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 103

Onderzoek van verschillen tussen muizenstammen naar de hoeveelheden ACTH in de hypophyse-voorkwab.

door

H.J. Prins en Prof. Dr J.Hemelrijk

1953

1.0 Inleiding.

Gezocht werd naar een methode om te bepalen, of verschillen tussen de hoeveelheden ACTH in de hypophyse-voorkwab (HVK) van enige muizenstammen aan te geven zijn en of deze in verband kunnen worden gebracht met de kankereigenschappen van deze stammen. In eerste instantie ligt het voor de hand een ijkingsmethode te zoeken om hoeveelheden ACTH te bepalen. Dit blijkt een biologische ijkingsmethode te moeten zijn.

Enige ijkingsmethoden worden beschreven door C.W.EMMENS [1] en in een artikel van M.A.SAYERS [2]. Deze methoden hebben alle het bezwaar de sterkte van een bepaalde werkingscomponent van het ACTH te bepalen, zodat monsters onderzocht met de ene methode mogelijk heel andere resultaten opleveren dan monsters onderzocht met een andere methode.

De door SAYERS [2] beschreven ijkingsmethode kwam uit de in de litteratuur beschreven methoden als vermoedelijk de nauwkeurigste naar voren en werd daarom door de onderzoeker gekozen. SAYERS gebruikt de daling van het ascorbinezuurgehalte van de bijnier van een rat na injectie met ACTH als reactiemaat. Deze rat moet van te voren gehypophysectomeerd zijn om te voorkomen dat de ACTH-productie van de hypophyse van invloed is. Bij deze ijking wordt een dosis ACTH per 100 gr. lichaamsgewicht van de rat ingespoten. Van te voren is de rechter bijnier verwijderd. Enige uren na inspuiting wordt de linker bijnier verwijderd. Het verschil in hoeveelheid ascorbinezuur in γ per mg weefsel van de bijnier tussen linker en rechter bijnier wordt als reactiemaar gebruikt. Deze reactiemaat is vrij goed lineair afhankelijk van de logarithme van de per 100 gr. lichaamsgewicht ingespoten dosis.

In het onderhavige geval werd, om een reactiemaat te krijgen evenredig met de hoeveelheid ACTH per gewichtseenheid HVK van een muis, de rat per 100 gr. lichaamsgewicht ingespoten met een met het HVK-gewicht van de muis omgekeerd evenredige verdunning van de uit deze HVK gemaakte oplossing.

1.1 Proefopzet.

Bij het vaststellen van de proefopzet kwamen twee problemen van geheel verschillende aard naar voren, nl.:

1. dat het opzetten van een volledige ijking van voldoende nauwkeurigheid volgens de methode van SAYERS, gezien de beperkte tijd die voor het onderzoek beschikbaar was, vrijwel ondoenlijk zou zijn; er zijn dan nl. uitgebreide voorproeven nodig om de te gebruiken ratten op bepaalde eigenschappen te onderzoeken. Bovendien zou, ook na dit onderzoek, de proef zelf op vrij grote schaal moeten worden opgezet, hetgeen bezwaarlijk was. Dit leidde

tot het opstellen van een proefschema, dat eenvoudiger is en dat minder waarnemingen vereist.

2. Verder kwam de vraag naar voren, welke maat men moest kiezen voor het vóórkomen van ACTH bij de verschillende muizenstammen. Hiervoor wordt in de litteratuur meestal de totale hoeveelheid ACTH in de hypophyse-voorkwab genomen. Het gebruik van deze grootheid alleen heeft echter het bezwaar, dat de gewichten der HVK's bij verschillende muizenstammen niet gelijk zijn. Indien nu de concentratie van het ACTH in de HVK voor deze stammen wel hetzelfde zou zijn, zou men eigenlijk door de gehele ijking slechts op ingewikkelde wijze het gewicht van de HVK der muizen bepalen. Daarom werd besloten het onderzoek ook te richten op de concentratie van ACTH in de HVK, hoewel men tegen het gebruik van deze grootheid alleen een soortgelijke tegenwerping zou kunnen maken als tegen het gebruik van de totale hoeveelheid als enige maat. Onder de concentratie wordt hier verstaan de hoeveelheid ACTH per gewichtshoeveelheid weefsel. Het is dan noodzakelijk de HVKen te wegen. Aangezien er niets bekend was over de mogelijkheid om verschillen in HVK-gewichten te bepalen en de spreiding van de wegingsuitkomsten bij de gebruikte balans ongeveerd 10% van het totale HVK-gewicht was, werd een voorproef gedaan om uit te maken of het wegen van de HVK wel voldoende goed uitvoerbaar was (zie par. 1.2).

1.2 Voronderzoek.

Om na te gaan of het wegen van de HVK behoorlijk uitvoerbaar was, werd uitgegaan van een veronderstelling, nl. dat jongere muizen een kleiner HVK-gewicht hebben dan oudere. Kan dit inderdaad worden aangetoond op grond van de gebruikte operatietechniek en met de beschikbare balans, dan volgt daaruit dat dez: nog zo slecht niet zijn; er is dan althans aangetoond, dat de meettechniek het ontdekken van verschillen mogelijk maakt.

Van 3 leeftijdsgroepen van eenzelfde stam werden ieder 18 dieren genomen [1]), waarvan de lichaamsgewichten en de HVK-gewichten bepaald werden (de eerste met een spreiding van ongeveer 3%, de tweede met een spreiding van ongeveer 10% van de te meten waarde). Op de zo verkregen HVK-gewichten werd de toets tegen verloop van T.J.TERPSTRA toegepast (zie memorandum S 47 (M 6) en S 73 (M 28)) [2]). Wij vonden daarbij een overschrijdingskans

---

1) Van één leeftijdsgroep 21 i.p.v. 18. Toen bleek dat van sommige dieren de leeftijd niet bekend was, werden enige dieren meer genomen waarvan de leeftijd geschat werd.

2) De vermelde memoranda zijn aan het einde van dit rapport toegevoegd.

0,046; de kans op een dergelijk resultaat zou, als er geen verloop in het HVK-gewicht is bij toenemende leeftijd, of ook als de onnauwkeurigheid van de gewichtsbepaling dit verloop volledig zou verdoezelen, kleiner dan 1 op 20 zijn geweest. Men beschouwt een dergelijke uitkomst gewoonlijk als een aanwijzing vóór de aanwezigheid van (in dit geval) een verloop, hetgeen dus betekent, dat weging van de HVK niet ondoenlijk behoeft te worden geacht.

Verder werd onderzocht of lichaamsgewicht en hypophysegewicht met elkaar in verband staan (men kan een positieve correlatie tussen deze twee gewichten verwachten). Hiertoe werd de methode der rangcorrelatie (zie bijgevoegd memorandum S 73 (M 13b)) gebruikt. De waarnemingen werden - ongeacht de leeftijdsgroep - gerangschikt naar hypophysegewicht en lichaamsgewicht. Dit gaf een overschrijdingskans van 0,003; op grond van deze kleine waarde der overschrijdingskans kunnen wij zonder bezwaar besluiten tot de aanwezigheid van de vermoede correlatie.

Uit deze twee onderzoekingen blijkt, dat de weging van de HVK's voldoende nauwkeurig is om bruikbaar te zijn voor het aantonen van systematische verschillen in HVK-gewicht, zoals die bij de gebruikte muizen voorkomen, zodat het inderdaad mogelijk moet zijn ook de concentraties aan ACTH in het onderzoek te betrekken.

2.0 Het gebruikte proefschema.

Om tegemoet te komen aan de in par. 1.1 gesignaleerde moeilijkheden werd een proefschema gebruikt waarbij de gehele proef in een aantal kleine proeven werd verdeeld, die ieder op één dag konden worden voltooid. Daardoor werd bereikt, dat eventuele "dagverschillen" op het uiteindelijk resultaat geen invloed meer konden uitoefenen. De resultaten werden namelijk slechts vergeleken met andere resultaten van dezelfde dag, terwijl ten slotte de uitkomsten der verschillende dagen werden gecombineerd. De gevolgde methode wordt hieronder nog verder beschreven.

Er werden 4 muizenstammen onderzocht, nl. O 20, A, C57 en dba. Van ieder der stammen werd dagelijks één muis gebruikt. Daar het er om ging te onderzoeken of de ACTH concentraties van de HVK bij de 4 onderzochte stammen verschillend zijn, kunnen wij de methode van m rangschikkingen (zie memorandum S 47 (M 14)) gebruiken, indien wij voor ieder der dagen een statistische schatting van de volgorde naar grootte van de ACTH-concentraties der 4 onderzochten muizen kunnen vaststellen. Hiervoor is het niet nodig de ijkingen volgens SAYERS volledig uit te voeren, hetgeen een groot voordeel is. Aan elk van de 4 muizen werd namelijk op aselecte wijze (met behulp van Latijnse vierkanten) een tijdstip

voor behandeling een een rat toegevoegd. Op de 4 ratten van één dagproef werd de methode van SAYERS toegepast. De volgorde van de bij deze 4 ratten gevonden reacties (vgl. par. 1.0) op de inspuiting met het HVK-extract van de bijbehorende muis is nu het gegeven dat verder wordt gebruikt. Omzetting van deze reacties in geschatte ACTH-hoeveelheden met behulp van een ijklijn, waarvan de bepaling veel waarnemingen zou vergen, verandert niets aan deze volgorde. Beperken wij ons dus tot statistische methoden waarbij slechts deze volgorde van belang is, dan wordt de bepaling van deze ijklijn overbodig. Dit is bij de methode der m rangschikkingen inderdaad het geval.

Om enig idee te krijgen hoeveel proeven met 4 muizen van verschillende stammen nodig zouden zijn om het veronderstelde effect inderdaad te vinden, werd de in de volgende paragraaf beschreven voorproef gedaan.

Aangezien de experimentator meer dan 4 ratten per dag kon verwerken werden er nog 2 in de proef opgenomen en ingespoten met bekende doses ACTH. Hierdoor werd het mogelijk een schatting van de ijklijn te maken en zodoende iets te weten te komen over de totale hoeveelheid ACTH in de HVK van de stammen en over de verschillen tussen de stammen in dit opzicht.

Om eventuele invloeden van het tijdstip van behandeling te elimineren werd de rat, ingespoten met de kleinste bekende dosis, de ene dag vóór alle andere, de volgende dag ná alle andere behandeld en de rat, ingespoten met de grootste bekende dosis ACTH, juist in omgekeerde zin.

2.1 Het tweede vooronderzoek.

Twee stammen, waartussen, op grond van tellingen van de (vermoedelijk) ACTH-producerende cellen in de HVK, het grootste verschil naar voorkomen van het ACTH verwacht werd, werden voor dit onderzoek gebruikt. Overigens werd het onderzoek geheel analoog aan het ten slotte uitgevoerde proefschema uitgevoerd (zie par. 2.0). Alleen het toewijzen van een tijdstip van behandeling en een rat op aselecte wijze werd achterwege gelaten.

Het onderzoek bestond uit enige proeven, waarbij, telkens een muis van iedere stam (O 20 en dba) gebruikt werd. De ratten werden ingespoten met een pareparaat van een vaste gewichtshoeveelheid (1/40 mg) HVK-weefsel van de bijbehorende muis. Bij iedere proef werd nagegaan welke rat de grootste reactie ondervond en tot welke stam de muis hoorde, waarvan het HVK-praeparaat bij deze rat was ingespoten.

Van de 7 proeven gaven in 5 gevallen de muizen van stam O 20 een lagere reactie dan die van stam dba. In één geval waren de reacties gelijk en in één geval was het effect omgekeerd. De toe-

gepaste statistische toets - de tekentoets (memorandum S 53 (M 22)) - leverde een tweezijdige overschrijdingskans 0,22. Het effect is niet duidelijk, maar bij een zo klein aantal proeven is dat ook niet te verwachten. De 5 van de 6 proeven, waarbij stam 20 een lagere uitslag gaf (in deze richting werd het effect inderdaad verwacht) doen echter vermoeden dat het niet zinloos is de proef op iets grotere schaal uit te voeren. Voor de definitieve proef was nog slechts een beperkte tijd beschikbaar. In totaal werden er met alle 4 muizenstammen nog 9 proeven voltooid.

Om uit te maken welke maat voor het voorkomen van ACTH bij de definitieve proeven het best gebruikt zou kunnen worden, werd nog nagegaan hoe de resultaten voor het HVK-gewicht bij het voor onderzoek lagen. In alle zeven gevallen bleek het HVK-gewicht bij de muizen van stam O 20 groter te zijn dan bij de muizen van stam dba (overschrijdingskans 0,006 bij toepassing van de tekentoets). Dit effect van het HVK-gewicht doet verwachten dat de verschillen in hoeveelheden ACTH geringer zouden zijn, dan de verschillen in ACTH concentraties. Daarom werd dan ook besloten de ratten in te spuiten met praeparaten van een vaste gewichtshoeveelheid HVK-weefsel (dit zijn dus doses ACTH evenredig met de concentratie) en niet met een praeparaat van hele HVK's (dos evenredig met de hoeveelheid ACTH in de HVK).

Voorts werd een onderzoek ingesteld om tot een schatting v het lineariteitsgebied van de ijklijn te komen. In het lineariteitsgebied verloopt de ijklijn nl. het steilst, zodat bij een bepaald doses-verschil het grootste reactieverschil optreedt.

Hiervoor werd gebruik gemaakt van een serie verdunningen (1/4, 1/8, 1/16, 1/32, 1/64 mg/100 gr [3])) van eenzelfde HVK-praeparaat. Het gunstigste gebied bleek in de omgeving van 1/40 mg/100 gr. [3]) te liggen. Daarom werd deze dosering bij de uitgevoerde proef gebruikt. Hiermee was het lineariteitsgebied dus ongeveer bepaald ten opzichte van de hoeveelheid ACTH in een muizen-HVK. Daar echter niet bekend was hoeveel ACTH de HVK ongeveer bevatte, was hiermee nog niet bepaald met hoeveel γ/100 zuiver ACTH een HVK-praeparaat van 1/40 mg/100 gr. [3]) correspondeert. De juiste doses voor de schatting van de ijklijn gedurende de proef, zoals beschreven aan het eind van par. 2.0, moesten dus gedurende de proef bepaald worden. Als eerste poging werd hierbij gebruik gemaakt van de waarden die SAYERS vond voor de

---

3) mg/100 gr. is een verkorte schrijfwijze voor: mg HVK per 100 gr. lichaamsgewicht van de rat.

hoeveelheid ACTH, waarbij de ijklijn lineair is. Deze doses ACTH bleken een te sterke reactie te geven in vergelijking met de reacties op de HVK-doses. Daarom werd direct na de eerste proef op kleinere doses ACTH overgegaan en deze werden voor de overige proeven gehandhaafd (zie ook tabel I, par. 2.2).

2.2 De waarnemingsresultaten van de uitgevoerde proef.

In tabel I is het bij de definitieve proef verkregen waarnemingsmateriaal schematisch weergegeven.

Tabel I

Resultaten van de uitgevoerde proef

volgorde der proefdagen		Stammen				reacties op bekende doses ACTH			
						doses ACTH			
		0 20	C57 bl	dba	A	0,1	1,0	0,2	2,0
	lich.gew.	27	31	26	28				
I	HVK-gew.	1,08	1,10	0,68	0,90				
	reactie	- 0,33	- 0,80	- 0,98	- 0,82	-	-	- 1,10	- 1,86
	lich.gew.	26	23	23	33				
II	HVK-gew.	1,22	1,12	0,68	0,80				
	reactie	- 0,29	- 0,89	- 0,92	- 1,22	- 1,04	- 1,60		
	lich.gew.	23	30	27	36				
III	HVK-gew.	1,04	1,28	0,66	0,92				
	reactie	- 0,29	- 0,93	- 0,93	- 1,92	- 0,80	- 1,88		
	lich.gew.	27	35	21	34				
IV	HVK-gew.	1,28	1,32	0,70	0,82				
	reactie	- 0,49	- 0,96	- 1,11	- 1,56	-	- 1,49		
	lich.gew.	25	34	23	33				
V	HVK-gew.	1,04	0,68	0,78	0,90				
	reactie	- 1,27	- 1,52	- 1,45	- 1,60	- 0,96	- 1,65		
	lich.gew.	26	29	26	30				
VI	HVK-gew.	1,50	0,98	0,58	0,64				
	reactie	- 0,36	- 1,64	- 1,44	- 0,87	-	-		
	lich.gew.	25	35	26	36				
VII	HVK-gew.	0,94	1,28	0,96	0,70				
	reactie	-	- 1,53	- 1,11	- 1,48	- 1,33	- 1,68		
	lich.gew.	25	34	23	27				
VIII	HVK-gew.	1,10	1,20	0,88	0,72				
	reactie	- 1,05	- 1,04	- 1,52	- 1,53	- 0,62	- 1,57		
	lich.gew.	25	31	23	25				
IX	HVK-gew.	1,04	1,20	0,82	0,90				
	reactie	- 0,97	- 1,65	- 1,80	- 0,93	- 0,95	- 1,41		

Opmerkingen.

1. Zoals boven reeds gezegd, werd bij ieder der proeven 1/40 mg HVK per 100 gr. rat ingespoten. Het lichaamsgewicht van de muis is in de tabel opgegeven in grammen, het HVK-gewicht in mg, de reactie is gegeven als verschil in hoeveelheid ascorbinezuur per gewichtshoeveelheid tussen linker en rechter bijnier van de rat in γ/mg.
2. Waarnemingen die op grond van afwijkende behandeling gedurende de proef, door de experimentator onbetrouwbaar worden geacht zijn weggelaten.
3. Enkele andere waarnemingen vielen uit doordat de betrokken rat dood ging gedurende de behandeling of doordat niet voldoende ratten behandeld konden worden.
4. Bij toepassing van de methode der m rangschikkingen wordt uitsluitend gebruik gemaakt van puntschattingen, zodat spreidingen ten gevolge van meetfouten voor de statistische bewerking onbelangrijk zijn.

3.0 Statistische bewerking van het materiaal.

Met de methode der m rangschikkingen werd onderzocht of uit het materiaal bleek dat de stammen te onderscheiden zijn naar verschillen in:

a) 1 de lichaamsgewichten,
   2 de HVK-gewichten,
   3 de ACTH-concentratie in de HVK,
   4 de hoeveelheid ACTH in de HVK,
   5 de hoeveelheid ACTH in de HVK gedeeld door het lichaamsgewicht.

Punt 5 werd onderzocht om een indruk te krijgen van de hoeveelheid ACTH die aan de verschillende weefsels wordt toegevoegd.

Voorts werd onderzocht

b) 1 of een verband bestond tussen HVK-gewicht en lichaamsgewicht en
   2 of overeenstemming in volgorde bestond bij rangschikking van de stammen naar concentratie van het ACTH en naar de totale hoeveelheid ACTH in de HVK.
   3 eveneens naar concentratie van het ACTH en naar de hoeveelheid ACTH in de HVK gedeeld door lichaamsgewicht.

a. Onderzoek met de methode der m rangschikkingen.

De grootheden, die voor de toepassing van de methode der m rangschikkingen zijn gebruikt zijn de volgende:

1. Voor het lichaamsgewicht van de muizen: deze lichaamsgewichten zelf;
2. Voor de HVK-gewichten: deze gewichten zelf;

3. Voor de ACTH-concentratie in de HVK: de reacties uit tabel I De volgorde naar grootte van deze reacties is nl. een schatting van de volgorde naar grootte van de ACTH-concentraties, omdat de ingespoten dosis evenredig met deze concentratie is.
4. Voor de totale hoeveelheid ACTH in de HVK moet gebruik gemaakt worden van een schatting van de ijklijn. Deze schatting werd verkregen door het gemiddelde van de reacties bij de dosis 0,1 γ resp. bij 1,0 γ ACTH te nemen en de zo verkregen punten door een rechte lijn te verbinden. Vervolgens kunnen de gevonden reacties in schattingen van de concentraties worden omgezet en deze, door vermenigvuldiging met het HVK-gewicht, in schattingen van de totale hoeveelheid ACTH in de HVK. Deze laatste schattingen kunnen dan gerangschikt worden voor toepassing van de methode der m rangschikkingen.
5. Voor de hoeveelheid ACTH in de HVK gedeeld door het lichaamsgewicht van de muis: de schattingen van de hoeveelheid, beschreven onder punt 4, gedeeld door het lichaamsgewicht.

Met de methode der m rangschikkingen wordt onderzocht of de bij de verschillende proeven gevonden rangschikkingen der stamme naar een bepaald kenmerk onderling overeenstemmen.

Statistisch komt dit neer op het toetsen van de nul-hypothe H_0: "de rangschikking naar grootte van de waarnemingen is bij ieder der proeven een toevallige" tegen de alternatieve hypothes H: "er bestaat bij de verschillende proeven een gemeenschappelij voorkeur voor een bepaalde rangschikking naar grootte der waarnemingen".

Bij verwerping van de nulhypothese levert volgorde naar grootte van de kolomtotalen uit het schema (zie memorandum S 47 (M 14)) een goede schatting voor de rangschikking der stammen naar het beschouwde kenmerk.

Men dient hierbij in het oog te houden, dat de gebruikte toetsingsmethode ook onderscheidend is voor gevallen waarin niet alle stammen naar het beschouwde kenmerk te onderscheiden zijn, maar één of meer ervan duidelijk verschillen van de overige. Een beschouwing van de kolomtotalen kan een aanwijzing geven welke verschillen tussen stammen of groep van stammen voornamelijk ve antwoordelijk zijn voor gevonden kleine overschrijdingskansen. Bij gering verschil tussen twee kolomtotalen zal men niet genei zijn te concluderen dat deze stammen naar bedoeld kenmerk te onderscheiden zijn. Verschillen de kolomtotalen echter zeer veel dan zal men geneigd zijn te veronderstellen dat een duidelijk onderscheid te maken is naar het beschouwde kenmerk tussen die stammen.

In tabel II zijn de resultaten van de statistische bewerkir

met de methoden der m rangschikkingen weergegeven. Voor ieder kenmerk zijn de kolomtotalen en de daaruit volgende schatting voor de volgorde gegeven. Onder punt 6 is de rangschikking gegeven, zoals die verwacht werd op grond van telling van de - vermoedelijk - ACTH producerende cellen in de HVK van muizen van de verschillende stammen. Het is niet duidelijk of deze rangschikking een verwachting is voor de rangschikking naar concentratie van het ACTH in de HVK of voor de rangschikking naar de hoeveelheid ACTH in de HVK.

Tabel II

Schematisch overzicht van het onderzoek met behulp van de methode der m rangschikkingen

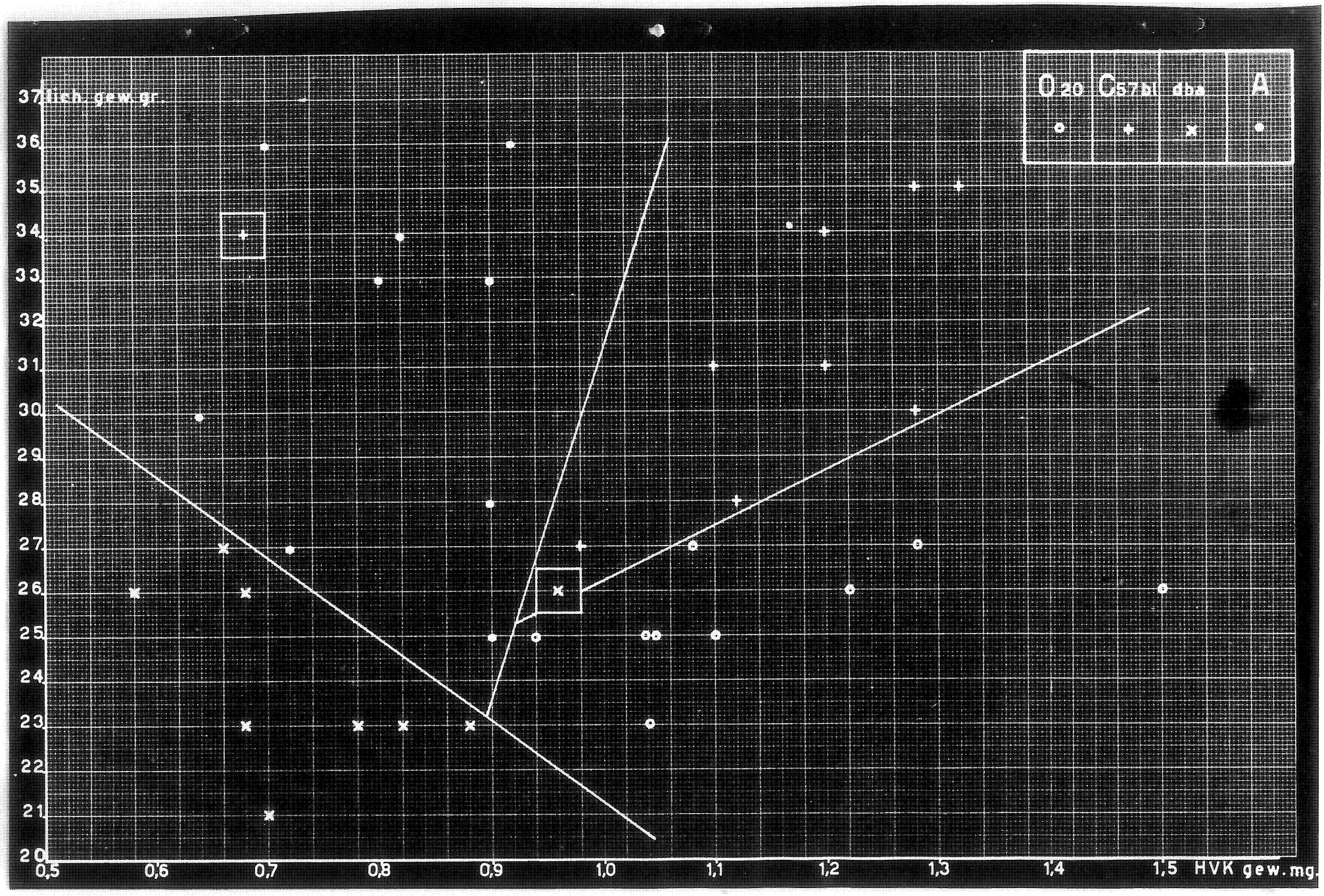
nr.	stam / kenmerk	O 20	C57 bl	dba	A	overschrijdingskans
1	lichaamsgewicht	16 2	32 4	11½ 1	30½ 3	10^{-4}
2	HVK-gewicht	29 3	31 4	13 1	17 2	0.0012
3	concentratie van het ACTH in de HVK	9½ 1	20 2	25 3	25½ 4	0.005
4	hoeveelheid ACTH in de HVK	9 1	24 3½	23 2	24 3½	0.006
5	hoeveelheid ACTH in HVK per gr. lich. gewicht	10 1	21 2	25 4	24 3	0.012
6	verwacht op grond van celtelling in de HVK	1	3	4	2	

Conclusies.

De resultaten van de statistische bewerking weergegeven in tabel II kan men op grond van de hierboven gegeven overwegingen op de hieronder beschreven wijze interpreteren.

Alle gevonden overschrijdingskansen zijn klein, zodat voor alle beschouwde kenmerken tot een systematische overeenstemming der proeven besloten kan worden. De kolomtotalen geven aan, dat dit resultaat als volgt verder gepreciseerd kan worden.

1. De stammen O 20 en dba hebben een kleiner lichaamsgewicht dan de stammen C57 bl. en A. Uit de proef is niet gebleken of de stammen van bovenstaande paren onderling verschillen.


O 20
C 57 bl
dpa
A
lich. gew. gr.
37
36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
0,5
0,6
0,7
0,8
0,9
1,0
1,1
1,2
1,3
1,4
1,5
HVK gew. mg.

2. De stammen dba en A hebben kleiner HVK-gewicht dan de stamme O 20 en C57 bl. Uit de proef is niet gebleken of een onderling verschil tussen de stam van een paar bestaat.

Zie ter illustratie van 1 en 2 ook figuur 1 onder punt b verderop.

3. De stam O 20 heeft een lagere concentratie van de ACTH in de HVK dan de stammen C57 bl., dba en A. Uit de proef bleek niet of de laatste drie onderling naar concentratie van het ACTH te onderscheiden zijn.

4. De stam O 20 heeft een lagere totale hoeveelheid ACTH in de HVK dan de stammen C57 bl., dba en A. Over verschillen tussen de laatste drie stammen bleek nog niets uit deze proef.

5. Ook hier valt de stam O 20 er alleen uit, terwijl de andere stammen niet te onderscheiden zijn.

Op te merken valt nog de overeenkomst tussen de rangschikkingen naar concentratie van het ACTH en naar hoeveelheid ACTH gedeeld door het lichaamsgewicht.

b. Onderzoek met andere methoden.

Een tweede onderzoek betrof het verband tussen de verschillende kenmerken.

1. In par. 1.2 werd gevonden dat een correlatie bestaat tussen HVK-gewicht en lichaamsgewicht voor de muizen van stam Q. Het materiaal van de proef is te klein om te toetsen of een dergelijk verband voor ieder der stammen bestaat. Om echter inzicht te krijgen in de ligging van deze kenmerken voor de verschillende stammen is in bijgaande figuur lichaamsgewicht tegen HVK-gewicht uitgezet, waarbij de waarnemingen van verschillende stammen door verschillende tekens zijn aangegeven. Uit de figuur blijkt duidelijk dat elk paar stammen duidelijk verschilt naar minstens één van deze beide kenmerken.

De verschillen kunnen getoetst worden met de toets van Wilcoxon (S 47 (M 7)) en leveren alle overschrijdingskansen kleiner dan 0.003.

2. Met de methode der rangcorrelatie (S 47 (M 13)) werd onderzocht of overeenstemming bestond binnen ieder der dagproeven afzonderlijk tussen de rangschikkingen naar ACTH-concentratie en naar ACTH-hoeveelheid. De 8 uitkomsten werden vervolgens tot één toets gecombineerd. De waarde van de toetsingsgrootheid gaf een overschrijdingskans van 10^{-4} aan bij een positieve correlatie, d.w.z. lage ACTH-concentratie gaat samen met een kleine hoeveelheid ACTH in de HVK.

Zoals te verwachten was (zie par. 2.1) gaf eenzelfde onderzoek voor het verband tussen ACTH-concentratie en HVK-gewicht een negatieve correlatie aan (overschrijdingskans 0.03).

3. Toepassing van dezelfde toets gaf voor de concentratie van het ACTH en de hoeveelheid ACTH per gr. lichaamsgewicht een positieve correlatie aan (overschrijdingskans 10^{-4}).

3.1 Samenvatting.

Samenvattend kan geconcludeerd worden:

1. Muizen van de stam O 20 hebben een kleinere concentratie van het ACTH in de HVK en een kleinere hoeveelheid ACTH in de HVK dan muizen van de andere stammen. Muizen van deze andere stammen zijn met het gegeven proefmateriaal niet verder te onderscheiden naar beide kenmerken.
2. De muizen van de stammen O 20 en dba hebben een kleiner lichaamsgewicht dan de muizen van de stammen C57 bl. en A. Die van de stammen dba en A hebben een kleiner HVK-gewicht dan die van de stammen O 20 en C57.
3. Er is niet uit te maken of rangschikking naar ACTH-concentratie of hoeveelheid ACTH enig verschil maakt, behalve dat de verschillen tussen de stammen voor de hoeveelheid ACTH minder uitgesproken zijn (zie par. 2.1 en par. 3.0 b.2).

Opmerking: Uit de schatting van de ijklijn is een orde van grootte berekend voor de ingespoten doses (1/40 mg HVK per 100 gr. rat) voor de stam O 20 en de andere stammen:

O 20 : 0.04 γ ACTH
andere stammen : 0.23 γ ACTH.

Aangezien de lineariteit van de ijklijn niet uitvoerig is onderzocht en de schatting van deze ijklijn slechts terloops is geschied, moeten deze schattingen met veel voorbehoud worden beschouwd. Zij geven slechts de orde van grootte aan.

Litteratuur:

[1] EMMENS, C.W., Hormone assay, N.Y. (1950).
[2] SAYERS, M.A., SAYERS, G. and WOODBURY, L.A., Endocrinology 42, 379 (1948).

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

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

De toetsing van een hypothese 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 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 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[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 H_0 op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, 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 een onrechte verwerping van de juiste hypothese; ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, 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 H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

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

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

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

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

Litteratuur:

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

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

Mathematisch Centrum,
2de Boerhaavestraat 49,
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 $\frac{1}{2}$ 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

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

De grootheid $\mu(\underline{U})$ is dus onafhankelijk van de waarden vast. Indien de hypothese H_o 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_o 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 (tweezijdige) overschrijdingskans k , behorende bij T, is gedefiniëerd als

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

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≤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,
A m s t e r d a m -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 x en 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 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

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 $|S| \geq S_o$, en bij éénzijdige toetsing van de vorm $S \geq S'_o$ (rechtszijdige toetsing) of $S \leq S''_o$ (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 $n \leq 10$ 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 $l = 1,2,\ldots,k_2$. σ_S kan dan gevonden worden uit de volgende formule:

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

$$+ \frac{1}{9n(n-1)(n-2)} \sum_{h=1}^{k_1} t_h(t_h-1)(t_h-2) \sum_{l=1}^{k_2} u_l(u_l-1)(u_l-2) +$$

$$+ \frac{1}{2n(n-1)} \sum_{h=1}^{k_1} t_h(t_h-1) \sum_{l=1}^{k_2} u_l(u_l-1).$$

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:

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

zodat:

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

en $\sigma_S = 4,8\mathcal{C}$ is.

Als alle t_h en alle u_l gelijk zouden 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 $n = 40,41,\ldots,100$ vindmen 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:

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

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.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.
Statistische Afdeling
S 73 (M 13b)

Spearman's correlatie-coëfficiënt, wanneer er geen gelijke waarnemingen optreden[1].

We beschouwen het geval dat twee stochastische grootheden $\underline{x}$ en $\underline{y}$[2] een simultane verdeling bezitten. Over deze verdeling zelf behoeft niets ondersteld te worden. (x_1,y_1), $(x_2,y_2)\ldots(x_n,y_n)$ 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_o te toetsen, dat $\underline{x}$ en $\underline{y}$ onafhankelijk verdeeld zijn heeft C.Spearman de volgende parametervrije toetsingsgrootheid ingevoerd:

Nummer de waarnemingen x_i $(i = 1,2,\ldots,n)$ 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 gekwadrateerd 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_o van onafhankelijkheid ten gunste van de alternatieve hypothese: $\underline{x}$ en $\underline{y}$ 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. 142[3]. Voor n = 9 en 10 in Biometrika [2] p. 133[3].

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

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

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 $\mathcal{E}\,\underline{S}(d^2)$ van $\underline{S}(d^2)$

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

en voor de variantie

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

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.

MATHEMATISCH CENTRUM
2de Boerhaavestr.49
A M S T E R D A M O.
Statistische Afdeling
S47 (M14)

Methode der m rangschikkingen [1])

Een duidelijk voorstelling van deze toetsingsmethode verkrijgt men door n elementen te beschouwen, die een bepaald kenmerk, eventueel in verschillende mate, bezitten. dit kenmerk wordt door m waarnemers beoordeeld en ieder van deze waarnemers rangschikt deze n elementen volgens zijn beoordeling naar opklimmende waardering. Op deze wijze ontstaan m rijen van rangschikkingen. We willen nu een maat aangeven voor de overeenstemming tussen deze rangschikkingen, m.a.w. een maat voor de overeenstemming tussen de m beoordelingen. De hypothese 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 $\frac{1}{2}nm(n+1)$. Onder de hypothese 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 −3 −4 5 −2 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_o is exact berekend voor een aantal waarden van n en m [1], terwijl voor grote m en n benaderingen bekend zijn.

De meeste 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 $m \to \infty$ 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

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

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

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

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

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

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

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

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

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

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

Statistische Afdeling
S 53 (M 22)

Tekentoets[1)]

Deze toets dient voor het toetsen van de hypothese H_o, dat een aantal grootheden $\underline{z}_1,\ldots,\underline{z}_n$ alle nul tot mediaan hebben, d.w.z. dat

$$P\left[\underline{z}_i > 0\right] = P\left[\underline{z}_i < 0\right] \qquad i = 1,\ldots,n,$$

is. De toets geldt zonder enige verdere beperking dan de eis, dat de grootheden $\underline{z}_i$ onderling onafhankelijk verdeeld zijn; zij behoeven niet dezelfde waarschijnlijkheidsverdeling te bezitten.

De toets berust op één waarneming van ieder der grootheden $\underline{z}_i$, dus op n waarnemingen $z_1,\ldots,z_n$. De waarnemingen, die de waarde 0 bezitten, laten wij buiten beschouwing[2)]. Als toetsingsgrootheid gebruiken wij nu $\underline{n}_1$, het aantal positieve waarnemingen. Zijn er m waarnemingen $\neq 0$, dan bezit $\underline{n}_1$ een binomiale verdeling, onderstellende, dat H_o juist is:

$$P\left[\underline{n}_1 = n_1 \mid H_o\right] = \binom{m}{n_1} 2^{-m}.$$

Als kritieke zône worden de grote en kleine waarden van n_1 genomen. De kritieke zône is, voor onbetrouwbaarheidsdrempels 0,01; 0,05; 0,10 en 0,25 en m =1 tot 100 getabelleerd door

W.J.Dixon and A.M.Mood, The statistical sign test, Jrn.Am.Stat.Ass. 41 (1946) p. 556-566.

Voor een groter aantal waarnemingen gebruikt men als benadering van de binomiale verdeling de aangepaste normale verdeling.

Opmerking: De toets wordt vaak gebruikt, indien men een aantal grootheden twee maal heeft waargenomen, voor en na een bepaalde gebeurtenis, om na te gaan of deze gebeurtenis invloed op de grootheden heeft uitgeoefend. Noemen wij de waarnemingen vóór het optreden der gebeurtenis $\underline{x}_i$ (i=1,...,n) en erna $\underline{y}_i$, dan hebben de grootheden $\underline{x}_i-\underline{y}_i$ alle o als mediaan, indien $\underline{x}_i$ dezelfde verdeling bezit als $\underline{y}_i$ (dus als de gebeurtenis geen invloed heeft gehad). De toets wordt nu toegepast op $\underline{z}_i=\underline{x}_i-\underline{y}_i$ (i=1,...,n).

---

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

2) In tegenstelling tot de gewoonte deze waarnemingen voor de helft bij de positieve en voor de helft bij de negatieve te tellen; de door ons gebruikte methode geeft de toets een groter onderscheidingsvermogen.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A M S T E R D A M - O.

Statistische Afdeling.
S 73 (M 28)

Een parametervrije toets tegen trend voor groepen waarnemingen[1)]

door T.J. Terpstra.

Januari 1952.

Wij beschouwen het geval, dat h onafhankelijke steekproeven van h stochastische grootheden gegeven zijn:

$x_{1,1}, x_{1,2}, \cdots, x_{1,n_1}$ van de grootheid $\underline{x}_1$,

$x_{2,1}, x_{2,2}, \cdots, x_{2,n_2}$ " " " $\underline{x}_2$,

. . . . . .

$x_{h,1}, x_{h,2}, \cdots, x_{h,n_h}$ " " " $\underline{x}_h$.

De uitgebreidheden van de steekproeven zijn dus $n_1, n_2, \cdots, n_h$ en de eerste index van de waarneming $x_{i,j}$ geeft aan uit welke steekproef deze waarneming afkomstig is, terwijl de tweede index het nummer der waarneming binnen die steekproef aangeeft.

De hypothese H_o, die wij wensen te toetsen luidt, dat al deze waarnemingen van dezelfde stochastische grootheid afkomstig zijn. Anders uitgedrukt: H_o houdt in, dat de stochastische grootheden $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_h$ onderling onafhankelijk verdeeld zijn en alle dezelfde waarschijnlijkheidsverdeling bezitten.

De alternatieve hypothese, waartegen wij H_o wensen te toetsen, is, dat de rij grootheden $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_h$ een (stijgende of dalende) trend vertonen. Een precieze definitie van "trend" zullen wij hier niet geven. Grofweg komt een stijgende trend hierop neer, dat van twee waarnemingen, één van $\underline{x}_i$ en één van de meer naarrechts in de rij voorkomende grootheid $\underline{x}_j$ (dus $j > i$), de laatste meer kans heeft om groter te zijn dan de eerste, dan andersom. Dit behoeft, strikt genomen , niet voor iedere i en j met $i < j$ te gelden, maar slechts voor het merendeel van dergelijke paren, maar daar gaan wij nu niet nader op in. Een analoge definitie geldt voor dalende trend.

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

De toetsingsgrootheid T wordt nu als volgt gedefiniëerd: Tel het aantal waarnemingen uit de 2^e, 3^e, ..., h^e steekproef, dat kleiner is dan $x_{1,1}$ (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1) en noem dit aantal $v_{1,1}$. Voer dezelfde telling uit voor $x_{1,2}$, $x_{1,3}$, ..., x_{1,n_1}, en noem de gevonden aantallen $v_{1,2}$, ..., v_{1,n_1}. Noem de som van deze aantallen V_1:

$$V_1 = v_{1,1} + v_{1,2} + \cdots + v_{1,n_1}.$$

Tel vervolgens het aantal waarnemingen uit de 3^e, 4^e, ..., h^e steekproef, dat kleiner is dan $x_{2,1}$ (weer $\frac{1}{2}$ tellen bij gelijkheid) en noem dit $v_{2,1}$; analoog $v_{2,2}$,,...., v_{2,m_2} en noem

$$V_2 = v_{2,1} + \cdots + v_{2,n_2}.$$

Bij deze tweede stap wordt dus de eerste steekproef buiten beschouwing gelaten. Bij de derde stap laten wij de eerste twee steekproeven buiten beschouwing en bepalen

$$V_3 = v_{3,1} + \cdots + v_{3,n_3}$$

op analoge wijze. Dit wordt voortgezet tot en met

$$V_{h-1} = v_{h-1,1} + \cdots + v_{h-1,n_{h-1}},$$

die uit de laatste twee steekproeven wordt bepaald. De toetsingsgrootheid is nu

$$T = V_1 + V_2 + \cdots + V_{h-1}.$$

Deze grootheid kan dus kortweg gedefiniëerd worden als het aantal paren $(x_{i,a}, x_{j,b})$ met $i < j$ en a en b willekeurig[2)], waarvoor $x_{i,a} > x_{j,b}$ is, vermeerderd met de helft van die paren, waarvoor $x_{i,a} = x_{j,b}$ is.

Het is duidelijk, dat T vooral grote waarden zal aannemen, indien er een dalende en kleine, indien er een stijgende trend is. Beschouwen wij de verzameling van alle bij het experiment mogelijke uitkomsten, dan bezit $\underline{T}$ op deze verzameling een waarschijnlijkheidsverdeling (vandaar de onderstreping van de letter T). Indien de hypothese H_o, inhoudende, dat alle waarnemingen van dezelfde stochastische grootheid afkomstig zijn, juist is, geldt voor deze waarschijnlijkheidsverdeling het volgende:

1. $\underline{T}$ is bij benadering normaal verdeeld.
2. Het gemiddelde van deze verdeling is:

2) Met uiteraard $1 \leqslant i < h$, $1 < j \leqslant h$, $1 \leqslant a \leqslant n_i$ en $1 \leqslant b \leqslant n_j$.

$$\mu = E(\underline{T}|H_o) = \tfrac{1}{4}\,(N^2 - \sum_{i=1}^{h} n_i^{\,2})$$

en het spreidingskwadraat is

$$\sigma^2 = \sigma^2(\underline{T}|H_o) = \frac{1}{72}\Big\{N(N+1)(2N+1) - \sum_{i=1}^{h} n_i(n_i+1)(2n_i+1)\Big\}$$

met

$$N = \sum_{i=1}^{h} n_i .$$

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

$$\frac{T-\mu}{\sigma} > \xi_\alpha \quad , \qquad 3)$$

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

$$k = \frac{2}{\sqrt{2\pi}} \int_{\left|\frac{T-\mu}{\sigma}\right|}^{\infty} e^{-\frac{1}{2}x^2}\, dx \qquad 3)$$

en kan dus ook in een tabel van de normale verdeling worden gevonden. Bij éénzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.

Opmerkingen.

1. De boven voor σ^2 gegeven formule geldt eigenlijk alleen, indien er geen gelijke waarnemingen zijn. Als er wel gelijke waarnemingen zijn, kan σ^2 op de volgende wijze gecorrigeerd worden. Beschouw alle waarnemingen tezamen; zijn er k groepen van gelijke waarnemingen onder, die resp. m_1, m_2, ..., m_k elementen

3) Deze formules berusten op de normale benadering van de verdeling van $\underline{T}$, als H_o vervuld is. De exacte verdeling van $\underline{T}$ onder H_o is slechts voor enkele gevallen ($h = 2$ en $n_i \leq 10$, of $n_i = 1$ voor $i = 1, \ldots, h$ en $h \leq 40$) bekend.

bevatten, dan wordt de reeds berekende σ^2 verminderd met het volgende bedrag:

$$\frac{1}{72}\sum_{j=1}^{k} m_j\,(m_j-1)(2m_j+5) +$$

$$-\frac{1}{36N(N-1)(N-2)}\left\{\sum_{i=1}^{h} n_i(n_i-1)(n_i-2)\right\}\left\{\sum_{j=1}^{k} m_j(m_j-1)(m_j-2)\right\} +$$

$$-\frac{1}{8N(N-1)}\left\{\sum_{i=1}^{h} n_i(n_i-1)\right\}\cdot\left\{\sum_{j=1}^{k} m_j(m_j-1)\right\} .$$

Deze correctie is steeds negatief en is, als de m_i niet te groot zijn, in het algemeen klein, zodat het dan van weinig belang is de berekening uit te voeren.

2. Strikt genomen is de geldigheid van de toets alleen bewezen voor het geval, dat er geen gelijke waarnemingen zijn. Zijn er wel gelijken, dan is het n.l. niet bekend, of de normale benadering nog goed is.

3. De toets is een toepassing van Kendall's theorie over de rangcorrelatiecoëfficiënt τ en kan anderzijds beschouwd worden als een generalisatie van Wilcoxon's toets voor het probleem van twee steekproeven.