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 164

Vergelijkend onderzoek naar het resultaat van één en twee injecties salvarsan en penicilline ten aanzien van de symptomen van framboesia.

door

Ir A.R. Bloemena
en
Rina Korswagen

1954

1. Onderzoek.

Het onderzoek geschiedde in Indonesië en had betrekking op 466 frambosia-patiënten. Bij het eerste onderzoek werden de patiënten doorlopend genummerd, waarna de diagnose werd gesteld. Daarop kregen ongeacht de klinische toestand van de patiënten de even nummers penicilline en de oneven nummers salvarsan. Er had zo dit mogelijk was een tweede en derde onderzoek plaats na resp. 1 en 2 weken. Bij het tweede onderzoek werd de verbetering genoteerd en alternerend aan de patiënten een injectie met hetzelfde geneesmiddel als de eerste maal resp. een physiologische zoutoplossing gegeven. Bij het derde onderzoek werd weer de verbetering t.o.v. de toestand tijdens het eerste onderzoek genoteerd.

De diagnose werd gecodeerd volgens het systeem van Dr H. Oomen, met kleine wijzigingen door de onderzoeker aangebracht in verband met locale verschillen. De cijfers 1,2, en 3 worden in dit systeem gebruikt om het stadium van de ziekte aan te geven. De cijfers worden gecombineerd met letters die in het door het cijfer aangegeven stadium bepaalde uitwendige verschijnselen aanduiden.

De graad van verbetering werd aangegeven door:

0: geen verbetering
1 geringe "
2: matige "
3: goede "
v: volledige "

Alle verbeteringen werden beschouwd ten opzichte van de status bij het eerste onderzoek.

Naar aanleiding van dit onderzoek hebben wij de volgende vragen beantwoord:

a) Is er verschil tussen de uitwerking van penicilline en die van salvarsan?

b) Is er, zowel voor penicilline als voor salvarsan een verschil tussen éénmaal en tweemaal een injectie?

2 Bewerking van het materiaal.

Uit de toegezonden lijsten werden tabellen opgesteld voor iedere behandelingswijze, die aangaven hoeveel patiënten een verbetering 0,1,2,3 of v vertoonden, onderverdeeld naar de 9 verschillende ziektebeelden: 1,2c,2p,21,3u,3b,3zu,3znu en 3n. Dit alles zowel voor de resultaten na 1 week als voor die na 2 weken.

Een groot aantal patiënten leed echter aan een combinatie van 2 of meer van deze ziektebeelden. Uit statistisch oogpunt bleek het daarom wenselijk om afzonderlijke tabellen te maken voor:

a) de enkelvoudige ziektebeelden: d.w.z. de patiënten die aan één en niet meer dan één ziektebeeld leden;

b) de meervoudige ziektebeelden: d.w.z. de patiënten die 2 of meer ziektebeelden vertoonden. In deze tabel wordt een patiënt die aan twee ziektebeelden lijdt tweemaal getabelleerd nl. één keer onder het eerste ziektebeeld en de daarbij behorende verbetering, de tweede keer onder het tweede ziektebeeld en de verbetering daarvan. Een patiënt met drie ziektebeelden komt driemaal in deze tabel voor. Hieruit volgt dat het materiaal in deze tabel niet onafhankelijk is.

c) a) en b) tezamen. Deze tabel is dus de som van a) en b) en dientengevolge zijn de gegevens in deze tabel ook niet onafhankelijk.

Aangezien de aantallen verkregen bij de ziektebeelden 1, 2p, 2l, 3b en 3u zeer klein zijn, werden alleen de gegevens voor de ziektebeelden 2c, 3u, 3zu en 3znu statistisch bewerkt.

De volgende notatie werd aangehouden:

P :	resultaten	na	1 week	van	penicilline
P_1:	"	"	2 weken	"	1 injectie penicilline
P_2:	"	"	2 "	"	2 injecties "
S :	"	"	1 week	"	salvarsan
S_1:	"	"	2 weken	"	1 injectie salvarsan
S_2:	"	"	2 "	"	2 injecties "

Een voorbeeld van de zo verkregen tabellen, waarvan bij de berekeningen werd uitgegaan is Tabel I, die voor de enkelvoudige en voor de meervoudige ziektebeelden apart de aantallen patiënten met een graad van verbetering 0,1,2,3 of v voor P en voor S aangeeft. (zie blz. 3)

De verschillen tussen twee waarnemingsreeksen zijn getoetst met behulp van de toets van WILCOXON. Een oriëntatie betreffende het gebruik hiervan is opgenomen in de memoranda S 47 (M 6 en 7) en S 145 (M 7a) die als aanhangsel zijn bijgevoegd. In alle gevallen is voor 2c en 3u de overschrijdingskans berekend volgens de methode van de normale benadering (zie M 7), voor 3zu en 3znu in de meeste gevallen volgens de exacte methode (zie M 7a)

De tabellen II en III geven de zo verkregen lijst van berekende tweezijdige overschrijdingskansen.

Tabel I

Resultaten van P en S na één week voor de enkelvoudige en meervoudige ziektebeelden; uitgedrukt in aantallen patiënten.

		2c		3u		3zu		3znu	
ziektebeeld:		enkelvoudig	meervoudig	enkelvoudig	meervoudig	enkelvoudig	meervoudig	enkelvoudig	meervoudig
P	0	-	-	1	1	-	-	21	33
	1	3	1	2	6	-	-	2	1
	2	12	9	11	9	-	3	-	-
	3	46	31	8	8	6	14	-	-
	v	6	1	-	-	-	-	-	-
totaal		67	42	22	24	6	17	23	34
S	0	-	-	2	1	-	2	13	19
	1	8	3	8	5	2	1	1	4
	2	20	6	12	6	1	4	-	2
	3	38	29	11	6	6	12	-	-
	v	2	4	-	-	-	-	-	-
totaal		68	42	22	18	9	19	14	25

Tabel II 1)

Tweezijdige overschrijdingskansen, gevonden bij de vergelijking van de enkelvoudige en de meervoudige ziektebeelden.

	2c	3u	3zu	3znu
P	0,52-	0,46-	0,54-	0,44-
S	0,02+	0,92-	0,88-	0,23+
P_1	0,72+	0,33+	1,00	0,68-
P_2	0,24-	0,007-	0,49-	0,51-
S_1	0,25+	0,97+	0,59-	0,23+
S_2	0,50-	0,81-	0,51+	0,55-

---

1) Een + teken bij een overschrijdingskans geeft aan, dat de eerstgenoemde reeks lagere graden van verbetering vertoont dan de tweede

Tabel III [1]

Tweezijdige overschrijdingskansen, gevonden bij de vergelijking van P en S en bij de vergelijking van 1 en 2 injecties bij

Enkelvoudige ziektebeelden					meervoudige ziektebeelden					alle ziektebeelden				
	2c	3u	3zu	3znu		2c	3u	3zu	3znu		2c	3u	3zu	3znu
P - S	0,009-	0,40-	0,23-	1,00	P - S	0,52+	0,86-	0,18-	0,014 +	P - S	0,07-	0,51-	0,05-	0,06+
P_1-S_1	0,09 -	0,54+	1,00	0,29+	P_1-S_1	0,42-	0,60-	1,00	0,07 +	P_1-S_1	0,08-	1,00	0,77+	0,77-
P_2-S_2	0,05 -	0,06-	0,24-	0,24+	P_2-S_2	0,33-	0,67+	0,56-	0,12 +	P_2-S_2	0,03-	0,33-	0,23-	0,07+
P_1-P_2	0,55 +	0,008+	1,00	0,85+	P_1-P_2	0,36-	0,61-	1,00	0,25 -	P_1-P_2	0,96-	0,17+	1,00	0,27-
S_1-S_2	0,55 +	0,58+	0,27-	0,07+	S_1-S_2	0,21-	0,63+	0,63-	0,90 -	S_1-S_2	0,70-	0,44+	0,12-	0,28+

---

1) Een + teken bij een overschrijdingskans geeft aan, dat de eerstgenoemde reeks lagere graden van verbetering vertoont dan de tweede.

3. Conclusies.

Bij het formuleren van onze conclusies hechten wij alleen waarde aan resultaten met overschrijdingskansen die ongeveer 0,05 of lager zijn. Al naar gelang de overschrijdingskansen zullen wij dan spreken van een aanwijzing voor een bepaald effect of zonder meer van een effect.

1. Bij het enkelvoudige ziektebeeld 2c blijkt een week na de eerste injectie penicilline betere resultaten te geven dan salvarsan. Een aanwijzing in dezelfde richting vindt men 14 dagen na de eerste injectie, in het bijzonder bij die patiënten die een tweede injectie hebben gehad. Men dient hierbij wel te overwegen, dat dit laatste resultaat afhankelijk is van het eerste, omdat het betrekking heeft op een deel van de patiënten, waarbij de eerste vergelijking is toegepast.

Bij het ziektebeeld 2c, optredend in combinatie met andere ziektebeelden is van dit effect niets meer te vinden.

2. Bij de ziektebeelden 3u en 3zu worden geen duidelijke aanwijzingen voor een verschil tussen penicilline en salvarsan gevonden, hoewel de vergelijking van P - S voor alle ziektebeelden van 3zu en de vergelijking $P_2 - S_2$ voor het enkelvoudige ziektebeeld 3u nogal kleine overschrijdingskansen in het voordeel van de penicilline vertonen.

3. Bij het ziektebeeld 3znu vindt men resultaten, die gunstiger zijn voor salvarsan. Alleen voor die gevallen, waar 3znu optreedt in combinatie met een ander ziektebeeld is deze aanwijzing na één week behandeling van enige kracht.

Samenvattend kan men zeggen, dat er aanwijzingen zijn, dat penicilline beter werkt voor de lichtere vorm 2c, wanneer zij enkelvoudig voorkomt, terwijl salvarsan wellicht de voorkeur verdient voor de ernstige meervoudige vormen met 3znu.

4. Er is slechts éénmaal een duidelijk verschil tussen de resultaten van één en van twee injecties. Dit is het geval bij het enkelvoudige ziektebeeld 3u bij behandeling met penicilline. Hier blijken 2 injecties een beter resultaat te geven dan 1 injectie. Er is verder een zwakken aanwijzing dat een beter resultaat verkregen wordt bij 2 x een salvarsan-injectie bij het enkelvoudige ziektebeeld 3znu dan 1 x een injectie. Over het algemeen kan men dus moeilijk zeggen dat een tweede injectie het resultaat duidelijk verbetert.

5. Het blijkt dat salvarsan betere resultaten geeft bij de gevallen van meervoudige 2c dan bij die van enkelvoudige 2c. Bij 3u vinden wij dat 2 x een injectie penicilline beter werkt bij het enkelvoudige ziektebeeld dan bij het meervoudige, een resultaat dat in de richting ligt van de tendens, die wij in de samenvatting onder 3 geconstateerd hebben.

Daar het waarnemingsmateriaal in verband met het vrij grote aantal ziektebeelden niet van voldoende grote omvang was, om de verwachting van zeer duidelijke resultaten te rechtvaardigen, is het niet te verbazen, dat de conclusies een voorzichtig karakter moeten dragen. Zij dienen hoofdzakelijk als aanwijzingen beschouwd te worden, waarvan desgewenst door hernieuwd onderzoek een bevestiging kan worden gezocht.

MATHEMATISCH CENTRUM
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 geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese 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 nevenvoorwaarden 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 overschrijdingskans k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van 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. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven littèratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van H_0 leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

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 4 (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 $x_1,\dots,x_n$ en $y_1,\dots,y_m$ 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 $x_1,\dots,x_n$ en $y_1,\dots,y_m$ 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 $x_3,x_4,\dots,x_n$ de aantallen $V_3,V_4,\dots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

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

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 $x_1,\dots,x_n$ en $y_1,\dots,y_m$ 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 $t_1,\ldots,t_k$ 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

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

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden ~~vast.~~ van ti.
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 gedefiniëerd 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 een tabel 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
2e Boerhaavestraat 49
A m s t e r d a m - O.
Statistische Afdeling
S 145 (M 7a).

Enige opmerkingen over de verdeling van de toetsingsgrootheid van WILCOXON.

In memorandum S 47 (M 7) is vermeld dat de verdeling van de toetsingsgrootheid $\underline{U}$ van WILCOXON, onder de hypothese H_0, dat de beide steekproeven afkomstig zijn uit dezelfde waarschijnlijkheidsverdeling, voor grote n en m bij benadering normaal is. Deze benadering geldt over het algemeen als er weinig of geen gelijke waarnemingen in beide steekproeven voorkomen. Als n en m klein zijn, kan men, zo er weinig gelijke waarnemingen voorkomen gebruik maken van exacte tabellen van de verdeling van $\underline{U}$ (litt. [1], [2]). Als dan echter de gegeven waarnemingsreeksen grote groepen van gelijke waarnemingen bevatten, zijn de exacte tabellen en de normale benadering niet altijd voldoende nauwkeurig. We zullen laten zien hoe men voor dergelijke gevallen de verdeling van $\underline{U}$ exact kan berekenen.

De gegeven waarnemingsreeks kan dan worden samengevat als aangegeven in tabel I.

Tabel I

Voorkomende steekproef waarden	Aantal malen dat z_i optreedt bij		totaal
	$\underline{x}$	$\underline{y}$	
z_1	$\underline{a}_1$	$\underline{b}_1$	$\underline{t}_1$
z_2	$\underline{a}_2$	$\underline{b}_2$	$\underline{t}_2$
.	.	.	.
.	.	.	.
.	.	.	.
z_k	$\underline{a}_k$	$\underline{b}_k$	$\underline{t}_k$
totaal	n	m	N

Hierin zijn $z_1, \dots, z_k$ de voorkomende steekproefwaarden, gerangschikt naar opklimmende grootte; n en m de steekproef-groottes; t_1, t_2, ..., t_k de grootten der groepen gelijke waarnemingen en $\underline{a}_i$ (resp. $\underline{b}_i$) het aantal malen dat de waarde z_i in de eerste (resp. tweede) steekproef optreedt.

De kans dat de combinatie, gegeven in tabel I, optreedt onder de hypothese H_0, en onder de voorwaarden $\underline{t}_1 = t_1$, $\underline{t}_2 = t_2, \ldots, \underline{t}_k = t_k$ wordt nu:

(1) $$P[\underline{a}_1 = a_1, \underline{a}_2 = a_2, \ldots, \underline{a}_k = a_k \mid t_1, t_2, \ldots, t_k ; H_0] =$$

$$= \frac{\binom{t_1}{a_1} \cdot \binom{t_2}{a_2} \cdot \ldots \cdot \binom{t_k}{a_k}}{\binom{N}{m}} . \qquad 1)$$

Nu worden voor de gegeven waarde van n, m en $t_1, t_2, \ldots, t_k$ alle andere mogelijke combinaties van de a's en de b's opgeschreven en voor ieder hiervan de $\tilde{u}$ berekend volgens

(2) $$\underline{\tilde{u}} = \tfrac{1}{2} \sum_{i<j} \sum (\underline{a}_i \underline{b}_j - \underline{a}_j \underline{b}_i).$$

Voorbeeld: Gegeven 4 waarnemingen van een stochastische grootheid $\underline{x}$ en 11 waarnemingen van een stochastische grootheid $\underline{y}$. De aangenomen steekproefwaarden zijn gerangschikt naar opklimmende grootte z_1, z_2, z_3 en z_4 en in de twee steekproeven tezamen genomen treden deze waarden resp. 1, 1, 2 en 11 maal op. Dus (zie tabel I): $n = 4$, $m = 11$, $t_1 = t_2 = 1$, $t_3 = 2$, $t_4 = 11$, $N = 15$.

Wij geven hieronder alle mogelijke combinaties met de corresponderende waarde van $\tilde{u}$.

	1)	2)	3)	4)	5)	6)
	0 1	0 1	0 1	1 0	0 1	0 1
	0 1	0 1	1 0	0 1	0 1	1 0
	0 2	1 1	0 2	0 2	2 0	1 1
	4 7	3 8	3 8	3 8	2 9	2 9
$\tilde{u}=$	-8	$-1\frac{1}{2}$	0	$+1$	$+5$	$+6\frac{1}{2}$

	7)	8)	9)	10)	11)	12)
	1 0	1 0	0 1	1 0	1 0	1 0
	0 1	1 0	1 0	0 1	1 0	1 0
	1 1	0 2	2 0	2 0	1 1	2 0
	2 9	2 9	1 10	1 10	1 10	1 10
$\tilde{u}=$	$+7\frac{1}{2}$	$+9$	$+13$	$+14$	$+15\frac{1}{2}$	$+22$

---

1) $\binom{t}{a} = \frac{t!}{a!\,(t-a)!}$ Tabellen voor deze binomiaal coëfficiënten vindt men o.a. in [3].

We berekenen vervolgens de waarschijnlijkheden van al deze combinaties volgens formule (1) (zie tabel II, kolom 5). Deze waarschijnlijkheidsverdeling is in figuur 1 getekend.

De linkseenzijdige overschrijdingskans van een gevonden waarde $\tilde{u}$ van $\underline{\tilde{u}}$ wordt gedefiniëerd als de som van de waarschijnlijkheden van die waarden van $\underline{\tilde{u}}$ die kleiner of gelijk zijn aan de gevonden waarde. In tabel II zijn in kolom 6 deze exacte linkseenzijdige overschrijdingskansen gegeven en in kolom 4 de linkseenzijdige overschrijdingskansen die men vindt met behulp van de normale benadering (zie ook figuur 2).

De rechtseenzijdige overschrijdingskans wordt (analoog aan de linkseenzijdige) gedefiniëerd als de som van de waarschijnlijkheden van die waarden van $\underline{\tilde{u}}$ die groter of gelijk zijn aan de gevonden waarde van $\underline{\tilde{u}}$.

Tabel II.

Vergelijking van de linkseenzijdige overschrijdingskans berekend volgens de benaderde en de exacte methode.

1	2	3	4	5	6
Combinatie	$\tilde{u}$	$\frac{\underline{\tilde{u}}}{\sigma}$	linkereenzijdige overschrijdingskans m.b.v. de normale benadering	exacte wh van de combinaties	exacte linker overschrijdingskans
1	− 8	−1,34	0,090	0,242	0,242
2	− $1\frac{1}{2}$	−0,25	0,401	0,242	0,484
3	0	0	0,500	0,121	0,605
4	+ 1	+0,17	0,568	0,121	0,726
5	+ 5	+0,84	0,800	0,040	0,766
6	+ $6\frac{1}{2}$	+1,09	0,862	0,080	0,847
7	+ $7\frac{1}{2}$	+1,26	0,896	0,080	0,927
8	+ 9	+1,51	0,935	0,040	0,967
9	+13	+2,18	0,985	0,008	0,975
10	+14	+2,35	0,991	0,008	0,983
11	+$15\frac{1}{2}$	+2,60	0,995	0,016	0,999
12	+22	+3,69	1,000	0,001	1,000

Zoals duidelijk blijkt is voor enkele van de genoemde combinaties de overschrijdingskans berekend volgens de methode van de normale benadering belangrijk verschillend van die gevonden met de exacte methode.

Een tweezijdige kritieke zone met onbetrouwbaarheidsdrempel α kan hier op verschillende manieren gedefiniëerd worden:

B.v. door een linkseenzijdige en een rechtseenzijdige kritieke zone te kiezen, ieder met onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$. Een bezwaar van deze methode is dat hij vaak leidt tot een groot verschil tussen α en de werkelijke onbetrouwbaarheid. Dit is vooral dan in sterke mate het geval, als de kleinste of de grootste waarde, die $\underline{\tilde{u}}$ aan kan nemen, een waarschijnlijkheid bezit, die $> \frac{1}{2}\alpha$ is. Dan wordt de kritieke zone bovendien éénzijdig. Als we b.v. in ons voorbeeld nemen $\alpha = 0{,}20$ dan bestaat de op deze wijze gedefiniëerde "tweezijdige" kritieke zone uit de waarden $\tilde{u} = 9, 13, 14, 15\frac{1}{2}$ en 22 en de onbetrouwbaarheid is $0{,}040 + 0{,}008 + 0{,}008 + 0{,}016 + 0{,}001 = 0{,}073$.

Een andere methode is, dat men de waarden van $\underline{\tilde{u}}$ met de kleinste waarschijnlijkheden bij elkaar zoekt totdat de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert. Deze methode heeft het voordeel boven de vorige dat in het algemeen de onbetrouwbaarheid dichterbij α ligt, maar bij verdelingen met meer dan één top het nadeel dat de kritieke zone geen aaneengesloten geheel is. Nemen we in ons voorbeeld weer $\alpha = 0{,}20$ dan bestaat deze tweezijdige kritieke zone uit de waarden $\tilde{u} = 5, 9, 13, 14, 15\frac{1}{2}$ en 22 en de onbetrouwbaarheid is $0{,}040 + 0{,}040 + 0{,}008 + 0{,}008 + 0{,}016 + 0{,}001 = 0{,}113$. De waarde $\tilde{u} = 5$ behoort dus wel tot de kritieke zone maar $\tilde{u} = 6\frac{1}{2}$ en $\tilde{u} = 7\frac{1}{2}$ niet en men kan deze kritieke zone bezwaarlijk tweezijdig noemen.

We definiëren nu de tweezijdige kritieke zone als volgt: We beginnen met van de kleinste en de grootste waarde van $\tilde{u}$ diegene te nemen met de kleinste waarschijnlijkheid. In ons voorbeeld dus $\tilde{u} = 22$. Dan kiezen we één van de waarden $\tilde{u} = -8$ en $\tilde{u} = 15\frac{1}{2}$ en wel zodanig dat het verschil tussen de som van de waarschijnlijkheden links en de som van de waarschijnlijkheden rechts in absolute waarde minimaal is. Nemen we $\tilde{u} = -8$ dan is dit absolute verschil $0{,}242 - 0{,}001 = 0{,}241$ en nemen we $\tilde{u} = 15\frac{1}{2}$ dan is dit absolute verschil $0{,}016 + 0{,}001 = 0{,}017$. We nemen dus $\tilde{u} = 15\frac{1}{2}$.

Vervolgens kiezen we één van de waarden $\tilde{u} = 14$ en $\tilde{u} = -8$. Nemen we $\tilde{u} = -8$ dan wordt het genoemde verschil tussen links en rechts in absolute waarde: $0{,}242 - (0{,}016 + 0{,}001) = 0{,}225$ en bij $\tilde{u} = 14$: $0{,}008 + 0{,}016 + 0{,}001 = 0{,}025$. We kiezen weer die waarde van $\tilde{u}$ die het kleinste absolute verschil tussen links en rechts geeft, dus $\tilde{u}$ = 14. Zo gaan we door met telkens links of rechts een waarde van $\tilde{u}$ bij de kritieke zone te nemen totdat de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert. Nemen we weer α =0,2 dan vinden we in ons voorbeeld voor deze kritieke zone de waarden $\tilde{u}$ = $7\frac{1}{2}$, 9, 13, 14, $15\frac{1}{2}$ en 22 en de onbetrouwbaarheid is 0,080 + 0,040 + 0,008 + 0,008 + 0,016 + 0,001 = 0,153. Op deze wijze bereikt men dus dat de onbetrouwbaarheid in het algemeen dichterbij α ligt dan volgens de eerste methode; bovendien verkrijgt men een aaneengesloten geheel voor het linker- en rechter deel der kritieke zone, terwijl de onbetrouwbaarheid zo goed als mogelijk is in gelijke mate over de twee delen verdeeld wordt.

De tweezijdige overschrijdingskans wordt nu gedefiniëerd als de onbetrouwbaarheid van de kleinste kritieke zone van deze aard, die de gevonden waarde van $\underline{\tilde{u}}$ bevat. B.v. de kleinste kritieke zone die het punt $\tilde{u}$ = $6\frac{1}{2}$ bevat bestaat uit de waarden $\tilde{u}$ = 22, $15\frac{1}{2}$, 14, 13, 9, $7\frac{1}{2}$, -8 en $6\frac{1}{2}$ en de tweezijdige overschrijdingskans van $\tilde{u}$ = $6\frac{1}{2}$ is dus in ons voorbeeld de som van de waarschijnlijkheden van deze waarden van $\underline{\tilde{u}}$ en dus 0,476.

In tabel II zijn voor ieder der mogelijke waarden van $\underline{\tilde{u}}$ deze exacte en tevens de benaderde tweezijdige overschrijdingskansen gegeven.

Tabel III.

Exacte en benaderde tweezijdige overschrijdingskansen.

$\tilde{u}$	tweezijdige overschrijdingskans exact	tweezijdige overschrijdingskans benaderd
-8 (7)*	0,395	0,180
$-1\frac{1}{2}$(11)	0,879	0,802
0 (12)	1,000	1,000
1 (10)	0.637	0,864
5 (9)	0,516	0,400
$6\frac{1}{2}$(8)	0,476	0,276
$7\frac{1}{2}$(6)	0,153	0,208
9 (5)	0,073	0,130
13 (4)	0,033	0,030
14 (3)	0,025	0,018
$15\frac{1}{2}$(2)	0,017	0,010
22 (1)	0,001	<0,001

* De cijfers tussen haakjes geven de volgorde aan, waarin de waarden van $\tilde{u}$ bij de kritieke zones getrokken worden.

Om te illustreren welke invloed de waarden van n , m ; t_1 , t_2 , ..., t_k op het resultaat hebben geven we hier nog een tweetal andere voorbeelden met de tweezijdige overschrijdingskansen.

A.

54	32	86
3	5	8
0	2	2
57	39	96

$\tilde{u} = + 139$
exact: $k = 0,06$.
benaderd $k = 0,04$

B.

13	19	32
1	4	5
0	2	2
14	25	39

$\tilde{u} = + 30\frac{1}{2}$
exact $k = 0,23$
benaderd $k = 0,18$

In het algemeen geldt dat de exacte verdeling van $\tilde{u}$ symmetrisch is als $n = m$ en ook als $t_1 = t_k$, $t_2 = t_{k-1}$, $t_3 = t_{k-2}$, enz. (zie Tabel I). In deze gevallen zal voor niet te kleine n en m , de normale benadering meestal vrij goed zijn.

Literatuur:

[1] H.R.VAN DER VAART: Gebruiksaanwijzing voor de toets van Wilcoxon, Rapport S 32 (M 4) van het Mathematisch Centrum, Statistische Afdeling.

[2] C.WHITE: The use of ranks in a test of significance for comparing two treatments Biometrics, Vol.8, number 1., March 1952.

[3] T.C.FRY: Probability and its engineering uses D.van Nostrand Company, New York 1928.

[4] Rapport S 47 (M7) van de Statistische Afdeling van het Mathematisch Centrum.

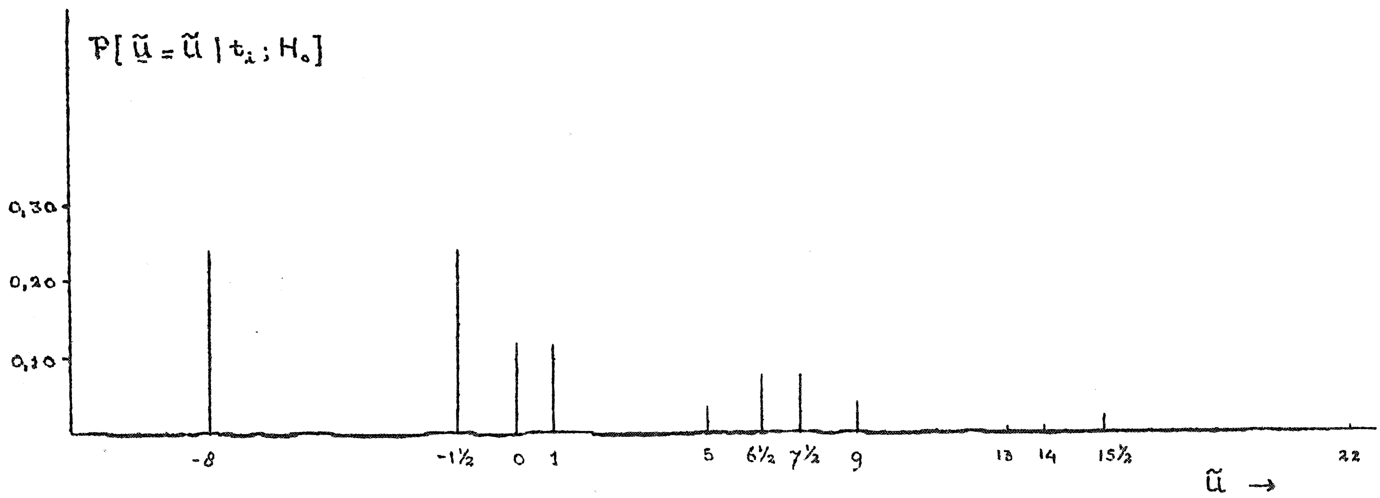

Fig. 1

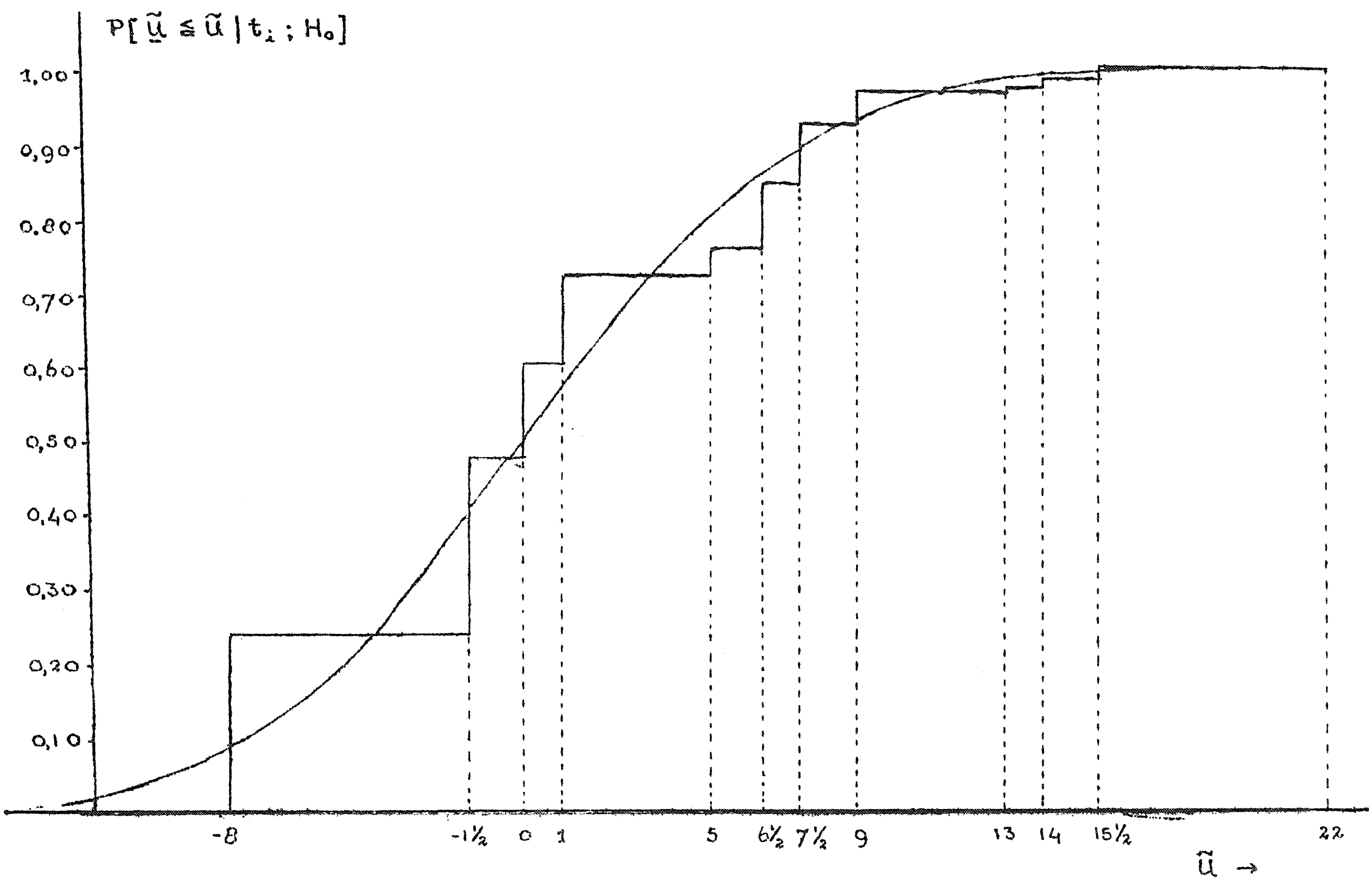

Fig. 2

$P[\underline{\tilde{u}} = \tilde{u} \mid t_i ; H_0]$ en $P[\underline{\tilde{u}} \leq \tilde{u} \mid t_i ; H_0]$, $n = 4$, $m = 11$, $t_1 = t_2 = 1$, $t_3 = 2$, $t_4 = 11$.