

STICHTING
MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Rapport S 262 (M 83)

Handleiding voor de rangcorrelatietoets van KENDALL

door

Dr Constance van Eeden

en

Rina Korswagen

December 1959

<u>Inhoud</u>	pag.
1. Inleiding.	1
2. Het waarnemingsmateriaal en de getoetste hypothese.	1
3. De toetsingsgrootheid.	3
4. Een grafische methode voor de berekening van S .	11
5. De verdeling van $\underline{S}$ onder de getoetste hypothese H_0 .	
5.1. De verdeling van $\underline{S}$ onder H_0 als in geen van beide reeksen gelijken voorkomen.	16
5.2. De verdeling van $\underline{S}$ onder H_0 als in één of beide reeksen gelijken voorkomen.	18
6. Kritieke zones en overschrijdingskansen.	20
7. Tabellen voor kritieke waarden en overschrijdingskansen.	21
8. Benaderingen voor kritieke waarden en overschrijdingskansen.	23
9. Uitwerking van de voorbeelden.	25
Literatuur	29

Tabellen

1. Waarden van $10^3 \cdot P[\underline{S} \geq S n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$	30
2. Waarden van $10^3 \cdot P[\underline{S} \geq S n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$	35
3. Exacte rechter kritieke waarden van S voor $n=4(1)40$ en $\alpha = 0,005; 0,01; 0,025; 0,05; 0,10$ (eenzijdig)	40
4. Benaderingen voor de rechter kritieke waarden van S voor $n=41(1)100$ en $\alpha = 0,005; 0,01; 0,025; 0,05; 0,10$ (eenzijdig)	41
5. Tabel van $\sigma^2 = \frac{1}{18} n(n-1)(2n+5)$ en σ voor $n = 1(1)100$	43
6. Tabel van t^2 en t^3 voor $t = 1(1)40$	44
7. Tabel van de normale verdeling	45

1. Inleiding.

In dit rapport wordt een beschrijving gegeven van de rangcorrelatietoets van M.G. KENDALL (1955, hoofdstuk 1 t/m 5). Deze toets is o.a. een verdelingsvrije toets voor de hypothese dat twee stochastische grootheden $\underline{x}$ en $\underline{y}$ ¹⁾ onderling onafhankelijk verdeeld zijn. "Verdelingsvrij" wil zeggen dat geen onderstellingen gemaakt worden over de vorm van de simultane verdeling van $\underline{x}$ en $\underline{y}$. Dit is het belangrijkste verschil met de onafhankelijkheidstoets gebaseerd op de product-moment-correlatiecoëfficiënt, die slechts toepasbaar is als $\underline{x}$ en $\underline{y}$ een twee-dimensionale normale verdeling bezitten.

Verder wordt bij de rangcorrelatietoets alleen gebruik gemaakt van de volgorde naar grootte van de waarnemingen. De toets kan dus ook worden toegepast op gegevens waarbij alleen een volgorde bepaald kan worden, zonder dat aan iedere waarneming een bepaalde waarde kan worden toegekend.

Tenslotte kan de toets ook worden toegepast als één van de twee grootheden $\underline{x}$ en $\underline{y}$ niet stochastisch is, maar een van tevoren gegeven objectieve volgorde bepaalt, bijv. opeenvolgende tijdstippen van waarneming der andere grootheid. Men kan dan onderzoeken of de andere grootheid een stijgend of dalend verloop vertoont.

Dit rapport bevat, behalve een beschrijving van de toets, een uitgebreide tabellenverzameling betreffende de exacte en een benaderende verdeling van de toetsingsgrootheid.

Een globale kennis van de gang van zaken bij het toetsen van hypothesen wordt voorondersteld.

2. Het waarnemingsmateriaal en de getoetste hypothese.

Het waarnemingsmateriaal bestaat uit n waarnemingsparen $(\underline{x}_1, \underline{y}_1), \dots, (\underline{x}_n, \underline{y}_n)$ en de te toetsen hypothese luidt

H_0 : bij gegeven volgorde van de x -waarnemingen zijn alle permutaties van de y -waarnemingen even waarschijnlijk.

Voorbeelden van situaties, waarin men deze hypothese kan

1) Stochastische grootheden worden onderscheiden van getallen (bijv. van de waarden, die zij bij een experiment aannemen) door hun symbolen te onderstrepen.

toetsen zijn de volgende:

1. Stel dat $(\underline{x}_1, \underline{y}_1), \dots, (\underline{x}_n, \underline{y}_n)$ onderling onafhankelijke waarnemingsparen zijn van de stochastische grootheden $\underline{x}$ en $\underline{y}$, die een simultane verdeling bezitten. Dan is H_0 vervuld als de hypothese

H'_0 : $\underline{x}$ en $\underline{y}$ zijn onderling onafhankelijk verdeeld

vervuld is. Men kan dus H'_0 toetsen door H_0 te toetsen.

In dit geval leidt de toets speciaal dan tot verwerping van H_0 als $\underline{x}$ en $\underline{y}$ positief of negatief gecorreleerd zijn.

2. Laat de x -waarnemingen een van tevoren gegeven volgorde bepalen, bijv. de tijdstippen van waarneming in de y -reeks. Laat verder $y_1, \dots, y_n$ resp. waarnemingen zijn van de grootheden $\underline{y}_1, \dots, \underline{y}_n$, dan is de hypothese H_0 vervuld als de hypothese

H''_0 : $\underline{y}_1, \dots, \underline{y}_n$ bezitten dezelfde verdeling

vervuld is. De hypothese H''_0 kan dus getoetst worden door H_0 te toetsen. In dit geval leidt de toets speciaal tot verwerping van de getoetste hypothese als $\underline{y}_1, \dots, \underline{y}_n$ een stijgend of dalend verloop vertonen.

De grootheden $\underline{y}_1, \dots, \underline{y}_n$ behoeven dus, om de toets te kunnen toepassen, niet onafhankelijk te zijn. Beschouw bijv. het geval van een vaas met n van 1 tot n genummerde briefjes. Deze briefjes worden zonder teruglegging uit de vaas getrokken en in volgorde van trekking neergelegd. Is nu $\underline{y}_i$ het resultaat van de i^e trekking dan zijn $\underline{y}_1, \dots, \underline{y}_n$ afhankelijk. Als echter de trekking aselekt geschiedt dan bezitten $\underline{y}_1, \dots, \underline{y}_n$ dezelfde verdeling, zodat men, door toepassing van de toets, de aseleetheid van de trekking kan toetsen. De toets is dan speciaal gevoelig voor selectheid in die zin, dat de kans om een hoog nummer te trekken toeneemt (of afneemt) bij stijgend nummer van trekking.

3. De toetsingsgrootte.

De toetsingsgrootte S wordt als volgt gedefiniëerd: De waarnemingsparen (x_i, y_i) en (x_j, y_j) met $i < j$ leveren tot S een bijdrage

$$s_{i,j} = \begin{cases} +1 & \text{als } (x_i - x_j)(y_i - y_j) > 0 \\ 0 & \text{als } (x_i - x_j)(y_i - y_j) = 0 \\ -1 & \text{als } (x_i - x_j)(y_i - y_j) < 0 \end{cases}$$

en de grootheid S is dus

$$S = \sum_{i < j} s_{i,j}.$$

Of, anders geformuleerd:

S is het aantal paren (i, j) met $i < j$, waarbij de volgorde naar grootte van de waarnemingen in de twee reeksen overeenstemt, verminderd met het aantal paren (i, j) met $i < j$, waarbij de volgorde naar grootte van de waarnemingen in de twee reeksen tegengesteld is. Paren (i, j) , waarbij de waarnemingen in één of beide reeksen gelijk zijn, leveren geen bijdrage tot S .

Komen in geen van beide waarnemingsreeksen gelijke waarnemingen voor, dan geldt:

$$(3;1) \quad -\frac{1}{2} n(n-1) \leq S \leq \frac{1}{2} n(n-1).$$

Immers: S is maximaal als de volgorde in grootte van de x -waarnemingen overeenstemt met de volgorde in grootte van de y -waarnemingen; dan is $S = (n-1) + (n-2) + \dots + 1 = \frac{1}{2}n(n-1)$. De grootheid S is minimaal als de volgorde in grootte van de x -waarnemingen tegengesteld is aan die van de y -waarnemingen; dan is $S = -(n-1) - (n-2) - \dots - 1 = -\frac{1}{2}n(n-1)$.

Behalve de grootheid S wordt door KENDALL de rangcorrelatiecoëfficiënt τ ingevoerd. Treden in geen van beide waarnemingsreeksen gelijken op dan is

$$(3;2) \quad \tau = \frac{2S}{n(n-1)}$$

en uit (3;1) volgt dan

$$(3;3) \quad -1 \leq \tau \leq 1.$$

Komen in één of beide reeksen gelijke waarnemingen voor dan wordt τ als volgt gedefiniëerd:

Laat de x -reeks (resp. de y -reeks) bestaan uit k (resp. h) groepen gelijke waarnemingen ter grootte $t_1, \dots, t_k$ (resp. $u_1, \dots, u_h$). Dan is

(3;4)

$$n = \sum_{\nu=1}^k t_{\nu} = \sum_{\kappa=1}^h u_{\kappa}.$$

Stel verder

(3;5)

$$C_1 = \sum_{\nu=1}^k t_{\nu}^2, \quad C_2 = \sum_{\kappa=1}^h u_{\kappa}^2$$

dan is

(3;6)

$$\tau = \frac{2S}{\sqrt{(n^2 - C_1)(n^2 - C_2)}}.$$

Ook in dit geval geldt $-1 \leq \tau \leq 1$, maar de waarden +1 en -1 kunnen niet altijd worden bereikt.

Formule (3;6) gaat, voor het geval in geen van beide reeksen gelijken voorkomen, over in (3;2). Immers dan zijn alle t_{ν} en alle u_{κ} gelijk aan 1; dus $C_1 = C_2 = n$.

De berekening van S en τ wordt nu aan de hand van enige voorbeelden nader toegelicht.

Voorbeeld 1.²⁾

Stel dat een aantal leerlingen gerangschikt is volgens hun bekwaamheid in wiskunde en muziek. De leerlingen worden aangegeven met de letters A,B,...,J; de rangschikking wordt aangegeven door rangnummers naar opklimmende bekwaamheid.

Laat het resultaat van deze rangschikking het volgende zijn:

Leerling	A	B	C	D	E	F	G	H	I	J
wiskunde	7	4	3	10	6	2	9	8	1	5
muziek	5	7	3	10	1	9	6	2	8	4

Als men nu wil nagaan of uit dit materiaal blijkt dat er een overeenstemming is tussen bekwaamheid in wiskunde en muziek, kan men KENDALL's rangcorrelatietoets toepassen.

Om de grootheid S te berekenen zet men eerst de kolommen in zodanige volgorde, dat één van de twee waarnemingsreeksen in de natuurlijke volgorde komt te staan. Door zulke permutaties verandert S niet en de berekening wordt eenvoudiger. In ons voorbeeld

2) Dit voorbeeld is ontleend aan M.G. KENDALL (1955, p.3).

krijgen we dan:

Leerling	I	F	C	B	J	E	A	H	G	D
wiskunde	1	2	3	4	5	6	7	8	9	10
muziek	8	9	3	7	4	1	5	2	6	10

Daar nu in de eerste reeks $x_i - x_j < 0$ is voor ieder paar (i, j) met $i < j$, wordt S gevonden door in de tweede reeks bij ieder rangnummer te tellen

- 1) het aantal rangnummers dat erachter staat en groter is,
- 2) het aantal rangnummers dat erachter staat en kleiner is.

Dit geeft

Leerling	I	F	C	B	J	E	A	H	G	D	totaal
groter	2	1	5	1	3	4	2	2	1	0	21 (=P)
kleiner	7	7	2	5	2	0	1	0	0	0	24 (=Q)

Noemen we de som van de aantallen in de eerste regel (dus het aantal paren (i, j) met $i < j$ en $y_i < y_j$) P en de som van de aantallen in de tweede regel (dus het aantal paren (i, j) met $i < j$ en $y_i > y_j$) Q , dan is

$$S = P - Q = 21 - 24 = -3.$$

Verder is, daar in geen van beide reeksen gelijken voorkomen,

$$\tau = \frac{2S}{n(n-1)} = -\frac{1}{15}.$$

Voorbeeld 2.

Laat de volgende 10 waarnemingsparen gegeven zijn:

x:	4	6	2	3	9	5	1	6	3	3
y:	12	23	12	23	5	20	16	7	18	20

Bij dit voorbeeld komen dus zowel in de x-reeks als in de y-reeks gelijken voor.

Om S te berekenen worden de kolommen in zodanige volgorde geplaatst dat $x_1 \leq \dots \leq x_n$; daar paren (i, j) met $x_i = x_j$ geen bijdrage tot S leveren, kunnen deze paren in een willekeurige volgorde geplaatst worden. Dit geeft

x:	1	2	3	3	3	4	5	6	6	9
y:	16	12	23	18	20	12	20	23	7	5

We tellen nu, voor iedere waarneming in de y-reeks

- het aantal waarnemingen, dat erachter staat en een positieve bijdrage geeft,
- het aantal waarnemingen, dat erachter staat en een negatieve bijdrage geeft.

Ter controle tellen wij ook het aantal waarnemingen dat een bijdrage 0 geeft:

											totaal
+	5	5	0	2	1	2	1	0	0	0	16 (=P)
-	4	2	4	3	3	2	2	1	1	0	22 (=Q)
0	0	1	3	1	1	0	0	1	0	0	7 (=R)
totaal	9	8	7	6	5	4	3	2	1	0	45

Hierbij dient men te bedenken dat een paar (i, j) met $x_i = x_j$ een bijdrage 0 geeft. Achter de tweede waarneming 23 b.v. staan twee waarnemingen; hiervan geeft er één een negatieve bijdrage (de waarneming 5) en één een bijdrage 0 (de waarneming 7, omdat voor dit paar (i, j) $x_i = x_j$ is).

Is nu P resp. Q het totaal van de aantallen in de eerste, resp. tweede regel, dan is

$$S = P - Q = 16 - 22 = -6.$$

Is verder R het totaal van de aantallen in de derde regel, dan is $P+Q+R$ het totaal aantal paren (i, j) met $i < j$, dus

$$P + Q + R = \frac{1}{2} n(n-1).$$

Tenslotte is de som van de aantallen in een kolom gelijk aan het aantal kolommen dat erachter staat, zodat in de onderste totaalregel de getallen $n-1, n-2, \dots, 0$ komen te staan.

Deze laatste berekeningen zijn niet nodig voor de berekening van S, maar wel nuttig ter controle.

Daar in beide reeksen gelijke waarnemingen voorkomen, wordt τ berekend volgens formule (3;6). In de eerste reeks komen 7 groepen gelijken voor, nl. de waarnemingen 1, 2, 3, 4, 5, 6 en 9. Dus

$$t_1 = t_2 = t_4 = t_5 = t_7 = 1, \quad t_3 = 3, \quad t_6 = 2, \quad C_1 = 5 \cdot 1^2 + 3^2 + 2^2 = 18.$$

In de tweede reeks komen eveneens 7 groepen gelijken voor, nl. de waarnemingen 5, 7, 12, 16, 18, 20 en 23 met

$$u_1 = u_2 = u_4 = u_5 = 1, \quad u_3 = u_6 = u_7 = 2, \quad C_2 = 4 \cdot 1^2 + 3 \cdot 2^2 = 16.$$

Dit geeft

$$\tau = - \frac{2 \cdot 6}{\sqrt{(100-18)(100-16)}} = -0.14.$$

Opmerkingen

1. Men kan ook ieder der reeksen waarnemingen eerst vervangen door hun rangnummers naar opklimmende grootte. Gelijke waarnemingen delen daarbij de hun toekomende rangnummers gelijk op (gemiddelde rangnummers). Dit geeft

$$\begin{array}{l} x: \quad 1 \quad 2 \quad 4 \quad 4 \quad 4 \quad 6 \quad 7 \quad 8\frac{1}{2} \quad 8\frac{1}{2} \quad 10 \\ y: \quad 5 \quad 3\frac{1}{2} \quad 9\frac{1}{2} \quad 6 \quad 7\frac{1}{2} \quad 3\frac{1}{2} \quad 7\frac{1}{2} \quad 9\frac{1}{2} \quad 2 \quad 1 \end{array}$$

2. Indien in slechts één van de twee waarnemingsreeksen gelijken voorkomen wordt S het eenvoudigst gevonden door de reeks zonder gelijken in de natuurlijke volgorde te zetten. De bijdragen tot S worden dan gevonden uit de andere reeks waarbij men niet meer, zoals in het tweede voorbeeld, op de eerste reeks behoeft te letten.

Voorbeeld 3.

Laat x een grootheid zijn, die slechts 3 waarden aanneemt (bijv. de waarden 1, 2 en 3) en y een grootheid, die slechts 4 waarden aanneemt (bijv. de waarden 1, 2, 3 en 4). Laat verder, bij 30 onderling onafhankelijke waarnemingsparen de volgende resultaten verkregen zijn:

$y \backslash x$	1	2	3	totaal
1	6	2	0	8
2	1	4	2	7
3	1	3	2	6
4	1	1	7	9
totaal	9	10	11	30

In deze tabel staan aantallen paren waarnemingen vermeld. Zo zijn er bijv. 6 paren (x_i, y_i) met $x_i = y_i = 1$; 2 paren (x_i, y_i) met $x_i = 3, y_i = 2$, enz. De randtotalen van de tabel geven de grootten van de groepen gelijken: in de x-reeks komen drie groepen gelijken voor met $t_1 = 9, t_2 = 10, t_3 = 11$; in de y-reeks vier groepen gelijken met $u_1 = 8, u_2 = 7, u_3 = 6, u_4 = 9$.

In dergelijke gevallen, waarbij grote aantallen gelijken voorkomen, berekent men S het eenvoudigst als volgt:

De bijdrage tot S van een vakje uit de bovenstaande tabel vindt men door de rij en de kolom, waartoe dit vakje behoort, te schrappen; de som van de aantallen waarnemingen in de vakjes rechts onder wordt nu verminderd met de som van de aantallen in de vakjes rechtsboven en dit verschil wordt vermenigvuldigd met het aantal waarnemingen in het betreffende vakje. De 3 paren (x_i, y_i) met $x_i = 2, y_i = 3$ bijv. geven dus een bijdrage

$$3\{ (7) - (2+0) \} = 15.$$

Zo krijgt men

$\begin{matrix} x \\ y \end{matrix}$	1	2	3
1	$6(19-0) = 114$	$2(11-0) = 22$	0
2	$1(13-2) = 11$	$4(9-0) = 36$	0
3	$1(8-8) = 0$	$3(7-2) = 15$	0
4	$1(0-13) = -13$	$1(0-4) = -4$	0

De grootte S wordt nu de som van de bijdragen van alle vakjes, dus

$$S = 114 + 11 - 13 + 22 + 36 + 15 - 4 = 181.$$

De rangcorrelatiecoëfficiënt τ wordt berekend volgens (3;6).

Hier is

$$C_1 = \sum_{v=1}^k t_v^2 = 9^2 + 10^2 + 11^2 = 302$$

$$C_2 = \sum_{x=1}^h u_x^2 = 8^2 + 7^2 + 6^2 + 9^2 = 230,$$

dus

$$\tau = \frac{2 \cdot 181}{\sqrt{(900 - 302)(900 - 230)}} = 0,57.$$

Voorbeeld 4.

Stel men wil S berekenen voor twee waarnemingsreeksen waarvan de ene zich leent voor een rangschikking met weinig of geen gelijken, terwijl de andere een indeling in twee klassen (een dichotomie) is. Bijv.:³⁾ 15 kandidaten, waaronder 7 meisjes (M) en 8 jongens (J) hebben een examen gedaan en zijn gerangschikt naar opklimmende resultaten:

rangnummer: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 geslacht : J J M J M M J J J M J M J M M

Rangschikken we de jongens vóór de meisjes dan krijgen de jongens het rangnummer $\frac{1}{2}(1+8) = 4\frac{1}{2}$ en de meisjes het rangnummer $\frac{1}{2}(9+15) = 12$. Dit geeft

x: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
 y: $4\frac{1}{2}$ $4\frac{1}{2}$ 12 $4\frac{1}{2}$ 12 12 $4\frac{1}{2}$ $4\frac{1}{2}$ $4\frac{1}{2}$ 12 $4\frac{1}{2}$ 12 $4\frac{1}{2}$ 12 12

en S wordt gevonden uit

																	totaal
+	7	7	0	6	0	0	4	4	4	0	3	0	2	0	0		37(=P)
-	0	0	6	0	5	5	0	0	0	2	0	1	0	0	0		19(=Q)
+	7	6	6	5	5	4	4	3	2	3	1	2	0	1	0		49(=R)
totaal	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0		105

Dus

$$S = P - Q = 37 - 19 = 18.$$

(Bij rangschikking van de meisjes vóór de jongens vindt men $S = -18$).

De rangcorrelatiecoëfficiënt τ wordt gevonden volgens (3;6). In de x-reeks komen geen gelijken voor, dus $C_1 = n=15$. De y-reeks bevat twee groepen gelijken ter grootte $u_1=8$ en $u_2=7$. Dus $C_2=8^2+7^2 = 113$. Hieruit volgt

$$\tau = \frac{2 \cdot 18}{\sqrt{(225 - 15)(225 - 113)}} = 0,23.$$

In gevallen zoals in dit voorbeeld beschreven, dus gevallen waarbij de x-reeks bestaat uit waarnemingen $x_1, \dots, x_n$ en de

3) Zie M.G. KENDALL (1955, p.41-42).

y-reeks uit een indeling in twee klassen, is KENDALL's rang-correlatietoets identiek met de toets voor twee steekproeven van WILCOXON (zie bijv. D. WABEKE en C. van EEDEN (1955)). De twee steekproeven bestaan hierbij uit de in de twee klassen gevonden waarnemingen x_i en de steekproefgrootten zijn dus de grootten van de twee groepen gelijke waarnemingen in de y-reeks (t_1 en t_2). Verder is, als W de toetsingsgrootte van WILCOXON's twee-steekproeven-toets is,

$$S = W - t_1 t_2.$$

In ons voorbeeld zijn de twee steekproeven dus

$$\begin{array}{ll} \text{M:} & 3, 5, 6, 10, 12, 14, 15 & t_1 = 7 \\ \text{J:} & 1, 2, 4, 7, 8, 9, 11, 13 & t_2 = 8. \end{array}$$

Hieruit vinden we: $W=74$, dus $S = 74 - 7 \times 8 = 18$.

In gevallen zoals deze kan men de toets beter in de vorm van WILCOXON's twee-steekproeven-toets toepassen. De daarvoor ontwikkelde berekeningsmethoden en formules zijn op dit speciale geval afgestemd en verlopen daarom het snelst.

Voorbeeld 5.

Een ander bijzonder geval treedt op als beide reeksen dichotomieën zijn. Men kan de waarnemingen dan als volgt rangschikken (zie ook voorbeeld 3):

	A	$\bar{A}$	totaal
B	a	c	u_1
$\bar{B}$	b	d	u_2
totaal	t_1	t_2	n

Er zijn hier dus a paren waarnemingen, waarbij de x -waarneming in klasse A en de y -waarneming in klasse B valt; verder zijn er b paren, waarbij de x -waarneming in klasse A en de y -waarneming in klasse $\bar{B}$ (= niet B) valt, enz.

Berekenen we S volgens voorbeeld 3 dan vinden we

$$S = ad - bc.$$

Verder is

$$n^2 - C_1 = (t_1 + t_2)^2 - (t_1^2 + t_2^2) = 2 t_1 t_2$$

$$n^2 - C_2 = (u_1 + u_2)^2 - (u_1^2 + u_2^2) = 2 u_1 u_2,$$

dus

$$\tau = \frac{ad - bc}{\sqrt{t_1 t_2 u_1 u_2}}.$$

In dit geval is KENDALL's rangcorrelatietoets identiek met de methode der 2x2-tabel (zie bijv. R.A. FISHER (1948, p.96) of C. van EEDEN (1953)) en

$$S = n \left(a - \frac{t_1 u_1}{n} \right),$$

waarin a de toetsingsgrootte van de methode der 2x2-tabel is.

4. Een grafische methode voor de berekening van S.

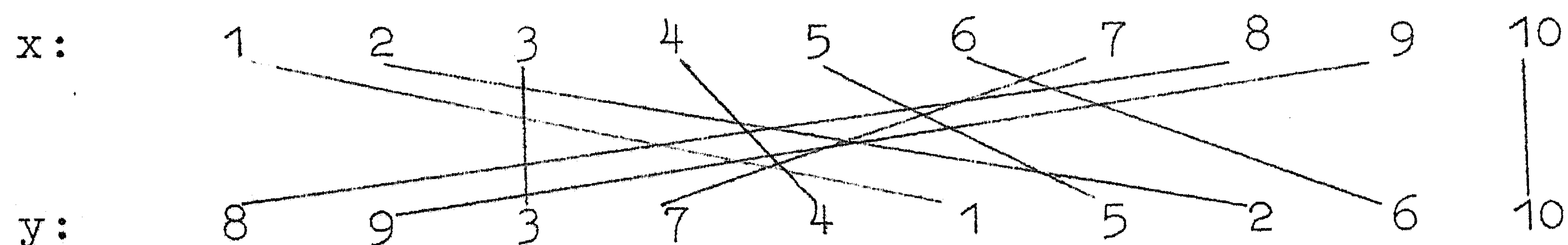
P. SANDIFORD beschreef in 1928 een van S.D. HOLMES afkomstige grafische methode voor het schatten van de produkt-moment-correlatiecoëfficiënt in kleine steekproeven. H.D. GRIFFIN (1958) toonde aan, dat de door SANDIFORD en HOLMES gebruikte schatting identiek is met KENDALL's τ voor het geval in geen van beide reeksen gelijke waarnemingen voorkomen. Tevens gaf GRIFFIN een uitbreiding van de grafische methode voor het geval er wel gelijken zijn.

We geven nu eerst een beschrijving van de methode voor het geval dat in geen van beide reeksen gelijken voorkomen:

In beide reeksen worden de waarnemingen vervangen door hun rangnummers naar opklimmende grootte en de kolommen worden in een zodanige volgorde gezet dat één van de twee reeksen in de natuurlijke volgorde staat. Laat nu s voorstellen het aantal snijpunten verkregen door ieder rangnummer uit de eerste reeks te verbinden met het daaraan gelijke rangnummer in de tweede reeks. Dan is

$$(4;1) \quad S = \frac{1}{2} n(n-1) - 2s.$$

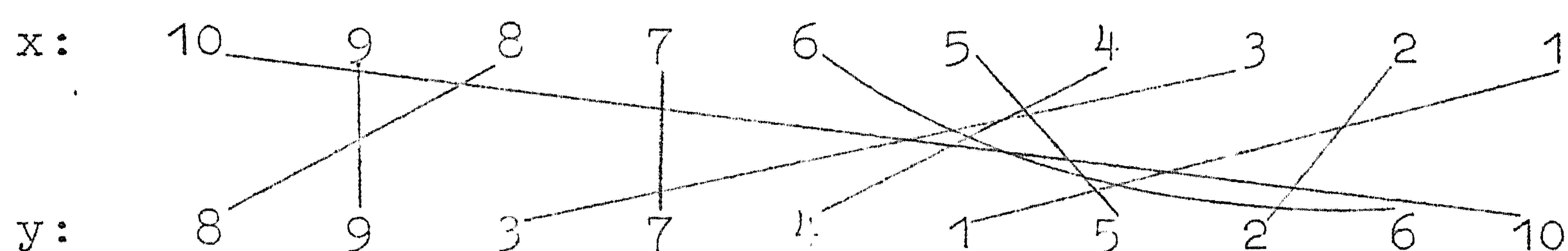
Bij voorbeeld 1 uit paragraaf 3 krijgen we



Het verdient aanbeveling de lijnen zodanig te trekken dat een lijn nooit door het snijpunt van twee andere gaat.

Het aantal snijpunten s , dat men het eenvoudigst vindt door turven tijdens het trekken der lijnen, is in ons voorbeeld 24, dus $S = \frac{1}{2} \cdot 10 \cdot 9 - 2 \cdot 24 = 45 - 48 = -3$.

Als we de volgorde van de eerste reeks omdraaien en die van de andere ongewijzigd laten, krijgen we



en het aantal snijpunten $s' = 21$. Dit geeft

$$S' = -S = \frac{1}{2} \cdot 10 \cdot 9 - 2 \cdot 21 = 45 - 42 = 3.$$

In gevallen, waarin het duidelijk is dat een negatieve correlatie optreedt, geeft dit omdraaien van de volgorde van één van de reeksen een aanzienlijke werkbesparing.

In het geval dat één van de twee reeksen gelijke waarnemingen bevat, gaat men als volgt te werk.

De kolommen worden in zodanige volgorde geplaatst, dat de reeks zonder gelijken in de natuurlijke volgorde komt te staan. De waarnemingen in de reeks zonder gelijken worden vervangen door hun rangnummers naar opklimmende grootte en men berekent twee waarden s_1 en s_2 van s als volgt: in de tweede reeks worden de waarnemingen in opklimmende grootte vervangen door de rangnummers $1, \dots, n$ en wel zodanig dat

- 1) binnen iedere groep gelijken de tot die groep behorende rangnummers in de natuurlijke volgorde staan. Dit geeft de waarde s_1 voor s ,
- 2) binnen iedere groep gelijken de tot die groep behorende rangnummers tegengesteld aan de natuurlijke volgorde staan. Dit geeft de waarde s_2 voor s .

Dan is

$$2S = s_1 + s_2.$$

Voorbeeld 6.

Laat de volgende twee reeksen waarnemingen gegeven zijn:

x: 1 2 3 4 5 6 7
y: 11 9 9 13 17 21 13.

Rangschikking van de tweede reeks volgens 1) geeft

x: 1 2 3 4 5 6 7
y: 3 1 2 4 6 7 5

met $s_1 = 4$.

Rangschikking volgens 2) geeft

x: 1 2 3 4 5 6 7
y: 3 2 1 5 6 7 4

met $s_2 = 6$. Dus

$$2S = s_1 + s_2 = 4 + 6 = 10$$

en

$$S = \frac{1}{2} n(n-1) - 2S = \frac{1}{2} \cdot 7 \cdot 6 - 10 = 11.$$

Komen in beide reeksen gelijken voor dan berekent men vier waarden s_1 , s_2 , s_3 en s_4 van s . Hiertoe plaatst men eerst de kolommen in zodanige volgorde dat $x_1 \leq \dots \leq x_n$; gelijke waarnemingen worden hierbij in een willekeurige volgorde gezet. In beide reeksen worden nu de waarnemingen in opklimmende grootte vervangen door de rangnummers $1, \dots, n$ en wel zodanig dat

- 1) in beide reeksen binnen iedere groep gelijken de tot die groep behorende rangnummers in de natuurlijke volgorde staan. Dit geeft de waarde s_1 voor s ,
- 2) in beide reeksen binnen iedere groep gelijken de tot die groep behorende rangnummers tegengesteld aan de natuurlijke volgorde staan. Dit geeft s_2 (nadat men de kolommen weer zodanig verwisseld heeft dat één van de twee reeksen in de natuurlijke volgorde staat).

Tenslotte vindt men s_3 door voor de eerste reeks de rangnummers te nemen, die voor de berekening van s_1 werden gebruikt en voor de tweede reeks de rangnummers, die voor de berekening van s_2 werden gebruikt; s_4 wordt gevonden door voor de eerste reeks de rangnummers te nemen die voor de berekening van s_2 werden gebruikt en voor de tweede reeks die welke voor de berekening van s_1 werden gebruikt. In dit laatste geval moet men de kolommen weer verwisselen zodanig dat één van de reeksen in de natuurlijke volgorde komt te staan.

Nu is

$$4S = s_1 + s_2 + s_3 + s_4.$$

Voorbeeld 7

Laat de volgende twee reeksen rangnummers gegeven zijn

$$\begin{array}{ccccccc} x: & 2 & 2 & 2 & 4\frac{1}{2} & 4\frac{1}{2} & 6\frac{1}{2} & 6\frac{1}{2} \\ y: & 4\frac{1}{2} & 2\frac{1}{2} & 1 & 4\frac{1}{2} & 6 & 2\frac{1}{2} & 7 \end{array}$$

Voor de berekening van s_1 gebruiken we de rangnummers

$$\begin{array}{ccccccc} x: & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ y: & 4 & 2 & 1 & 5 & 6 & 3 & 7 \end{array}$$

dus $s_1 = 6$.

Voor s_2 krijgen we

$$\begin{array}{ccccccc} x: & 3 & 2 & 1 & 5 & 4 & 7 & 6 \\ y: & 5 & 3 & 1 & 4 & 6 & 2 & 7 \end{array}$$

Herrangschikking van de paren zodat de x-reeks in de natuurlijke volgorde komt te staan, geeft

$$\begin{array}{ccccccc} x: & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ y: & 1 & 3 & 5 & 6 & 4 & 7 & 2 \end{array}$$

dus $s_2 = 7$.

Voor de berekening van s_3 krijgen we

$$\begin{array}{ccccccc} x: & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ y: & 5 & 3 & 1 & 4 & 6 & 2 & 7 \end{array}$$

dus $s_3 = 8$.

Bij s_4 gebruiken we

x:	3	2	1	5	4	7	6
y:	4	2	1	5	6	3	7

Herrangschikking geeft:

x:	1	2	3	4	5	6	7
y:	1	2	4	6	5	7	3

dus $s_4 = 5$ en

$$4s = 6+7+8+5 = 26$$

$$S = \frac{1}{2}n(n-1) - 2s = \frac{1}{2} \cdot 7 \cdot 6 - 13 = 8.$$

We zullen nu laten zien, dat de op deze wijze berekende grootheid S identiek is met de in paragraaf 3 gedefiniëerde S .⁴⁾ We beschouwen eerst het geval dat in geen van beide reeksen gelijken voorkomen. Dan is

$$P + Q = \frac{1}{2} n(n-1)$$

en dus

$$S = P - Q = \frac{1}{2} n(n-1) - 2Q,$$

zodat we in dit geval moeten bewijzen dat $Q=s$ is.

Laat de kolommen zodanig gerangschikt zijn dat $x_1 < \dots < x_n$ is.

Laat verder in beide reeksen de waarnemingen vervangen zijn door hun rangnummers naar opklimmende grootte en laat, voor $i=1, \dots, n$, r_i het rangnummer van y_i voorstellen.

Laat nu, voor $i=1, \dots, n$, L_i de verbindingslijn van de rangnummers r_i in de twee reeksen voorstellen, dan is het aantal snijpunten s gelijk aan het aantal paren (i, j) met $i < j$ waarvoor L_i en L_j elkaar snijden. Daar de x -reeks in de natuurlijke volgorde staat snijden L_i en L_j voor $i < j$ elkaar dan en slechts dan als in de tweede reeks $r_i > r_j$ is. Dus is s gelijk aan het aantal paren (i, j) met $i < j$, waarvoor in de tweede reeks $r_i > r_j$ is en dit is (zie voorbeeld 1) de grootheid Q .

⁴⁾ Dit bewijs wijkt af van het door GRIFFIN gegeven, niet geheel duidelijke, bewijs.

We beschouwen nu het geval dat in één van beide reeksen gelijken voorkomen. Stel

$$S_{\lambda} = \frac{1}{2} n(n-1) - 2s_{\lambda} \quad \lambda = 1, 2,$$

waarin s_1 resp. s_2 de aantallen snijpunten voorstellen bij toekenning van de rangnummers volgens 1) resp. 2). Dan is (volgens het bovengegeven bewijs) S_1 resp. S_2 KENDALL's toetsingsgrootheid bij toekenning van de rangnummers volgens 1) resp. 2). In dit geval is het dus voldoende te bewijzen dat

$$S = \frac{S_1 + S_2}{2}.$$

Evenzo is het, voor het geval dat in beide reeksen gelijken voorkomen, voldoende te bewijzen dat S het gemiddelde is van de vier waarden $S_1, \dots, S_4$, die gevonden worden bij toekenning van de rangnummers op de vier bovenbeschreven manieren.

Dat S gelijk is aan het gemiddelde der S_{λ} kan men als volgt inzien (zie ook KENDALL (1955), p.37):

Een paar $(x_i, x_j), (y_i, y_j)$ met $i < j$ dat tot S een bijdrage $\neq 0$ geeft, geeft tot ieder der S_{λ} eenzelfde bijdrage en wel dezelfde bijdrage als tot S .

Voor een paar $(x_i, x_j), (y_i, y_j)$ met $i < j$, dat een bijdrage 0 tot S geeft, geldt

$$x_i = x_j \quad \text{en(of)} \quad y_i = y_j.$$

De bijdragen van dit paar tot de S_{λ} zijn +1 of -1 en wel is het aantal gevallen, waarin +1 optreedt gelijk aan het aantal gevallen, waarin -1 optreedt. Het gemiddelde van de bijdragen die dit paar tot ieder der S_{λ} geeft is dus 0 en dus gelijk aan de bijdrage tot S .

5. De verdeling van $\underline{S}$ onder de getoetste hypothese H_0 :

5.1. De verdeling van $\underline{S}$ onder H_0 als in geen van beide reeksen gelijken voorkomen.

Voor kleine waarden van n kan de verdeling van $\underline{S}$ onder de getoetste hypothese H_0 eenvoudig als volgt berekend worden:

Als $n=2$ is zijn er, bij gegeven volgorde van de waarnemingen in de x -reeks, twee mogelijke volgorden voor de y -reeks:

$x: \quad 1 \quad 2$ $x: \quad 1 \quad 2$
 $y: \quad 1 \quad 2$ $y: \quad 2 \quad 1$

en

Deze twee mogelijkheden zijn onder H_0 even waarschijnlijk (zie paragraaf 2) en de hierbij optredende waarden van S zijn resp. $S=1$ en $S=-1$, dus

$$P[S = 1 | n=2, H_0] = P[S = -1 | n=2, H_0] = \frac{1}{2}.$$

Voor $n=3$ zijn er 6 mogelijkheden

$x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=3$ $x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=1$ $x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=1$
 $y: \quad 1 \quad 2 \quad 3$ $y: \quad 1 \quad 3 \quad 2$ $y: \quad 2 \quad 1 \quad 3$

 $x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=-1$ $x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=-1$ $x: \quad 1 \quad 2 \quad 3 \quad \left. \vphantom{\begin{matrix} x \\ y \end{matrix}} \right\} S=-3$
 $y: \quad 2 \quad 3 \quad 1$ $y: \quad 3 \quad 1 \quad 2$ $y: \quad 3 \quad 2 \quad 1$

Deze 6 mogelijkheden zijn onder H_0 even waarschijnlijk, dus

$$P[S = 3 | n=3, H_0] = P[S = -3 | n=3, H_0] = \frac{1}{6}$$

$$P[S = 1 | n=3, H_0] = P[S = -1 | n=3, H_0] = \frac{1}{3}.$$

Volgens dit principe kan ook voor grotere waarden van n de verdeling van S berekend worden. Eenvoudiger is het echter gebruik te maken van een recursieformule waarbij de verdeling van S bij n paren (x_i, y_i) uitgedrukt wordt in die bij $n-1$ paren (zie KENDALL (1955, hoofdstuk 5)).

Een voorbeeld van de verdeling van S vindt men in de onderstaande figuur, waarbij $n=6$ is. Op de X-as zijn de waarden

$$P[S = S | n=6, H_0]$$

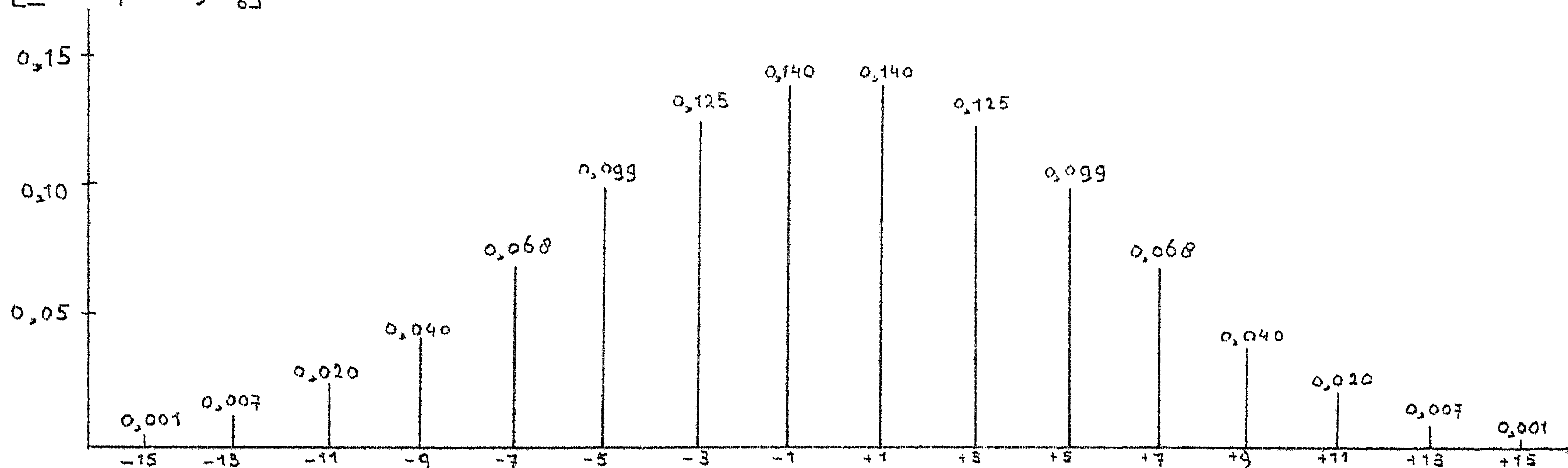


Fig. 5.1: Verdeling van S onder de hypothese H_0 voor $n=6$.

die $\underline{S}$ aan kan nemen uitgezet en op de Y-as de kans dat $\underline{S}$ de betreffende waarde aanneemt als H_0 juist is. Bijv.

$$P[\underline{S} = 9 | n=6; H_0] = 0,040.$$

Verder is bijv.

$$P[\underline{S} \geq 11 | n=6; H_0] = P[\underline{S} = 11 | n=6; H_0] + P[\underline{S} = 13 | n=6; H_0] + \\ + P[\underline{S} = 15 | n=6; H_0] = 0,020 + 0,007 + 0,001 = 0,028,$$

dus de kans dat $\underline{S}$ de waarde 11 of een grotere waarde aanneemt is, als H_0 juist is, gelijk aan 0,028.

De verdeling van $\underline{S}$ onder de getoetste hypothese H_0 heeft de volgende eigenschappen:

1. $\underline{S}$ neemt slechts even (resp. oneven) waarden aan als $\frac{1}{2}n(n-1)$ even (resp. oneven) is. De afstand tussen de opeenvolgende waarden die $\underline{S}$ aan kan nemen is gelijk aan 2.
2. de verdeling van $\underline{S}$ is symmetrisch ten opzichte van nul; voor de verwachting (het theoretisch gemiddelde) μ van de verdeling geldt dus

$$(5.1;1) \quad \mu = 0.$$

3. de variantie van de verdeling wordt gegeven door

$$(5.1;2) \quad \sigma^2 = \frac{1}{18} n(n-1)(2n+5).$$

4. $\frac{\underline{S}}{\sigma}$ is, voor grote waarden van n , bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

De bewijzen van deze eigenschappen vindt men in KENDALL (1955).

5.2. De verdeling van $\underline{S}$ onder H_0 als in één of beide reeksen gelijken voorkomen.

Komen in één of beide reeksen gelijken voor dan kan men de verdeling van $\underline{S}$ voor kleine waarden van n weer berekenen door voor alle mogelijke volgorden van de y-reeks, bij gegeven volgorde van de x-reeks, de waarde van $\underline{S}$ te berekenen. De verdeling van $\underline{S}$ volgt dan uit het feit dat deze volgorden alle even waarschijnlijk zijn. Voor grote waarden van n kan men de verdeling echter sneller vinden door gebruik te maken van de door L.J. SMID (1956) voor dit geval afgeleide recursie-formule.

Een voorbeeld van de verdeling van S voor het geval in beide reeksen

gelijken voorkomen vindt men in onderstaande figuur. Hierin is $n=4$; de x-reeks bestaat uit de rangnummers $1, 2, 3\frac{1}{2}, 3\frac{1}{2}$; de y-reeks uit de rangnummers $1\frac{1}{2}, 1\frac{1}{2}, 3, 4$.

$$P[\underline{S} = S | n=4; H_0]$$

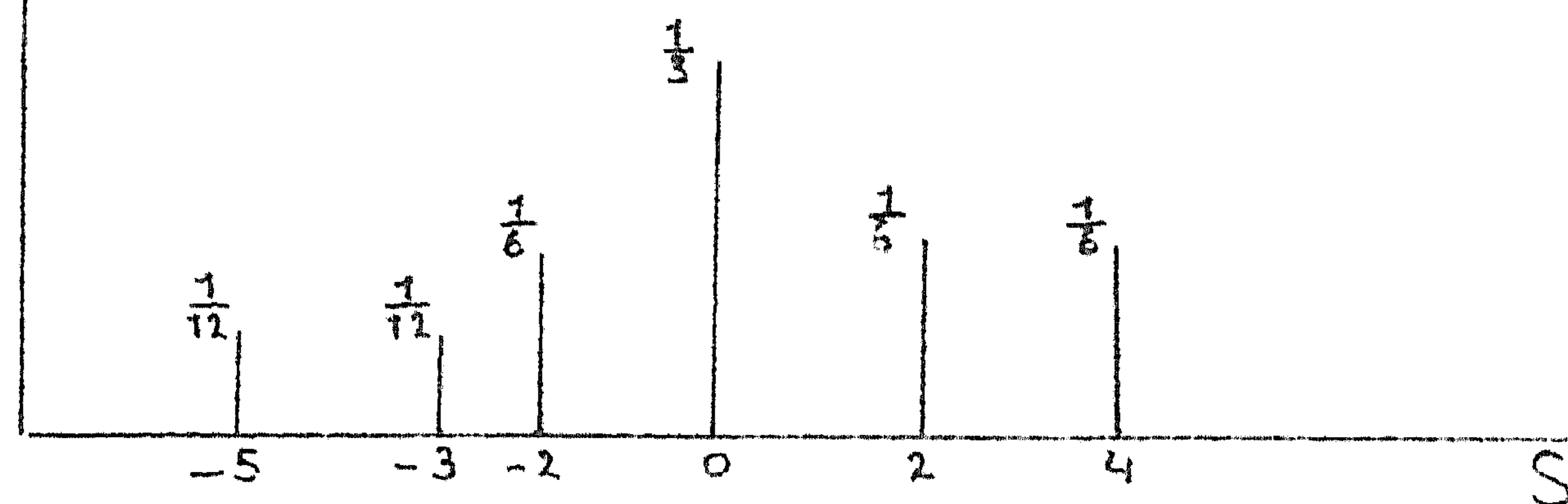


Fig. 5;2: Verdeling van $\underline{S}$ onder H_0 voor $n=4$ met gelijken in beide reeksen.

Uit deze figuur zien we (vgl. paragraaf 5;1)

1. dat de afstand tussen de opeenvolgende waarden die $\underline{S}$ aanneemt in het algemeen niet constant is
2. dat de verdeling van $\underline{S}$ in het algemeen niet symmetrisch is.⁵⁾

Voor de verwachting μ van de verdeling blijft echter gelden

$$\mu = 0$$

Laat verder (zie paragraaf 3)

$$\begin{aligned} C_1 &= \sum_{v=1}^k t_v^2 & , & & C_2 &= \sum_{x=1}^h u_x^2 \\ D_1 &= \sum_{v=1}^k t_v^3 & , & & D_2 &= \sum_{x=1}^h u_x^3 \end{aligned}$$

5) Het al of niet symmetrisch zijn van de verdeling van $\underline{S}$ hangt af van de grootten van de groepen gelijken in de twee waarnemingsreeksen. Of de verdeling symmetrisch is kan men (zonder de verdeling te berekenen) als volgt nagaan: Men plaatst de x-waarnemingen in zodanige volgorde dat $x_1 \leq \dots \leq x_n$ en kent in dezelfde volgorde de symbolen $t_1, \dots, t_k$ aan de groepen gelijken toe. Op analoge wijze kent men de symbolen $u_1, \dots, u_h$ toe aan de groepen gelijken in de y-reeks. Dan is de verdeling van $\underline{S}$ symmetrisch als

- a) de rij $t_1, \dots, t_k$ identiek is met de rij $t_k, \dots, t_1$
 - of b) de rij $u_1, \dots, u_h$ identiek is met de rij $u_h, \dots, u_1$.
- Hieraan is o.a. voldaan als hoogstens één van de twee waarnemingsreeksen gelijken bevat.

dan wordt de variantie van $\underline{S}$ gegeven door

$$(5.2;1) \quad \sigma^2 = \frac{2 \{n^3 - D_1 - 3(n^2 - C_1)\} \{n^3 - D_2 - 3(n^2 - C_2)\} + 9(n-2)(n^2 - C_1)(n^2 - C_2)}{18 n(n-1)(n-2)}, \quad 6)$$

Treden in geen van beide reeksen gelijken op, dan is

$$C_1 = C_2 = D_1 = D_2 = n$$

en formule (5.2;1) gaat over in (5.1;2).

Komen alleen in de y-reeks gelijke waarnemingen voor dan is $C_1 = D_1 = n$ en (5.2;1) gaat over in

$$(5.2;2) \quad \sigma^2 = \frac{1}{9} (n^3 - D_2) + \frac{1}{6} (n^2 - C_2).$$

Tenslotte is ook in dit geval de verdeling van $\frac{S}{\sigma}$ voor grote waarden van n bij benadering normaal verdeeld, mits zowel de $t_1, \dots, t_k$ als de $u_1, \dots, u_h$ onderling niet te veel in grootte verschillen.

6. Kritieke zones en overschrijdingskansen.

Uit het voorafgaande zal duidelijk zijn, dat de grootheid $\underline{S}$ in het algemeen grote positieve waarden aanneemt als (zie paragraaf 2) $\underline{x}$ en $\underline{y}$ positief gecorreleerd zijn resp. als $\underline{y}_1, \dots, \underline{y}_n$ een stijgend verloop vertoont en grote negatieve waarden als $\underline{x}$ en $\underline{y}$ negatief gecorreleerd zijn resp. als $\underline{y}_1, \dots, \underline{y}_n$ een dalend verloop vertoont.

Wenst men nu H_0 te toetsen tegen de alternatieve hypothese

$$H_1: \begin{cases} \underline{x} \text{ en } \underline{y} \text{ zijn positief gecorreleerd} \\ \text{resp.} \\ \underline{y}_1, \dots, \underline{y}_n \text{ vertoont een stijgend verloop,} \end{cases}$$

dan kiest men een rechtseenzijdige kritieke zone Z_r bestaande uit grote positieve waarden van S .

Een linkseenzijdige kritieke zone Z_l , bestaande uit grote negatieve waarden van S , gebruikt men als men H_0 wil toetsen tegen de -----

6) De hier gegeven formule voor de variantie is een vereenvoudiging van de door KENDALL (1955, p. 55) gegeven formule en werd afgeleid door A. BENARD.

alternatieve hypothese

$$H_2: \begin{cases} \underline{x} \text{ en } \underline{y} \text{ zijn negatief gecorreleerd} \\ \text{resp.} \\ \underline{y}_1, \dots, \underline{y}_n \text{ vertoont een dalend verloop.} \end{cases}$$

Een tweezijdige kritieke zone Z , bestaande uit een links- en een rechtseenzijdige, gebruikt men als men H_0 wil toetsen tegen de alternatieve hypothese

$$H: \begin{cases} \underline{x} \text{ en } \underline{y} \text{ zijn gecorreleerd} \\ \text{resp.} \\ \underline{y}_1, \dots, \underline{y}_n \text{ vertoont een verloop} \end{cases}$$

Wordt een waarde van S gevonden, die in de gebruikte kritieke zone ligt, dan wordt H_0 verworpen en wel ten gunste van H_1 als $S > 0$ is en ten gunste van H_2 als $S < 0$ is.

De kans dat, als H_0 juist is, $\underline{S}$ een waarde in de kritieke zone aanneemt (dus de kans dat H_0 ten onrechte verworpen wordt) heet de onbetrouwbaarheid α_0 van de kritieke zone (of van de toets). Men kiest nu de kritieke zone zodanig, dat α_0 hoogstens gelijk is aan een van tevoren gekozen getal α , de onbetrouwbaarheidsdrempel. De tweezijdige kritieke zone kiest men zodanig dat de twee eenzijdige, waaruit hij bestaat, ieder een onbetrouwbaarheid $\frac{1}{2}\alpha$ hebben.

Het resultaat van een toets kan nog iets vollediger uitgedrukt worden door de overschrijdingskansen op te geven. Voor de rechts- (resp. links-) eenzijdige toets wordt de overschrijdingskans aangegeven met k_r (resp. k_l) en gedefiniëerd als de kans op het optreden (onder de hypothese H_0) van een minstens even grote (resp. hoogstens even grote) waarde van S als de gevondene. De tweezijdige overschrijdingskans wordt met k aangegeven en is gelijk aan tweemaal de kleinste eenzijdige overschrijdingskansen. Het vinden van een overschrijdingskans $\leq \alpha$ is dan identiek met het vinden van een waarde van S in de kritieke zone.

7. Tabellen voor kritieke waarden en overschrijdingskansen.

De tabellen voor overschrijdingskansen en kritieke waarden, die aan dit rapport zijn toegevoegd (tabellen 1 t/m 4, p. 30 t/m 42)

gelden voor het geval er in geen van beide reeksen gelijke waarnemingen voorkomen. De verdeling van $\underline{S}$ is dan symmetrisch ten opzichte van 0 (zie paragraaf 5.1), zodat slechts de helft behoeft te worden opgenomen. Hiervoor is de rechterhelft gekozen.

De tabellen 1 en 2 bevatten voor $n = 4(1)40$ de exacte rechtseenzijdige overschrijdingskansen van S .⁷⁾ Daar $\underline{S}$ slechts even (resp. oneven) waarden aanneemt als $\frac{1}{2}n(n-1)$ even (resp. oneven) is, bevat tabel 1 de waarden van n waarvoor $\frac{1}{2}n(n-1)$ even is (dus $n \equiv 4,5 \pmod{4}$) en tabel 2 de waarden van n waarvoor $\frac{1}{2}n(n-1)$ oneven is (dus $n \equiv 6,7 \pmod{4}$). Voor een negatieve waarde van S is, wegens de symmetrie van de verdeling, de linkseenzijdige overschrijdingskans gelijk aan de rechtseenzijdige overschrijdingskans van $|S|$. De tweezijdige overschrijdingskans is gelijk aan tweemaal de rechtseenzijdige overschrijdingskans van $|S|$.

De tabellen 3 en 4 bevatten de rechter kritieke waarden S_r van S ,⁸⁾ behorende bij de eenzijdige onbetrouwbaarheidsdrempels 0,005; 0,01; 0,025; 0,05 en 0,10. Dit betekent, dat de rechter kritieke zone Z_r met de gekozen onbetrouwbaarheidsdrempel bestaat uit alle waarden van $S \geq S_r$, als S_r het in de tabel gevonden getal is. De linker kritieke zone Z_l met dezelfde α , bestaat uit alle waarden van $S \leq -S_r$. De tweezijdige kritieke zone met onbetrouwbaarheidsdrempel 2α wordt gevonden door Z_r en Z_l samen te nemen. Toetst men bijv. tweezijdig met onbetrouwbaarheidsdrempel 0,05 dan bestaat de kritieke zone uit alle S met $|S| \geq S_r$, als S_r de rechter kritieke waarde behorende bij $\alpha = 0,025$ is. Toetst men rechtseenzijdig met $\alpha = 0,05$ dan wordt H_0 verworpen als $S \geq S_r$ is, waarbij S_r de rechter kritieke waarde behorende bij $\alpha = 0,05$ is. Analooq voor linkseenzijdig.

De tabellen 1 t/m 4 gelden strikt genomen alleen als in geen van de beide reeksen gelijken voorkomen, maar zij geven voor gevallen met weinig gelijken veelal een goede benadering. De exacte verdeling van $\underline{S}$ voor het geval in één of beide reeksen gelijken

7) Deze tabellen werden vervaardigd door de Rekenafdeling van het Mathematisch Centrum (zie L. KAARSEMAKER en A. van WIJNGAARDEN (1952)).

8) Tabel 3 bevat exacte kritieke waarden, gevonden met behulp van de tabellen 1 en 2. In tabel 4 staan benaderingen voor de kritieke waarden vermeld, gevonden met behulp van de normale benadering voor de verdeling van $\underline{S}$ (zie par. 5.1).

voorkomen is moeilijk te tabelleren, daar deze verdeling niet alleen afhangt van $t_1, \dots, t_k$ en $u_1, \dots, u_h$, maar ook van hun volgorde. Verkeert men echter in één der volgende situaties dan kan men, ondanks het voorkomen van gelijken, toch langs eenvoudige weg de exacte overschrijdingskansen vinden.

1. Indien in slechts één van de twee reeksen gelijken voorkomen dan kan men, als $n \leq 10$ is en als bovendien de grootte van alle groepen gelijken hoogstens 3 is, de exacte overschrijdingskansen vinden in de tabel van G.P.SILLITTO (1947).
2. Bestaat de ene reeks uit twee groepen gelijken, terwijl de andere bestaat uit waarnemingen $x_1, \dots, x_n$, dan is, zoals wij reeds in paragraaf 3 zagen, KENDALL's rangcorrelatietoets identiek met WILCOXON's twee steekproeventoets. Komen dus onder de $x_1, \dots, x_n$ geen gelijken voor dan kan men de exacte overschrijdingskansen vinden met behulp van tabellen van de exacte verdeling van W .
3. Bestaan beide reeksen uit twee groepen gelijken dan is (zie paragraaf 3) KENDALL's rangcorrelatietoets identiek met de methode der 2x2-tabel. In dit geval kan men de exacte overschrijdingskansen dus berekenen zoals bijv. aangegeven door FISHER (1948) en C. van EEDEN (1953). Men kan hier ook gebruik maken van tabellen van de exacte verdeling van de toetsingsgrootte van de 2x2-tabel (zie bijv. D. MAINLAND, L. HERRERA, M.I. SUTCLIFFE (1956)).

8. Benaderingen voor kritieke waarden en overschrijdingskansen.

Als n groot is en als bovendien $t_1, \dots, t_k$ onderling en $u_1, \dots, u_h$ onderling niet te veel in grootte verschillen, kan men gebruik maken van het feit, dat de grootte S , als H_0 juist is, bij benadering normaal verdeeld is met verwachting 0 en spreiding σ , die berekend wordt volgens één der formules (5.1;2), (5.2;1) of (5.2;2). De variantie σ^2 en de spreiding σ volgens formule (5.1;2) staan voor $n = 1(1)100$ vermeld in tabel 5. Een tabel van t^2 en t^3 voor $t = 1(1)40$ vindt men in tabel 6.

De rechtseenzijdige overschrijdingskans is nu bij benadering het oppervlak van de normale verdeling met verwachting 0 en spreiding

ding 1 rechts van

$$u_1 = \frac{S-1}{\sigma}$$

en kan dus in een tabel van de standaard normale verdeling worden opgezocht (zie tabel 7).

Daar hier een discrete verdeling benaderd wordt met een continue wordt een z.g. continuïteitscorrectie toegepast. Hiervoor neemt men meestal de helft van de afstand tussen de opeenvolgende waarden, die de toetsingsgrootte aan kan nemen. Deze afstand is, als er in geen van beide reeksen gelijken optreden, gelijk aan 2, zodat in dit geval de correctie gelijk is aan 1. Dit is de 1 in de teller van bovenstaande formule.

Treden wel gelijken op dan is de afstand tussen de opeenvolgende waarden van S in het algemeen niet meer constant. Een algemene regel voor de correctie is dus moeilijk te geven. Daar echter de overschrijdingskans groter wordt door het aanbrengen van een correctie doet men beter ook in gevallen met gelijken een correctie ter grootte 1 toe te passen.⁹⁾

Analoog is de linkseenzijdige overschrijdingskans bij benadering het oppervlak der normale verdeling links van $u_1 = \frac{S+1}{\sigma}$

9) Als de x-reeks bestaat uit twee groepen gelijken (dus als de toets identiek is met WILCOXON's twee steekproeventoets) dan is de afstand tussen de opeenvolgende waarden van S constant als (zie SMID (1956)) voor de groepen gelijken in de y-reeks geldt dat $u_x + u_{x+1}$ constant is. (Hierbij zijn de symbolen $u_1, \dots, u_h$ toegekend aan de groepen gelijken zoals aangegeven in voetnoot 4).

De afstand is dan gelijk aan $u_x + u_{x+1}$.

Aan deze voorwaarden is o.a. voldaan

- 1) als $h=2$, dus als het schema der waarnemingen overgaat in een 2x2-tabel. De afstand tussen de opeenvolgende waarden van S is dan $u_1 + u_2 = n$.
- 2) als $u_x = u$, dus als de u_x zelf constant zijn, d.w.z. als de groepen gelijken in de y-reeks alle dezelfde omvang hebben. De afstand is dan $2u$. Dit geval treedt dus o.a. op als in de y-reeks geen gelijken voorkomen; dan zijn alle $u_x=1$ en de afstand wordt gelijk aan 2.

en de tweezijdige overschrijdingskans bij benadering tweemaal het oppervlak rechts van

$$u = \frac{|S| - 1}{\sigma}.$$

Opmerkingen

3. Als er gelijke waarnemingen voorkomen is de variantie volgens de voor dit geval geldende formules (5.2;1) of (5.2;2) altijd kleiner dan die volgens (5.1;2). Als men dus, bij de berekening van de variantie, de gelijke waarnemingen niet in rekening brengt vindt men een te grote waarde voor deze variantie en dus een te grote waarde voor de overschrijdingskans. Is deze te grote waarde voor de overschrijdingskans $\leq \alpha$ dan wordt H_0 verworpen en heeft men de berekening van de kwadraten en derde machten in (5.2;1) of (5.2;2) niet uit te voeren.

4. De benaderende rechter kritieke waarden, die in tabel 4 vermeld staan, zijn gevonden met behulp van de bovengenoemde normale benadering. Deze tabel geldt voor het geval zonder gelijken, dus

$$\sigma^2 = \frac{1}{18} n(n-1)(2n+5).$$

De rechter kritieke waarde met onbetrouwbaarheidsdrempel α is dan bij benadering die waarde van S , waarvoor de benaderende rechteroverschrijdingskans $\leq \alpha$ is, dus de grootste waarde van S , die voldoet aan

$$\frac{S-1}{\sigma} \leq \xi_{\alpha}$$

waarbij ξ_{α} die waarde van de $N(0,1)$ -verdeelde grootheid is waarvoor de rechter overschrijdingskans $\leq \alpha$ is. Dus

$$\frac{1}{\sqrt{2\pi}} \int_{\xi_{\alpha}}^{\infty} e^{-\frac{1}{2}u^2} du = \alpha.$$

Daar S alleen even (resp. oneven) waarden aanneemt als $\frac{1}{2}n(n-1)$ even (resp. oneven) is, zijn de met bovengenoemde formule gevonden waarden van S naar boven afgerond op het dichtstbijzijnde even (resp. oneven) getal als $\frac{1}{2}n(n-1)$ even (resp. oneven) is.

9. Uitwerking van de voorbeelden.

Voorbeeld 1:

In dit voorbeeld is $n=10$ en $S=-3$, terwijl in geen van beide

reeksen gelijken voorkomen. De exacte overschrijdingskans wordt dus gevonden in tabel 2. Wij vinden dan (tweezijdig) $k=0,86$, zodat er op grond van deze gegevens geen reden is om aan te nemen dat er een correlatie bestaat tussen aanleg voor wiskunde en muziek.

Voorbeeld 2:

Bij dit voorbeeld is $n=10$ en $S=-6$. In beide reeksen treden gelijken op, zodat de exacte overschrijdingskans niet in een tabel kan worden opgezocht.

Een benadering voor de overschrijdingskans kan men vinden door tabel 2 te gebruiken. Nu komen bij $n=10$ zonder gelijken alleen oneven waarden van S voor. Tabel 2 bevat dus de rechtseenzijdige overschrijdingskansen van $S=5$ en $S=7$. Voor de rechtseenzijdige overschrijdingskans van $S=6$ kan men dan het gemiddelde nemen van deze twee, dus $\frac{0,364 + 0,300}{2}$. De tweezijdige overschrijdingskans is dan tweemaal deze eenzijdige, dus $0,66$.

Men kan ook de normale benadering toepassen. De variantie wordt dan berekend volgens formule (5.2;1). Wij zagen reeds (zie paragraaf 3) dat

$$\begin{aligned} t_1 = t_2 = t_4 = t_5 = t_7 = 1, \quad t_3 = 3, \quad t_6 = 2 & \quad C_1 = 18 \\ u_1 = u_2 = u_4 = u_5 = 1, \quad u_3 = u_6 = u_7 = 2 & \quad C_2 = 16. \end{aligned}$$

Verder is

$$\begin{aligned} D_1 &= \sum_{v=1}^k t_v^3 = 5 \cdot 1^3 + 3 \cdot 3^3 + 2 \cdot 2^3 = 40 \\ D_2 &= \sum_{x=1}^k u_x^3 = 4 \cdot 1^3 + 3 \cdot 2^3 = 28. \end{aligned}$$

Dus

$$\begin{aligned} n^2 - C_1 &= 82 & n^2 - C_2 &= 84 \\ n^3 - D_1 &= 960 & n^3 - D_2 &= 972. \end{aligned}$$

Ingevuld in formule (5.2;1) geeft dit

$$\sigma^2 = \frac{2(960 - 3 \cdot 82)(972 - 3 \cdot 84) + 9 \cdot 8 \cdot 82 \cdot 84}{18 \cdot 10 \cdot 9 \cdot 8} = 117,60 \quad \sigma = 10,84$$

dus

$$u = \frac{|S| - 1}{\sigma} = \frac{5}{10,84} = 0,46.$$

De tweezijdige overschrijdingskans wordt dan bij benadering (zie tabel 7) $k = 0,65$, wat zeer goed overeenstemt met de in tabel 2 gevonden overschrijdingskans.

Als we bij de berekening van de variantie de gelijken buiten be-

schouwing laten (zie opmerking 3) dan vinden we (zie tabel 5)
 $\sigma = 11,18$, dus

$$u = \frac{5}{11,18} = 0,45.$$

De invloed van de gelijke waarnemingen is hier dus zeer klein.

Tenslotte kan men ook nog een benadering vinden door de tabellen van SILLITTO te gebruiken. Daar deze tabellen slechts gelden voor gevallen met gelijken in slechts één reeks, moet men de gelijken in één van de twee reeksen verwaarlozen. Neemt men hiervoor de x-reeks dan vindt men $k=0,66$. Verwaarlozing van de gelijken in de y-reeks geeft eveneens $k=0,66$.

Voorbeeld 3.

Hier is $n=30$ en $S=181$, terwijl zeer veel gelijken optreden. We passen dus de normale benadering toe met variantie volgens (5.2;1). In paragraaf 3 vonden we reeds

$$\begin{array}{ll} t_1 = 9, t_2 = 10, t_3 = 11 & C_1 = 302 \\ u_1 = 8, u_2 = 7, u_3 = 6, u_4 = 9 & C_2 = 230 \end{array}$$

$$\begin{array}{l} \text{Verder is } D_1 = 9^3 + 10^3 + 11^3 = 3060 \\ D_2 = 8^3 + 7^3 + 6^3 + 9^3 = 1800, \end{array}$$

dus

$$\begin{array}{ll} n^2 - C_1 = 598 & n^2 - C_2 = 670 \\ n^3 - D_1 = 23940 & n^3 - D_2 = 25200. \end{array}$$

Dit geeft

$$\sigma^2 = \frac{2(23940-3.598)(25200-3.670)+9.28.598.670}{18.30.29.28} = 2572,75$$

$$\text{dus } \sigma = 50,72 \text{ en } u = \frac{|S|-1}{\sigma} = \frac{180}{50,72} = 3,55.$$

De tweezijdige overschrijdingskans, opgezocht in tabel 7, wordt dus $< 0,0004$.

Als we bij de berekening van de spreiding de gelijken verwaarlozen, dan vinden we (zie tabel 5) $\sigma = 56,05$. Dus

$$u = \frac{180}{56,05} = 3,21,$$

wat (zie tabel 7) een overschrijdingskans van 0,0014 geeft. Berekening van de spreiding volgens (5.2;1) was hier dus over-

bodig (zie opmerking 3).

Voorbeeld 4.

Bij dit voorbeeld was KENDALL's rangcorrelatietoets identiek met WILCOXON's twee steekproeventoets. We vonden als toetsingsgrootte van WILCOXON $W=74$ bij de steekproefgrootten 7 en 8. Daar onder de x-waarnemingen geen gelijken voorkomen kan de exacte overschrijdingskans worden gevonden in een tabel van de verdeling van $\underline{W}$ (zie bijv. D. WABEKE en C. van EEDEN (1955)); dit geeft (tweezijdig) $k=0,34$.

Passen we de bovengegeven normale benadering toe dan kunnen we, daar slechts in de y-reeks gelijken voorkomen, formule (5.2;2) voor de variantie gebruiken. Dus

$$\sigma^2 = \frac{1}{9} (n^3 - D_2) + \frac{1}{6} (n^2 - C_2),$$

met $u_1=7$ en $u_2=8$. Dus

$$n^2 - C_2 = 225 - (49+64) = 112$$

$$n^3 - D_2 = 3375 - (343+512) = 2520$$

en

$$\sigma^2 = \frac{1}{9} \cdot 2520 + \frac{1}{6} \cdot 112 = 298\frac{2}{3} \quad \sigma = 17,28.$$

Dit geeft

$$u = \frac{S-1}{\sigma} = 0,98.$$

De benaderende tweezijdige overschrijdingskans is dus (zie tabel 7) $k=0,33$, wat zeer goed overeenstemt met de bovengegeven exacte waarde.

Voorbeeld 6.

Bij dit voorbeeld komen slechts in één reeks gelijken voor, terwijl $n=7$ is. Verder is $S=11$. Daar ieder der groepen gelijken een omvang ≤ 2 heeft kan de exacte overschrijdingskans gevonden worden in de tabellen van SILLITTO (1947). Dit geeft (tweezijdig) $k=0,13$.

Een benadering voor k kan men, daar er slechts weinig gelijken voorkomen, vinden in tabel 2. Dit geeft $k=0,14$.

Voorbeeld 7.

Hier is $n=7$ en $S=8$, terwijl in beide reeksen gelijken voor-

komen. De exacte overschrijdingskans kan dus niet in een tabel gevonden worden.

Een benadering kan in de eerste plaats gevonden worden uit tabel 2. Dit geeft (tweezijdig) $k=0,31$.

Een andere benadering kan men vinden met behulp van de tabellen van SILLITTO. Daar deze slechts gelden voor gevallen met gelijken in één reeks, moet men de gelijken in één van de twee reeksen verwaarlozen. Neemt men hiervoor de y-reeks, dan geeft de tabel van SILLITTO $k=0,28$. Verwaarlozing van de gelijken in de x-reeks geeft $k=0,30$.

Literatuur.

- VAN EEDEN, C., (1953), Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Statistica 7, 141-162.
- FISHER, R.A., (1948), Statistical methods for research workers, Oliver and Boyd, London.
- GRIFFIN, H.D., (1958), Graphic computation of tau as a coefficient of disarray, Journ. Amer. Stat. Ass. 53, 441-447.
- KAARSEMAKER, L. and A. VAN WIJNGAARDEN, (1952), Tables for use in rankcorrelation, Rapport R 73 van de Rekenafdeling van het Mathematisch Centrum, Amsterdam.
- KENDALL, M.G., (1955), Rank Correlation Methods, Ch.Griffin, London.
- MAINLAND, D., L. HERRERA and M.I. SUTCLIFFE, (1956), Tables for use with binomial samples, Dept. of Medical Statistics, New York College of Medicine.
- SANDIFORD, P., (1928), A graphical method of estimating R for small groups, Appendix B of "Educational psychology", New York, Longmans, Green and Co.
- SILLITTO, G.P., (1947), The distribution of Kendall's coefficient of rank correlation in rankings containing ties, Biometrika 34, 36-40.
- SMID, L.J., (1956), On the distribution of the test statistics of Kendall and Wilcoxon's two sample test when ties are present, Statistica Neerlandica 10, 205-214.
- WABEKE, D. en C. VAN EEDEN, (1955), Handleiding voor de toets van Wilcoxon, Rapport S 176 (M 65) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

Tabel 1

Waarden van $10^3 \cdot P[S \geq s | n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$

$\begin{matrix} S \\ n \end{matrix}$	0	2	4	6	8	10	12	14	16	18
4	625	375	167	042	0	0	0	0	0	0
5	592	408	242	117	042	008	0	0	0	0
8	548	452	360	274	199	138	089	054	031	016
9	540	460	381	306	238	179	130	090	060	038
12	527	473	420	369	319	273	230	190	155	125
13	524	476	429	383	338	295	255	218	184	153
16	518	482	447	412	378	345	313	282	253	225
17	516	484	452	420	388	358	328	299	271	245
20	513	487	462	436	411	387	362	339	315	293
21	512	488	464	441	417	394	371	349	327	306
24	510	490	471	451	432	413	394	375	356	338
25	509	491	472	454	436	418	400	382	364	347
28	508	492	477	461	446	430	415	400	385	370
29	507	493	478	463	448	434	419	405	390	376
32	506	494	481	468	455	442	430	417	405	392
33	506	494	482	469	457	445	433	421	409	397
36	505	495	484	473	462	452	441	430	420	409
37	505	495	484	474	464	453	443	433	423	413
40	505	495	486	477	468	459	449	440	431	422

$\begin{matrix} S \\ n \end{matrix}$	20	22	24	26	28	30	32	34	36	38
8	007	002	001	000	000	0	0	0	0	0
9	022	012	006	003	001	000	000	000	000	0
12	098	076	058	043	031	022	016	010	007	004
13	126	102	082	064	050	038	029	021	015	011
16	199	175	153	133	114	097	083	070	058	048
17	220	196	174	154	135	118	102	088	076	064
20	271	250	230	211	193	176	159	144	130	117
21	285	265	246	228	210	193	177	162	147	134
24	320	303	286	270	254	238	223	209	195	181
25	330	314	297	282	266	251	237	222	209	196
28	355	341	326	312	298	285	272	259	246	234
29	362	348	334	321	308	295	282	270	257	246
32	380	368	356	344	332	320	309	298	287	276
33	385	373	362	350	339	328	317	306	295	285
36	399	388	378	368	358	347	338	328	318	308
37	403	393	383	373	363	353	344	334	325	315
40	413	404	395	386	377	369	360	351	343	334

Tabel 1 (vervolg)

Waarden van $10^3 \cdot P[S \geq s | n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$

$\begin{matrix} s \\ n \end{matrix}$	40	42	44	46	48	50	52	54	56	58
12	003	002	001	000	000	000	000	000	000	000
13	007	005	003	002	001	001	000	000	000	000
16	039	032	026	021	016	013	010	008	006	004
17	054	046	038	032	026	021	017	014	011	009
20	104	093	082	073	064	056	049	043	037	032
21	121	109	098	088	079	070	062	055	049	043
24	169	156	145	134	123	113	104	095	087	079
25	183	171	159	148	138	128	118	109	101	093
28	222	211	200	189	178	168	158	149	140	131
29	234	223	212	201	191	181	171	162	153	144
32	265	255	244	234	224	215	206	197	188	179
33	274	264	254	244	235	225	216	207	199	190
36	299	290	280	271	262	254	245	237	228	220
37	306	297	288	279	271	262	254	245	237	229
40	326	318	309	301	293	285	277	270	262	255

$\begin{matrix} s \\ n \end{matrix}$	60	62	64	66	68	70	72	74	76	78
16	003	002	002	001	001	001	000	000	000	000
17	007	005	004	003	002	002	001	001	001	000
20	027	023	020	017	014	012	010	008	007	006
21	037	032	028	024	021	018	015	013	011	009
24	072	066	059	054	048	044	039	035	031	028
25	085	078	071	065	059	054	049	044	040	036
28	123	115	108	101	094	087	081	075	070	065
29	136	128	120	112	105	099	092	086	080	075
32	171	163	155	147	140	133	126	119	113	107
33	182	174	166	158	151	144	137	130	124	117
36	212	204	197	189	182	175	168	161	155	148
37	222	214	206	199	192	185	178	171	165	158
40	247	240	233	226	219	212	205	199	192	186

Tabel 1 (vervolg)

Waarden van $10^3 \cdot P[S \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	80	82	84	86	88	90	92	94	96	98
20	005	004	003	002	002	002	001	001	001	001
21	008	007	005	005	004	003	002	002	002	001
24	025	022	019	017	015	013	011	010	009	007
25	032	029	026	023	021	018	016	014	013	011
28	060	055	051	047	043	039	036	033	030	027
29	070	065	060	056	052	048	044	041	037	034
32	101	095	090	085	080	075	070	066	062	058
33	111	106	100	095	090	085	080	075	071	067
36	142	136	130	124	119	114	108	103	099	094
37	152	146	140	134	129	123	118	113	108	103
40	180	174	168	162	156	151	146	140	135	130

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	100	102	104	106	108	110	112	114	116	118
21	001	001	001	001	000	000	000	000	000	000
24	006	006	005	004	003	003	003	002	002	001
25	010	009	008	007	006	005	004	004	003	003
28	025	023	021	019	017	015	014	012	011	010
29	031	029	026	024	022	020	018	017	015	014
32	054	051	048	044	041	039	036	033	031	029
33	063	059	055	052	049	046	043	040	037	035
36	089	085	081	077	073	069	066	062	059	056
37	098	094	090	085	081	077	074	070	067	063
40	125	121	116	111	107	103	099	095	091	087

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	120	122	124	126	128	130	132	134	136	138
24	001	001	001	001	001	000	000	000	000	000
25	002	002	002	001	001	001	001	001	001	000
28	009	008	007	006	006	005	004	004	003	003
29	012	011	010	009	008	007	007	006	005	005
32	027	025	023	021	019	018	016	015	014	013
33	032	030	028	026	024	023	021	019	018	017
36	053	050	047	044	042	039	037	035	033	031
37	060	057	054	051	048	046	043	041	039	037
40	083	080	076	073	070	067	064	061	058	055

Tabel 1 (vervolg)

Waarden van $10^3 \cdot P [S \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	140	142	144	146	148	150	152	154	156	158
28	003	002	002	002	002	001	001	001	001	001
29	004	004	003	003	003	002	002	002	002	001
32	012	011	010	009	008	007	007	006	006	005
33	015	014	013	012	011	010	009	008	008	007
36	029	027	025	024	022	021	020	018	017	016
37	034	032	031	029	027	025	024	022	021	020
40	053	050	048	046	043	041	039	037	035	034

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	160	162	164	166	168	170	172	174	176	178
28	001	001	000	000	000	000	000	000	000	000
29	001	001	001	001	001	001	001	000	000	000
32	005	004	004	003	003	003	002	002	002	002
33	006	006	005	005	004	004	004	003	003	003
36	015	014	013	012	011	010	010	009	008	008
37	018	017	016	015	014	013	012	011	011	010
40	032	030	029	027	026	024	023	022	020	019

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	180	182	184	186	188	190	192	194	196	198
32	002	001	001	001	001	001	001	001	001	001
33	002	002	002	002	002	001	001	001	001	001
36	007	006	006	006	005	005	004	004	004	003
37	009	009	008	007	007	006	006	005	005	005
40	018	017	016	015	014	014	013	012	011	011

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	200	202	204	206	208	210	212	214	216	218
33	001	001	001	001	001	000	000	000	000	000
36	003	003	003	002	002	002	002	002	001	001
37	004	004	004	003	003	003	003	002	002	002
40	010	009	009	008	008	007	007	006	006	005

Tabel 1 (vervolg)

Waarden van $10^3 \cdot P[S \geq s \mid n; H_0]$ voor $\frac{1}{2}n(n-1)$ even en $4 \leq n \leq 40$

$\begin{matrix} S \\ n \end{matrix}$	220	222	224	226	228	230	232	234	236	238
36	001	001	001	001	001	001	001	001	001	000
37	002	002	002	001	001	001	001	001	001	001
40	005	005	004	004	004	004	003	003	003	003

[illegible][illegible]

Tabel 2

Waarden van $10^3 \cdot P[S \geq s | n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$

$\begin{matrix} s \\ n \end{matrix}$	1	3	5	7	9	11	13	15	17	19
6	500	360	235	136	068	028	008	001	0	0
7	500	386	281	191	119	068	035	015	005	001
10	500	431	364	300	242	190	146	108	078	054
11	500	440	381	324	271	223	179	141	109	082
14	500	457	415	374	334	295	259	225	194	165
15	500	461	423	385	349	313	279	248	218	190
18	500	470	441	411	383	354	327	300	275	250
19	500	473	445	418	391	365	339	314	290	267
22	500	478	456	434	412	390	369	348	328	308
23	500	479	458	438	417	397	377	357	338	319
26	500	483	465	448	431	414	397	380	363	347
27	500	484	467	451	434	418	402	386	371	355
30	500	486	472	458	444	430	416	402	389	375
31	500	487	473	460	446	433	420	407	394	381
34	500	488	477	465	453	442	430	418	407	396
35	500	489	478	466	455	444	433	422	411	400
38	500	490	480	470	460	450	440	431	421	411
39	500	490	481	472	462	452	443	433	424	414

$\begin{matrix} s \\ n \end{matrix}$	21	23	25	27	29	31	33	35	37	39
10	036	023	014	008	005	002	001	000	000	000
11	060	043	030	020	013	008	005	003	002	001
14	140	117	096	079	063	050	040	031	024	018
15	164	141	120	101	084	070	057	046	037	029
18	227	205	184	165	147	130	115	100	088	076
19	245	223	203	184	166	149	133	119	105	093
22	289	270	252	234	217	201	186	171	157	144
23	301	283	265	248	232	216	201	187	173	160
26	331	316	300	285	270	256	242	229	216	203
27	340	325	310	296	281	268	254	241	228	216
30	362	349	336	323	310	298	286	274	262	251
31	368	355	343	331	318	306	295	283	272	261
34	384	373	362	351	340	329	319	308	298	288
35	389	378	368	357	347	336	326	316	306	296
38	401	392	382	373	363	354	345	336	327	318
39	405	396	387	377	368	359	350	341	333	324

Tabel 2 (vervolg)

Waarden van $10^3 \cdot P[\underline{S} \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	41	43	45	47	49	51	53	55	57	59
14	013	010	007	005	003	002	002	001	001	000
15	023	018	014	010	008	006	004	003	002	001
18	066	056	048	041	034	029	024	020	016	013
19	082	072	062	054	047	040	034	029	025	021
22	131	120	109	099	089	080	072	064	058	051
23	147	135	124	114	104	094	086	078	070	063
26	191	179	168	157	147	137	127	118	110	102
27	204	192	181	170	160	150	141	132	123	115
30	239	228	218	208	198	188	178	169	160	152
31	250	239	229	219	209	199	190	181	172	164
34	278	268	259	249	240	231	222	213	205	196
35	286	277	267	258	249	240	232	223	215	206
38	309	300	291	283	274	266	258	250	242	234
39	315	307	298	290	282	274	266	258	250	243

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	61	63	65	67	69	71	73	75	77	79
15	001	001	000	000	000	000	000	000	000	000
18	011	009	007	005	004	003	003	002	001	001
19	017	014	012	010	008	006	005	004	003	003
22	045	040	035	031	027	024	021	018	015	013
23	057	051	046	041	036	032	028	025	022	019
26	094	087	080	073	067	062	057	052	047	043
27	107	099	092	085	079	073	067	062	057	052
30	144	136	128	121	114	107	100	094	088	083
31	155	147	140	132	125	118	112	105	099	093
34	188	180	173	165	158	151	144	137	131	125
35	198	191	183	176	168	161	154	148	141	135
38	227	219	212	205	198	191	184	177	171	165
39	235	228	221	214	207	200	193	187	180	174

Tabel 2 (vervolg)

Waarden van $10^3 \cdot P[\underline{S} \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	81	83	85	87	89	91	93	95	97	99
18	001	001	000	000	000	000	000	000	000	000
19	002	002	001	001	001	001	000	000	000	000
22	011	010	008	007	006	005	004	003	003	002
23	017	015	013	011	009	008	007	006	005	004
26	039	035	032	029	026	023	021	019	017	015
27	048	044	040	036	033	030	027	025	022	020
30	077	072	067	063	059	054	051	047	043	040
31	088	082	077	072	068	063	059	055	052	048
34	119	113	107	102	097	092	087	082	078	074
35	129	123	117	112	107	101	096	092	087	083
38	158	152	147	141	135	130	125	120	115	110
39	168	162	156	150	145	139	134	129	124	119

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	101	103	105	107	109	111	113	115	117	119
22	002	002	001	001	001	001	001	000	000	000
23	004	003	003	002	002	001	001	001	001	001
26	013	012	010	009	008	007	006	005	005	004
27	018	016	015	013	012	010	009	008	007	006
30	037	034	032	029	027	025	023	021	019	017
31	045	041	038	036	033	031	028	026	024	022
34	070	066	062	058	055	052	049	046	043	040
35	078	074	070	066	063	059	056	053	050	047
38	105	101	096	092	088	084	080	076	073	069
39	114	109	105	101	096	092	088	084	081	077

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	121	123	125	127	129	131	133	135	137	139
26	004	003	003	002	002	002	001	001	001	001
27	006	005	004	004	003	003	003	002	002	002
30	016	014	013	012	011	010	009	008	007	006
31	020	019	017	016	014	013	012	011	010	009
34	038	035	033	031	029	027	025	023	022	020
35	044	042	039	037	034	032	030	028	026	025
38	066	063	060	057	054	051	049	046	044	041
39	074	070	067	064	061	058	055	053	050	048

Tabel 2 (vervolg)

Waarden van $10^3 \cdot P[S \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	141	143	145	147	149	151	153	155	157	159
26	001	001	001	000	000	000	000	000	000	000
27	001	001	001	001	001	001	001	000	000	000
30	006	005	005	004	004	003	003	003	002	002
31	008	007	007	006	006	005	004	004	004	003
34	019	017	016	015	014	013	012	011	010	009
35	023	022	020	019	017	016	015	014	013	012
38	039	037	035	033	031	029	028	026	025	023
39	045	043	041	039	037	035	033	031	029	028

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	161	163	165	167	169	171	173	175	177	179
30	002	002	001	001	001	001	001	001	001	001
31	003	003	002	002	002	002	001	001	001	001
34	008	008	007	007	006	005	005	005	004	004
35	011	010	010	009	008	007	007	006	006	005
38	022	021	019	018	017	016	015	014	013	012
39	026	025	023	022	021	020	018	017	016	015

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	181	183	185	187	189	191	193	195	197	199
31	001	001	001	001	001	000	000	000	000	000
34	003	003	003	003	002	002	002	002	002	001
35	005	005	004	004	003	003	003	003	002	002
38	011	011	010	009	009	008	008	007	007	006
39	014	014	013	012	011	010	010	009	009	008

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	201	203	205	207	209	211	213	215	217	219
34	001	001	001	001	001	001	001	001	001	000
35	002	002	002	001	001	001	001	001	001	001
38	006	005	005	004	004	004	004	003	003	003
39	007	007	006	006	006	005	005	004	004	004

Tabel 2 (vervolg)

Waarden van $10^3 \cdot P[S \geq S | n; H_0]$ voor $\frac{1}{2}n(n-1)$ oneven en $6 \leq n \leq 39$

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	221	223	225	227	229	231	233	235	237	239
35	001	001	001	001	000	000	000	000	000	000
38	003	002	002	002	002	002	002	001	001	001
39	004	003	003	003	003	002	002	002	002	002

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	241	243	245	247	249	251	253	255	257	259
38	001	001	001	001	001	001	001	001	001	000
39	002	001	001	001	001	001	001	001	001	001

$\begin{smallmatrix} S \\ n \end{smallmatrix}$	261	263	265	267	269	271	273	275	277	279
39	001	001	001	001	000	000	000	000	000	000

Tabel 3

Exacte rechter kritieke waarden van S voor $n=4(1)40$ ¹⁰⁾

n	$\alpha=0,005$	$\alpha=0,01$	$\alpha=0,025$	$\alpha=0,05$	$\alpha=0,10$	n
4	-	-	-	6	6	4
5	-	10	10	8	8	5
6	15	13	13	11	9	6
7	19	17	15	13	11	7
8	22	20	18	16	12	8
9	26	24	20	18	14	9
10	29	27	23	21	17	10
11	33	31	27	23	19	11
12	38	36	30	26	20	12
13	44	40	34	28	24	13
14	47	43	37	33	25	14
15	53	49	41	35	29	15
16	58	52	46	38	30	16
17	64	58	50	42	34	17
18	69	63	53	45	37	18
19	75	67	57	49	39	19
20	80	72	62	52	42	20
21	86	78	66	56	44	21
22	91	83	71	61	47	22
23	99	89	75	65	51	23
24	104	94	80	68	54	24
25	110	100	86	72	58	25
26	117	107	91	77	61	26
27	125	113	95	81	63	27
28	130	118	100	86	68	28
29	138	126	106	90	70	29
30	145	131	111	95	75	30
31	151	137	117	99	77	31
32	160	144	122	104	82	32
33	166	152	128	108	86	33
34	175	157	133	113	89	34
35	181	165	139	117	93	35
36	190	172	146	122	96	36
37	198	178	152	128	100	37
38	205	185	157	133	105	38
39	213	193	163	139	109	39
40	222	200	170	144	112	40

10) "-" betekent, dat, voor de betreffende waarde van n en α , voor geen enkele waarde van S de rechtséénzijdige overschrijdingskans $\leq \alpha$ is.

Tabel 4

Benaderingen voor de rechter kritieke waarden van S voor n=41(1)100

n	$\alpha=0,005$	$\alpha=0,01$	$\alpha=0,025$	$\alpha=0,05$	$\alpha=0,10$	n
41	232	210	176	148	116	41
42	239	217	183	153	121	42
43	249	225	189	159	125	43
44	256	232	196	164	128	44
45	266	240	202	170	134	45
46	275	247	209	175	137	46
47	283	255	215	181	141	47
48	292	264	222	188	146	48
49	300	272	230	192	150	49
50	309	281	237	199	155	50
51	319	289	243	205	159	51
52	328	296	250	210	164	52
53	338	306	258	216	170	53
54	347	313	265	223	173	54
55	357	323	271	229	179	55
56	366	332	280	234	184	56
57	376	340	286	240	188	57
58	387	349	295	247	193	58
59	395	357	301	253	197	59
60	406	366	310	260	202	60
61	416	376	316	266	208	61
62	427	385	325	273	213	62
63	437	395	333	279	219	63
64	446	404	340	286	224	64
65	456	412	348	292	228	65
66	467	423	357	299	233	66
67	479	431	365	305	239	67
68	488	442	372	312	244	68
69	500	452	380	320	250	69
70	511	461	389	327	255	70
71	521	471	397	333	261	71
72	532	480	406	340	266	72
73	542	490	414	348	272	73
74	553	501	421	355	277	74
75	565	511	431	361	283	75
76	576	520	440	368	288	76
77	588	530	448	376	294	77
78	599	541	457	383	299	78
79	611	551	465	391	305	79
80	622	562	474	398	310	80
81	634	572	482	406	316	81
82	645	583	491	413	323	82
83	657	593	501	421	327	83
84	668	604	510	428	334	84
85	680	614	518	436	340	85

Tabel 4 (vervolg)

Benaderingen voor de rechter kritieke waarden van S voor n=41(1)100

n	$\alpha=0,005$	$\alpha=0,01$	$\alpha=0,025$	$\alpha=0,05$	$\alpha=0,10$	n
86	693	625	527	443	345	86
87	705	637	537	451	351	87
88	716	648	546	458	358	88
89	728	658	556	466	364	89
90	741	669	565	473	369	90
91	753	681	573	481	375	91
92	766	692	584	490	382	92
93	778	702	592	498	388	93
94	791	715	603	505	395	94
95	803	725	611	513	401	95
96	816	736	622	522	406	96
97	828	748	630	530	414	97
98	841	759	641	537	429	98
99	855	771	651	547	425	99
100	866	784	660	554	432	100

Tabel 5

Tabel van $\sigma^2 = \frac{1}{18} n(n-1)(2n+5)$ en σ voor $n=4(1)100$

n	σ^2	σ	n	σ^2	σ
1	0	0	51	15158,3333	123,11918
2	1,0000	1,00000	52	16059,3333	126,72542
3	3,6667	1,91485	53	16995,3333	130,36615
4	8,6667	2,94392	54	17967,0000	134,04104
5	16,6667	4,08248	55	18975,0000	137,74977
6	28,3333	5,32291	56	20020,0000	141,49205
7	44,3333	6,65833	57	21102,6667	145,26757
8	65,3333	8,08290	58	22223,6667	149,07604
9	92,0000	9,59166	59	23383,6667	152,91719
10	125,0000	11,18034	60	24583,3333	156,79073
11	165,0000	12,84523	61	25823,3333	160,69640
12	212,6667	14,58310	62	27104,3333	164,63394
13	268,6667	16,39105	63	28427,0000	168,60308
14	333,6667	18,26655	64	29792,0000	172,60359
15	408,3333	20,20726	65	31200,0000	176,63522
16	493,3333	22,21111	66	32651,6667	180,69772
17	589,3333	24,27619	67	34147,6667	184,79087
18	697,0000	26,40076	68	35688,6667	188,91444
19	817,0000	28,58321	69	37275,3333	193,06821
20	950,0000	30,82207	70	38908,3333	197,25195
21	1096,6667	33,11596	71	40588,3333	201,46546
22	1257,6667	35,46360	72	42316,0000	205,70853
23	1433,6667	37,86379	73	44092,0000	209,98095
24	1625,3333	40,31542	74	45917,0000	214,28252
25	1833,3333	42,81744	75	47791,6667	218,61305
26	2058,3333	45,36886	76	49716,6667	222,97235
27	2301,0000	47,96874	77	51692,0000	227,36021
28	2562,0000	50,61620	78	53720,3333	231,77647
29	2842,0000	53,31041	79	55800,3333	236,22094
30	3141,6667	56,05057	80	57933,3333	240,69344
31	3461,6667	58,83593	81	60120,0000	245,19380
32	3802,6667	61,66577	82	62361,0000	249,72185
33	4165,3333	64,53939	83	64657,0000	254,27741
34	4550,3333	67,45616	84	67008,6667	258,86032
35	4958,3333	70,41543	85	69416,6667	263,47043
36	5390,0000	73,41662	86	71881,6667	268,10757
37	5846,0000	76,45914	87	74404,3333	272,77158
38	6327,0000	79,54244	88	76985,3333	277,46231
39	6833,6667	82,66599	89	79625,3333	282,17961
40	7366,6667	85,82929	90	82325,0000	286,92333
41	7926,6667	89,03183	91	85085,0000	291,69333
42	8514,3333	92,27315	92	87906,0000	296,48946
43	9130,3333	95,55278	93	90788,6667	301,31158
44	9775,3333	98,87029	94	93733,6667	306,15954
45	10450,0000	102,22524	95	96741,6667	311,03322
46	11155,0000	105,61723	96	99813,3333	315,93248
47	11891,0000	109,04586	97	102949,3333	320,85719
48	12658,6667	112,51074	98	106150,3333	325,80720
49	13458,6667	116,01149	99	109417,0000	330,78241
50	14291,6667	119,54776	100	112750,0000	335,78267

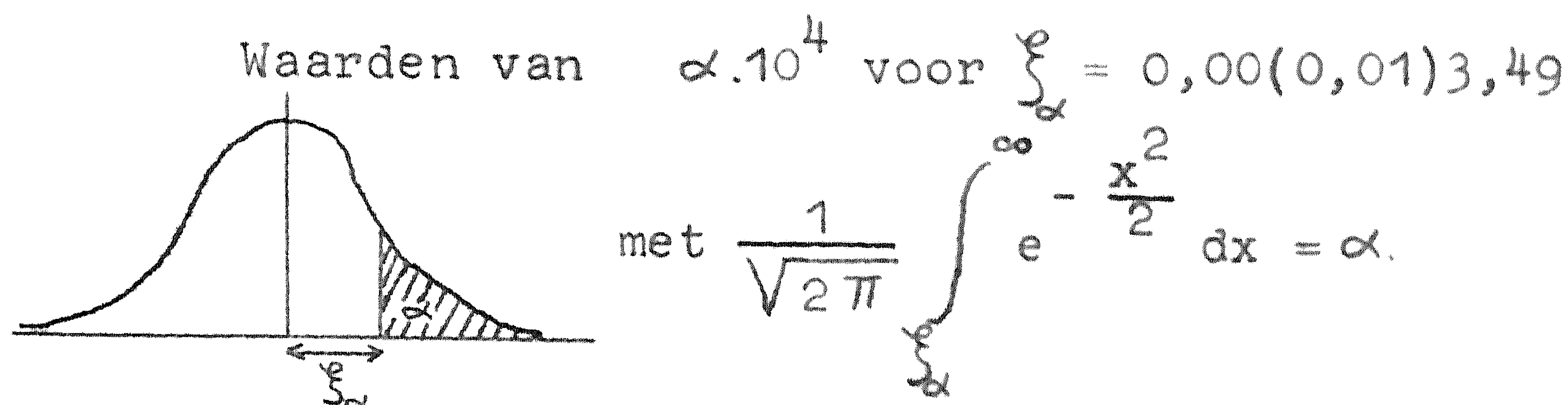
Tabel 6

Tabel van t^2 en t^3 voor $t = 1(1)40$

t	t^2	t^3	t	t^2	t^3
1	1	1	21	441	9261
2	4	8	22	484	10648
3	9	27	23	529	12167
4	16	64	24	576	13824
5	25	125	25	625	15625
6	36	216	26	676	17576
7	49	343	27	729	19683
8	64	512	28	784	21952
9	81	729	29	841	24389
10	100	1000	30	900	27000
11	121	1331	31	961	29791
12	144	1728	32	1024	32768
13	169	2197	33	1089	35937
14	196	2744	34	1156	39304
15	225	3375	35	1225	42875
16	256	4096	36	1296	46656
17	289	4913	37	1369	50653
18	324	5832	38	1444	54872
19	361	6859	39	1521	59319
20	400	8000	40	1600	64000

Tabel 7

TABEL VAN DE NORMALE VERDELING



	0	1	2	3	4	5	6	7	8	9
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3,0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3,3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3,4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Een 5 betekent, dat bij afronding naar het dichtsbijgelegen oneven cijfer afgerond moet worden. Een 5 wordt naar het dichtstbijzijnde even cijfer afgerond.