Statistische Afdeling
van het
Mathematisch Centrum,
Amsterdam

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 94

Onderzoek naar een snelle wisseling van het histaminegehalte bij schizophrenen

door

Mej. C. van Eeden.

Augustus 1952

1. Inleiding:

Om te onderzoeken of bij schizophrenen een snelle wisseling van het histaminegehalte optreedt, werd de volgende proef opgezet:

Bij ieder van een zevental schizophrenen werden tien bloedmonsters van ieder 2 cc afgenomen met tussenpozen van 1 minuut. In deze tussenpozen werd de naald in de vena gelaten; het bloed druppelt dan langzaam uit de naald en wordt opgevangen in een buis, waarin enige druppels heparine zijn gebracht. Nadat de tien bloedmonsters zijn afgenomen wordt uit de buis 10 x 2 cc gezogen.

Op deze wijze krijgt men 20 bloedmonsters: nl. 10 monsters om de minuut afgenomen[1)] en 10 monsters uit eenzelfde hoeveelheid bloed[2)].

Bij vier van de zeven personen werden de twintig monsters bewerkt volgens de nieuwe methode (zie rapport S 75 van de Statistische Afdeling van het Mathematisch Centrum); bij de overige drie werd de oude methode gebruikt.

2. Methode van onderzoek en resultaten:

2.1 Monsters bewerkt volgens de nieuwe methode:

We onderzoeken hier eerst met behulp van de methode der m rangschikkingen (zie bijlage S 47 (M 14)) de overeenstemming tussen de vier volgordeschattingen en wel voor ieder der tijd- en contrôlereeksen apart.

In tabel I vinden we de hierbij gevonden kolomtotalen en overschrijdingskansen voor de tijdreeks; tabel II geeft deze resultaten voor de contrôlereeks.

Tabel I

Kolomtotalen en overschrijdingskansen gevonden met de methode der m rangschikkingen, toegepast op de vier volgordeschattingen voor de bloedmonsters van de tijdreeksen:

proef- datum	kolomtotalen voor bloedmonsters no										overschrijdingskans
	1	2	3	4	5	6	7	8	9	10	
16-1-'52	19	27	6	20	17	5	7	15	19	-[3)]	$\approx 10^{-4}$
30-1-'52	28	14,5	33	18	21,5	38	7,5	12,5	16,5	32,5	10^{-4}
18-2-'52	11	20	5	9,5	10	-[3)]	17	11,5	24	27	0,0004
5-3-'52	24,5	13	30	14	13	7	17	10	23,5	13	0,013

1) Wij zullen deze monsters in het volgende aanduiden met: de monsters van de tijdreeks.
2) Deze monsters duiden we aan met: de monsters v.d.contrôlereeks.
3) "-" betekent: "bepalingen ontbreken".

Tabel II

Kolomtotalen en overschrijdingskansen gevonden met de methode der m rangschikkingen toegepast op de vier volgordeschattingen voor de bloedmonsters van de contrôlereeksen:

proef-	kolomtotalen voor bloedmonsters no.											overschrij-
datum	11	12	13	14	15	16	17	18	19	20	21	22	dingskans
16-1-'52	30	6,5	11	16	3,5	18	13	21	20	26	- [3]	-	$<10^{-4}$
30-1-'52	-	-	16,5	-	25	25	26	12	11,5	14	15	35	0,01
18-2-'52	17	13	-	-	-	21	4.5	8	15	5,5	-	-	10^{-4}
5-3-'52	20	14,5	14	16	14	5	2	5	8	11,5	-	-	0,006

We vinden dus zowel bij de tijd- als bij de contrôlereeksen een duidelijk verschil tussen de histaminegehalten der bloedmonsters. Wat de contrôlereeksen betreft, wijst dit op een verschil in de bloedmonsters na de opwerking. Daar deze monsters voor ieder der contrôlereeksen alle uit één buis zijn genomen, is dit te wijten aan een inhomogene verdeling van het histamine in de buis, of aan de opwerking. Wat de tijdreeksen betreft is de vraag, of hier misschien ook snelle physiologische schommelingen mee hebben gewerkt, nog niet beslist.

2.2 Onderzoek naar een verschil in spreiding tussen tijd- en contrôlereeks:

De vraag waar het bij dit onderzoek om gaat is: Treedt er bij schizophrenen een snelle wisseling van het histaminegehalte op? Of wel: Is er een verschil in spreiding tussen de waarnemingen van de tijdreeks en die van de corresponderende contrôlereeks? Voor wij dit verschil in spreiding onderzoeken, gaan wij eerst na of er een niveauverschil is tussen tijd- en contrôlereeks.

Hiertoe is de toets van Wilcoxon toegepast. Bij de personen, waarbij de monsters bewerkt zijn volgens de oude methode, is deze toets toegepast op de door de onderzoekers zelf gegeven schattingen van het histaminegehalte. Bij de anderen is Wilcoxon's toets toegepast op de kolomtotalen, die we voor tijd- en contrôlereeks vinden, als we de methode der m rangschikkingen toepassen op de twee reeksen tezamen (hierbij is dus één rangschikking afkomstig van die 20 monsters van tijd- en contrôlereeks, die t.a.v. dezelfde bekende bepaald zijn).

In tabel III vinden we de resultaten van dit onderzoek naar niveauverschil:

Tabel III[4)]

Onderzoek naar een niveauverschil tussen tijd- en contrôlereeks:

bepalings-methode	proef-datum	n	m	U	overschrij-dingskans
oud	21-1-'52	9	9	18,5	0,06 +
	14-2-'52	9	8	56	0,06 -
	18-2-'52	10	9	80	0,003-
nieuw	16-1-'52	10	9	40,5	0,75 +
	30-1-'52	10	9	53	0,54 -
	18-2-'52	9	9	52	0,34 -
	5-3-'52	10	10	35	0,28 -

Bij de monsters bewerkt volgens de nieuwe methode is er dus geen reden om aan te nemen, dat er een verschil in niveau tussen de tijdreeksen en de corresponderende contrôlereeksen is. Bij de oude methode vinden we éénmaal (proefdatum 18-2-'52) een duidelijk verschil en tweemaal een aanwijzing voor een verschil, in tegengestelde richtingen. Een niveauverschil tussen de waarnemingen van een tijd- en een contrôlereeks wijst erop, dat deze twee reeksen ten gevolge van de proefopzet of van de bepalingsmethode, niet als voldoende vergelijkbaar kunnen worden beschouwd. Immers men zou verwachten, dat de gemiddelden van de bepalingen uit de beide reeksen ongeveer gelijk zouden zijn. Wij hebben daarom de rest van het onderzoek slechts bij die monsters uitgevoerd, waarbij geen niveauverschil gevonden is (dat zijn dus de bloedmonsters, die bewerkt zijn volgens de nieuwe methode). Hiervoor zullen wij het verschil in spreiding tussen tijd- en contrôlereeks onderzoeken.

Dit onderzoek is als volgt gedaan:

De n kolomtotalen (dezelfde als die waarop wij de toets van Wilcoxon toepasten) worden gerangschikt naar opklimmende grootte en verdeeld in twee groepen, ieder ter grootte $\frac{1}{2}$n. De ene groep bestaat uit de middelste $\frac{1}{2}$n kolomtotalen in de gerangschikte rij, de overige kolomtotalen (dat zijn dus de eerste $\frac{1}{4}$n en de laatste $\frac{1}{4}$n) vormen de tweede groep[5)].

---

4) Het teken + bij een overschrijdingskans betekent, dat de waarnemingen van de tijdreeks hoger liggen, dan die van de contrôlereeks; het teken - betekent het tegengestelde.

5) Indien het niet mogelijk is deze aantallen precies $\frac{1}{4}$n, $\frac{1}{2}$n en $\frac{1}{4}$n te nemen, dan zorgen we, dat ze zo dicht mogelijk bij $\frac{1}{4}$n, $\frac{1}{2}$n en $\frac{1}{4}$n liggen.

We hebben nu de kolomtotalen onderscheiden naar: behorend tot de middelste $\frac{1}{2}n$ en behorend tot de buitenste $\frac{1}{2}n$. Een tweede onderscheiding is die naar tijd- en contrôlereeks. Met behulp van de methode der dubbele dichotomie (zie bijlage S 53 (M 23)) toetsen wij nu de hypothese, dat bovengenoemde onderscheidingen onafhankelijk van elkaar zijn met als alternatieve hypothese, dat de kans om tot de middelste $\frac{1}{2}n$ te behoren bij de tijdreeks kleiner is dan bij de contrôlereeks. We toetsen hier éénzijdig, omdat een grotere spreiding bij de contrôlereeks geen argument voor snelle physiologische wisselingen zou zijn. We vinden nu de volgende aantallen:

Tabel IV

Onderscheiding van de kolomtotalen naar tijd- en contrôlereeks en naar buitenste en middelste kolomtotalen:

proefdatum	reeks	buitenste kolomtotalen	middelste kolomtotalen
16-1-'52	tijd	6	3
	contrôle	4	6

proefdatum	reeks	buitenste kolomtotalen	middelste kolomtotalen
30-1-'52	tijd	7	3
	contrôle	2	7

proefdatum	reeks	buitenste kolomtotalen	middelste kolomtotalen
18-2-'52	tijd	3	6
	contrôle	6	3

proefdatum	reeks	buitenste kolomtotalen	middelste kolomtotalen
5-3-'52	tijd	5	5
	contrôle	6	4

De (éénzijdige) overschrijdingskansen zijn resp. 0,24; 0,05; 0,97; 0,50.

De resultaten voor deze vier personen hebben wij op de volgende wijze gecombineerd:

Noem het aantal kolomtotalen van de tijdreeks in de buitenste $\frac{1}{2}n$ kolomtotalen $\underline{a}$, dan hebben wij voor ieder der personen:

1. een waarde voor $\underline{a}$,
2. een waarde voor $\mu(\underline{a})$,
3. een waarde voor $\sigma^2(\underline{a})$.

We noemen deze resp. a_i, μ_i en σ_i^2, waarbij de index i het nummer der proefpersoon is. Indien de te toetsen hypothese H_0 juist is, zal $a = \sum_i a_i$ bij benadering normaal verdeeld zijn met gemiddelde $\mu = \sum_i \mu_i$ en variantie $\sigma^2 = \sum_i \sigma_i^2$; $\frac{a-\mu}{\sigma}$ is dus bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

Is H_0 onjuist, dan zal $\frac{a-\mu}{\sigma}$ groot zijn, dus de kritieke zône bestaat uit grote waarden van $\frac{a-\mu}{\sigma}$.

De resultaten vinden we in tabel V:

Tabel V

Onderzoek naar verschil in spreiding tussen tijd- en contrôlereeks, bij de vier personen tezamen:

a	μ	σ	$\frac{a-\mu}{\sigma}$	overschrijdingskans
21	19,48	2,23	0,68	0,25

Uit tabel V zien we dus, dat er geen reden is om aan te nemen, dat de spreiding bij de tijdreeksen groter is dan die bij de corresponderende contrôlereeksen.

3. Conclusies:

1. Bij personen, waarvan de monsters bewerkt zijn volgens de oude methode, vinden wij niveauverschillen tussen de gehalten in de tijd- en de contrôlereeksen. Deze verschillen liggen niet alle in dezelfde richting. Dit maakt deze waarnemingen ongeschikt voor verder onderzoek.

2. Bij de personen, waarvan de bloedmonsters bewerkt zijn volgens de nieuwe methode, vinden wij een wisselend histaminegehalte, zowel binnen de tijdreeksen als binnen de contrôlereeksen. Bij deze laatste kan dit slechts een gevolg van de bepalingsmethode zijn, b.v. van de opwerking. Hetzelfde effect moet dus ook in de tijdreeks verwacht worden. Indien daar nog een physiologische schommeling bij zou komen, zou dit de spreiding der waarnemingen in de tijdreeks groter moeten maken dan in de contrôlereeksen. Hiervoor werd echter geen aanwijzing gevonden.

4. Opmerkingen:

De onvolmaaktheid van de bepalingsmethode is bij dit onderzoek opnieuw gebleken uit de storende verschijnselen (niveauverschil tussen de beide reeksen bij de oude methode en schommelingen in het gehalte na de opwerking, bij de nieuwe methode geconstateerd). Dientengevolge kan dit onderzoek niet beschouwd worden als beslissend over de vraag, of er snelle wisselingen zijn. Dit is te meer het geval, daar er onder de vier proeven van tabel IV één is (de derde), waarbij slechts één in plaats van vier rangschikkingen ter beschikking stond; bij de overige bepalingen waren tijd- en contrôlereeks niet t.o.v. dezelfde bekende getest, waardoor zij niet in één rij gerangschikt konden worden. Deze reeks gaf een uitkomst, die juist in de tegengestelde richting van de eerste twee proeven ligt en die bij de gevolgde methode even zwaar wordt meegeteld als die van de andere proeven, die op meer dan één rangschikking berusten. Zonder deze proef zou de, uit de overige drie proeven tezamen berekende, overschrijdingskans 0,06 zijn geweest. Weliswaar zou ook dit geenszins een overtuigend resultaat zijn geweest, maar een voortzetting van het onderzoek op iets grotere schaal wordt hierdoor toch wel gerechtvaardigd. Daarbij valt te overwegen de proef (met geruime tussenpozen) te herhalen bij die patiënten, waarbij het resultaat nu gunstig was; dit zijn de patiënten van de proefdata 16-1-'52 en 30-1-'52. Immers het is zeer wel mogelijk, dat snelle wisselingen bij de ene patiënt wel en bij de andere patiënt niet optreden. Een analoog onderzoek dient dan uiteraard bij normale proefpersonen te worden verricht, om na te gaan, of daar géén snelle wisselingen gevonden worden. Dit contrôle-onderzoek behoeft echter slechts ter hand genomen te worden, indien het onderzoek bij schizophrenen tot een positief resultaat heeft geleid.

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 $x_1, x_2, \ldots, x_n$, 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 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 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 ≥ 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 (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 een tabel 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 $\underline{k}$ voor de uit de steekproef bepaalde waarde U van $\underline{U}$ (zie [2] en [4]). Dergelijke tabellen bestaan echter niet voor het geval, dat gelijke waarnemingen optreden.

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

Litteratuur:

1.	F.Wilcoxon,	Individual comparisons by ranking methods, Biometrics 1 (1945), p.80-83.
2	H.B.Mann and D.R.Whitney	On a test of whether one of two random variables is stochastically larger than the other, Amer.Math.Stat. 18 (1947),p. 50-60.
3	H.R.van der Vaart	Some remarks on the power function of Wilcoxon's test for the problem of two samples, Proceedings van de Kon. Ned.Ak.v.Wet., 53 (1950),p. 494-520.
4	H.R.van der Vaart	Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en m$\leq$10, Rapport S32 (M4) (1950).
5	H.R.van der Vaart	De toets van Wilcoxon voor het probleem van twee steekproeven. (Cursus "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 (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

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

---

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

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

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

In ons voorbeeld is $W = \frac{12 \times 64}{16 \times 210} = 0{,}229$.
W varieert dus tussen 0 en 1.

De verdeling van $\underline{S}$ onder de hypothese H_0 is exact berekend voor een aantal waarden van n en m [1], terwijl voor grote m en n benaderingen bekend zijn.

De meest gebruikelijke benaderingen zijn de volgende.

1°. De χ^2-benadering:

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

2°. De z-benadering:

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

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

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

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

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

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

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

Literatuur: [1] M.G.Kendall, Rank correlation methods, London 1948, Hoofdstuk 6, pag. 80.
Tabel van de verdelingsfunctie van S voor:
$n = 3$ $m = 2$ t/m 10
$n = 4$ $m = 2$ t/m 6
$n = 5$ $m = 3$
op pag. 146-149.
Tabel van de waarden van S, waarvan de overschrijdingskansen onder de hypothese H_0 gelijk zijn aan 0,05 of 0,01, berekend met behulp van de z-benadering voor:
$n = 3$ $m = 8,9,10,12,14,15,16,18,20$
$n = 4$ $m = 4,5,6,8,10,15,20$
$n = 5$ t/m 7 $m = 3,4,5,6,8,10,15,20$
op pag. 150.

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

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

Statistische Afdeling.
S 53 (M 23).

Toetsing van de hypothese $p_1 = p_2$ met behulp van een 2 x 2-tabel[1].

Wij beschouwen twee reeksen van onafhankelijke experimenten, waarbij ieder experiment van de ene reeks één van de twee resultaten A of $\bar{A}$ (non-A) heeft en ieder experiment van de tweede reeks één van de beide resultaten B of $\bar{B}$ (hierbij kan A=B zijn). Daarbij wordt ondersteld, dat bij ieder der experimenten van de ene reeks de kans op A gelijk aan p_1 (en dus de kans op $\bar{A}$ gelijk aan $1-p_1$) is en bij ieder der experimenten van de tweede reeks de kans op B gelijk aan p_2 (en dus de kans op $\bar{B}$ gelijk aan $1-p_2$). De te toetsen hypothese luidt nu:

$$H_0: \; p_1 \neq p_2.$$

Indien de eerste reeks uit n en de tweede reeks uit m waarnemingen bestaat, waaronder n_1 (resp. m_1) maal A (resp. B) voorkomt, kunnen deze gegevens in de volgende 2 x 2-tabel worden samengevat:

	A resp. B	$\bar{A}$ resp. $\bar{B}$	totaal
eerste reeks	n_1	$n-n_1$	n
tweede reeks	m_1	$m-m_1$	m
totaal	r	$n+m-r$	n+m

Als toetsingsgrootheid wordt n_1, het aantal malen A in de eerste reeks waarnemingen, gebruikt. Indien H_0 juist is bezit deze grootheid onder de voorwaarde, dat r de bij het experiment gevonden waarde aanneemt, de volgende waarschijnlijkheidsverdeling: de kans, dat een bepaalde waarde n_1 aangenomen wordt, is gelijk aan:

$$\frac{\binom{n}{n_1}\binom{m}{m_1}}{\binom{n+m}{r}}$$

Als kritieke zône worden de waarden van n_1 met de kleinste waarschijnlijkheden bijeengezocht, tot de gekozen betrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert (bij éénzijdige toetsing bestaat de kritieke zône uitsluitend uit grote of uitsluitend uit kleine waarden van n_1).

---

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

De overschrijdingskans, behorende bij de gevonden waarde van n_1, is gedefiniëerd als de som van alle waarschijnlijkheden van bovenstaande verdeling, die hoogstens gelijk aan de waarschijnlijkheid van de gevonden waarde zijn (bij éénzijdige toetsing echter gelijk aan de som van de waarschijnlijkheden van alle waarden die groter of gelijk aan de gevondene, of van alle waarden, die kleiner of gelijk aan de gevondene zijn). Deze exacte toetsingsmethode voor H_0 is afkomstig van R.A. Fisher.

Indien n en m zo groot zijn, dat deze exacte berekening te omslachtig wordt, maakt men gebruik van de volgende benadering:

Gemiddelde en spreiding van de grootheid n_1 zijn (indien H_0 juist is):

$$\frac{nr}{n+m} \quad \text{resp.} \quad \sqrt{\frac{nmrs}{(n+m)^2(n+m-1)}}$$

Men gebruikt dan in plaats van de exacte waarschijnlijkheidsverdeling van n_1 de normale verdeling met hetzelfde gemiddelde en dezelfde spreiding en in plaats van de gevonden waarde van n_1 neemt men het getal, dat $\frac{1}{2}$ dichter bij het gemiddelde ligt dan deze gevonden waarde (dit laatste is de z.g. "continuïteitscorrectie", die bij toenemende n en m weldra verwaarloosd kan worden). Met behulp van de benadering gaat men dan verder te werk als boven beschreven, daarbij gebruik makende van een tabel van de normale verdeling.

Litteratuur:

R.A.Fisher, Statistical Methods for Research Workers, London 1948, p. 96. Opmerking: Fisher gebruikt hier de éénzijdige overschrijdingskans.

J.Hemelrijk, Waarschijnlijkheidsrekening en Statistiek, Vacantiecursus Mathematisch Centrum, Amsterdam 1950, § 4.