

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 124

Diagnoseschema voor reumapatiënten gebaseerd op serologische reacties

door

A. Benard

en

Mevr. G. Klerk-Grobbe

1953

1. Inleiding.

In het proefschrift van Mevrouw P.E. van LOCHEM-LANGEREIS "Serologische reacties bij reumatische ziekten", 1950, verder aan te geven met "de dissertatie", worden o.a. de volgende drie ¹⁾ serologische tests behandeld, die van nut kunnen zijn bij de diagnose van rheuma:

1. de antistreptolysine-test (afkorting AST)
2. de L-agglutinine-test (afkorting L)
3. de gewijzigde Rose-test (afkorting R)

De drie gevallen, die bij de onderzochte personen in dit verband van het meeste belang waren, zijn:

1. gezond (afkorting G)
2. reumatoïde **arthritis** (afkorting RA)
3. acuut rheuma (afkorting A).

Deze afkorting met hoofdletter (G, RA en A) zullen we gebruiken voor diagnoses welke op andere gronden dan de bovengenoemde serologische tests tot stand zijn gekomen; deze diagnosemethode zullen we verder "diagnose D" noemen.

De bedoeling is nu een nieuwe diagnose-methode voor deze drie gevallen op te stellen, die gebaseerd is op de uitslag van de drie serologische tests en zo goed mogelijk aansluit bij de resultaten van diagnose D en de eigenschappen van dit diagnose-schema, dat we "diagnose d" zullen noemen, te analyseren. De volgens diagnose d aangewezen gevallen geven we ter onderscheiding van de uitslag van diagnose D met overeenkomstige kleine letters aan (dus resp. g, ra en a) en de eventueel overblijvende twijfelgevallen met t.

Bij de ontwikkeling van een dergelijk schema gaat men er in feite min of meer van uit, dat diagnosemethode D een goede methode is, immers deze methode zal de grondslag moeten vormen voor het opstellen van schema d en men zal trachten ervoor te zorgen, dat D en d zo vaak mogelijk dezelfde uitslag geven. Men kan dan achteraf, d.w.z. bij nieuwe gevallen, D aanvullen of controleren met d of zelfs D in bepaalde gevallen door d vervangen.

De uitkomst van de drie tests bij een bepaalde persoon bestaat uit drie getallen AST, L en R, die dus tezamen voorgesteld

-
- 1) De vierde test, de bepaling van de O-agglutinetiter, werd in overleg met Mevrouw van Lochem, buiten beschouwing gelaten, daar deze minder geschikt leek voor de diagnose van rheuma, teneinde het onderzoek te vereenvoudigen.

kunnen worden als een punt T (titerpunt) in een driedimensionale ruimte met deze drie getallen als coördinaten. Onderzoekt men een groot aantal personen, behorende tot de drie groepen G, RA en A, dan vormen voor ieder van deze groepen de titerpunten T een "puntenwolk". Indien deze puntenwolken voor G, RA en A ongeveer hetzelfde zijn, dan is geen bruikbaar schema op te stellen, daar dan de serologische tests geen geschikt middel zijn om de drie groepen van elkaar te onderscheiden.

Liggen de drie puntenwolken echter min of meer gescheiden dan is het mogelijk in de ruimte drie gebieden g, ra en a aan te geven, zodanig dat een titerpunt in gebied g meestal behoort bij een persoon uit de groep G en analoog voor de gebieden ra en a en de groepen RA en A. Een verdeling van de ruimte waarbij de diagnoses g, ra en a precies overeenstemmen met de diagnoses G, RA en A, is niet mogelijk, daar de puntenwolken gedeeltelijk over elkaar bleken te vallen. In het bijzonder bleek de scheiding tussen G en A nogal onscherp te zijn. Om deze reden is dan ook bij diagnose d nog een twijfelgebied t tussen g en a ingelast.

De dissertatie geeft belangrijke aanwijzingen over de wijze waarop grenzen tussen g, ra en a getrokken kunnen worden; door het ontbreken van gegevens over zowel AST als L en R bij iedere persoon afzonderlijk, moet echter voor een nauwkeuriger bepaling van deze grenzen en een analyse van diagnoseschema d (aangeven van punt- en intervallschattingen voor de kansen, dat bepaalde D-diagnoses niet met de d-diagnoses overeenstemmen) gebruik worden gemaakt van gegevens uit materiaal dat na het verschijnen van de dissertatie is verzameld en dat we hieronder met "het nieuwe materiaal" zullen aanduiden.

Opmerking. In de dissertatie worden de resultaten van L-en R-test weergegeven door breuken met teller 1. Wij geven er de voorkeur aan de noemers van deze breuken als maat te gebruiken. In dit verslag zal dus b.v. $R = 40$ betekenen: de gewijzigde Rose titer is $\frac{1}{40}$ en b.v. $L \geq 20$: de L-agglutinetiter is hoogstens $\frac{1}{20}$. Op deze wijze wordt bereikt dat voor gezonde personen de titers alle drie over het algemeen op een lager niveau liggen dan de titers van zieke personen.

We willen nu achtereenvolgens de gegevens uit de dissertatie en het nieuwe materiaal bespreken en aan de hand daarvan een diagnoseschema d opstellen. Daarna volgt nog een analyse van bijzonderheden uit het materiaal en een samenvatting van alle conclusies.

2. Gegevens uit de dissertatie en conclusies hieruit over diagnoseschema d.

De gegevens uit de dissertatie die voor het opstellen van een diagnoseschema van betekenis zijn, zijn in tabel I weergegeven.

Tabel I
Overzicht van enige gegevens uit de dissertatie

	gezonde (G)		acuut rheuma (A)		rheumatoïde arthritis (RA)	
	aantal sera	percentage	aantal sera	percentage	aantal sera	percentage
AST < 200	129	89,6	9	14,3	196	78,1
200 ≤ AST ≤ 400	12	8,3	21	33,3	46	18,3
400 ≤ AST	3	2,1	33	52,4	9	3,6
totaal	144	100	63	100	251	100
L ≤ 20	94	93,1	58	92,1	17	4,8
L = 40	5	5,0	2	3,2	16	6,4
L = 80	2	2,0	1	1,6	40	15,9
L ≥ 160	0	0	2	3,2	183	72,9
totaal	101	100	63	100	251	100
R ≤ 4	64	95,5	29	100	22	20,4
R = 8	3	4,5	0	0	5	4,6
R = 16	0	0	0	0	7	6,5
R ≥ 32	0	0	0	0	74	68,5
totaal	67	100	29	100	108	100

We moeten hierbij opmerken dat de onderzochte sera niet alle van verschillende personen genomen zijn. Zo waren de 251 sera van RA-patiënten, waarvan de AST en de L bepaald zijn, afkomstig van 203 patiënten. De verschillende waarnemingen kunnen dus niet als onderling onafhankelijk beschouwd worden, zodat de opgegeven percentages niet geheel betrouwbaar zijn. Met behulp van deze tabel is het niet mogelijk de titerpuntwolken van de drie groepen G, RA en A te bepalen, omdat men niet beschikt over de resultaten der drie tests van iedere patiënt afzonderlijk.

Toch geeft tabel I een aantal belangrijke aanwijzingen. Het algemene beeld van de L- en R-tests is voor de hier beschouwde groepen hetzelfde. De gezonden en acuut-rheumapatiënten hebben bijna uitsluitend een lage L en R terwijl de rheumatoïde arthritispatiënten overwegend een hoge L en R bezitten.

Het onderscheid tussen gezond en acuut rheuma zal in de

AST-waarde gezocht moeten worden, hoewel dit onderscheid vermoedelijk niet zo scherp is. Personen met zowel een hoge AST- als een hoge L- en R-titer zullen maar weinig voorkomen. Voor een preciezere afbakening van de verschillende gebieden zullen we het nieuwe materiaal in de beschouwing moeten betrekken.

3. Het nieuwe materiaal.

Dit materiaal bestond uit de resultaten van drie tests voor elke persoon afzonderlijk van een aantal verschillende groepen mensen. Tabel II geeft een overzicht van dit materiaal met enkele bijzonderheden als datum en leeftijd voor zover bekend.

Tabel II
Overzicht van het nieuwe materiaal

	datum v/h experiment	plaats van herkomst	leeftijd	aantal
<u>a</u> gezonde vrouwen	14-12-'51	Uitkoorn	diverse (bekend)	64
	15- 1-'52	Overschie	"	48
	1- 2-'52	Schoonhoven	"	66
<u>b</u> gezonde mannen	October '50	"	20-22 j.	86
<u>c</u> A-patiënten	2e helft '50	"	-	47
<u>d</u> RA-patiënten	2e helft '50	"	-	156

Bij enkel personen ontbraken een of twee van de drie bepalingen; deze zijn waar de drie bepalingen nodig zijn (titerpuntwolken) weggelaten. In de groepen c en d zijn bij verschillende patiënten meerdere malen de drie tests uitgevoerd; voor deze patiënten is dan alleen van het eerste drietal bepalingen gebruik gemaakt teneinde voor alle personen gelijkwaardige gegevens te verwerken.

We moeten nu trachten een goede scheiding tussen gezonden en acuut rheumatici te vinden op grond van de AST-bepaling. Daartoe zijn in figuur 1 tegen de AST-waarde, χ , uitgezet:

- 1) het percentage der gezonden dat een AST-waarde groter dan χ heeft, en
- 2) het percentage der acuut rheumatici dat een AST-waarde kleiner dan χ heeft.

Bij ieder percentage is bovendien de bovenste betrouwbaarheidsgrens voor dit percentage aangegeven met een onbetrouwbaarheid 0,05. Dit betekent het volgende (voor de gezonden en ana-

loog voor de andere groep; zie ook memorandum S 47 (M 18) ²⁾): Indien de waargenomen groep gezonden beschouwd kan worden als een steekproef uit de collectie gezonden ten aanzien waarvan men zijn conclusies wil trekken (b.v. alle gezonde Nederlanders), dan is de kans dat deze betrouwbaarheidsgrens inderdaad groter is dan het percentage van gezonden in genoemde collectie met een AST-waarde groter dan χ , voor iedere χ afzonderlijk, 0,95, of anders gezegd: de gebruikte methode heeft bij iedere χ een kans 0,05 om een te kleine bovengrens te geven. Het bedoelde percentage is dus een onbekende grootheid; ons gevonden percentage is een waarneming hiervan. Bij de berekening van deze grens gaan we uit van een binomiale waarschijnlijkheidsverdeling voor het aantal gezonden met een AST-waarde groter dan χ (of bij A-patiënten een AST-waarde kleiner dan χ). De gevonden betrouwbaarheidsgrenzen behorende bij verschillende waarden van χ zijn niet onafhankelijk. Dit impliceert hier: er is wel is waar een kleine kans (0,05) dat onze bovengrens voor een of andere χ te laag ligt, doch als dit het geval is, zal zij niet alleen voor die waarde maar voor een groot interval van χ -waarden te laag kunnen zijn.

We merken nog op dat er tussen de in de figuur aangegeven waarden van χ , geen AST-bepalingen voorkomen, zodat dus de gevonden percentages en grenzen ook voor de tussen deze waarden gelegen punten gelden.

Op grond van deze resultaten lijkt ons nu de volgende scheiding tussen gezond en acuut rheuma volgens AST-waarde aanbevelenswaardig (dit is nog afgezien van de patiënten met rheumatoïde arthritis):

$$\begin{aligned} \text{AST} < 220 &\rightarrow \text{gezond} \\ 220 \leq \text{AST} \leq 320 &\rightarrow \text{twijfelgebied} \\ 320 < \text{AST} &\rightarrow \text{acuut rheuma} \end{aligned}$$

De percentages G en A welke in het nieuwe materiaal in de verschillende gebieden vallen zijn in tabel III samengevat.

Tabel III

Percentages G en A in de verschillende AST-gebieden volgens de steekproef

	AST < 220	220 ≤ AST ≤ 320	320 < AST	totaal
G	83,6	14,1	2,3	100
A	2,1	27,7	70,2	100

2) Enkele memoranda waarnaar in dit rapport wordt verwezen, zijn aan dit rapport toegevoegd.

De betrouwbaarheidsgrenzen (onbetrouwbaarheid 0,05) voor de corresponderende percentages in de collectie zijn in tabel IV gegeven (≥ 79 betekent 79% is een ondergrens, ≤ 18 betekent 18% is een bovengrens).

Tabel IV

Eenzijdige betrouwbaarheidsgrenzen corresponderend met de percentages uit tabel III

	AST < 220	220 \leq AST \leq 320	320 < AST
G	≥ 80	≤ 18	≤ 4
A	≤ 9	≤ 40	≥ 58

Zoals we in deze tabel zien, mogen we verwachten dat maar zelden het titerpunt van een A-patiënt in het gebied g of het titerpunt van een gezonde in gebied a terecht zal komen.

Voor het onderscheid tussen gezonden en acut rheuma-patiënten enerzijds en rheumatoïde arthritis-patiënten anderzijds zullen de waarden van L en R bekeken moeten worden. De gegevens uit de dissertatie doen een correlatie tussen de L- en R-waarden bij RA-patiënten vermoeden, welke in het nieuwe materiaal ook inderdaad aangetoond kan worden (paragraaf 4). De titerpunten van het RA-patiënten met een bepaalde AST-waarde zullen dus een puntenwolk van de in figuur 2 geschetste gedaante vertonen, terwijl de titerpuntenwolk bij gezonden (of acut rheumatici) met dezelfde AST-waarde overwegend bij lage waarden

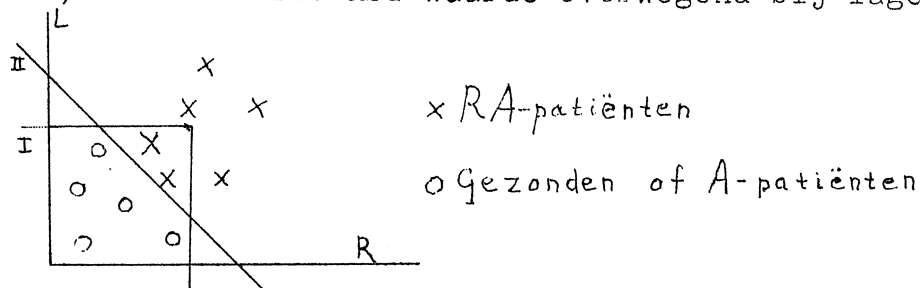


fig. 2. Schets van de titerpuntenwolk bij een gegeven AST.

van L en R gelocaliseerd is. Op grond hiervan komt het ons waarschijnlijk voor om te verwachten dat een schuine lijn als II in figuur 2 een doeltreffender scheidingslijn zal zijn tussen gezonden (of A-patiënten) en RA-patiënten dan een lijn van de gedaante I.

Door het discrete karakter van de waarnemingen van L en R zullen deze grenzen het beste met trapvormige lijnen kunnen worden weergegeven.

Het op deze vrij vage motieven opgebouwde diagnoseschema d, wordt, zoals we hieronder zullen zien, door vergelijking met

de diagnoses volgens D verder gerechtvaardigd.

De figuren 3, 4 en 5 geven voor ieder van de drie intervallen van AST-waarden ($AST < 220$; $220 \leq AST \leq 320$ en $320 < AST$) een L-R-diagram van de groepen G, RA en A. De in deze figuren getrokken lijnen worden als grenzen tussen de gebieden g en ra, t en ra en a en ra voorgesteld. In figuur 6 is nog eens beknopter weergegeven, hoe de verschillende gebieden van de drie-dimensionale AST-L-R-ruimte, waar diagnose d op berust, liggen. Daar uit de dissertatie bleek dat de RA-patiënten in meerderheid een lage AST-waarde bezitten, zijn voor de begrenzing tussen ra en de andere gebieden de met de AST toenemende waarden van L en R gekozen.

Tabel V geeft een overzicht van de aantallen titerpunten van G, A en RA welke in het nieuwe materiaal in deze gebieden vallen, terwijl in tabel VI deze aantallen uitgedrukt in percentages van de totalen weergegeven zijn.

Tabel V

Aantallen titerpunten van G, A en RA in de gebieden g, a, ra en

D \ d	g	a	ra	t	totaal
G	205	6	14	37	262
A	1	33	0	13	47
RA	3	1	145	4	153

Tabel VI

Percentages van G, A en RA in de verschillende gebieden

D \ d	g	a	ra	t	totaal
G	78,2	2,3	5,3	14,1	100
A	2,1	70,2	0	27,7	100
RA	2,0	0,65	94,75	2,6	100

In tabel VII zijn op dezelfde manier als bij tabel IV is besproken, betrouwbaarheidsgrenzen corresponderende met deze percentages samengevat.

Tabel VII

Betrouwbaarheidsgrenzen corresponderende met de percentages uit tabel VI (onbetrouwbaarheid 0,05)

D \ d	g	a	ra	t
G	≥ 73	≤ 4	≤ 8	≤ 18
A	≤ 9	≥ 58	≤ 6	≤ 40
RA	≤ 4	≤ 2	≥ 91	≤ 5

Uit deze tabel blijkt dat slechts weinig diagnoses volgens schema d fout zullen zijn in vergelijking met schema D; wel treden nog al veel twijfelgevallen op.

4. Statistisch vooronderzoek van het waarnemingsmateriaal.

Bij de opstelling van schema d is gebruik gemaakt van een aantal uitkomsten van statistische toetsen welke we hieronder kort laten volgen.

Verder is bij het onderzoek gebleken dat in het waarnemingsmateriaal enkele onregelmatigheden voorkomen, welke bij de beoordeling van het in de vorige paragraaf behandelde niet uit het oog verloren moeten worden (vooral bij de berekening van de grenzen der betrouwbaarheidsintervallen). Hieronder volgt een nadere uiteenzetting van deze onregelmatigheden. Het merendeel van de toetse heeft betrekking op de AST-waarden. De overeenkomstige toetsen op de L en R-waarden konden bij gezonden en acuut rheumatici niet worden uitgevoerd in verband met het te veelvuldig optreden van de waarde 0.

Voor de algemene gang van zaken bij het toetsen van een statistische hypothese en voor de betekenis van enkele bij de toetsingstheorie gebruikelijke termen (zoals: overschrijdingskans, welke we steeds met k zullen aanduiden), verwijzen we naar memorandum S 47 (M 6). Voor verschillende uitgevoerde toetsen worden bovendien speciale memoranda bijgevoegd.

Correlaties tussen AST, L en R.

Bij de opstelling van diagnoseschema d is gebruik gemaakt van het bestaan van een positieve correlatie tussen L en R voor RA-patiënten, waarvoor reeds aanwijzingen bestaan in de gegevens van de dissertatie (zie tabel I). Deze correlatie, alsmede de correlaties tussen AST en L en tussen AST en R werden in het nieuwe materiaal getoetst met behulp van de rangcorrelatietoets van SPEARMAN (memorandum S 73 (M 13b)). Voor AST en L en voor AST en R gaaf deze toets bij de RA-patiënten geen aanwijzing voor correlatie. Uitslagen van de toets:

$$\text{RA-patiënten} \left\{ \begin{array}{l} \text{AST - L, } k \approx 0,27 \\ \text{AST - R, } k \approx 0,79 \\ \text{L - R, } k \approx 10^{-9} \end{array} \right.$$

Vergelijking van de gegevens uit de dissertatie met het nieuwe materiaal bij gezonden.

Met de toets van WILCOXON (zie memorandum S 47 (M 7)) zijn zowel voor AST als voor L en voor R de gegevens uit de dissertatie met die uit het nieuwe materiaal vergeleken. Hierbij werden

de volgende resultaten gevonden:

- 1) AST in het nieuwe materiaal is hoger ($\bar{k} \approx 10^{-10}$)
- 2) L-waarde in het nieuwe materiaal is lager ($\bar{k} \approx 10^{-4}$)
- 3) R-waarden wijzen geen verschillen tussen dissertatie en nieuw materiaal aan ($\bar{k} = 0,87$)

Inhomogeniteit van de AST-waarden.

Het verschil in de AST-waarden is nader onderzocht. In de eerste plaats werden aanwijzingen voor een afhankelijkheid van AST met de leeftijd gevonden. Dit is onderzocht bij de drie groepen gezonde vrouwen, waarvan de leeftijden bekend waren, met behulp van de rangcorrelatietoets van SPEARMAN (zie memorandum S 73 (M 13b)), met de resultaten:

- Vrouwen uit Uithoorn, daling met de leeftijd ($\bar{k} = 0,001$)
- Vrouwen uit Schoonhoven, slechts zwakke daling ($\bar{k} = 0,12$)
- Vrouwen uit Overschie, daling met de leeftijd ($\bar{k} = 0,018$).

Om dit leeftijdseffect te elimineren bij het onderzoek naar een eventueel tijdsverloop, is de groep soldaten met behulp van de toets van WILCOXON vergeleken met een groep gezonde vrouwen van ongeveer dezelfde leeftijd (20-22 jaar). De AST-waarden bij soldaten lagen duidelijk lager volgens deze toetsing ($\bar{k} < 10^{-6}$).

Een verklaring hiervoor kan zijn een stijging van de AST-waarden met het tijdstip van bepaling (b.v. door een verloop in de bepalingstechniek). De AST werd bij de soldaten een jaar eerder bepaald dan bij de vrouwen (zie tabel II), maar ook een systematisch verschil tussen mannen en vrouwen wat de AST-waarde betreft zou de uitslag van deze toets kunnen verklaren. Bovendien werden de AST-waarden van de soldaten vergeleken met de gegevens uit de dissertatie, welke laatste nog weer eerder verzameld werden. Het resultaat van deze toetsing (eveneens met de toets van WILCOXON) was een verdere ondersteuning van de eerstgenoemde onderstelling:

AST-waarden van soldaten hoger dan die uit de dissertatie (overschrijdingskans $\bar{k} = 0,027$).

Ten slotte werden de groepen gezonde vrouwen uit de verschillende plaatsen nog eens onderling vergeleken. Het leeftijds-effect werd geëlimineerd door de totale groep gezonde vrouwen in 7 groepen van 25 te verdelen, zodat binnen een groep de leeftijd ongeveer constant is. Op deze 7 groepen van waarnemingen uit de drie plaatsen Uithoorn, Overschie en Schoonhoven is een

generalisatie ³⁾ van de methode van m rangschikkingen onderzocht of een rangschikking van de drie plaatsen naar de AST-waarden in alle leeftijdsgroepen al dan niet overeenstemt.

Het resultaat was een overschrijdingskans 0,04, dus een lichte aanwijzing voor overeenstemming, terwijl de gevonden volgorde van stijging van de AST-waarden, nl. Uithoorn, Overschie, Schoonhoven overeenkomt met de volgorde van het tijdstip van behandeling, zij het ook dat deze tijdstippen zeer dicht bij elkaar lagen.

Vatten we de resultaten uit deze paragraaf samen dan kunnen we zeggen:

1. De L- en R-waarden bij RA-patiënten zijn sterk gecorreleerd; dit resultaat is gebruikt bij de opstelling van diagnose d.

2. Tussen het nieuwe materiaal en de gegevens uit de dissertatie zijn belangrijke verschillen gevonden, vooral in de AST-waarden, maar ook in de L-waarden: de AST is gestegen, de L-titer is gedaald.

3. Een mogelijke verklaring voor dit verschil in AST-waarden ligt in een verloop met het tijdstip van bepaling (wellicht een verloop in de waarnemingstechniek?) en wel een stijging van de AST-bepalingen.

4. Bovendien werd bij de AST-waarden een sterke aanwijzing voor leeftijdsafhankelijkheid gevonden.

5. Het gedrag van de L- en R-waarden bij gezonden en acut rheumatici kon niet nader onderzocht worden in verband met de zeer vele waarden 0 die bij deze groepen optreden.

Bij de beoordeling van diagnoseschema d en vooral van de betrouwbaarheidsgrenzen voor de verschillende percentages uit paragraaf 3 moet in verband hiermee wel het nodige voorbehoud worden gemaakt. Met name het gedrag van de AST-bepalingen kan wellicht leiden tot een verschuiving van de grenzen tussen de verschillende gebieden en aanleiding zijn tot een frequentere

3) Bij de gewone methode van m rangschikkingen (zie mem. S 47 (M 14)) is sprake van m waarnemers, die ieder van n elementen één keer waarnemen en op grond daarvan rangnummers toekennen aan de elementen. In de hier toegepaste gegeneraliseerde methode, wordt toegelaten dat de waarnemers, ieder element 0 of meer dan één keer waarnemen. Deze toets is te gecompliceerd voor behandeling in een memorandum. Zie: A. Benard en Ph. van Elteren, A generalization of the method of m rankings, Indagationes 15 (1953), p. 358 en Proceedings Kon. Ned. Ak. v. Wet. 56 (1953).

overschrijding van deze grenzen dan volgens de betrouwbaarheids grenzen voor de kansen in de verschillende gebieden terecht te komen uit tabel VII mag worden verwacht.

5. Samenvatting.

Uitgaande van de bestaande diagnose D (G, A en RA) werd een nieuw schema d (g, a, ra en t) opgesteld, gebaseerd op de metingen van de AST, L en R-titers bij de te onderzoeken patiënt. Uit de gegevens van de dissertatie kon reeds worden afgeleid dat de scheiding tussen g en a op de AST-waarde zou moeten berusten (a bij hoge AST). Daar de scheiding tussen de AST-waarden van acuut rheumatici en gezonden niet zo scherp is, werd een overgangsgebied, t (twijfelgebied), tussen g en a ingeschakeld, waarin dus volgens diagnoseschema d nog geen pertinente uitspraak mogelijk is.

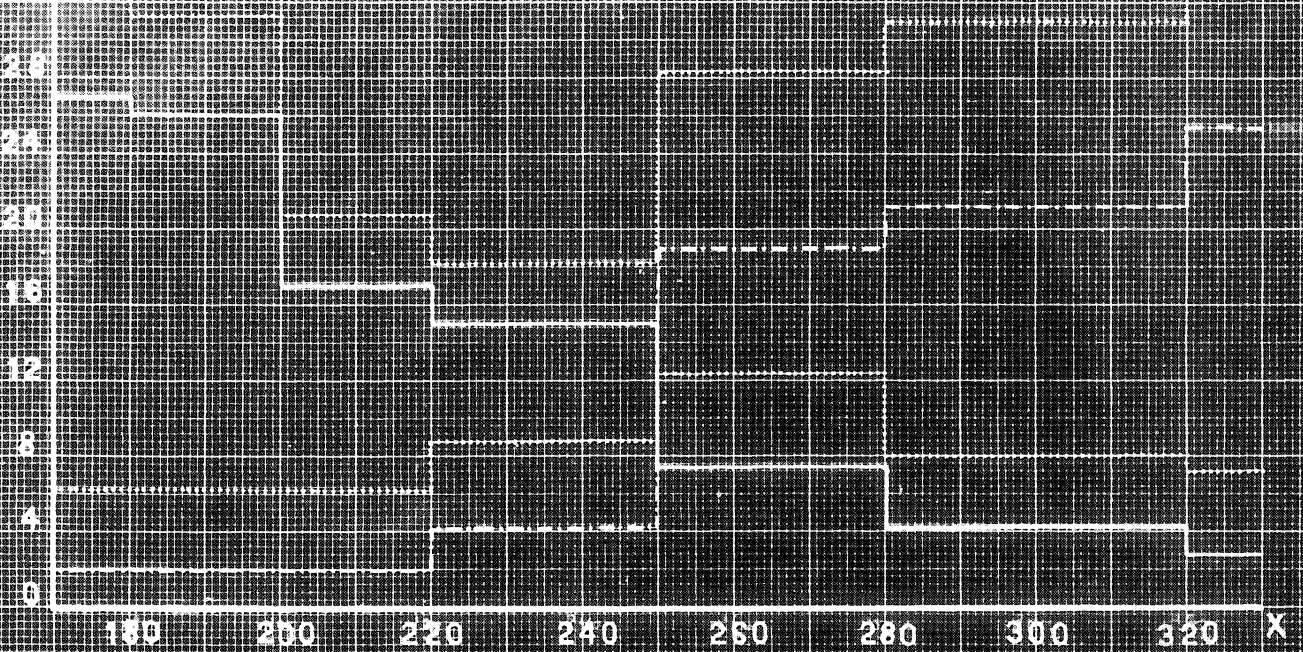
Verder bleek uit de dissertatie dat RA-patiënten voor het overgrote deel verhoogde L- en R-titers bezitten, zodat voor diagnose d bij iedere AST-waarde een grote L- en R-waarde tot de uitspraak ra zal moeten leiden.

Uit de dissertatiegegevens is dus de verdeling van de titerpuntenruimte (AST, L en R als coördinaten) in gebieden g, a, t en ra voor diagnose d reeds in grote trekken bepaald.

Voor een nauwkeurigere bepaling van de grenzen tussen de verschillende gebieden (fig. 5) en van betrouwbaarheids grenzen voor de percentages G, RA- en A-personen met titerpunten in de verschillende gebieden (tabel VII), werden de gegevens van het nieuwe materiaal gebruikt.

In paragraaf 4 werd gewezen op het voorbehoud dat bij de beoordeling van schema d gemaakt moet worden in verband met gevonden onregelmatigheden van de bepalingen. In het bijzonder blijkt er een verloop met de leeftijd en vermoedelijk ook met de tijd in de AST-titer voor te komen.

Het zal van grote betekenis zijn de waarde van diagnoseschema d te onderzoeken met nieuwe gegevens, opgezet als in "het nieuwe materiaal" en hierbij tevens aandacht te besteden aan extra factoren (dus naast de omstandigheden: gezond, acuut rheuma of rheumatoïde arthrititis) welke een invloed op de bepalingen kunnen uitoefenen. Misschien kan dan bij een eventuele verfijning van diagnoseschema d ook van de O-agglutinetiter gebruik worden gemaakt.



———— percentage AST-bepalingen $\geq X$ by G (totaal aantal 264)
 - - - - - percentage AST-bepalingen $< X$ by A (totaal aantal 47)
 bovengrenzen voor de percentages (zie tekst)

fig. 1 Percentages G en A met AST \geq resp $<$ X met bovengrenzen

X 25 BA

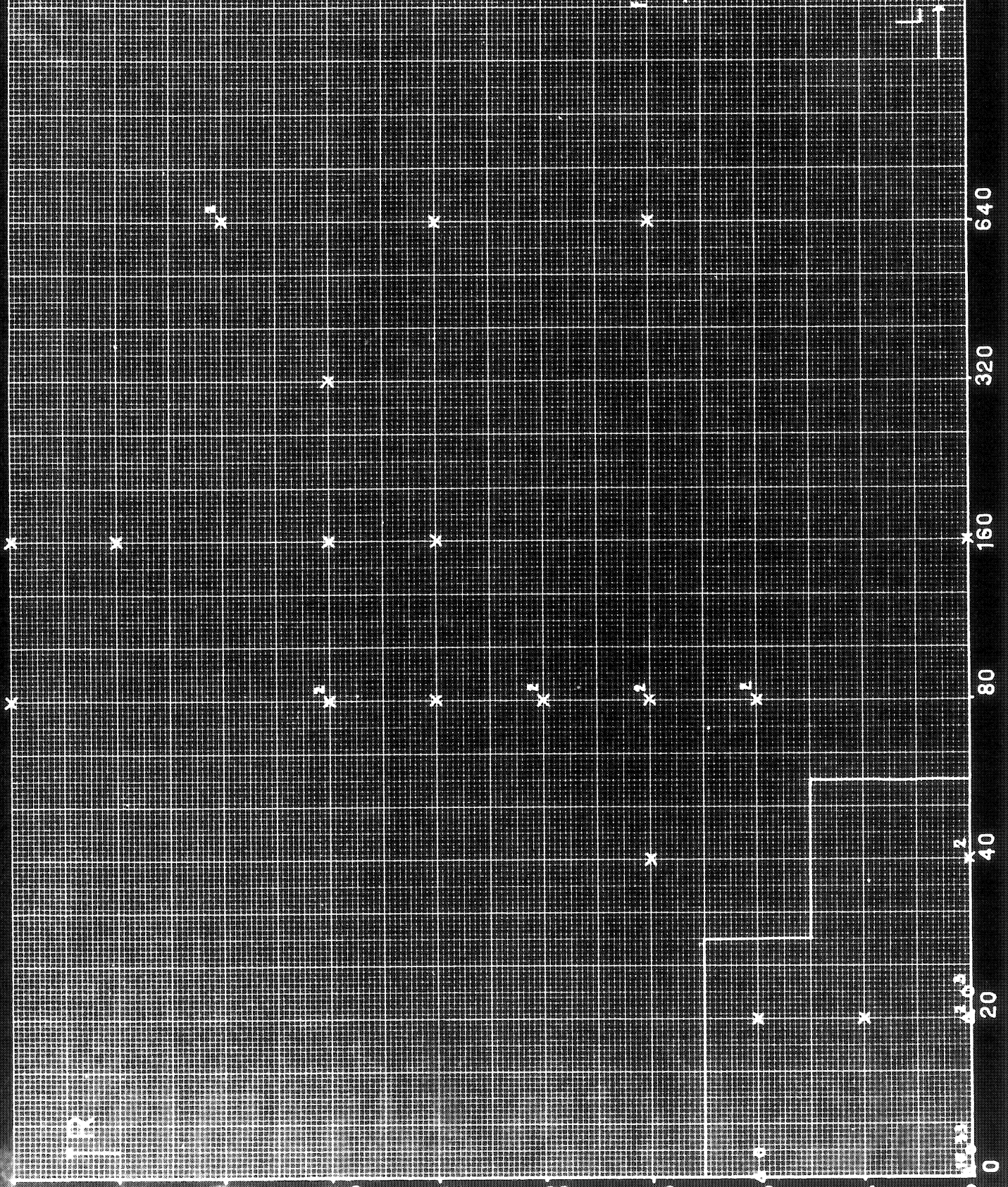
0 37 G

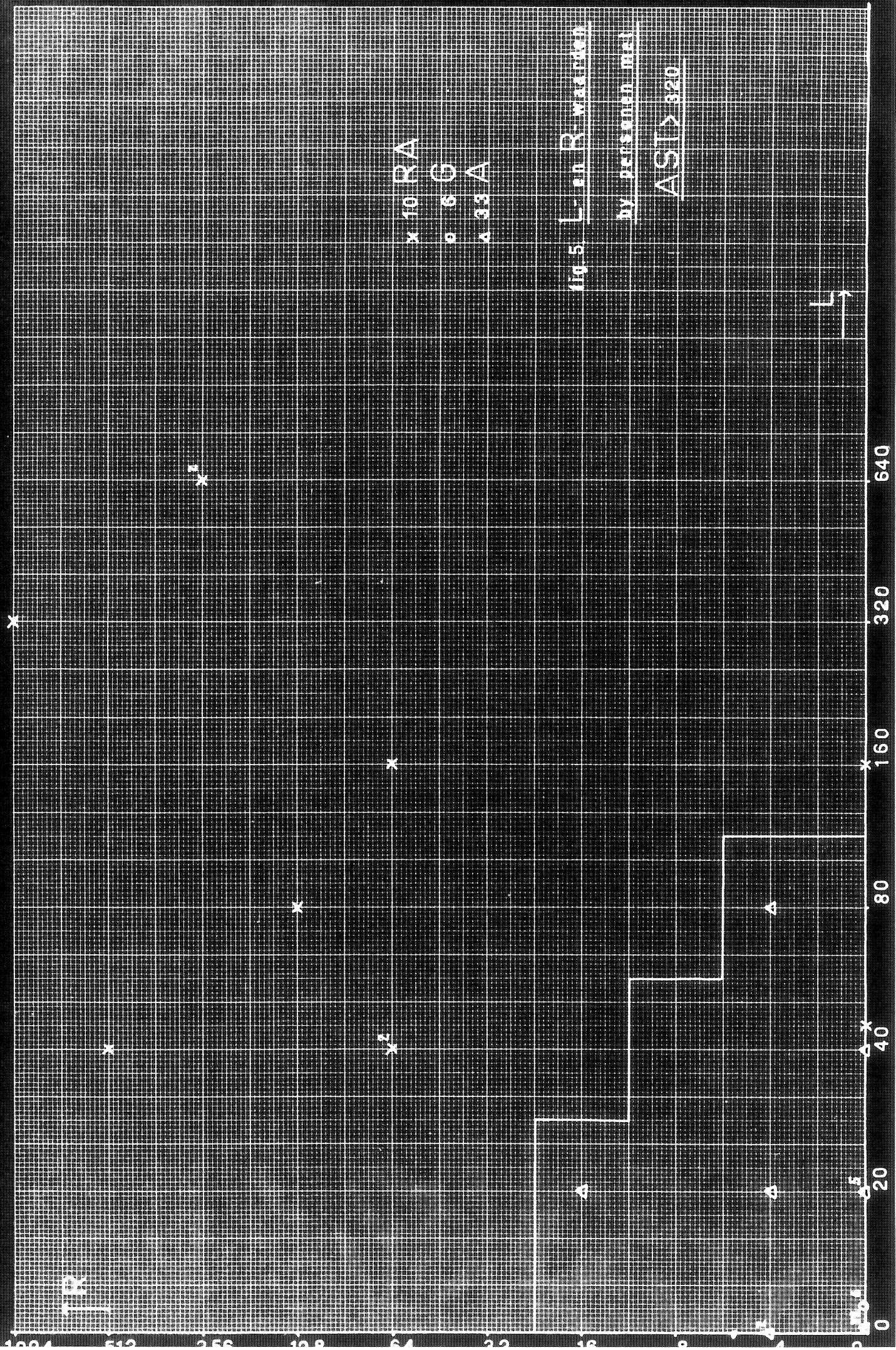
A 13 A

FIG. 1. L-113. 3. 1961

BY PERSEPOLIS

2000 AS 1000





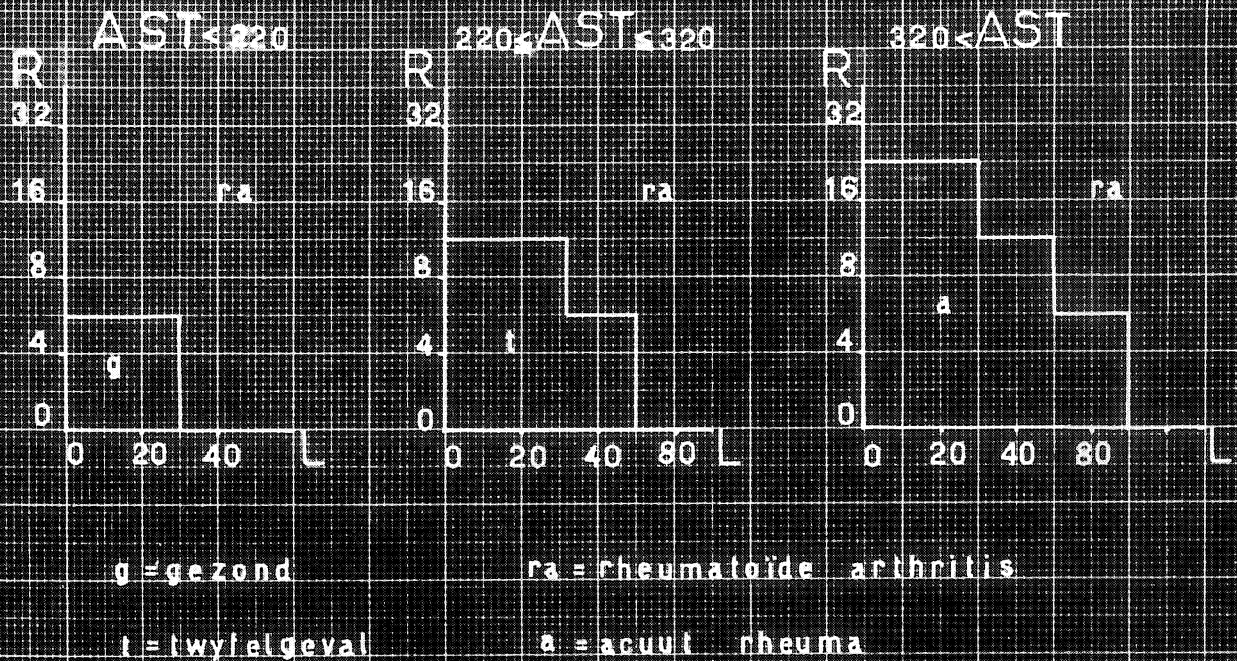


Fig 6 Scheiding tussen gezonden, acuut rheuma- en reumatoïde artritis-patiënten volgens diagnose schema d.

Algemene gang van zaken bij het toetsen van een ¹⁾
hypothese.

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen x_1, x_2, \dots, x_n van één of meer stochastische grootheden ²⁾, of op enige groepen van waarnemingen (bv. twee steekproeven).

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

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van u , en wel op zodanige wijze, dat de kans, dat u een in Z gelegen waarde aanneemt, onder de hypothese H_0 , gelijk is aan een gegeven getal α , zodat Z dus van α afhankelijk is. Z heet de kritieke zone van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwierpt nu H_0 op grond van de waarnemingen x_1, x_2, \dots, x_n , indien de bij deze waarnemingen behorende waarde van u in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α . Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0,05$ resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

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

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

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 zone Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α , dan wordt verworpen, indien $k \leq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijkte 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.

De toets van Wilcoxon.¹⁾

Deze methode dient tot het toetsen van de hypothese H_0 , inhoudende, dat twee steekproeven x_1, \dots, x_n en y_1, \dots, y_m afkomstig zijn uit één collectie (ook wel populatie of universum genaamd).

Voor het toetsen van de hypothese H_0 wordt gebruik gemaakt van een toetsingsgrootte \underline{U} ²⁾, die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen x_1, \dots, x_n en y_1, \dots, y_m naar opklimmende grootte gerangschikt zijn, dan bepalen we eerst het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de kleinste waarneming x_1 uit de eerste steekproef (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1). Noem dit aantal V_1 . Vervolgens wordt het aantal waarnemingen uit de tweede steekproef bepaald, dat kleiner is dan de op één na kleinste waarneming x_2 uit de eerste steekproef (bij gelijkheid wordt weer een $\frac{1}{2}$ in plaats van 1 geteld). Dit aantal noemen we V_2 . Evenzo worden met betrekking tot x_3, x_4, \dots, x_n de aantallen V_3, V_4, \dots, V_n bepaald. De waarde U van de toetsingsgrootte \underline{U} wordt voor de twee steekproeven dan gegeven door

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

Wanneer onder de waarnemingen niet te veel gelijken voorkomen, kan bewezen worden, dat de toetsingsgrootte \underline{U} onder de hypothese H_0 voor grote waarden van n en m (beide ≥ 10) bij benadering een normale verdeling bezit. De waarnemingen x_1, \dots, x_n en y_1, \dots, y_m tezamen genomen vallen uiteen in een aantal groepen van gelijke waarnemingen. Noem het aantal van deze groepen k , dan is k minstens 1 (als alle waarnemingen gelijk zijn) en hoogstens $m+n$ (als alle waarnemingen verschillend zijn).

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

2) Stochastische grootheden worden door onderstreping aangeduid.

Zijn t_1, \dots, t_k de aantallen waarnemingen in deze groepen van gelijken, dan worden het gemiddelde μ en de variantie σ^2 van de toetsingsgrootheid \underline{U} gegeven door

$$\mu(\underline{U}) = \frac{1}{2}nm,$$

en

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

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. Indien de hypothese H_0 niet vervuld is, zal de grootheid \underline{U} grote of kleine waarden bezitten, al naar gelang \underline{y} systematisch kleiner of groter is dan \underline{x} .

De (tweezijdige) toets bestaat nu daarin, dat men H_0 verworpt indien de gevonden waarde U van \underline{U} te sterk van μ afwijkt, d.w.z. als

$$\frac{|U - \mu|}{\sigma} > \frac{k}{\frac{1}{2}\alpha}, \quad 2)$$

waarin α de onbetrouwbaarheidsdrempel is en $\frac{k}{\frac{1}{2}\alpha}$ volgt uit

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{k}{\frac{1}{2}\alpha}}^{\infty} e^{-\frac{1}{2}x^2} dx = \frac{1}{2}\alpha,$$

en in een tabel van de normale verdeling kan worden opgezocht.

De (tweezijdige) overschrijdingskans k , behorende bij T , is gedefiniëerd als

$$k = \frac{2}{\sqrt{2\pi}} \int_{\left| \frac{U - \mu}{\sigma} \right|}^{\infty} e^{-\frac{1}{2}x^2} dx \quad 2)$$

en kan ook in eentabel van de normale verdeling worden gevonden.

Bij eenzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.

Een bijzonder geval van het bovenstaande is, dat onder de waarnemingen voor \underline{x} en \underline{y} in 't geheel geen gelijken voorkomen. In dat geval kan de uitdrukking voor de variantie herleid worden tot

$$\sigma^2 = \frac{1}{12} nm(n+m+1).$$

1) Deze formule is een door T.J.Terpstra gegeven vereenvoudiging van de door J.Hemelrijk ([5] en [7]) afgeleide formule. De afleiding van deze vereenvoudigde formule zal nog gepubliceerd worden.

2) Deze formules berusten op de normale benadering van de verdeling van \underline{U} .

Indien n en m kleiner zijn dan 10, zijn tabellen beschikbaar voor het berekenen van de overschrijdingskans k voor de uit de steekproef bepaalde waarde U van \underline{U} (zie [2] en [4]).

Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

Opmerking. Men kan gemakkelijk bewijzen, dat de variantie van \underline{U} door het optreden van gelijke waarnemingen vermindert. Het verschil, dat door deze gelijken optreedt, is echter in het algemeen gering. Men kan daarom in eerste instantie deze correctie op σ^2 verwaarlozen. De overschrijdingskansen, die men dan vindt, zijn iets te groot.

Litteratuur:

1. F.Wilcoxon, Individual comparisons by ranking methods, *Biometrics* 1 (1945), p.80-83.
- 2 H.B.Mann and D.R.Whitney On a test of whether one of two random variables is stochastically larger than the other, *Amer.Math.Stat.* 18 (1947), p. 50-60.
- 3 H.R.van der Vaart Some remarks on the power function of Wilcoxon's test for the problem of two samples, *Proceedings van de Kon. Ned.Ak.v.Wet.*, 53 (1950), p. 494-520.
- 4 H.R.van der Vaart Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en $m \leq 10$, *Rapport S32 (M4)* (1950).
- 5 H.R.van der Vaart De toets van Wilcoxon voor het probleem van twee steekproeven. (Cursus "Parameter vrije Methoden", 1951-'52).
- 6 D.van Dantzig Kadercursus Mathematische Statistiek, *Math. Centrum, Amsterdam* (1947-'50), hoofdst. 6, § 3.
- 7 J.Hemelrijk Note on Wilcoxon's two sample test, when ties are present, *Ann.Math.Stat.* 23 (1952) no. 2.

Rangcorrelatie¹⁾

1. Beschrijving van de methode.

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_i, y_i) ($i = 1, \dots, n$), zijn onafhankelijke waarnemingsparen van deze stochastische grootheden

Voorbeeld:

$i =$	1	2	3	4	5	6
x_i	0,11	0,12	0,10	0,11	0,15	0,13
y_i	3,4	3,0	3,2	3,5	3,5	3,5

Wij zeggen dat de waarnemingsparen (x_i, y_i) en (x_j, y_j) positief gecorreleerd zijn, als de volgorde van x_i en x_j hetzelfde is als die van y_i en y_j (bv. $x_i < x_j$ en $y_i < y_j$); zij zijn negatief gecorreleerd als de volgorde van x_i en x_j tegengesteld is aan de volgorde van y_i en y_j (bv. $x_i > x_j$ en $y_i < y_j$) en zij zijn niet gecorreleerd als $x_i = x_j$ of $y_i = y_j$.

In tabel 1 hebben wij van alle tweetallen (x_i, y_i) en (x_j, y_j) uit ons voorbeeld nagegaan of zij positief, negatief dan wel niet gecorreleerd zijn. Een positieve correlatie is aangeduid met +1, een negatieve met -1, terwijl het ontbreken van correlatie wordt aangegeven door een 0.

De toetsingsgrootte van de methode van rangcorrelatie is nu het aantal positief gecorreleerde tweetallen verminderd met het aantal negatief gecorreleerde, of wel de som van de getallen, die in tabel 1 in de kolom "correlatie" voorkomen.

De verdeling van S voor het geval dat x en y onafhankelijk zijn is bekend (zie § 2). De hypothese dat x en y onafhankelijk

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

Tabel 1

Berekening van S
voor het voorbeeld

<u>i</u>	<u>j</u>	<u>Correlatie</u>
1	2	-1
1	3	+1
1	4	0
1	5	+1
1	6	+1
2	3	-1
2	4	-1
2	5	+1
2	6	+1
3	4	+1
3	5	+1
3	6	+1
4	5	0
4	6	0
5	6	0

$$S = +5$$

zijn, kan dus getoetst worden.

Is de hypothese van onafhankelijkheid niet vervuld, dan is de waarschijnlijkheid van grote positieve of grote negatieve waarden van S groter, dan wanneer dit wel het geval is. De kritieke zône is daarom van de vorm $|S| \geq S_0$, en bij ééNZijDige toetsing van de vorm $S \geq S'_0$ (rechtszijdige toetsing) of $S \leq S''_0$ (linkszijdige toetsing).

2. Verdeling van S als x en y onafhankelijk zijn.

Als er noch bij de x_i noch bij de y_j gelijke waarden voorkomen kunnen wij gebruik te maken van exacte tabellen, die voorkomen in [1] pg 141 (n = 4 t/m, 10) en in [2] (tables I and II, n = 4 t/m 40). Bovendien vindt men in [2] table III de kleinste waarden van \underline{S} , waarvan de overschrijdingskansen onder de hypothese van onafhankelijkheid hoogstens gelijk zijn aan α voor $\alpha = 0,005; 0,01; 0,025; 0,05$ en $0,10$ en $n = 4, 5, 6, \dots, 40$.

Als er bij de x_i óf bij de y_i , doch niet bij beide tweetallen of drietallen gelijken voorkomen, kan men voor $n \leq 10$ gebruik maken van de tabel van Sillitto [4].

Voor grote waarden van n is de verdeling van $\frac{\underline{S}}{\sigma_{\underline{S}}}$ (waarin

$\sigma_{\underline{S}}$ de spreiding van \underline{S} is, die uit een hieronder op te geven formule berekend kan worden) bij benadering normaal met gemiddelde 0 en spreiding 1. Hiervan kunnen we gebruik maken om de hypothese van onafhankelijkheid te toetsen in de gevallen waar de exacte verdeling niet getabelleerd is. Dit geschiedt dan, door in een tabel van de normale verdeling de

overschrijdingskans op te zoeken, die behoort bij de gevonden waarde van $\frac{S}{\sigma_S}$.

Om σ_S te berekenen, nemen wij in de rij der waarnemingen x_i de gelijke waarnemingen in groepen bij elkaar. De aantallen waarnemingen in die groepen duiden wij aan met t_h , waarin $h = 1, 2, \dots, k_1$. Evenzo doet men in de rij der waarnemingen y_j , waar we de overeenkomstige aantallen aanduiden met u_l , waarin $l = 1, 2, \dots, k_2$. σ_S kan dan gevonden worden uit de volgende formule:

$$(1) \sigma_S^2 = \frac{1}{18} \left\{ n(n-1)(2n+5) - \sum_{h=1}^{k_1} t_h(t_h-1)(2t_h+5) - \right. \\ \left. - \sum_{l=1}^{k_2} u_l(u_l-1)(2u_l+5) \right\} + \\ + \frac{1}{9n(n-1)(n-2)} \sum_{h=1}^{k_1} t_h(t_h-1)(t_h-2) \sum_{l=1}^{k_2} u_l(u_l-1)(u_l-2) + \\ + \frac{1}{2n(n-1)} \sum_{h=1}^{k_1} t_h(t_h-1) \sum_{l=1}^{k_2} u_l(u_l-1).$$

In ons voorbeeld van § 1 komt in de rij x_i één tweetal gelijken (dus $k_1=1$ en $t_1=2$) en in de rij y_j één drietal gelijken ($k_2=1$, $u_1=3$) voor. Dus geldt:

$$\begin{aligned} t_1(t_1-1)(2t_1+5) &= 2 \cdot 1 \cdot 9 = 18 \\ u_1(u_1-1)(2u_1+5) &= 3 \cdot 2 \cdot 11 = 66 \\ t_1(t_1-1)(t_1-2) &= 0, \quad u_1(u_1-1)(u_1-2) = 0 \\ t_1(t_1-1) &= 2 \cdot 1 = 2 \\ u_1(u_1-1) &= 3 \cdot 2 = 6 \\ n(n-1)(2n+5) &= 6 \cdot 5 \cdot 17 = 510 \\ n(n-1) &= 6 \cdot 5 = 30 \end{aligned}$$

zodat:

$$\sigma_S^2 = \frac{1}{18} \{ 510 - 18 - 66 \} + \frac{1}{60} \times 2 \times 6 = 23,87$$

en $\sigma_S = 4,89$ is.

Als alle t_h en alle u_l gelijk zijn en er dus in geen van beide rijen gelijken voorkomen, gaat formule (2) over in:

$$(2) \sigma_S = \sqrt{\frac{1}{18} n(n-1)(2n+5)}$$

Een tabel van deze functie voor $n = 40, 41, \dots, 100$ vindt men in [2] (table IV).

3. Rangcorrelatiecoëfficiënt

Als maat voor de correlatie in de rij van waarnemingsparen $(x_1, y_1), \dots, (x_n, y_n)$ heeft Kendall de coëfficiënt τ gedefinieerd, die +1 is als de volgorden der waarnemingen in beide rijen x_1, \dots, x_n en y_1, \dots, y_n volledig overeenstemmen en -1 is, als deze volgorden volkomen tegengesteld zijn. De definitie van τ is:

$$(3) \tau = \frac{2S}{\left\{n(n-1) - \sum_{h=1}^{k_1} t_h(t_h-1)\right\}^{\frac{1}{2}} \left\{n(n-1) - \sum_{j=1}^{k_2} u_j(u_j-1)\right\}^{\frac{1}{2}}}$$

Als er in geen van beide rijen gelijke waarnemingen voorkomen wordt deze formule:

$$(4) \tau = \frac{2S}{n(n-1)}.$$

Literatuur:

- [1] M.G. Kendall. Rank correlation Methods London 1948, Hoofdstuk 1.
- [2] L. Kaarsemaker en A. van Wijngaarden. Tables for use in rank correlation. (1952)
Report R 73 of the Computation Department of the Mathematical Centre.
- [3] J. Hemelrijk. Kendall's rangcorrelatie-coëfficiënt. Hoofdstuk I der cursus "Parameter vrije Methoden" Rapport S 59 (1951) Mathematisch Centrum, blz. 1-17.
- [4] G.P. Sillitto. "The Distribution of Kendall's coefficient of rankcorrelation in rankings containing ties. Biometrika 34 (1947) p. 36-40.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - 0.

Statistische Afdeling.

S 73 (M 13a)

Trend-toets met behulp van rangcorrelatie¹⁾

Men kan de methode der rangcorrelatie toepassen om na te gaan of er in een reeks waarnemingen van een stochastische grootheid 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 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_0 , 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, \dots, y_n van de stochastische grootheid 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_0 . Onder deze hypothese heeft S dezelfde verdeling als de overeenkomstige grootte der rangcorrelatie-toets 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:

$y_{1,1}, y_{1,2}, \dots, y_{1,n_1}$

$y_{2,1}, y_{2,2}, \dots, y_{2,n_2}$

$y_{3,1}, y_{3,2}, \dots, y_{3,n_3}$

\vdots

$y_{m,1}, y_{m,2}, \dots, y_{m,n_m}$

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 $(1, y_{11})(2, y_{12}) \dots \dots (n_1, y_{1n_1})$ en noemen deze S_1 . Evenzo behandelen wij de tweede rij en verkrijgen zo S_2 etc.

Wij bepalen dan de som

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

Eveneens berekenen wij bij de eerste rij de spreiding σ_1 , volgens formule (1) of formule (2) van memorandum S 47 (M 13). Evenzo $\sigma_2, \sigma_3, \dots, \sigma_m$.

Wij weten nu, dat onder de hypothese H_0 voor rij 1 S_1 bij benadering normaal verdeeld is met gemiddelde 0 en spreiding σ_1 , indien H_0 geldt voor rij 2 is 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 S_{tot} bij benadering normaal verdeeld is met gemiddelde 0 en spreiding:

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

Wij kiezen nu als kritieke zône weer $|S_{\text{tot}}| \geq S_0$ bij tweezijdige toetsing, $S_{\text{tot}} \geq S'_0$ bij rechtszijdige en $S_{\text{tot}} \leq S''_0$ 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.

Spearman's correlatie-coëfficiënt, wanneer er geen gelijke waarnemingen optreden¹⁾.

We beschouwen het geval dat twee stochastische grootheden x en y ²⁾ een simultane verdeling bezitten. Over deze verdeling zelf behoeft niets ondersteld te worden. $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ zijn n onafhankelijke waarnemingsparen van deze stochastische grootheden.

Voorbeeld.

$i =$	1	2	3	4	5	6
x_i	0,107	0,123	0,101	0,111	0,154	0,132
y_i	3,39	3,01	3,17	3,54	3,47	3,49

Om op grond van deze waarnemingen de hypothese H_0 te toetsen, dat x en y onafhankelijk verdeeld zijn heeft C.Spearman de volgende parameter vrije toetsingsgrootheid ingevoerd:

Nummer de waarnemingen x_i ($i = 1, 2, \dots, n$) naar opklimmende grootte, evenals de y_i .

Voorbeeld.

	$i =$	1	2	3	4	5	6
rangnummer van x_i		2	4	1	3	6	5
rangnummer van y_i		3	1	2	6	4	5
verschillen der rangnummers, d :		1	3	1	3	2	0
d^2 :		1	9	1	9	4	0

Van de paren rangnummers wordt het verschil bepaald en deze verschillen worden gekwadraterd en opgeteld. Hun som is $\underline{S}(d^2)$. In dit voorbeeld dus 24.

$\underline{S}(d^2)$ varieert van 0, bij volkomen overeenstemming, tot $1/3(n^3 - n)$, bij volkomen tegengestelde nummering der beide rijen.

We verwerpen de hypothese H_0 van onafhankelijkheid ten gunste van de alternatieve hypothese: x en y zijn positief resp. negatief gecorreleerd, als $\underline{S}(d^2)$ een zeer

-
- 1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
 - 2) De onderstreping geeft aan, dat de toetsingsgrootheid stochastisch is, d.w.z. een waarschijnlijkheidsverdeling bezit.

lage resp. zeer hoge waarde aanneemt.

Tabellen van de kans, dat een zekere waarde van $\underline{S}(d^2)$ bereikt of overschreden wordt voor $n = 4$ t/m 8 zijn te vinden in Kendall [1] p. 142³⁾. Voor $n = 9$ en 10 in Biometrika [2] p. 133³⁾.

Voor $n > 10$ kan men gebruik maken van het feit, dat

$$\underline{t} = \{n^3 - n - 6\underline{S}(d^2)\} \sqrt{\frac{n-2}{12\underline{S}(d^2) \{n^3 - n - 3\underline{S}(d^2)\}}}$$

bij benadering een Studentverdeling heeft met $n-2$ vrijheidsgraden.

Voor nog grotere n kan men ook van de in dit geval minder nauwkeurige ^{normale} benadering gebruik maken. Daarbij geldt dan voor het gemiddelde $\underline{C}^e \underline{S}(d^2)$ van $\underline{S}(d^2)$

$$\underline{C}^e \underline{S}(d^2) = \frac{1}{6}(n^3 - n) = \binom{n+1}{3},$$

en voor de variantie

$$\sigma^2 \{ \underline{S}(d^2) \} = \frac{1}{36} (n^5 + n^4 - n^3 - n^2) = \frac{1}{3} \binom{n+1}{3} \binom{n+1}{2}.$$

Om de mate van afhankelijkheid van \underline{x} en \underline{y} als een getal tussen -1 en $+1$ uit te drukken, gebruikt men in dit verband de grootheid

$$\rho = 1 - \frac{6\underline{S}(d^2)}{n^3 - n}.$$

De boven ingevoerde grootheid \underline{t} kan dan als volgt in ρ worden uitgedrukt;

$$\underline{t} = \rho \sqrt{\frac{n-2}{1-\rho^2}}.$$

3) Zie literatuuropgave achteraan.

Literatuur:

[1] M.G.Kendall, Rank Correlation Methods, London, 1948.
[2] S.T.David, M.G.Kendall and A.Stuart, Some questions of distribution in the theory of rank correlation, Biometrika 38 (1951) p. 131-140.

Methode der m rangschikkingen ¹⁾

Een duidelijke 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 coëfficiënt van overeenstemming

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

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

W varieert dus tussen 0 en 1.

De verdeling van S onder de hypothese H_0 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:

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

2°. De z -benadering:

$V = (m-1) \frac{W}{1-W}$ is bij benadering verdeeld als $F = e^{2z}$

(F is de F van Snedecor, z de z van Fisher) met

$$V_1 = n-1 - \frac{2}{m}$$

$$V_2 = (m-1) V_1 \quad \text{vrijheidsgraden ([1] pg. 84 [2] pg. 33-36).}$$

Met behulp van de verdelingen van S of W onder de hypothese H_0 , kan deze hypothese getoetst worden, waarbij H_0 verworpen wordt als W waarden dichtbij 1 (resp. S dichtbij $\frac{1}{12} m^2 (n^3 - n)$) aanneemt, de kritieke z one is dus van de vorm $W \geq W_0$ (resp. $S \geq S_0$).

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 S nu verandert, moeten wij een correctie op de formule voor 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:

$$n = 3 \quad m = 2 \text{ t/m } 10$$

$$n = 4 \quad m = 2 \text{ t/m } 6$$

$$n = 5 \quad m = 3$$

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:

$$n = 3 \quad m = 8, 9, 10, 12, 14, 15, 16, 18, 20$$

$$n = 4 \quad m = 4, 5, 6, 8, 10, 15, 20$$

$$n = 5 \text{ t/m } 7 \quad m = 3, 4, 5, 6, 8, 10, 15, 20$$

op pag. 150.

[2]

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

S47(M18)

Betrouwbaarheidsintervallen (algemeen).¹⁾

Zij \underline{x} een stochastische grootte, die een verdelingsfunctie bezit die, op een onbekende parameter θ na, geheel bekend is (θ kan bv. het gemiddelde van \underline{x} zijn, of de spreiding of iets dergelijks), dan kan men de vraag stellen uit een aantal waarnemingen van \underline{x} een schatting voor θ af te leiden.

Een betrouwbaarheidsinterval \mathcal{I} voor θ is een interval, waarvan de grenzen afhankelijk zijn van de waarnemingen x_1, \dots, x_n van \underline{x} , en dat de eigenschap bezit, behoudens een zekere gegeven onbetrouwbaarheid α , de juiste waarde van θ te bevatten. Dit betekent, dat bij een serie bepalingen van betrouwbaarheidsintervallen slechts in ongeveer een fractie α van deze gevallen het interval \mathcal{I} zo zal uitvallen, dat het θ niet bevat. Hierbij is dus θ constant en het interval \mathcal{I} veranderlijk (en wel stochastisch). Hierin ligt het grote verschil met een zgn. voorspellingsinterval, d.i. een gegeven vast interval, waar een stochastisch punt met een zekere waarschijnlijkheid in valt.

Het algemene principe ter bepaling van een betrouwbaarheidsinterval is het volgende: zij \mathcal{T} een toets voor de hypothese $\theta = \theta_0$ (vgl. S47(M6)), dan is \mathcal{J} de verzameling van die waarden θ_0 die bij toepassing van \mathcal{T} op grond van de gevonden waarnemingen x_1, \dots, x_n niet voor verwerping in aanmerking komen. Is \mathcal{T} toegepast met een onbetrouwbaarheidsdrempel α , dan is dit ook de onbetrouwbaarheidsdrempel van het betrouwbaarheidsinterval.

Litteratuur:

- M.G. Kendall, The Advanced Theory of Statistics, London 1946, deel II, p.62-84.
A.M. Mood, Introduction to the theory of Statistics, London 1950, p.220.
J. Neyman, First course in probability and statistics, N.Y.1950.

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