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 138

Regeneratie van de rattenlever

door

Emily C. Bos-Levenbach

1954

1. Inleiding.

Dit rapport bevat de statistische analyse van een experiment betreffende de regeneratie van de lever van ratten.

Wanneer bij een rat ongeveer 3/4 van de lever operatief verwijderd wordt, treedt een snelle groei van het overgebleven gedeelte op, zodat na één week het oorspronkelijke levergewicht bijna, doch niet geheel, hersteld is.

Bovengenoemd experiment werd uitgevoerd om na te gaan of het terugbrengen, in fijngemaakte toestand, van het weggenomen levergedeelte in de buikholte, invloed uitoefent op het regeneratieproces.

De gang van zaken was als volgt:
Vier ratten van hetzelfde geslacht en vrijwel gelijk gewicht werden gelijktijdig en onder dezelfde omstandigheden geopereerd. Bij twee hiervan werd de fijngesneden lever weer in de buikholte teruggebracht, bij de andere twee niet.

Van deze ratten werd het gewicht van het weggenomen levergedeelte bepaald. Na zeven dagen, doorgebracht onder gelijke levens- en voedselomstandigheden, werd van de overlevende ratten de gehele lever verwijderd en gewogen. Deze gehele proef werd een aantal malen uitgevoerd.

De ons gestelde vraag was, of er na zeven dagen een systematisch verschil in levergewicht is tussen de ratten, waarbij de fijngesneden lever weer in de buikholte is teruggebracht, en de ratten waarbij dit niet gebeurd is.

2. Methode van onderzoek.

2.1. Indeling van het materiaal.

Bij het uitwerken van de gegevens kon geen gebruik gemaakt worden van het feit, dat de proeven gedaan zijn met groepen van vier ratten van gelijk geslacht en ongeveer gelijk gewicht, omdat uit bijna al deze groepen één of meer ratten gestorven zijn.

De wenselijkheid van het werken met deze groepen staat ook nog niet vast, omdat verschillen in lichaamsgewicht en/of geslacht der ratten nog geen verschillen in het gewicht van hun lever behoeven te veroorzaken.

We onderscheiden daarom in hoofdzaak twee groepen ratten en wel:

Groep I: De ratten, waarbij de lever in fijngemaakte toestand in de buikholte teruggebracht is en

Groep II: De ratten, waarbij dit niet gebeurd is, en die dus bij dit experiment de controlegroep vormen.

Het beschikbare materiaal bestaat nu uit 21 ratten in groep I en 24 ratten in groep II, echter is in groep I slechts van 17 ratten het gewicht der weggenomen lever bekend en van 20 dat van de lever na zeven dagen. Er is door de onderzoeker een correctie toegepast om een eventuele invloed van het gewicht van de ratten zo goed mogelijk uit te schakelen.

Deze correctie is verkregen door het relatieve levergewicht in te voeren. Hieronder verstaan wij het levergewicht per gewichtseenheid van de rat.

In een vooronderzoek werd nu nagegaan, of wij de groepen I en II zonder meer tegen elkaar kunnen toetsen om de in de inleiding gestelde vraag te beantwoorden.

2.2. Vooronderzoek.

Bij een vergelijking van de groepen I en II, wat betreft hun levergewicht na zeven dagen, moeten we rekening houden met de volgende factoren, die de uitkomst zouden kunnen beïnvloeden:

1e. het verschil in geslacht;

2e. het verschil in lichaamsgewicht van de rat;

3e. het verschil in het tijdstip waarop de proeven uitgevoerd werden;

4e. het verschil in gewicht van de weggenomen lever.

Voor we dus de groepen I en II vergelijken, zullen we eerst moeten nagaan of, en in hoeverre deze verschillen tussen de ratten onderling invloed hebben op het levergewicht na zeven dagen.

De vragen die zich hierbij voordoen zijn dus:
Hebben de verschillen in bovengenoemde factoren invloed op het resultaat?
Zo ja, kan hiervoor gecorrigeerd worden, zodat de gecorrigeerde grootheden niet meer afhankelijk zijn van deze factoren?

Opmerking.

Bij het onderzoek van het materiaal wordt gebruik gemaakt van statistische methoden, die berusten op het toetsen van bepaalde hypothesen (zie bijlage S 47 (M 6)). Voor de speciale methoden, die werden gebruikt, wordt in de tekst naar de desbetreffende bijlagen verwezen.

3. Resultaten.

3.1. Vooronderzoek.

3.1.a. Invloed van het geslacht.

We beginnen met na te gaan of er verschil is tussen mannelijke en vrouwelijke ratten wat betreft hun gewicht. Hiertoe passen wij de toets van WILCOXON (zie bijlage S 47 (M 7)) toe en wel op de ratten van de groepen I en II gezamenlijk.

Deze toets is zodanig uitgevoerd, dat de waarnemingen betreffende de mannelijke ratten steeds corresponderen met de waarnemingen aangeduid met $x_1, \ldots, x_n$ in de bijlage, en de waarnemingen betreffende de vrouwelijke ratten met $y_1, \ldots, y_m$. Dit is ook het geval bij de andere toepassingen van de toets van WILCOXON in dit rapport, waarbij mannelijke met vrouwelijke ratten vergeleken worden. Indien het resultaat $U > \frac{1}{2}mn$ is, betekent dit, dat in het waarnemingsmateriaal in meer gevallen een mannelijke rat zwaarder is dan een vrouwelijke dan omgekeerd. Wanneer in het eerste geval de overschrijdingskans kleiner is dan 0,05 hechten we aan dit resultaat de conclusie, dat de mannelijke ratten systematisch zwaarder zijn dan de vrouwelijke en dat het geconstateerde verschil niet aan het toeval kan worden geweten. Als $U < \frac{1}{2}mn$ is, geldt uiteraard het tegengestelde.

De resultaten staan vermeld in tabel I.

Tabel I

Vergelijking van mannelijke en vrouwelijke ratten wat betreft hun gewicht

U	$\frac{1}{2}$mn	overschrijdingskans
240	250	0,82

Conclusie:

We zien dus dat er geen reden is om aan te nemen dat er een verschil bestaat tussen mannelijke en vrouwelijke ratten wat betreft hun gewicht.

Wij zullen nu onderzoeken of er een verschil tussen mannelijke en vrouwelijke ratten is, wat betreft het levergewicht na zeven dagen. We passen weer de toets van WILCOXON toe en wel nu voor de groepen I en II apart, omdat het verschil in behandeling tussen de beide groepen invloed op het levergewicht kan hebben. De bedoeling van de aparte behandeling der twee groepen is, deze (eventuele) invloed bij het onderzoek van het onderhavige punt te elimineren. De resultaten voor deze twee groepen worden gecombineerd op de in bijlage S 102 (M 17b) aangegeven wijze. (We kiezen daarbij $c_i = \frac{1}{N_i}$.). Tabel II geeft de resultaten.

Tabel II

Vergelijking van mannelijke en vrouwelijke ratten wat betreft het levergewicht na zeven dagen

groep	U	$\frac{1}{2}$mn	overschrijdingskans
I	53	50	0,84
II	70	70	1,--
gecombineerd			0,88

Er is dus ook geen reden om aan te nemen, dat er verschil is tussen mannelijke en vrouwelijke ratten wat betreft hun levergewicht na zeven dagen.

Conclusie:

Er is geen verband gevonden tussen het geslacht en het lichaamsgewicht van de rat, evenmin tussen het geslacht en het levergewicht na zeven dagen.

Bij het verdere onderzoek hebben wij daarom voorlopig geen onderscheid meer gemaakt tussen mannelijke en vrouwelijke ratten.

3.1.b. Invloed van het lichaamsgewicht.

Vervolgens onderzoeken we of er verband bestaat tussen het gewicht van de rat en het levergewicht na zeven dagen.

We passen daartoe de methode der rangcorrelatie (zie bijlage S 47 (M 13)) toe en wel voor de groepen I en II apart. De resultaten voor de twee groepen worden weer gecombineerd zoals is aangegeven in bijlage S 102 (M 17b) en staan vermeld in tabel III. (We nemen weer $c_i = \frac{1}{N_i}$.)

Tabel III

Onderzoek naar een verband tussen het gewicht van de rat en het levergewicht na zeven dagen

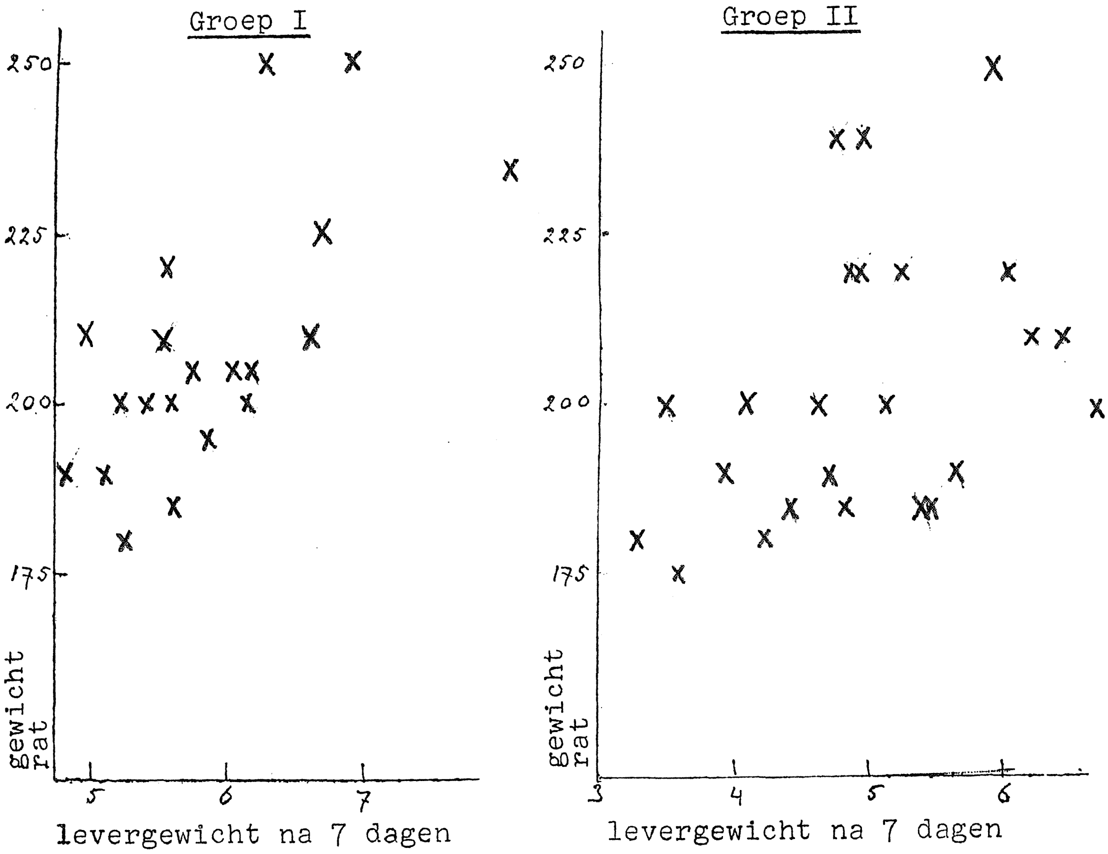
groep	S	overschrijdingskans
I	+ 88	0,004
II	+ 72	0,07
gecombineerd		0,001

Conclusie:

Uit tabel III zien we dat er een verband bestaat tussen het gewicht van de rat en het levergewicht na zeven dagen en, aangezien de gevonden waarden van S beide positief zijn, dat dit levergewicht in het algemeen groter is naarmate het gewicht van de rat groter is. Ter illustratie van dit verband diene figuur 1.

Figuur 1.

Het gewicht van de rat en het levergewicht na zeven dagen



De vraag doet zich nu voor, of de door de onderzoeker gebruikte correctie, die bestaat uit het werken met het relatieve levergewicht in plaats van het absolute, deze afhankelijkheid opheft.

Om dit na te gaan toetsen we de onafhankelijkheid van het gewicht van de rat en het relatieve levergewicht na zeven dagen.

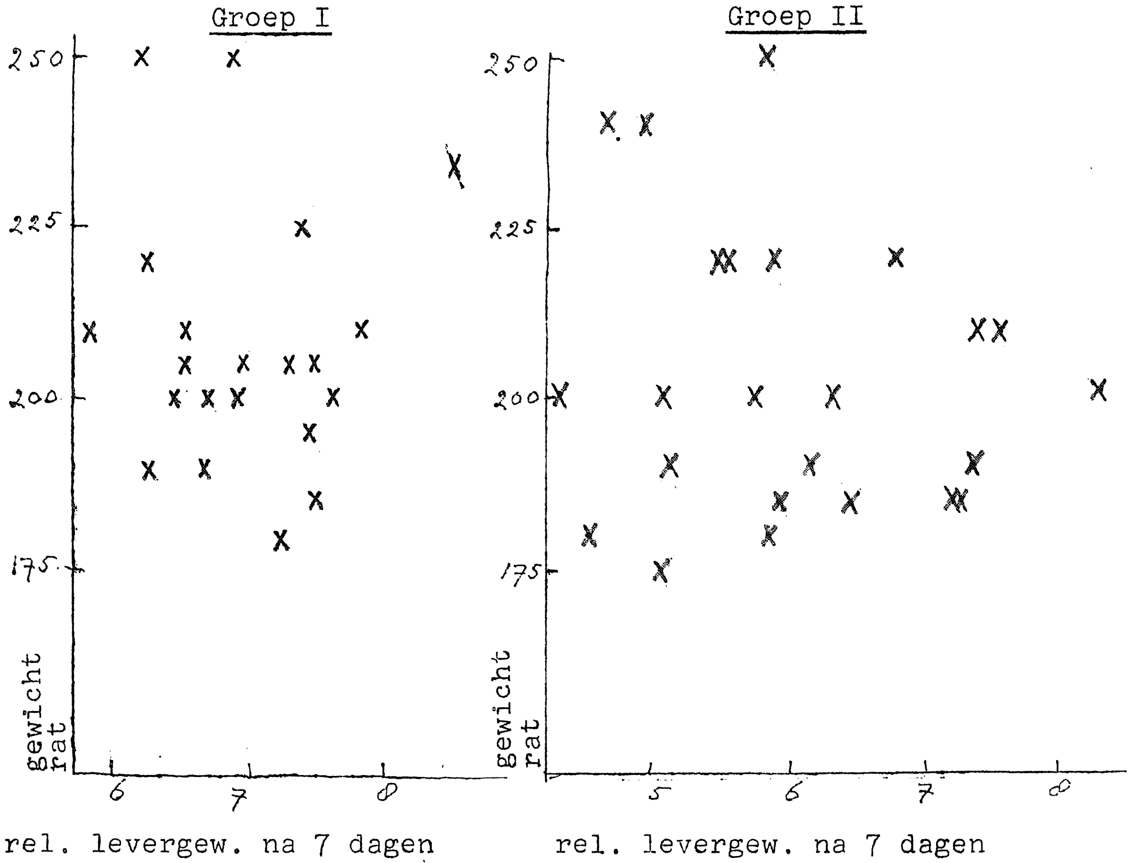
Tabel IV

Onderzoek naar een verband tussen het gewicht van de rat en het relatieve levergewicht na zeven dagen

groep	S	overschrijdingskans
I	- 14	0,67
II	- 15	0,70
gecombineerd		0,56

Conclusie:
We zien dus dat er geen reden is om aan te nemen, dat er een verband bestaat tussen het gewicht van de rat en het relatieve levergewicht na zeven dagen. De nu ontstane situatie wordt geïllustreerd door figuur 2.

Figuur 2.
Het gewicht van de rat en het relatieve levergewicht na 7 dagen



Gaan we nu het relatieve levergewicht na zeven dagen gebruiken om de groepen I en II te vergelijken, dan zullen we eerst moeten nagaan, of mannelijke en vrouwelijke ratten op dit punt niet verschillen; in dat geval zouden we mannelijke en vrouwelijke ratten apart moeten onderzoeken.

Deze vergelijking geschiedt wederom met de toets van WILCOXON (zie tabel V).

Tabel V

Vergelijking van mannelijke en vrouwelijke ratten wat betreft het relatieve levergewicht na zeven dagen

groep	U	$\frac{1}{2}$mn	overschrijdingskans
I	$48\frac{1}{2}$	55	0,67
II	74	70	0,84
gecombineerd			0,88

Conclusie:

Op grond van dit resultaat is er geen aanleiding om onderscheid te maken tussen mannelijke en vrouwelijke ratten wat betreft het relatieve levergewicht na zeven dagen.

3.1.c. Invloed van de proefdata.

We zullen nu onderzoeken of er een verband bestaat tussen de datum waarop de proef is uitgevoerd en het relatieve levergewicht na zeven dagen. Hiertoe wordt, weer voor de groepen I en II apart, de methode der rangcorrelatie toegepast. De resultaten voor de twee groepen worden gecombineerd zoals boven is aangegeven. We vinden dan:

Tabel VI

Onderzoek naar een verband tussen proefdatum en relatief levergewicht na zeven dagen

groep	S	overschrijdingskans
I	- 21	0,53
II	- 6	0,88
gecombineerd		0,59

Conclusie:

We vinden geen verband tussen proefdatum en relatief levergewich na zeven dagen.

3.1.d. Invloed van het gewicht van de weggenomen lever.

Wij zullen nu nog onderzoeken of er een verband is tussen het relatieve levergewicht na zeven dagen en het relatieve gewicht der weggenomen lever. We passen weer de methode der rangcorrelatie toe voor de groepen I en II apart en combineren de resultaten zoals boven aangegeven.

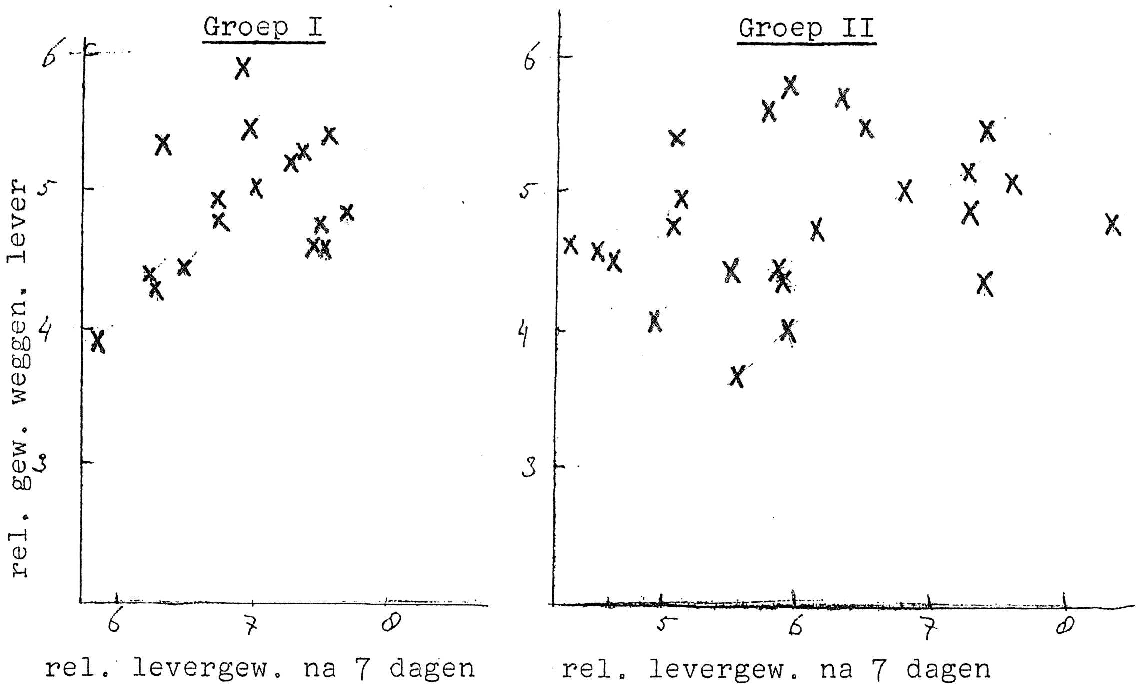
Tabel VII

Onderzoek naar een verband tussen relatief levergewicht na zeven dagen en relatief gewicht der weggenomen lever

groep	S	overschrijdingskans
I	+ 34	0,16
II	+ 46	0,26
gecombineerd		0,08

Conclusie:

In tabel VII kunnen we, wegens de betrekkelijk kleine waarde van de gecombineerde overschrijdingskans, een (zeer zwakke) aanwijzing zien voor de mogelijkheid dat het relatieve levergewicht na zeven dagen positief gecorreleerd is met het relatieve gewicht der weggenomen lever. Deze correlatie zou b.v. kunnen ontstaan, doordat er bij elke operatie naar gestreefd is, 3/4 van de lever weg te nemen, zodat er van ratten met een relatief grote lever in verhouding tot het lichaamsgewicht meer weggenomen wordt dan van ratten met een relatief kleine lever, terwijl na een week ook de ratten, die oorspronkelijk de grootste levers hadden weer grotere hebben dan de ratten, die oorspronkelijk kleine hadden. De situatie wordt geïllustreerd door figuur 3.



Figuur 3. Relatief levergewicht na zeven dagen en relatief gewicht weggenomen lever.

3.1.e. Vergelijking van de relatieve gewichten van het weggenomen deel van de lever.

Indien wij rekening willen houden met de mogelijkheid, dat dit verband inderdaad zou bestaan, is het van belang na te gaan of de te vergelijken groepen I en II systematisch verschillen in het relatieve levergewicht bij het begin van de proef. Is dit nl. het geval, dan zou een verkeerde conclusie bij de vergelijking van de groepen I en II het gevolg kunnen zijn.

De vergelijking van de relatieve levergewichten werd uitgevoerd met de toets van WILCOXON, waarbij nu de getallen $x_1,\dots,x_n$ uit bijlage S 47 (M 7) corresponderen met de waarnemingen van groep I en $y_1,\dots,y_m$ met die van groep II. $U > \frac{1}{2}mn$ betekent nu dus, dat de waarnemingen in groep I systematisch hoger liggen dan in groep II en v.v.

Tabel VIII

Vergelijking der groepen I en II wat betreft het relatieve gewicht der weggenomen lever

U	$\frac{1}{2}$mn	overschrijdingskans
219	204	0,70

Conclusie:

Er is geen reden om aan te nemen dat er een verschil is tussen de groepen I en II wat betreft het relatieve gewicht der weggenomen lever.

3.2. Vergelijking van de relatieve levergewichten der groepen I en II na zeven dagen regeneratie.

Op grond van het vooronderzoek is het gerechtvaardigd de relatieve levergewichten van de groepen I en II na een week regeneratie met elkaar te vergelijken met behulp van de toets van WILCOXON. Het resultaat vindt men in tabel IX.

Tabel IX

Vergelijking der groepen I en II wat betreft het relatieve levergewicht na zeven dagen.

U	$\frac{1}{2}$mn	overschrijdingskans
391	252	$<$ 0,002

Uit tabel VIII blijkt (aangezien $U > \frac{1}{2}mn$ is) dat het relatieve levergewicht na zeven dagen bij groep I groter is dan bij groep II.

Conclusie:

Uit de kleine overschrijdingskans, die hier gevonden wordt, concluderen wij, dat inderdaad het relatieve levergewicht na zeven dagen regeneratie bij de ratten van groep I groter is dan van groep II. Tezamen met de resultaten, samengevat in de tabellen VII en VIII, leidt dit tot de conclusie, dat dit systematische verschil een gevolg moet zijn van de invloed van het terugbrengen van het weggenomen levergedeelte in de buikholte.

Ten slotte werd, om een indruk te krijgen van het quantitatieve verschil tussen de groepen I en II, voor beide groepen het gemiddelde relatieve levergewicht voor en na de proef berekend, waarbij dit gewicht voor de proef geschat werd op 4/3 van het relatieve gewicht van het weggenomen levergedeelte. De uitkomsten staan vermeld in tabel X.

Tabel X

Gemiddelde relatieve levergewichten

groep	gemiddelde relatieve gewicht van het weggenomen gedeelte van de lever	geschat gemiddelde relatieve gewicht van de oorspronkelijke lever	gemiddelde relatieve gewicht van de lever na zeven dagen regeneratie
I	4,90	6,53	6,93 [1)]
II	4,82	6,43	6,05

Uit tabel X zien we dat de aangegroeide lever in groep I gemiddeld groter is dan het geschatte gemiddelde bij het begin van de proef. Daar wellicht de mogelijkheid bestaat, dat in werkelijkheid gemiddeld meer dan 3/4 van de lever weggenomen is aan het begin van de proef, kunnen wij op dit punt niet tot een definitieve conclusie komen. Het waarnemingsmateriaal biedt niet de mogelijkheid om deze vraag nader te onderzoeken. Daarvoor zou een andere proefopzet nodig zijn.

----

1) Het gemiddelde relatieve gewicht van de lever na zeven dagen bij groep I is slechts berekend over de 17 ratten, waarvan ook het gewicht van de weggenomen lever bekend is.

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 $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \ldots, x_n$ 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 kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (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 "significant" 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 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.

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 $\mathcal{H}_0$ 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 $\mathcal{H}_0$ 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 $\mathcal{H}_0$ 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 Boerhaavestr. 49,
A m s t e r d a m - O.

Statistische Afdeling
Rapport S 102 (M 17b)

Het combineren van onafhankelijke toetsen (aanvulling) [1].

In memorandum S 73 (M 17a) wordt een methode voor combinatie van onafhankelijke toetsen behandeld, waarbij het nodig is de overschrijdingskans van iedere toets te bepalen. In vele gevallen kan men de combinatie ook direct op de afzonderlijke toetsingsgrootheden baseren en dit verdient zelfs de voorkeur.

Wij beschouwen hier het geval, dat een bepaalde toets moet worden toegepast op een heterogeen materiaal. Dit materiaal wordt dan eerst verdeeld in h homogeen geachte groepen. Het aantal waarnemingen van de i^e groep zij n_i en de toetsingsgrootheid $\underline{t}_i$ [2]. Laat verder gegeven zijn, dat de verdeling van $\underline{t}_i$ onder de getoetste hypothese (voor de i^e groep aangeduid door H_i) voor grote n_i asymptotisch normaal [3] is, met bekende verwachting μ_i en bekende spreiding σ_i. Aan deze voorwaarden is o.a. voldaan, indien wij te doen hebben met toetsen van WILCOXON, rangcorrelatietoetsen van KENDALL of SPEARMAN, tekentoetsen enz.

Wij toetsen met al de hier te behandelen gecombineerde methoden de hypothese H, dat voor iedere groep de desbetreffende hypothese H_i geldt, terwijl de groepen onderling onafhankelijk zijn. De toetsen verschillen echter ten aanzien van de alternatieve (van H afwijkende) hypothesen waarvoor zij gevoelig [4] zijn.

De meest gebruikelijke toetsingsgrootheden van gecombineerde toetsen zijn van de gedaante:

$$\underline{T} = \sum_{i=1}^{h} c_i(\underline{t}_i - \mu_i)$$

waarin de letters c_i ($i = 1, 2, \ldots, h$) constanten voorstellen, die voor ieder van de combinatiemethoden op een bepaalde wijze

---

1) Dit memorandum dient slechts ter oriëntatie en streeft niet naar volledigheid of volledige exactheid. Het is bedoeld als een aanvulling op Rapport S 73 (M 17a).

2) De onderstreping geeft aan dat een toetsingsgrootheid stochastisch is, d.w.z. een waarschijnlijkheidsverdeling bezit.

3) Dit houdt in dat $\underline{t}_i$ een waarschijnlijkheidsverdeling heeft, die als n_i toeneemt, steeds minder van een normale verdeling (verdeling van Gauss) afwijkt.

4) Een toets van hypothese H is gevoelig ten opzichte van een alternatieve hypothese H', als de kans dat H verworpen wordt, indien H' juist is, groot is.

gekozen worden. Onder de hypothese H zal $\underline{T}$ asymptotisch (voor grote h en/of grote n_i) normaal verdeeld zijn met verwachting 0 en spreiding $\sigma = \sqrt{\sum_{i=1}^{h} c_i^2 \sigma_i^2}$. De dubbele overschrijdingskans van een gevonden waarde T van $\underline{T}$ is dus bij benadering gelijk aan:

$$\frac{2}{\sqrt{2\pi}} \int_{\left|\frac{T}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2}\, dx$$

en kan bepaald worden met behulp van een tabel van de normale verdeling. Indien de dubbele overschrijdingskans kleiner is dan de onbetrouwbaarheidsdrempel α , zal men H verwerpen.

Wij geven hier 3 combinatiemethoden van dit type:

<u>Methode 1</u>: $c_1 = c_2 = \cdots\cdots = c_h = 1$

dus: $\underline{T} = \sum_{i=1}^{h} \underline{t}_i - \sum_{i=1}^{h} \mu_i$; $\sigma = \sqrt{\sum_{i=1}^{h} \sigma_i^2}$

<u>Methode 2</u>: $c_1 = \frac{1}{\sigma_1}$, $c_2 = \frac{1}{\sigma_2}$, ..., $c_h = \frac{1}{\sigma_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{\sigma_i}$; $\sigma = \sqrt{h}$

<u>Methode 3</u>: $c_1 = \frac{1}{n_1}$, $c_2 = \frac{1}{n_2}$,....., $c_h = \frac{1}{n_h}$

dus: $\underline{T} = \sum_{i=1}^{h} \frac{\underline{t}_i - \mu_i}{n_i}$; $\sigma = \sqrt{\sum_{i=1}^{h} \left(\frac{\sigma_i}{n_i}\right)^2}$

Deze methoden zijn alleen gevoelig ten aanzien van alternatieve hypothesen volgens welke de grootheden $\underline{t}_i$ verdelingen hebben die, voor zover zij afwijken van de verdelingen onder de corresponderende hypothesen H_i, dit over het algemeen in dezelfde richting doen. Men zal dan <u>methode 1</u> bij voorkeur toepassen als men aan de $\underline{t}_i$ met een kleine spreiding (in de regel zullen dat de $\underline{t}_i$ van kleine groepen zijn) een geringer gewicht wil toekennen dan aan de $\underline{t}_i$ met een grote spreiding. De <u>methoden 2 en 3</u> zijn te gebruiken als men aan de verschillende groepen waarnemingen, ongeacht hun grootte, een ongeveer gelijke invloed op het resultaat wil toekennen. De keuze tussen deze twee methoden hangt verder van hier niet te behandelen theoretische overwegingen af (zie literatuur [1]).

Indien men verwacht dat mogelijke verschuivingen van de verdelingen der $\underline{t}_i$ in beide richtingen kunnen liggen, verdient het de voorkeur om gebruik te maken van de volgende toetsingsgrootheid:

$$\sum_{i=1}^{h} \left(\frac{\underline{t}_i - \mu_i}{\sigma_i}\right)^2 \qquad \text{(\underline{methode 4})}$$

Deze grootheid is onder de hypothese H asymptotisch verdeeld volgens een χ^2-verdeling met h vrijheidsgraden. De overschrijdingskans van een gevonden waarde van deze grootheid kan dus met behulp van een tabel van de χ^2-verdeling bepaald worden.

De toets, behandeld in memorandum S 73 (M 17a) par. 1, waarbij men het product van linkszijdige en product van alle rechtszijdige overschrijdingskansen bepaalt en het kleinste van deze twee producten gebruikt, heeft betrekking op dezelfde gevallen als de hier behandelde methoden 2 of 3, terwijl de methode, behandeld in S 73 (M 17a) par. 2, berustend op het product van de tweezijdige overschrijdingskansen, meer overeenkomt met methode 4. Men mag echter verwachten, dat, zo aan de asymptotische normaliteit der $\underline{t}_i$ voldaan is, de in dit memorandum behandelde methodenscherper zijn dan de toetsen behandeld in S 73 (M 17a).

Literatuur:

1 C.van Eeden, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Rapport S 115 (M 45) van het Mathematisch Centrum (1953).

2 -----------, Trendtoets met behulp van rangcorrelatie, Memorandum S 73 (M 13a). (Voorbeeld van toepassing van methode 1.)

3 Dr J.Hemelrijk, Het combineren van onafhankelijke toetsen, Memorandum S 73 (M 17a).

Mathematisch Centrum,
2de Boerhaavestraat 49,
Amsterdam O.
Statistische Afdeling,
S47 (M7).

Maart, 1952.

De toets van Wilcoxon.[1])

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

Voor het toetsen van de hypothese H_o wordt gebruik gemaakt van een toetsingsgrootheid $\underline{U}$ [2]), die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ 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 $x_3,x_4,\ldots,x_n$ de aantallen $V_3,V_4,\ldots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

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

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

---

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

[2]) Stochastische grootheden worden door onderstreping aangeduid.

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

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

en

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

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

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

$$\frac{|U-\mu|}{\sigma} > \xi_\alpha\,, \qquad ^{2)}$$

waarin α de onbetrouwbaarheidsdrempel is en ξ_α volgt uit

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

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

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

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

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

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

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

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

---

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

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

Indien n en m kleiner zijn dan 10, zijn tabellen beschikbaar voor het berekenen van de overschrijdingskans k voor de uit de steekproef bepaalde waarde U van U (zie [2] en [4]). Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

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

Litteratuur:

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

MATHEMATISCH CENTRUM
2de Boerhaavestr. 49,
Amsterdam -O.
Statistische Afdeling
S47(M13)

Rangcorrelatie[1)]

1. Beschrijving van de methode.

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

De stochastische grootheden $\underline{x}$ en $\underline{y}$ bezitten een simultane verdeling. Over deze verdeling zelf behoeft niets ondersteld te worden.

(x_i, y_i) $(i = 1,\ldots,n)$, 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 toetsingsgrootheid 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 $\underline{S}$ voor het geval dat $\underline{x}$ en $\underline{y}$ onafhankelijk zijn is bekend (zie § 2). De hypothese dat $\underline{x}$ en $\underline{y}$ 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

i	j	Correlatie
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 $|S| \geqq S_0$, en bij éénzijdige toetsing van de vorm $S \geqq S_0'$ (rechtszijdige toetsing) of $S \leqq S_0''$ (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 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.

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

Voor grote waarden van n is de verdeling van $\frac{S}{\sigma_S}$ (waarin σ_S de spreiding van 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 $h = 1,2,\ldots,k_1$. Evenzo doet men in de rij der waarnemingen y_j, waar we de overeenkomstige aantallen aanduiden met u_ℓ waarin $\ell = 1,2,\ldots,k_2$. σ_S kan dan gevonden worden uit de volgende formule:

$$(1)\quad \sigma_S^2 = \frac{1}{18}\Big\{ n(n-1)(2n+5) - \sum_{h=1}^{k_1} t_h(t_h-1)(2t_h+5) - \sum_{\ell=1}^{k_2} u_\ell(u_\ell-1)(2u_\ell+5)\Big\} + \frac{1}{9n(n-1)(n-2)}\sum_{h=1}^{k_1} t_h(t_h-1)(t_h-2)\sum_{\ell=1}^{k_2} u_\ell(u_\ell-1)(u_\ell-2) + \frac{1}{2n(n-1)}\sum_{h=1}^{k_1} t_h(t_h-1)\sum_{\ell=1}^{k_2} u_\ell(u_\ell-1).$$

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:

$t_1(t_1-1)(2t_1+5) = 2.1.9 = 18$
$u_1(u_1-1)(2u_1+5) = 3.2.11 = 66$
$t_1(t_1-1)(t_1-2) = 0$, $u_1(u_1-1)(u_1-2) = 6$
$t_1(t_1-1) = 2.1 = 2$
$u_1(u_1-1) = 3.2 = 6$
$n(n-1)(2n+5) = 6.5.17 = 510$
$n(n-1) = 6.5 = 30$

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 aan 1 en er dus in geen van beide rijen gelijken voorkomen, gaat formule (2) over in:

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

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

3. Rangcorrelatiecoëfficiënt

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

$$(3)\quad \tau = \frac{2S}{\left\{n(n-1)-\sum_{h=1}^{k_1} t_h(t_h-1)\right\}^{\frac{1}{2}}\left\{n(n-1)-\sum_{l=1}^{k_2} u_l(u_l-1)\right\}^{\frac{1}{2}}}.$$

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

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

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 "Parametervrije 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. Biometrica 34 (1947) p. 36-40.