

STATWORKS : Statistics with Graphics for the Macintosh

P. Debets *

Samenvatting

Het gebruik van statistische programmatuur op een microcomputer lijkt in de praktijk, in kursussen, in evaluaties en softwarebesprekingen bijna onlosmakelijk verbonden met IBM PC's of compatibles (zie bijv. Gaaff & De Boer, 1985; Keller, 1986; Fridlund, 1986 en Nijburg & De Grefte, 1987).

Ook bij de discussie of statistiek op een mainframe danwel op een microcomputer plaats moet vinden (zie De Boer, 1986) heeft men veelvuldig geschermd met het argument dat 8 of 16 bits computers niet nauwkeurig kunnen rekenen en daarmee de Macintosh met 32 bits buiten beschouwing gelaten.

De oorzaak is simpel: er is nog maar weinig statistische programmatuur voor de Macintosh beschikbaar. Het aantal groeit echter. Er is zelfs al een evaluatie voor een aantal programma's gedaan (Lehman, 1987).

In dit artikel wordt het programma StatWorks besproken. Hierbij is in eerste instantie gelet op werkwijze en mogelijkheden. Er zijn niet allerlei tests voor snelheid of nauwkeurigheid gedraaid.

*

Dienst voor Informatieverwerking, Software-ontwikkeling en Computeronderwijs
Universiteit van Amsterdam, Herengracht 196, 1016 BS Amsterdam, 020-5252221/2074

1. Inleiding

Als men het over statistische programmatuur voor de micro heeft, dan denkt men meestal aan een IBM PC of compatible en de vele daarvoor beschikbaar zijnde programma's. Dat is niet verbazingwekkend want de meeste statistische programma's, waaronder de grote bekenden SPSS, SAS, BMDP, STATGRAPHICS en MINITAB zijn enkel nog beschikbaar voor IBM en compatibles. Ik heb er zelf ook lang aan getwijfeld of een Macintosh hier wel voor te gebruiken is. Als ik mensen achter de Macintosh zag zitten dan waren die meestal bezig met tekstverwerking, tekenen, spelletjes of muziek. Zeker voor de eerste twee activiteiten is de Macintosh uitermate geschikt; teksten met allerlei formules en symbolen en tekeningen kunnen op een vrij eenvoudige manier tot één manuscript worden samengevoegd.

De twijfel werd voornamelijk veroorzaakt door de werkwijze met de Macintosh:

" Met een muis (die ook nog eens een keer van de tafel kan rollen of tegen de Macintosh botsen) wijst de gebruiker (die 'niets' van computers behoeft af te weten en geen opdrachten hoeft te kennen of te onthouden) met een pijltje op een soort van elektronisch bureau (met een echte prullenmand) aan wat hij wil.

De gebruiker sleept, klikt, klikt dubbel en heeft menu's ter beschikking waarmee hij mappen of dokumenten op dat bureau kan bekijken of bewerken.

Hij kijkt door vensters die verplaatst, vergroot of verkleind, actief gemaakt of gesloten kunnen worden. "

Hoe kun je op zo'n manier nu de voor de statistiek zo onontbeerlijke data invoeren, transformeren, manipuleren en er statistische berekeningen op uitvoeren.

Mijn eerste kennismaking met statistische programmatuur voor de Macintosh betreft het programma StatWorks waarover een aantal leden van de Microclub van de Sektie Statistische Programmatuur nogal positief en enthousiast waren. Met behulp van de criteria van Keller (1986) zijn de mogelijkheden van dit programma bekeken.

2. Prijs, benodigde configuratie en installatie

StatWorks versie 1.2, geschreven door James Rafferty, Richard Norling, Robert Tamaru, Charles McMath en Dave Morganstein, wordt gedistribueerd door Heyden & Son Inc. USA maar is gelukkig ook verkrijgbaar in Nederland (Guide Systems Rotterdam). Het programma kost fl. 499,- excl. BTW, hierbij is de handleiding inbegrepen.

StatWorks draait op een Macintosh met 128k geheugen. Op de geleverde diskette staan: de systeem map (zodat vanaf deze schijf kan worden opgestart), het programma en een aantal voorbeeld datasets. Er behoeft verder niets geïnstalleerd te worden.

3. De dokumentatie

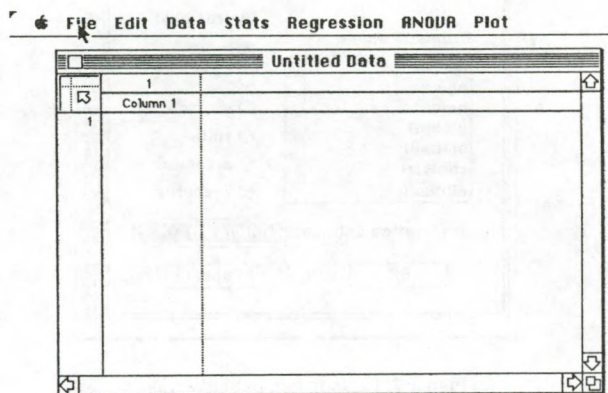
De bijbehorende Engelstalige handleiding (A5 formaat) van 95 pagina's bevat:

- een "quick menu guide" met een overzicht van alle menu's, de daarbij te kiezen kommando's en een korte uitleg;
- een korte introductie;
- een korte uitleg van de werkwijze met StatWorks en de Macintosh;
- uitleg over data en data-windows;
- per menu een uitvoerige uitleg;
- en tot slot worden de formules gegeven voor de gehanteerde statistische berekeningen.

Het boekje is verfraaid met invoer- en uitvoer-voorbeelden. Kortom een duidelijke en overzichtelijke handleiding. Zoals later blijkt heeft men deze nauwelijks nodig.

4. Het gebruik van StatWorks

Na het opstarten van StatWorks krijgt men een aantal menu's en het data-window (figuur 1).



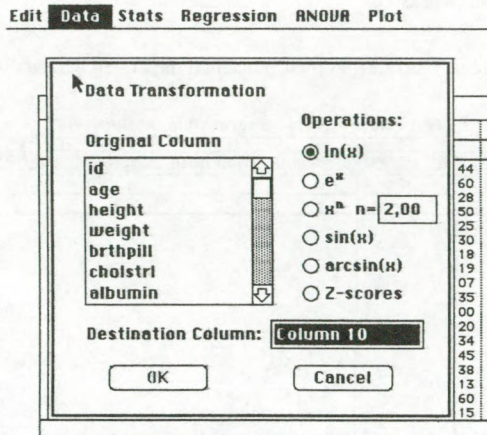
figuur 1: StatWorks menu's en data-window

Op de menubalk staan van links naar rechts een aantal menu's:

- Apple** een verzameling van hulp programma's die aan het systeem zijn toegevoegd.
- File** voor het inlezen en wegschrijven van datasets, het sturen van resultaten naar de printer en het beëindigen van het programma.
- Edit** voor het verplaatsen/kopiëren van rijen of kolommen binnen een dataset of van de ene dataset naar de andere.

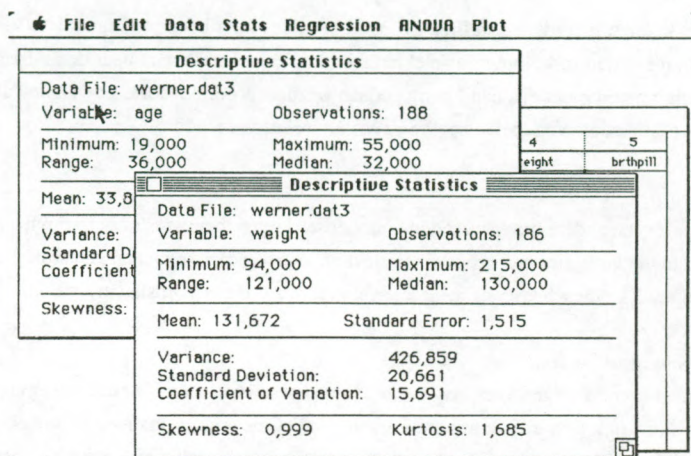
Data	voor datamanipulaties.
Stats	voor beschrijvende statistiek en toetsen.
Regression	voor regressie analyse.
Anova	voor variantie analyse.
Plot	voor een aantal grafische mogelijkheden.

Om met StatWorks te werken heeft men geen stuurtaal te leren. Het enige vereiste is dat men iets weet over de werkwijze met een Macintosh. Bovendien moet men, omdat StatWorks een statistisch programma is, op de hoogte zijn van de statistische termen die gebruikt worden. StatWorks maakt gebruik van zgn. "dialog boxes", waarin de voor een taak of analyse benodigde informatie wordt opgevraagd. Nadat men uit het menu een bepaalde optie heeft gekozen verschijnt zo'n dialog box. In figuur 2 is deze gegeven voor Transform; de procedure voor het transformeren van variabelen in de dataset.



Figuur 2: Dialoog box in StatWorks

In de box of het venster selecteert men de variabele die men wil transformeren (de original column) en de transformatie (operations). Men krijgt hierbij een overzicht van de variabelen die in de dataset zitten. Het resultaat wordt automatisch opgeslagen in de eerstvolgende niet-gebruikte kolom. Het selecteren gaat op de Macintosh manier: met de muis aanwijzen, klikken en slepen. Op eenzelfde manier specificeert men eveneens de statistische bewerkingen. De uitvoer van de procedures verschijnt eveneens in vensters waarvan de inhoud naar de printer gestuurd kan worden. In figuur 3 staat als voorbeeld de uitvoer van de procedure Descriptive.



Figuur 3: Uitvoer van de procedure Descriptive

5. Data management

Import / Export

De data (numeriek of alfanumeriek) moeten in StatWorks via het data-window (zie figuur 1) worden ingevoerd. In dit window zijn de kolommen de variabelen en de rijen de cases. Het maximaal aantal kolommen in StatWorks is 35.

Via het data-window kan men ook de data "editen" (verbeteren, uitbreiden).

Ingevoerde data kan men als StatWorks file bewaren met het kommando "Save" in het File menu. Een StatWorks file wordt ingevoerd door deze file aan te wijzen en "dubbel te klikken" of via het kommando "Open" in het file-menu.

Het is mogelijk standaard ASCII tekst files in te lezen, echter niet in de meest gebruikte formaten free-field (met spaties of komma's tussen de getallen) en fixed format (vaste kolommen) maar met tabs als getalscheiders. Als de data afkomstig zijn uit een (statistisch) programma op een IBM PC zal men een aantal kunstgrepen moeten uithalen om hieraan te voldoen. Versie 1.2 van StatWorks kan geen ASCII files inlezen die groter zijn dan 32k, men moet de data dan opsplitsen. Het inlezen van ASCII files neemt nogal wat tijd in beslag. Voor een dataset van 188 cases en 9 variabelen (9k) heeft StatWorks ongeveer 2 minuten nodig, voor een dataset van 32k bijna 8 minuten.

De echte liefhebber kan met behulp van twee voorbeeld programma's in de handleiding vanuit eigen Pascal of BASIC programma's StatWorks files lezen of schrijven.

Kombineren/Sorteren

In StatWorks kunnen meerdere datafiles geopend zijn, de laatst geopende is actief. Via het klembord kan men rijen of kolommen selekteren en kopiëren van de ene naar de andere file. Bestanden combineren moet dus handmatig gedaan worden. Voor het sorteren van de data op basis van één of meerdere variabelen heeft StatWorks wel een procedure.

Selekteer/Subgroep

StatWorks heeft geen andere mogelijkheden voor het selekteren van cases of het aanwijzen van subgroepen dan via het klembord. Voor elke selectie/subgroep dient men een nieuw bestand te maken. Bij ingewikkelde selecties of meerdere subgroepen is dit vrij omslachtig.

Compute/Kommando-file/Taal

StatWorks heeft voorzieningen voor het berekenen van nieuwe variabelen zoals rekenkundige transformaties (zie ook figuur 2), combinaties van twee of meer kolommen en herkoderingen. De herkoderingen zijn nogal omslachtig, in de procedure recode kan men slechts één herkodering voor één variabele doen. Indien men een numerieke variabele naar bijv. 6 categorieën wil transformeren dan zal men de procedure zes keer moeten uitvoeren. StatWorks maakt geen gebruik van een stuurtaal; en heeft dus ook geen mogelijkheden voor kommando-files, sprongopdrachten of andere programmeerfaciliteiten.

Labels/Missing/gewichten

De default namen voor de variabelen Column 1, Column 2 etc. kunnen gewijzigd worden in meer zinvolle teksten. Aangezien StatWorks alfanumerieke variabelen van maximaal 15 tekens toestaat is het gebruik van value labels overbodig.

Als missing value geldt de "return" of de "tab". Met deze missing values wordt bij berekeningen en in de procedures rekening gehouden.

In StatWorks kunnen geen ophooggewichten opgegeven worden.

6. Beschrijvende Statistiek en Toetsen

Op dit gebied heeft StatWorks de volgende mogelijkheden:

<i>Descriptive</i>	voor berekening van minimum, maximum, range, mediaan, gemiddelde, standaardfout, variantie, standaarddeviatie, variatie-coëfficiënt, skewness en kurtosis.
<i>Cross-Tabulation</i>	twee dimensionale kruistabellen met percentages en associatiematen.
<i>t-Test</i>	toetsen van gemiddeldes
<i>Normality Test</i>	Kolmogorov-Smirnov toets op normaliteit van verdelingen.
<i>Kolmogorov-Smirnov</i>	Kolmogorov-Smirnov toets op gelijkheid van verdelingen.

<i>Mann-Whitney U</i>	niet-parametrische t-toets voor onafhankelijke steekproeven.
<i>Wilcoxon Signed Rank</i>	niet-parametrische t-toets voor gepaarde steekproeven.
<i>Spearman Correlation</i>	rangorde correlatie volgen Spearman.
<i>Kendall Correlation</i>	rangorde correlatie volgens Kendall.

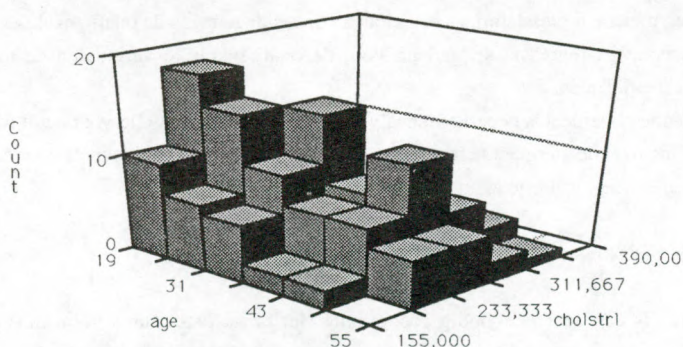
StatWorks heeft geen procedure voor frekwentie verdelingen; voor de gewone Pearson correlatiecoëfficiënt moet men de regressie procedure gebruiken.

Wat betreft de kruistabellen valt op dat de klassen niet in de "juiste" volgorde staan. StatWorks gaat er vanuit dat er alleen van categorische (nominale) gegevens kruistabellen gemaakt worden; ook de associatiematen zijn voor variabelen met dit meetnivo. De volgorde van de waarden in de tabel wordt bepaald door de volgorde in de datamatrix.

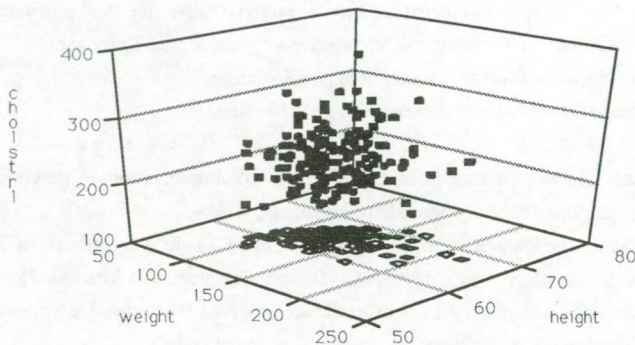
De kruistabellen procedure is beperkt tot twee-dimensionale kruistabellen. Het maximaal aantal categorieën van de variabelen bedraagt 10, echter bij meer dan 5 categorieën voor één van de variabelen past de tabel niet meer op het scherm en moet men hem laten printen wil men hem helemaal zien.

7. Graphics

De grafische mogelijkheden in StatWorks zijn bedoeld voor numerieke gegevens: strooi-diagram, lijngrafiek, histogram, boxplot en 3-dimensionaal spreidingsdiagram of histogram. Met name deze laatste zijn erg fraai (zie figuur 4 en 5). Labels en intervallen worden door het programma bepaald. Met behulp van het "grab picture" kommando kunnen de tekeningen naar tekstverwerkers overgeheveld worden.



Figuur 4: Drie-dimensionaal histogram



Figuur 5: Drie-dimensionaal spreidingsdiagram

8. Multivariate Analyses

Wat betreft multivariate analyses is StatWorks beperkt tot regressie- en variantie-analyse.

In het regressie onderdeel kan men kiezen uit enkelvoudige, polynome of meervoudige regressie-analyse. Bij elke analyse kan men opgeven of de regressielijn door de oorsprong moet gaan. Residuen, gestandaardiseerde residuen en de voorspelde y-waarde kunnen bewaard en later gebruikt worden voor diverse plotjes.

De uitvoer bestaat uit een venster met de regressiecoëfficiënten en hun significantienivo en een venster met de ANOVA-tabel, de correlatiecoëfficiënt en de Durbin-Watson statistic. Bij een enkelvoudige of polynome analyse krijgt men bovendien een spreidingsdiagram van de beide variabelen met de regressielijn, en een plot van de residuen tegen de onafhankelijke variabele. Bij meervoudige regressie krijgt men voor de onafhankelijke variabelen de matrix met correlatiecoëfficiënten.

Het variantie onderdeel is beperkt tot analyses met 1 of 2 factoren. Hiervoor heeft men ook de beschikking over de niet-parametrische alternatieven: de Kruskal-Wallis toets voor de eenwegs analyse en de toets van Friedman voor de tweewegs analyse.

Evaluatie en discussie

StatWorks is een zeer eenvoudig programma Het bevat procedures voor vaak gebruikte statistische technieken zoals: beschrijvende statistiek, kruistabellen met associatiematen, toetsen, regressie-analyse, variantie-analyse en grafieken.

Het programma is zeer eenvoudig te gebruiken. Het leren van een of andere stuurtaal is voor

dit programma niet nodig; de gehele kommunikatie met het programma gaat via menu's en vensters, waarin met behulp van de muis de keuzes gemaakt worden. De goed verzorgde handleiding heeft men waarschijnlijk alleen de eerste keer nodig.

Data-invoer en data-editing in het programma zijn vrij eenvoudig. Het programma bevat een groot aantal datatransformatie mogelijkheden, welke met behulp van de menu's vrij eenvoudig zijn te verwezenlijken, alleen het indelen in klassen is nogal omslachtig. Helaas ontbreken ook de "case-selektie" mogelijkheden waardoor het niet mogelijk is voor een bepaalde groep een procedure te doen; voor de betreffende groep moet eerst een nieuw bestand gemaakt worden. Wat betreft de statistische procedures ontbreekt er eigenlijk alleen een procedure voor frekwentieverdelingen van categorische (nominaal of ordinaal) gegevens. In de meeste analyses is dit een van de eerste stappen. Het grafische onderdeel zou nog uitgebreid moeten worden met grafieken (zoals staafdiagrammen of taartdiagrammen) voor dit soort gegevens.

Het programma StatWorks heeft mij er in elk geval van overtuigd dat statistische bewerkingen op een Macintosh mogelijk zijn. En niet alleen dat. Statistische bewerkingen zijn op eenzelfde 'speelse' manier mogelijk als tekstverwerking en tekenen. De bewerkingen worden daarmee niet minder serieus en niet minder nauwkeurig, zoals de rekentests in de evaluatie van een aantal Macintosh programma's door Lehman (1987) aantonen.

Lehman bespreekt in deze evaluatie de programma's Data Desk Professional, Stat80, Statfast, Statview, Statview512+, StatWorks, SYSTAT, TrueSTAT en WormStat. Lehman beveelt Stat80 of SYSTAT (een pakket dat ook voor de IBM PC beschikbaar is) aan voor het zwaardere werk, StatWorks behoort tot de programma's voor diegenen die niet dagelijks met een statistisch programma behoeven te werken. In het algemeen konkludeert Lehman dat ook Macintosh gebruikers de band met het mainframe kunnen verbreken en al hun analyses op hun bureau kunnen doen.

Literatuur

- Boer, Th. W. de (1986). "Micro versus mainframe". Kwantitatieve Methoden, 7, 1986, nr. 20, p. 65-79.
- Fridlund, A.J. (1986). "Statistics Software". InfoWorld, September 1, 1986, p. 31-40.
- Gaaff, A., en de Boer, Th.W. (1985). "Statistische programmatuur voor personal computers geteld, gewogen en getest.". In: Bundel bij het Symposium Statistische Software, 1985, p. 55-82. Technisch Centrum FSW, Universiteit van Amsterdam.
- Keller, W.J. (1986). Statistische programmatuur voor de personal computer. Stichting Teleac, Utrecht 1986/1987.
- Lehman, R.S. (1987). "Statistics on the Macintosh". Byte, july 1987, p. 207-215.
- Nijburg, D.A. en de Grefte, I. (1987). "Statistische Software". Software Magazin, april 1987, p. 14-30.