MATHEMATISCH CENTRUM
AMSTERDAM

Januari 1952.

Statistische Afdeling
Rapport S 75
door
Dr J.Hemelrijk Jr
en Mej. C.van Eeden.

De vergelijking van twee methoden ter bepaling van het histaminegehalte van bloed.

1. Inleiding.

1.1. Het probleem. De bepaling van het histaminegehalte van het bloed berust op een biologische methode, waarbij gebruik gemaakt wordt van metingen van de contractie van een stukje caviadarm. Bij de statistische verwerking van waarnemingsmateriaal, dat op deze wijze verkregen was bleek, dat de nauwkeurigheid van deze bepalingen geenszins constant is, terwijl er vermoedelijk subjectieve factoren in het geding zijn, die de betrouwbaarheid der bepalingen ongunstig beïnvloeden. In het materiaal werden tegenstrijdigheden gevonden, die misschien aan deze zwakheden in de bepalingstechniek te wijten zijn. Het niet constant zijn van de nauwkeurigheid veroorzaakt bovendien grote moeilijkheden bij de statistische analyse: de voor de toepassing van klassieke methoden, zoals de variantie-analyse, noodzakelijke voorwaarden zijn niet vervuld. In overleg met de onderzoekers werd daarom door de Statistische Afdeling van het Mathematisch Centrum een wijziging in de bepalingstechniek voorgesteld, die erop gericht was de boven beschreven moeilijkheden althans voor toekomstige waarnemingsreeksen op te heffen. Daar er nog steeds proeven genomen worden, was het van belang zo snel mogelijk na te gaan, of deze nieuwe methode aan de verwachtingen voldoet, terwijl bovendien vergelijking met de gangbare methdde van belang was. Daartoe werd door de onderzoekers een proefexperiment uitgevoerd, dat in § 3 beschreven is. Het doel van dit onderzoek is dus, na te gaan of het wenselijk is bij toekomstige proeven de voorgestelde wijziging in de bepalingen aan te brengen of niet. Verder worden in dit rapport enkele mogelijkheden van toepassing van de nieuwe methode aangegeven.

2. De beide methoden.

2.1. De gangbare methode. Bij de bepaling van het histaminegehalte van bloed wordt gebruik gemaakt van een stukje caviadarm, dat contraheert onder invloed van histamine. De contracties onder invloed van de vloeistof met onbekend gehalte (kortweg de "onbekende vloeistof" of de "onbekende" genoemd) worden vergeleken met die onder invloed van bekende histamineoplossingen. De contracties worden geregistreerd op een draaiende trommel. Vóór de eigenlijke bepaling geschiedt, ondergaat het bloed verschillende bewerkingen (tezamen de "opwerking" van het bloed genoemd), die we hier niet zullen beschrijven[1]. Bij de eigenlijke bepaling registreert men de contracties onder invloed van enige bekende histamineoplossingen en daartussendoor de contracties onder invloed van de onbekende vloeistof. Men krijgt dan b.v. iets als figuur 1:

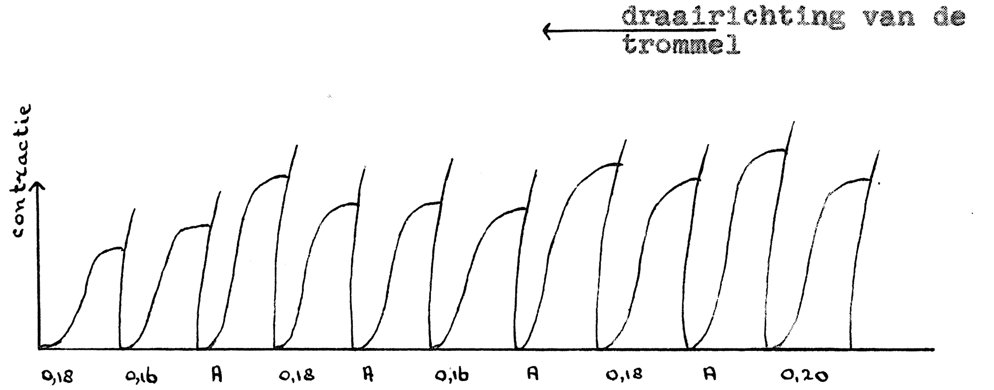


fig. 1.
Contracties bij een histamine-bepaling.

De getallen onder de getekende contractiekromme geven de concentratie van de toegevoegde bekende vloeistof aan, terwijl de onbekende door A aangegeven is. Na iedere contractie wordt de trommel stilgezet; deze tijdstippen zijn te herkennen aan de tot boven de contractiekromme doorlopende verticale strepen.

Uit een dergelijke figuur, waarbij de onbekende niet meer dan vier maal gebruikt kan worden wegens de geringe hoeveelheid, die hiervan beschikbaar is, geeft men een visueel bepaalde schatting van het histaminegehalte van A.

---

1) Voor een uitvoerige beschrijving verwijzen wij naar: J.A.H.Gooszen: Onderzoekingen over de histaminestofwisseling bij schizophrenen, Diss., Utrecht 1951, Hoofdstuk V en Th.Strengers: Een vergelijkend onderzoek over de histaminespiegel van het bloed bij mens, paard en rund, Diss., Utrecht 1946, Hoofdstuk III.

Bovenbeschreven bepalingen worden o.a. verricht om te onderzoeken of het histaminegehalte van het bloed snelle wisselingen vertoont en of het histaminegehalte stijgt onder invloed van arbeid.

2.2. Nadelen van deze methode. Bij de statistische verwerking doen zich enige moeilijkheden voor:

1. Het is bekend, dat de nauwkeurigheid der bepalingen afhankelijk is van de caviadarm, daar deze soms regelmatig contraheert en bij ontspanning weer op het oorspronkelijke niveau terugkeert, maar zich soms ook zeer onregelmatig gedraagt. Hierdoor wordt het gebruik van vele klassieke analyse-methoden onverantwoord, terwijl bovendien de spreiding van een aantal bepalingen van het histaminegehalte van verschillende sera weinig zegt over de spreiding van de te bepalen gehalten van deze sera.

2. Het gebeurt vrij vaak dat een bepaling mislukt, waardoor men in een variantie-analyse of analoog schema open plaatsen krijgt, hetgeen de statistische analyse eveneens moeilijker en minder efficiënt maakt.

3. In deze bepaling schuilt een subjectief element, gelegen in de keuze der bekende gehalten en in de visuele schattingsmethode en waarvan de invloed niet gemakkelijk na te gaan is. Deze invloed kan bovendien van bepaling tot bepaling sterk wisselend zijn.

Door het Mathematisch Centrum werd daarom een wijziging in de bepalingsmethode voorgesteld, die in § 2.3 besproken zal worden.

2.3. De nieuwe methode[2]. De wijzigingen, die worden voorgesteld, zijn er in hoofdzaak op gericht zgn. "ranginvariante" methoden - d.w.z. methoden, waarbij alleen gebruik gemaakt wordt van de volgorde naar grootte der waarnemingen - op de waarnemingsuitkomsten toe te kunnen passen, zonder dat deze toepassing in gevaar wordt gebracht door subjectieve factoren in de bepaling of door de wisseling der nauwkeurigheid van de waarnemingsuitkomsten, veroorzaakt door de veranderlijke gevoeligheid van de caviadarm. Deze rangordemethoden berusten hierop, dat men één of meer rijen rangnummers beschouwt, waarvan onder aanname van de te toetsen hypothese, alle permutaties gelijke waarschijnlijkheid bezitten. Voorlopig

---

2) Deze nieuwe methode berust op statistische toetsingsmethoden, die in de aan het einde van dit verslag toegevoegde memoranda kort zijn beschreven.

zullen wij onderstellen, dat de te toetsen hypothese inhoudt, dat een aantal (n) bloedmonsters alle hetzelfde histaminegehalte hebben. Deze hypothese is b.v. adaequaat indien men van één persoon snel na elkaar een aantal bloedmonsters neemt, om te onderzoeken of het histaminegehalte al dan niet onder invloed van arbeid of andere omstandigheden, aan snelle wisselingen onderhevig is. Verderop zullen wij beschrijven, hoe ook andere dan de hier genoemde hypothesen met de nieuwe methode kunnen worden getoetst (zie § 4.5).

Wij dienen er nu dus zorg voor te dragen, dat àls de te toetsen hypothese vervuld is, dus als alle bloedmonsters hetzelfde histaminegehalte bezitten, alle permutaties van de rijen waarnemingsuitkomsten, die wij verkrijgen, gelijke waarschijnlijkheid hebben, ook als de gevoeligheid van de caviadarm wisselt. Tevens moet de invloed van eventueel aanwezige subjectieve elementen in de bepaling de gelijkheid van deze waarschijnlijkheden niet verstoren. Voor de toetsingsmethoden, die dan kunnen worden toegepast, is het voldoende, de uitkomsten in de vorm van rijen rangnummers naar opklimmende grootte te geven; quantitatieve bepalingen zijn overbodig. Daar men er echter vaak toch op gesteld is, deze laatste te bezitten, wordt tevens beschreven, hoe deze ook bij de nieuwe methode verkregen kunnen worden. Wij wijzen er echter nu reeds op, dat het subjectieve element in deze numerieke schattingen niet vermeden kan worden, terwijl dit bij de volgordebepaling wel vrijwel volledig mogelijk is.

De volgende methode wordt nu voorgesteld. Loot, met behulp van een "aselecte getallenrij" (zgn. tabel van "random numbers"[3]) een volgorde der getallen $1,\dots,n$ (er zijn n opgewerkte bloedmonsters, die wij met $M_1,\dots,M_n$ aangeven). Zet de n buisjes met bloedmonsters in de gelote volgorde gereed (wij zullen deze volgorde aangeven als $i_1, i_2,\dots, i_n$, zodat dus de monsters in de volgorde $M_{i_1}, M_{i_2},\dots, M_{i_n}$ op een rijtje staan). Nu wordt een reeks contracties gemeten die begint met 4 x een bekende oplossing B_1, dan één maal M_{i_1}, vervolgens

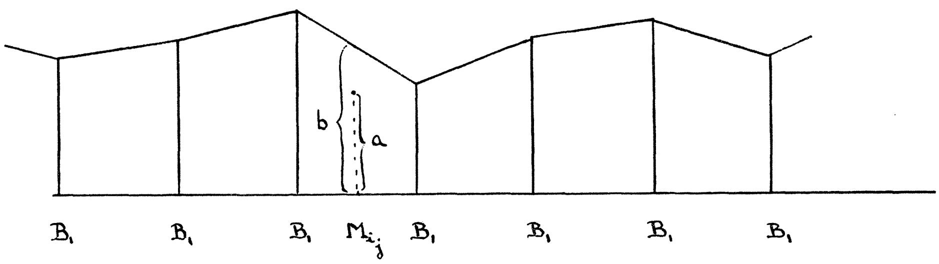
---

3) B.v. M.G.Kendall and B.Babington Smith, Tables of random sampling numbers, Tracts for Computers XXIV, Cambridge 1939.

B_1 weer, M_{i_2}, B_1, etc. Dus:

B_1, B_1, B_1, B_1, M_{i_1}, B_1, M_{i_2}, B_1,....., M_{i_n}, B_1, B_1, B_1, B_1.

Hierbij wordt steeds dezelfde bekende oplossing gebruikt. Daar de gevoeligheid van de caviadarm niet constant is, zullen de contracties bij B_1 verschillend zijn. Om nu uit de rij der verkregen contracties een volgorde naar grootte der waarnemingsuitkomsten behorend bij de onbekende monsters te verkrijgen, kan men b.v. de grootte der contracties op gelijke afstanden naast elkaar uitzetten (zie fig. 2) en de punten, die bij de bekende oplossing horen door een gebroken lijn met elkaar verbinden. De relatieve hoogte der "onbekende" punten ten opzichte van deze lijn bepaalt dan een volgorde naar grootte.



<u>fig. 2.</u> Relatieve hoogte = $\frac{a}{b}$

Volgordeschatting volgens de nieuwe methode.

Daar nu vooraf door loting de volgorde $i_1, \ldots, i_n$ van de monsters is bepaald, is voldaan aan de voorwaarde, dat alle permutaties der bij $M_1, \ldots, M_n$ gevonden rangnummers gelijke waarschijnlijkheden bezitten, indien althans aan de te toetsen hypothese voldaan is, die inhoudt dat de gehalten van alle monsters gelijk zijn.

Daar uit ieder monster 4 maal de hoeveelheid genomen kan worden, die voor een contractie-proef wordt gebruikt, kunnen wij de beschreven rij bepalingen nog 3 maal herhalen, waarbij bij iedere reeks opnieuw door loting de volgorde van de bepalingen der onbekenden gekozen wordt. Dit behoeft <u>niet</u> met dezelfde caviadarm te gebeuren en evenmin met dezelfde bekende. Wel is het wenselijk, in iedere reeks bepalingen slechts met één bekende en met één caviadarm te werken. Door 4 bekenden $B_1, \ldots, B_4$ te gebruiken voor de 4 reeksen waarnemingen, kan men bereiken dat van ieder der monsters apart een numerieke schatting van het gehalte verkregen kan worden, die op een analoge wijze visueel bepaald wordt, als bij de in § 2.1 beschreven methode het geval is. Hiertoe licht men uit ieder der 4 contractie-reeksen de contractie der beschouwde onbekende en der direct voorafgaande en volgende bekenden. Zet men deze 4 drietallen naast elkaar (zie b.v. fig. 3), dan is het zeer wel mogelijk, al is dit voor de statistische verwerking overbodig, een numerieke schatting van het onbekende gehalte te geven met een precisie, die ongeveer overeenkomt met de bij de vroegere methode bereikte.

De keuze van de eerste bekende oplossing kan wellicht moeilijkheden veroorzaken. Deze keuze zou echter gemakkelijker gemaakt kunnen worden door het afnemen van één extra bloedmonster, waarvan het gehalte dan op de oude wijze zou kunnen worden geschat. Indien de proef betrekking heeft op een aantal proefpersonen kan men hiervoor wellicht een gemengd monster van een aantal van deze personen gebruiken.

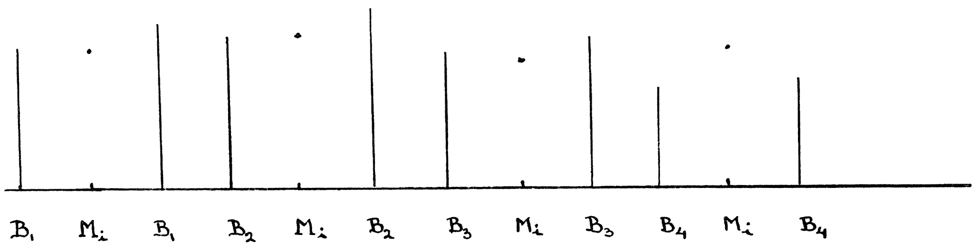


fig. 3. Numerieke schatting volgens de nieuwe methode.

2.4. Voordelen boven de oude methode. De voordelen die door deze wijzigingen verkregen kunnen worden, indien de methode behoorlijk uitvoerbaar blijkt te zijn, zijn oa.:

1. De onderstellingen, nodig voor een mathematisch verantwoorde statistische analyse, zijn nu vervuld. De conclusies kunnen dus met veel minder voorbehoud worden gegeven, dan bij de oude methode noodzakelijk is.

2. Het aantal gegevens, dat aan een reeks monsters ontleend wordt, is aanzienlijk groter dan bij de oude methode. In plaats van één rij getallen, waarop men op zichzelf slechts zeer weinig toetsen toe kan passen, beschikt men over 4 rijen rangnummers (en eventueel ook nog een rij getal-schattingen), die voor verschillende toetsingen geschikt zijn. Zo kan men b.v. reeds met behulp van één reeks bloedmonsters toetsen of er verschil in gehalte tussen deze monsters is, hetgeen in het algemeen niet mogelijk is, indien men uit ieder monster slechts één getal afleidt.

3. Het verschil in nauwkeurigheid veroorzaakt door het verschillende gedrag van caviadarm is niet hinderlijk meer voor de statistische verwerking. Ook het subjectieve element in de waarnemingen is sterk verminderd. Dit is in iedere rij rangnummers voor alle bloedmonsters gelijk (n.l. de keuze van de bekende) en brengt daarom de geldigheid van de statistische toetsingen niet meer in gevaar. Het blijft echter aanwezig in de numerieke schattingen.

Opmerkingen. De vergelijkbaarheid van door verschillende onderzoekers of op verschillende tijden verrichte quantitatieve uitkomsten voor de gehalten is bij deze opzet niet beter dan bij de oude. Men zal er naar moeten streven experimenten op te zetten, die gericht zijn op de beantwoording van een zo precies mogelijk gestelde vraag.

3. Vergelijking der twee methoden.

3.1. Beschrijving van de proef. Om de uitkomsten der twee methoden te vergelijken werd de volgende proef uitgevoerd:

Een hoeveelheid bloed werd in twaalf gelijke delen verdeeld. Aan deze twaalf bloedmonsters werden verschillende bekende hoeveelheden histamine toegevoegd, waarna ieder monster in twee gelijke delen werd verdeeld; vervolgens vond de opmerking plaats.

Op deze wijze werden twee parallel lopende series bloedmonsters verkregen, waarvan de ene volgens de oude methode werd behandeld (curve 3[4]) en de andere volgens de nieuwe methode (curve 4). Als bekende histamineoplossingen werden bij curve 4 de volgende gehalten gebruikt: 0,20 γ , 0,14 γ , 0,24 γ en 0,30 γ . De hypothese, die getoetst (en verworpen) moet worden, is in dit geval precies de boven beschrevene, inhoudende dat de gehalten der bloedmonsters gelijk zijn.

Bij beide reeksen werd allereerst door zes verschillende personen (assistenten en medewerkers van het Mathematisch Centrum), onafhankelijk van elkaar en zonder enige kennis van de toegevoegde hoeveelheden histamine een visuele schatting van het histaminegehalte der bloedmonsters gegeven. Op deze uitkomsten werd in beide geval-

---

4) Het woord curve wordt gebruikt omdat de waarnemingen in de vorm van een contractiecurve, zoals in fig. 1 getekend, worden verkregen.

len de methode der m rangschikkingen (zie bijlage S 47 (M 14)) toegepast, om na te gaan of de schattingen van deze zes personen behoorlijk overeenstemden voor zover het de volgorde van grootte der gehalten betreft. Deze overeenstemming bleek zeer goed te zijn. Hier komen we in § 4.1 op terug.
Van curve 3 waren bovendien schattingen beschikbaar, die door de onderzoekers zelf waren verricht.

We voeren nu de volgende notatie in:

I : Curve 3 volgens oude methode (schattingen der onderzoekers zelf).

I': Curve 3 volgens de methode der m rangschikkingen (schattingen van het Mathematisch Centrum met m=6).

II^A: Curve 4 volgens nieuwe methode (dus volgens de in § 2.3 beschreven methode met 4 rangschikkingen).

II^B: Curve 4 volgens de methode der m ranschikkingen (schattingen van het Mathematisch Centrum, m=6).

Hiervan is dus alleen de reeks onder II^A geheel vrij van subjectieve factoren.

3.2 <u>Resultaten van de proef</u>. De uitkomsten van I vinden wij in tabel I.

Tabel I

Curve 3 volgens oude methode (I)

monster no.	histamine toevoeging	hist. geh.	volgorde
2	0,00 γ	4,02 γ	3
4	0,00 γ	3,26 γ	1
6	0,01 γ	4,27 γ	4
8	0,02 γ	4,79 γ	9
10	0,03 γ	3,88 γ	2
12	0,04 γ	mislukt	-
14	0,05 γ	4,3 γ	5
16	0,06 γ	4,68 γ	8
18	0,07 γ	4,46 γ	7
20	0,08 γ	4,42 γ	6
22	0,09 γ	5,32 γ	10
24	0,10 γ	5,46 γ	11

Tabel II geeft de met de methode der m rangschikkingen bij I' gevonden kolomtotalen en tevens de uit deze kolomtotalen volgende volgorde in grootte van het histaminegehalte der bloedmonsters:

Tabel II

Curve 3 volgens 6 rangschikkingen methode (I')

no.	2	4	6	8	10	14	16	18	20	22	24
kolomtotalen	36	10,5	31,5	53,5	8,5	31,5	18,0	38,5	43,0	65,5	59,5
volgorde	6	2	4,5	9	1	4,5	3	7	8	11	10

Bij II^A werd op de in § 2.3 beschreven manier voor ieder der vier reeksen een volgorde naar grootte voor de bloedmonsters bepaald. Op deze 4 rijen werd de methode der m ranschikkingen toegepast, met m=4.

In tabel III vinden wij de voor iedere reeks gevonden volgorde, de kolomtotalen en de uit de kolomtotalen volgende volgorde naar gehalte der bloedmonsters.

Tabel III

Curve 4 volgens nieuwe methode (II^A)

reeks \ no.	1	3	5	7	9	11	13	15	17	19	23
1	6	11	3,5	1	5	3,5	7	2	10	9	8
2	6	8	3	5	4	1	7	9	11	10	2
3	6	4	3	10	2	1	8	5	11	9	7
4	6	5	2	1	4	7	9	3	8	11	10
kolomtotalen	24	28	11,5	17	15	12,5	31	19	40	39	27
volgorde	6	8	1	4	3	2	9	5	11	10	7

Daar in de eerste reeks no. 21 ontbreekt hebben wij in tabel III in alle reeksen no. 21 weggelaten. Geven wij in de eerste reeks no. 21 als rangnummer het gemiddelde van de rangnummers uit de drie andere reeksen dan vinden wij:

Tabel IV

Curve 4 volgens nieuwe methode, met no. 21 ($II^{A'}$)

no.	1	3	5	7	9	11	13	15	17	19	21	23
kolomtotalen	24	29	11,5	18	15	12,5	31	19	42	42	40	28
volgorde	6	8	1	4	3	2	9	5	11,5	11,5	10	7

Bij II^B werd evenals bij I' de methode der m rangschikkingen toegepast op de door zes personen geschatte waarden voor het histaminegehalte van de twaalf bloedmonsters. De resultaten (kolomtotalen en volgorde) staan vermeld in tabel V:

Tabel V

Curve 4 volgens 6 rangschikkingen methode (II^B)

no.	1	3	5	7	9	11	13	15	17	19	21	23
kolomtotalen	38	23,5	17	32,5	7,5	32,5	40,5	25,5	64	69	62,5	55,5
volgorde	7	3	2	5,5	1	5,5	8	4	11	12	10	9

In tabel VI vinden wij een overzicht van de bij I, I', II^A, $II^{A'}$ en II^B gevonden volgorden naar grootte van het histaminegehalte der bloedmonsters.

Tabel VI

Overzicht van de gevonden volgorden

no.	I	I'	no.	II^A	$II^{A'}$	II^B
2	3	6	1	6	6	7
4	1	2	3	8	8	3
6	4	4,5	5	1	1	2
8	9	9	7	4	4	5,5
10	2	1	9	3	3	1
12	-	-	11	2	2	5,5
14	5	4,5	13	9	9	8
16	8	3	15	5	5	4
18	7	7	17	11	11,5	11
20	6	8	19	10	11,5	12
22	10	11	21	-	10	10
24	11	10	23	7	7	9

Opmerking: Door vergelijking van de kolommen II^A en $II^{A'}$ blijkt dat het opvullen van één enkele lege plaats in het schema van 4 rangschikkingen, zoals dat hier is uitgevoerd de gevonden volgorde slechts zwak heeft beïnvloed.

4. Statistische analyse der resultaten.

4.1. Overeenstemming der schattingen van verschillende waarnemers.

Zoals in § 3.1 werd opgemerkt hebben wij bij I' en II^B de methode der m rangschikkingen toegepast op de door zes personen gegeven schattingen van het histaminegehalte der bloedmonsters om na te gaan of de schattingen van deze zes personen behoorlijk overeenstemmen wat betreft de volgorde naar grootte der gehalten. In de tabellen II en V gaven wij de hierbij gevonden kolomtotalen. De overschrijdingskansen, die bij de methode der m rangschikkingen gevonden zijn vinden wij in tabel VII.

Tabel VII

Overschrijdingskansen[5] gevonden met de methode der m rangschikkingen bij I' en II^B

	overschrijdingskans
I'	< 0,0001
II^B	< 0,0001

Hieruit zien wij dus dat er een zeer duidelijke overeenstemming is tussen de door de zes personen gegeven schattingen. Het subjectieve element in de schatting uit contractiecurve is dus niet zeer storend. De invloed daarvan is blijkbaar klein in vergelijking met de verschillen tussen de monsters. Dit zegt echter niet over de invloed van de keuze der bekenden bij iedere onbekende, en evenmin over de invloed van de wisselende gevoeligheid van de caviadarm.

Om te onderzoeken of de door de onderzoekers zelf gegeven schattingen overeenstemmen met de schattingen van het Mathematisch Centrum (wat betreft de volgorde naar grootte), zijn deze twee volgorden (I en de bij I' uit de kolomtotalen volgende volgorde) vergeleken met behulp van de methode der rangcorrelatie (zie bijlage S 47 (M 13)). In tabel VIII vinden wij de hierbij gevonden waarde van S en de overschrijdingskans.

Tabel VIII

Vergelijking van I en I' met de rangcorrelatie-methode.

n	S	overschrijdingskans
11	34	0,009

---

5) De in dit rapport opgegeven overschrijdingskansen zijn tweezijdig, indien niet uitdrukkelijk vermeld staat, dat zij éénzijdig zijn. In de hier voorkomende gevallen is, indien tweezijdige toetsing mogelijk is, de tweezijdige overschrijdingkans twee maal zo groot als de éénzijdige.

Uit tabel VIII zien wij dus dat de bij I en I' gevonden volgorden zeer goed overeenstemmen. Daar de reeks I' beter dan I vergelijkbaar is met II^B, daar I' en II^B visuele schattingen door dezelfde 6 onderzoekers zijn, hebben wij verder met reeks I' gewerkt in plaats van met I.

4.2. Toetsen tegen trend. Aan de bloedmonsters zijn in de volgorde waarin zij in tabel VI staan toenemende hoeveelheden histamine toegevoegd (zie ook tabel I). Wij wensen nu na te gaan of deze stijging ook in de resultaten terug te vinden is en of dit met de beide methoden even goed gaat of niet. Daartoe staan ons een aantal toetsingsmethoden ter beschikking, die ieder hun voor- en nadelen bezitten en die hieronder staan opgenoemd. Dergelijke toetsen noemen wij toetsen tegen trend (trend = stijgende of dalende tendentie).

α. Op de reeksen II^A en $II^{A'}$ kon in dit geval, daar bekend is, dat het om het ontdekken van een stijging gaat, de uitsluitend voor stijging of daling apart geschikte, door Ph.van Elteren op grond van de methode der m rangschikkingen ontwikkelde trend-toets met m+1 rangschikkingen worden toegepast, die in bijlage S 73 (M 14a) beschreven is. Deze toets blijkt voor het onderhavige geval zeer geschikt te zijn, zoals uit de zeer kleine overschrijdingskans, die hiermee gevonden wordt, blijkt (zie tabel IX). Voor toepassing hiervan dient bovendien echter nog de voorwaarde vervuld te zijn, dat er in het schema der m rangschikkingen (waaraan als $(m+1)^e$ rangschikking de rij 1,2,3,... wordt toegevoegd) geen open plaatsen voorkomen. Bij toepassing op II^A werd daarom het bloedmonster no. 21 buiten beschouwing gelaten.

β. Op II^A kan ook een andere trend-toets worden toegepast, afkomstig van G.Elfring en J.H.Whitloch die op het combineren van een aantal enkelvoudige rangcorrelaties berust en die in bijlage S 73 (M 13a) beschreven is. Deze bezit het voordeel ook toegepast te kunnen worden, indien niet bekend is of de eventueel aanwezige trend een stijging dan wel een daling is. Dit wordt vaak uitgedrukt door te zeggen, dat hier tweezijdige toetsing mogelijk is. Bovendien zijn open plaatsen in het schema nu voor de toepassing niet hinderlijk, zodat het bloedmonster no. 21 nu niet weggelaten behoeft te worden. Wij vermoeden echter, dat het onderscheidingsvermogen van deze toets iets geringer is dan van de vorige. De (éénzijdige) overschrijdingskans, vermeld in tabel IX onder II^A (rangcorrelatie) wijst ook in deze richting.

γ. Op de reeksen I' en II^B konden wij slechts een enkelvoudige rangcorrelatie met de natuurlijke getallen 1,2,3... toepassen, daar in deze beide gevallen slechts één reeks getallen beschikbaar is. Deze rangcorrelatiemethode staat (voor 2 willekeurige rijen getallen, waarvan er hier dus één door 1,2,3... vervangen wordt) beschreven in bijlage S 47 (M 13). Deze toets, die evenals de vorige ook tweezijdig kan worden toegepast, zou ook met de kolomtotalen van II^A en $II^{A'}$ kunnen worden uitgevoerd. Dit hebben wij echter nagelaten, daar het onderscheidingsvermogen veel geringer is dan van de twee hierboven beschreven toetsen. Daar wij voor I' en II^B geen andere mogelijkheid bezitten, waren wij wel gedwongen het voor deze twee reeksen bij deze toets te laten. De in tabel IX vermelde éénzijdige overschrijdingskansen zijn dan ook veel groter dan bij de twee andere reeksen.

Tabel IX

Toepassing van trend-toetsen op I', II^B, $II^{A'}$ en II^A
(éénzijdige overschrijdingskansen)

reeks	methode	n	overschrijdingskans
I'	γ	11	0,035
II^B	γ	12	0,025
$II^{A'}$	α	12	<0,0001
II^A	α	11	<0,0001
II^A	β	12	0,002

Conclusies:

1. De kunstmatig veroorzaakte stijging wordt in alle gevallen bij éénzijdige toetsing teruggevonden. Bij de nieuwe methode (reeksen II^A en $II^{A'}$) echter veel overtuigender dan bij de oude (I'). Ook de numerieke schattingen volgens de nieuwe methode (II^B) voldoen minder goed dan de op II^A en $II^{A'}$ toegepaste methoden.

2. Bij vele experimenten, waarbij geen kunstmatige stijging is aangebracht, maar waar het erom gaat een eventuele trend te ontdekken (b.v. bij arbeidscurven of bij opeenvolgende bloedafnemingen onder veranderende omstandigheden) zal het voorkomen dat de richting van de trend, die op kan treden, niet van te voren bekend is. De toetsing dient dan tweezijdig te geschieden,

zodat de onder α genoemde methode niet toegepast kan worden. Indien de onderhavige gegevens van een dergelijk experiment afkomstig waren geweest, zouden dus de derde en de vierde regel uit tabel IX vervallen, terwijl de overschrijdingskansen in de overblijvende regels verdubbeld zouden moeten worden. In dat geval is de uitkomst bij II^A nog duidelijk (overschrijdingskans 0,004), maar bij I' en II^B onduidelijk (overschrijdingskansen 0,07 en 0,05). Ook in dergelijke gevallen mag men dus verwachten, dat de rangordemethoden, waarvoor de nieuwe methode is opgezet, goed zullen voldoen.

3. De bij II^A en $II^{A'}$ gevonden overschrijdingskansen bij methode α zijn ongeveer gelijk. Hier blijkt dus opnieuw, dat het invullen van de ene open plaats in het schema der 4 rangschikkingen geen merkbare invloed op het resultaat heeft gehad. Bij meer open plaatsen zou dit echter wel verwacht kunnen worden, zodat dan methode β, ook bij éénzijdige toetsing, te prefereren is.

4.3. Overeenstemming tussen I', $II^{A'}$ en II^B. De overeenstemming tussen I', $II^{A'}$ en II^B werd onderzocht door ze twee aan twee te vergelijken met de methode der rangcorrelatie. De reeks II^A werd hier niet gebruikt, omdat in deze reeks weer een andere waarneming ontbreekt dan in I'. Het aantal waarnemingen wordt daardoor nog kleiner dan het al is. Daar tegengesteld gedrag van de te vergelijken methoden, hetgeen met negatieve correlatie overeen zou komen, volslagen onaannemelijk geacht moet worden, is het verantwoord in dit geval met éénzijdige toetsing te volstaan. De hierbij gevonden waarden van S en de éénzijdige overschrijdingskansen vinden we in tabel X.

Tabel X

Overeenstemming tussen I', $II^{A'}$ en II^B
(éénzijdige overschrijdingskansen)

	n	S	overschrijdingskans
I' - $II^{A'}$	11	17	0,11
I' - II^B	11	26	0,025
$II^{A'}$ - II^B	12	44	0,001

Hieruit zien we dus:

1. dat er een duidelijke overeenstemming is tussen $II^{A'}$ en II^B.

2. dat er geen aanwijzing voor overeenstemming tussen I' en $II^{A'}$ gevonden wordt.
3. dat er een duidelijke aanwijzing is voor een overeenstemming tussen I' en II^{B}.

Opmerking. Het gebrek aan overeenstemming tussen I' en $II^{A'}$, dat wellicht vreemd zal lijken, kan verschillende oorzaken hebben. Ten eerste is de extra histamine vóór de opwerking der bloedmonsters toegevoegd, zodat (volgens inlichtingen van de onderzoekers) een gedeelte daarvan verloren kan gaan en vermoedelijk ook inderdaad verloren gaat. Hierdoor is het niet zeker, dat de volgorde der gehalten na de opwerking in de reeksen I en II geheel dezelfde is, ook al zal de kunstmatig aangebrachte stijgende tendentie niet geheel verloren gaan (zoals bij reeks II overtuigend is aangetoond). Op dit punt komen wij bij de opmerkingen in de laatste paragraaf nog terug. Ten tweede schuilt in de eerste bepalingsreeks het subjectieve element van de keuze der bekende oplossingen, die bij de verschillende praeparaten verschillend kan zijn. De invloed van deze factor is uit de onderhavige gegevens niet na te gaan. Ten derde is het aantal waarnemingen in de twee vergeleken reeksen slechts klein (n.l. 11) en bezit de toegepaste toets geen groot onderscheidingsvermogen, zodat overeenstemming slechts gevonden zal worden indien deze sterk is. Het feit, dat tussen I' en II^{B} en tussen $II^{A'}$ en II^{B} wel een overeenstemming gevonden wordt, doet bovendien vermoeden dat aan de grote overschrijdingskans o,11 in dit geval niet al te veel betekenis moet worden gehecht.

4.4. Toepassing van de methode der m rangschikkingen op II^{A}.

Eén van de doelstellingen van het ontwerpen van de nieuwe methode is de mogelijkheid te scheppen om te onderzoeken of er snelle wisselingen in het histaminegehalte van bepaalde personen optreedt, ook indien de omstandigheden bij de reeks bloedafnemingen van die aard is, dat er geen trend te verwachten is.

Indien wij de bloedmonsters, die volgens de nieuwe methode behandeld zijn, in een willekeurige volgorde zetten, hebben wij een reeks verkregen, waarbinnen inderdaad verschillen in gehalte bestaan en die dus aan een dergelijk experiment beantwoord. Wij hebben daarom nagegaan of deze wisselingen nu te vinden zijn. Daarbij kan dan dus geen trend-toets worden toegepast, maar wel kan de gewone methode der m rangschikkingen (zie S 47 (M 14); in dit geval met m=4, daar de in § 2.3 beschreven methode ons 4 rangschikkingen geeft) op reeks II^A (of $II^{A'}$) worden toegepast. Hiervoor kan tabel III worden gebruikt. De kolommen van deze tabel zouden dan in een willekeurige volgorde moeten worden gezet, maar dit maakt voor de uitkomst der m rangschikkingen-toets geen verschil, zodat dit ook nagelaten kan worden. De bij tabel III behorende overschrijdingskans blijkt $< 0{,}01$ te zijn, terwijl $II^{A'}$ een nog kleinere overschrijdingskans geeft. Wij zien dus, dat ook als er niet naar een trend gezocht wordt, in materiaal van deze aard een verschil tussen de bloedmonsters te ontdekken is. De mogelijkheid dit te onderzoeken bij één reeks monsters ontbreekt bij de oude methode geheel.

4.5. <u>Toepassing van de nieuwe methode op experimenten van andere aard.</u>

Wij vermelden ten slotte nog de wijze, waarop deze methode op experimenten van andere aard kan worden toegepast. Deze toepassing berust daarop, dat de kolomtotalen van de in § 4.4 beschreven methode der m rangschikkingen een schatting vormen van de volgorde naar grootte van de gehalten der bloedmonsters, indien deze gehalten niet gelijk zijn.

Indien men er nu b.v. op uit is een verschil in het histaminegehalte bij twee groepen A en B van personen te ontdekken, kan men als volgt te werk gaan. Men neemt op aselecte wijze ("at random") een aantal (b.v.5) personen uit ieder der groepen en van ieder van deze personen neemt men onder zo goed mogelijk gelijk gemaakte omstandigheden een bloedmonster. Deze bloedmonsters zullen nu in het algemeen niet meer hetzelfde gehalte bezitten, daar zij van verschillende personen afkomstig zijn. Zowel op de collectie gevormd door de groep A van personen (of op de collectie, waaruit deze groep een steekproef is) als op de bij B behorende collectie bezit het histaminegehalte een waarschijnlijkheidsverdeling. De hypothese, die nu getoetst moet worden luidt, dat deze twee waarschijnlijkheidsverdelingen gelijk zijn. Behandelen wij nu de reeks van 10 bloedmonsters volgens de nieuwe methode en passen wij de methode der m rangschikkingen op de 4 verkregen rangschikkingen toe, dan verkrijgen wij een rij van 10 kolomtotalen, die een schatting van de volgorde naar grootte der 10 histaminegehalten geven. Van deze kolomtotalen behoren er 5 bij groep A en 5 bij groep B. De bovengenoemde hypothese, dat de waarschijnlijkheidsverdeling van het histaminegehalte bij A en B dezelfde is, kan nu getoetst worden door de toets van Wilcoxon voor het probleem van twee steekproeven toe te passen op de twee groepen van 5 kolomtotalen, die bij A resp. B behoren. Daar steekproeven van 5 waarnemingen klein zijn, kan men de gehele proef één- of meermalen herhalen met andere vijftallen personen uit A en B en de bij deze proeven verkregen resultaten kunnen dan gemakkelijk worden gecombineerd tot één gezamenlijke toetsing van de bovengenoemde hypothese.

Indien men de oude methode gebruikt, kan men eveneens de toets van Wilcoxon toepassen. De wisselende gevoeligheid van de caviadarm en de subjectieve keuze van de bekende oplossingen verzwakken dan echter de fundering van deze toets en verminderen bovendien het onderscheidingsvermogen. Aan deze moeilijkheden wordt nu door de nieuwe methode tegemoetgekomen, doordat
1^e^ de gevoeligheidswisselingen door loting op toevallige wijze over de monsters verdeeld wordt,
2^e^ dit 4 maal geschiedt bij 1 reeks monsters, zodat deze verschillen elkaar grotendeels zullen opheffen en
3^e^ het genoemde subjectieve element voor alle monsters van een reeks gelijk is.

Het is niet moeilijk nog andere proeven te geven, waarbij de nieuwe methode van nut kan zijn. Wij zullen het echter bij het bovenstaande illustratieve voorbeeld laten.

5. Conclusies en opmerkingen.

5.1. Conclusies.

1. De stijging der histaminegehalten in de reeksen I en II is in II (met de nieuwe methode) zeer veel duidelijker aantoonbaar dan in I (oude methode); zie tabel IX.
2. De overeenstemming tussen de resultaten verkregen met de oude en de nieuwe methode is behoorlijk, maar niet volkomen (zie tabel X). In verband met het feit, dat bij de nieuwe methode met de wisselende gevoeligheid van de caviadarm en met de subjectiviteit van de keuze der bekende oplossingen rekening is gehouden, zodat deze factoren de geldigheid der toegepaste statistische analyse niet in gevaar brengen, verdienen de resultaten der nieuwe methode een groter vertrouwen dan die van de oude.

3. De overeenstemming tussen de (voor de statistische verwerking overbodige) numerieke schattingen, verkregen volgens de beide methoden, is, wat de volgorde naar grootte betreft, bevredigend. De numerieke schattingen van reeks II (nieuwe methode) stemmen wat volgorde betreft zeer bevredigend overeen met de kolomtotalen van de 4 rangschikkingen, die bij de nieuwe methode behoren (zie tabel X, laatste regel).

4. De nieuwe methode schept mogelijkheden van statistische verwerking, die bij de oude niet bestaan (zie § 4.4).

5.2. Opmerkingen.

1. Terugkomende op het gedeeltelijk verdwijnen van de toegevoegde histamine bij de opwerking, merken wij op, dat dit ook met in het bloed aanwezige histamine het geval zou kunnen zijn. Het is dus eigenlijk niet juist de gevonden numerieke schattingen zonder nader onderzoek gelijk te stellen met het histaminegehalte van het bloed. Zolang over de oorzaken van dit verdwijnen van histamine bij de opwerking niets naders bekend is, zullen zich bij het formuleren van conclusies steeds moeilijkheden blijven voordoen. Onderstel b.v., dat er in het bloed de één of andere stof S aanwezig is, die het histamine bij de opwerking beschermt, dan zou een gevonden systematisch verschil in uitkomsten bij 2 personen of groepen van personen even goed kunnen wijzen op een verschil in histaminegehalte als op een verschil van gehalte van de stof S of op beide. Er zijn verschillende proeven te bedenken om althans iets meer te weten te komen over deze kwestie. In de eerste plaats zou het verrichten van een experiment als het nu geanalyseerde, maar dan zonder toevoeging van histamine, met daarop aansluitend de in § 4.4. beschreven methode van 4 rangschikkingen, de mogelijkheid geven om te toetsen of de opwerking zelf al een merkbaar verschil tussen de uiteindelijke histaminegehalten van het bloed geeft. In de tweede plaats zou men kunnen nagaan of histamine, toegevoegd na de opwerking wel geheel wordt teruggevonden. Daarbij zou van de numerieke schattingen gebruik moeten worden gemaakt, bij voorkeur toch weer volgens de nieuwe methode of volgens beide methoden. Zo zouden gemakkelijk nog meer experimenten te bedenken zijn, die het inzicht in de waarnemingstechniek zouden kunnen verdiepen.

2. Ten slotte willen wij nogmaals wijzen op de aan het eind van § 2.2 reeds gemaakte opmerking dat er in het

vergelijken van de numerieke schattingen van verschillende onderzoekingen een gevaar schuilt. Dit gevaar wordt door de nieuwe methode niet opgeheven. Bij de opzet van een experiment kan hiermee rekening worden gehouden door van te voren de probleemstelling scherp te formuleren en de proef hierbij aan te passen. De resultaten van iedere proef apart bezitten in de boven uiteengezette statistische verwerking een zeer exacte basis voor het trekken van conclusies. Combinatie van de uitkomsten van een aantal op deze wijze opgezette experimenten van gelijksoortige aard is dan zeer wel mogelijk.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.
S47(M6)

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

De toetsing van een hypothese $\mathcal{H}_0$ berust steeds op een aantal waarnemingen x_1, x_2, ..., x_n van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

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

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat $\mathcal{H}_0$ juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese $\mathcal{H}_0$, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de <u>kritieke zône</u> van de toets, α de <u>onbetrouwbaarheidsdrempel</u> (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu $\mathcal{H}_0$ op grond van de waarnemingen x_1, x_2, ..., x_n, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "<u>significant</u>" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien $\mathcal{H}_0$ juist is, gelijk aan α. Derhalve is α de <u>kans op ten onrechte verwerping van de juiste hypothese</u>, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, <u>waarin de hypothese</u> die men toetst juist is, deze toch verwerpen.

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

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

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

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

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

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

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de overschrijdingskans k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leqq \alpha$ is.

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

Litteratuur:

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

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

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

S 47 (M 13)

RANGCORRELATIE [1]

De door M.G. Kendall ontwikkelde methode der rangcorrelatie is toepasbaar op de volgende situatie:

De stochastische grootheden x en y bezitten een simultane verdeling. Over deze verdeling zelf behoeft niets ondersteld te worden.

(x,y) $(i=1,\ldots,n)$, zijn onafhankelijk waarnemingsparen van deze stochastische grootheden.

De x_i worden gerangschikt naar opklimmende grootte en vervolgens vervangen door hun rangnummers $1,2\ldots n$. De y_i verkrijgen op die manier een rangschikking, die niet overeen behoeft te komen met die volgens opklimmende grootte en worden nu vervangen door de rangnummers, die bij deze rangschikking naar opklimmende grootte behoren.

Voorbeeld:

x_i :	0,28	0,42	1,14	1,16	1,21	1,76	1,96	2,53	3,12	3,48
y_i :	36,4	38,1	28,2	34,1	29,7	25,3	31,3	26,8	32,7	40,2

rangnummers der x_i :	1	2	3	4	5	6	7	8	9	10
" " y_i :	8	9	3	7	4	1	5	2	6	10

De rangcorrelatiecoëfficiënt τ dient als maat voor de graad van overeenkomst tussen deze twee rijen rangnummers

Deze coëfficiënt wordt zo gedefinieerd dat bij volledige overeenstemming tussen de twee rijen $\tau = +1$ en bij precies tegenovergestelde rangschikking $\tau = -1$ is.

Voor iedere andere rangschikking ligt τ tussen de grenzen $+1$ en -1.

We definiëren daarvoor eerst de grootheid P als volgt:
P is de som van het aantal paren rangnummers der (y_i, y_j) dat in de normale volgorde staat, dus waarbij het linker rangnummer het laagste is.

----

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

We berekenen P als volgt: (zie v.b.)
We beginnen met het eerste rangnummer van de rij der y_i: dit is 8 en staat vóór de hogere rangnummers 9 en 10. Het geeft dus een bijdrage 2 aan P.

Het tweede rangnummer van de rij, 9, staat vóór 10 en geeft dus een bijdrage 1 aan P.

Het derde van de rij, 3, staat vóór 7, 4, 5, 6, 10 en geeft dus een bijdrage 5 aan P, enz.

$P = 2 + 1 + 5 + 1 + 3 + 4 + 2 + 2 + 1 = 21$

P wordt maximaal als de y_i dezelfde rangorde bezitten als de x_i; dan is $P = \frac{1}{2}n(n-1)$.

We definiëren nu $\tau = \frac{2P}{\frac{1}{2}n(n-1)} - 1$.

In dit voorbeeld vinden we: $\tau = \frac{2 \times 21}{\frac{1}{2}.10.9} - 1 = -0.07$.

Voor het geval dat $\underline{x}$ en $\underline{y}$ onafhankelijk zijn is de verzameling van τ door M.G.Kendall berekend. De hypothese, dat $\underline{x}$ en $\underline{y}$ onafhankelijk zijn, kan dus getoetst worden. Bij deze toets wordt gewoonlijk niet de grootheid τ, doch

$$S = \tfrac{1}{2}n(n-1)\tau = 2P - \tfrac{1}{2}n(n-1)$$

gebruikt.

Is de hypothese van onafhankelijkheid niet vervuld, dan is de waarschijnlijkheid van grote waarden van S groter, dan wanneer dit wel het geval is. De kritieke zône is daarom van de vorm $|S| \geqq S_0$, en bij éénzijdige toetsing van de vorm $S \geqq S_0'$ (rechtszijdige toetsing) of $S \leqq S_0''$ (linkszijdige toetsing). De bij dit systeem van kritieke zônes behorende overschrijdingskansen vinden wij voor $n \leqq 10$ getabelleerd in [1] (table 1), en voor $n \leqq 40$ in [2] bijlage 1. Voor $n > 10$ maken wij gebruik van het feit, dat de verdeling van $\underline{S}$ dan goed benaderd wordt door een normale verdeling met gemiddelde 0 en spreiding:

$$\sigma_{\underline{S}} = \sqrt{\frac{1}{18}n(n-1)(2n+5)} \tag{1}$$

Het kan gebeuren, dat onder de x_i of y_i gelijke waarnemingen voorkomen. We geven deze gelijke waarnemingen dan ook gelijke rangnummers en wel op de volgende wijze: Veronderstel, dat wij aan het 2e t/m 7e rangnummer geen rangverschil kunnen toekennen dan kiezen wij als rangnummer voor ieder van deze waarnemingen het gemiddelde van deze rangnummers 2, 3,..., 7, dus

$$\frac{1}{6}(2 + 3 + 4 + 5 + 6 + 7) = 4\tfrac{1}{2}.$$

Daar het maximum van P nu kleiner wordt, moeten wij een correctie op de formule voor τ toepassen, die in [1], pag. 25 en in [2] pag. 15 beschreven staat. Ook voor dit geval is de verdeling van $\underline{S}$ bekend. Indien in één van beide rijen tweetallen op drietallen gelijken voorkomen kan men voor $n \leq 10$ gebruik maken van de tabel van Sillitto [3] . Voor $n > 10$ is ook in dit geval, indien er niet te veel series gelijken voorkomen benadering met behulp van een normale verdeling mogelijk. De spreiding $\sigma_{\underline{S}}$ moet dan echter berekend worden uit de volgende formule:

$$(2)\quad \sigma_{\underline{S}}^2 = \frac{1}{18}\left\{n(n-1)(2n+5) - \sum_t t(t-1)(2t+5) - \sum_u u(u-1)(2u+5)\right\} + \\ + \frac{1}{9n(n-1)(n-2)} \sum_t t(t-1)(t-2) \sum_u u(u-1)(u-2) + \\ + \frac{1}{2n(n-1)} \sum_t t(t-1) \sum_u u(u-1)$$

waarin n de lengte van de twee rijen aangeeft en t (resp. u) het aantal rangnummers in ieder der groepen gelijke rangnummers voorkomende in de eerste (resp. tweede) rij voorstelt.

Voorbeeld:

1^e rij:	$1\frac{1}{2}$	$1\frac{1}{2}$	$3\frac{1}{2}$	$3\frac{1}{2}$	$5\frac{1}{2}$	$5\frac{1}{2}$	7	$8\frac{1}{2}$	$8\frac{1}{2}$	10	11	12
	$2\frac{1}{2}$	$2\frac{1}{2}$	$2\frac{1}{2}$	$9\frac{1}{2}$	5	$6\frac{1}{2}$	$6\frac{1}{2}$	$9\frac{1}{2}$	$9\frac{1}{2}$	$2\frac{1}{2}$	$9\frac{1}{2}$	12

In de eerste rij komen vier paren gelijken voor (t=2), in de tweede rij 2 groepen van 4 gelijken (u=4) en 1 paar gelijken (u=2).

Dus geldt:
$$\sum t(t-1)(2t+5) = 4 \times (2.1.9) = 72$$
$$\sum u(u-1)(2u+5) = 2 \times (4.3.13) + 2.1.9 = 330$$
$$\sum t(t-1)(t-2) = 0$$
$$\sum t(t-1) = 4.2 = 8$$
$$\sum u(u-1) = 2 \times (4.3) + 2 = 26$$

Dus is:
$$\sigma_{\underline{S}}^2 = \frac{1}{18}\{2.11.29 - 72 - 330\} + \frac{8.26}{2.12.11} = 191$$
$$\sigma_{\underline{S}} = \sqrt{191} = 13{,}8$$

Litteratuur:

[1] M.G.Kendall, Rank Correlation Methods, London 1948, hoofdstuk 1.
Tabel verdelingsfunctie van $\underline{S}$ voor n = 4 t/m 10, pag. 141.

[2] J.Hemelrijk, Kendall's rangcorrelatie-coëfficiënt τ , hoofdstuk I der cursus "Parametervrije Methoden", Rapport S 59 Mathematisch Centrum (1951) Blz. 1-17.
Bijlage I: Tabel van de kleinste waarden van $\underline{S}$ waarvan de overschrijdingskansen onder de hypothese van onafhankelijkheid hoogstens gelijk zijn aan α voor α = 0,005; 0,01; 0,025; 0,05 en 0,10 en n=4,5,6...,40

[3] G.P.Sillitto: "The distribution of Kendall's coefficient of rankcorrelation in rankings containing ties, Biometrika $\underline{34}$ (1947) p. 36-40.

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

Statistische Afdeling.
S 73 (M 13a)

Trend-toets met behulp van rangcorrelatie[1)]

Men kan de methode der rangcorrelatie toepassen om na te gaan of er in een reeks waarnemingen van een stochastische grootheid $\underline{y}$ een trend, d.w.z. een stijgend of dalend verloop aanwezig is. Men maakt daarbij gebruik van het feit, dat bij een systeem van n waarnemingsparen (x_i, y_i) de grootte van S[2)] en de verdeling van $\underline{S}$ onder de hypothese van onafhankelijkheid niet verandert als men de volgorde van de paren wijzigt. Men kan dus bij de berekening van S de volgorde der paren $(x_i y_i)$ zodanig kiezen, dat de rangnummers van bijvoorbeeld de x_i de rij 1, 2, 3,..., n vormen. De hypothese van onafhankelijkheid komt dan overeen met de hypothese H_o, dat alle mogelijke permutaties van de rangnummers der y_i even waarschijnlijk zijn.

Indien wij nu willen onderzoeken of er een trend aanwezig is in de rij waarnemingen $y_1,...,y_n$ van de stochastische grootheid $\underline{y}$, dan voegen wij de rij 1, 2, ..., n toe en bepalen vervolgens S zoals aangegeven in memorandum S 47 (M 13) en toetsen bovengenoemde hypothese H_o. Onder deze hypothese heeft S dezelfde verdeling als de overeenkomstige grootheid der rangcorrelatietoets onder de hypothese van onafhankelijkheid; wij gebruiken hier ook dezelfde kritieke zônes. Indien S een positieve (resp. negatieve) waarde heeft behorende tot de kritieke zône spreken wij van een significant stijgende (resp. dalende) trend.
Voor literatuur over deze toets zie men [1] .

Deze toets kan worden uitgebreid tot meerdere onafhankelijke reeksen waarnemingen, die niet alle even groot behoeven te zijn. Bij voorbeeld:

$$
\begin{array}{l}
y_{1,1},\ y_{1,2},\ldots\ldots,\ y_{1,n_1} \\
y_{2,1},\ y_{2,2},\ldots\ldots,\ y_{2,n_2} \\
y_{3,1},\ y_{3,2},\ldots\ldots,\ y_{3,n_3} \\
\vdots \\
y_{m,1},\ y_{m,2},\ldots\ldots,\ y_{m,n_m}
\end{array}
$$

----

1) Dit memorandum is een aanvulling op memorandum S 47 (M 13). Het is slechts bedoeld ter orientatie en streeft niet naar volledigheid of volledige exactheid.

2) Zie voor de definitie van S memorandum S 47 (M 13).

Wij bepalen nu de S voor het stelsel paren $(1,y_{11})(2,y_{12})\ldots$ $\ldots(n_1,y_{1n_1})$ en noemen deze S_1. Evenzo behandelen wij de tweede rij en verkrijgen zo S_2 etc.

Wij bepalen dan de som

$$S_{tot} = S_1 + S_2 + \ldots + S_m.$$

Eveneens berekenen wij bij de eerste rij de spreiding σ_1, volgens formule (1) of formule (2) van memorandum S 47 (M 13). Evenzo σ_2, $\sigma_3,\ldots,\sigma_m$.

Wij weten nu, dat onder de hypothese H_0 voor rij 1 $\underline{S}_1$ bij benadering normaal verdeeld is met gemiddelde 0 en spreiding σ_1, indien H_0 geldt voor rij 2 is $\underline{S}_2$ bij benadering normaal verdeeld met gemiddelde 0 en spreiding σ_2 etc. Indien hypothese H_0 geldt voor alle rijen, dus indien in alle rijen alle permutaties der rangnummers van de waarnemingen even waarschijnlijk zijn, en indien deze rijen bovendien onafhankelijk zijn, dan geldt dat $\underline{S}_{tot}$ bij benadering normaal verdeeld is met gemiddelde 0 en spreiding:

$$\sigma_{tot} = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_m^2}.$$

Wij kiezen nu als kritieke zône weer $|S_{tot}| \geqq S_0$ bij tweezijdige toetsing, $S_{tot} \geqq S_0'$ bij rechtszijdige en $S_{tot} \leqq S_0''$ bij linkszijdige toetsing. Voor literatuur over deze toets zie men [2] .

Literatuur: [1] H.B.Mann: Non parametric tests against trend, Econometrica 13 (1945) blz. 245-259.

[2] G.Elfving en J.H.Whitlock: A simple trend-test with application to erythrocyte data, Biometrics 6 (1950) Blz. 282-288.

ATHEMATISCH CENTRUM
de Boerhaavestr.49
MSTERDAM O.
tatistische Afdeling
47 (M14)

Methode der m rangschikkingen [1]

Een duidelijk voorstelling van deze toetsingsmethode verkrijgt men door n elementen te beschouwen, die een bepaald kenmerk, eventueel in verschillende mate, bezitten. dit kenmerk wordt door m waarnemers beoordeeld en ieder van deze waarnemers rangschikt deze n elementen volgens zijn beoordeling naar opklimmende waardering. Op deze wijze ontstaan m rijen van rangschikkingen. We willen nu een maat aangeven voor de overeenstemming tussen deze rangschikkingen, m.a.w. een maat voor de overeenstemming tussen de m beoordelingen. De hypothese $\mathcal{H}_0$, die met deze methode getoetst kan worden, houdt in dat er geen overeenstemming tussen de waarnemers bestaat; precieser gezegd, dat alle rangschikkingen onafhankelijk van elkaar op toevallige wijze zijn ontstaan. Dit is b.v. het geval, als het betrokken kenmerk in werkelijkheid voor alle elementen dezelfde waarde bezit.

We kunnen de afleiding voor de maat van overeenstemming het eenvoudigst geven aan de hand van een voorbeeld.

elementen		A	B	C	D	E	F
rangnummers toegekend door waarnemer	a	5	4	1	6	3	2
	b	2	3	1	5	6	4
	c	4	1	6	3	2	5
	d	4	3	2	5	1	6
		15	11	10	19	12	17

De som van alle rangnummers is $\frac{1}{2} n m(n+1)$. Onder de hypothese $\mathcal{H}_0$ is het theoretische gemiddelde van iedere kolom: $\frac{1}{2} m(n+1)$

We beschouwen nu de afwijkingen van dit gemiddelde. In ons voorbeeld is het theoretisch kolomgemiddelde gelijk aan 14. De afwijkingen daarvan zijn

$$1 \quad -3 \quad -4 \quad 5 \quad -2 \quad 3$$

---

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

De som der kwadraten van deze afwijkingen noemen wij S.
In ons voorbeeld is S = 64.
Als alle m rangschikkingen gelijk zijn wordt het maximum van S bereikt.
Dit maximum is $\frac{1}{12}\, m^2(n^3-n)$.
We definiëren nu als <u>coëfficiënt van overeenstemming</u>

$$W = \frac{12\ S}{m^2(n^3-n)}$$

In ons voorbeeld is $W = \frac{12 \times 64}{16 \times 210} = 0{,}229$.

W varieert dus tussen 0 en 1.

De verdeling van $\underline{S}$ onder de hypothese H_o is exact berekend voor een aantal waarden van n en m [1] , terwijl voor grote m en n benaderingen bekend zijn.

De meest gebruikelijke benaderingen zijn de volgende.

1°. De χ^2-benadering:

$\chi_r^2 = m(n-1)\underline{W} = \frac{12\ \underline{S}}{mn(n+1)}$ heeft voor $m \to \infty$ een χ^2-verdeling met n-1 vrijheidsgraden ([1] pg. 84 [2] pg. 36-37).

2°. De z-benadering:

$\underline{V} = (m-1)\,\frac{\underline{W}}{1-\underline{W}}$ is bij benadering verdeeld als $\underline{F} = e^{2\underline{z}}$
($\underline{F}$ is de $\underline{F}$ van Snedecor, $\underline{z}$ de $\underline{z}$ van Fisher) met

$\nu_1 = n-1-\frac{2}{m}$

$\nu_2 = (m-1)\,\nu_1$ vrijheidsgraden ([1] pg. 84 [2] pg.33-36).

Met behulp van de verdelingen van $\underline{S}$ of $\underline{W}$ onder de hypothese H_o, kan deze hypothese getoetst worden, waarbij H_o verworpen wordt als $\underline{W}$ waarden dichtbij 1 (resp. $\underline{S}$ dichtbij $\frac{1}{12}\, m^2(n^3-n)$) aanneemt, de kritieke zône is dus van de vorm $W \geq W_o$ (resp. $S \geq S_o$).

Het kan voorkomen dat de waarnemers geen onderscheid ontdekken in de mate waarin verschillende elementen het kenmerk bezitten. Ze geven deze elementen dan gelijke rangnummers.

Veronderstel, dat door een waarnemer geen onderscheid wordt gemaakt tussen de elementen, die de rangnummers 3 t/m 6 moeten dragen. Dan wordt als rangnummer van ieder van deze elementen het gemiddelde van de rangnummers $\frac{1}{4}\,(3 + 4 + 5 + 6) = 4\frac{1}{2}$ gebruikt.

Daar het maximum van $\underline{S}$ nu verandert, moeten wij een correctie op de formule voor $\underline{W}$ toepassen. Deze vindt men in [1] (pg.82) en [2] (pg. 28-30). Eveneens veranderen dan de formules voor de χ^2-benadering ([1] pg. 86, [2] pg. 37) en voor de z-benadering ([1] pg. 86 [2] pg. 34), doch deze correcties zijn van weinig betekenis, tenzij het aantal gelijken groot is.

Literatuur: [1] M.G.Kendall, Rank correlation methods, London 1948, Hoofdstuk 6, pag. 80.

Tabel van de verdelingsfunctie van S voor:

n = 3 m = 2 t/m 10
n = 4 m = 2 t/m 6
n = 5 m = 3
op pag. 146-149.

Tabel van de waarden van S, waarvan de overschrijdingskansen onder de hypothese H_0 gelijk zijn aan 0,05 of 0,01, berekend met behulp van de z-benadering voor:

n = 3 m = 8,9,10,12,14,15,16,18,20
n = 4 m = 4,5,6,8,10,15,20
n = 5 t/m 7 m = 3,4,5,6,8,10,15,20
op pag. 150.

[2] Ph.van Elteren, Methode der m rangschikkingen, Cursus "Parametervrije Methoden", Hoofdstuk II, Rapport S 59, Mathematisch Centrum (1951), Blz. 18-45.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.

Januari 1952.

Statistische Afdeling
S 73 (M 14a)
door
Ph.van Elteren.

Trendtoets met behulp van m+1 rangschikkingen[1].

Het kan voorkomen, dat men de methode der m rangschikkingen wenst toe te passen om na te gaan of er een "trend", d.w.z. een gemeenschappelijk stijgend of dalend verloop aanwezig is in m reeksen waarnemingen van n objecten. Wij kunnen de toets voor dit doel verscherpen, door aan het rangnummerschema der waarnemingen de rij 1,2,...,n of de rij n,n-1,...,1 toe te voegen, al naar gelang de trend, die wij wensen op te sporen, stijgend dan wel dalend is.

Wij maken hierbij gebruik van de volgende feiten:

1°. De grootheid S van de methode der m rangschikkingen verandert niet, indien wij de kolommen permuteren.

2°. De verdeling van $\underline{S}$ onder de hypothese H_o[2] verandert niet indien wij de kolommen permuteren.

3°. Wij kunnen dus de verdeling van $\underline{S}$ onder de hypothese H_o berekenen, door een schema van m rijen te beschouwen, bestaande uit de getallen 1,2,...,n, waarvan een rij vast ligt (b.v.: in de volgorde 1,2,...,n) terwijl alle permutaties van de overige rijen even waarschijnlijk en onderling onafhankelijk zijn.

Hieruit volgt dat indien wij aan een schema van m rijen waarvoor de hypothese H_o geldt, de vaste rij 1,2,...,n of n,n-1,...,1 toevoegen, de grootheid $\underline{S}$ van het uitgebreide schema verdeeld zal zijn als de $\underline{S}$ van een schema met m+1 rangschikkingen onder de hypothese H_o.

Indien in het oorspronkelijke schema H_o niet geldt, doch een gemeenschappelijk stijgende trend aanwezig is, dan zal door toevoeging van de rij 1,2,..,n de stijgende trend in de kolomtotalen versterkt worden, waardoor men kan verwachten, dat er spoediger een significant resultaat zal optreden.

Evenzo kan door toevoeging van de rij n,n-1,...,1 een dalende trend van de kolomtotalen versterkt worden.

---

1) Dit memorandum is een aanvulling op memorandum S 47 (M 14). Het is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) H_o houdt in, dat in iedere rangschikking alle permutaties der rangnummers gelijke waarschijnlijkheid bezitten en dat de rangschikkingen stochastisch onafhankelijk zijn.

Wij wijzen er met nadruk op, dat de richting van de trend, waartegen H_0 getoetst wordt. gegeven moet zijn. De toets is één-zijdig en is nog niet op geschikte wijze tot een tweezijdige toets uitgebreid.

Indien nu b.v. tegen stijgende trend wordt getoetst, dus de rij 1,2,...,n wordt toegevoegd en de m+1 rangschikkingen een significante overeenstemming vertonen, zou het zeer goed denkbaar zijn dat de kolomtotalen toch geen stijgende trend zouden vertonen. Indien nl. de m gegeven rangschikkingen zeer goed overeenstemmen en m niet te klein is, blijft deze significantie bij toevoeging van iedere rij rangnummers behouden. In een dergelijk geval zou men dus zowel bij toevoeging van de rij 1,2,...,n als van de tegengestelde rij n,n-1,...,1 een significante overeenstemming tussen de m+1 rangschikkingen kunnen vinden. Wij stellen daarom nog een extra eis, waaraan voldaan moet zijn, willen wij H_0 verwerpen ten gunste van de hypothese van een stijgende trend (voor een dalende trend analoog). Niet alleen eisen wij, dat er een significante overeenstemming (met onbetrouwbaarheidsdrempel α) bestaat tussen de m+1 rangschikkingen, die verkregen zijn na toevoeging van de rij 1,2,...,n, maar bovendien moet de overeenstemming tussen deze m+1 rangschikkingen beter zijn dan tussen de oorspronkelijke m rangschikkingen. Dit laatste valt aldus te preciseren: de overschrijdingskans, behorende bij de toets voor overeenstemming tussen de rangschikkingen, moet door toevoeging van de rij 1,2,...,n kleiner worden[3].

Door de toevoeging van deze eis daalt de onbetrouwbaarheid van de toets beneden de opgegeven onbetrouwbaarheidsdrempel α . De grootte van deze daling is niet bekend; wel ziet men gemakkelijk in, dat deze daling niet groter dan $\frac{1}{2}\alpha$ kan zijn. Indien de werkelijke onbetrouwbaarheid bekend was, zou de toets op bruikbare wijze tweezijdig gemaakt kunnen worden, hetgeen nu nog niet mogelijk is.

---

3) In plaats daarvan kan men b.v. ook eisen, dat de rij der kolomtotalen van m rangschikkingen positieve rangcorrelatie vertoont met de rij 1,2,...,n (vgl. S 47 (M 13)); dit maakt weinig of geen verschil met het bovenstaande.