

TOEPASSING VAN HET COMBINEREN VAN ENKELE VERDELINGSVRIJE TOETSEN IN EENVOUDIGE FACTORIELE SCHEMA'S

P. van der Laan *) en J. Weima **)

SAMENVATTING

Voor een aantal praktische situaties, die o.a. in landbouwkundig onderzoek voorkomen, worden enkele snelle en doorzichtige statistische analysemethoden aangegeven. Hierbij wordt gebruik gemaakt van enkele bekende verdelingsvrije procedures. Veronderstellingen over de vorm van de onderliggende verdelingen zijn dan niet nodig. Ook veronderstellingen als onafhankelijkheid en gelijkheid van varianties kunnen soms achterwege gelaten worden.

1. INLEIDING

Een gebied in de statistiek dat reeds grote bekendheid geniet en waarvoor de belangstelling nog steeds groter wordt, is dat van de verdelingsvrije methoden, ook wel parameter vrije methoden genoemd.

Verdelingsvrije methoden zijn ruwweg statistische analysemethoden waarbij de geldigheid niet afhangt van de onderliggende kansverdelingen der waarnemingen. Als deze vorm wel bekend is, wat zelden het geval is, is men vaak in staat om technieken af te leiden en toe te passen, die hierop gebaseerd zijn. Vooral bij meer ingewikkelde proefopzetten, waarvoor (nog) geen verdelingsvrije toets bekend is, zal men soms sneller naar een klassieke methode grijpen dan misschien wel verantwoord is. De onderzoeker zal dan veronderstellingen moeten maken, bv. de veronderstelling dat de verschillende waarnemingen afkomstig zijn uit normale populaties, waarvan de juistheid vaak een vraag is waarop de onderzoeker geen antwoord weet. Verdelingsvrije methoden, indien bekend voor het voorkomende probleem, vormen dus een middel om aan bovenstaande onzekerheid tegemoet te komen.

Een ander voordeel van verdelingsvrije methoden is dat zij, althans voor een groot gedeelte, op eenvoudige wijze zijn te presenteren en zeer doorzichtig zijn. Voor niet te grote aantallen waarnemingen zijn zij vaak snel uit te voeren zonder dat men gebruik hoeft te maken van een computer. Belangrijke aspecten bij het vergelijken van verdelingsvrije en klassieke toetsen zijn natuurlijk geldigheid (werkelijke onbetrouwbaarheid komt overeen met opgegeven onbetrouwbaarheid) en onderscheidingsvermogen.

*) Landbouwhogeschool, Vakgroep Wiskunde, De Dreijen 8, 6703 BC Wageningen.

**) Department of Crop Science, P.O. Box 30197, University of Nairobi, Kenya.

Afwijkingen van bijvoorbeeld normaliteit hebben geen invloed op de geldigheid van een verdelingsvrije toets; echter wel op de geldigheid van een klassieke toets.

Anderzijds zal bij juistheid van de veronderstellingen, nodig voor de geldigheid van de klassieke toets, het onderscheidingsvermogen van klassieke toetsen in het algemeen beter zijn dan dat van verdelingsvrije toetsen.

Bij onjuistheid van de veronderstellingen kan een verdelingsvrije toets daarentegen weer een iets hoger onderscheidingsvermogen hebben dan een klassieke toets. Nog een voordeel van verdelingsvrije toetsen gebaseerd op rangnummers, die in dit artikel toegepast zullen worden, is dat voor geldigheid van de nulhypothese formeel slechts symmetrie van de variabelen in hun simultane verdelingsfunctie geëist hoeft te worden. Vooral in landbouwkundig onderzoek, waarbij vlak bij elkaar gelegen proefveldjes verloot worden over de verschillende behandelingen, is dit een prettige bijkomstigheid. De veelal gebruikte t- en F-toetsen vereisen eigenlijk onafhankelijkheid. Bij de gebruikelijke factoriële schema's in de landbouw kunnen deze toetsen eigenlijk slecht als benaderingen voor de wel geldige permutatietoetsen gebruikt worden. Veelal zijn voor ingewikkelder factoriële proeven en vraagstellingen geen verdelingsvrije toetsen beschikbaar. Het doel van dit artikel is om enkele betrekkelijk eenvoudige bekende verdelingsvrije procedures door middel van combineren toe te passen in iets ingewikkelder situaties. Het streven hierbij is niet een optimale procedure te geven, wel een redelijke analyse die niet te veel tijd in beslag neemt.

In de eerstvolgende paragraaf zullen we een korte samenvatting geven van enkele in het vervolg te gebruiken statistische toetsen waarbij niet naar volledigheid van alle mogelijke informatie gestreefd wordt^{*}. De meeste toetsen zullen voornamelijk gevoelig zijn voor locatieverschillen.

In de dan volgende paragrafen zullen we voor een aantal experimenten waarbij twee factoren optreden, verdelingsvrije analyses bespreken. Hierbij kan de ene factor een indeling naar blokken zijn.

In het algemeen zullen we veronderstellen dat we met continue verdelingen te maken hebben, zodat in principe geen gelijke waarnemingen voorkomen.

In de laatste paragraaf zullen we tenslotte nog enkele slotopmerkingen maken en enkele verwijzingen geven naar andere methoden.

^{*}) Betrekkelijk eenvoudige boeken over verdelingsvrije toetsen zijn

Bradley (1968), Lehmann (1975), Hollander en Wolfe (1973) en Conover (1971).

2. ENKELE BENODIGDE VERDELINGSVRIJE TOETSEN

2.1 *Vergelijking van twee behandelingen met verschuivingsalternatieven: de toets van Wilcoxon.*

Gegeven zijn $X_1, \dots, X_m$ en $Y_1, \dots, Y_n$, twee onafhankelijke steekproeven bestaande uit onderling onafhankelijke waarnemingen, afkomstig van twee continue verdelingen. De te toetsen nulhypothese H_0 houdt in dat deze twee verdelingen identiek zijn.

De toetsingsgrootte W van Wilcoxon is nu gedefinieerd als de som van de rangnummers der Y -waarnemingen, na rangschikking naar opklimmende grootte van alle $N = m+n$ waarnemingen. Deze toets is in het bijzonder gevoelig voor verschuivingen. Voor nadere details (zoals tabellen, normale benadering, continuïteitscorrectie en behandeling van gelijke waarnemingen of knopen) wordt verwezen naar de literatuur (bv. Lehmann (1975)).

De voorwaarde van onafhankelijkheid der waarnemingen is een voldoende voorwaarde. Bij de verdeling van W onder H_0 wordt alleen gebruik gemaakt van het feit dat de rangnummers van de Y 's: $r_1, \dots, r_n$ op te vatten zijn als een aselechte steekproef zonder teruglegging uit de populatie van rangnummers $1, \dots, N$. Hieruit volgt dat we de toets van Wilcoxon kunnen toepassen voor het vergelijken van twee behandelingen A en B in de situatie waarbij de behandeling A toegekend is aan m objecten, aselekt uit N objecten getrokken, en de behandeling B aan de resterende n objecten (het zg. randomisatiemodel). Zie ook de opmerking in de inleiding dat slechts symmetrie in de gezamenlijke verdelingsfunctie geëist wordt. Wel moet hierbij opgemerkt worden dat het "extrapoleren" van een conclusie tot een grotere populatie moeilijk en niet van risico's ontbloot is.

2.2 *Vergelijking van meer dan twee behandelingen met verschuivingsalternatieven: de toets van Kruskal-Wallis.*

Nu worden I onafhankelijke steekproeven ter grootte n_i ($i = 1, \dots, I$) getrokken uit I continue verdelingen. Deze steekproeven bestaan uit onderling onafhankelijke waarnemingen (ook hier is onafhankelijkheid een sterke eis; men kan ook weer aan een randomisatiemodel denken). De nulhypothese houdt in dat alle I verdelingen gelijk zijn. De alternatieve hypothese houdt in dat de verdelingen niet identiek zijn, waarbij dan vooral aan verschuivingen gedacht moet worden.

Een geschikte rangnummertoeets is die van Kruskal - Wallis.

De toetsingsgrootheid K van Kruskal - Wallis is gebaseerd op het gemiddelde rangnummer per behandeling na rangschikking van alle $N = \sum_{i=1}^I n_i$ waarnemingen:

$$K = \frac{12}{N(N+1)} \sum_{i=1}^I n_i \left(\frac{R_i}{n_i} - \frac{1}{2}(N+1) \right)^2$$

met R_i : som van rangnummers behorend bij behandeling i .

Voor nadere details wordt verwezen naar Lehmann (1975). Uitgebreide tabellen zijn te vinden in Iman, Quade en Alexander (1975). Voor $I = 2$ is eenvoudig aan te tonen dat deze toets overeenkomt met de tweezijdige uitvoering van de toets van Wilcoxon.

2.3 *Vergelijking van meer dan twee behandelingen met een geordend alternatief; de toets van Jonckheere.*

In een analoge uitgangssituatie als bij de toets van Kruskal - Wallis kunnen we in plaats van een omnibusalternatief denken aan een geordend alternatief. Een geordend alternatief houdt in dat de tweede verdeling ruwweg naar rechts verschoven is ten opzichte van de eerste verdeling, de derde verdeling ruwweg naar rechts verschoven is ten opzichte van de tweede verdeling, enz. Uiteraard moet er minstens één strikte ongelijkheid van verdelingen zijn.

Een geschikte toets is de toets van Jonckheere met toetsingsgrootheid

$$S = \sum_{i < j} U_{ij}$$

waarbij U_{ij} de toetsingsgrootheid van Wilcoxon voor behandelingen i en j is, gedefinieerd als het aantal malen dat een waarneming bij behandeling i kleiner is dan een waarneming bij behandeling j (eigenlijk toets van Mann - Whitney; Relatie met Wilcoxon: $U_{ij} = W - \frac{1}{2}n_i(n_i+1)$).

Voor $I=2$ gaat de toets, zoals eenvoudig te zien is, over in de éénzijdige toets van Wilcoxon voor twee steekproeven. Voor nadere details wordt weer verwezen naar Lehmann (1975).

3. HET VERGELIJKEN VAN MEERDERE BEHANDELINGEN OP J NIVO'S TEGELIJK

De beschreven toetsen zullen gebruikt worden om schema's met twee (nevengeschikte) factoren te analyseren. De verschillende (vaste) verschijningsvormen van een factor in een experiment heten de trap-pen of klassen. Bij de eerste factor A zullen we denken aan een kwalitatieve behandelingsfactor, zoals variëteiten, bemestingssoorten, geneesmiddelen, enz. Bij de tweede factor B wordt gedacht aan een kwantitatieve of geordende kwalitatieve classificatie factor waarbij de klassen te ordenen zijn, zoals verschillende bemestingsnivo's of verschillende tijdstippen.

Veronderstel dat A uit I verschillende behandelingen bestaat en dat iedere behandeling op J nivo's (factor B) wordt bestudeerd. Aan cel (i,j) doen we n_{ij} metingen en geven met Y_{ijk} de uitkomst van de k-de herhaling aan.

We gaan uit van het model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijk}$$

$$(i = 1, \dots, I; j = 1, \dots, J \text{ en } k(=k_{ij}) = 1, \dots, n_{ij})$$

waarbij μ, α_i, β_j en γ_{ij} respectievelijk de algemene verwachtingswaarde, het effect van de i-de rij, het effect van de j-de kolom en de interactie tussen i-de rij en j-de kolom weergeven.

Verder veronderstellen we

$$\sum_{i=1}^I \alpha_i = 0, \quad \sum_{j=1}^J \beta_j = 0 \text{ en } \sum_i \gamma_{ij} = 0 = \sum_j \gamma_{ij}.$$

In dit hoofdstuk gaan we de nogal strenge nulhypothese toetsen dat alle behandelingen A gelijke effecten hebben voor iedere trap van B.

In termen van parameters krijgen we : $\alpha_i = 0 \forall i$ en $\gamma_{ij} = 0 \forall i, j$.

De procedures die we gebruiken vereisen slechts dat de storingstermen E_{ijk} verwachtingswaarde nul hebben en dat zij per trap van B symmetrisch zijn in hun gezamenlijke verdelingsfunctie.

Er wordt dus geen normaliteit van de storingstermen verondersteld en de varianties van de storingstermen mogen per trap van B verschillen, iets wat in de praktijk nogal eens voorkomt. Een hogere stikstoftrap van bemesting zou bijvoorbeeld best een grotere variantie kunnen veroorzaken.

3.1 Het vergelijken van twee behandelingen op J nivo's.

We veronderstellen dat $I = 2$ en beschouwen het toetsen van de nulhypothese

$$H_0 : \alpha_1 = \alpha_2 = 0; \gamma_{ij} = 0 \quad (i = 1, 2; j = 1, \dots, J)$$

tegen één van de mogelijke alternatieve hypothesen:

$$H_1 : \alpha_1 + \gamma_{11} \geq \alpha_2 + \gamma_{21} ; \dots ; \alpha_1 + \gamma_{1J} \geq \alpha_2 + \gamma_{2J}$$

$$H_2 : \alpha_1 + \gamma_{11} \geq \alpha_2 + \gamma_{21} ; \dots ; \alpha_1 + \gamma_{1J} \geq \alpha_2 + \gamma_{2J}$$

$$\text{of } \alpha_1 + \gamma_{11} \leq \alpha_2 + \gamma_{21} ; \dots ; \alpha_1 + \gamma_{1J} \leq \alpha_2 + \gamma_{2J}$$

$$H_3 : \alpha_1 + \gamma_{11} \neq \alpha_2 + \gamma_{21} ; \dots ; \alpha_1 + \gamma_{1J} \neq \alpha_2 + \gamma_{2J}$$

met steeds tenminste één strikte ongelijkheid.

Op ieder nivo j van factor B rangschikken we nu de $n_{1j} + n_{2j}$ waarnemingen en geven deze rangnummers. Zij nu W_j de som van de rangnummers behorend bij behandeling 1 op nivo j .

Voor het toetsen van H_0 tegen H_1 of H_2 nemen we als toetsingsgrootheid

$$T_1 = W_1 + \dots + W_k.$$

Bij beschouwing van H_1 dient H_0 verworpen te worden voor grote waarden van T_1 . Bij beschouwing van H_2 dient H_0 verworpen te worden voor grote en kleine waarden van T_1 .

Kritieke waarden van deze toetsingsgrootheid zijn getabelleerd voor het geval dat alle n_{ij} 's gelijk zijn en $n_{ij} = 1(1)30$ en $J = 1(1)12$. Voor deze tabellen wordt verwezen naar Nelson (1970).

Voor ongelijke n_{ij} 's kan, voor grote waarden van k en/of n_{ij} voor alle i en j gewerkt worden met de normale benadering. Hiervoor moet men berekenen:

$$E\{T_1\} = \sum_{j=1}^J \frac{1}{2} n_{1j} (n_{1j} + n_{2j} + 1)$$

$$\text{var}\{T_1\} = \sum_{j=1}^J \frac{1}{12} n_{1j} n_{2j} (n_{1j} + n_{2j} + 1).$$

Opmerking:

Voor ongelijke n_{ij} 's kan men soms een onderscheidender toets krijgen door niet slechts de W_i 's bij elkaar op te tellen, maar een gewogen som van de W_i 's te nemen. Voor een nadere bespreking en details wordt verwezen naar Van Elteren (1960).

Voor het toetsen van H_0 tegen H_3 nemen we als toetsingsgrootheid

$$T_2 = \sum_{j=1}^J \frac{12}{n_{1j} n_{2j} (n_{1j} + n_{2j} + 1)} \{W_j - \frac{1}{2} n_{1j} (n_{1j} + n_{2j} + 1)\}^2.$$

Deze grootheid heeft, onder H_0 , voor grote waarden van n_{1j} en n_{2j} ($j = 1, \dots, J$) bij benadering een chikwadraatverdeling met J vrijheidsgraden. Kritieke waarden kunnen daarvoor uit de geschikte tabellen gehaald worden.

3.2 Vergelijken van I behandelingen op J nivo's.

De nulhypothese luidt nu: alle I behandelingen gelijk voor alle nivo's, dus:

$$H_0 : \alpha_i = 0 \text{ en } \gamma_{ij} = 0 \text{ voor } i = 1, \dots, I; j = 1, \dots, J.$$

We beschouwen de alternatieve hypothesen H_1 en H_2 . Hierbij is H_1 gedefinieerd als

$$H_1 : \alpha_{i_1} + \gamma_{i_1 j} \geq \dots \geq \alpha_{i_I} + \gamma_{i_I j} \quad (j = 1, \dots, J)$$

voor een van te voren gespecificeerde permutatie $\{i_1, \dots, i_I\}$ van $\{1, \dots, I\}$. Hierbij moet minstens één van de ongelijkheden strikt zijn.

Verder is H_2 gedefinieerd als:

$$H_2 : \alpha_{i_1} + \gamma_{i_1 j} \geq \dots \geq \alpha_{i_I} + \gamma_{i_I j} \quad (j = 1, \dots, J)$$

voor een niet gespecificeerde permutatie $\{i_1, \dots, i_I\}$ van $\{1, \dots, I\}$ die voor iedere j weer anders mag zijn ($j = 1, \dots, J$). Ook hier moet minstens één van de $J(I-1)$ ongelijkheden strikt zijn.

H_1 houdt dus in dat er een (van te voren bekende) ordening in de behandelingen zit, die voor alle nivo's geldt. H_2 is erg algemeen, de behandelingen zijn niet alle gelijk en per nivo kan de ongelijkheid wel heel verschillend zijn.

Voor het uitvoeren van deze toetsingsproblemen rangschikken we op ieder nivo j de $n_{1j} + \dots + n_{Ij}$ waarnemingen. Voor het toetsen van H_0 tegen H_1 rekenen we de toetsingsgrootheid van Jonckheere uit op ieder nivo j (zie par. 2.3). Neem nu als toetsingsgrootheid

$$T_3 = S_1 + \dots + S_J.$$

Ook hier lijkt een nader onderzoek over een gewogen som van de toetsingsgroottheden ter verhoging van het onderscheidingsvermogen gewenst.

H_0 dient verworpen te worden voor grote waarden van T_3 . Kritieke waarden dan wel overschrijdingskansen kunnen bepaald worden met behulp van de normale benadering (voor grote waarden van J en/of n_{ij} voor alle i en j).

Onder H_0 geldt

$$E\{T_3\} = \sum_{j=1}^J \frac{1}{4} (N_j^2 - \sum_{i=1}^I n_{ij}^2)$$

en

$$\text{var}\{T_3\} = \sum_{j=1}^J \frac{1}{72} \{N_j^2 (2N_j + 3) - \sum_{i=1}^I n_{ij}^2 (2n_{ij} + 3)\}$$

waarbij

$$N_j = \sum_{i=1}^I n_{ij} \quad (j = 1, \dots, J).$$

Voor het toetsen van H_0 tegen H_2 kennen we op ieder nivo j rangnummers toe aan de waarnemingen en bepalen daar de grootheid van Kruskal - Wallis K_j (zie par. 2.2). Neem nu als toetsingsgrootheid

$$T_4 = \sum_{j=1}^J K_j.$$

Onder H_0 heeft T_4 , voor grote waarden van n_{ij} voor alle i en j , bij benadering een chikwadraatverdeling met $J(I-1)$ vrijheidsgraden.

De nulhypothese H_0 dient verworpen te worden voor grote waarden van T_4 . In de hoofdstuk hebben we de zeer beperkte nulhypothese getoetst, dat de behandelingen gelijk zijn en wel voor elke trap van factor B. Bij verworping van deze nulhypothese kan alleen geconstateerd worden dat sommige behandelingen een globale tendens hebben in de richting van een beter resultaat dan andere (in geval van alternatief H_1) of dat er op één of andere manier verschillen zijn bij tenminste één nivo van B (in geval van alternatief H_2). Waardoor deze verschillen ontstaan, hoofdeffect of interactie of beide, wordt niet gespecificeerd door de toetsingsgroottheden. Er kan bijvoorbeeld in het geval van H_1 , geconcludeerd worden dat tenminste één behandeling over de hele linie van ingestelde nivo's van B beter is dan de andere, maar niet of op sommige nivo's van B het verschil groter is dan bij andere nivo's (wat overeenkomt met interactie). Bij aanwezigheid van interactie geldt uiteraard sterk de voorwaarde "van ingestelde nivo's van B".

Als in werkelijkheid alleen hoofdeffecten een rol spelen (en er dus geen interactie is), zodat de volgorde in H_2 voor iedere j hetzelfde is, is de toets T_4 ook wel gevoelig. Er zijn dan echter wel geschiktere rangnummer-toetsen beschikbaar, namelijk de toets van Friedman (bij gelijk aantal waarnemingen per cel) en de toets van Bernard en Van Elteren (1954) of die van De Kroon en Van der Laan (1983) voor ongelijk aantal waarnemingen per cel. Bij gezamenlijke werking van hoofdeffecten en interacties heeft het geen zin om te toetsen of de hoofdeffecten nul zijn (dus te beschouwen $H_0 : \alpha_i = 0$ voor $i = 1, \dots, I$). Wel heeft het zin om te toetsen of er geen interactie is.

In het volgende hoofdstuk worden enkele snelle verdelingsvrije procedures besproken voor deze meer samengestelde nulhypothese ($\gamma_{ij} = 0 \quad \forall i, j$).

4. HET TOETSEN VAN INTERACTIE

We gaan weer uit van hetzelfde model als in hoofdstuk 3 maar toetsen nu $H_0 : \gamma_{ij} = 0 \quad \forall i \text{ en } j$ tegen enkele alternatieven.

Net als in hoofdstuk 3 maken we weer onderscheid voor de gevallen $I = 2$ en $I > 2$.

4.1. $I = 2$

We beschouwen het toetsen van de nulhypothese:

$$H_0 : \gamma_{ij} = 0 \quad \text{voor } i = 1, 2; j = 1, \dots, J$$

of in woorden: de responsiecurven van de beide behandelingen lopen evenwijdig

tegen de volgende alternatieve hypothesen:

$$H_1 : \gamma_{1j} \text{'s } (j = 1, \dots, J) \text{ zijn niet allemaal gelijk}$$

of in woorden: de responsiecurven lopen niet evenwijdig

$$H_2 : \gamma_{11} \leq \gamma_{12} \leq \dots \leq \gamma_{1J}$$

met tenminste één strikte ongelijkheid. In woorden: de responsiecurven gaan steeds verder uit elkaar lopen.

Voor de nu volgende analyse veronderstellen we dat $n_{1j} = n_{2j}$ ($j = 1, \dots, J$) en dat zowel de storingstermen E_{1jk} ($j = 1, \dots, J; k = 1, \dots, n_{1j}$) identiek verdeeld zijn als de storingstermen E_{2jk} ($j = 1, \dots, J; k = 1, \dots, n_{2j}$).

We constateren dat we nu onder andere gelijkheid van varianties van de storingstermen binnen rijen veronderstellen in tegenstelling tot de veronderstelling in het vorige hoofdstuk.

We gaan nu door middel van loten paren vormen (binnen de kolommen) van een waarneming van behandeling 1 met één van behandeling 2.

(Voor de eenvoud veronderstellen we even dat de waarnemingen in een cel in een willekeurige, d.i. aselechte volgorde staan, zodat het niet expliciet nodig is te loten).

Definieer nu de volgende variabelen:

$$V_{jk} = Y_{1jk} - Y_{2jk} \quad \text{met } j = 1, \dots, J; k = 1, \dots, n_{1j}.$$

Op deze wijze krijgen we J series grootheden en er geldt

$$V_{jk} \stackrel{\sim}{=} (\alpha_1 - \alpha_2) + (Y_{1j} - Y_{2j}) + (E_{1jk} - E_{2jk}).$$

Onder $H_0 (Y_{1j} = Y_{2j} = 0)$ hebben al deze verschillen dus dezelfde verdeling.

Onder H_1 of H_2 zijn er verschuivingen van kolommen t.o.v. elkaar.

Voor het toetsen van H_0 tegen H_1 rangschikken we de $n_{11} + \dots + n_{1J}$ verschillen en kennen rangnummers toe. Daarna voeren we de toets van Kruskal - Wallis uit voor J onafhankelijke steekproeven. De toetsingsgrootheid wordt aangegeven met T_5 . Voor kleine waarden van J en n_{1j} kan voor het bepalen van kritieke waarden van exacte tabellen voor deze toets gebruik worden gemaakt. Voor grotere aantallen heeft T_5 onder H_0 bij benadering een chikwadraatverdeling met $J-1$ vrijheidsgraden. In het geval $J = 2$ kunnen we de tweezijdige uitvoering van de toets van Wilcoxon gebruiken.

Voor het toetsen van H_0 tegen H_2 kunnen we op de J groepen verschillen de toets van Jonckheere toepassen. De toetsingsgrootheid geven we aan met T_6 . Ook hiervoor zijn exacte tabellen en benaderingen beschikbaar. Voor het geval $J = 2$ kan de éénzijdige uitvoering van de tweesteekproeventoets van Wilcoxon gebruikt worden.

In het geval dat de storingstermen per kolom dezelfde verdeling hebben terwijl de veronderstelling van gelijke verdelingen binnen rijen niet geldt kunnen we het schema op een andere manier analyseren. Een voldoende voorwaarde hiervoor is dat de aantallen waarnemingen per cel binnen de eerste rij gelijk zijn (stel n_1) en ook de aantallen binnen de tweede rij gelijk zijn (stel n_2).

Veronderstel verder dat de factor B een even aantal nivo's heeft, dus $J = 2M$ met M een geheel getal (indien J oneven is, dan één nivo weglaten. Dit moet gebeuren buiten de waarnemingen om. Voor een kwantitatieve faktor zal dit vaak het middelste nivo zijn).

Voor beide behandelingen loten we nu waarnemingen van nivo 1 en nivo $M + 1$ bij elkaar. In sommige situaties kan het beter zijn om op een andere manier nivo's bij elkaar te kiezen. Dat verandert echter niets aan de essentie van de door ons voorgestelde procedure. We berekenen de volgende verschillen (aannemende dat de waarnemingen in iedere cel al in aselechte volgorde staan):

$$\text{behandeling 1: } Y_{1(M+1)1} - Y_{111}, \dots, Y_{1(M+1)n_1} - Y_{11n_1}$$

$$\text{behandeling 2: } Y_{2(M+1)1} - Y_{211}, \dots, Y_{2(M+1)n_2} - Y_{21n_2}.$$

We rangschikken deze $n_1 + n_2$ verschillen en kennen rangnummers $1, \dots, n_1 + n_2$ toe. Zij nu W_{11} de som van de n_1 rangnummers bij behandeling 1. Hetzelfde doen we bij de nivo's 2 én $M+2$, waarbij we de som van de rangnummers bij behandeling 1, aangeduid met W_{12} , berekenen. Analoog krijgen we $W_{13}, \dots, W_{1M}$. Voor het toetsen van H_0 tegen H_1 nemen we dan als toetsingsgrootheid:

$$T_7 = \sum_{j=1}^M \frac{12}{n_1 n_2 (n_1 + n_2 + 1)} (W_{1j} - \frac{1}{2} n_1 (n_1 + n_2 + 1))^2$$

welke onder H_0 bij benadering, voor grote waarden van n_1 en n_2 , een chikwadraatverdeling heeft met M vrijheidsgraden. Grote waarden van T_7 leiden tot verwerpen van H_0 .

Voor het toetsen van H_0 tegen H_2 nemen we als toetsingsgrootheid

$$T_8 = \sum_{j=1}^M W_{1j}.$$

Voor tabellen en benaderingen, zie par. 3.1. Grote waarden van T_8 leiden tot verwerpen van H_0 .

4.2 $I > 2$

Vele procedures zijn in de literatuur al voorgesteld voor deze situatie. De meeste zijn alleen asymptotisch of conditioneel verdelingsvrij. Bovendien zijn ze vaak niet erg inzichtelijk en wordt veel rekenwerk vereist. Op deze procedures wordt nader ingegaan in hoofdstuk 6.

In deze paragraaf stellen we een procedure voor die verband houdt met wat in de vorige hoofdstukken gedaan is en die echt verdelingsvrij is. Hierbij moet verondersteld worden dat de storingstermen per kolom dezelfde verdeling hebben en dat de aantallen waarnemingen binnen iedere rij gelijk zijn. De nulhypothese luidt

$$H_0 : \gamma_{ij} = 0 \quad \text{voor} \quad i = 1, \dots, I; j = 1, \dots, J$$

in woorden: de I responsiecurven lopen evenwijdig

en de te beschouwen alternatieve hypothesen zijn

H_1 : de responsiecurven lopen niet evenwijdig

en

H_2 : de responsiecurven gaan steeds verder uit elkaar lopen, in de zin van dat de verschillen tussen de responsiecurven bij de behandelingen i en j ($i < j$) steeds groter worden.

Veronderstel weer dat J even is (anders één nivo weglaten) en zij $J = 2M$. Nu wordt bij iedere behandeling i voor elke waarneming binnen nivo j een waarneming uit nivo $M+j$ geloot en het verschil bepaald. Dit wordt gedaan voor $i = 1, \dots, I$. Zo krijgen we I groepjes verschillen van omvang $n_1, \dots, \text{resp. } n_I$.

Voor het alternatief H_1 berekenen we de toetsingsgrootheid van Kruskal-Wallis. Dit doen we voor iedere $j \in \{1, \dots, M\}$. De M verkregen toetsingsgrootheden tellen we nu bij elkaar op. De resulterende som T_9 heeft onder H_0 bij benadering, voor grote waarden van n_i ($i = 1, \dots, I$), een chikwadraatverdeling met $M(I-1)$ vrijheidsgraden. Voor het alternatief H_2 berekenen we voor iedere j de toets van Jonckheere op de verschillen. De M verkregen

toetsingsgrootheden worden daarna bij elkaar opgeteld. Met behulp van de normale benadering, voor grote n_i ($i = 1, \dots, I$) en/of grote M , kunnen voor de zo verkregen toetsingsgrootheid T_{10} weer rechteroverschrijdingskansen berekend worden.

5. VOORBEELD

Een onderzoeker wenst de invloed van stikstofbemesting op de groei van de bladeren van de suikerbiet na te gaan. Als factor A is gekozen de stikstofvorm (ammoniumsulfaat, kalksalpeter) en als factor B het bemestingsnivo (50, 100, 150 en 200 kg/ha).

Alle combinaties van de twee factoren worden in 6-voud toegepast. Hiervoor is een homogeen proefveld beschikbaar met 48 veldjes. De toewijzing van de veldjes aan de factorcombinaties gebeurt door loting.

Op een bepaald tijdstip werd op elk veldje het gemiddeld bladoppervlak per plant bepaald. In de volgende tabel zijn de (fictieve) waarnemingen weergegeven (in cm^2/plant).

	50		100		150		200	
ammoniumsulfaat	2158	2301	2441	2510	2234	2304	2169	2283
	2253	2341	2498	2532	2267	2288	2118	2251
	2279	2199	2583	2668	2405	2193	2301	2143
kalksalpeter	2340	2215	2568	2843	2689	2668	2731	2698
	2368	2150	2612	2621	2401	2749	2587	2641
	2381	2312	2910	2740	2651	2765	2771	2862

We toetsen bijvoorbeeld eerst H_0 : "geen verschil tussen de behandelingen" tegen alternatief H_2 in par. 3.1.

Rangnummers levert op:

	50		100		150		200	
ammoniumsulfaat	2	7	1	3	2	5	3	5
	5	10	2	4	3	4	1	4
	6	3	6	9	7	1	6	2
kalksalpeter	9	4	5	11	16	9	10	9
	11	1	7	8	6	11	7	8
	12	8	12	10	8	12	11	12

Zij T_1 : som van de rangnummers behorend bij ammoniumsulfaat.

Dit levert als uitkomst 101 op.

Onder H_0 geldt: $E\{T_1\} = 4 * \frac{1}{2} * 6 * (6 + 6 + 1) = 156$

$$\text{var}\{T_1\} = 4 * \frac{1}{12} * 6 * 6 * 13 = 156.$$

Tweezijdige overschrijdingskans : $2P(\underline{X} < \frac{101 - 156}{\sqrt{156}}) = 2P(\underline{X} > 4.40) \approx 0.000.$

H_0 wordt verworpen. Er is een zeer duidelijke aanwijzing dat H_2 juist is. De getallen geven de indruk dat kalksalpeter het over de hele linie beter doet dan ammoniumsulfaat.

We kunnen nog onderzoeken of er interactie bestaat tussen de twee factoren. We zullen toetsen tegen het in paragraaf 4.1 genoemde alternatief H_1 . We veronderstellen gelijkheid van de verdelingen van de storingstermen binnen rijen. Voor het model wordt verwezen naar paragraaf 3. We berekenen nu in iedere kolom de verschillen van een waarneming met kalksalpeter en één met ammoniumsulfaat (de waarnemingen staan reeds in gelote volgorde).

	50	100	150	200
verschillen	182(11)	127(9)	455(18)	562(22)
	115(8)	114(7)	134(10)	469(20)
	102(5)	327(13)	246(12)	470(21)
	-86(2)	333(14)	364(15)	415(17)
	-91(1)	89(4)	461(19)	390(16)
	113(6)	72(3)	572(23)	719(24)

Tussen haakjes staan de rangnummers na rangnummering van de 24 verschillen naar opklimmende volgorde.

De toetsingsgrootheid T_5 uit paragraaf 4.1 heeft dan als uitkomst

$$\frac{12}{24(24+1)} * \frac{1}{6} \{ (33-75)^2 + (50-75)^2 + (97-75)^2 + (120-75)^2 \} = 16.33.$$

Rechterkritieke waarde bij $\gamma = 0.005$ gebaseerd op de χ^2_3 - benadering is 12.8. H_0 wordt verworpen. Er is duidelijk aangetoond dat er interactie bestaat tussen de 2 factoren.

Het lijkt erop dat het verschil tussen kalksalpeter en ammoniumsulfaat steeds groter wordt naarmate het bemestingsnivo toeneemt.

Waren we van te voren van dit alternatief (H_2) uitgegaan dan was het beter geweest om T_6 te gebuiken. Bij veronderstelling van gelijke verdeling van de storingstermen binnen bemestingsnivo's zouden we T_7 of T_8 moeten gebruiken.

6. DISCUSSIE

Het combineren van toetsen, die voor eenvoudige situaties zijn afgeleid, leidt tot vrij inzichtelijke procedures in meer ingewikkelde schema's.

Aan het verdelingsvrij toetsen op interactie zoals beschreven in hoofdstuk 4 is veel aandacht besteed in de literatuur.

Mehra en Sen (1969) stellen een procedure voor, waarbij de waarnemingen eerst getransformeerd moeten worden. Deze toets is slechts voorwaardelijk verdelingsvrij. Bhapkar en Gore (1974) komen met een procedure die alleen asymptotisch verdelingsvrij is.

Lemmer en Stoker (1967) werkten met procedures die gebaseerd waren op rangnummers die verkregen waren na het rangschikken van de hele verzameling waarnemingen. Dergelijke toetsen zijn ook slechts asymptotisch verdelingsvrij en bovendien erg gevoelig voor hoofdeffecten.

Het in hoofdstuk 4 gebruikte lotingsprincipe is ook gebruikt door Crouse (1960) voor een 2^k -proef.

De Kroon en Van der Laan (1981) geven een toets die alleen geschikt is voor een bepaald soort interactie (de zogenaamde rank-interaction). De door ons gebruikte procedures in hoofdstuk 4 zijn echt verdelingsvrij en onbetrouwbaarheid en onderscheidingsvermogen zijn niet afhankelijk van eventuele hoofdeffecten. Men kan bezwaren hebben tegen het bij elkaar loten van waarnemingen, daar verschillende onderzoekers met hetzelfde getallenmateriaal tot verschillende conclusies kunnen komen. Een loting kan zo "ongelukkig" uitvallen dat significantie wordt bereikt

terwijl de proefuitkomsten het tegendeel suggereren en omgekeerd. In diverse landen zijn overheidsinstanties eerder geneigd conclusies van onderzoekswerk door de industrie te accepteren als er een navolgbaar "gegevensspoor" is, inclusief het narekenbaar zijn van statistische berekeningen.

Ook zou men kunnen denken dat door het bij elkaar loten van waarnemingen veel informatie overboord gegooid wordt waardoor het onderscheidingsvermogen klein zal zijn in vergelijking met de gebruikelijke F-toets.

Om van dit laatste een indruk te krijgen hebben we de toets T_5 vergeleken met de klassieke F-toets op interactie in het geval dat de storingstermen allen dezelfde verdeling hebben. Dit is voor een aantal verschillende onderliggende verdelingen gebeurd. Een vergelijking is gemaakt met behulp van de Asymptotische Relatieve Efficiency (A.R.E.) van de twee toetsen ten opzichte van elkaar.

Als bijvoorbeeld de A.R.E. van twee toetsen T_1 en T_2 gelijk is aan 1.5 betekent dit dat voor grote steekproeven in de "buurt" van de nulhypothese ongeveer 1.5. keer zoveel waarnemingen nodig zijn voor T_2 als voor T_1 om hetzelfde onderscheidingsvermogen te bereiken bij gelijke onbetrouwbaarheid. Voor preciezere definitie zie Lehmann (1975) of Randless en Wolfe (1979).

Voor een aantal verdelingen van de storingstermen zijn in de volgende tabel de A.R.E.'s van T_5 t.o.v. de klassieke F-toets voor een 2 * J-tabel weergegeven.

storingsterm	uniform	normaal	logistisch	dubbel exponentieel	exponentieel
A.R.E. (T_5, F)	0.889	0.955	1.026	1.172	1.500

Hieruit blijkt dat zelfs in de meest ideale situatie voor de F-toets (nl. onderliggende verdeling normaal) het verlies van T_5 ten opzichte van F slechts in de orde van 4.5% ligt. Bij afwijkingen van normaliteit kan onze lotingsprocedure het zelfs beter doen dan de F-toets. Ervaringen met andere vergelijkingen d.m.v. A.R.E. hebben geleerd dat de A.R.E. ook een duidelijke indicatie geeft voor het onderscheidingsvermogen bij kleine steekproefomvang en/of verder van de nulhypothese verwijderde alternatieven. Niettemin zou een simulatieonderzoek voor kleine steekproefomvang erg nuttig zijn. We hopen in een ander artikel hierop terug te komen.

LITERATUUR

- Benard, A. and Van Elteren, Ph. (1954). A generalization of the method of m rankings,
Indagationes Mathematicae 16, 358 - 369.
- Bhappkar, V.P. and Gore, A.P. (1974). A nonparametric test for interaction in two-way layouts,
Sankhyā, 36 A, 261 - 272.
- Bradley, J.V. (1968). *Distribution-free Statistical Tests*,
 Prentice-Hall, Englewood Cliffs, N.J.
- Conover, W. (1971). *Practical nonparametric statistics*,
 Wiley & Sons, New York - London.
- Crouse, C.F. (1968). A distribution-free method of analyzing a 2^m factorial experiment,
S. Afr. Statist. J., 2, 101 - 108.
- Elteren, Ph. van (1960). On the combination of independent two-sample tests of Wilcoxon,
Bulletin Inst. Intern. Statist. 37, 351 - 361.
- Hollander, M. and Wolfe, D.A. (1973). *Nonparametric statistical methods*.
 Wiley & Sons, New York - London.
- Iman, R.L., Quade, D. and Alexander, D. (1975). Exact probability levels for the Kruskal - Wallis test,
Selected Tables in Math. Statistics, Vol. 3(H.L. Harter & D.B. Owen, Eds.).
- Kroon, J. de and Laan, P. van der (1981). Distribution-free test procedures in two-way layouts; a concept of rank-interaction,
Statistica Neerlandica 35, 189 - 213.
- Kroon, J. de and Laan, P. van der (1983). A generalization of Friedman's rank statistic.
Statistica Neerlandica 37, 1 - 14.
- Lehmann, E.L. (1975). *Nonparametrics: Statistical methods based on ranks*.
 McGraw-Hill International Book Company, New York - London.
- Lemmer, H.H. and Stoker, D.J. (1967). A distribution-free analysis of variance for the two-way classification,
S. Afr. Statist. J., 1, 67 - 74.

- Mehra, K.L. and Sen, P.K. (1969). On a class of conditionally distribution-free tests,
Ann. Math. Statist. 40, 658 - 664.
- Nelson, L.S. (1970). Tables for Wilcoxon's rank sum test in randomized blocks.
Journal of Quality Technology 2, 207 - 218.
- Randles, R.H. and Wolfe, D.A. (1979). Introduction to the theory of nonparametric statistics,
Wiley & Sons, New York.