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 98

Onderzoek naar de invloed van twee wasmiddelen op de treksterkte van katoen

door

Mej. C.van Eeden

1952

1. Inleiding

Om te onderzoeken of er een verschil is in de slijtage van wasgoed, die veroorzaakt wordt door twee wasmiddelen, waarin dit wasgoed gewassen is, werd de volgende proef opgezet:

Een lange rol katoen (zie fig. 1) werd in vier gelijke stukken verdeeld.

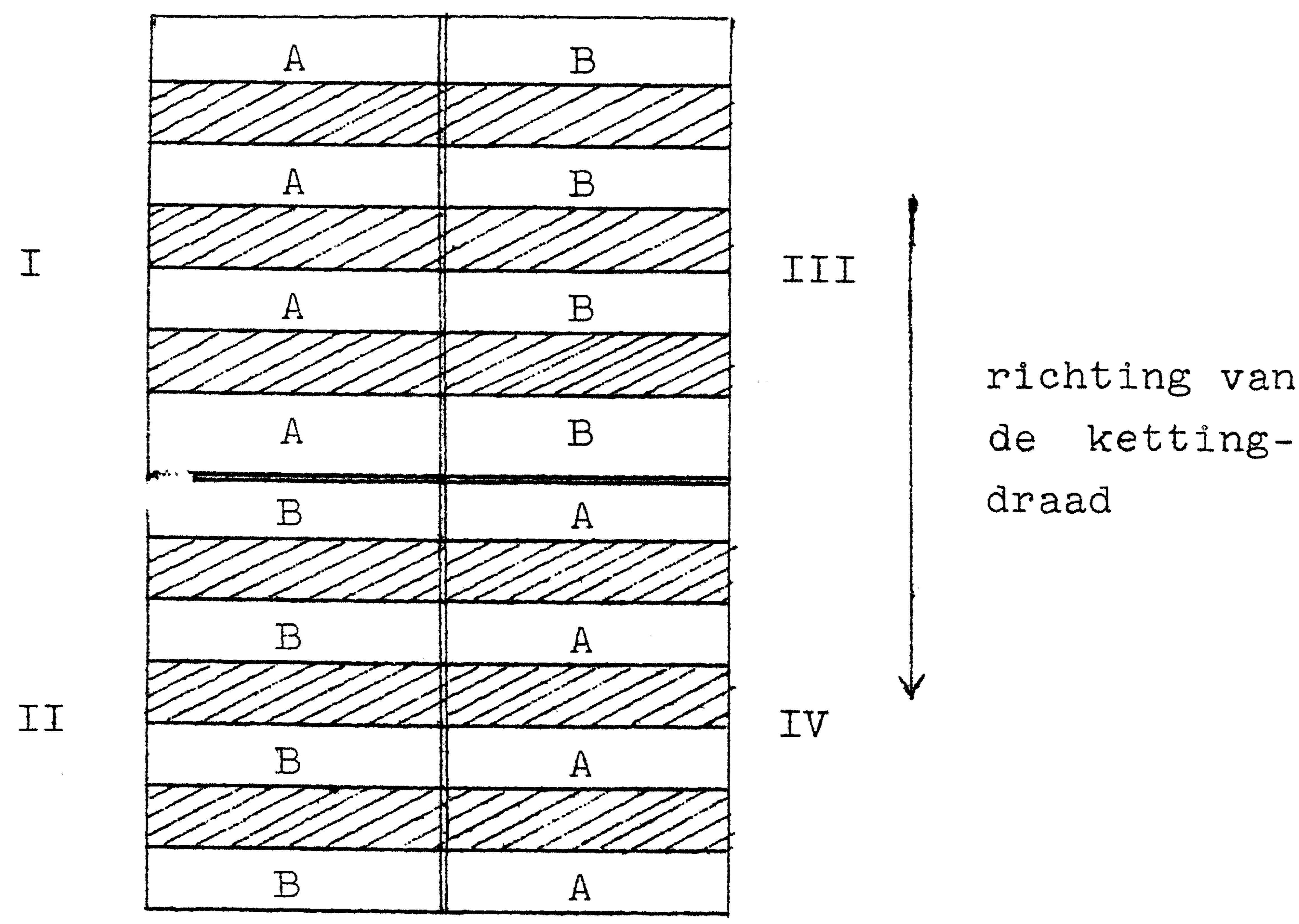

Fig. 1

Ieder van deze vier stukken werd nu in zeven even grote stroken verdeeld.
De in fig. 1 gearceerde stroken werden ongewassen gelaten; de niet gearceerde stroken werden 24x gewassen. Bij dit wassen werden twee wasmiddelen gebruikt, die we resp. A en B zullen noemen. De stroken links boven en rechtsonder werden gewassen met wasmiddel A; die links onder en rechts boven met wasmiddel B.

Na het wassen werden alle stroken (dus ook degene die niet gewassen zijn) in vijf even grote stukken verdeeld in de richting van de kettingdraad en van deze 140 stukken werd de treksterkte bepaald.
De treksterktebepalingen zijn alle op dezelfde dag door één persoon met één machine verricht. De volgorde waarin deze bepalingen verricht zijn is bekend.

Gevraagd wordt nu:
1. Is er een verschil tussen de wasmiddelen wat betreft de slijtage die ze veroorzaken?
2. Voor ieder der wasmiddelen een betrouwbaarheidsinterval te bepalen voor de gemiddelde treksterktedaling.

2. Methode van onderzoek en resultaten

2.1 Onderzoek naar de homogeniteit der ongewassen stroken en naar de homogeniteit der stroken die met eenzelfde wasmiddel gewassen zijn.

In de eerste plaats is, voor ieder van de delen links boven, links onder, rechts boven en rechts onder[1)] apart, onderzocht of er een verschil in treksterkte is tussen de drie ongewassen stroken. Hiertoe is de toets van Kruskal[2)] toegepast; een steekproef bestaat hier uit de vijf waarnemingen van de treksterkte van een strook.
De resultaten vinden wij in tabel I:

Tabel I
Onderzoek naar een verschil tussen de drie ongewassen stroken van ieder der delen I, II, III en IV

	H[3)]	C-1	Overschrijdingskans
I	1,26	2	0,53
II	1,37	2	0,50
III	1,09	2	0,58
IV	0,56	2	0,76

Voor geen van de vier delen I, II, III en IV vinden wij dus een verschil tussen de ongewassen stroken.

We zullen nu onderzoeken of er een verschil is tussen I, II, III en IV wat de ongewassen stroken betreft. Hiertoe is weer de toets van Kruskal toegepast; een steekproef bestaat nu uit de 15 waarnemingen van de treksterkte van de ongewassen stroken van een der delen I t/m IV.

De resultaten staan vermeld in tabel II:

----

1) Wij zullen deze vier delen in het volgende aanduiden met resp. I, II, III en IV (zie fig. 1).
2) W.Kruskal, A non-parametric analogue based upon ranks of one-way analysis of variance, Ann. Math. Stat. 23 (1952) 140.
3) H bezit bij benadering een χ^2-verdeling met C-1 vrijheidsgraden (C is het aantal groepen).

Tabel II

Onderzoek naar een verschil tussen I, II, III en IV wat betreft de ongewassen stroken

R_i [4]				H	C-1	Overschrijdingskans
I	II	III	IV			
582,5	522,5	212,5	512,5	18,1	3	0,0004

Uit tabel II zien we dus dat er tussen I, II, III en IV een duidelijk verschil is wat de treksterkte der ongewassen stroken betreft en wel in het bijzonder dat de ongewassen stroken rechts boven een lage treksterkte hebben.

Een volgende vraag is of, voor ieder der delen I t/m IV, de gewassen stroken verschillend zijn. We passen weer, voor ieder der delen I t/m IV apart, de toets van Kruskal toe; een steekproef bestaat dan uit de vijf waarnemingen van de treksterkte van een gewassen strook.

Tabel III geeft de resultaten:

Tabel III

Onderzoek naar een verschil tussen de vier gewassen stroken voor ieder der delen I t/m IV

	H	C-1	Overschrijdingskans
I	0,08	3	0,99
II	3,43	3	0,32
III	9,88	3	0,02
IV	7,58	3	0,06

We zien dus dat er voor I en II (dat is dus links boven en links onder) geen reden is om aan te nemen dat er een verschil is tussen de gewassen stroken. Rechtsboven vinden we dit verschil wel, terwijl er rechtsonder een aanwijzing voor een verschil is.

Uit het bovenstaande blijkt dus dat de rechter helft niet homogeen is; we vinden verschillen tussen de groepen van ongewassen stroken en tussen de stroken, die met eenzelfde wasmiddel gewassen zijn.

---

4) R_i is de som van de rangnummers van de waarnemingen uit de i-de groep. Een lage waarde van R_i betekent dus, dat de desbetreffende groep waarnemingen laag ligt ten opzichte van de over-

Om de linkerhelft nog nader te onderzoeken hebben wij de ongewassen stroken van I vergeleken met de ongewassen stroken van II. Hiertoe is de toets van Wilcoxon (zie bijlage S 47 (M 7)) toegepast, die hetzelfde is als de toets van Kruskal voor twee groepen van waarnemingen. De resultaten vinden wij in tabel IV:

Tabel IV[5)]

Vergelijking van de ongewassen stroken van I en II

n	m	U	Overschrijdingskans
15	15	138,5	0,28(+)

Uit tabel IV blijkt dus dat er geen reden is om aan te nemen dat er een verschil is tussen I en II wat de ongewassen stroken betreft.

In verband met de inhomogeniteit van de rechter helft zullen we nu nog onderzoeken of de waarnemingen een verloop vertonen in de volgorde waarin de bepalingen verricht zijn. Dit onderzoek is uitgevoerd voor de ongewassen, de met A gewassen en de met B gewassen stroken apart.

Om dit verloop te onderzoeken is in de eerste plaats een rangcorrelatie (zie bijlage S 47 (M 13)) toegepast op de gemiddelden van de waarnemingen van ieder der stroken met de volgorde waarin de bepalingen verricht zijn. Dit is voor de ongewassen stroken en voor de met ieder der wasmiddelen behandelde stroken apart uitgevoerd. Op deze wijze verkrijgen wij een duidelijk overzicht over de situatie. De resultaten staan vermeld in tabel V (zie blz. 4a).

We vinden dus alleen bij de ongewassen stroken een uitgepsroken daling van de treksterkte; de twee andere gevallen liggen echter wel in dezelfde richting.

Deze drie uitkomsten hebben wij op de volgende manier gecombineerd:

We vonden voor de ongewassen, de met A gewassen en de met B gewassen stroken ieder:

----

5) Het teken + bij de overschrijdingskans betekent dat de waarnemingen van de treksterkte van I groter uitgevallen zijn dan die van II.

Tabel V[6)]

Onderzoek naar een verloop in de strookgemiddelden, in de volgorde waarin de bepalingen verricht zijn, verricht met de methode der rangcorrelatie

<table>
<tr><td colspan="49">Treksterkte</td><td rowspan="2">Overschrijdingskans</td></tr>
<tr><td>Deel</td><td colspan="12">I</td><td colspan="12">II</td><td colspan="12">IV</td><td colspan="12">III</td></tr>
<tr><td>volgorde van bepalen</td><td colspan="4">1</td><td colspan="4">2</td><td colspan="4">3</td><td colspan="4">4</td><td colspan="4">5</td><td colspan="4">6</td><td colspan="4">7</td><td colspan="4">8</td><td colspan="4">9</td><td colspan="4">10</td><td colspan="4">11</td><td colspan="4">12</td><td rowspan="2">0,004 -</td></tr>
<tr><td>onge-wassen</td><td colspan="4">68,7</td><td colspan="4">69,2</td><td colspan="4">69,3</td><td colspan="4">69,0</td><td colspan="4">68,8</td><td colspan="4">67,8</td><td colspan="4">68,6</td><td colspan="4">68,7</td><td colspan="4">67,8</td><td colspan="4">64,4</td><td colspan="4">65,4</td><td colspan="4">66,0</td></tr>
<tr><td>volgorde van bepalen</td><td colspan="3">1</td><td colspan="3">2</td><td colspan="3">3</td><td colspan="3">4</td><td colspan="12"></td><td colspan="3">5</td><td colspan="3">6</td><td colspan="3">7</td><td colspan="3">8</td><td colspan="12"></td><td rowspan="2">0,40 -</td></tr>
<tr><td>gewassen met A</td><td colspan="3">59,7</td><td colspan="3">59,2</td><td colspan="3">58,9</td><td colspan="3">59,3</td><td colspan="12"></td><td colspan="3">59,4</td><td colspan="3">60,5</td><td colspan="3">57,2</td><td colspan="3">55,4</td><td colspan="12"></td></tr>
<tr><td>volgorde van bepalen</td><td colspan="12"></td><td colspan="3">1</td><td colspan="3">2</td><td colspan="3">3</td><td colspan="3">4</td><td colspan="12"></td><td colspan="3">5</td><td colspan="3">6</td><td colspan="3">7</td><td colspan="3">8</td><td rowspan="2">0,18 -</td></tr>
<tr><td>gewassen met B</td><td colspan="12"></td><td colspan="3">61,1</td><td colspan="3">63,7</td><td colspan="3">63,4</td><td colspan="3">62,9</td><td colspan="12"></td><td colspan="3">65,2</td><td colspan="3">61,3</td><td colspan="3">60,4</td><td colspan="3">59,4</td></tr>
</table>

----

6) Het teken - bij een overschrijdingskans betekent dat de waarnemingen een daling vertonen in de volgorde waarin de bepalingen verricht zijn.

1. een waarde voor de toetsingsgrootheid $\underline{S}$,
2. een waarde voor $\sigma^2_{\underline{S}}$.

We noemen deze resp. S_i en $\sigma^2_{\underline{S}_i}$. Indien de te toetsen hypothese juist is zal $\underline{S} = \sum_i \underline{S}_i$ bij benadering normaal verdeeld zijn met gemiddelde nul en variantie $\sigma^2 = \sum_i \sigma^2_{\underline{S}_i}$. $\frac{\underline{S}}{\sigma}$ zal dus bij benadering N(0,1) verdeeld zijn.

Indien H_0 onjuist is zal de gevonden waarde van $\frac{\underline{S}}{\sigma}$ sterk van nul afwijken; de kritieke zône bestaat dus uit grote waarden van $\left|\frac{S}{\sigma}\right|$. We vinden: $\frac{S}{\sigma} = -3{,}02$ en een overschrijdingskans van 0,0027.

We vinden dus een duidelijke daling van de treksterkte in de volgorde waarin de bepalingen verricht zijn.

Daar de bepalingen van de treksterkte van de linkerhelft 's ochtends en die van de rechterhelft 's middags verricht zijn zullen we nog onderzoeken of deze daling zowel in de linker als in de rechter helft optreedt of alleen in de rechter helft. Hiertoe is de linker helft nader onderzocht en wel is op de ongewassen stroken, de met A gewassen en de met B gewassen stroken apart de parametervrije toets tegen trend van Terpstra toegepast (zie bijlage S 73 (M 28)). Deze drie resultaten hebben wij gecombineerd zoals boven is aangegeven. We hebben dus de gevonden waarden voor $\underline{T}$, de waarden van $\mu_{\underline{T}}$ en die van $\sigma^2_{\underline{T}}$ opgeteld en vinden dan $\frac{T-\mu}{\sigma} = -0{,}66$ en een overschrijdingskans van 0,51.

Hieruit zien wij dus dat er geen reden is om aan te nemen dat er in de linker helft een daling der treksterkte optreedt in de volgorde waarin de bepalingen verricht zijn. De waarnemingen liggen echter wel in de richting van een daling. Passen wij deze zelfde toets toe op de rechter helft dan vinden wij: $\frac{T-\mu}{\sigma} = -4{,}16$ en een overschrijdingskans $< 0{,}0001$.

Tot nu toe hebben wij dus gevonden:

1. Een lage treksterkte van de ongewassen stroken van rechts boven.
2. Een inhomogeniteit van de gewassen stroken rechts boven.
3. Een daling van de treksterkte in de volgorde waarin de bepalingen verricht zijn, die hoofdzakelijk veroorzaakt wordt door een daling in de rechter helft. Dit blijkt uit de bovengevonden overschrijdingskansen van 0,51 voor de linker helft en $< 0{,}0001$ voor de rechter helft en uit de tabellen II en V.

Op grond van het bovenstaande hebben wij de rechter helft in het volgende buiten beschouwing gelaten en het onderzoek dus uitgevoerd met de gegevens der linker helft.

2.2 Onderzoek naar een verschil tussen de twee wasmiddelen.

Wij beschouwen dus alleen de linker helft der rol katoen. Om de twee wasmiddelen te vergelijken hebben wij, met behulp van de toets van Wilcoxon, de gewassen stroken van I vergeleken met de gewassen stroken van II. We vinden dan:

Tabel VI[7]

Vergelijking van de gewassen stroken van I en II

n	m	U	Overschrijdingskans
20	20	59	0,0001 (-)

Wij zien dus dat er een zeer duidelijk verschil is tussen de ·e met A en de met B gewassen stroken en wel dat de met B gewassen stroken (deel II) een grotere treksterkte hebben dan de met A gewassen stroken (deel I).

2.3 Bepaling van een betrouwbaarheidsinterval voor de gemiddelde treksterktedaling.

We zullen voor ieder der wasmiddelen een betrouwbaarheidsinterval bepalen voor de gemiddelde treksterktedaling. Dit geschie.. als volgt[8]:

Stel de treksterkte van de ongewassen katoen $\underline{x}$ en die van de gewassen katoen $\underline{y}$. Dan bestaat het betrouwbaarheidsinterval voor de gemiddelde treksterktedaling uit de verzameling van al die waarde van d waarvoor de hypothese dat $\underline{x}$-d en $\underline{y}$ dezelfde verdeling bezitten op grond van het waarnemingsmateriaal niet voor verwerping in aanmerking komt. De hypothese dat $\underline{x}$-d en $\underline{y}$ dezelfde verdeling bezitten wordt daarbij getoetst met behulp van de toets van Wilcoxon.

We vinden dan de volgende betrouwbaarheidsintervallen (onbetrouwbaarheidsdrempel 0,05):

wasmiddel A: $8,3 \leqq d \leqq 11,5$

wasmiddel B: $4,2 \leqq d \leqq 7,3$.

---

7) Het teken - bij de overschrijdingskans betekent dat de stroken I een lagere treksterkte hebben dan de stroken van II.

8) Zie ook bijlage S 47 (M 18).

(Voor de gemiddelde treksterkte van de ongewassen stroken der linker helft werd gevonden: 68,8).

3. Samenvatting der resultaten

We hebben dus gevonden:

1. In de rechter helft der rol katoen een daling van de treksterkte in de volgorde waarin de bepalingen verricht zijn. In de linker helft vinden wij deze daling niet.

2. Een inhomogeniteit van de vier groepen ongewassen stroken en wel in het bijzonder een lage treksterkte van de ongewassen stroken rechts boven.

3. Een inhomogeniteit van de gewassen stroken rechts boven en een aanwijzing hiervoor rechts onder; in de linker helft vinden wij deze inhomogeniteit niet.

4. Een duidelijk verschil tussen de met A en met B gewassen stroken, waarbij de met B gewassen stroken een grotere treksterkte (dus een kleinere treksterkte daling) hebben dan de met A gewassen stroken.

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 $\mathcal{H}_0$ 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.

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 eentabel van de normale verdeling worden gevonden.

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

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

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

---

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

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

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

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

Litteratuur:

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

Rangcorrelatie[1)]

1. Beschrijving van de methode.

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

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

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

Voorbeeld:

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

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

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

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

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

---

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

Tabel 1

Berekening van S voor het voorbeeld

i	j	Correlatie
1	2	-1
1	3	+1
1	4	0
1	5	+1
1	6	+1
2	3	-1
2	4	-1
2	5	+1
2	6	+1
3	4	+1
3	5	+1
3	6	+1
4	5	0
4	6	0
5	6	0
		S = +5

zijn, kan dus getoetst worden.

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

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

Als er noch bij de x_i noch bij de y_j gelijke waarden voorkomen kunnen wij gebruik te maken van exacte tabellen, die voorkomen in [1] pg 141 (n = 4 t/m,10) en in [2] (tables I and II, n = 4 t/m 40). Bovendien vindt men in [2] table III de kleinste waarden van 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 $\dfrac{\underline{S}}{\sigma_{\underline{S}}}$.

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

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

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

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

zodat:

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

en $\sigma_{\underline{S}} = 4{,}89$ is.

Als alle t_h en alle u_ℓ gelijk zijn aan 1 en er dus in geen van beide rijen gelijken voorkomen, gaat formule (2) over in:

$$(2)\qquad \sigma_{\underline{S}} = \sqrt{\frac{1}{18}\, n(n-1)(2n+5)}$$

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

3. Rangcorrelatiecoëfficiënt

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

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

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

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

Literatuur:

[1] M.G. Kendall. Rank correlation Methods London 1948, Hoofdstuk 1.

[2] L. Kaarsemaker en A. van Wijngaarden. Tables for use in rank correlation . (1952) Report R 73 of the Computation Department of the Mathematical Centre.

[3] J. Hemelrijk. Kendall's rangcorrelatie-coëfficiënt . Hoofdstuk I der cursus "Parametervrije Methoden" Rapport S 59 (1951) Mathematisch Centrum, blz. 1-17.

[4] G.P. Sillitto. "The Distribution of Kendall's coefficient of rankcorrelation in rankings containing ties. Biometrica 34 (1947) p. 36-40.

Betrouwbaarheidsintervallen (algemeen).[1)]

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

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

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

Litteratuur:

M.G. Kendall, The Advanced Theory of Statistics, London 1946, deel II, p.62-84.

A.M. Mood, Introduction to the theory of Statistics, London 1950, p.220.

J. Neyman, First course in probability and statistics, N.Y.1950.

---

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

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:

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

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

. . . .

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

De uitgebreidheden van de steekproeven zijn dus $n_1, n_2, \ldots, n_h$ 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 $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_h$ onderling onafhankelijk verdeeld zijn en alle dezelfde waarschijnlijkheidsverdeling bezitten.

De alternatieve hypothese, waartegen wij H_o wensen te toetsen, is, dat de rij grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_h$ 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 naar rechts 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:

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

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,n_2} en noem

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

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

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

op analoge wijze. Dit wordt voortgezet tot en met

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

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

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

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_0, 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 $1 \leq i < h$, $1 < j \leq h$, $1 \leq a \leq n_i$ en $1 \leq b \leq n_j$.

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

en de spreidingskwadraat is

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

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

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

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. $m_1, m_2, \ldots, m_k$ 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:

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

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

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

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.