MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 125

Statistische analysemethoden voor wasproeven
(Vertrouwelijk)

door

Constance van Eeden

1953

1. Inleiding.

Om twee wasmiddelen V en C te vergelijken, werden door het Proefstation voor de Wasindustrie de volgende proeven uitgevoerd:

1. In ieder van de twee wasmiddelen werd een gelijk aantal gemiddeld even vuile keukendoeken gewassen. Na het wassen werd de vuilheidsgraad van deze doeken beoordeeld door ze in gedachten in acht velden te verdelen (vier op iedere zijde der doek) en het aantal niet geheel schoongewassen achtste doeken te tellen.
   Deze proef werd 24 maal uitgevoerd; de keukendoeken die voor verschillende wasproeven gebruikt zijn, zijn niet altijd even vuil. Het verschil is soms aanzienlijk. De aantallen keukendoeken, die per proef en per wasmiddel gewassen zijn, zijn bekend en variëren tussen 10 en 44.
2. In ieder van de twee wasmiddelen werden vier kunstmatig vuilgemaakte doeken, die vóór het wassen gemiddeld dezelfde helderheid vertoonden, gewassen. Na het wassen werd de helderheid van deze doeken opnieuw gemeten.
   De proef werd 12 maal uitgevoerd; de helderheid van de doeken, die voor de verschillende proeven gebruikt werden, was echter niet altijd dezelfde. Verder werden deze proeven nog uitgevoerd voor twee verschillende soorten kunstmatig vuilgemaakte doeken, die aangeduid worden als serie P en serie B. De helderheid van de doeken vóór het wassen was voor deze twee series verschillend.

In dit rapport worden enkele methoden beschreven om te onderzoeken of er een verschil is tussen de twee wasmiddelen.

2. Methode van onderzoek.

Indien de doeken, die bij één wasproef gewassen zijn, vóór het wassen niet te veel in vuilheidsgraad verschillen, kan men, zoals bij de in de inleiding beschreven proeven gedaan is, deze doeken systematisch over de twee wasmiddelen verdelen, zodanig dat de doeken die met het ene wasmiddel gewassen zijn gemiddeld even vuil zijn als degene, die met het andere wasmiddel gewassen zijn. Om te onderzoeken of er een verschil is tussen de twee wasmiddelen kan men nu als volgt de toets van WILCOXON (zie bijlage S 47 (M 7)) toepassen:

Voor ieder der wasproeven heeft men een aantal waarnemingen (een steekproef) van de vuilheidsgraad na het wassen (nl. voor iedere doek het aantal niet geheel schoongewassen achtste doeken resp. de helderheid der doeken) voor ieder der wasmiddelen.

Op de twee steekproeven van één wasproef passen we nu de toets van WILCOXON toe. Dit geeft voor iedere wasproef:

1. een waarde voor de toetsingsgrootheid $\underline{u}$,
2. een waarde voor μ ,
3. een waarde voor σ^2,

die we resp. aangeven met u_i , μ_i en σ_i^2 ; de index i duidt het nummer der proef aan.

We berekenen nu de grootheid:

$$\underline{W} = \sum_i \frac{u_i - \mu_i}{N_i} ,$$

waarin N_i het aantal doeken is dat bij de i^e wasproef voor de twee wasmiddelen tezamen gebruikt is.

Als de te toetsen hypothese H_o, inhoudende dat er geen verschil is tussen de twee wasmiddelen, juist is, is $\underline{W}$ bij benadering normaal verdeeld met gemiddelde $\mu = 0$ en variantie:

$$\sigma^2 = \sum_i \frac{\sigma_i^2}{N_i^2} .$$

$\frac{\underline{W}}{\sigma}$ is dus bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

De tweezijdige kritieke zone bestaat uit grote waarden van $\frac{|W|}{\sigma}$ 1).

Indien echter de doeken, die bij één wasproef gebruikt zijn, vóór het wassen veel in vuilheidsgraad verschillen, kan men, als men de bovenbeschreven toets wil toepassen, beter deze doeken aselect over de twee wasmiddelen verdelen, daar een systematische verdeling van de doeken het onderscheidingsvermogen van de toets verkleint. Men kan in dit geval echter ook de doeken systematisch over de twee wasmiddelen verdelen (en wel weer zodanig dat de doeken, die met het ene wasmiddel gewassen worden vóór het wassen gemiddeld even vuil zijn als degene, die met het andere wasmiddel worden gewassen) en dan als volgt de symmetrietoets toepassen:

Men berekent voor iedere wasproef en voor ieder wasmiddel de som (of het gemiddelde) van de bij die wasproef en dat wasmiddel gedane waarnemingen. Vervolgens berekent men voor iedere wasproef het verschil van deze sommen (of gemiddelden) voor de twee wasmiddelen.

Als de te toetsen hypothese H_o, inhoudende dat er geen verschil is tussen de twee wasmiddelen, juist is, zijn deze ver-

---

1) Ter verdere toelichting is memorandum S 102 (M 17b) aan dit verslag toegevoegd.

schillen symmetrisch om nul verdeeld. Deze symmetrie kan men toetsen met behulp van de symmetrietoets T_2 (zie bijlage S 47 (M 10)) [2]. Noemen wij, bij één wasproef, het resultaat (totaal of gemiddeld) van het ene wasmiddel X en van het andere Y, dan kan men de symmetrietoets ook toepassen op de grootheden $\frac{X-Y}{X+Y}$ in plaats van op $X-Y$ zelf. Deze methode valt aan te bevelen als het verschil in vuilheidsgraad na het wassen (dus $X+Y$) bij de verschillende proeven nog groot is in vergelijking met de voor $X-Y$ gevonden waarden. Hierbij kan men voor X en Y in het geval van de theedoeken het totale aantal geheel schone achtste doeken of ook het totale aantal niet schone achtste doeken nemen. Bij toepassing van de toets op $X-Y$ maakt dit geen verschil; bij toepassing op $\frac{X-Y}{X+Y}$ is de uitkomst niet geheel dezelfde, maar in de regel zal het verschil niet groot zijn. Zijn de aantallen doeken in twee te vergelijken groepen (dus bij één wasproef) niet gelijk, dan neme men voor X en Y niet totale aantallen, maar gemiddelden. Het is echter beter dit zoveel mogelijk te vermijden.

Opmerking.

De bovenbeschreven toetsen kan men alleen dan op deze wijze toepassen als er geen systematisch verschil is tussen de twee wasketels.

Werkt men met twee wasketels, waartussen wel een verschil bestaa dan zal men er bij iedere wasproef om moeten loten welke wasketel men voor het ene en voor het andere wasmiddel gebruikt.

In dit geval kan men echter de toets van WILCOXON niet meer toepassen, de symmetrietoets wel. Bij toepassing van de toets van WILCOXON zou dan nl. de kans op een foute conclusie groter kunnen worden dan de opgegeven onbetrouwbaarheidsdrempel.

3. Resultaten.

Passen we de in par. 2 beschreven toetsen toe op het ons verstrekte waarnemingsmateriaal dan vinden we:

---

2) Niet gepubliceerde tabellen van deze toets zijn op het Mathematisch Centrum aanwezig en kunnen schriftelijk geconsulteerd worden.

Tabel I

Vergelijking der wasmiddelen met behulp van de toets van WILCOXON

	serie	overschrijdingskans [3]
kunstmatig vuilgemaakte doeken	P	0,43 +
	B	< 0,0001 +
theedoeken	-	0,46 +

Tabel II

Vergelijking der wasmiddelen met behulp van de symmetrietoets

	serie	overschrijdingskans [4]
kunstmatig vuilgemaakte doeken	P	0,21 +
	B	0,003 +
theedoeken	-	0,40 +

We zien dus dat de twee toetsen in dit geval dezelfde resultaten geven: voor de kunstmatig vuilgemaakte doeken, serie B, vinden we dat wasmiddel C beter is dan V, terwijl de resultaten voor de andere proeven in dezelfde richting liggen, maar op zichzelf niet tot een dergelijke conclusie zouden leiden.

Toepassing van de symmetrietoets op $\frac{x-y}{x+y}$ in plaats van op $x-y$ geeft in dit geval vrijwel dezelfde uitkomsten als de in tabel II vermelde.

---

3) Het teken + bij een overschrijdingskans betekent dat wasmiddel C betere resultaten geeft dan wasmiddel V. De overschrijdingskansen zijn tweezijdig.

4) Zie voetnoot 3.

Amsterdam
Statistische Afdeling.

S47(M6)

Algemene gang van zaken bij het toetsen van een hypothese.[1)]

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2)], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \ldots, x_n$ een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese H_0, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3)]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α. Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $\leq \alpha$ is.

De toetsingstheorie biedt in het algemeen <u>geen</u> mogelijkheid om tot <u>aanvaarding</u> van een hypothese te komen. Indien een bepaalde hypothese $\mathcal{H}_0$ niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze <u>nevenvoorwaarden</u> dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de <u>overschrijdingskans</u> k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van $\mathcal{H}_0$ zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. <u>Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.</u>

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven litteratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van $\mathcal{H}_0$ leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

<u>Litteratuur:</u>

J. Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

J. Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica <u>4</u> (1950) p.54-66.

Mathematisch Centrum,
2de Boerhaavestraat 49,
Amsterdam O.
Statistische Afdeling,
S47 (M7).

Maart, 1952.

De toets van Wilcoxon.[1])

Deze methode dient tot het toetsen van de hypothese H_o, inhoudende, dat twee steekproeven $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ afkomstig zijn uit één collectie (ook wel populatie of universum genaamd).

Voor het toetsen van de hypothese H_o wordt gebruik gemaakt van een toetsingsgrootheid $\underline{U}$ [2]), die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ naar opklimmende grootte gerangschikt zijn, dan bepalen we eerst het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de kleinste waarneming x_1 uit de eerste steekproef (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1). Noem dit aantal V_1. Vervolgens wordt het aantal waarnemingen uit de tweede steekproef bepaald, dat kleiner is dan de op één na kleinste waarneming x_2 uit de eerste steekproef (bij gelijkheid wordt weer een $\frac{1}{2}$ in plaats van 1 geteld). Dit aantal noemen we V_2. Evenzo worden met betrekking tot $x_3,x_4,\ldots,x_n$ de aantallen $V_3,V_4,\ldots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

$$U=V_1+V_2+\ldots+V_n.$$

Wanneer onder de waarnemingen niet te veel gelijken voorkomen, kan bewezen worden, dat de toetsingsgrootheid $\underline{U}$ onder de hypothese H_o voor grote waarden van n en m (beide $\geq$ 10) bij benadering een normale verdeling bezit. De waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ tezamen genomen vallen uiteen in een aantal groepen van gelijke waarnemingen. Noem het aantal van deze groepen k, dan is k minstens 1 (als alle waarnemingen gelijk zijn) en hoogstens m+n (als alle waarnemingen verschillend zijn).

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Stochastische grootheden worden door onderstreping aangeduid.

Zijn $t_1,\dots,t_k$ de aantallen waarnemingen in deze groepen van gelijken, dan worden het gemiddelde μ en de variantie σ^2 van de toetsingsgrootheid $\underline{U}$ gegeven door

$$\mu(\underline{U}) = \tfrac{1}{2}nm,$$

en

$$\sigma^2 = \mathrm{Var}\,(\underline{U}) = \frac{1}{12}\,\frac{nm}{(n+m)(n+m-1)}\left\{(n+m)^3 + \right.$$
$$\left. - (t_1^3 + t_2^3 + \dots + t_k^3)\right\} \qquad ^{1)}$$

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. Indien de hypothese H_0 niet vervuld is, zal de grootheid $\underline{U}$ grote of kleine waarden bezitten, al naar gelang $\underline{y}$ systematisch kleiner of groter is dan $\underline{x}$.

De (tweezijdige) toets bestaat nu daarin, dat men H_0 verwerpt indien de gevonden waarde U van $\underline{U}$ te sterk van μ afwijkt, d.w.z. als

$$\frac{|U-\mu|}{\sigma} > \xi_\alpha \,, \qquad ^{2)}$$

waarin α de onbetrouwbaarheidsdrempel is en ξ_α volgt uit

$$\frac{1}{\sqrt{2\pi}}\int_{\xi_\alpha}^{\infty} e^{-\frac{1}{2}x^2}\,dx = \tfrac{1}{2}\alpha \,,$$

en in een tabel van de normale verdeling kan worden opgezocht.

De (tweezijdige) overschrijdingskans k , behorende bij T, is gedefinieerd als

$$k = \frac{2}{\sqrt{2\pi}}\int_{\left|\frac{U-\mu}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2}\,dx \qquad ^{2)}$$

en kan ook in eenttabel van de normale verdeling worden gevonden.

Bij eenzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.

Een bijzonder geval van het bovenstaande is, dat onder de waarnemingen voor $\underline{x}$ en $\underline{y}$ in 't geheel geen gelijken voorkomen. In dat geval kan de uitdrukking voor de variantie herleid worden tot

$$\sigma^2 = \frac{1}{12}\,nm(n+m+1).$$

---

1) Deze formule is een door T.J.Terpstra gegeven vereenvoudiging van de door J.Hemelrijk([5] en [7]) afgeleide formule. De afleiding van deze vereenvoudigde formule zal nog gepubliceerd worden.

2) Deze formules berusten op de normale benadering van de verdeling van $\underline{U}$.

Indien n en m kleiner zijn dan 10, zijn tabellen beschikbaar voor het berekenen van de overschrijdingskans k voor de uit de steekproef bepaalde waarde U van U (zie [2] en [4]). Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

Opmerking. Men kan gemakkelijk bewijzen, dat de variantie van U door het optreden van gelijke waarnemingen vermindert. Het verschil, dat door deze gelijken optreedt, is echter in het algemeen gering. Men kan daarom in eerste instantie deze correctie op σ^2 verwaarlozen. De overschrijdingskansen, die men dan vindt, zijn iets te groot.

Litteratuur:

1. F.Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945), p.80-83.
2. H.B.Mann and D.R.Whitney, On a test of whether one of two random variables is stochastically larger than the other, Amer.Math.Stat. 18 (1947), p. 50-60.
3. H.R.van der Vaart, Some remarks on the power function of Wilcoxon's test for the problem of two samples, Proceedings van de Kon. Ned.Ak.v.Wet., 53 (1950), p. 494-520.
4. H.R.van der Vaart, Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en m $\leq$ 10, Rapport S32 (M4) (1950).
5. H.R.van der Vaart, De toets van Wilcoxon voor het probleem van twee steekproeven. (Cursus "Parametervrije Methoden", 1951-'52).
6. D.van Dantzig, Kadercursus Mathematische Statistiek, Math. Centrum, Amsterdam (1947-'50), hoofdst. 6, § 3.
7. J.Hemelrijk, Note on Wilcoxon's two sample test, when ties are present, Ann.Math.Stat. 23 (1952) no. 2.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m O.
Statistische Afdeling.

S 47 (M 10)

Symmetrietoets[1].

Hypothese H_0: de waarnemingen $z_1,\ldots,z_{n'}$ zijn afkomstig van n' onafhankelijk verdeelde stochastische grootheden, die alle symmetrisch ten opzichte van 0 verdeeld zijn[2]. Van deze toets bestaan meerdere versies $T_1,\ldots,T_2''$. We bespreken eerst T_1 en T_2.

Toetsingsgrootheden. Deze worden als volgt uit $z_1,\ldots,z_{n'}$ afgeleid:

1e. de waarnemingen, die gelijk aan 0 zijn worden weggelaten. Stel er blijven over: $z_1,\ldots,z_n$.

2e. Hieruit worden de positieve waarnemingen gezocht. Stel dit zijn $x_1,\ldots,x_{n_1}$, dus n_1 in aantal.

3e. De overblijvende negatieve waarnemingen worden van teken veranderd, zodat zij ook positief worden. Stel dit zijn dan $y_1,\ldots,y_{n_2}$.

4e. De grootheden $x_1,\ldots,x_{n_1}$ en $y_1,\ldots,y_{n_2}$ worden, door elkaar, in afdalende grootte-volgorde opgeschreven. Stel dit geeft: $w_1,\ldots,w_n$. (Komen er gelijken voor, dan worden deze in willekeurige volgorde geplaatst.)

5e. De groep waarden $w_1,\ldots,w_n$ wordt verdeeld in twee groepen $w_1,w_2,\ldots,w_r$ en $w_{r+1},\ldots,w_n$, waarbij $w_r \neq w_{r+1}$ is en r zo dicht mogelijk bij de waarde $\frac{1}{2}n$ genomen wordt. Is n even, dan wordt $r=\frac{1}{2}n$, indien althans $w_{\frac{1}{2}n} \neq w_{\frac{1}{2}n+1}$ is. Zijn er twee mogelijk keuzen voor r, beide op gelijke afstand van $\frac{1}{2}n$, dan nemen wij $r > \frac{1}{2}n$. Is b.v. n oneven en $w_{\frac{1}{2}n-\frac{1}{2}} \neq w_{\frac{1}{2}n+\frac{1}{2}}$, dan nemen wij $r=\frac{1}{2}n+\frac{1}{2}$. Wij geven de waarden $w_1,\ldots,w_r$ aan als groep A (die dus r elementen bevat) en de overigen als groep B. Alle elementen van A zijn dus groter dan ieder element van B[3].

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Zetten wij hier a in plaats van 0, dan geldt H_0 voor $x_1-a,\ldots,x_{n'}-a$.

3) In de oorspronkelijke publicaties over deze toets (zie de litteratuurverwijzingen aan het einde van dit memorandum) is een enigzins minder algemene definitie van r gegeven.
Alle stellingen blijven echter gelden, indien de hier gegeven definitie gebruikt wordt.

6e. Het aantal waarden van $x_1,\ldots,x_n$ die in A voorkomen noemen wij u.

De toetsingsgrootheden zijn $\underline{n_1}$ en $\underline{u}$, r is een hulpgrootheid.

<u>V.B.</u> 22 waarden z_i: 7,4/6,3/3,6/3,5/3,4/2,9/2,5/1,1/0/0/-1,3/-2,5/-3,2/-4,6/-4,6/-4,6/-4,8/-6,3/-7,0/-7,9/-8,0/-8,7.

$\therefore$ $r = 11$, $n_1 = 8$

$u = 2$

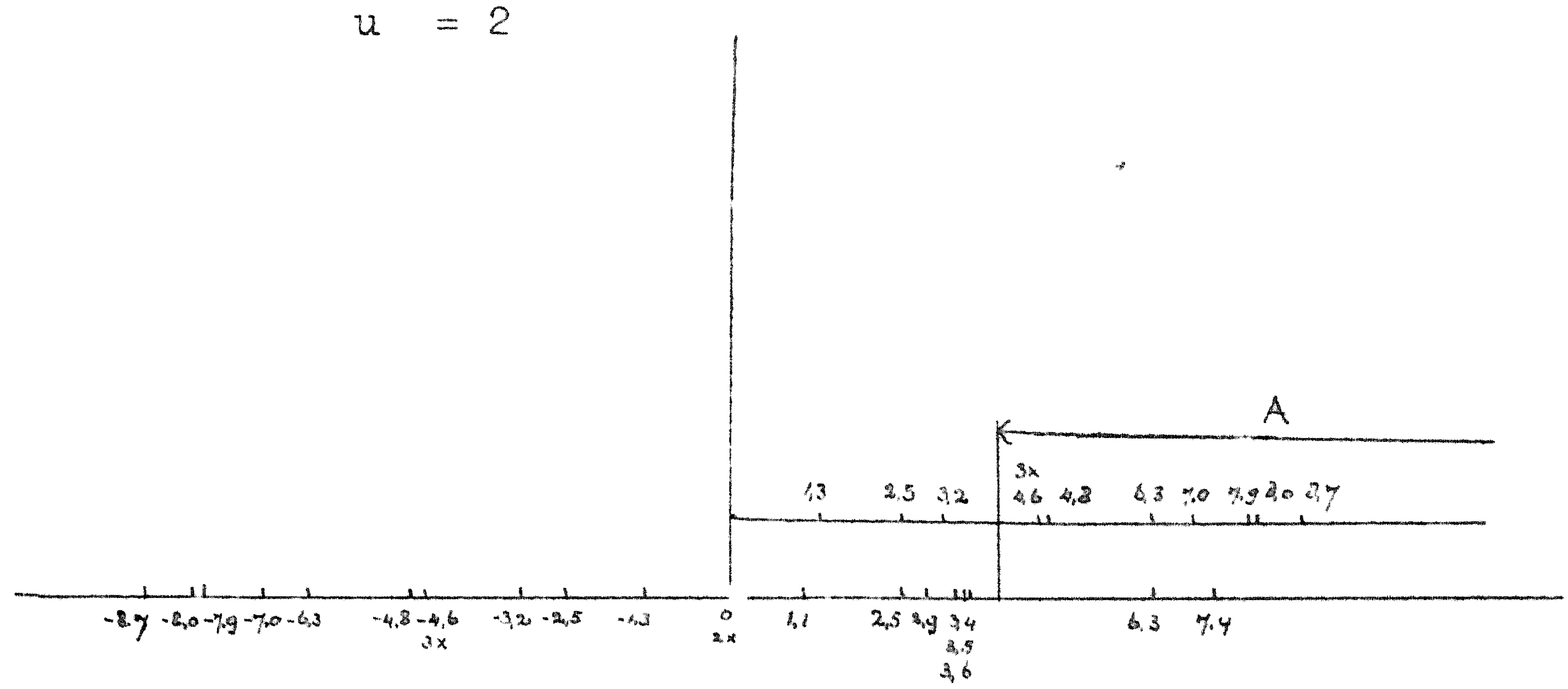

fig. 1

<u>Kritieke zônes</u>. Waarden van n_1, die dicht bij 0 of dicht bij n liggen, zullen, als H_0 juist is weinig, maar als H_0 onjuist is vaker, voorkomen. Grote resp. kleine waarden van u zullen eveneens, als H_0 juist is weinig voorkomen. Hierop berust de keuze van de bij T_1 ~~en~~ T_2 behorende kritieke zônes Z_1 resp. Z_2. Z_1 bevat grote en kleine waarden van n_1 en grote en kleine waarden van u, terwijl Z_2 bij grote waarden van n_1 in hoofdzaak grote waarden van u en bij kleine waarden van n_1 in hoofdzaak kleine waarden van u bevat. T_1 leidt bij voldoende grote n vrijwel steeds tot verwerping als de hypothese niet is vervuld. T_2 leidt echter alleen tot verwerping van H_0 als er veel <u>positieve</u> (resp, negatieve) waarden zijn, die <u>verder</u> van 0 verwijderd liggen dan de <u>negatieve</u> (resp. positieve). In sommige gevallen is het juist van belang om deze laatste afwijkingen van H_0 te ontdekken. In dat geval gebruikt men T_2 liever dan T_1. In fig. 2 is een schematisch voorbeeld gegeven van een serie waarnemingen, waarbij het aantal positieve groter is dan het aantal negatieve, terwijl deze positieve dichter bij 0 liggen dan de negatieve, zodat T_2 niet tot verwerping leidt.

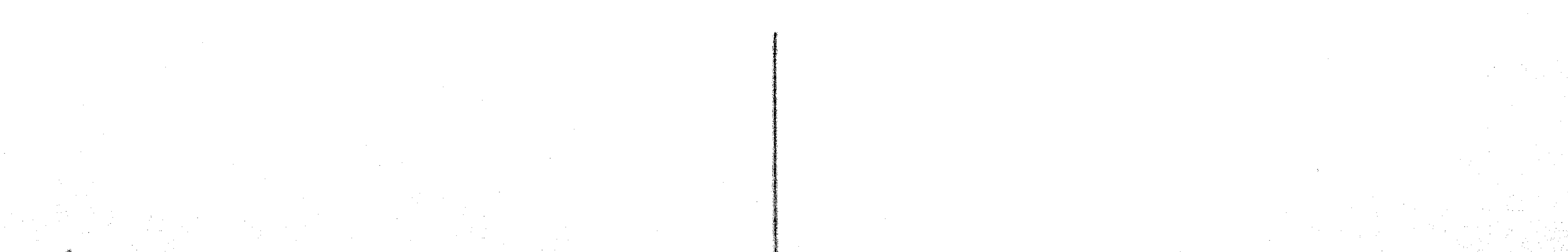

fig. 2

Van T_1 en T_2 bestaan ook éénzijdige versies, waarvan de beschrijving te ver zou voeren.

T_1' en T_2'.

Toetsingsgrootheden.

1e, 2e en 3e als boven. Eerste toetsingsgrootheid: $\underline{n}_1$.

4e: op $x_1,\ldots,x_{n_1}$ en $y_1,\ldots,y_{n_2}$ wordt de toets van Wilcoxon toegepast (vgl. S 47 (M 8)). De toetsingsgrootheden zijn $\underline{n}_1$ en de $\underline{U}$ van deze toets van Wilcoxon.

Kritieke zônes. Overwegingen analoog aan die voor T_1 en T_2 (met U in plaats van u) leiden tot analoge kritieke zônes Z_1' en Z_2', behorend bij T_1' en T_2'.

Opmerkingen: T_1 en T_2 zijn bijzonder geschikt voor een niet te groot aantal waarnemingen. Zij gelden ook voor niet continue verdelingen. T_1' en T_2' zijn alleen geschikt, als er geen of weinig paren (x_i,y_j) met $x_i=y_j$ zijn. Voor grote aantallen zijn T_1' en T_2' geschikter dan T_1 en T_2. Er is ook een versie voor grote aantallen (T_1'' en T_2''), die geheel analoog is met T_1' en T_2' met dien verstande, dat $\underline{u}$ in plaats van $\underline{U}$ wordt gebruikt (vgl. b.v. [2], blz. 77, § 6.4.5).

Litteratuur:

[1] J.Hemelrijk, A family of parameterfree tests for symmetry with respect to a given point, I, II. Proceedings van de Kon.Ned.Ak.v.Wet. 53 (1950), p.945-955. Indagationes Mathematicae 12 (1950), p. 340-350.

[2] - " - , Symmetrietoetsen, Diss., Den Haag 1950, Excelsior.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.

Statistische Afdeling
Rapport S 102 (M 17b)

Het combineren van onafhankelijke toetsen (aanvulling) [1].

In memorandum S 73 (M 17a) wordt een methode voor combinatie van onafhankelijke toetsen behandeld, waarbij het nodig is de overschrijdingskans van iedere toets te bepalen. In vele gevallen kan men de combinatie ook direct op de afzonderlijke toetsingsgrootheden baseren en dit verdient zelfs de voorkeur.

Wij beschouwen hier het geval, dat een bepaalde toets moet worden toegepast op een heterogeen materiaal. Dit materiaal wordt dan eerst verdeeld in h homogeen geachte groepen. Het aantal waarnemingen van de i^e groep zij n_i en de toetsingsgrootheid $\underline{t}_i$ [2]. Laat verder gegeven zijn, dat de verdeling van $\underline{t}_i$ onder de getoetste hypothese (voor de i^e groep aangeduid door H_i) voor grote n_i asymptotisch normaal [3] is, met bekende verwachting μ_i en bekende spreiding σ_i. Aan deze voorwaarden is o.a. voldaan, indien wij te doen hebben met toetsen van WILCOXON, rangcorrelatietoetsen van KENDALL of SPEARMAN, tekentoetsen enz.

Wij toetsen met al de hier te behandelen gecombineerde methoden de hypothese H, dat voor iedere groep de desbetreffende hypothese H_i geldt, terwijl de groepen onderling onafhankelijk zijn. De toetsen verschillen echter ten aanzien van de alternatieve (van H afwijkende) hypothesen waarvoor zij gevoelig [4] zijn.

De meest gebruikelijke toetsingsgrootheden van gecombineerde toetsen zijn van de gedaante:

$$\underline{T} = \sum_{i=1}^{h} c_i(\underline{t}_i - \mu_i)$$

waarin de letters c_i ($i = 1, 2, \ldots, h$) constanten voorstellen, die voor ieder van de combinatiemethoden op een bepaalde wijze

---

1) Dit memorandum dient slechts ter oriëntatie en streeft niet naar volledigheid of volledige exactheid. Het is bedoeld als een aanvulling op Rapport S 73 (M 17a).

2) De onderstreping geeft aan dat een toetsingsgrootheid stochastisch is, d.w.z. een waarschijnlijkheidsverdeling bezit.

3) Dit houdt in dat $\underline{t}_i$ een waarschijnlijkheidsverdeling heeft, die als n_i toeneemt, steeds minder van een normale verdeling (verdeling van Gauss) afwijkt.

4) Een toets van hypothese H is gevoelig ten opzichte van een alternatieve hypothese H', als de kans dat H verworpen wordt, indien H' juist is, groot is.

gekozen worden. Onder de hypothese H zal $\underline{T}$ asymptotisch (voor grote h en/of grote n_i) normaal verdeeld zijn met verwachting 0 en spreiding $\sigma = \sqrt{\sum_{i=1}^{h} c_i^2 \sigma_i^2}$. De dubbele overschrijdingskans van een gevonden waarde T van $\underline{T}$ is dus bij benadering gelijk aan:

$$\frac{2}{\sqrt{2\pi}} \int_{\left|\frac{T}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2} dx$$

en kan bepaald worden met behulp van een tabel van de normale verdeling. Indien de dubbele overschrijdingskans kleiner is dan de onbetrouwbaarheidsdrempel α , zal men H verwerpen.

Wij geven hier 3 combinatiemethoden van dit type:

<u>Methode 1</u>: $c_1 = c_2 = \cdots\cdots = c_h = 1$

dus: $\underline{T} = \sum_{i=1}^{h} \underline{t}_i - \sum_{i=1}^{h} \mu_i$; $\sigma = \sqrt{\sum_{i=1}^{h} \sigma_i^2}$

<u>Methode 2</u>: $c_1 = \frac{1}{\sigma_1}$, $c_2 = \frac{1}{\sigma_2}$, $\ldots\ldots$, $c_h = \frac{1}{\sigma_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{\sigma_i}$; $\sigma = \sqrt{h}$

<u>Methode 3</u>: $c_1 = \frac{1}{n_1}$, $c_2 = \frac{1}{n_2}$, $\ldots\ldots$, $c_h = \frac{1}{n_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{n_i}$; $\sigma = \sqrt{\sum_{i=1}^{h} \left(\frac{\sigma_i}{n_i}\right)^2}$

Deze methoden zijn alleen gevoelig ten aanzien van alternatieve hypothesen volgens welke de grootheden $\underline{t}_i$ verdelingen hebben die, voor zover zij afwijken van de verdelingen onder de corresponderende hypothesen H_i, dit over het algemeen in dezelfde richting doen. Men zal dan <u>methode 1</u> bij voorkeur toepassen als men aan de $\underline{t}_i$ met een kleine spreiding (in de regel zullen dat de $\underline{t}_i$ van kleine groepen zijn) een geringer gewicht wil toekennen dan aan de $\underline{t}_i$ met een grote spreiding. De <u>methoden 2 en 3</u> zijn te gebruiken als men aan de verschillende groepen waarnemingen, ongeacht hun grootte, een ongeveer gelijke invloed op het resultaat wil toekennen. De keuze tussen deze twee methoden hangt verder van hier niet te behandelen theoretische overwegingen af (zie literatuur [1]).

Indien men verwacht dat mogelijke verschuivingen van de verdelingen der $\underline{t}_i$ in beide richtingen kunnen liggen, verdient het de voorkeur om gebruik te maken van de volgende toetsingsgrootheid:

$$\sum_{i=1}^{h} \left(\frac{\underline{t}_i - \mu_i}{\sigma_i}\right)^2 \qquad \text{(methode 4)}$$

Deze grootheid is onder de hypothese H asymptotisch verdeeld volgens een χ^2-verdeling met h vrijheidsgraden. De overschrijdingskans van een gevonden waarde van deze grootheid kan dus met behulp van een tabel van de χ^2-verdeling bepaald worden.

De toets, behandeld in memorandum S 73 (M 17a) par. 1, waarbij men het product van linkszijdige en product van alle rechtszijdige overschrijdingskansen bepaalt en het kleinste van deze twee producten gebruikt, heeft betrekking op dezelfde gevallen als de hier behandelde methoden 2 of 3, terwijl de methode, behandeld in S 73 (M 17a) par. 2, berustend op het product van de tweezijdige overschrijdingskansen, meer overeenkomt met methode 4. Men mag echter verwachten, dat, zo aan de asymptotische normaliteit der $\underline{t}_i$ voldaan is, de in dit memorandum behandelde methodenscherper zijn dan de toetsen behandeld in S 73 (M 17a).

Literatuur:

1 C.van Eeden, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Rapport S 115 (M 45) van het Mathematisch Centrum (1953).

2 -----------, Trendtoets met behulp van rangcorrelatie, Memorandum S 73 (M 13a). (Voorbeeld van toepassing van methode 1.)

3 Dr J.Hemelrijk, Het combineren van onafhankelijke toetsen, Memorandum S 73 (M 17a).