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 170

Tweede voorlopige rapport over een vlamstralingsonderzoek (Derde rapport over het onderdeel P.T. 5).

door

Mevr. E.C. Bos-Levenbach
en
R. Doornbos

1955

1. Inleiding

In een vorig rapport (Rapport S 140) werd een analyse van de resultaten van de vijfde proevenserie in de reeks experimenten betreffende de warmtestraling van vlammen met behulp van de methode der variantieanalyse besproken, welke methode, in een enigszins andere vorm, ook werd gevolgd in het rapport van Dr E.F. Drion (T.N.O.-A.B.W.) over hetzelfde onderwerp. Hierbij werd de wenselijkheid naar voren gebracht dit onderzoek, ter bevestiging van de resultaten, voor zover mogelijk te herhalen met andere methoden, waarbij minder onderstellingen omtrent de onbekende waarschijnlijkheidsverdelingen gemaakt behoefden te worden. Dit dient dan tevens ter introductie van deze methoden in het statistisch onderzoek der vlamstralingsprocessen. In verband hiermede voerden wij dit onderzoek opnieuw uit, nu met gebruikmaking van een verdelingsvrije methode, nl. de toets van WILCOXON (zie het bijgevoegde memorandum S 47 (M 7)).

Voor de beschrijving van het experiment verwijzen wij naar het eerste rapport, S 140, par. 1.

2. Vooronderzoek

In rapport S 140, par. 2, werd reeds met behulp van een verdelingsvrije methode een onderzoek naar eventuele afhankelijkheid ingesteld.

De conclusie, die wij hier nog even herhalen, was: "De waarnemingen voor de verschillende spleten van dezelfde vlam zijn, uitgezonderd misschien de e-waarden, onderling afhankelijk." We mogen dus niet de uitkomsten van de toetsen bij de verschillende spleten op de gebruikelijke wijze met elkaar combineren, daar de hiertoe gewoonlijk gevolgde methoden op onafhankelijkheid van de waarnemingen onderling berusten.

3. Toetsing van verschillende hypothesen en betrouwbaarheidsintervallen voor verschillen.

We toetsen hier evenals in rapport S 140, par. 3, de volgende hypothesen:

α: De soort brandstof heeft geen invloed op de straling.

β: De symmetrie of asymmetrie van de vlam beïnvloedt de straling niet.

De hypothese γ, inzake de interactie, kan met behulp ~~van~~ deze methoden niet getoetst worden. Voor ~~het onderzoek van~~ hypothese α werd de toets van Wilcoxon per spleet toegepast op de waarnemingen 1c tegen de waarnemingen 2c(I) en op de waarnemingen 1d tegen de waarnemingen 2d(II). Vervolgens werden beide toetsen op nader aan te geven wijze gecombineerd.

Voor het onderzoek van hypothese β werden de waarnemingen 1c getoetst tegen de waarnemingen 1d(I) en die van 2c tegen die van 2d(II). Vervolgens werden I en II gecombineerd.

De combinatie van de beide toetsen I en II geschiedde op de volgende wijze. De toetsingsgrootheid $\underline{U}$ van Wilcoxon, zoals deze gedefinieerd wordt in memorandum S 47 (M 7) heeft een verdeling die exact bekend is in het geval van 2 steekproeven die elk uit 4 waarnemingen bestaan, zoals bij de toetsen I en II het geval is. Noemen we de toetsingsgrootheden van I en II $\underline{U}_1$ en $\underline{U}_2$, dan kan de verdeling van $\underline{U} = \underline{U}_1 + \underline{U}_2$ eveneens exact berekend worden.

Wij krijgen dus, voor iedere spleet, van beide hypothesen, voor R_1, R_2, R_3 en e de gevonden waarden U_1 en U_2. De overschrijdingskans van $U = U_1 + U_2$ die dus de waarden 0,1,...,32 kan aannemen, wordt dan berekend met behulp van de exact bepaalde verdeling van $\underline{U}$ (zie voor details de appendix, waarin een voorbeeld volledig uitgewerkt is).

Voor de het minst van de verwachting afwijkende gehele waarden van U die onder de nulhypothese met kansen hoogstens gelijk aan 0,05; 0,01 en 0,001 worden onder- resp. overschreden vinden we 5(27); 3(29) en 0(32).

Een overzicht van de gevonden waarden van U en de bijbehorende overschrijdingskansen vinden we in tabel I. Overschrijdingskansen in de intervallen 0,05-0,01; 0,01-0,001 en < 0,001 zijn in deze tabel aangegeven met resp. I,II en III om de vergelijking met de overeenkomstige tabel 2 in S 140 te vergemakkelijken. 1<2 betekent, dat de gevonden waarde van $U = U_1 + U_2$ kleiner is dan de verwachting hiervan onder de getoetste hypothese.

Daar alleen de volgorde van de waarnemingen hier van belang is, en b.v. een zeer grote of een zeer kleine waarde weinig invloed heeft op de resulterende waarde van U, kan het gevonden ongelijkheidsteken in een enkel geval andersom liggen dan in de overeenkomstige tabel van rapport S 140. Inderdaad vinden we iets dergelijks bij spleet 7, R_1 en R_2, c tegen d, gevallen, waarvoor echter geen van beide toetsen op een systematisch effect wijzen.

Tabel I

Overschrijdingskansen gevonden met de gecombineerde toets van Wilcoxon voor de hypothesen β : c en d geven dezelfde straling, en α : 1 en 2 geven dezelfde straling.

		2			3			4		
		u	o.k.		u	o.k.		u	o.k.	
R_1	α	0	0,0004	III 1<2	0	0,0004	III 1<2	0	0,0004	III 1<2
	β	5½	0,04	I c<d	32	0,0004	III c>d	32	0,0004	III c>d
R_2	α	0	0,0004	III 1<2	0	0,0004	III 1<2	1	0,0012	II 1<2
	β	2	0,003	II c<d	31½	0,001	II c>d	32	0,0004	III c>d
R_3	α	1	0,0012	II 1<2	2	0,003	II 1<2	2	0,003	II 1<2
	β	6	0,05	c<d	7	0,08	c<d	13½	0,69	c<d
e	α	1½	0,002	II 1<2	0	0,0004	III 1<2	4	0,016	I 1<2
	β	13	0,62	c<d	24	0,13	c>d	31	0,0012	II c>d

		5			6			7		
		u	o.k.		u	o.k.		u	o.k.	
R_1	α	1	0,0012	II 1<2	19½	0,55	1>2	23½	0,16	1>2
	β	32	0,0004	III c>d	26½	0,04	I c>d	15½	0,96	c<d
R_2	α	3½	0,01	I 1<2	26½	0,04	I 1>2	17½	0,84	1>2
	β	29½	0,005	II c>d	28	0,016	I c>d	16	1,00	c=d
R_3	α	6½	0,07	1<2	19	0,62	1>2	23½	0,16	1>2
	β	21	0,37	c>d	24	0,128	c>d	24	0,128	c>d
e	α	5	0,029	I 1<2	14½	0,84	1<2	25	0,08	1>2
	β	31	0,0012	II c>d	20½	0,43	c>d	13	0,62	c<d

In de figuren 1-8 zijn de betrouwbaarheidsintervallen met onbetrouwbaarheid 0,05 nu bepaald met behulp van de toets van WILCOXON (zie bijgevoegd memorandum S 145(M 52)), getekend voor de verschillen van de gemiddelden voor brandstof 2 en 1 resp. de symmetrische en de asymmetrische vlam.

Zoals te verwachten was zijn deze intervallen over het algemeen iets ruimer dan de corresponderende intervallen in de figuren 1-8 van rapport S 140. Hierbij geldt weer de in rapport S 140, blz. 5 gemaakte opmerking, dat tengevolge van de gevonden afhankelijkheid tussen de waarnemingen van verschillende spleten van dezelfde vlam de betrouwbaarheidsintervallen niet simultaan met dezelfde betrouwbaarheid geldig zijn

Dezelfde analyse, zoals aangegeven aan het begin van par. 3, is uitgevoerd voor de gemiddelde waarde van de straling over de gehele vlam evenals voor de eerste en de tweede helft van de vlam, benevens voor e .

De uitkomsten worden gegeven in tabel II

Tabel II

Toets van Wilcoxon voor gemiddelde over alle of over de helft van de spleten.

		1e helft			2e helft			gehele vlam		
		u	k.		u	k.		u	k.	
R_1	α	0	0,0004	III 1<2	15	0,92	1<2	0	0,0004	III 1<2
	β	32	0,0004	III c>d	31	0,001	II c>d	32	0,0004	III c>d
R_2	α	0	0,0004	III 1<2	12	0,49	1<2	2	0,003	II 1<2
	β	32	0,0004	III c>d	29	0,007	II c>d	30	0,003	II c>d
R_3	α	2	0,003	II 1<2	19	0,62	1>2	6	0,05	1<2
	β	11½	0,43	c<d	24	0,13	c>d	17	0,92	c>d
e	α	0	0,0004	III 1<2	15	0,92	1<2	3	0,007	II 1<2
	β	27	0,03	I c>d	21½	0,32	c>d	26	0,05	c>d

Vergelijking van deze uitkomsten met die van rapport S 140 (tabel 3, blz. 6) laat zien, dat de resultaten vrijwel dezelfde zijn.

In tabel III vinden wij de resultaten van de toets van WILCOXON toegepast op de bodemstraling. Er ontbreekt echter in het materiaal een waarneming in groep 2c, 1e en 2e waarnemingspunt. In verband hiermede moeten we nu de exacte verdeling van U berekenen voor U_1 $(n_1 = 4,\ m_1 = 3)$ gecombineerd met U_2 $(n_2 = 4,\ m_2 = 4)$. Voor de het minst van het gemiddelde afwijkende gehele waarden van U die onder de nulhypothese met kansen hoogstens gelijk aan 0,05; 0,01 en 0,001 worden onder- resp. overschreden vinden we 4(24), 2(26) en 0(28).

Tabel III

Toets van Wilcoxon voor ieder der drie bodempunten afzonderlijk.

		1			2			3		
B	α	0	0,0008	III 1<2	$8\frac{1}{2}$	0,16	1<2	5	0,029	I 1<2
	β	10	0,45	c<d	16	0,75	c>d	$21\frac{1}{2}$	0,32	c>d

De resultaten blijken dezelfde te zijn als die van rapport S 140 (tabel 4, blz. 7.)

In de figuren 9 en 10 zijn de betrouwbaarheidsintervallen voor de verschillen van de gemiddelde B-waarden getekend.

De gelijkmatigheid van de vlam is in rapport S 140 reeds met een verdelingsvrije methode onderzocht (tabel 6, blz. 8). Dit behoeft hier dus niet meer beschouwd te worden.

4. Conclusies

Hoewel de overschrijdingskansen iets groter en de betrouwbaarheidsintervallen iets ruimer zijn dan met onderstelling van normaliteit werd gevonden worden alle conclusies van rapport S 140 door het voorgaande onderzoek bevestigd.

De conclusies zijn dus iets minder uitgesproken, maar hier staat tegenover, dat we de onderstelling, dat de waarnemingen normaal verdeeld zijn, een onderstelling waarvan we de juistheid vrijwel niet kunnen toetsen, hier niet behoeven te maken.

Een ernstig bezwaar is echter, dat er geen verdelingsvrije toetsingsmethode is voor het onderzoek van de interactie, zodat de interactie tussen brandstof en symmetrie op deze wijze niet getoetst kan worden.

Appendix.

In par. 3 van dit rapport toetsen wij

a. Hypothese α: I waarnemingen 1c-2c

II " 1d-2d.

b. Hypothese β: I " 1c-1d.

II " 2c-2d.

Aan de hand van een voorbeeld zullen we nu laten zien hoe de toetsingsgrootheden van I en II gecombineerd kunnen worden.

We kiezen daartoe de waarnemingen van spleet 4, R_2, gerangschikt naar opklimmende grootte.

Tabel IV

De waarnemingen van spleet 4, R_2.

1c	1d
8,60	6,85
8,85	6,90
8,95	7,--
9,35	7,75
2c	**2d**
10,--	7,55
10,15	7,85
10,75	8,80
11,--	9,15

Met behulp van de reeds eerder vermelde toets van WILCOXON vinden we de toetsingsgrootheden

a. I. $U_1 = 0$ $\quad U = 1$ $\quad$ b. I. $U_1 = 16$ $\quad U = 32$.

II. $U_2 = 1$ $\quad\quad\quad\quad$ II. $U_2 = 16$

We willen nu de tweezijdige overschrijdingkans van $U = U_1 + U_2$ bepalen. Daartoe gaan we uit van de exacte tabel voor éénzijdige overschrijdingskansen voor deze toets, vermeld in H.R. VAN DER VAART [1], tabel I, $n=4, m=4$. Hierin vinden we de eenzijdige overschrijdingskansen $P[\underline{U}_i \leq U_i]$ 1), waaruit eenvoudig de kansen $P[\underline{U}_i = U_i]$ afgeleid kunnen worden door 2 opeenvolgende

---

1) $P[\underline{x} \leq x]$ betekent: de kans dat de stochastische grootheid $\underline{x} \leq x$ is.

$P[\underline{x} = x]$ betekent: de kans dat de stochastische grootheid $\underline{x} = x$ is.

termen van elkaar af te trekken.

Van alle mogelijke combinaties van U_1 en U_2 die samen een bepaalde waarde U geven, bepalen we de bijbehorende P_1 en P_2, en vervolgens

$$\sum P_1 P_2 \; ; \text{ d.i } P[\underline{U} = U] \quad \text{voor} \quad U = 0, 1, \ldots, 16.$$

Hieruit vinden we door optellen $P[\underline{U} \leqq U]$.

Een methode die op hetzelfde neerkomt is het eerst aangegeven in WILCOXON [2] . Indien de gevonden waarde van U kleiner is dan de verwachte waarde onder de getoetste hypothese (i.c. 16, daar U tussen 0 en 32 ligt en er geen gelijke waarnemingen zijn), vinden wij de tweezijdige overschrijdingskans door de kans $P[\underline{U} \leq U]$ met twee te vermenigvuldigen.

Is $U > 16$, dan vinden we de rechtereenzijdige overschrijdingskans van $32-U$ te bepalen. Vanwege de symmetrie van de verdeling van $\underline{U}$ onder de getoetste hypothese ($\underline{U}$ is altijd symmetrisch verdeeld als er geen gelijke waarnemingen zijn) geeft dit het juiste antwoord. De gevonden kans vermenigvuldigen we dan weer met 2 om de tweezijdige overschrijdingskans te verkrijgen. In tabel V is de berekening der overschrijdingskansen uitgevoerd voor $U = 0, 1, 2$ en 3.

Tabel V

Berekening der tweezijdige overschrijdingskans van U.

$U = U_1 + U_2$	U_1	U_2	$P[\underline{U}_1 = U_1] = P_1$	$P[\underline{U}_2 = U_2] = P_2$	$P[\underline{U} = U] = \sum P_1 P_2$	$P[\underline{U} \leqq U]$	tweezijdige overschrijdingskans
0	0	0	1/70	1/70	1/4900	1/4900	2/4900 = 0,0004
1	0	1	1/70	1/70	2/4900	3/4900	6/4900 = 0,0012
	1	0	1/70	1/70			
2	0	2	1/70	2/70	5/4900	8/4900	16/4900 = 0,0033
	1	1	1/70	1/70			
	2	0	2/70	/70			
3	0	3	1/70	3/70	10/4900	18/4900	36/4900 = 0,0073
	1	2	1/70	2/70			
	2	1	2/70	1/70			
	3	0	3/70	1/70			

In het geval van ons voorbeeld vinden we dus:

a. Hypothese α.

$$\mathcal{U} = 1, \qquad k = 0{,}0012.$$

Conclusie: brandstof 1 geeft in deze situatie systematisch geringere stralingsintensiteit dan brandstof 2.

b. Hypothese β.

$$\mathcal{U} = 32, \qquad 32 - \mathcal{U} = 0, \qquad k = 0{,}0004.$$

Conclusie: de symmetrische vlam geeft in deze situatie systematisch een grotere stralingsintensiteit dan de asymmetrische.

Komen er gelijken onder de waarnemingen voor (in het boven uitgewerkte voorbeeld was dit niet het geval), dan kunnen we waarden voor $\mathcal{U}$ vinden die niet geheel zijn en dus niet in de tabel voorkomen waarvan tabel V het begin vormt. We hebben dan als benaderde overschrijdingskans de door lineaire interpolatie verkregen waarde genomen. Behalve door de in memorandum S 47(M 7) genoemde oorzaak zal ook door deze interpolatie, de gevonden overschrijdingskans iets groter zijn, zodat onze conclusies "aan de veilige kant" blijven, d.w.z. wij zullen niet vaker dan door de onbetrouwbaarheidsdrempel aangegeven wordt, ten onrechte tot verwerping van de getoetste hypothese overgaan.

Bij de bodemstraling waar we een ontbrekende waarneming hebben, zodat de ene toets toegepast wordt op 4 tegen 3 waarnemingen gaat de berekening van de exacte verdeling van $\mathcal{U}$ geheel analoog aan het voorgaande.

Literatuur:

[1] H.R. VAN DER VAART — Gebruiksaanwijzing voor de toets van Wilcoxon, Mathematisch Centrum, Statistische Afdeling, Rapport S 32 (M 4), 2e druk, 1952.

[2] F. WILCOXON — Individual Comparisons of Grouped Data by ranking Methods, Journal of Entomology, 39 (1946). 269-270.

FIG. 1

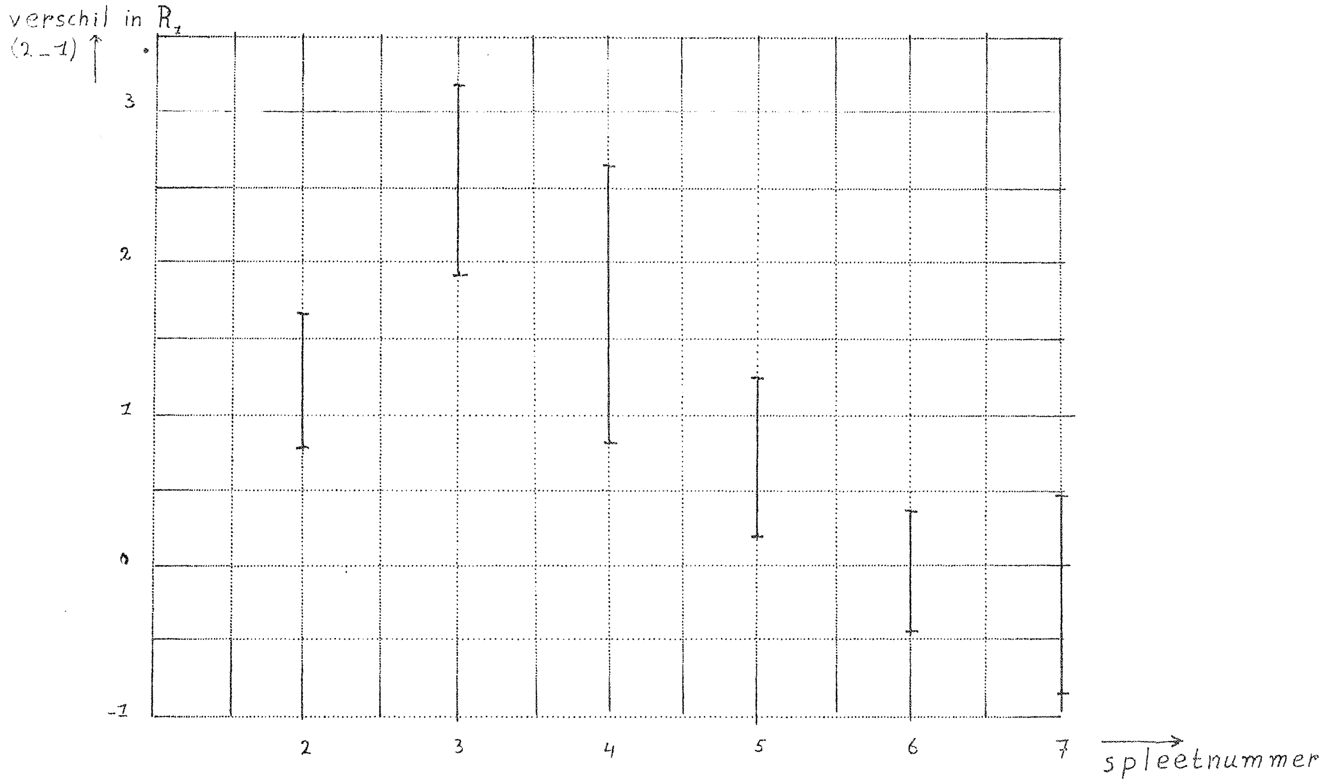

verschil in R_1
(2 – 1)
3
2
1
0
-1
2
3
4
5
6
7
spleetnummer

FIG. 2

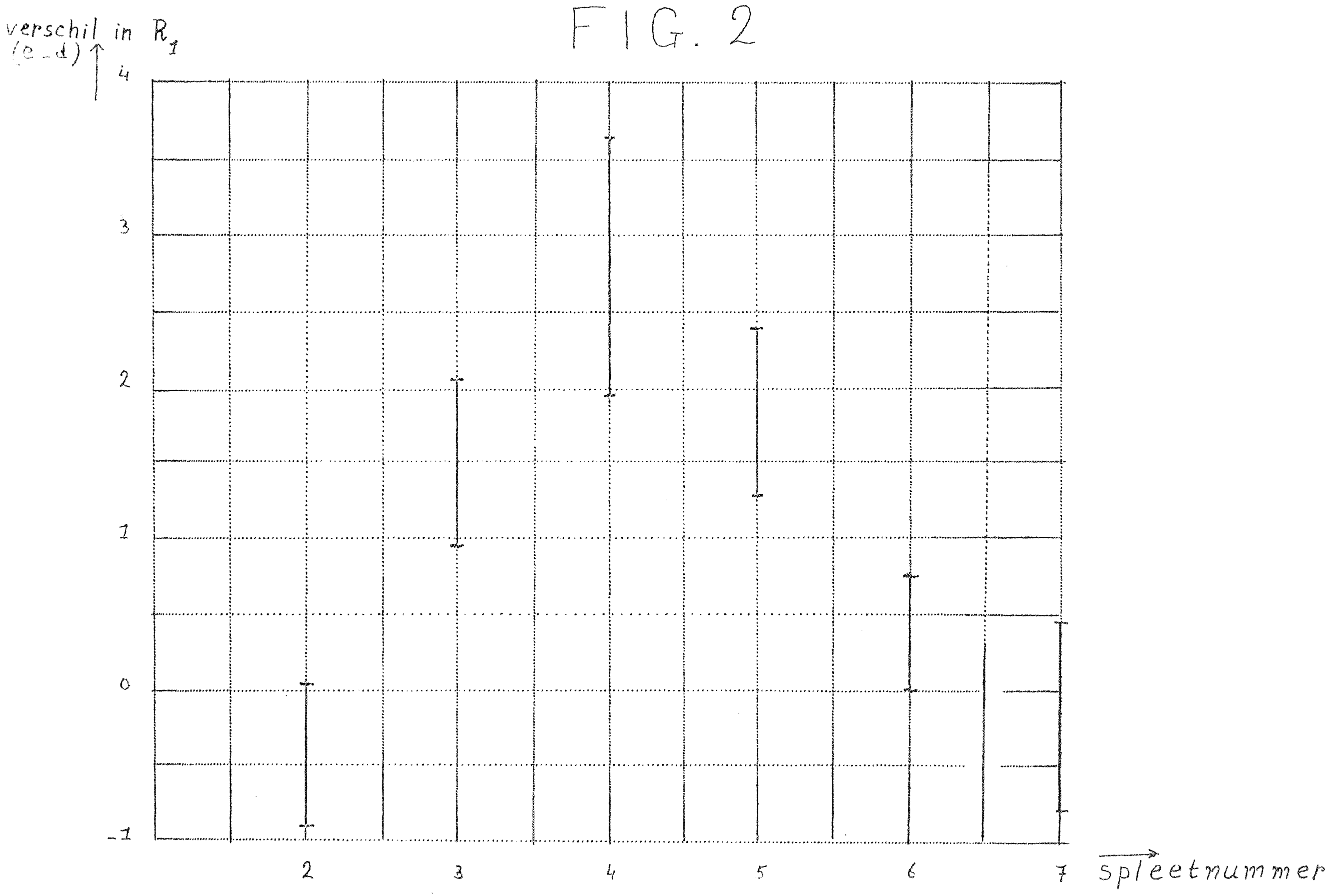
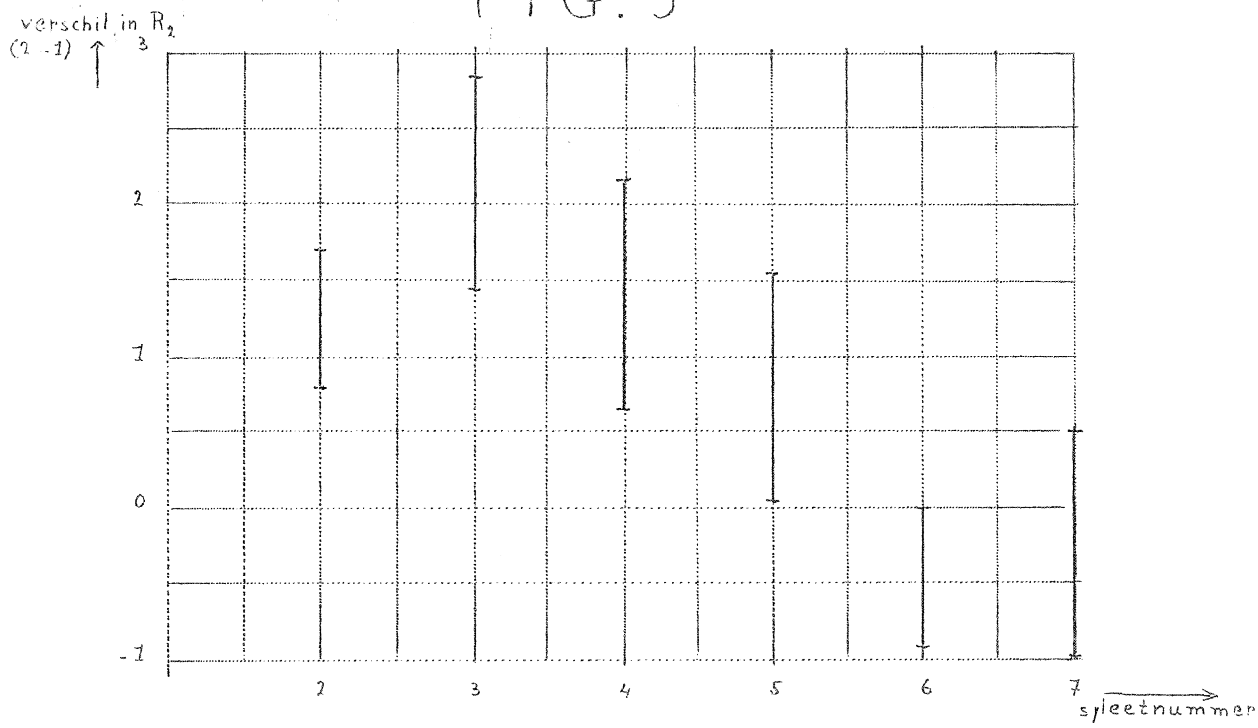
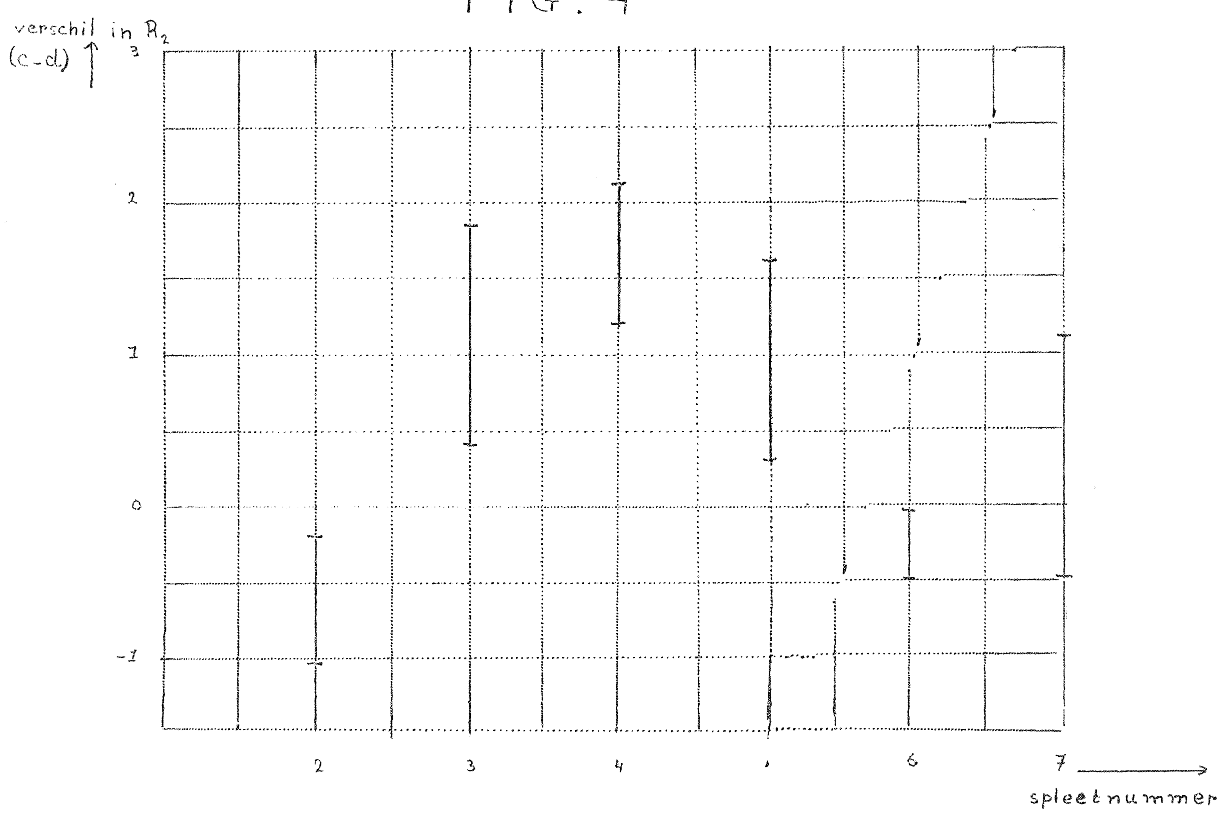

verschil in R_1
(e – d)
4
3
2
1
0
-1
2
3
4
5
6
7
spleetnummer

FIG. 3



verschil in R_2
(2 - 1)
3
2
1
0
-1
2
3
4
5
6
7
spleetnummer

FIG. 4



verschil in R_2
(c - d)
3
2
1
0
-1
2
3
4
6
7
spleetnummer

FIG. 5

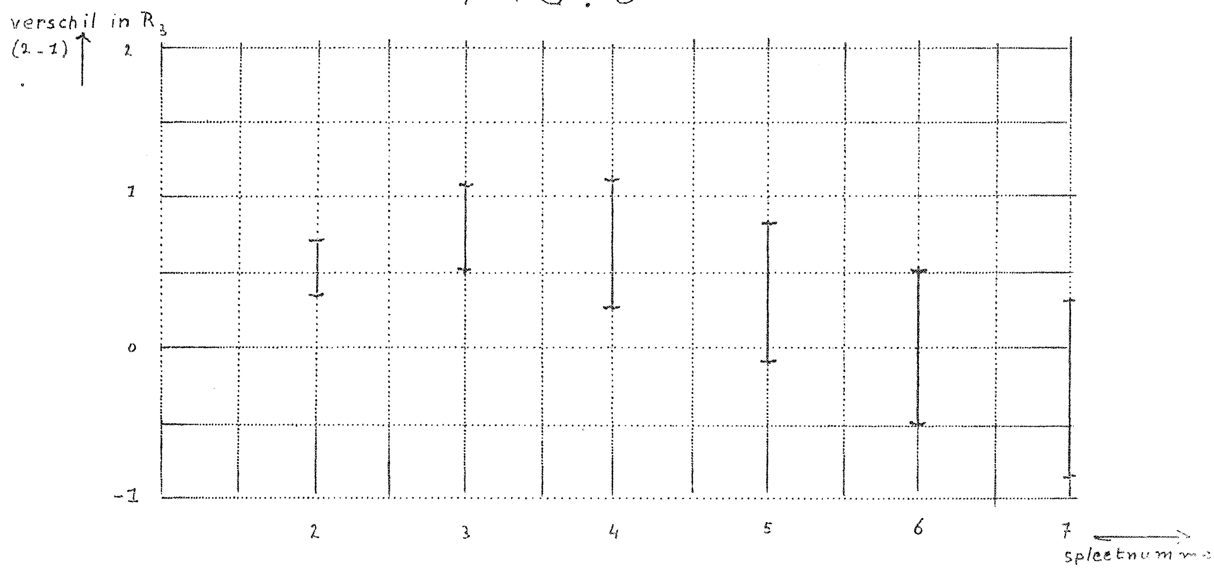
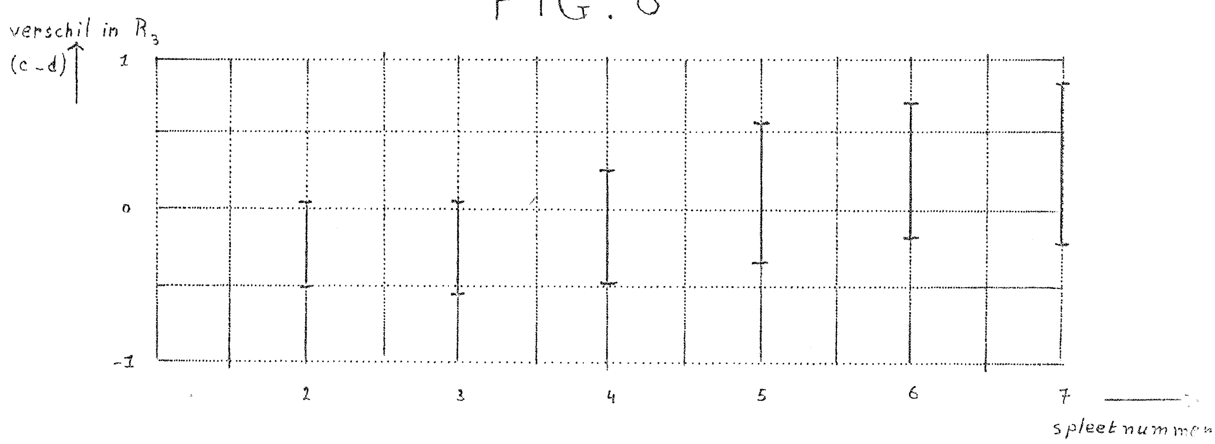
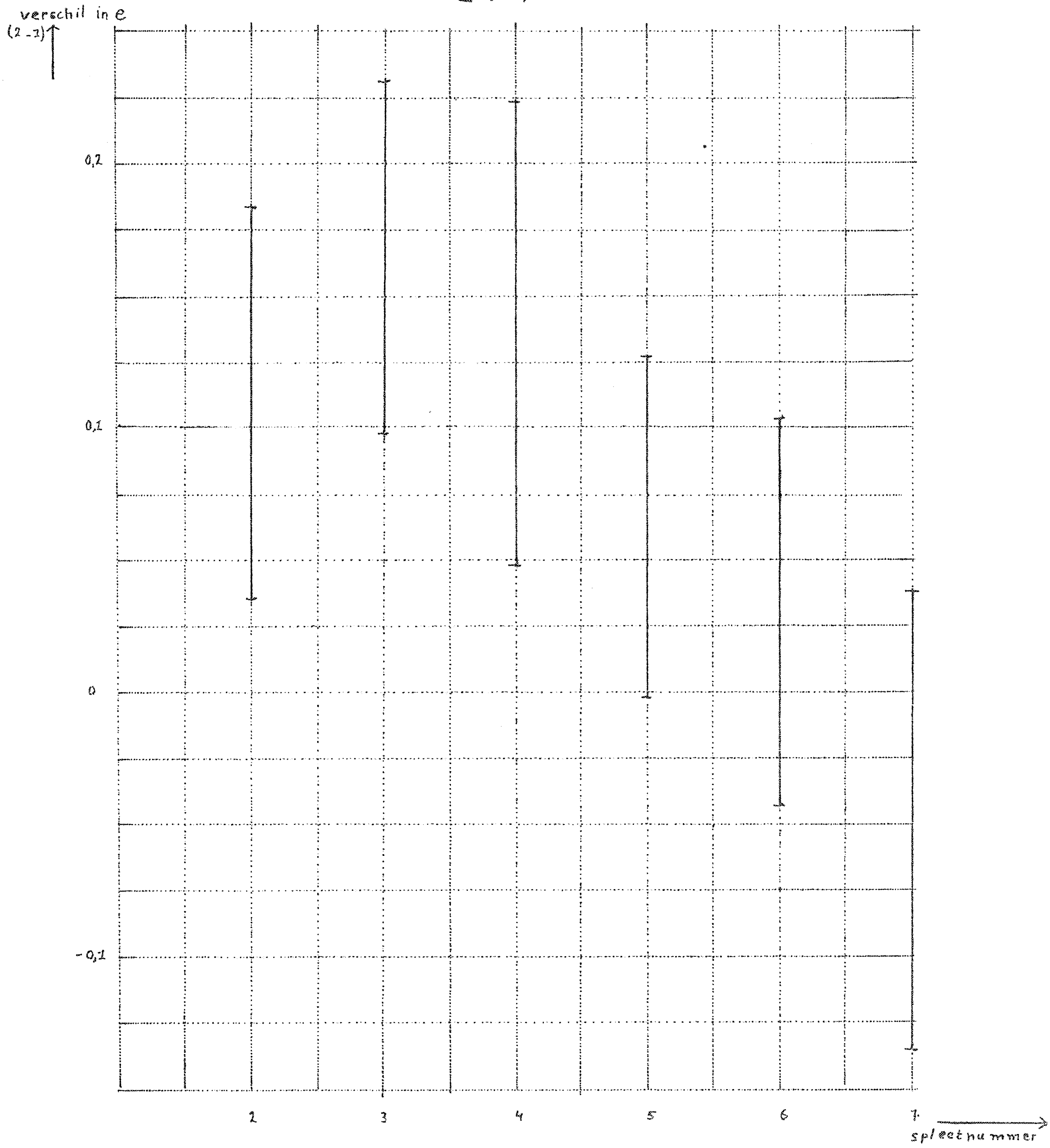

verschil in R_3
(2 - 1)
2
1
0
-1
2
3
4
5
6
7
spleetnummer

FIG. 6


verschil in R_3
(c - d)
1
0
-1
2
3
4
5
6
7
spleetnummer

F I G. 7



F I G. 8

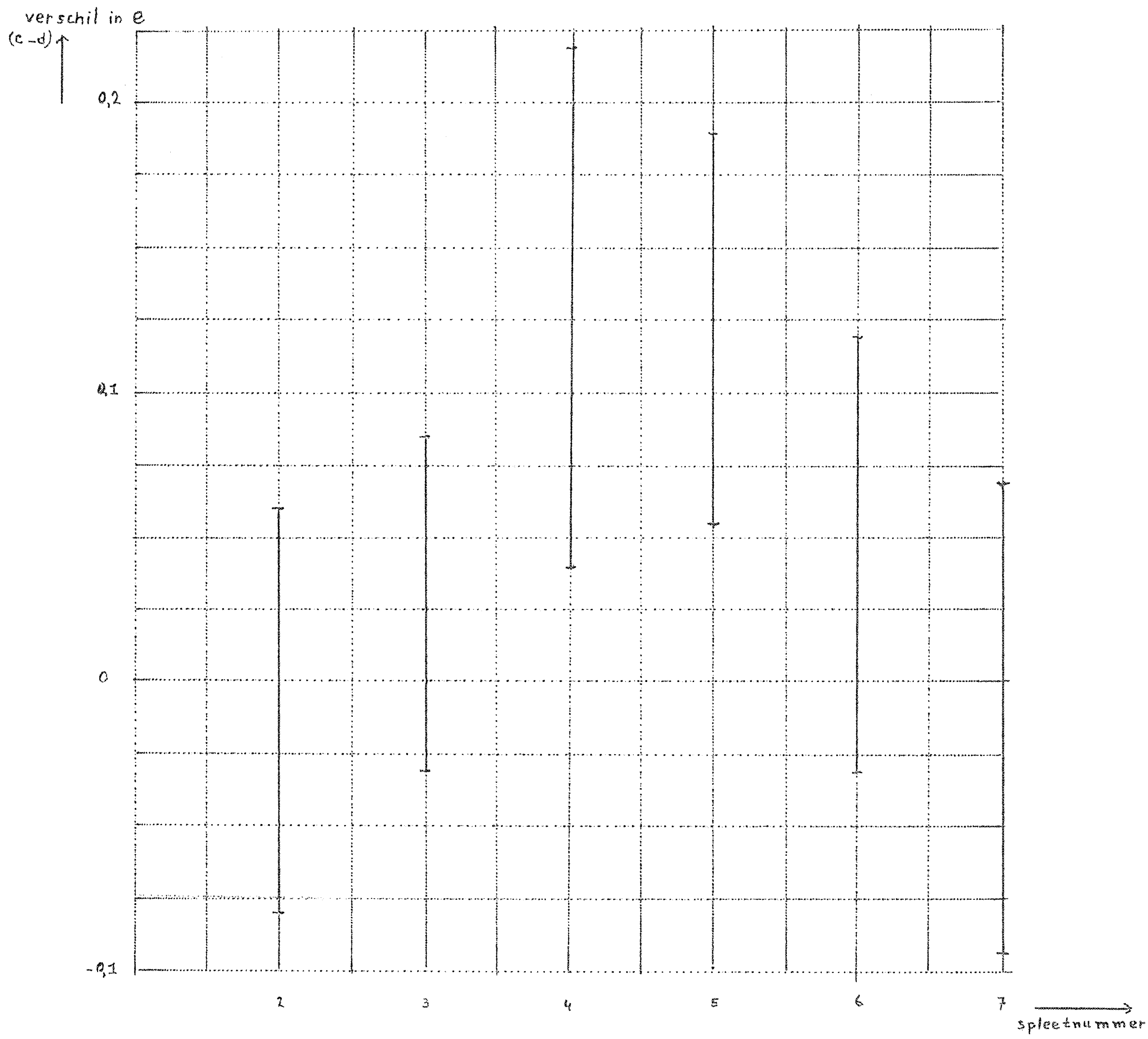


FIG. 9

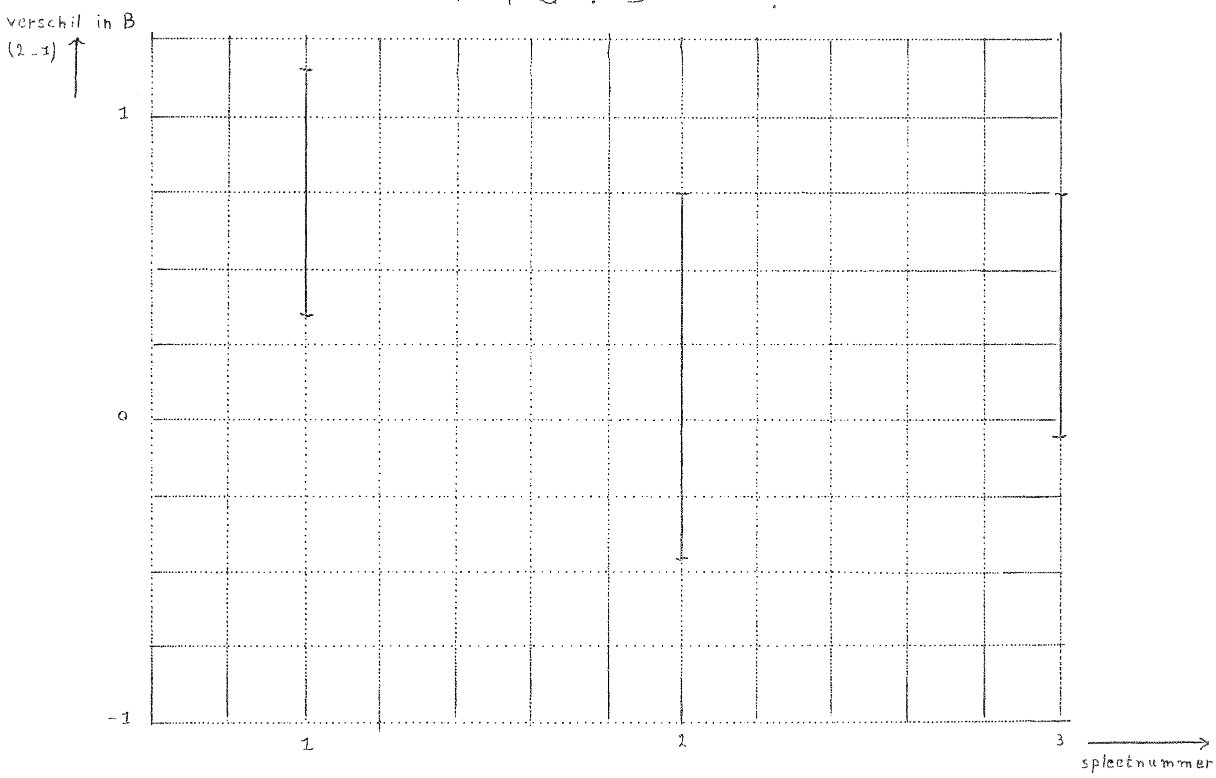
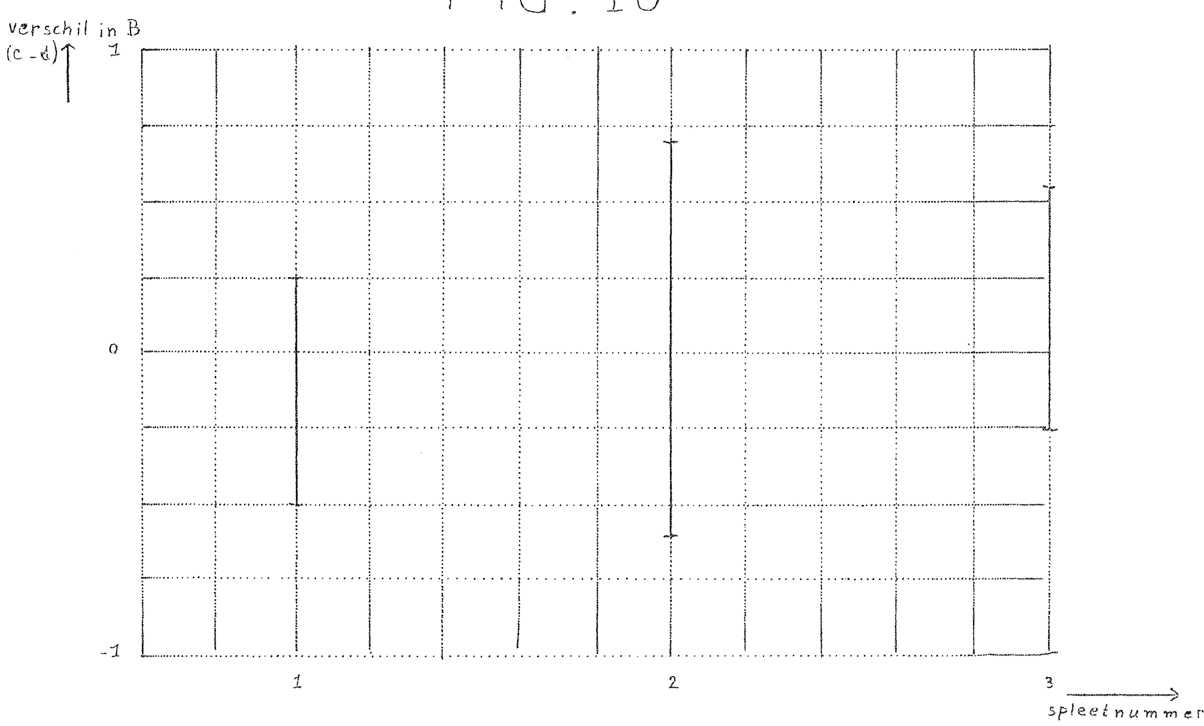

verschil in B
(2 - 1)
1
0
-1
1
2
3
spleetnummer

FIG. 10


verschil in B
(c - d)
1
0
-1
1
2
3
spleetnummer

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

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

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. 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 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 Boerhaavestraat 49,
A m s t e r d a m - O.
Rapport S 145 (M 52)
door Ph. van Elteren.

Constructie van betrouwbaarheidsintervallen met behulp van de toets van WILCOXON.

1. Betrouwbaarheidsinterval bij 2 steekproeven.

Gegeven zijn twee steekproeven:

$x_1, \ldots\ldots, x_n$ van de stochastische variabele $\underline{x}$ en
$y_1, \ldots\ldots, y_m$ " " " " $\underline{y}$.

Er wordt ondersteld dat $\underline{x}$ en $\underline{y}$ verdelingen hebben, die alléén in hun gemiddelden kunnen verschillen. Dus:

$$\mathcal{E}\underline{x} - \mathcal{E}\underline{y} = \delta$$

en $\underline{x}$ en $\underline{y} + \delta$ hebben dezelfde verdeling.

Gevraagd wordt op grond van de steekproeven een betrouwbaarheidsinterval met onbetrouwbaarheidsdrempel α voor δ te bepalen.

Deze vraag kan worden beantwoord met behulp van de toets van WILCOXON, welke wordt behandeld in memorandum S 47 (M 7) van het Mathematisch Centrum. Daarin wordt de toetsingsgrootheid $\underline{U}$ van WILCOXON voor steekproeven van de uitgebreidheid n van $\underline{x}$ en van de uitgebreidheid m van $\underline{y}$ gedefinieerd en de verdeling van $\underline{U}$ beschreven onder de hypothese H_o dat $\underline{x}$ en $\underline{y}$ dezelfde verdeling hebben. Als de verdeling van $\underline{U}$ onder H_o bekend is, kan de kritieke waarde U_ε worden bepaald, gedefinieerd als de grootste waarde, die $\underline{U}$ kan aannemen waarvoor geldt:

$$P\left\{\underline{U} \leqq U_\varepsilon \mid H_o\right\}^{1)} \leqq \varepsilon .$$

Voor de bepaling van het betrouwbaarheidsinterval gaan wij nu als volgt te werk:

1. Bepaal alle m n verschillen van een waarneming van $\underline{x}$ en een waarneming van $\underline{y}$, die men uit het waarnemingsmateriaal kan vormen,
2. Rangschik deze verschillen naar opklimmende grootte,
3. Bepaal voor ieder verschil $x_i - y_j$ de volgende aantallen:
   a = het aantal verschillen, gelijk aan dat verschil.
   b = het aantal verschillen, groter dan dat verschil.

Als wij de grens van een naar boven begrensd eenzijdig betrouwbaarheidsinterval willen bepalen,

1) Hieronder te verstaan de kans dat $\underline{U} \leqq U_\varepsilon$ is, als H_o geldt.

zoeken wij 2 opeenvolgende verschillen: d_1 en d_1' ($d_1 < d_1'$) zódanig dat $b_1 + \frac{1}{2}a_1 > U_\alpha$ en $b_1' + \frac{1}{2}a_1' \leqq U_\alpha$ is, waarbij a_1 en b_1 resp. a_1' en b_1' de bij d_1 resp. d_1' behorende aantallen a en b gedefinieerd in punt 3 zijn en α de gekozen onbetrouwbaarheidsdrempel is. Is nu $b_1 \leqq U_\alpha$, dan wordt het betrouwbaarheidsinterval $\delta \leq d_1$ en is dus d_1 een bovengrens, die bij het interval behoort. Is echter voor d_1: $b_1 > U_\alpha$, dan wordt het interval $\delta < d_1'$ en is dus d_1' een bovengrens die niet bij het interval behoort.
Als wij echter de grens van een naar beneden begrensd éénzijdig betrouwbaarheidsinterval willen bepalen, zoeken wij 2 opeenvolgende verschillen d_2 en d_2' ($d_2 > d_2'$) (met aantallen a en b, in 3 gedefinieerd, gelijk aan a_2 en b_2 resp. a_2' en b_2') zódanig dat: $b_2 + \frac{1}{2}a_2 < mn - U_\alpha$ en $b_2' + \frac{1}{2}a_2' \geqq mn - U_\alpha$ is.

Als nu $b_2' \geqq mn - U_\alpha$ wordt het betrouwbaarheidsinterval $\delta \geqq d_2$ en is dus d_2 een benedengrens, die bij het interval behoort, is daarentegen $b_2' < mn - U_\alpha$, dan wordt het interval $\delta > d_2'$ en is dus d_2' een benedengrens, die niet bij het interval behoort.

Indien men een tweezijdig begrensd betrouwbaarheidsinterval wil verkrijgen, bepaalt men als boven aangegeven een bovengrens en een benedengrens, maar dan ieder bij een onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$, als α de voorgeschreven onbetrouwbaarheid is.

Voorbeeld:

Waarnemingen van $\underline{x}$: 18,2; 16,2; 20,3; 17,4; 18,9; 21,4
" " $\underline{y}$: 16,7; 12,5; 15,2; 14,3; 15,9; 17,8.

Gevraagd wordt een tweezijdig begrensd betrouwbaarheidsinterval voor $\delta = \mathcal{E}\underline{x} - \mathcal{E}\underline{y}$ bij een onbetrouwbaarheidsdrempel $\alpha = 0{,}05$.

Oplossing: Wij vinden in een tabel van de verdeling van de toetsingsgrootheid $\underline{U}$ van WILCOXON (zie b.v. [1] van de literatuurlijst):

$$U_{0,025} = 5$$

$$mn - U_{0,025} = 36 - 5 = 31$$

Wij rangschikken nu de x- en de y-waarnemingen naar opklimmende grootte en geven de verschillen $x_i - y_j$ in onderstaande tabel:

Tabel der verschillen $x_i - y_j$ in het voorbeeld

		waarnemingen van $\underline{x}$					
		16,2	17,4	18,2	18,9	20,3	21,4
waarnemingen van $\underline{y}$	12,5	+3,7	+4,9	+5,7	+6,4	+7,8	+8,9
	14,3	+1,9	+3,1	+3,9	+4,6	+6,0	+7,1
	15,2	+1,0	+2,2	+3,0	+3,7	+5,1	+6,2
	15,9	+0,3	+1,5	+2,3	+3,0	+4,4	+5,5
	16,7	-0,5	+0,7	+1,5	+2,2	+3,6	+4,7
	17,8	-1,6	-0,4	+0,4	+1,1	+2,5	+3,6

Wij rangschikken nu de verschillen naar opklimmende grootte en bepalen de aantallen a en b boven vermeld:

	a	b		a	b		a	b		a	b
-1,6	1	35	+1,9	1	25	+4,4	1	12	+6,4	1	3
-0,5	1	34	+2,2	2	23	+4,6	1	11	+7,1	1	2
-0,4	1	33	+2,3	1	22	+4,7	1	10	+7,8	1	1
+0,3	1	32	+2,5	1	21	+4,9	1	9	+8,9	1	0
+0,4	1	31	+3,0	2	19	+5,1	1	8			
+0,7	1	30	+3,1	1	18	+5,5	1	7			
+1,0	1	29	+3,6	2	16	+5,7	1	6			
+1,1	1	28	+3,7	2	14	+6,0	1	5			
+1,5	2	26	+3,9	1	13	+6,2	1	4			

Wij vinden nu:

voor $d_1 = 6{,}0$: $b_1 + \frac{1}{2}a_1 = 5\frac{1}{2}$

voor $d_1' = 6{,}2$: $b_1' + \frac{1}{2}a_1' = 4\frac{1}{2}$

$b_1 = 5 = U_{0{,}025}$ dus $\delta \leqq d_1 = 6{,}0$

voor $d_2 = 0{,}7$: $b_2 + \frac{1}{2}a_2 = 30\frac{1}{2}$

voor $d_2' = 0{,}4$: $b_2' + \frac{1}{2}a_2' = 31\frac{1}{2}$

$b_2' = 31 = mn - U_{0{,}025}$ dus $\delta \geqq d_2 = 0{,}7$

Het tweezijdig begrensde betrouwbaarheidsinterval wordt dus:

$$+0{,}7 \leqq \delta \leqq +6{,}0.$$

N.B. De feitelijke eenzijdige overschrijdingskans van U = 5 bij m = n = 6 is 0,02. Bij een onbetrouwbaarheidsdrempel 0,04 zouden wij dus hetzelfde interval gevonden hebben. Het verdient aanbeveling om deze kleinst mogelijke onbetrouwbaarheid op te geven, aangezien deze aanzienlijk lager kan zijn dan de voorgeschreven waarde.

In vele gevallen kan de waarde van U_ε alleen bij benadering bepaald worden (Zie [1]) In die gevallen gelden de betrouwbaarheidsgrenzen uiteraard ook alleen bij benadering.

2. Betrouwbaarheidsinterval bij aan aantal paren steekproeven.

Gegeven zijn k paren stochastische grootheden $\underline{x}_i, \underline{y}_i$ $(i = 1,\ldots\ldots,k)$. Wij beschikken over onderling onafhankelijke steekproeven van ieder van de grootheden:

$$\left.\begin{array}{l} x_{i,1},\ldots.x_{i,n_i} \text{ van } \underline{x}_i \\ y_{i,1},\ldots.y_{i,m_i} \text{ van } \underline{y}_i \end{array}\right\} (i = 1,\ldots\ldots,k)$$

Er wordt ondersteld, dat $\underline{x}_i$ en $\underline{y}_i$ verdelingen hebben, die alleen in hun gemiddelden kunnen verschillen, terwijl deze verschillen dan voor alle i gelijk zijn. Dus:

$$\mathcal{E}\underline{x}_i - \mathcal{E}\underline{y}_i = \delta \quad (i = 1,\ldots\ldots,k)$$

en $\underline{x}_i$ en $\underline{y}_i + \delta$ hebben dezelfde verdeling (Deze verdeling behoeft voor verschillende waarden van i niet hetzelfde te zijn).

Gevraagd wordt op grond van de steekproeven een betrouwbaarheidsinterval voor δ, met onbetrouwbaarheidsdrempel α te bepalen.

Deze vraag kan worden beantwoord met behulp van een combinatie van k onderling onafhankelijke toetsen van WILCOXON. De methode van het combineren van onderling onafhankelijke toetsen wordt behandeld in memorandum S 102 (M 17). Men kan daarin het volgende vinden: Zij c_i $(i = 1,\ldots\ldots,k)$ een gegeven constante en U_i $(i = 1,\ldots\ldots,k)$ de toetsingsgrootheid U van WILCOXON voor het paar steekproeven $x_{i,1},\ldots,x_{in_i}$; $y_{i,1}\ldots\ldots,y_{in_i}$, dan heeft

$$\underline{T} = \sum_{i=1}^{k} c_i \underline{U}_i$$

onder de hypothese H_0 dat voor iedere i $\underline{x}_i$ en $\underline{y}_i$ dezelfde verdeling hebben, bij benadering een normale verdeling met gemiddelde: $\sum_{i=1}^{k} c_i \mu_i$ en variantie $\sum_{i=1}^{k} c_i^2 \sigma_i^2$,

waarin $\mu_i = \frac{1}{2} m_i n_i$

en $\sigma^2_i = \frac{1}{12} m_i n_i \,(m_i + n_i + 1)$

(eventueel met een correctie van gelijke waarnemingen, zie memorandum S 47 (M 7)), de verwachting resp. de variantie onder H_0 voorstellen van de toetsingsgrootheid $\underline{U}_i$. Indien wij kiezen $c_1 = c_2 = \ldots = c_k = 1$, kan men hieruit onmiddelijk de verdeling van $\underline{T} = \sum_i \underline{U}_i$ onder de hypothese H_0 afleiden, dus ook de grootste waarde T_ε, die $\underline{T}$ kan aannemen en waarvoor geldt:

$$P\{\underline{T} \leqq T_\varepsilon \mid H_0\} \leqq \varepsilon .$$

Voor de bepaling van het betrouwbaarheidsinterval gaan wij als volgt te werk:

1. Bepaal alle $\sum_{i=1}^{k} m_i n_i$ verschillen van een waarneming van $\underline{x}_i$ en een waarneming van de overeenkomstige stochastische grootheid $\underline{y}_i$,
2. Rangschik de verschillen naar opklimmende grootte,
3. Bepaal voor ieder verschil $x_{ij} - y_{il}$:

   a = het aantal verschillen, gelijk aan dat verschil,

   b = het aantal verschillen, groter dan het verschil.

   Men gaat verder geheel te werk als beschreven is in par. 1 voor één paar steekproeven, met dien verstande dat men voor U_α moet lezen $\underline{T}_\alpha$.

<u>Voorbeeld:</u>

k = 2

waarnemingen van $\underline{x}_1$:	5,65	6,20	5,05	5,50
" " $\underline{y}_1$:	3,00	3,15	2,85	3,20
" " $\underline{x}_2$:	7,05	8,40	7,35	7,60
" " $\underline{y}_2$:	3,95	5,65	4,35	3,95

<u>Gevraagd:</u> wordt weer een tweezijdig begrensd betrouwbaarheidsinterval voor $\delta = \mathcal{E}\underline{x}_i - \mathcal{E}\underline{y}_i$ bij een onbetrouwbaarheidsdrempel $\alpha = 0{,}05$.

<u>Oplossing:</u> Wij vinden met behulp van een tabel van de verdeling van de toetsingsgrootheid $\underline{U}$ voor de kritieke waarden:

$$T_{0{,}025} = 4 \quad \text{en} \quad \sum_{i=1}^{2} m_i n_i - T_{0{,}025} = 32 - 4 = 28 .$$

(deze kritieke waarden zijn niet bepaald met behulp van bovenbeschreven benadering, doch afgeleid uit de exacte verdeling van $\underline{U}_1 + \underline{U}_2$, die in dit geval gemakkelijk te bepalen is).

Wij rangschikken de x_i en y_i waarnemingen weer naar opklimmende grootte en bepalen de verschillen:

Tabel der verschillen $\underline{x}_{ij} - \underline{y}_{il}$.

	waarnemingen van y_1: 2,85	3,--	3,15	3,20
waarnemingen van $\underline{x}_1$: 5,05	2,20	2,05	1,90	1,85
5,50	2,65	2,50	2,35	2,30
5,65	2,80	2,65	2,50	2,45
6,20	3,35	3,20	3,05	3,--

	waarnemingen van y_2: 3,95 (2)	4,35	5,65
waarnemingen van $\underline{x}_2$: 7,05	3,10 (2)	2,70	1,40
7,55	3,60 (2)	3,20	1,90
7,60	3,65 (2)	3,25	1,95
8,40	4,45 (2)	4,05	2,75

(Waarnemingen en verschillen waarachter het cijfer (2) vermeld is komen 2 maal voor).

Wij rangschikken nu de verschillen naar opklimmende grootte en bepalen de aantallen a en b gedefinieerd in par.1:

	a	b		a	b		a	b		a	b
1,40	1	31	2,30	1	24	2,75	1	16	3,25	1	8
1,85	1	30	2,35	1	23	2,80	1	15	3,35	1	7
1,90	2	28	2,45	1	22	3,--	1	14	3,60	2	5
1.95	1	27	2,50	2	20	3,05	1	13	3,65	2	3
2,05	1	26	2,65	2	18	3,10	2	11	4,05	1	2
2,20	1	25	2,70	1	17	3,20	2	9	4,45	2	0

Wij vinden nu:

voor $d_1 = 3{,}60 \quad b_1 + \frac{1}{2}a_1 = 6$

" $d_1' = 3{,}65 \quad b_1' + \frac{1}{2}a_1' = 4$

$b_1 = 5 > T_{0{,}025} = 4$ dus $\delta < d_1' = 3{,}65$

voor $d_2 = 1{,}95 \quad b_2 + \frac{1}{2}a_2 = 27\frac{1}{2}$
" $d_2' = 1{,}90 \quad b_2' + \frac{1}{2}a_2' = 29$
$b_2' = 28 = \sum_{i=1}^{2} m_i n_i - T_{0{,}025} = 28$ dus $\delta \geqq d_2 = 1{,}95$.

Het tweezijdig begrensde betrouwbaarheidsinterval wordt dus:

$$1{,}95 \leqq \delta < 3{,}65$$

Literatuur:

[1] H.R. VAN DER VAART Gebruiksaanwijzing voor de toets van WILCOXON, Rapport S32 (M 4) van het Mathematisch Centrum 2e druk 1952.