

STAT34

Statistiekpakket voor het onderwijs

J.M. Buhrman¹

Gegevens

Auteur/distributeur: J.T. Geilenkirchen, Etten-Leur

Besturingssystemen: MS-DOS, Macintosh

Omvang: 119 kB, harde schijf niet nodig; enkele voorbeeldbestanden van 3 tot 6 kB

Prijs:

- f 195,- voor individuele gebruikers

- f 195,- per jaar voor een licentie voor een beknopte versie voor een hele school
waar het bijbehorende **Werkboek StatistiekPracticum** (los verkrijgbaar, f 29,40,
Stenfort Kroese) is voorgeschreven

Samenvatting

STAT34 met werkboek is een combinatie waar de leerling de allereerste kennismaking met statistiek per computer kan opdoen. De mogelijkheden zijn beperkt, de layout van 'grafieken' en tabellen is pover, de nauwkeurigheid is maar net voldoende. De werking is niet in alle gevallen even grondig. Voor wat het programma kan, is de prijs hoog, zeker als de hele school het boekje heeft gekocht.

Het gebruik van de Nederlandse taal kan voor sommigen een voordeel zijn. Ook is het prettig dat er zo goed als geen eisen aan de PC worden gesteld. Heeft men geen moeite met Engels en heeft men nieuwere apparatuur, dan kan beter direct met een echt pakket worden begonnen.

¹ Hogeschool voor Economische Studies, Postbus 295, 1000 AG Amsterdam, tel. 020 - 5236057

1 Algemeen

Dit programma is bedoeld om leerlingen kennis te laten maken met het idee "statistisch pakket". Het is hiervoor geschikt, mits men enige eigenaardigheden voor lief neemt en een leraar deze zó goed kent, dat hij de leerlingen over de hobbels heen kan helpen. Verder moet men aan de mogelijkheden en nauwkeurigheid geen al te hoge eisen stellen. Als uitvoermedium kan men kiezen tussen scherm, printer en file; een combinatie is niet mogelijk. De keuze dient van te voren voor de gehele opdracht te worden gemaakt. Na elke voltooide opdracht kan men de uitvoerbestemming wijzigen.

Na aanroep met 'STAT34' verschijnt het volgende menu:

```

***** S T A T 3 4 *****
                          HOOFDMENU

Frequentieverdeling      Selecteren
Kruistabel                Selectie ongedaan maken
Maatstaven               Classificeren
Subtabel                 Transformeren
                          Sorteren

Correlatiematrix
Regressie                Bestand afdrukken
Puntenwolk               Bestandsopbouw afdrukken
                          Waarnemingen toevoegen
                          Variabele toevoegen
Binomiale verdeling      Gegevens veranderen
Poisson-verdeling        Namen/labels veranderen
Normale verdeling        Waarden 'onbekend' veranderen
Inverse normale verdeling Waarnemingen verwijderen
T-verdeling              Variabele verwijderen
Chi-kwadraatverdeling
F-verdeling
Indexcijfers
Tijdreeksanalyse

                          Uitvoerbestemming
                          Gegevens opslaan/inlezen

Stoppen

Maak uw keuze m.b.v. cursors en type <CR>

```

De keuze uit de taken kan uitsluitend worden gemaakt door met de cursorbesturings-toetsen - alleen de pijltjes - (door het pakket de "cursors" genoemd) de lichtbalk (hierna te noemen: cursor) op de gewenste keuze te brengen.

De keuzen zijn globaal te verdelen in de volgende groepen:

- berekeningen van kansen
- bewerkingen met een gegevensbestand (opbouw, data invoeren, e.d.)
- analyse van een gegevensbestand (berekeningen, figuren maken)

Deze keuzen zijn in het menu niet op deze wijze ingedeeld, wat een onverzorgde indruk maakt. Onhandig is het, dat bij eerste aanroep de cursor zich linksboven in het menu bevindt, terwijl - afgezien van de kansberekeningen - de enige zinvolle eerste taak (inlezen van een bestand) rechtsonder in het menu staat. De argeloze gebruiker die het telkens oplichtende 'frequentietabel' intikt, wordt geconfronteerd met de

mededeling 'geen gegevens !'. Dat is niet bemoedigend voor de student die de gegevens voor een frequentietabel van zijn papiertje had willen intikken. Had hij de instructies in het werkboek gelezen, dan had hij geweten wat hem door het programma niet verteld wordt: er moet altijd worden begonnen met de keuze 'bestand opslaan/inlezen'. Zulke dingen vallen onder de hobbels die ik in de eerste alinea noemde. Het zou bovendien prettig zijn geweest als de cursor op deze voor de hand liggende keuze staat direct na het aanroepen van het programma.

Bij het invoeren van gegevens kan men kiezen uit zelf intikken en het inlezen van een STAT34-bestand. Het inlezen van een ASCII-file (bijvoorbeeld uitvoer van een ander programma) behoort niet tot de mogelijkheden.

2 De kansverdelingen

Voor onderwijsdoeleinden is de nauwkeurigheid voldoende. De lespraktijk wordt wellicht gevolgd als bij de vraag naar de kans

$$P(x < 27 \mid \text{bin}(100 ; 0.30))$$

het programma volstaat met het advies de normale benadering te gebruiken, en vervolgens terugkeert naar het hoofdmenu. Het is jammer dat het programma niet zowel de kans geeft als het advies om de normale benadering ook eens te proberen, en dat de cursor na dit advies niet vanzelf op de normale kansverdeling terecht komt.

Het opgegeven aantal decimalen is soms een beetje groter dan de nauwkeurigheid rechtvaardigt, maar een praktisch bezwaar is dat niet.

3 Het verwerken van een datamatrix

Het uitvoeren van sommige bewerkingen verliep geheel zoals verwacht. Dat was echter helaas lang niet altijd het geval. Zelfs in het bijgeleverde bestand VOORBEEL gebeurden er enkele rare dingen. Zo gaf de opdracht 'Kruistabel' bij de variabelen 'geboortjaar' en 'salarisschaal' op, dat er 21 waarnemingen 'niet geclassificeerd' waren. Nu is het wel denkbaar dat voor het maken van een kruistabel eerst een indeling moet worden opgegeven, maar het programma vraagt daar niet om. Wordt er niet om gevraagd, dan mag de gebruiker er vanuit gaan, dat het programma zelf een klasse-indeling ontwerpt. Overigens heeft het classificeren tot gevolg, zo meldt de handleiding eerlijk, dat de waarnemingen worden vervangen door nummers van de klasse-indeling. Van het bewerkte gegevensbestand moet dus maar liever een copie worden bewaard.

In uitvoer in tabelvorm (bijvoorbeeld de variantieanalyse-tabel) staan de decimale punten niet netjes boven elkaar. De getallen hebben een vast aantal posities, en daarmee worden de kolommen opgebouwd. Hoewel dit lelijk en ongebruikelijk is, kan men aanvoeren dat men zo het grootst mogelijke aantal significante cijfers krijgt met zo min mogelijk ruimte.

Labels worden niet optimaal gebruikt. In de uitgevoerde tabellen zijn alleen de codes af te lezen. Aan de andere kant moeten bij de invoer van een nieuw bestand wel

telkens namen worden opgegeven, ook als men dit overbodig vindt. In het stadium van zijn studie waarin de leerling STAT34 gebruikt, kan het gebruik van labels bovendien beter helemaal achterwege blijven.

Bij de regressieanalyse wordt men op een gegeven moment geconfronteerd met de vraag 'Storingen in bestand opnemen'. Met 'Storingen' worden de residuen bedoeld, en met 'opnemen' het vervangen van de oorspronkelijke waarnemingen door deze residuen. Wil men daarna de regressiecoëfficiënten nog eens zien (waarvoor de hele opdracht opnieuw dient te worden uitgevoerd), dan ziet de analyse er geheel anders uit, want dan zijn de residuen de waarnemingen geworden. Wie storingen aanvraagt, kan ze krijgen. Had u het bestand nu maar eerst even opgeslagen!

4 Tijdreeksen en indexcijfers

De tijdreeksanalyse beperkt zich tot het berekenen van de voortschrijdende gemiddelden en seizoenpatronen volgens het additieve of multiplicatieve model. Daarbij moet de student eerst de handleiding of zijn leraar raadplegen. Het programma vertelt namelijk niets over de wijze waarop de gegevens in het 'bestand' dienen te worden opgenomen, en deze is vanzelfsprekend anders dan bij een regressieanalyse.

Voor de berekening van indexcijfers geldt hetzelfde. Zonder leraar berekent de student vermoedelijk eerst een paar keer indexcijfers over de nummers van de waarnemingen (die dan als prijzen of hoeveelheden worden opgevat) of berekent hij de voortschrijdende gemiddelden van de jaartallen in plaats van van de omzetten.

5 Diversen

Bij de kansberekeningen trof mij de noodzaak kansen in percentages op te geven onaangenaam. Weliswaar staat het %-teken er telkens bij, maar toch keek ik de eerste keer raar op toen ik bij een binomiale kansberekening ongeloofwaardige uitkomsten kreeg nadat ik $p = 0.35$ had opgegeven. Werde namelijk gelezen als 0.0035 en dat maakt wel wat uit. Overigens kan ik mij voorstellen dat andere gebruikers (anders dan ik) kansen in percentages prettiger vinden dan als fracties tussen 0 en 1.

Aldoor in de uitvoer wordt de mogelijkheid geboden: * = terug. Wie denkt dat hij hiermee het vorige scherm van de output te zien krijgt, komt bedrogen uit. Met ruwe hand wordt hij teruggeworpen in het hoofdmenu. Dat is trouwens de enige manier om de vorige bladzijde te zien te krijgen: de hele opdracht van voren af aan afwerken en één bladzij eerder tot stilstand komen. Overigens ligt het gebruik van de ESC-toets voor een sprong naar het hoofdmenu méér voor de hand.

6 Het werkboek

Het werkboek is een combinatie van een verzameling opdrachten en een handleiding voor het pakket. De inhoud is voldoende overzichtelijk, maar een register ontbreekt. Dat geeft aan dat het boekje als naslagwerk niet voldoet, en eigenlijk alleen geschikt

is om - net als het pakket - één keer te worden doorgewerkt. De vragen zijn niet al te moeilijk en door de redelijke leerling te beantwoorden. Bij sommige vragen gaat het er alleen om de uitvoer van de monitor te interpreteren en op de juiste manier over te schrijven, soms moet iets dieper worden nagedacht. Het niveau lijkt redelijk.

7 Wensen

Uit het bovenstaande zal duidelijk zijn geworden dat programma aan kwaliteit zou winnen als de volgende eigenschappen werden aangebracht:

- betere indeling van het hoofdmenu
- plaats cursor in het hoofdmenu blijft behouden na uitvoeren van een opdracht
- mogelijkheid tot terugbladeren door de uitvoer
- gebruik ESC-toets voor het afbreken van een opdracht
- aantal decimalen aangepast aan de nauwkeurigheid
- betere aanduidingen bij tabellen (bijvoorbeeld dat de andere helft van de tabel op het volgende scherm zichtbaar wordt)
- betere aanduidingen bij plaatjes (bijvoorbeeld een minder primitieve schaalverdeling bij histogrammen en puntenwolk)
- mogelijkheid tot het inlezen van andere ASCII-gegevensbestanden
- bestandsopbouw afhankelijk van de taak die moet worden uitgevoerd (de antwoorden bij een enquête onder proefpersonen worden gewoonlijk anders geordend dan gegevens voor de berekening van indexcijfers)
- losse handleiding (en dan wel met register)

