features

of ENVIRONMENTALSTATS for S-PLUS include:

- **Probability Distributions**

Compute densities, probabilities, quantiles, and random numbers for the following distributions (underlined distributions are available only in ENVIRONMENTALSTATS for S-PLUS):

*Continuous Distributions*

Beta	<u>3-Parameter Lognormal</u>
Cauchy	<u>Mixture of Two Lognormals</u>
<u>Chi (square root of a chisquare)</u>	<u>Truncated Lognormal</u>
Chisquare	Normal
<u>Empirical</u>	<u>Mixture of Two Normals</u>
Exponential	<u>Truncated Normal</u>
<u>Extreme Value</u>	<u>Pareto</u>
<u>Generalized Extreme Value</u>	Stable
F (central and non-central)	Student's t (central and <u>non-central</u> )
Gamma	<u>Triangular</u>
Logistic	Uniform
Lognormal	Weibull

*Discrete Distributions*

Binomial	Negative Binomial
<u>Empirical</u>	Poisson
Geometric	Wilcoxon
Hypergeometric	

*Mixed Distributions*Zero-Modified Lognormal

(Also called the Delta Distribution; Lognormal with positive mass at 0),

Zero-Modified Normal (Normal with positive mass at 0)

- **Probability Density and Cumulative Distribution Plots**
- **Q-Q Plots for All Probability Distributions**  
Includes Standard Q-Q Plots and Tukey Mean-Difference Plots
- **Q-Q Plot Gestalt Function That Produces Numerous "Typical" Q-Q Plots for a Specified Distribution**  
Allows the User to Build Up a Visual Memory of "Typical" Q-Q Plots
- **Estimation of Distribution Parameters and Quantiles**  
Several Estimation Methods Available  
Maximum Likelihood, Minimum Variance Unbiased,  
Method of Moments, Method of L-Moments, etc.
- **Results Printed in "Nice" Format**  
Data Set Name, Sample Size, Method of Estimation, Optional Confidence Interval
- **Confidence Intervals for Distribution Parameters**

Binomial	exact, normal approximation
Exponential	exact
Extreme Value	normal approximation
Lognormal	exact (Land, 1971), Parkin et al.'s (1990) approximation, Cox's approximation (Land, 1972), normal approximation
Three-Parameter Lognormal	normal approximation, likelihood profile, zero-skewness (Royston, 1992b)
Normal	exact
Poisson	exact, Pearson-Hartley approximation, normal approximation
Zero-Modified Lognormal	normal approximation
Zero-Modified Normal	normal approximation

- **Confidence Intervals for Distribution Quantiles**

- Lognormal
- Normal
- Poisson
- Nonparametric

- **Goodness-of-Fit Tests**

- Chi-Square, Kolmogorov-Smirnov,  
Probability Plot Correlation Coefficient,  
Shapiro-Francia, Shapiro-Wilk
- Allow User to Estimate the Distribution Parameters
- Results Printed in "Nice" Format  
Data Set Name, Hypothesized Distribution,  
Estimated Parameters, Test Method
- Results Can Be Plotted. Optional Plots Include:  
Histogram with Overlaid Fitted Distribution,  
Q-Q Plot, CDF Plots of Observed and Fitted Distribution, Test Results

- **Prediction and Tolerance Intervals**

- Lognormal
- Normal
- Poisson
- Nonparametric

- **Methods for Type I Singly and Multiply Censored Data**
  - Empirical Cumulative Distribution Plots
  - Quantile-Quantile (Probability) Plots
  - Goodness-of-Fit Tests
  - Estimation, Hypothesis Testing, and Confidence Intervals
- **Special Hypothesis Tests**
  - Seasonal Kendall Test for Trend
  - Quantile Test (Detects Shifts in Tail of Distribution)
- **Sample Size and Power Calculations and Plots for Standard Confidence Intervals and Hypothesis Tests**
- **Tools for Probabilistic Risk Assessment**
  - Simple Random Sampling and Latin Hypercube Sampling
  - Generate Random Numbers from a Multivariate Normal Distribution
  - Generate a Multivariate Matrix from One or More Specified Distributions with a Specified Rank Correlation
  - Create an Output Distribution of Exposure or Risk
- **Built-In Data Sets**
  - Data Sets Appearing in Selected EPA Guidance Documents
  - Selected Data Sets from the Environmental Statistics Literature
- **Extensive Hypertext Help System**
  - Cross-Referenced Help Files that Clearly Explain Each Procedure and Provide Specific, Detailed Examples
  - Detailed Abstracts of Selected Literature in Environmental Statistics
  - A Fully Cross-Referenced, Hypertext Glossary of Statistical and Environmental Terms

## APPENDIX B INLEIDING IN DE S-PLUS COMMANDO TAAL

In deze bijlage staat op de S-PLUS commando taal centraal.

S-PLUS is een *object-geöriënteerde taal*. Dit houdt in dat er niets gedeclareerd hoeft te worden, maar ook dat het zonder meer mogelijk is reeds bestaande objecten te overschrijven. Bij het opstarten zijn er optioneel verschillende windows te openen, waaronder het commando-window. In deze appendix worden de in te typen opdrachten of S-PLUS functies **vet** weergegeven

*Toekenningen* gebeuren in het commando-window met < - , d.w.z. het < -teken (kleiner teken) en daaraan vast het minteken. Andere mogelijkheden om toekenningen te doen zijn

- "->", minteken en groter teken, dit is handig als men vergeten is op te geven de objectnaam op de geven waaraan de toekenning dient te gebeuren.
- "\_", het liggend streepje, Hierbij dient men op te letten dat de toekenning gebeurt aan het object links van het liggend streepje.
- assign** - commando. Deze toekenning werkt ook als men binnen een functie een toekenning wil doen naar een variabele buiten de functie.

Let bij toekenningen er op dat

- bestaande objecten niet ongewenst overschreven worden
- er geen namen van reeds bestaande functies gebruikt worden zoals de functies *c*, *t* en *q*. Gebeurt dit toch dan krijgt men een waarschuwing te zien.


Bij twijfel is dit gemakkelijk te controleren door gewoon de gewenste naam in te typen en een return/enter te geven. Als het object bestaat, krijgt men de inhoud te zien (ook als het een functie is) en anders de melding dat het object niet gevonden kan worden.

In de nu volgende S-PLUS sessie wordt e.e.a. kort gedemonstreerd.

Links staan (**vetgedrukt**) de opdrachten en na de : eventueel commentaar of vragen. Verdere uitleg staat soms ter onderbreking van een serie opdrachten over de hele paginabreedte.

### B.1 Het commando window

Start S-PLUS op via de NAL of via Start / Programs / S-PLUS 2000 (R3) / S-PLUS 2000 (zie ook 4.2)

Open, indien nodig, het commando-window via de volgende knop  of via het menu Window / Command Window.

- |                |   |  |
|----------------|---|--|
| <b>ls ()</b>   | : | deze functie toont de inhoud van <code>_Data</code> , of juist de inhoud van de huidige werkdirectory.     |
| <b>"Hallo"</b> | : | levert [1] Hallo<br>[1] betekent dat het eerstvolgende getal of string het eerste element in de vector is. |
| <b>Hallo</b>   | : | Error: Object "Hallo" not found<br>dus ...dit object bestaat niet.   |
| <b>1 : 50</b>  | : | levert een vector 1,2,...50 op. Let op de [.] indices links.   |

- a ← "Goedendag"** : de eerste toekenning
- ls()** : geeft nogmaals de inhoud van de werkdirectory.
- a** : geeft inhoud object "a"
- a ← c(a, "dit", "is", "begin")** : levert een vector van 4 strings op, kijk maar met de volgende opdracht.
- a** : met de functie **c** kunnen dus objecten aan elkaar worden "geplakt" of veranderd zoals in de volgende opdracht.
- a ← c(a[1:3], "het", a[4])** : de rij 1:3 wordt nu als index voor de vector a gebruikt.

**a**

In de nu volgende opdrachten worden aan object *a* allerlei verschillende soorten datastructuren toegekend. Merk overigens op dat spaties er niet toe doen in S-PLUS, behalve bij strings. S-PLUS onderscheidt wel hoofdletters van kleine letters.

- a ← 1 : 50** : toekenning van getallen
- a** : a is nu een numerieke vector
- a ← q** : toekenning van functie q
- a** : a is nu een functie
- a ← matrix (1:20, ncol=4)** : toekenning van een matrix
- a** : merk op dat er nu ook indices boven de matrix staan; [ ,1] betekent kolom 1, a[r,k] = het element [r,k] van a dat op rij r en kolom k staat.

Bedenk bij de volgende twee opdrachten wat er uit zal komen, en voer ze vervolgens uit.

- a ← matrix (1:20, ncol=5)**
- a ← matrix (1:20, ncol=6)**
- rm (a)** : object a wordt nu verwijderd

Zoals uit deze opdrachten blijkt, hoeft er niets gedeclareerd te worden in S-PLUS. Dit maakt dat er gemakkelijk gerekend kan worden en dat gebeurt vectorgewijs, zodat for-loops bijna nooit nodig zijn. Omdat S-PLUS een *interpreter*<sup>6</sup> is, is het aan te raden om loops te vermijden.

## **B.2 S-PLUS als rekenapparaat**

- a ← 4:0**
- b ← 6:10**
- a + b**

---

<sup>6</sup> Een *interpreter*, is een computertaal die steeds één commando regel inleest, de opdracht in computercode vertaalt en deze dan pas uitvoert. Omdat dit vertalen steeds weer opnieuw gebeurt, is dit langzamer dan wanneer het programma vooraf gecompileerd wordt, d.w.z. in machinecode omgezet wordt. Bij interactief gebruik merkt men hier echter niets van.

**a \* b**  
**b / a** : Inf = infinite, in wiskunde aangegeven met  $\infty$   
**Inf/Inf** : NA = missing value = NOT AVAILABLE  
**sin (pi \* a / 4 )**  
**log10 (10^a)**  
**log (0)** : ook met  $-\infty$  kan S-PLUS overweg

Het is trouwens eenvoudig om lege objecten aan te maken. Dit gaat als volgt:

**a ← numeric (0)**  
**a**  
**length(a)** : geeft de lengte van a aan.  
**a[2] ← 1**  
**a**  
**a ← character(0)**  
**a**  
**a ← c(letters[1:3],LETTERS[1:3]):** letters en LETTERS zijn bestaande S-PLUS objecten

Een belangrijk onderdeel bij het verwerken van gegevens is het selecteren van cases. Soms wil men een eigenschap voor alleen vrouwen onderzoeken of gegevens uit een deel van Nederland. Het maken van selecties wordt nu op eenvoudige wijze geoefend.

**a ← 4:0**  
**b ← 6:10**  
**a < b** : levert een logische vector op: T is true, F is false  
**a ← 1: 50**

Wat is het verschil tussen de volgende twee opdrachten ?

**a < 10**  
**a[a<10]**  
**a [a <35 & a > 20]** : levert meteen de selectie op.

Ook op resultaten van functies kan men direct weer logische expressies loslaten, zoals de volgende opdracht laat zien. Bedenk vooraf wat het resultaat zal zijn, **abs** is de functie absolute waarde, die van alle negatieve getallen positieve getallen maakt door "de min weg te laten".

**(-15:15)[ abs (-15 : 15) > 10 ]**

De volgende functies zijn handig bij het maken van vectoren. Met de functie **args** kunnen de argumenten van de functies bekeken worden.

**c , rep, seq**

**c -** is al aan de orde geweest: merk op dat de data van hetzelfde type moeten zijn, zoals uit de volgende 3 opdrachten blijkt.

**a ← c (letters [1:5], LETTERS [21:26])**  
**b ← c(10:1)**  
**c (a,b)**  
**args(rep)**

**rep** - repeteert een opgegeven serie x op 2 mogelijke manieren:

times = n keer of

length.out = n keer waarbij x cyclisch wordt gebruikt totdat lengte n bereikt is.

**rep (a,2)**

**rep (a,,5)**

**rep (a, length = 5)**

: S-PLUS accepteert zowel de parameter naam als de volgorde in de parameterlijst aangegeven via komma's. Merk tevens op dat de officiële parameter naam length.out is maar dat length al voldoet,

**rep (c(3,5,7), 3)**

**rep (c(3,5,7), length = 10)**

**rep(c(2, rep(1,3)), 2)**

**args (seq)**

**seq** - hiermee kunnen ook vectoren worden gemaakt, de argumenten zijn:

from : startwaarde

to : eindwaarde

by : afstand tussen de waarden onderling

length.out: lengte van de vector

along : een ander object

**seq (1, 10, 1)** : dit is 1:10

**seq (-1, 1, 0.01)**

**seq (2, by = 2, length = 10)** : een rijtje even getallen

**seq (along = a)** : dit is gelijk aan 1: length (a)

**seq (a)**

**x ← rep (c(seq(-5, 5, 1), rep (6, 5), NA), 2)**

**x**

Hier staan nog een aantal functies die niet behandeld worden, maar die zeker wel van pas zullen komen. Vervolgens komen er nog wat oefeningen om zelf te maken.

sort	unique	mean	ceiling	summary
rev	duplicated	var	floor	order
rank	min	cor	trunc	
length	max	sum	round	help
match	quantile	cumsum	signif	args

Oefening<sup>7</sup> 1: Data van de New Jersey Pick-it lottery zijn als systeem data beschikbaar in S-PLUS, onder de objectnamen *lottery.number* en *lottery.payoff*.

- Welke loten hebben een uitbetaling van meer dan \$500?
- Lottery.number bevat alle winnende nummers. Hoeveel zijn het er en hoeveel is er uitbetaald?
- Zou je met één opdracht m.b.v. **summary**, **mean** en **var** kunnen onderzoeken of de uitbetalingen normaal verdeeld zijn?
- Controleer dit met de grafische functie **hist**.

<sup>7</sup> Deze oefeningen zijn speciaal bedoeld voor diegenen die de commando taal van S-PLUS nog niet kennen.



- e. Zoek de 10 laagste uitbetalingen met de bijbehorende nummers. Gebruik hierbij de functie **order**.

Oefening 2: Op de vorige pagina is een object *x* aangemaakt met missing values

- Hoeveel missing values zijn er, gebruik **is.na ()** en **sum ()**?
- Op welke plaatsen zitten deze missing values?
- Wat levert **mean (x)** op en hoe kun je het gemiddelde uitrekenen van alle niet missing values m.b.v. **is.na** en het **uitroepteken** (dat betekent het "logische "niet").
- Bekijk met **args** de functie **mean** en bereken nogmaals het gemiddelde.

Oefening 3: Trek met de functie **rnorm** 1000 getallen uit  $N(5, 100)$ .

Doe hetzelfde met **runif** uit uniform (0,1).

- Onderzoek of de resultaten kloppen.
- Zijn het minimum en maximum van de uniforme verdeling min of meer volgens de theorie?

### **B.3 De functie "options" en het gebruik van een editor**

Nu worden in het kort de functie **options** en de on-line editor besproken.

De functie **options** geeft een lijst van S-PLUS instellingen. Elke optie heeft een **\$optienaam** met vervolgens de waarde van **\$optienaam**.

<b>options ()</b>	:	levert de volledige lijst
<b>options ()\$width</b>	:	levert de waarde van de ingestelde scherm breedte
<b>options ()\$length</b>	:	levert de waarde van de ingestelde scherm lengte

De opties zijn te wijzigen met

**options (width= $n_1$ , length = $n_2$ )** : waarbij de gebruiker getallen voor  $n_1$  en  $n_2$  opgeeft

Het aanpassen van deze instellingen kan nodig zijn om het commando-scherm zo optimaal te benutten.

<b>options()\$editor</b>	:	geeft aan welke editor wordt gebruikt standaard: <i>notepad</i> .
--------------------------	---	--

Tot zover enkele opties, de andere komen vanzelf aan de orde. Merk op dat uit de laatste opdracht blijkt dat notepad wordt gebruikt. De editor wordt voornamelijk gebruikt voor het schrijven van functies. In het nu volgende voorbeeld wordt een numerieke vector gemaakt via notepad.

<b>y ← numeric (0)</b>	:	maak numerieke lege vector <i>y</i> aan
<b>fix (y)</b>	:	nu wordt een nieuw window met notepad geopend.

Voor numerieke data kan men achter elkaar, door een spatie gescheiden, de gewenste getallen intypen.

Voor character data dient men elke string op een nieuwe regel te beginnen. De quotes zijn dan niet nodig.

Oefening: type een aantal plaatsnamen in.

Men sluit het editen af door het bestand te bewaren en het window te verlaten. Men kan dit in een keer doen met de toetscombinatie **Alt-F4**.

Het intypen van een matrix gaat wat moeilijker, in die zin dat het niet rechtstreeks als matrix kan.

```
y ← matrix (rep(1,9), ncol=3)
fix(y)
```

Dit levert iets op waar men niet snel mee uit de voeten kan. Gelukkig is met de volgende opdracht van een vector snel een matrix te maken.

Oefening: Maak via de editor een 3x3 matrix door in een vector ter lengte 9 rij na rij de matrix te typen en verlaat dan de editor. Type vervolgens:

```
naammatrix ← matrix (naamvector, ncol=3, byrow=T)
```

Men kan dus in de editor de matrix intypen d.w.z. 3 rijen van 3 kolommen.

Oefening: doe het bovenstaande in één regel d.w.z. naammatrix ←...

Oefening voor de fijnproevers:

**rbind** is een functie die twee matrices met hetzelfde aantal kolommen rij-gewijs aan elkaar plakt, zeg **rbind (x,y)**. Hoe kan men in één regel aan naammatrix via de editor nog 2 rijen van 3 kolommen toevoegen?

#### **B.4 Gebruik grafisch scherm**

Nu wordt nader ingegaan op de grafische mogelijkheden van S-PLUS, waarbij onder men zelf eerst een grafisch scherm opent via de S-PLUS opdracht **graphsheet**. Met een aantal voorbeelden en oefeningen komen diverse functies aan bod, maar het zal niet mogelijk zijn in een kort tijdsbestek alles te behandelen. Indien nodig doe eerst:

```
library("RIVMfun",first=T,lib.loc="O:/milieu/applics/splus 2000/RIVM")
```

of de analoge aankoppeling zoals aangegeven voor thuisgebruik. Type vervolgens

```
graphsheet()      : hiermee wordt een grafisch scherm geopend. Zet het
                  : naast het commando scherm
usa()             : levert de kaart van de USA.
NL()              : Levert een keuze menu voor de kaart van Nederland.
```

Het inkleuren van de provincies.

```
ls("nl.*",2)     : laat als het goed is alle bestanden in RIVMfun zien die
                  : met nl. beginnen. Mocht deze opdracht niet werken
                  : doe dan:
search()         : dit geeft de lijst met aangekoppelde S-PLUS
                  : directories. Als RIVMfun daar op een andere positie
                  : staat, vervang de 2 in de vorige opdracht dan door de
                  : juiste index.
polygon(nl.brab) : maakt de provincie Noord-Brabant zwart.
polygon(nl.over,col=2) : kleurt de provincie Overijssel met een andere kleur.
polygon(nl.gron,angle=45,density=30,col=3) : geeft weer meer mogelijkheden van inkleuren.
```

In een eerdere opgave zijn 2 vectoren gemaakt met normaal en uniform verdeelde grootheden. Met de functie **summary** is op primitieve wijze onderzocht of de vectoren min of meer de juiste verdeling bezaten. Dit gaat grafisch veel beter:

Oefening: Onderzoek de gesimuleerde data m.b.v. de functies **hist** , **boxplot** , **density** en **qqnorm**.

Oefening:

```
x ← 1 : 100
y ← x ^ 2 - 1000
plot (x,y)
plot (x,y, type ="1")
plot (x,y, type ="n", main = "Er komt wel wat")
lines (x,y)
abline (h=0, lty=2)
points (Z←seq(1,100,10), Z^2-1000*Z+Z,pch="o")
```

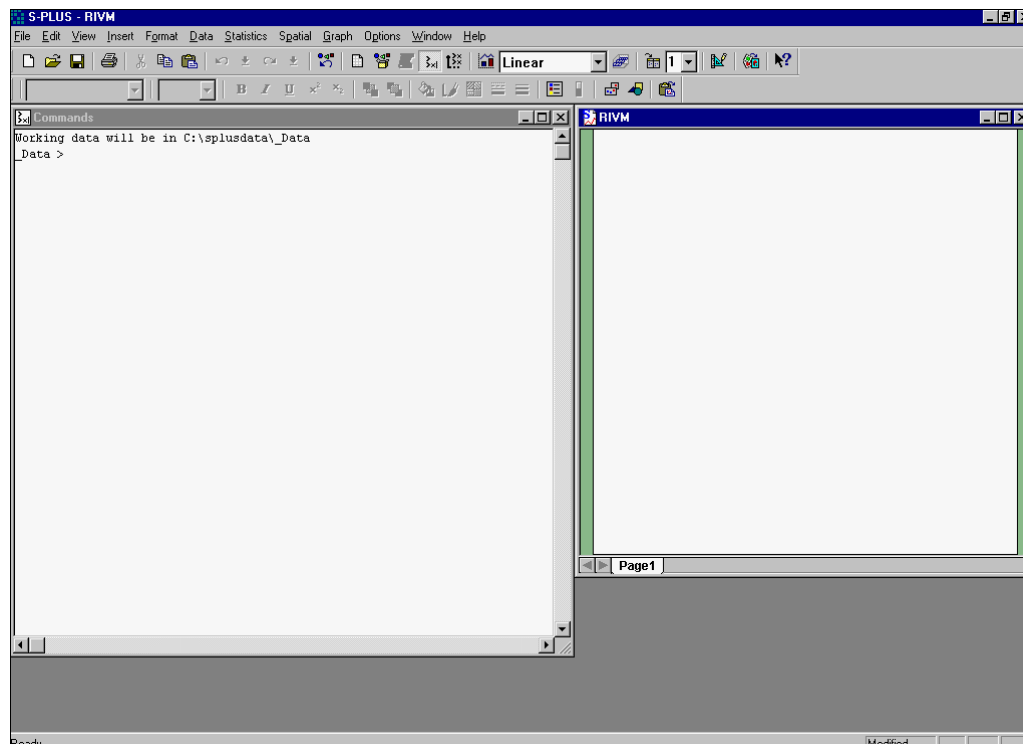
Oefening: plot de functie  $f(x) = (x-50)^3$  met als hoofdtitel: "Dit is een 3e graads fie" en als subtitel: " $f(x) = (x-50)^3$ ". Teken de raaklijnen voor  $x=25$  en  $x=75$

Oefening: Onderzoek of de winnende nummers in *lottery-number* wel eerlijk over alle nummers verdeeld zijn. Evenzo voor de uitbetalingen.

In de functie **.First** wordt ook een grafisch scherm geopend. Dit gebeurt met de volgende twee commando's. Deze geven een aantal mogelijkheden van **graphsheet** aan.

```
if(dev.cur() == 1) {
  guiSetOption(option.name = ImageColorTableStd",
    value.string = get(paste("ImageColorTable.", ncolors, sep = "")))
  graphsheet(18, 18, "normal", 0.45, 0.75, 0.55, 0, 12, "portrait", 2.54
    , pages = "within expression", color.scheme = "Standard",
    image.color.scheme = "Standard", Name = "RIVM")
}
```

Met deze opdrachten en de juiste grootte van het commando scherm krijgt men standaard de volgende schermindeling bij het openen van S-PLUS.



## B.5 Verschillende DATASTRUCTUREN in S-PLUS

Nu worden systematisch enkele datastructuren binnen S-PLUS behandeld en een aantal functies die daarmee verband houden. Doordat S-PLUS object-geïënteerd is, kan één functie afhankelijk van het datatype verschillend reageren. Een heel duidelijk voorbeeld hiervan is de functie **summary**.

### B.5.1 Vectoren en factors

Vectoren zijn "atomic" structuren d.w.z. de waarden die ze bevatten zijn allen van hetzelfde type ook "mode" genoemd in S-PLUS. Een aantal types zijn inmiddels de revue gepasseerd. Het complete lijstje is:

```
logical   -   T of F (≡ 1 resp. 0 als men ermee rekent)
numeric  -   integer of real
complex
character
```

De enige uitzondering op de regel dat de waarden van hetzelfde type moeten zijn, is met NA-values, waarden die niet beschikbaar zijn. Een vector kan ook leeg zijn en toch al een mode hebben.

```
zz ← character (0)
mode (zz)           }
                    } de functies mode en length geven de betreffende
length (zz)         } attributen-info over een object
zz[3] ← "c"         } Wat houdt deze berekening in voor de length en de
                    } waarden van zz?

length (zz)
zz
```

Een derde attribuut is **class**. De "class" van een object zorgt ervoor dat allerlei functies verschillend reageren op verschillende classes. Hieronder volgt een eenvoudig voorbeeld van een tijdreeks in de S-PLUS datasets. In feite wordt door de naam van een object interactief te typen de print functie gebruikt die verschillend reageert op verschillende "classes":

```
ship
ship [1:24]
attributes (ship)      : laat de attributen van ship zien
```

### Factors en Categories

Een category is een vector-object dat gebruikt wordt om een discrete classificatie van de waarden van een vector te specificeren.

Een factor doet dit ook maar is van de klasse factor, wat inhoudt dat functies als **aoV** etc. speciaal hierop reageren.

```
prov ← c("nhol", "zhol", "utre", "gron", "dren", "over", "utre", "gron", "gron", "geld",
"geld")
prov.cat ← category (prov)
prov.cat
prov.fac ← factor (prov)
prov.fac
unclass (prov.fac)
```

**attributes(prov.fac)**  
**attr (prov.fac, "levels")** : de functie **attr** geeft de mogelijkheid om attributen op te vragen. De functie *attributes* geeft alle attributen van een object.

Een belangrijke functie m.b.t. factoren is de functie **tapply**. Onderstaand voorbeeld laat zien hoe men in één keer via **tapply** de groepsgemiddelden of wat dan ook van de verschillende groepen kan berekenen.

```
inkomen ← c(20, 25, 18, 15, 17, 17, 12, 20, 17.5, 10, 20)
tapply (inkomen, prov.fac, mean)
tapply (inkomen, prov.fac, sum)
```

De functie **tapply** werkt ook met meerdere factoren tegelijkertijd.

### B.5.2 Matrices en arrays

Een matrix is een 2 dimensionaal array. We zullen ons voor de eenvoud tot matrices beperken.

```
mat ← matrix (1:20, ncol=5) : aanmaak array
mat
mat [,1] : 1e kolom geselecteerd
mat[3, ] : 3e rij geselecteerd
mat [4,5] : element 4e rij, 5e kolom
i ← matrix(c(1:3, 3:1), ncol=2,byrow=F)
i : indexarray
mat[i] : selecteer de elementen via index array
      bedenk van tevoren het antwoord van mat[i]
mat [i] ← 0 : en ken er 0 aan toe
mat
```

Voor verdere info: zie de on-line manual *Programmer's Guide*.

### B.5.3 Lists

De functie **options()** heeft een list als resultaat. Een list is een object dat bestaat uit een geordende verzameling van componenten die van verschillende types kunnen zijn. Een list wordt net als functies en expressies een "recursive structure" genoemd i.p.v. "atomic structure" zoals vectoren en matrices. De functie **list** wordt gebruikt om een list te maken:

```
alles ← list(char = character (0), : Na een komma kan gewoon een enter gegeven
      nume = numeric (0), : worden. De prompt wordt dan een + .
      logi = logical (0),
      mat = matrix (1:20, ncol=10)
alles
attributes (alles)
alles $ mat [,3] : via $ naam kunnen de diverse componenten
alles $ char ← c("a", "b", "c") : benaderd worden
alles $ nume ← rep(1,5)
alles $ logi ← alles $ mat > 12
```

```
alles [[1]] : ook via alles [[indexnummer]] kun je
alles [[1]] [3] : component indexnummer bereiken. Merk op dat ook
      hier
alles [[4]] [1] : spaties er niet toe doen.
```

**alles \$ toegevoegd ← "klaar"**      het toevoegen van nieuwe componenten is nu eenvoudig

**mode (alles)**  
**length (alles)**

Met de **c** functie (concatenate) kunnen lists aan elkaar geplakt worden. Functies reageren, zoals eerder gezegd, verschillend op verschillende classes. Bijvoorbeeld **plot** en **lines** hebben o.a. x en y als variabelen. Bij een list met een x en y component is de objectnaam zelf al voldoende zoals het volgende laat zien:

```
alles $x ← 1:100
alles $ y ← (alles $x^2)/100 + sin (alles $ x)
plot (alles, type = "l")
point (alles)
```

Tenslotte zijn er veel functies die als argument een list gebruiken. Het volgende voorbeeld waarin namen aan de kolommen en rijen van een matrix worden gegeven, is daarvan een voorbeeld.

```
mat ← matrix (1:15, ncol =3)
dimnames (mat) ← list (letters [1:5], c("een", "twee", "drie"))
mat
dimnames (mat) ← list (NULL, c("een", "twee", "drie"))
mat
dimnames (mat) ← list (rep(" ",5), c("een", "twee", "drie"))
mat
```

#### **B.5.4 Data.frames**

Een data.frame is een list van de klasse data.frame. Er zijn enkele restricties om van een list een data.frame te kunnen maken.

- \* De componenten van de list dienen vectoren te zijn (numeriek, character of logisch), factoren, numerieke matrices, andere data.frames, of ook weer lists die aan de juiste voorwaarden voldoen.
- \* Matrices, lists en data.frames voegen net zoveel variabelen aan de data.frame toe als ze resp. kolommen, elementen of variabelen hebben.
- \* Numerieke vectoren en factoren worden als zodanig opgenomen; niet numerieke vectoren worden tot factoren omgevormd met als levels de unieke waarden aanwezig in de vector.
- \* Alle vectoren die als variabelen in de data.frame worden opgenomen, moeten dezelfde lengte hebben en de matrices hetzelfde aantal rijen.

In veel opzichten kan een data.frame als een matrix beschouwd worden waarvan de kolommen uit verschillende modes of attributen mogen bestaan. Een data.frame kan als een matrix geprint worden en rijen en kolommen kunnen geselecteerd worden zoals bij matrices.

Een belangrijk verschil met matrices is dat de kolommen en rijen van data.frames géén gelijke labels mogen hebben!

De **attach** functie is reeds gebruikt om S-PLUS directories aan te koppelen. Deze functie kan ook gebruikt worden om de *data.frames* aan te koppelen zodat men dan verder kan rekenen met de variabelen, voor zover de variabelen niet in de werkdirectory voorkomen. Dit wordt in onderstaande oefening onder de aandacht gebracht.

Data.frames zijn vooral aan te bevelen voor o.a. regressie-analyse en variantie analyse, omdat de regressieformules dan eenvoudig aan te geven zijn.

```
z ← c(rnorm(10,mean=2), rnorm(10), rnorm (10,-2))
```

```
z.let ← c(rep("a",10), rep("b",10), rep("c",10))
```

```
z.fac ← factor (z.let)
```

```
z.data ← data.frame (lft=z, fct=z.fac, nrm =z)
```

```
z.data
```

Merk op dat **z.let** tot dezelfde factor in data.frame wordt omgezet als de functie **factor** dat doet. Nu volgt een zelfde selectie van de laatste 2 kolommen als bij een matrix en vervolgens worden de functies **summary** (die werkt nu per kolom), **plot.factor** en **aov** (anova-analyse) toegepast.

```
z.data ← z.data [,2:3]
```

```
summary (z.data)
```

```
summary (z.data [,2])
```

```
plot.factor (z.data)
```

```
aov (nrm ~ fct , z.data)
```

Tenslotte nog een korte oefening met *attach* en *detach* m.b.t. een data.frame.

```
x1 ← rep (1,10)
```

```
x2 ← rep (10,10)
```

```
x.data ← data.frame (x1, x2)
```

```
x.data
```

```
attach (x.data)           : merk op dat er géén quotes om de objectnaam staan
```

```
search()                 : x.data staat nu op de tweede plaats
```

```
x1 ← x1 + x2/10
```

```
x1                       : dit gebeurt niet in de werkdirectory!
```

```
x.data                   : dit wel
```

```
detach (1)
```

```
x.data
```

Het werken met directories en data.frames wordt in het volgende hoofdstuk verder uitgelegd.

### **B.5.5 Tijdreeksen**

Heel kort iets over tijdreeksen. Tijdreeksen zijn van mode "numeric". Het aardige van tijdreeksen is dat de lay-out van het printen en het plotten automatisch een tijdreeks geeft.

Oefening: Maak met de functie **ts** een tijdreeks van **(1:120) + rnorm (120,,10)** die maandgegevens bevat vanaf 1980 maand juni.

Plot deze reeks met **ts.plot**.

Plot dezelfde reeks voor 1985 t/m 1989.

Voor verdere functies m.b.t. tijdreeksen kan men in de on-line help kijken bij de categorie Time Series.

Een overzicht van alle type objecten in S-PLUS kan eenvoudig verkregen worden door in de S-PLUS Help onder *Topic objects* te typen.

## B.6 Functies in S-PLUS

De reproduceerbaarheid van het onderzoek wordt in S-PLUS gewaarborgd door functies. Deze functies kunnen resultaten als volgt vastleggen

1. rechtstreeks naar het scherm
2. in ASCII bestanden, die vanuit het programma aangemaakt worden
3. naar grafische windows
4. in grafische bestanden (o.a. postscript, WMF of BMP bestanden) met grafieken
5. in nieuwe S-PLUS objecten, waarmee dan weer verder gerekend kan worden.

In het nu volgende wordt stapsgewijs ingegaan op het maken van functies. Doel is het dusdanig generiek maken van functies dat een functie zowel interactief (zonder voorkennis) gebruikt kan worden alsook dat de functie met opgegeven parameters in weer een andere functie gebruikt kan worden, zonder verdere interactie.

### B.6.1 Het maken van functies

Het schrijven van functies is vrij eenvoudig en gebeurt het veiligst via een editor zodat fouten achteraf gecorrigeerd kunnen worden. Merk op dat S-PLUS een ingebouwd programma heeft dat na het verlaten van de editor de ingetypte tekst van een functie volgens een vastgelegde standaard opmaakt en controleert op syntaxfouten. Indien er fouten inzitten wordt dit gemeld en worden de wijzigingen wel in het functie-object bewaard, maar er wordt geen lay-out opgemaakt. Met de opdracht **fix()** -zonder argument !- kan men dan verbeteringen aanbrengen.

Omdat zowel gegevens, figuren als functies objecten zijn, is het handig om een functienaam altijd met **f.** te laten beginnen. Met **ls("f.\*")** kan men dan meteen een overzicht krijgen van alle beschikbare functies in de werkdirectory.

**fix(f.som)** : **f.som** is nu van het type functie en voor de rest leeg.  
Tussen de haakjes dient de functienaam te komen.

Maak volgens onderstaande opzet de functie **f.som** en verlaat dan de editor.

```
function (x,y)
{
#-----
# Deze functie telt 2 getallen of vectoren op
#
# Met een hekje kan men commentaar tussenvoegen
# en door het commentaar in te kaderen wordt het
# programma overzichtelijk.
#-----
                                x + y
}
```

Ga na of deze functie werkt voor zowel getallen als vectoren. De uitvoer  $x+y$  wordt op het scherm afgedrukt omdat het de laatste toekenning is. Wil men ook de invoer op het scherm, dan dient men deze te laten afdrukken via de opdracht **print**.

**fix()** : automatisch wordt f.som weer in de editor geladen als laatst gebruikte functie door de editor.

```
function (x, y)
{
#-----
# Deze functie telt 2 getallen of vectoren op
```



```

#
# Met een hekje kan men commentaar tussenvoegen
# en door het commentaar in te kaderen wordt het
# programma overzichtelijk.
#-----
#met cat kunnen teksten worden afgedrukt
cat ("\n f.som : uw steun en toeverlaat bij sommeren\n\n")
cat (paste("De som van ", x, "en", y, "is", x+y, "\n"))
print(x) # als men interactief werkt is het voldoende om x in te typen
print(y) # en de inhoud van x komt op het scherm.
# Binnen functies dient men het commando print te gebruiken
print(x+y)
return(invisible())          #
#deze regel betekent dat de functie niets uitvoert als resultaat
}

```

Oefening: werk een paar keer met **f.som**. Wat gebeurt er als de laatste regel in bovenstaande versie weggelaten wordt ?

Wil men het resultaat van  $x+y$  aan een nieuw object toekennen, dan dient  $x+y$  expliciet uitgevoerd te worden. Dit gebeurt door de laatste regel te wijzigen in **return(x+y)** De toekenning gebeurt via de bekende toekenningen

**res ← f.som(1:3,4:6)**

Men kan ook **defaultwaarden** instellen. Dit gaat als volgt.

Oefening:

1. Verander de eerste regel in function ( $x = 0$ ,  $y = 0$ ), waarbij nul nu de defaultwaarde is.
2. Verlaat de editor en type `f.som()`

Het kan echter nog mooier. Hoe is deze "ingewikkelde" functie eenvoudig door een leek te gebruiken? Juist, door er een volledig interactief gestuurde functie van te maken. De werkwijze daarvoor staat in de volgende oefening beschreven.

Oefening:

1. Voeg na de eerste opdracht `cat` het volgende toe
 

```

if(x=0 & y=0)
{x ← eval (parse(prompt = "Geef x s.v.p. "))
y ← eval (parse(prompt = "Geef y s.v.p. "))}

```
2. Verlaat de editor en ga verder met:
 

```

f.som()
f.som(0,i)

```
3. Edit de functie en maak er een syntax fout in. Verlaat dan de editor en los het op.

De huidige versie van **f.som** voldoet eigenlijk nog niet helemaal, omdat **f.som()** het resultaat 0 geeft, terwijl er eigenlijk niets wordt opgeteld. Wijzig daarom én de eerste regel én het **if** statement in resp.

**function(x=numeric(0),y=numeric(0)**

**if(length(x)=0 & length(y)=0) :** merk het verschil tussen = en = =  
 = wordt gebruikt om defaultwaarden in te stellen  
 = = om te checken of een variabele gelijk is aan iets

### B.6.2 Menu's

Met de volgende oefening wordt het programma verder menugestuurd gemaakt. De geproduceerde functie **f.som** is daarnaast ook veel te eenvoudig voor een onderzoeker. Vandaar de volgende oefening met als doel **f.reken**, waarmee de vier basisbewerkingen, optellen aftrekken, delen en vermenigvuldigen, uitgevoerd kunnen worden.

**f.reken** ← **f.som**  
**fix(f.reken)** : en vervang de 2e cat door

```
cat ("\n Welke berekening wil je? \n")
ans ← menu(c("optellen", "aftrekken", "vermenigvuldigen", "delen"))
switch(ans,
  {res ← x+y
  cat(paste("\nf.reken : ", x, "+", y, "=", res, "\n"))},
  {res ← x-y
  cat(paste("\n f.reken :", x, "-", y, "=", res, "\n"))},
  {res ← x*y
  cat(paste("\n f.reken :", x, "*", y, "=", res, "\n"))},
  {res ← x/y
  cat(paste("\n f.reken :", x, "/", y, "=", res, "\n"))})
```

Oefening : Schrijf nu zelf een nieuwe functie, waarin **f.reken** wordt aangeroepen. Geef daarbij in de nieuwe functie al x en y waarden op. Het vervelende is nu dat **f.reken** telkens vraagt welke bewerking er gedaan dient te worden. Los dit op door **f.reken** te wijzigen met een extra variabele **bew** van bewerking.

Oefening : Het menu kan ook in een window verschijnen. Daarvoor dient het grafisch device geopend te zijn. Probeer dit uit via **args(menu)**.

## APPENDIX C INLEIDING IN HET GEBRUIK VAN DE GUI

In deze bijlage worden enkele oefeningen gegeven in het gebruik van de S-PLUS menu's. Een aantal veel gebruikte handelingen komt hierbij aan de orde. Het importeren van data, het selecteren van cases en analyseren van de gegevens kan voor een deel gemakkelijk met de menu's in S-PLUS. Aan de hand van een ASCII databestand wordt e.e.a. gepresenteerd. In de daarop volgende oefening dient men zelf weer andere data binnen te halen. Als geheugensteun volgen nu een aantal menu's die gebruikt gaan worden. Het te importeren bestand heet **knmi.txt** en staat in **o:/milieu/applics/splus 2000/RIVM/TXT**

**Klik op de Object Browser knop**  dit is de Filemanager van S-PLUS en controleer of het bestand **knmi.txt** er al staat (zie Map Data).


Als het goed is dient het nog geïmporteerd te worden, via het menu systeem. Het is echter raadzaam om **knmi.txt** eerst met notepad of WORD te bekijken. Bij het importeren dient men namelijk op te geven of er in de eerste regel de kolomnamen staan, of de TAB (\t) het kolomscheidingsteken is of een spatie etc. Ga vervolgens in S-PLUS naar

**File / Import Data / From file ...** : Let op tabblad OPTIONS !

Er wordt nu een spreadsheet geopend met 7 kolommen, waarvan de eerste twee de x- en y coördinaten bevatten van de KNMI meetpunten. De overige meetpunten bevatten de totale neerslag per aangegeven jaar.

**Klik met linkermuisknop op x en y en houd Shift ingedrukt** : Selecteer x en y kolom (om grafiek te maken) : gebruik Shift en Ctrl toets is als in Excel en WORD.

Klik op 2D knop 

kies linksboven knop  : de figuur geeft een indruk van Nederland !

Een andere manier om dit te doen, gaat via de menu-balk.

**Selecteer nogmaals x en y kolom  
Graph / 2D Menu / Scatter**

Omdat er een handige functie is in de RIVM S-PLUS bibliotheek RIVMfun wordt deze bibliotheek, indien nog niet beschikbaar, aangekoppeld via het commando

**library("RIVMfun",first=T,lib.loc="O:/milieu/applics/splus 2000/RIVM")**

Geef nu de volgende opdrachten

**f.gs** : maakt graphsheet in handig formaat aan  
**NL(1)** : tekent kaart van Nederland [deze functie is ook te gebruiken als NL()] waarop een eenvoudig menu verschijnt.  
**points(knmi)** : tekent de punten.

De ligging van de meetpunten is misschien interessant. Maar veel interessanter is de verdeling van de meetwaarden over Nederland. Een eerste manier om deze te bekijken is via 3D plots. Vervolgens wordt de functie `f.image.NL` gepresenteerd om een beter inzicht in de ruimtelijke verdeling van de meetwaarden te krijgen. Verder komen aan bod onder menu statistics: **summary** en de lineaire regressie om een eventuele trend uit de gegevens te halen.

**Selecteer in knmi weer x, y, en jr93 en maak een duidelijke 3D plot via de 3D knop** 

**f.image.NL(knmi)**

**args(f.image.NL)** : Laat de optionele argumenten van de functie zien.  
 Omdat deze niet voor zich spreken, komt hier een eerste deel van de functie tekst.  
**f.image.NL** : toont de inhoud van het functie-object

```
function(x, y = numeric(0), z = numeric(0), type = "original", xo = numeric(0), yo
= numeric(0), mn = "", sb = "", cex = 1, zmin = numeric(0), zmax = numeric(0),
block = F, blocksize = 5000)
{
#-----
# f.image.NL :
#-----
#
# function for making image plot of the Netherlands
#
# x = x-coordinates in Nieuwe Amersfoortse coordinates in (m) or (km)
# y = y-coordinates in Nieuwe Amersfoortse coordinates in (m) or (km)
# z = values at (x,y)
#
# type = "interp"
#       uses interp on xo and yo if omitted
#
#       "original" (DEFAULT VALUE)
#       puts the z-values on a given grid, where the sum of
#       z-values is taken when more z-values at the same gridcell
#       Default grid is the grid of LLO
#
# OPTIONAL ARGUMENTS
#
# xo      = ascending sequence of x-coord. for the grid (see interp)
# yo      = ascending sequence of y-coord. for the grid (see interp)
#
# mn      = main title
# sb      = sub title
# cex     = cex (height of characters)
# zmin    = minimum value of z-values to be used (for scaling)
# zmax    = maximum value of z-values to be used (for scaling)
# block   = Boolean block = T designs squared boxes around blocks
#          DEFAULT VALUE block=F
# blocksize = blocksize in meters DEFAULT VALUE blocksize=5000
```

**f.image.NL(knmi,block=T,mn="knmi 1993")**:levert een nog wat duidelijke beeld op

#### statistics / data summaries / summary statistics

**kies knmi en de jr93 kolom** : maak een summary, is een log-transformatie op zijn plaats? zo ja, voer deze uit en maak nogmaals de summary.

**statistics / regression / linear** : lineaire regressie van jr93 op x en y en van log(jr93) op x en y. Bewaar de residuen in knmi.res en voeg met de volgende opdrachten de x- en y-coördinaten toe.

#### open knmi.res / selecteer residuals /

**klik rechts / insert column / x & knmi\$x / apply**  
**idem voor kolom y**

**Maak nu weer een 3D-plot en een f.image.plot van de residuen.**

Voor veel meer informatie over het gebruik van de GUI: zie het boekje

*An introduction to S-PLUS for Windows*

Longhow Lam

Candiensten, 1999

## APPENDIX D BEGINNEN MET S-PLUS VOOR SPSS GEBRUIKERS

Deze appendix is geschreven door Hans de Wit ([h.de.wit@sfk.nl](mailto:h.de.wit@sfk.nl)) toenmalig medewerker van het Sociaal Cultureel Planbureau, die zowel SPSS als S-PLUS gebruikte. Nu werkzaam bij de Stichting Farmaceutische Kengetallen, Postbus 30460, 2500 GL DEN HAAG.

De onderwerpen die aan de orde komen zijn

- |                            |                       |                 |                    |
|----------------------------|-----------------------|-----------------|--------------------|
| 1. Inleiding               | 6. display labels     | 11. recode      | 16. crosstabs      |
| 2. Bestanden en variabelen | 7. compute            | 12. if en do if | 17. add files      |
| 3. Naar een data-frame     | 8. (add) value labels | 13. rank ntiles | 18. match file     |
| 4. get file                | 9. variable labels    | 14. count       | 19. missing values |
| 5. save file               | 10. rename            | 15. frequencies |                    |

### D.1 Inleiding

S-PLUS heeft erg veel aan boord. De documentatie is overweldigend en er is nog van alles bijgeschreven om het nog weer gemakkelijker en beter te maken (zie hoofdstuk 8 en E).

S-PLUS heeft een GUI, een grafische gebruikersinterface. Al klikkend kom je een heel eind. Toch is dat niet helemaal wat je wilt. Vooral in de datamanipulatie-fase kom je daarmee te kort. SPSS blinkt uit in de eenvoud waarmee je gegevens kunt veranderen. Dit document probeert het SPSS idioom te vertalen in S-PLUS.

Naar mijn mening moet het manipuleren van data met een programma, zodat je achteraf nog kunt zien wat er allemaal is gebeurd. Ik ga er in dit document van uit dat je aan de slag wilt met S-PLUS en alleen de commando's wilt weten die je in SPSS altijd gebruikte. De commando's van SPSS en S-PLUS binnen dit document zal ik weergeven in een proportioneel lettertype. Al het moois dat S-PLUS biedt gaat grotendeels aan dit document voorbij. Daar zijn echter al die boeken voor die ik al heb genoemd.

Twee tips voor je met commando's aan de slag gaat:

- S-PLUS is gevoelig voor het gebruik van hoofd- en kleine letters: case-sensitive. Als je begint maak je erg veel fouten die daartoe te herleiden zijn. Ga dus eerst kijken of je de variabele naam of de aanroep van een functie wel goed geschreven hebt!
- Van elke functie zijn de benodigde argumenten te achterhalen met de functie `args()`. Voorbeeld: `args(lm)`. Met `?lm` kun je hulp krijgen over de functie `lm`, en met `lm` zie je de programmatuur van de functie. Dit laatste geldt niet alleen voor functies, maar voor alles wat je tegenkomt. Geeft de naam in als commando en S-PLUS kiest iets passends om weer te geven.

### D.2 Bestanden en variabelen

SPSS bestanden zijn georganiseerd in systeemfiles, S-PLUS bestanden in data-frames. In een data-frame zitten 1 of meer variabelen. Elk van de variabelen heeft, net als in SPSS, hetzelfde aantal "cases". S-PLUS doet niet aan "cases", maar noemt het rows. Een data-frame is maar 1 van de mogelijkheden om gegevens op te slaan, S-PLUS heeft ook vectoren, matrices en lists. Die kom je vanzelf tegen.

Er zijn een groot aantal typen variabelen die je in een dataframe kunt stoppen. De belangrijkste twee zijn: factor en numeric. Over numerical is niet veel te vertellen. Hoewel er een onderscheid is tussen natuurlijke getallen en reële getallen zijn het gewoon getallen. Het type factor representeert nominale variabelen, in enquête onderzoek is dit het

overheersend type. Een factor bestaat uit 2 delen informatie: de codes en de levels. De codes zijn te vergelijken met de getalletjes (1,2,1,2,3....) en de levels herbergen de value labels van SPSS. Een uitbreiding op het type factor is de ordered factor. Is er een ordening te herkenning, bijvoorbeeld laag, midden, hoog, dan kun je dat al in het datatype aanbrengen, komt ik zo op terug.

Een belangrijk verschil met SPSS is, dat alle veranderingen op een dataframe permanent zijn. In SPSS had je de werkfile die weer verloren ging, tenzij je hem bewaarde. S-PLUS laat alles direct doorwerken in het dataframe.

### **D.3 Naar een data-frame**

Voor je met je datamanipulatie en je analyses kunt beginnen heb je een dataframe nodig. Er zijn twee belangrijke manieren om gegevens naar S-PLUS te brengen:

- Vanuit SPSS. S-PLUS kan SPSS-systemfiles lezen, maar pas op: geen gecomprimeerde bestanden en geen SPSS 8 bestanden. Wel kan het .por, SPSS portable files lezen, gemaakt met het SPSS commando export. Dat gaat prima, het pakket maakt van de nominale variabelen meteen factoren als die variabelen maar value labels hebben. Dat is dan ook het criterium. Heb je bij een interval variabele een enkele waarde toch een omschrijving gegeven (bijvoorbeeld 999="jonger dan 16, missing"), dan maakt S-PLUS er meteen een factor van. Wil je hem toch als interval variabele importeren, dan moet je de value labels er van te voren afhalen. Het gemakkelijkst is het om het importeren toch met de GUI te doen: File Import-data enz. Het kan ook met een commando (let op de dubbele \ achter de schijfnaam, die zijn verplicht, ook bij een subdirectory):  

```
import.data(FileName = "t:\\t$.por", FileType = "SPSS_POR",
DataFrame = "testframe").
```
- Vanuit ASCII. Wil je het in één keer goed doen, dan moet je de ASCII file uitgeschreven hebben met de value labels. Dus "man" en "vrouw". Verder moet de eerste regel van het bestand de variabele namen bevatten. De ASCII file moet netjes geordend zijn, in kolommen met spaties ertussen. Je kunt hem dan in 1 keer inlezen met het commando `testframe<-read.table("invoer.txt")`. De gegevens worden uit `invoer.txt` in het dataframe `testframe` geplaatst en de factoren zijn meteen gedefinieerd. Het kan op vele andere manieren, maar dit gaat vrij handig. Overigens is het goed mogelijk om in plaats van "man", "vrouw" getallen 1, 2 in te lezen. De informatie "man", "vrouw" moet je dan later toevoegen met het levels() commando.

### **D.4 get file**

Je kunt direct bij alle variabelen van alle frames. Als je een variabele leeftijd in een dataframe `testframe` hebt, dan kun je die variabele direct "pakken" met `testframe$leeftijd`. Je kunt ook naar een dataframe toe met `attach(testframe)`. Vervolgens adresseer je de variabelen direct met hun naam: `leeftijd`

### **D.5 save file**

deze opdracht is binnen S-PLUS niet nodig, alle resultaten worden bewaard tenzij je dat niet wilt. S-PLUS vraagt bij het afsluiten of je de nieuw gecreëerde objecten (en dus ook de dataframes) van de sessie wilt bewaren. Alle spulletjes staan standaard in je default `_Data` directory. Het wordt aangeraden per project een `_Data` directory op een andere plek te openen (zie 7.2 en 7.3). Naar die directory kun je "attach-en", net als naar een dataframe.

## D.6 *display labels*

Zoals bekend krijgt u met `display labels` in SPSS een overzicht van de te gebruiken variabelen. In S-PLUS kun je in de objectbrowser even kijken. Daar klik je wat en zie je het gebeuren. In de commandotaal geef je: `names(testframe)`. Je krijgt echter alleen de namen van variabelen in het frame, daar S-PLUS geen variable labels heeft.

## D.7 *compute*

Met een `<-`, (een `<` gevolgd door een `-`) ken je een waarde aan een variabele toe. `x<-2`. In SPSS is dat voldoende om een hele variabele te vullen. S-PLUS neemt daar geen genoegen mee, de variabele is slechts een scalar. Wil je een hele variabele vullen, dan is een vector met evenveel rows als het dataframe kent noodzakelijk. Deze vector moet je dan aan de variabele toekennen. Daarbij zijn twee functies erg handig, `c()` en `rep()`. Met `c()`, concatenate, maak je van een aantal scalars een vector. Het resultaat van `c(1,1,1,1,1,1)` is een vector. Met `x<-c(1,1,1,1,1,1)` ken je aan `x` een vector toe. Enquêtebestanden zijn echter nogal groot, het is nogal onpraktisch om dat 1,1,1,1,1.. een keer of 2000 voort te zetten. Daar is de functie `rep()` voor uitgevonden. `rep(1,2000)` produceert een vector van 2000 1-en. Dus met `x<-rep(1,2000)` ken je een vector met 2000 1-en aan `x` toe.

Met de functie `dim()` kun je vragen hoeveel rows en variabelen er in je dataframe zitten. De functie geeft een vector terug met twee entries. De eerste geeft hoeveel rows er zijn, de tweede hoeveel variabelen. Als je alleen wilt weten hoeveel "cases" er zijn, dan kun je dat als volgt doen:

```
dim(testframe) [1] . rep(1,dim(testframe) [1])
```

geeft net zoveel 1-en als er rijen in het testframe zijn.

Vooralsnog is het equivalent voor `compute x=1` daarmee

```
x<-rep(1,dim(testframe) [1]).
```

We zijn er echter nog niet: `x` is een onafhankelijke vector, hij is nog niet aan het dataframe toegevoegd. Dat doe je met de functie `data.frame`:

```
testframe<-data.frame(testframe,x).
```

Bedenk echter dat S-PLUS de veranderingen permanent zijn. Voeg je `x` toe, dan is die niet meer verdwenen na afloop van de sessie. Doe dit dus alleen als het echt nodig is.

Gelukkig was dit een grapje, het kan veel makkelijker. Om een vector met 1-en aan een dataframe toe te voegen kun je het commando `cbind()` (column bind), gebruiken: `testframe<-cbind(1,testframe)`. Je hebt in ieder geval nu kennis gemaakt met `rep,c` en `dim, :)`?

Heb je eenmaal een vector `x`, dan kun je gewoon rekenen: `y<-0.5*x+3`. De vermenigvuldiging en de optelling worden elementsgewijs uitgevoerd. Na aanvankelijk wat onhandig te hebben gemanoeuvreerd, kun je gewoon aan het werk.

## D.8 *(add) value labels*

Zoals al is gesteld, een factor kent codes en levels. De levels zijn de value labels van SPSS. Er is op dit punt echter een belangrijk nuance verschil tussen SPSS en S-PLUS. In S-PLUS werk je met de levels, niet met de codes. Kom ik zo op terug.

Met `levels(sexe)` krijg je de value labels te zien, `levels(sexe) [1]` geeft het eerste value label. Je kunt ze anders definiëren door een vector toe te kennen aan de levels:

```
levels(sexe) <- c("hermafrodiet", "transsexueel")
```

Je kunt ook 1 level anders beschrijven:

```
levels(sexe) [1] <- "slak"
```

Nu het grapje: `levels(sexe) <- c("h", "h")`. Als je nu `sexe` laat weergeven (commando: `sexe`) komen er alleen maar h'tjes. Dat is zoals verwacht, S-PLUS geeft altijd de levels weer. Laat je de codes afdrukken (commando: `codes(sexe)`), dan schrik je even. Er staan alleen maar 1-en. `Levels(sexe) <- c("h", "h")` is daarmee verworpen tot een recode commando, je werkt met de levels en niet met de codes.

### D.9 variable labels

S-PLUS kent geen variable labels. Wel kun je een hele lange variable name maken, met spaties er in. Daar heb je dan vervolgens heel erg veel last van, want die moet je elke keer weer intypen.

### D.10 rename

In S-PLUS kun je met het commando `names()` variabele namen en met de functie `row.names` row-namen ("case labels") geven. Bij de functie moet je opgeven over welk dataframe het gaat. Een voorbeeld doet hopelijk wonderen:

- `names(testframe) <- c("sexe", "leeftijd in jaren")`
- `row.names(testframe) <- c("pietje", "puk", "jan", "klaas", "willem")`

Je kunt 1 van die namen veranderen met het volgnummer van de variabele in het frame. Zou je "leeftijd in jaren" willen veranderen in "age, vraag102", dan lukt dat op de volgende manier: `names(testframe) [2] <- "age, vraag102"`

### D.11 recode

Bij de recode moeten we onderscheid maken tussen numericals en factors. Een zeer handige mogelijkheid van de recode was om leeftijd in klassen te verdelen: `recode leeftijd(16 thru 34=1) (35 thru 64=2) (65 thru hi=3) (else=sysmis) into age3`. S-PLUS doet dit met de functie `cut.cut()` maakt een factor uit een numerieke variabele en voegt zelf overzichtelijke levels (value labels) toe.

SPSS:

```
recode leeftijd(16 thru 34=1) (35 thru 64=2) (65 thru
hi=3) (else=sysmis) into age3
value labels age3 1 "16-34" 2 "35-64" 3 ">=65"
```

S-PLUS:

```
age3 <- cut(leeftijd, c(16, 34, 64, Inf))
```

De levels die het oplevert komen te voorschijn met het commando

```
levels(age3):1 "15+ thru 34" 2 "35+ thru 64" 3 "65+ thru Inf"
```

S-PLUS heeft dus geen hi en geen lo, maar een Inf en een -Inf. Het heeft ook een system missing, de NA (not available). Als leeftijd lager zou zijn dan 16, dan zou het resultaat in `age3 NA` zijn.

Cut kun je overigens ook een getal x opgeven om leeftijd te verdelen in x intervallen van gelijke lengte. `age3 <- cut(leeftijd, 3)`



Nu de recode van factoren. Er is al een manier genoemd, je kunt nieuwe levels in vector vorm toekennen. `recode age3 (1,2=0) (3=1)` wordt dan:

```
levels(age3) <- c("65-", "65-", "65+")
```

Een waarschuwing is op zijn plaats, alweer. Zoals gezegd, S-PLUS brengt de wijzigingen in zijn databases aan. Je kunt een dergelijk programma dus geen twee keer draaien. Het is beter om alvorens een levels "recode" te doen, een nieuwe variabele te maken.

```
jongoud <- age3
```

```
levels(jongoud) <- c("65-", "65-", "65+")
```

### **D.12 if en do if**

Afhankelijk van de waarde van een variabele wordt er gemanipuleerd met alle rows van een andere variabele. Dit kan in S-PLUS heel erg elegant met het sub-script mechanisme:

SPSS:

```
compute v1=0
```

```
if v2=1 v1=100
```

S-PLUS (laatste regel alleen even):

```
v1[v2==1]=100
```

Dit lees je als: de v1's waarvoor v2=1 moeten 100 worden.

Do if gaat precies zo, je hebt alleen geen if then else constructie nodig, het gaat direct met een aantal van die commando's onder elkaar:

```
v1[v2==1]
```

```
v3[v2==1]
```

```
v5[v2==1]
```

### **D.12 rank ntiles**

Om een variabele in stukken met gelijke aantallen respondenten te verdelen kun je de quantile functie in samenwerking met de cut functie gebruiken. Ik bespreek het aan de hand van het volgende voorbeeld:

```
test <- rnorm(1000) (trek 1000 normaal verdeelde getallen met gemiddelde 0 en sd 1)
```

```
x <- quantile(test, c(.20, .40, .60, .80))
```

```
z <- cut(x, c(-Inf, x, Inf))
```

De quantile functie geeft een vector terug met de getallen die test in stukken met elk 20% rows verdeeld. Vervolgens kun je die gebruiken om de rows toe te kennen aan een nieuwe variabele z. De onderverdeling kent de einden nog niet (-Inf en Inf) en die plakken we er even aan met c()

### **D.14 count**

In SPSS kun je met count tellen hoe vaak een 1 of een 2 voorkomt in een aantal variabelen. Ingewikkeld voorbeeld:

```
geteld = count(v1(2), v2(3 thru 6), v3(7 thru hi))
```

In S-PLUS kun je een conditie testen. Voorbeelden: `v1=="Vrouw"`, `3<=v2<=6`, `7<=v3<Inf`. De tests mogen bij een factor zowel op de levels als met de codes uitgevoerd worden, en dat vind ik raar. De ongelijkheden kunnen alleen berekend worden voor numericals en voor ordered factoren. Bij een ordered factor heb je aangegeven welke orde de levels hebben. Dat gaat met de functie `ordered()`, waarbij je het beste wel de

rangordening kunt aangeven. Het kan ook zonder dat je iets aangeeft, maar dan doet S-PLUS maar wat. Het gaat zo:

```
opleiding<-ordered(opleiding,c("laag","midden","hoog"))
```

Vervolgens weet S-PLUS dat bij de variabele opleiding de volgende relatie geldt: laag<midden<hoog.

Het antwoord op een test is een boolean, T of F. Wil je booleans optellen, dan moet je ze eerst tot integer promoveren. Een T wordt daarbij een 1, een F een 0. Het gaat zo:

```
as.integer(v1=="Vrouw")
```

Nu het gehele probleem: In S-PLUS los je dat op door per onderdeel te vragen of het true or false is. De onderdelen converteer je naar integers met as.integer(). De resultaten tel je op.

```
geteld<-
as.integer(v1=="Vrouw")+as.integer(3<=v2<=6)+as.integer(7<=v3<Inf)
```

### **D.15 frequencys en descriptives**

Heel eenvoudig, ben je aan een dataframe "geattached", dan `summary(opleiding)`. Ben je dat niet, dan moet je eerst het dataframe specificeren:

`summary(testframe$opleiding)`. Voor een interval variabele genereert de summary standaard quartielen, mediaan, een mean, een minimum en een maximum. Bij een factor genereert het een frequentietabel.

Het mooie van het commando summary is, dat je het op van alles en nog wat kan loslaten. Zo kun je alle frequencys van een heel frame in 1 keer krijgen: `summary(testframe)` en `hoppa`.

Je kunt ook een `summary()` op data-type list loslaten. Dat is erg belangrijk, want de uitvoer van statistische procedures komt in de vorm van een list. Een list herbergt meerdere soorten data. Je kunt zowel een vector als een matrix als een dataframe in een list opslaan. `testlist<-list(testframe,x,amatrix)`. Heeft een statistische procedure een list voor je gemaakt, dan krijg je de uitvoer het mooist met `summary(): summary(testlist)`. De maker van de statistische procedure heeft voor je nagedacht wat je zou willen zien en dat in de summary, de default afdruk, van de procedure gestopt. Verder is het nog aardig dat je bij die list kan, en dus bij de uitvoer van je statistische procedures. Zou testlist de uitvoer van een lineaire regressie zijn, dan kun je met `testlist[1]` bij de coëfficiënten. Gelukkig mag het ook met `testlist$coefficients` en met `coef(testlist)` en `coefficients(testlist)`.

Voorbeeld: `x<-coef(testlist)`, `y<-x[1]*x[1]` (intercept gekwadrateerd)

Er is een tweede manier om een frequency te maken: `table(age3)`. Vervelend alleen is dat deze functie standaard de missings niet laat zien. Dat moet je afdwingen door de functie `na.include: table(na.include(age3))`

### **D.16 crosstabs**

De functie `table` kun je ook weer gebruiken voor een kruistabel:

```
table(na.include(age3),na.include(geslacht))
```

Table drukt echter geen percentages af en de notatie is ook nogal ongelukkig. Er is een wat handiger functie: `crosstabs`

```
crosstabs(~age3+geslacht, na.action=na.omit)
```

geeft een kruistabel met alle percentages, maar zonder NA

```
crosstabs(~age3+geslacht, na.action=na.include)
```

geeft ook de NA's meegepercenteerd.

Als je de `na.action` niet specificceert, dan kiest hij de default voor die `na.action` en vervelend vaak is dat `na.fail`.... S-PLUS voert de `crosstabs` dan dus niet uit. Vooral in het begin is dat erg vervelend.

Overigens, het slingertje moet, het maakt deel uit van de formule-motor van S-PLUS. Snap je die, dan is bovenstaande formule geen enkel probleem. In de S-PLUS Guide to Statistics, hoofdstuk 2, staat die formule-motor duidelijk uitgelegd.

### **D.17 add files**

In SPSS gebeurt het volgende bij een `add files`: overeenkomende variabelen krijgen in het resultaat overal vulling. Als een variabele in een bestand niet voorkomt dan krijgen de cases van het bestand `system missing` voor die variabele.

S-PLUS heeft het commando `rbind` (`rowbind`). Met dit commando plak je de rijen van twee dataframes onder elkaar. Het komt echter wat preciezer dan bij SPSS. Er moeten precies evenveel variabelen in elk frame zitten en ze moeten precies dezelfde namen hebben.

```
voorbeeld: culxx<-rbind(cul94,cul95)
```

Met het commando `cbind` plak je twee dataframes kolomsgewijs achter elkaar. Ze moeten dan wel precies hetzelfde aantal rows hebben.

### **D.18 match file**

Alles wat `match file` kan, kan S-PLUS ook met het commando `merge`. `Merge` gebruikt 2 frames en levert een nieuwe, waarbij gekoppeld is op een aantal variabelen. De gebruikelijke database relaties zijn daarbij mogelijk (1 op 1, 1 op m, m op 1 en zelfs m op n). In de help van S-PLUS staat het commando `merge` uitgebreid en redelijk helder beschreven. Hier slechts een voorbeeld uit die help:

```
authorbooks<-merge(authors, books, by.x = c("FirstName","LastName"),  
by.y = c("AuthorFirstName","AuthorLastName"), all.x = T)
```

De dataframes `authors` en `books` worden gekoppeld. De sleutelvariabelen aan de `authors` kant zijn `"Firstname"` en `"Lastname"`, de sleutelvelden aan de `books` kant zijn `"AuthorFirstname"` en `"AuthorLastname"`. Met `all.x=T` wordt aangegeven dat ook als een auteur geen boek heeft geschreven, of als een boek geen auteur heeft het resultaat terug te vinden moet zijn. De `merge` wordt in dit geval geplaatst in `authorbooks`. Nogmaals: de help is uitgebreider en zelfs helderder.

### **D.19 missing values**

S-PLUS kent maar 1 soort missing value, de NA, oftewel de `system-missing`. User missings kunnen niet. Je kunt, voor zover ik weet 3 dingen met de NA's doen in een statistische procedure:

1. Laat ze lekker meedoen: `na.action=na.include`
2. Verwijder ze: `na.action=na.omit`
3. Laat de procedure stoppen: `na.action=na.fail`

In de volgende Windows versie van S-PLUS worden multiple imputation methoden opgenomen.



## **APPENDIX E BOEKBESPREKINGEN VAN CANDIENSTEN**

### **An Introduction to S-PLUS for Windows**

*Longhow Lam*

Deze Engels talige inleiding is bedoeld om beginnende S-PLUS gebruikers op weg te helpen in het S-PLUS data analyse en visualisatie systeem. Het behandelt onder andere het editen van data, het maken van grafieken en het uitvoeren van statistische routines. Er wordt kort iets verteld over de S taal. Ook ervaren command-line gebruikers van S-PLUS, die over het algemeen de grafische gebruikersinterface weinig gebruiken, zullen merken dat een gedegen kennis van de grafische gebruikersinterface hun S-PLUS gebruik efficiënter maakt. Ook zullen zij in staat zijn om de door hun zelf ontwikkelde functies op een eenvoudige manier ter beschikking te stellen aan anderen.

### **The Basics of S and S-PLUS**

*A. Krause & M. Olson*

S-PLUS is a powerful tool for interactive data analysis, creating graphs, and implementing customized routines. Originating as the S language of AT&T Bell Laboratories, its modern language and flexibility make it appealing to data analysts from many scientific fields.

This book explains the basics of S-PLUS in a clear style at a level suitable for people with little computing or statistical knowledge. Unlike the S-PLUS manuals, it is not comprehensive, but instead introduces the most important ideas of S-PLUS through the use of many examples. Each chapter also includes a collection of exercises which are accompanied by fully worked-out solutions and detailed comments. The volume is rounded off with practical hints on how efficient work can be performed in S-PLUS. The book is well-suited for self-study and as a textbook.

Contents: 1. Introduction; 2. System Design; 3. A First Session; 4. A Second Session; 5. Graphics; 6. Exploring Data; 7. Statistical Modeling; 8. Programming; 9. Input and Output; 10. Useful Hints and Techniques; 11. Special Topics; 12. References.

### **Modern and Applied Statistics with S-PLUS**

*W.N. Venables & B.D. Ripley*

S-PLUS is a powerful environment for the statistical and graphical analysis of data. It provides the tools to implement many statistical ideas that have been made possible by the widespread availability of workstations having good graphics and computational capabilities. This book is a guide to using S-PLUS to perform statistical analyses and provides both an introduction to the use of S-PLUS and a course in modern statistical methods

Throughout, the emphasis is on presenting practical problems and full analyses of real data sets. Many of the methods discussed are state-of-the-art approaches to topics such as linear, non-linear and smooth regression models, tree-based methods, multivariate analysis and pattern recognition, survival analysis, time series and spatial statistics. Throughout, modern techniques such as robust methods, non-parametric smoothing and bootstrapping are used where appropriate.

## **Mixed Effects models in S and S-PLUS**

*José Pinheiro & Douglas Bates*

This book provides an overview of the theory and application of linear and nonlinear mixed-effects models in the analysis of grouped data, such as longitudinal data, repeated measures, and multilevel data. A unified model-building strategy for both linear and nonlinear models is presented and applied to the analysis of over 20 real datasets from a wide variety of areas, including pharmacokinetics, agriculture, and manufacturing. A strong emphasis is placed on the use of graphical displays at the various phases of the model-building process, starting with exploratory plots of the data and concluding with diagnostic plots to assess the adequacy of a fitted model. Over 170 figures are included in the book.

The NLME library for analyzing mixed-effects models in S, S-PLUS and R, developed by the authors, provides the underlying software for implementing the methods presented in the text, being described and illustrated in detail throughout the book.

The balanced mix of real data examples, modeling software, and theory makes this book a useful reference for practitioners who use, or intend to use, mixed-effects models in their data analyses. It can also be used as a text for a one-semester graduate-level applied course in mixed-effects models. Researchers in statistical computing will also find this book appealing for its presentation of novel and efficient computational methods for fitting linear and nonlinear mixed-effects models.

*José C. Pinheiro* has been a member of the technical staff in statistics research at Bell Laboratories since 1996. He received his Ph.D. in Statistics from the University of Wisconsin-Madison in 1994 and worked for two years in the Department of Biostatistics at the UW--Madison. The author of several articles in mixed-effects models, he is a member of the American Statistical Association and the Biometric Society.

*Douglas M. Bates* is Professor of Statistics at the University of Wisconsin--Madison. The author, with Donald G. Watts, of ``Nonlinear Regression Analysis and Its Applications'', he is a Fellow of the American Statistical Association and a former chair of its Statistical Computing Section.

## **S Programming**

*Venables & Ripley*

This book provides an in- depth guide for those writing software in the S language. The authors have written several software libraries which enhance S- PLUS; these and all the datasets used are available on the Internet in versions for Windows and UNIX. There are extensive online complements covering advanced material, user-contributed extensions, further exercises and new features of S-PLUS as they are introduced.

## **Modeling Survival Data, Extending the Cox Model**

*Terry M. Therneau & Patricia M. Grambsch*

This is a book for statistical practitioners, particularly those who design and analyze studies for survival and event history data. Its goal is to extend the toolkit beyond the basic triad provided by most statistical packages: the Kaplan-Meier estimator, log-rank test, and Cox regression model.

The focus of the book is on actual data examples, the analysis and interpretation of the results, and computation. The methods are now readily available in SAS and S-PLUS, and this book gives a hands-on introduction, showing how to implement them in both packages,

with worked examples for many data sets. The authors call on their extensive experience and give practical advice, including pitfalls to be avoided.

Contents: Introduction - Estimating the Survival and Hazard Functions - The Cox Model - Residuals - Functional Form - Testing Proportional Hazards - Influence - Multiple Events per Subject - Frailty Models - Expected Survival

## **Stat Labs**

*Deborah Nolan and Terry Speed*

Springer 0-387-98974

Inleiding statistiek met S-PLUS. In de woorden van Brian Ripley: "... wonderfully refreshing"!!

Inleiding statistiek met S-PLUS. In de woorden van Brian Ripley : "It is 'designed for use in a calculus-based introductory statistics course'. It is command-line not GUI based. I have only saw the book last week, but read it at one sitting, found it wonderfully refreshing and very competently done, so I very highly recommend it. Those teaching at a more advanced level will also learn from it: I did.





## **APPENDIX F AANTEKENINGEN**

Deze appendix is bedoeld voor aantekeningen van de S-PLUS gebruikers.

*De auteur stelt het zeer op prijs als ideeën, werkwijzen, opmerkingen, databestanden en/of functies, die van belang kunnen zijn voor het hele RIVM, aan hem per e-mail worden opgestuurd. Ze kunnen dan worden opgenomen in de help-functie van RIVMfun of op andere wijze beschikbaar worden gesteld aan de S-PLUS gebruikersgroep.*





