STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

Rangcorrelatie en de Schattingsproef van Varangot

J. Hemelrijk

S 41
(M 5)

1950

Overdruk uit Statistica, Jaargang 4, no. 5, 1950

Rangcorrelatie en de M^2-schattingsproef van Varangot *)

door J. Hemelrijk
Mathematisch Centrum, Amsterdam

S u m m a r y

Rank correlation methods applied to an experiment on estimation of V. Varangot.

The results of an experiment on estimation described by V. V a r a n g o t [1] *are analysed with the method described in M. G. K e n d a l l' s ,,Rank correlation methods'' under the title ,,The problem of* m *rankings''. The main advantage of this method is the possibility of applying a significance test to the result.*

§ 1. *Inleiding.*

De door V. V a r a n g o t [1] beschreven schattingsproef, waarbij, door een veertigtal proefpersonen verrichte, aflezingen van een aantal wijzerstanden werden vergeleken, is, zowel door hem zelf als door H. C. H a m a k e r [2], uitvoerig van commentaar voorzien. Het is niet de bedoeling, hier deze beschouwingen aan kritiek te onderwerpen, of te trachten deze langs dezelfde lijnen nog verder voort te zetten. In verband echter met de eveneens in dit nummer verschijnende boekbespreking van M. G. K e n d a l l 's ,,Rank correlation methods'' [3] is het interessant na te gaan, welke resultaten men bereiken kan door toepassing van deze methoden op de door V a r a n g o t verzamelde gegevens **). Men kan, zoals uit de volgende paragrafen blijkt, met behulp van de in hoofdstuk 6 en 7 van K e n d a l l 's boekje beschreven methode der m rangschikkingen (,,problem of m rankings'') een beoordeling krijgen zowel van de schattingscapaciteiten der proefpersonen, als van de ,,moeilijkheid'' der te schatten wijzerstanden. Daar men in de theorie der rangcorrelatie niet werkt met de waarde der waarnemingen zelf, maar met hun rangnummers bij rangschikking naar grootte, verkrijgt men als resultaat ook geen in getallen uitgedrukte waardering van schattingscapaciteit resp. moeilijkheid, maar slechts een *volgorde* van deze capaciteiten, resp. van de ,,moeilijkheid'' van het schatten der wijzerstanden. Dit is, uiteraard, een ernstige beperking. Daartegenover staat het grote voordeel, dat men bij deze theorie de beschikking heeft over toetsingsmethoden, die de mogelijkheid scheppen, om na te gaan of de gevonden volgorde redelijkerwijze als een toevallig ontstane kan gelden of niet. Dit is van groot belang. Immers, ook

---

*) Zie V. V a r a n g o t [1] en H. C. H a m a k e r [2].

**) C. S p e a r m a n, wiens door J. H. E n t e r s [4] besproken criterium de aanleiding geweest is voor de door V a r a n g o t uitgevoerde proef, heeft een aanzienlijk aandeel gehad in de ontwikkeling van de door K e n d a l l besproken methoden.

als de proefpersonen alle even bekwaam (of onbekwaam) zijn mag men niet verwachten, dat zij bij iedere wijzerstand alle dezelfde schatting zullen geven. Legt men nu de één of andere maatstaf voor hun bekwaamheid aan, dan zal men in het algemeen toch verschillen in bekwaamheid constateren. Men kan dan, indien men over geschikte toetsingsmethoden beschikt, nagaan, of deze uitkomst betekenis heeft of zuiver toevallig is. Beschikt men niet over dergelijke toetsingsmethoden, dan blijft men daarover in het onzekere.

In de volgende paragrafen zullen wij de toepassing van de methode der m rangschikkingen op een gedeelte van het materiaal van de schattingsproef uitvoeren zonder de bewijzen van de toegepaste stellingen te geven. De theoretische achtergrond kan men in de monographie van K e n d a l l [3] naslaan. Uit de 40 proefpersonen van de proef van V a r a n g o t zijn, om ruimte en berekeningen te sparen, een twaalftal gekozen, waaronder no. 20 en no. 5, die door deze methode als de beste resp. de slechtste schatter worden aangewezen. De berekeningen kunnen echter geheel analoog voor alle 40 proefpersonen worden uitgevoerd.

§ 2. *De beoordeling van de schattingscapaciteiten der proefpersonen.*

In tabel I vindt men dat gedeelte der proefresultaten van V a r a n g o t samengevat, dat wij in onze beschouwingen betrekken. De letters *a*, ..., *j* geven de wijzerstanden aan, de nummers in de eerste kolom zijn de nummers van de gekozen proefpersonen. De laatste regel bevat de ware wijzerstanden.

TABEL I

De door twaalf proefpersonen geschatte wijzerstanden

Proefpers No.	Wijzerstand									
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
	Geschatte wijzerstand									
3	65	45	33	11	88	72	98	23	3	58
5	70	40	30	10	90	70	95	20	5	65
14	65	45	34	13	87	68	97	29	4	54
16	65	40	35	10	90	75	98	25	2	55
17	61	45	35	10	90	75	97	25	5	55
18	65	45	30	10	90	70	98	25	4	55
20	65	47	34	11	89	70	98	26	3	58
24	65	48	40	10	90	70	98	25	4	57
25	65	47	30	12	88	71	97	27	4	56
26	68	46	30	10	90	72	97	30	5	60
34	66	47	39	12	88	75	98	25	4	59
38	65	45	30	12	88	76	97	28	3	60
Ware wijzerst.	64	46	35	11	89	72	98	26	3	57

Indien slechts één wijzerstand door de proefpersonen geschat was, zou men, zoals ook V a r a n g o t en H a m a k e r opmerken, de absolute waarde van de afwijking als maat voor de schattingsfout kunnen nemen. Doen wij dit, dan vinden wij voor iedere wijzerstand een rangschikking der proefpersonen naar hun schattingscapaciteit. Het principe van de methode der *m* rangschikkingen bestaat nu daarin, dat men deze rangorde voor ieder der wijzerstanden apart opstelt en de resultaten vervolgens combineert tot één samenvattend oordeel.

Daartoe maken wij eerst tabel II op, waarin in plaats van de schattingen de absolute afwijkingen van de ware wijzerstanden zijn genoteerd.

TABEL II

De verschillen tussen geschatte en ware wijzerstand in absolute waarde

Proefpers. No.	Wijzerstand									
	a	*b*	*c*	*d*	e	*f*	*g*	*h*	*i*	*j*
	Verschil									
3	1	1	2	0	1	0	0	3	0	1
5	6	6	5	1	1	2	3	6	2	8
14	1	1	1	2	2	4	1	3	1	3
16	1	6	0	1	1	3	0	1	1	2
17	3	1	0	1	1	3	1	1	2	2
18	1	1	5	1	1	2	0	1	1	2
20	1	1	1	0	0	2	0	0	0	1
24	1	2	5	1	1	2	0	1	1	0
25	1	1	5	1	1	1	1	1	1	1
26	4	0	5	1	1	0	1	4	2	3
34	2	1	4	1	1	3	0	1	1	2
38	1	1	5	1	1	4	1	2	0	3

Voor ieder der wijzerstanden apart wordt nu een volgorde der proefpersonen bepaald: de proefpersoon, die de kleinste absolute fout heeft gemaakt, krijgt rangnummer 1, de volgende 2, enz. Indien gelijken voorkomen (wat hier veelvuldig het geval is) delen zij de hun toekomende rangnummers gelijk op. Bij wijzerstand *a* b.v. is de kleinste fout gelijk aan 1 en deze is gemaakt door 8 proefpersonen; deze krijgen derhalve alle als rangnummers het getal $\frac{1}{8}(1 + 2 + \ldots + 8) = 4{,}5$. Bij wijzerstand *b* zijn de 2e, 3e, . . ., 9e der naar grootte gerangschikte afwijkingen gelijk, die dus alle het rangnummer $(9 + 2)/2 = 5{,}5$ krijgen, en verder de 11e en de 12e, die beide het rangnummer 11,5 krijgen. De op deze wijze verkregen rangnummers zijn samengevat in tabel III, waarin iedere rij de bij één wijzerstand behorende rangnummers voorstelt en iedere kolom de door één proefpersoon behaalde rangnummers.

TABEL III

De rangnummers der proefpersonen bij verschillende wijzerstanden

Proef-pers. No.		*3*	*5*	*14*	*16*	*17*	*18*	*20*	*24*	*25*	*26*	*34*	*38*
Wijzer-stand	*a*	4,5	12	4,5	4,5	10	4,5	4,5	4,5	4,5	11	9	4,5
	b	5,5	11,5	5,5	11,5	5,5	5,5	5,5	10	5,5	1	5,5	5,5
	c	5	9,5	3,5	1,5	1,5	9,5	3,5	9,5	9,5	9,5	6	9,5
	d	1,5	7	12	7	7	7	1,5	7	7	7	7	7
	e	6,5	6,5	12	6,5	6,5	6,5	1	6,5	6,5	6,5	6,5	6,5
	f	1,5	5,5	11,5	9	9	5,5	5,5	5,5	3	1,5	9	11,5
	g	3,5	12	9	3,5	9	3,5	3,5	3,5	9	9	3,5	9
	h	9,5	12	9,5	4,5	4,5	4,5	1	4,5	4,5	11	4,5	8
	i	2	11	6,5	6,5	11	6,5	2	6,5	6,5	11	6,5	2
	j	3	12	10	6,5	6,5	6,5	3	1	3	10	6,5	10
Som		42,5	99	84	61	70,5	59,5	31	58,5	59	77,5	64	73,5
Rangnr.		2	12	11	6	8	5	1	3	4	10	7	9
d		−22,5	34	19	−4	5,5	−5,5	−34	−6,5	−6	12,5	−1	8,5

d = Som — Gemiddelde som = Som — 65.

Voor ieder der proefpersonen wordt nu de som der aldus verkregen rangnummers berekend, en deze som bepaalt de samenvattende beoordeling van de volgorde van bekwaamheid der proefpersonen. Volgens deze methode is proefpersoon no. 20 dus de bekwaamste en no. 5 de onbekwaamste. Een indruk van het verschil in bekwaamheid van de verschillende proefpersonen verkrijgt men uit de som der rangnummers.

Uiteraard schuilt ook in deze methode, evenals in alle andere, een arbitrair element. Wij zullen nu echter nagaan, of de verkregen beoordeling het gevolg is van een behoorlijke overeenstemming tussen de voor ieder der wijzerstanden apart gevonden beoordelingen, of niet. Dit gebeurt in de vorm van de bepaling van een *coëfficiënt van overeenstemming*, W, die gelijk aan 1 is, indien er volledige overeenstemming bestaat tussen de 10 beoordelingen afgeleid voor ieder der 10 wijzerstanden afzonderlijk en die klein is (met minimale waarde gelijk nul) indien deze overeenstemming slecht is. Met behulp van een aan deze coëfficiënt verwante grootheid χ_r^2 kunnen wij vervolgens de hypothese H toetsen, dat de beoordelingen behorende bij de verschillende wijzerstanden, onderling onafhankelijk van elkaar op toevallige wijze tot stand zijn gekomen, dat wil zeggen, dat alle $(n!)^m$ mogelijke permutaties van de toegekende rangnummers a priori even waarschijnlijk geacht moeten worden *).

*) m is het aantal rijen, dus m = 10 in tabel III, en n het aantal kolommen, dus n = 12 in tabel III.

Dit is b.v. het geval, indien alle proefpersonen even bekwame schatters zijn voor ieder der wijzersanden. Onder deze hypothese is de verdeling van de coëfficiënt χ_r^2 bij benadering bekend *). W wordt als volgt gedefinieerd:

Zij

$$S = \Sigma d_i^2, \tag{1}$$

waarin d_i de afwijkingen van het gemiddelde zijn in de rij van tabel III, die de som der door de proefpersonen behaalde rangnummers aangeeft.

Komen er in één der rijen $a, \ldots, j$, t gelijke rangnummers voor, dan wordt de grootheid

$$\tfrac{1}{12}\, t\,(t^2 - 1)$$

berekend; voor iedere rij worden eventueel meerdere dergelijke termen berekend. Zij nu

$$T' = \tfrac{1}{12}\, \Sigma\, t\,(t^2 - 1) \tag{2}$$

de som van deze termen voor één rij; in de eerste rij van tabel III b.v. neemt t eenmaal de waarde 8 aan en vier maal de waarde 1; dus $\Sigma\, t\,(t^2 - 1) = 8\,(64 - 1) + 4 \times 1\,(1 - 1) = 504$; evenzo vinden we voor de zesde rij $t = 4, 3, 2, 2$, en 1, waaruit volgt $\Sigma\, t\,(t^2 - 1) = 96$.

Nu is de bovengenoemde *coëfficiënt van overeenstemming*

$$W = \frac{S}{\tfrac{1}{12}\, m^2\,(n^3 - n) - m\,\Sigma\, T'} \tag{3}$$

waarin m het aantal rijen aangeeft (dus in ons geval $m = 10$), n het aantal kolommen (dus $n = 12$), terwijl de som $\Sigma\, T'$ op de m rijen betrekking heeft. Men kan nu bewijzen dat bij volledige overeenstemming tussen alle rijen, W inderdaad juist gelijk 1 wordt.

Is echter de bovengenoemde hypothese juist, d.w.z. staat in iedere rij een willekeurige permutatie van de in die rij voorkomende getallen, dan zal W veel kleinere waarden aannemen, en we zullen de juistheid van de hypothese kunnen beoordelen op grond van de frequentieverdeling van W. Geschikter kan daartoe de grootheid

$$\chi_r^2 = m\,(n - 1)\,W$$

dienen, omdat kan worden bewezen dat deze grootheid bij benadering een χ^2-verdeling bezit met $n - 1$ vrijheidsgraden. Bij te grote waarde van χ_r^2 zal men dan de hypothese verwerpen.

Passen we deze methode toe op tabel III dan vinden we $S = \Sigma d_i^2 = 3563{,}5$;

*) Voor kleine m en n en voor het geval er geen gelijken zijn, is de verdeling van χ_r^2 exact bekend.

$\Sigma T' = 330$, en hiermede $W = 0{,}324$; $\chi_r^2 = 35{,}6$. Daar het aantal vrijheidsgraden 11 bedraagt, is deze waarde van χ_r^2 zeer hoog; de kans, dat een nog grotere waarde van χ_r^2 gevonden zou worden, terwijl de hypothese H juist is, (de z.g. overschrijdingskans) is ongeveer 0,0002. Wij kunnen de hypothese H dus met stelligheid verwerpen.

Hoe luidt nu de conclusie? Populair uitgedrukt kan men zeggen, dat de verkregen volgorde van bekwaamheid, hoewel er zeker nog wel toevallige factoren in het spel zijn, mede door een systematische factor tot stand is gekomen. Deze systematische factor zou men nu juist de ,,bekwaamheid'' kunnen noemen, indien tenminste de overige omstandigheden voor alle proefpersonen gelijk zijn geweest. Exact uitgedrukt houdt het toepassen van de methode het geven van een *definitie* van bekwaamheid (eigenlijk volgorde van bekwaamheid) in. Aannemende, dat voor ieder der proefpersonen de schattingsfout een, eventueel, voor iedere wijzerstand verschillende, waarschijnlijkheidsverdeling bezit (zodat er in totaal $m \cdot n$ dergelijke foutenverdelingen in de methode betrokken zijn), bezitten alle rangnummers uit tabel III een verwachtingswaarde. In het bijzonder is dit dus het geval met de rij rangnummers, die de uiteindelijke volgorde van bekwaamheid aangeeft. De met het experiment verkregen rij van rangnummers is nu een *schatting* van deze rij van verwachtingswaarden van rangnummers. De door de verwachtingswaarden aangegeven volgorde van bekwaamheid zouden wij de ,,ware'' volgorde kunnen noemen. De gevonden volgorde is dan een schatting van deze ,,ware''. Wij komen hierop in § 4 nog terug.

Het feit, dat de hypothese H verworpen wordt houdt nu in, dat de verwachtingswaarden van de rangnummers voor de bekwaamheid niet voor alle proefpersonen gelijk zijn, met andere woorden, dat er een ,,ware'' volgorde in bovengenoemde zin is, waarbij niet alle proefpersonen als even bekwaam gelden.

Dit betekent dus, dat het zinvol is, deze volgorde te schatten, hetgeen niet het geval is, indien hypothese H juist is. Het betekent echter niet, dat men dan van een willekeurig *tweetal* proefpersonen reeds bepaald zou hebben, wie de beste schatter is. De proefpersonen, no. 18, 24 en 25 b.v. verschillen slechts zeer weinig bij de einduitslag en men mag verwachten, dat dezelfde methode, toegepast op de gegevens van deze proefpersonen alléén geen significante uitkomst zou hebben. Voor de vergelijking van twee proefpersonen zijn echter andere dan de hier beschreven methoden beter geschikt. Hierover vindt men een korte opmerking in § 5.

§ 3. *De beoordeling van de moeilijkheid der wijzerstanden.*

Op geheel analoge wijze kunnen wij nu voor ieder der proefpersonen

apart de wijzerstanden naar volgorde van moeilijkheid rangschikken. Wij leiden daartoe dus uit tabel II een tabel af op dezelfde wijze als tabel III, maar nu met 12 rijen (één voor iedere proefpersoon) en 10 kolommen voor de wijzerstanden. Dit levert tabel IV.

TABEL IV

Rangnummers der wijzerstanden bij verschillende proefpersonen

Wijzer-stand		*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
	3	6,5	6,5	9	2,5	6,5	2,5	2,5	10	2,5	6,5
	5	8	8	6	1,5	1,5	3,5	5	8	3,5	10
	14	3	3	3	6,5	6,5	10	3	8,5	3	8,5
Proef-	*16*	5	10	1,5	5	5	9	1,5	5	5	8
pers.	*17*	9,5	4	1	4	4	9,5	4	4	7,5	7,5
No.	*18*	4,5	4,5	10	4,5	4,5	8,5	1	4,5	4,5	8,5
	20	7,5	7,5	7,5	3	3	10	3	3	3	7,5
	24	5	8,5	10	5	5	8,5	1,5	5	5	1,5
	25	5	5	10	5	5	5	5	5	5	5
	26	8,5	1,5	10	4	4	1,5	4	8,5	6	7
	34	7,5	4	10	4	4	9	1	4	4	7,5
	38	4	4	10	4	4	9	4	7	1	8
Som		74	66,5	88	49	53	86	35,5	72,5	50	85,5
Rangnr.		7	5	10	2	4	9	1	6	3	8
d		8	0,5	22	−17	−13	20	−30,5	6,5	−16	19,5

Wijzerstand *g* wordt blijkens tabel IV aangewezen als de gemakkelijkste en *c* als de moeilijkste. Verder vinden wij: $m = 12$; $n = 10$; $S = \Sigma d_i^2 = 3015$; $\Sigma T' = 173$; $W = 0{,}307$ en $\chi_r^2 = 33{,}2$, hetgeen, daar het aantal vrijheidsgraden 9 is, overeenkomt met een overschrijdingskans 0,0001. Ook hier mogen wij dus constateren, dat de verkregen volgorde van moeilijkheid niet toevallig ontstaan is, maar dat er, althans voor de gebruikte groep proefpersonen, een sterk significante overeenstemming van de volgorde van grootte der bij de verschillende wijzerstanden gemaakte fouten bestaat.

§ 4. *Verdere bewerking.*

Een van de vragen, waarop men zich heeft te beraden, is, of het wel gerechtvaardigd is bij de boven beschreven methode aan alle rijen gelijk gewicht toe te kennen. Men zou geneigd zijn aan de moeilijkere wijzerstanden een groter gewicht te geven dan aan de gemakkelijke en eveneens, bij de beoordeling van de moeilijkheid der schatting van de wijzerstanden, aan het oordeel

van een goede schatter grotere waarde te hechten dan aan het oordeel van een slechte.

Het toekennen van ongelijke gewichten houdt echter een *wijziging* in *van de definitie* van de „ware'' volgorde van bekwaamheid (resp. moeilijkheid van schatten), die in § 3 gegeven is. De vraag, of men gewichten wenst toe te kennen of niet, hangt dus af van de keuze van de definitie. En deze keuze is afhankelijk van tal van overwegingen, veelal van niet-statistische aard waarover een algemene uitspraak ternauwernood mogelijk is. In het bijzonder hangt deze keuze af van het *doel*, dat men zich stelt. Wenst men uit de proefpersonen diegene te kiezen, die het beste *moeilijk* te schatten of juist *gemakkelijk* te schatten punten kan aflezen? Wenst men een schatter, die géén voorkeur voor even cijfers heeft, of niet geneigd is op vijftallen af te ronden, of is dit niet van belang? Zonder precisering van deze doelstelling, of van de te gebruiken definitie van bekwaamheid bepaald op andere gronden, kan men over de keuze van gewichten niet beslissen. Is de definitie van bekwaamheid gepreciseerd, dan kan men bovendien bij de keuze der wijzerstanden met die definitie rekening houden en daardoor het onderzoek nauwer bij het doel aansluiten.Verder blijft dan nog de vraag, welke van de door V a r a n g o t en H a m a k e r of van de hier besproken methoden de beste methode is ter bereiking van het gestelde doel: een zo goed mogelijke schatting te verkrijgen van de nu scherper gedefinieerde „ware'' volgorde van bekwaamheid. K e n d a l l bewijst dat de door de methode der m rangschikkingen verkregen schatting in zekere zin als een „beste'' schatting kan worden beschouwd, daar deze schatting een zekere overeenkomst vertoont met volgens de methode der kleinste kwadraten verkregen schattingen bij andere problemen.

Verder zou men, aangezien het hier weer om een waardering van schattingscapaciteiten gaat, de besproken theorie (of één der andere) misschien op deze schattingsmethoden zelf kunnen toepassen.

Om enige indruk te verkrijgen van de *invloed*, die de toekenning van gewichten *) op de einduitslag voor de bekwaamheid der proefpersonen zou hebben, hebben wij de wijzerstanden gesplitst in twee groepen, t.w.

groep A : g, d, i, e en b en groep B : h, a, j, f en c ,

die volgens de vorige paragraaf de 5 gemakkelijkste resp. de moeilijkste wijzerstanden zijn. Het is overbodig de berekeningen hier wederom geheel

*) Toekenning van gewichten kan geschieden door de verkregen rangnummers van iedere rangschikking met het bijbehorende gewicht te vermenigvuldigen. De toetsingsmethode voor de hypothese H is echter alleen uitgewerkt voor gelijke gewichten.

te geven. Volgens het in § 2 gevolgde systeem verkrijgen wij nu de in tabel V samengevatte beoordelingen.

TABEL V

Volgorde van bekwaamheid der proefpersonen, bij verschillende groepen wijzerstanden

No.	*20*	*3*	*24*	*25*	*18*	*16*	*34*	*17*	*38*	*26*	*14*	*5*
$A + B$	1	2	3	4	5	6	7	8	9	10	11	12
A	1	2	6	7,5	3,5	8	3,5	10	5	7,5	11	12
B	1	2	4	3	6	5	8	7	11	10	9	12

Hierin stelt de eerste regel de volgorde van bekwaamheid der proefpersonen voor, die verkregen wordt, indien alle wijzerstanden gebruikt en de tweede en derde regel de volgorden, indien groep A resp. B wordt gebruikt. Men ziet, dat de twee bekwaamste en de minst bekwame proefpersonen hun plaatsen onbetwist behouden, terwijl de overigen bij de 3 verschillende beoordelingen nogal verschillend uit de bus komen. De overschrijdingskansen der χ_r^2, die bij de beoordelingen met behulp van groep A en groep B behoren, zijn echter vrij groot (nl. ongeveer 0,03 resp. 0,05), waaruit blijkt, dat de overeenstemming veel geringer is dan bij de beoordeling met de groepen A en B tezamen. Mogelijkerwijs is het aantal wijzerstanden in deze groepen te gering om de niet zeer sterk uiteenlopende bekwaamheid van de middelmatig bekwame proefpersonen behoorlijk te onderscheiden. Men zou dit kunnen nagaan, door de proefpersonen no. 20, 3 en 5 weg te laten en de berekeningen met de overige gegevens te herhalen. Dit zou ons echter te ver voeren. Wij hebben slechts een indruk willen geven van de middelen, die door de theorie der rangcorrelatie voor dergelijke problemen aan de hand worden gedaan.

§ 5. *Opmerking.*

Aan het einde van § 2 is reeds opgemerkt, dat men, om de bekwaamheid van *twee* proefpersonen te vergelijken, gevoeliger methoden kan toepassen. Men kan dan n.l. de hypothese toetsen, dat beide personen even bekwaam zijn, en wel op de volgende wijze:

Indien beide personen even bekwaam zijn, zal bij gegeven wijzerstand de absolute fout in de schatting voor beide personen dezelfde waarschijnlijkheidsverdeling bezitten. Noemen wij het verschil van beide absolute fouten bij de k^e wijzerstand: z_k, dan is, indien de proefpersonen onafhankelijk van elkaar werken, z_k voor iedere k symmetrisch om nul verdeeld, waarbij de verdeling overigens voor iedere k verschillend kan zijn, nl. afhankelijk van de wijzerstand. Van iedere z_k is dan één waargenomen waarde beschikbaar. De hypothese, dat z_k voor iedere k symmetrisch om nul is verdeeld, kan

nu, op grond van deze waargenomen waarden der z_k, getoetst worden met een onlangs op het Mathematisch Centrum ontworpen parametervrije methode [5]. Leidt deze toetsingsmethode tot verwerping van de gestelde hypothese, dan wijst dit op een significant grotere bekwaamheid van één van beide proefpersonen.

De berekeningen, die in dit artikel zijn vermeld, zijn grotendeels uitgevoerd door de heer M. de Vries, die ik daarvoor mijn hartelijke dank betuig.

LITERATUUR:

1) V. Varangot, Een schattingsproef, Statistica **3**, 144—153, 1949.
2) H. C. Hamaker, Systematische en toevallige fouten bij het aflezen van de stand van een wijzer op een schaal, Statistica **3**, 209—223, 1949.
3) M. G. Kendall, Rank correlation methods, Griffin, London, 1948; zie ook de hierachter volgende boekbespreking.
4) J. H. Enters, De betrouwbaarheid van stukscontrôle, Statistica **1**, 40—44, 1946.
5) J. Hemelrijk, A family of parameterfree tests for symmetry with respect to a given point, Proc. Kon. Ned. Ak. **53**, 945—955, 1950; ook Indagationes Mathematicae **12**, 340—350, 1950.

Boekbesprekingen

Rank correlation methods, M. G. Kendall, *Griffin and Co., London* 1948, 160 pag., 18/—.

Rang-invariante methoden, dat zijn methoden, waarbij alleen de *volgorde* van grootte van de waarnemingen in de statistische beschouwingen betrokken wordt, bezitten boven de ,,klassieke'' methoden grote voordelen en grote nadelen. In het algemeen kan men rang-invariante methoden toepassen onder zeer zwakke onderstellingen; soms is men nog wel genoodzaakt te eisen, dat de stochastische variabelen, waarop de waarnemingen betrekking hebben, een *continue* waarschijnlijkheidsverdeling bezitten, terwijl ook bepaalde eisen van onafhankelijkheid noodzakelijk zijn, maar verder behoeft men gewoonlijk niet te gaan. Tegenover dit voordeel staat het nadeel, dat de methoden meestal minder ,,scherp'' (minder ,,efficiënt'') zullen zijn dan de klassieke methoden, hetgeen zijn oorzaak daarin vindt, dat bij de klassieke methoden voor de waarschijnlijkheidsverdeling van de variabelen een bepaalde vorm wordt aangenomen. De gehele klassieke correlatie-rekening berust b.v. op de aanname van een 2- of meer-dimensionale *normale* verdeling van de stochastische variabelen. Over de voor- en nadelen van de rang-invariante methoden zijn de meningen nogal verdeeld. Velen achten het verlies aan efficiency ernstig en maken liever extra onderstellingen, ook al zijn deze wellicht niet, of niet geheel (of zelfs geheel niet) vervuld. Anderen zijn in de eerste plaats gesteld op exactheid en vermijden bij voorkeur onderstellingen over de vorm der waarschijnlijkheidsverdelingen, vooral in die gevallen, waarin deze onderstellingen slechts worden gemaakt om de wiskundige beheersing van de theorie te vergemakkelijken. De situatie wordt kernachtig samengevat door J. Wolfowitz in een artikel (,,Non-parametric statistical inference'') in de ,,Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability'', Berkeley 1949, p. 93:

> '' . . . . if the functional forms of the distribution functions are known or if there is good ground for assuming them, it is a loss not to make use of this information. Where this information is not at hand, statistical inference must properly proceed without it. In the latter event the criticism of some statisticians that non-parametric tests are ,,inefficient'' is not valid, because ,,efficiency'' (in the colloquial sense) implies thorough use of available resources, and it cannot be inefficient not to make use of unavailable information''.

De door Kendall in zijn boek ,,Rank correlation methods'' samengevatte methoden der rangcorrelatie houden zich, evenals de normale correlatierekening, bezig met de afhankelijkheid van twee (of meer) variabele grootheden. Het begrip ,,lineariteit'', dat in de klassieke correlatie- en regressie-rekening een centrale plaats inneemt, wordt echter in de rang-invariante correlatierekening vervangen door het begrip ,,monotonie''.

De noodzaak van deze vervanging wordt duidelijk, als men overweegt, dat niet met de grootte der waarnemingen wordt gewerkt, maar met hun grootte-volgorde en wel in het bijzonder met hun rangnummer bij rangschikking naar grootte. Een lineair verband tussen een rij van getallen-paren $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is, op die wijze, niet te onderscheiden van een monotoon verband.

Het eerste probleem der rangcorrelatie is het definiëren van een rangcorrelatiecoëfficiënt voor een dergelijke rij getallenparen (x_i, y_i) $(i = 1, \ldots, n)$, die een maat moet zijn voor de overeenstemming tussen de twee rijen der rangnummers van $x_1, \ldots, x_n$ en $y_1, \ldots, y_n$, bij rangschikking naar grootte van ieder der beide rijen. Indien de overeenstemming volledig is, d.w.z. indien voor iedere i het rangnummer van x_i in de rij der x-waarden het-

zelfde is als het rangnummer van y_i in de rij der y-waarden, wenst men dat de rangcorrelatiecoëfficiënt de waarde 1 aanneemt, en indien de rangnummers van x_i en y_i voor iedere i tezamen gelijk aan $n + 1$ zijn, dus als de rangschikking naar grootte voor de x_i en de y_i tegengesteld is, dient de waarde -1 aangenomen te worden, terwijl verder slechts waarden tussen -1 en $+1$ mogelijk zijn. Verschillende van deze coëfficiënten zijn ontwikkeld en twee daarvan, Spearman's coëfficiënt ρ en Kendall's coëfficiënt τ, worden door Kendall steeds naast elkaar behandeld. Wij zullen hier niet ingaan op de vraag, welke van deze twee coëfficiënten te prefereren is. Beide bezitten hun voor- en nadelen, die in het boek besproken worden. Indien $x_1, \ldots, x_n$ en $y_1, \ldots, y_n$ onafhankelijke waarnemingen zijn van stochastische variabelen x resp. y, is de waarschijnlijkheidsverdeling van ρ, evenals die van τ, bekend. Voor kleine n is deze exact berekend: Kendall geeft hem voor $n \leqq 10$; de Rekenafdeling van het Mathematisch Centrum heeft de verdeling berekend voor $n \leqq 40$ en zal de resultaten binnenkort publiceren in de Proceedings van de Koninklijke Nederlandse Akademie van Wetenschappen; voor grote n zijn benaderingen bekend. Zowel ρ als τ zijn asymptotisch normaal verdeeld. Met behulp van deze gegevens kan men dus de onafhankelijkheid van x en y toetsen, met als alternatieve hypothese een monotoon verband tussen x en het voorwaardelijk gemiddelde van y bij deze waarde van x. De onderstellingen zijn hierbij tot een minimum gereduceerd; men behoeft slechts aan te nemen, dat x en y een simultane waarschijnlijkheidsverdeling bezitten en dat de waarnemingen onafhankelijk zijn. De waarschijnlijkheidsverdeling behoeft *niet continu* te zijn; ook als gelijke waarnemingen voorkomen, kan de toetsingsmethode worden toegepast. Het is wellicht nuttig op dit punt een voorbeeld in te lassen.

In de kwartaalverslagen van het Bureau van Statistiek der Gemeente Amsterdam vindt men, onder het hoofd „Branden", de volgende gegevens over het aantal „Baldadige alarmeringen, waarop de brandweer is uitgerukt":

	Jan.	Feb.	Mrt	Apr.	Mei	Juni	Juli	Aug.	Sept.	Oct.	Nov.	Dec.
1948	1	5	1	3	6	4	5	8	12	2	3	2
1949	1	4	1	2	1	3	4	4	5	1	3	4

De overeenkomst voor de beide jaren is opvallend. Indien men nu op grond van deze gegevens wil nagaan, of het aantal „baldadige alarmeringen" samenhangt met de maand van het jaar, zal men zich met de klassieke correlatietheorie weinig gelukkig voelen, daar er in een dergelijk geval wel bijzonder weinig reden is, een *normale* waarschijnlijkheidsverdeling voor dit aantal aan te nemen. Daar men echter de hypothese toetst, dat er géén verband is tussen de maand en het aantal, is er weinig bezwaar om aan te nemen, dat dit aantal een of andere, geheel ongespecificeerde, maar uiteraard discrete, waarschijnlijkheidsverdeling bezit. Bepaalt men nu de rangcorrelatiecoëfficiënt τ van de beide rijen getallen, dan blijkt deze de waarde 0,54 te bezitten, terwijl de tweezijdige overschrijdingskans de waarde 0,02 bezit, zodat men de neiging heeft, aan de overeenstemming betekenis toe te kennen. Hierbij dient echter het voorbehoud gemaakt te worden, dat de onderstelling der onafhankelijkheid van de opeenvolgende waarnemingen zeer aanvechtbaar is, terwijl bovendien het verschijnsel onderzocht is, omdat het bij beschouwing van de gegevens in het oog sprong. Vergelijking met de gegevens van 1950 zou dus zeer gewenst zijn, maar is nog niet mogelijk.

Zouden deze gegevens van 1950 of van nog meer jaren beschikbaar zijn (voor de jaren 1946 en 1947 zijn zij wel beschikbaar, maar de aantallen zijn voor die jaren zo laag, dat ze voor dit doel practisch onbruikbaar zijn), dan neemt het probleem een andere vorm

aan, daar wij dan niet met *paren*, maar met drietallen, of in het algemeen met m-tallen, waarnemingen te maken hebben (voor iedere maand een m-tal waarnemingen, als de gegevens van m jaar beschikbaar zijn). Ook voor dit geval is in de theorie der rangcorrelatie al veel bereikt. Wij zullen dit hier niet uitvoerig bespreken daar elders in dit nummer van ,,Statistica'' een voorbeeld van toepassing van deze ,,methode der m rangschikkingen,'' (,,problem of m rankings'') wordt gegeven.

In het bovenstaande is slechts één van de problemen aangegeven, die K e n d a l l in zijn boek behandelt. Het is uiteraard ondoenlijk alle daar besproken problemen uit te stippelen. De theorie van de boven aangegeven problemen wordt behandeld in de eerste 6 hoofdstukken van het boek. Er volgt dan een hoofdstuk over partiële rangcorrelatie, een zeer interessant onderwerp, waarin men echter nog niet ver gevorderd is. Vervolgens behandelt K e n d a l l het verband tussen de rangcorrelatie en de normale correlatie in het geval van een tweedimensionale normale verdeling, terwijl de laatste twee hoofdstukken gewijd zijn aan de theorie der paarsgewijze vergelijking van objecten, die speciaal voor de psychologie van groot belang is. Een aantal voor de toepassing nodige tabellen is aan het eind van het boek opgenomen, tezamen met een uitgebreide litteratuuropgave. Het is jammer, dat bij deze tabellen de tabel van G. P. S i l l i t o voor de exacte verdeling van K e n d a l l's τ bij het optreden van gelijken onder de waarnemingen niet is opgenomen.

Het boekje vormt een uitstekende samenvatting van de belangrijkste op dit gebied behaalde resultaten, die tot nog toe slechts zeer verspreid in de litteratuur te vinden waren. Het zal er zeker toe bijdragen het gebruik en de ontwikkeling van deze methoden te bevorderen. Om de stof ook voor minder wiskundig geschoolden toegankelijk te maken, heeft K e n d a l l de bewijzen van de stellingen, die hij gebruikt om de methoden uiteen te zetten, in aparte hoofdstukken samengevat, een methode die aanbeveling verdient. Een groot aantal voorbeelden tussen de tekst heeft ten doel de berekeningsmethoden aanschouwelijk te maken. Dit alles heeft, evenals de onderhoudende stijl, waarin het boek geschreven is, ten doel de leesbaarheid te bevorderen. In zeker opzicht schiet de schrijver echter hierin zijn doel voorbij. Want, hoewel bij eerste lezing al deze middelen aan hun doel beantwoorden, is het, indien men naar het boekje grijpt, om één van de daarin beschreven methoden toe te passen, vaak moeilijk, snel te vinden, welk gedeelte men nodig heeft. Door de verhalende stijl wordt men veelal gedwongen grote stukken opnieuw te lezen, voordat men weer voldoende in de stof is ingewerkt, om aan de slag te gaan.

De wiskundige achtergrond van de theorie is, in tegenstelling tot de practische toepassingen, geenszins eenvoudig. De hoofdstukken, waarin de gebruikte stellingen worden bewezen, zijn dan ook veruit de moeilijkste, ondanks het feit, dat de schrijver erin geslaagd is alle bewijzen met elementaire middelen te voeren. De duidelijkheid van de bewijzen laat hier en daar te wensen over; men kan ze niet lezen, maar is genoodzaakt ze geheel te reconstrueren, wil men een duidelijk begrip van de juistheid van de stelling verkrijgen. Een typisch begeleidend verschijnsel is, dat de meeste van het kleine aantal drukfouten, dat aan de correctie is ontsnapt, in deze bewijzen te vinden zijn.

Alles tezamen genomen is deze monographie van een niet te onderschatten waarde. Het ware te wensen, dat meer dergelijke monographieën werden geschreven over onderwerpen, die zich daartoe lenen (en die zijn er genoeg); en desnoods wat minder ,,Inleidingen tot de Mathematische Statistiek'', die tegenwoordig als paddestoelen uit de grond rijzen. De toepassingsmogelijkheden van methoden, zoals die der rangcorrelatie, worden pas goed duidelijk, indien de theorie zo vlot toegankelijk is gemaakt, dat men bij allerlei problemen gaat overwegen, of men deze theorie niet zou kunnen toepassen. Het lijdt daarom geen twijfel, dat de schrijver zal slagen in zijn bedoeling, deze methoden ook op ander gebied dan de psychologie (die de stoot tot hun ontwikkeling heeft gegeven) ingang te doen vinden.

J. H e m e l r i j k