Statistische Afdeling
van het
Mathematisch Centrum,
Amsterdam

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

Rapport S 81.

Door Mej. C. van Eeden.

Vergelijking van wasmiddelen en van wasketels.

Maart 1952.

1. Inleiding.

Door het laboratorium van het Proefstation voor de wasindustrie werden ons de volgende gegevens verstrekt:

Gedurende 12 dagen werden, zowel 's morgens als 's middags, een aantal keukendoeken gewassen in ieder der wasmiddelen E, C en I. Deze doeken werden vóór het wassen in drie gelijke hoopjes van ongeveer gelijke vuilheidsgraad verdeeld; in ieder der drie wasmiddelen werd één zo'n hoopje wasgoed gewassen.

Bij dit wassen werden drie wasketels A, B en C gebruikt en wel zodanig, dat iedere combinatie van wasketel met wasmiddel acht maal voorkomt in de gehele proef.

De volgorde, waarin de wasproeven werden uitgevoerd, was 's morgens E, C, I en 's middags I, C, E.

Na het wassen werden de doeken beoordeeld; hierbij werd iedere doek in gedachten in acht velden verdeeld (vier op iedere zijde der doek). Als beoordelingscijfer voor de wasproef van één ochtend of middag werd nu opgegeven het totaal aantal geheel schone achtste doeken.

De bovenbeschreven wasproeven werden ook verricht voor slopen, lakens en badhanddoeken.

De vraag is nu:

1. of de variaties in de resultaten van dag tot dag misschien veroorzaakt worden door verschillen tussen de wasketels en door de volgorde, waarin de proeven iedere dag werden uitgevoerd.
2. of het mogelijk is de variantie van dag tot dag te splitsen in:
   a variantie ten gevolge van het verschil in voor de wasproef gebruikte ketel,
   b variantie veroorzaakt door verschil in uitvoeringsvolgorde,
   c restvariantie, in hoofdzaak veroorzaakt door verschil in vuilheidsgraad tussen de drie hoopjes op één morgen of één middag gebruikt wasgoed en door beoordelingsfouten;
   en of het mogelijk zou zijn, dat de overschrijdingskansen voor de gevonden verschillen tussen E, C en I zouden afnemen, als de variantie van dag tot dag in deze delen wordt gesplitst.

2. Methode van onderzoek.

Een splitsing van de variantie van dag tot dag in de in § 1 genoemde delen is wel mogelijk, maar heeft weinig zin, daar we op de uitkomsten niet de in de varianteanalyse gebruikelijke z-toets van Fisher kunnen toepassen. Aan de onderstelling van normaliteit en gelijke spreiding is nl. zeker niet voldaan.

Om te onderzoeken of er een verschil is tussen de wasketels, hebben wij de volgende methoden toegepast:

1. De methode der m rangschikkingen: (zie bijlage S 47 (M 14))

Hierbij was iedere rangschikking afkomstig van de beoordelingscijfers, die op een bepaalde ochtend of middag gegeven waren aan doeken, die in de wasketels A, B en C gewassen waren. (Hierbij wordt dus het eventuele verschil tussen de wasmiddelen verwaarloosd).

2. De toets van Wilcoxon: (zie bijlage S 47(M 7))

Voor ieder der wasmiddelen E, C en I (en ook voor de drie wasmiddelen tezamen) zijn de wasketels A, B en C onderling vergeleken met behulp van de toets van Wilcoxon.

Daar (bij de keukendoeken) het aantal gewassen doeken niet voor alle ochtenden en middagen hetzelfde is, hebben wij hier eerst alle beoordelingscijfers gedeeld door het aantal gewassen doeken, waarop deze cijfers betrekking hebben.

Bovenbeschreven methoden zijn toegepast voor de keukendoeken, de slopen en de badhanddoeken en op alle drie tezamen. Daar bij de lakens zeer veel gelijke waarnemingen optreden, hebben wij hier een andere methode toegepast en wel de volgende:

3. Methode der dubbele dichotomie: (zie bijlage S 53 (M 23)).

Als beoordelingscijfer van een ochtend of middag werd het aantal geheel schone lakens opgegeven. Daar er steeds twee lakens gewassen werden, krijgen we als beoordelingscijfers 0, 1 of 2.

De wasketels zijn nu twee aan twee vergeleken met

behulp van de methode der dubbele dichotomie, waarbij de beoordelingscijfers 0 en 1 samengenomen zijn. De te toetsen hypothese houdt dus in, dat de kans op 0 of 1 bij de ene ketel gelijk is aan de kans op 0 of 1 bij de tweede ketel.

De drie hier beschreven methoden kunnen mutatis mutandis worden toegepast om te onderzoeken of er een verschil is tussen de wasmiddelen.

3. Resultaten en conclusies.

3.1. Methode der m rangschikkingen:

De kolomtotalen en overschrijdingskansen gevonden met de methode der m rangschikkingen staan vermeld in tabel I (voor de wasketels) en tabel II (voor de wasmiddelen):

Tabel I:

Kolomtotalen en overschrijdingskansen gevonden met de methode der m rangschikkingen, voor de wasketels:

	Kolomtotalen voor de wasketels			Overschrijdingskans
	A	B	C	
keukendoeken	42,0	45,0	51,0	0,38
slopen	38,5	45,0	42,5	0,52
badhanddoeken	49,5	39,0	49,5	0,18
keukendoeken + slopen + badhanddoeken	130	129	143	0,36

Tabel II:

Kolomtotalen en overschrijdingskansen gevonden met de methode der m rangschikkingen voor de wasmiddelen:

	Kolomtotalen voor de wasmiddelen			Overschrijdingskans
	E	C	I	
keukendoeken	50,5	46,0	41,5	0,40
slopen	46,0	43,5	36,5	0,25
badhanddoeken	54,0	43,5	40,5	0,09
keukenhdd.+ slopen + badhanddd.	150,5	133,0	118,5	0,02

Met behulp van de tekentoets (zie bijlage S 53 (N 22)) zijn de wasketels, evenals de wasmiddelen, onderling vergeleken. De toets is toegepast op de verschillen tussen de beoordelingscijfers van twee hoopjes wasgoed, die op een zelfde ochtend of zelfde middag gewassen zijn (die dus voor het wassen van gelijke vuilheidsgraad waren) en waarvan het ene in de (het) ene en het andere in de (het)andere wasketel (wasmiddel) gewassen is.

In tabel III vinden we de resultaten voor de wasketels; in tabel IV die voor de wasmiddelen:

Tabel III:[1]

Overschrijdingskansen, gevonden bij vergelijking der wasketels met de tekentoets.

	Wasketels		
	A - B	A - C	B - C
keukendoeken	0,52 -	0,52 -	0,29 -
slopen	0,33 -	0,81 -	1,00
badhanddoeken	0,19 +	1,00	0,19 -
keukendoeken + slopen + badhanddoeken	0,90 +	0,53 -	0,15 -

Tabel IV:[2]

Overschrijdingskansen, gevonden bij vergelijking der wasmiddelen met de tekentoets.

	Wasmiddelen		
	[illegible]	B - I	C - I
keukendoeken	0,40 +	0,52 +	0,38 +
slopen	1,00	0,17 +	0,45 +
badhanddoeken	0,38 +	0,03 +	1,00
keukendoeken + slopen + badhanddoeken	0,19 +	0,007 +	0,30 +

---

1) Een + teken bij een overschrijdingskans betekent dat de waarnemingen bij de eerste wasketel hoger liggen dan bij de tweede.
2) Een + teken bij een overschrijdingskans betekent dat de waarnemingen bij het eerste wasmiddel hoger liggen dan bij het tweede.

3.2. **Toets van Wilcoxon:**

In tabel V vinden we de overschrijdingskansen, die bij de vergelijking der wasketels gevonden zijn met de toets van Wilcoxon.

Tabel VI geeft de overschrijdingskansen gevonden bij de vergelijking der wasmiddelen.

Tabel V:[3]

Vergelijking der wasketels met de toets van Wilcoxon:

Wasmiddelen	B			C			I			B+C+I		
Wasketels	A-B	A-C	B-C	A-B	A-C	B-C	A-B	A-C	B-C	A-B	A-C	B-C
keukendoeken	0,54 +	0,54 +	0,87 +	0,20 -	0,09 -	0,90 -	0,90 -	0,65 -	0,68 -	0,67 -	0,48 -	0,72 -
slopen	0,38 -	0,80 -	0,71 +	0,57 +	0,45 +	0,62 +	0,78 -	0,90 -	0,91 +	0,75 -	0,76 +	0,36 +
badhanddoeken	0,43 +	0,10 -	0,09 -	0,36 -	0,36 -	0,78 -	0,24 +	0,30 +	0,92 +	0,52 +	0,51 -	0,18 -
slopen + keukendoeken + badhanddoeken	0,70 +	0,44 -	0,51 -	0,29 -	0,24 -	0,99 -	0,62 +	0,81 +	0,93 -	0,98 -	0,50 -	0,78 -

Opmerking: Onder het hoofd B+C+I zijn de uitkomsten vermeld bij het tezamennemen van de waarnemingsresultaten ongeacht het gebruikte wasmiddel. In plaats daarvan kunnen ook, voor ieder der 3 vergelijkingen A-B, A-C en B-C apart, de onder B, C en I afzonderlijk gevonden resultaten worden gecombineerd door optelling van de voor de toetsingsgrootheid van Wilcoxon gevonden waarden. Voor deze som kan dan, met behulp van de normale benadering, een kritieke zone en dus ook een overschrijdingskans berekend worden. Deze bewerking werd wel uitgevoerd, maar de resultaten daarvan stemden in zo sterke mate met de in Tabel V reeds vermelde overeen, dat zij geen enkel nieuw gezichtspunt opleverden en daarom niet werden vermeld. Hetzelfde geld voor A+B+C in tabel VI.

---

3) Zie voetnoot 1.

Tabel VI:[4]

Vergelijking der wasmiddelen met de toets van Wilcoxon:

Wasketels	A			B			C			A+B+C		
Wasmiddelen	E-C	E-I	C-I	E-C	E-I	C-I	E-C	E-I	C-I	E-C	E-I	C-I
Keukendoeken	0,09 +	0,17 +	0,96	0,87 -	0,40 +	0,51 +	0,27 -	0,92	0,57 +	0,54 +	0,15 +	0,42 +
Slopen	0,61 -	0,73 +	0,41 +	0,29 +	0,19 +	0,62 +	0,83 +	0,61 +	0,75 +	0,70 +	0,17 +	0,27 +
Badhanddoeken	0,19 +	0,78 -	0,43 -	0,65 -	1,00	0,96	0,14 +	0,02 +	0,40 +	0,14 +	0,12 +	0,99
Keukendoeken + Slopen + Badhanddoeken	0,15 +	0,37 +	0,92 -	0,86 +	0,19 +	0,42 +	0,73 +	0,12 +	0,25 +	0,15 +	0,01 +	0,28 +

3.3. Methode der dubbele dichotomie.

De aantallen waarnemingen 0 en 1 en de aantallen waarnemingen 2 bij de beoordeling der lakens vinden we in tabel VII:

Tabel VII:

Aantallen waarnemingen 0 en 1 en aantallen waarnemingen 2 bij de beoordeling der lakens:

	Wasmiddelen			Wasketels		
	E	C	I	A	B	C
0+1	9	5	12	10	5	11
2	14	18	11	13	18	12

De overschrijdingskansen gevonden met de methode der dubbele dichotomie vinden we in tabel VIII:

Tabel VIII:[5]

Overschrijdingskansen gevonden bij vergelijking van de wasmiddelen en van de wasketels met de methode der dubbele dichotomie:

	Wasmiddelen			Wasketels		
	E-C	E-I	C-I	A-B	A-C	B-C
Overschrijdingskansen	0,20 -	0,88 +	0,04 +	0,10 -	0,76 +	0,08 +

---

4) Zie voetnoot 2.

5) Zie voetnoot 1 en 2.

3.4. Conclusies:

Uit het bovenstaande blijkt dus:

1. dat we met geen der methoden een verschil tussen de wasketels kunnen ontdekken.
2. dat wasmiddel I slechter is dan E(zie tabel IV en VI onder A+B+C) en dat er een aanwijzing is, dat I ook slechter is dan C (overal is het resultaat van C minstens even goed en bij de lakens wordt een overschrijdingskans 0,04 gevonden; zie tabel VIII).

Opmerking.

Hoewel er geen reden is, om aan te nemen, dat er een verschil tussen de wasketels is, kan men wel opmerken, dat de opzet van de proef niet zeer geschikt is, om een dergelijk verschil te ontdekken, daar de verschillen tussen de wasmiddelen eventueel aanwezige kleine verschillen tussen de wasketels kunnen verdoezelen. Ook het duidelijke verschil in vuilheidsgraad der gewassen goederen bemoeilijkt het onderzoek. Een nauwkeurig onderzoek naar de gelijkheid der wasketels kan beter worden uitgevoerd op de in rapport S 74 (Vergelijking van 4 wasmiddelen, door Mej. C. van Eeden) voor wasmiddelen beschreven wijze. Daarbij zou men dan met één wasmiddel kunnen werken (of de proef in zijn geheel met verschillende wasmiddelen kunnen herhalen). Indien ook bij een dergelijke proef geen verschil tussen de ketels gevonden zou worden, kan men de ketels voor deze proeven zonder bezwaar als gelijkwaardig beschouwen.

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. 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 $\alpha = 0{,}05$ 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 $k \leq \alpha$ 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 $\mathcal{H}_0$ 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.

De toets van Wilcoxon.[1)]

Deze methode dient tot het toetsen van de hypothese $\mathcal{H}_0$, inhoudend dat twee steekproeven $x_1, \ldots, x_n$ en $y_1, \ldots, y_m$ afkomstig zijn uit één collectie (ook populatio of universum genaamd). Zij is strict genomen, toepasbaar onder de voorwaarde, dat er geen enkel paar waarden (x_i, y_j) is met $x_i = y_j$. Verdere voorwaarden zijn voor de toepassing niet nodig, terwijl ook de zojuist genoemde, indien er niet teveel dergelijke paren zijn, de toepassing van de toets weinig hindert.

De toetsingsgrootheid $\mathcal{U}$ is het aantal paren (x_i, y_j) waarvoor $x_i > y_j$ is (het aantal "inversies"). Daar er nm dergelijke paren zijn, kan $\mathcal{U}$ alle gehele waarden van 0 tot en met nm aannemen. Is $\mathcal{U}$ groot, dan liggen er veel waarden x_i verder naar rechts dan waarden y_j , is $\mathcal{U}$ klein, dan juist weinig.

De kritieke zône $\mathcal{K}$ neemt nu daarom de kleine en de grote waarden van $\mathcal{U}$ en wel van beide zoveel, dat de gekozen onbetrouwbaarheidsdrempel α niet overschreden wordt.

Voor éénzijdige toetsing, te onderscheiden in linker- en rechter-éénzijdige toetsing, gebruikt men kritieke zônes $\mathcal{K}_1$, resp. $\mathcal{K}_2$, die geheel bestaan uit kleine, resp. grote waarden van $\mathcal{U}$.

Verwerping van $\mathcal{H}_0$ ten gevolge van het vinden van een grote (resp. kleine) waarde van $\mathcal{U}$ wijst erop, dat $x_1, \ldots, x_n$ en $y_1, \ldots, y_m$ steekproeven uit verschillende collecties zijn, waarbij de op de x-collectie aangenomen waarden systematisch groter (resp. kleiner) dan de op de y-collectie aangenomen waarden zijn.

Litteratuur:

F. Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945), p. 80-83.

H.B. Mann and D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18 (1947), p. 50-60. Bevat tabellen voor n en m ≦ 8.

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.

H.R. van der Vaart, Gebruiksaanwijzing voor de toets van Wilcoxon, met tabellen voor n en m ≦ 10, Rapport S32 (M_4)(1950).

D. van Dantzig, Kadercursus Mathematische Statistiek, Math. Centrum, Amsterdam(1947-50), hoofdstuk 6, 3.

---

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

Methode der m rangschikkingen [1])

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

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

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

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

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

1 −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 <u>coëfficiënt van overeenstemming</u>

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

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

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

De meest gebruikelijke benaderingen zijn de volgende.

1°. De χ^2-benadering:

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

2°. De z-benadering:

$\underline{V} = (m-1)\ \frac{\underline{W}}{1-\underline{W}}$ is bij benadering verdeeld als $\underline{F} = e^{2\underline{z}}$
($\underline{F}$ is de $\underline{F}$ van Snedecor, $\underline{z}$ de $\underline{z}$ van Fisher) met
$\nu_1 = n-1-\frac{2}{m}$
$\nu_2 = (m-1)\,\nu_1$ vrijheidsgraden ([1] pg. 84 [2] pg.33-36).

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

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.

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:

$$H_o: \quad p_1 = p_2.$$

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_o 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 gedefinieërd 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_o 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_o 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.

<u>Litteratuur:</u>

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

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