

Toen de muis nog in een hol zat

## WESP: Het eerste algemene statistische pakket bij de RUG

Leo van der Weele [l.t.van.weele@rc.rug.nl](mailto:l.t.van.weele@rc.rug.nl)

Tegenwoordig klikt de SPSS-gebruiker op een icoontje, krijgt dan een afrolmenu, kiest daarin de methode, specificeert variabelen, bepaalt nog wat opties en krijgt overzichtelijke tabellen en fraaie grafieken, die hij probleemloos in een tekstverwerker kan opnemen. Alle informatie staat overzichtelijk op het scherm en de aansturing gebeurt met de muis en een beetje typewerk. Dat is wel anders geweest. Bijvoorbeeld in het geval van het eerste algemene statistische pakket dat bij de RUG werd gebruikt: WESP (Waarlijk Eenvoudig Statistisch Pakket). Een verhaal over ponskaarten, wachten op de regeldrukker-uitvoer, foutmeldingen, weer wachten en resultaten, veel getallen en weinig tekst, die voor de publicatie moesten worden overgetypt, op een schrijfmachine.

### Voorgeschiedenis

#### De ZEBRA

Al rond 1960 werden op de ZEBRA, de eerste computer bij de RUG, statistische programma's gedraaid, bijvoorbeeld voor de berekening van correlatie-coëfficiënten. De berekening van de samenhang van een tiental vragen van een psychologische test afgenomen bij enige tientallen proefpersonen, kostte een kwartier. Een pc voert dezelfde berekening nu uit binnen een seconde. Even spectaculair is de verandering in het gebruiksgemak: bij de ZEBRA werden de gegevens met ponsband ingevoerd en werden de verschillende wegen in een programma gekozen door het omzetten van een schakelaar. Nu kan de gebruiker menugestuurd gegevens invoeren en keuzen opgeven. Een andere belangrijke ontwikkeling is dat in vroeger jaren voor elke methode een apart programma bestond met specifieke eisen aan de vorm waarin de data moesten worden ingevoerd, terwijl in de huidige tijd vrijwel alle statistische berekeningen met één enkel pakket, bijvoorbeeld SPSS, kunnen worden uitgevoerd. Door de grote inspanning die het kostte een programma op de ZEBRA te ontwikkelen en de beperkte geheugen- en rekencapaciteit, bleef het statistische werk op die machine uiterst beperkt.

#### ZAEHL

In de TR4-tijd, 1964 - 1972, heeft het in Duitsland ontwikkelde programma ZAEHL gezorgd voor een spectaculaire groei van het statistisch rekenen. Achteraf moet dit programma, dat eigenlijk slechts een tabelleer-programma was waaraan de Chi<sup>2</sup>-toets en de exacte toets van Fisher waren toegevoegd, als uiterst beperkt en zeer gebruiksonvriendelijk worden gekwalificeerd. Maar in die tijd -de zestiger jaren- speelde het programma een wezenlijke rol in onderzoek, vooral voor enquêteverwerking. De gebruikers van het programma kwamen veelal uit de hoek van de Sociale Wetenschappen.

### WESP

De overgang in 1972 naar Control Data-computers heeft een significante invloed gehad op het statistisch rekenen bij de RUG. Voor het bepalen van het beleid op dit gebied werd in 1971 een overleg tussen Rekencentrum en gebruikers, later geformaliseerd in de LISTORgroep, gestart. Dit was een vroege vorm van inspraak van de gebruiker. Het resultaat van dit overleg was de ontwikkeling van het algemene statistische pakket GRONSTA (GRONingen STATistiek), later omgedoopt in WESP. Dit is het eerste -en enige- algemene statistische pakket dat ooit in Nederland is ontwikkeld. In die tijd kon dat nog, nu is dit soort omvangrijke programmatuur slechts voorbehouden aan grote, commerciële instellingen.

De overleggroep stelde vast welke statistische methoden in het pakket moesten worden opgenomen, waarna Rekencentrum-medewerkers de stuurtaal voor het pakket definieerden en de programmering verzorgden in ALGOL60, met stukken machinecode. De basis voor het programma werd gevormd door een aantal ALGOL60-programma's die al voor de TR4 waren ontwikkeld, zoals programma's voor correlatie-coëfficiënten, factoranalyse, regressie, de Student t-toets en de Mann-Whitney U-toets. Veel aandacht werd besteed aan controles op correctheid van de ingevoerde gegevens en aan datatransformatie. Variabelen werden gedefinieerd door een (volg)nummer in plaats van een naam en de stuurtaal was wat primitief, maar -voor die tijd- waarlijk eenvoudig te hanteren. Er werd een uitgebreide handleiding voor het gebruik van WESP gemaakt. Zeer spoedig ontstond de behoefte aan cursussen WESP, waarvoor het 'Cursusboek WESP' werd

ontwikkeld. De cursus duurde vier dagen, waarin de opdrachten van WESP werden geleerd en veel praktische oefeningen werden gedaan. Het cursusmateriaal bevatte gegevens van leerlingen uit de laatste klas van de lagere school waarmee de invloed van het sociaal milieu op school- en beroepsloopbanen werd geanalyseerd. Een voorbeeld van een opdracht ging als volgt:  
Toets de volgende hypothesen:

- a. Jongens hebben andere schoolvoorkeuren dan meisjes.
- b. Er is verband tussen de opleiding van de vader en de schoolwens van zijn kind.
- c. Jongens blijven vaker zitten dan meisjes

Het materiaal paste uitstekend in de geest van die tijd, grote aandacht voor sociale aspecten van de maatschappij. De constructie van een variabele ?milieu? in 馱 van de opdrachten leidde soms tot heftige discussies. Een cursist liep weg van de cursus omdat hij het ongepast vond zo over het milieu te praten? .



**De invoer van instructies ging door middel van ponskaarten**

Het ponsen en corrigeren van de ?jobs? kostte veel tijd omdat de fout in de opdrachten -meestal ging het eerst een paar keer verkeerd- pas werd ontdekt na het inlezen, verwerken en printen van de cyclus. De cursisten werden ook ingewijd in de computerwereld door een diabandserie, waarin onder meer allerlei apparatuur werd getoond.

Stelling 7 bij het proefschrift van de Socioloog Peschar luidt: ?Het onderwijs in statistische computerprogrammapakketten dient verder te worden ge egreerd in de studieprogramma's van sociale wetenschappen.? Generaties van studenten hebben in die tijd dan ook hun eerste stappen op het gebied van het statistisch rekenen gedaan met WESP.

WESP ging uit van invoer van instructies en data via ponskaarten. Bij een beetje enqu馱 ging het om verscheidene bakken met kaarten. Een bestand met 1.000 individuen en 400 variabelen betekende zo'n 6.000 ponskaarten. Het werd dan ook dringend aanbevolen een kenmerk in de kaart te ponsen opdat de volgorde met een sorteermachine kon worden hersteld als de bak met kaarten viel. En dat gebeurde nogal eens.

## TTWESP

Al snel ontstond door de opkomst van de terminal behoefte aan een interactieve versie van het pakket, waarmee op een vraag-en-antwoord-manier met het programma kon worden gecommuniceerd. De data werden daarbij via de terminal ingevoerd of -bij voorkeur- eerst op een file gezet. Deze versie werd TTWESP genoemd (TT stond voor Teletype Terminal). Hoe primitief TTWESP (achteraf) ook was, in die tijd van ?batch-verwerking? was het een zeer moderne aanpak.. (TT)WESP werd nog omgezet in Fortran en er is nog jaren gewerkt aan uitbreiding van de mogelijkheden van WESP. Het (TT)WESP-pakket is ook nog gebruikt bij verschillende Nederlandse universiteiten en wetenschappelijke instituten, zelfs bij de universiteit van Stuttgart, met een handleiding in het Duits.

## Wat kon WESP?

De opgenomen statistische methoden dekten goed de behoeften van die tijd. Een overzicht van de statistische methoden die waren opgenomen in het pakket: frequentieverdelingen en kruistabellen, met een Chi2-toets; gemiddelde, standaarddeviatie en co馱ici馱en van scheefheid en kurtosis; een-weg variantieanalyse en de toets van Kruskal-Wallis; parametrische en niet-parametrische correlatie-co馱ici馱en; parti馱 correlatie en factoranalyse met diverse rotatie-methoden en de berekening van factorscores; meervoudige lineaire regressie; student t-toets voor onafhankelijke steekproeven en voor gepaarde waarnemingen; de Tekentoets, de Mann-Whitney U-toets en de Wilcoxon symmetrie-toets; de een- en tweesteekproeven toets van Kolmogorov-Smirnov. Alle methoden hadden voorzieningen voor missing data, ?list-wise? en ?pair-wise?, een aspect waaraan veel aandacht werd besteed. Een sterk punt van (TT)WESP was dat de datamatrix in groepen kon worden onderverdeeld, waarop de analyses afzonderlijk konden worden uitgevoerd.

De datatransformatie omvatte op bestandsniveau het toevoegen van individuen en variabelen aan het databestand en het transponeren van de datamatrix. Er kon een aselechte steekproef uit het databestand worden getrokken. Ook kon missing data worden vervangen door het gemiddelde van de variabele. Op een tamelijk primitieve manier konden variabelen worden getransformeerd met een rekenkundige uitdrukking, bijvoorbeeld de logaritme. Een reeks van variabelen kon worden gesommeerd, minimum en maximum van de

reeks konden worden bepaald en het aantal keren dat de score gelijk was aan een criteriumwaarde. Ook konden variabelen worden gecombineerd tot nieuwe variabelen met behulp van logische operatoren.

### Hoe zag de instructie van WESP eruit?

Om een beeld te geven van het pakket, volgen hier een aantal voorbeelden van instructies.

#### *Inlezen van data*

```
READDATA
filenaam, NCARD=2, IDENT=1, CARDN=3
I9AI99I9ADEI
I9B BBBI9I9
< ponskaarten met scores >
END
```

De data waren geponst in twee kaarten (NCARD=2), met een identificatienummer in de posities 1 en 2 (IDENT=1) en een kaartnummer, 1 resp. 2, in kolom 3 (CARDN=3). Een letter betekent het begin van een veld en de maximale waarde die in de kolom mag voorkomen. I99 betekent dat de maximale waarde van de variabele 999 is, D betekent dat waarden groter dan 4 niet zijn toegestaan. Na het inlezen krijgen de variabelen een volgnummer: 1 - 11.

#### *Kruistabellen*

```
CROSSTAB
3-4, 5, 7, 8, 7, 8, 9 - 4, 5
END
```

Kruistabellen worden berekend voor variabele 3 met 4, 5, 7 en 8 en voor de variabelen 7, 8 en 9 met zowel 4 als 5.

#### *Correlatie-coëfficiënten, per groep en voor de groepen 'gepooled' samengenomen*

```
CORRELATE
GROUP, POOLED
1, 2, 10 - 11
END
```

Voor alle groepen apart en voor de groepen tezamen genomen, wordt de correlatiematrix voor de variabelen 1, 2, 10 en 11 uitgerekend.

#### *Datatransformatie*

```
MANIPULATE
filenaam
RECODE
4 = IF 4=3, 4 THEN 2.
5 = IF 5=3 THEN 2 ELSE IF 5=4, 5 THEN 3.
ADD 12 = 2, 10, 11.
END
```

De codes 3 en 4 worden bij variabele 4 omgezet in de code 2. Bij variabele 5 wordt de code 3 omgezet naar 2 en de codes 4 en 5 naar 3. De (nieuwe) variabele 12 wordt de som van de variabelen 2, 10 en 11. Bij de datatransformatie wordt een nieuwe file gevormd, waaraan een naam moet worden gegeven.

### Bemind, maar ook gehaat

Met WESP zijn door medewerkers en studenten van alle faculteiten zeer veel data geanalyseerd, in de toptijd enige honderden, vaak langdurige, jobs per dag. Met name de sociale wetenschappers maakten daarmee de faam van de kwantitatieve aanpak van hun discipline bij de RUG zeer waar. De ponsmachines werden lange

tijd bezet voor het ponsen van grote databestanden, de kaartlezer deed er vele minuten over voordat de bakken met kaarten waren ingevoerd, de printer was zeer lang bezig met de omvangrijke uitvoer. En de wachtrij voor verwerking door de computer vermeldde veel WESP-jobs. Dit tot ongenoegen van gebruikers van de Faculteit der Wiskunde en Natuurwetenschappen. Een van hen ontwierp dan ook een ?banner-page? met de veelzeggende tekst ?KILL ALL WESP?. Maar WESP heeft ook een wezenlijke bijdrage geleverd aan veel onderzoek. Een greep uit de honderden onderzoeken waarbij WESP is gebruikt: enquêtes zijn verwerkt door sociologen en sociaal-geografen en er zijn psychologische test gevalideerd. Economen hebben CBS-gegevens verwerkt. Bij de afdeling Orthodontie is een onderzoek naar kaakgewrichtafwijkingen bij kinderen geanalyseerd met WESP. De afdeling Longziekten heeft de statistische analyse van diverse onderzoeken bij CARA-patiënten uitgevoerd. Door de afdeling Traumatologie zijn ongevalgegevens verwerkt. Voor de meldkamer van de politie zijn analyses uitgevoerd van binnengekomen meldingen. Totaal zijn er meer dan een half miljoen WESP-jobs op achtereenvolgende Control Data-machines verwerkt. Tot eind tachtiger jaren de concurrentie met het mondiale pakket SPSS, de pc en de inspanning voor conversie naar een nieuwe computer te groot werden. Bij het afscheid vertoonde de ?banner-page? een wespje dat op zijn rug lag ?

▲ Begin pagina ▲

---

[index](#) Pictogram 4