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 118

Onderzoek van kokmeeuweieren

door

Ph. van Elteren
en
J. F. van Haastrecht.

1953

INHOUD blz.

1. Inleiding.

1.1. Materiaal.

In Januari 1952 ontvingen wij van de heer P.J.H.van Bree, verbonden aan het Zoölogisch Museum te Amsterdam, een uitgebreid materiaal betreffende de afmetingen van kokmeeuweieren. Van deze eieren was met behulp van een schuifpasser de lengte (grootste diameter, aan te duiden met de letter ℓ) en de breedte (kleinste diameter, aan te duiden met de letter b) tot in tiende millimeters nauwkeurig gemeten. Dit is gedaan bij 855 eieren geraapt in een kokmeeuwenkolonie op Texel in het jaar 1950 en bij 318 kokmeeuweieren geraapt op Texel in 1951. De eieren van 1951 zijn in twee groepen verdeeld, corresponderend met twee gebieden in de kolonie waaruit zij werden geraapt. Een deel van de kolonie ligt namelijk op slikgrond en wordt bij hoge springvloed overstroomd. De verzameling eieren uit het waarnemingsmateriaal van 1951, afkomstig uit dit gebied, wordt aangeduid met "groep 1951 A". Het andere gedeelte ligt op zandgrond en wordt alleen bij stormvloeden overstroomd. Hiermee correspondeert een "groep 1951 B". Er is aan dit verslag een luchtfoto toegevoegd waarop de beide gebieden duidelijk te onderscheiden zijn (zie hieronder).

Luchtfoto van de kolonie op Texel

gebied B
" A

De eieren van 1951 zijn bovendien onderscheiden naar de nesten, waaruit zij geraapt zijn. Behoudens één ei, waren alle eieren geraapt uit nesten met 2 of nesten met 3 eieren en deze nesten zijn volkomen leeggehaald. De aantallen waren als volgt:

1951 A

31 nesten van 2 eieren:	62 eieren	
31 " " 3 " :	93 "	
	155 eieren,	

1951 B

33 nesten van 2 eieren:	66 eieren	
32 " " 3 " :	96 "	
	162 eieren.	

Een dergelijke indeling is niet gemaakt bij de eieren van 1950. Wel wist men ons mee te delen dat deze eieren voornamelijk uit één gebied van de kolonie waren geraapt en wel van slikgrond, terwijl de nesten geheel waren leeggehaald.

In September 1952 werden ons enige aanvullende inlichtingen vertrekt, verwerkt in het bovenstaande en tevens nieuw materiaal betreffende eieren geraapt in 1952. Deze eieren zijn gedeeltelijk in gebied A, gedeeltelijk in gebied B geraapt, doch er is niet opgegeven in welk gebied. Splitsing in 1952 A en 1952 B is dus niet mogelijk. Behalve lengte en breedte, werd hier van een aantal eieren eveneens de inhoud bepaald, door onderdompeling in water. Behoudens één ei, werden hier alle eieren geraapt uit nesten van 2 of van 3 eieren, die wederom geheel leeggehaald werden. Eén nest van 2 eieren moesten wij bij een gedeelte van het onderzoek buiten beschouwing laten, omdat kennelijk de breedte van één der eieren door het volume was vervangen. Behalve dit ene nest hadden wij:

1952

25 nesten van 2 eieren:	50 eieren	
8 " " 3 " :	24 "	
	74 eieren.	

Het volume is bepaald bij de eieren van 18 nesten van 2 en 7 nesten van 3, dus in totaal van 57 eieren.

1.2. Inhoud van de opdracht.

Er werd ons verzocht aan de hand van het bovenbeschreven materiaal een antwoord te geven op de volgende vragen:

1) Kan men aannemen, dat de verdelingen van de lengten en van de breedten van kokmeeuweieren normaal zijn. Is de simultane verdeling van lengte en breedte als normaal te beschouwen. (2) [1]).

2) Zijn de spreidingen van de afmetingen van de eieren binnen

---

1) Achter iedere vraag is vermeld in welke paragraaf van het verslag zij behandeld wordt.

de nesten kleiner dan de spreidingen van dezelfde afmetingen genomen over alle eieren. (3).

3) Bestaat er verschil tussen de afmetingen van de eieren uit nesten van 2 en uit nesten van 3. (4).

4) Bestaat er een verband tussen de lengte en de breedte van de eieren. (5).

5) Bestaan er verschillen tussen de afmetingen van de eieren van beide groepen geraapt in 1951. (6).

6) Bestaan er verschillen tussen de afmetingen van de eieren, die in de verschillende jaren geraapt zijn. (6).

7) Bestaat er verband tussen de verhouding van lengte en breedte van de eieren en hun inhoud. (7).

8) Het vermoeden bestaat, dat het eerst gelegde ei van de eieren uit een nest een meer langgerekte vorm heeft dan de latere. Gevraagd wordt of een bevestiging van dit vermoeden uit het materiaal is te verkrijgen (7).

Een overzicht van de antwoorden op deze vragen zal gegeven worden in 8 (samenvatting der conclusies).

2. Onderzoek van de verdeling der waargenomen grootheden in het bijzonder op normaliteit.

2.1. Normaliteit van de simultane verdeling van lengte en breedte.

Wij hebben de simultane verdeling van lengte en breedte onderzocht met behulp van de toets beschreven in memorandum S 102 (M 40), 6 [2]). De steekproef bestaat hier uit de paren waarnemingen (ℓ_i , b_i) waarbij ℓ_i de lengte en b_i de breedte van ei i voorstelt ($i = 1, \ldots, n$). De grootheden s_1 en s_2 worden hier aangeduid met s_ℓ en s_b ; $\bar{x}_1$ en $\bar{x}_2$ met $\bar{\ell}$ en $\bar{b}$. Hier geldt dus:

$$\bar{\ell} = \frac{1}{n}\sum_{i=1}^{n} \ell_i \qquad s_\ell = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\ell_i - \bar{\ell})^2}$$

$$\bar{b} = \frac{1}{n}\sum_{i=1}^{n} b_i \qquad s_b = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(b_i - \bar{b})^2}$$

$$r = \frac{1}{n-1}\,\frac{\sum_{i=1}^{n}(\ell_i - \bar{\ell})(b_i - \bar{b})}{s_\ell\, s_b}$$

De toets is toegepast bij het materiaal van 1950. De resultaten vindt men in de tabellen I en II.

---

2) Dit memorandum is toegevoegd aan dit rapport. Men leze ook het bijgevoegde memorandum S 32 (M 6).

Tabel I

Bepaling concentratie-ellipsen bij het materiaal van 1950.

$n = 855$

$\bar{b} = 36{,}687$ mm $\bar{\ell} = 51{,}494$ mm

$s_b = 1{,}177$ mm $s_\ell = 1{,}982$ mm

$r = 0{,}2702$ (correlatiecoëfficiënt)

$t = 0{,}2343$ (helling orthogonale regressielijn).

In het onderstaande tabelletje komen de kolommen overeen met de concentratie-ellipsen; ε stelt de kans voor dat een punt (b, ℓ) binnen de ellips ligt, volgens de aangepaste tweedimensionale normale verdeling.

	$\varepsilon =$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$
ellips	halve korte as	0,6720	1,0021	1,3102	1,6495	2,1065
	halve lange as	1,2192	1,8181	2,3771	2,9927	3,8219
	afstand middelp.-brandp.	1,0172	1,5170	1,9834	2,4971	3,1889

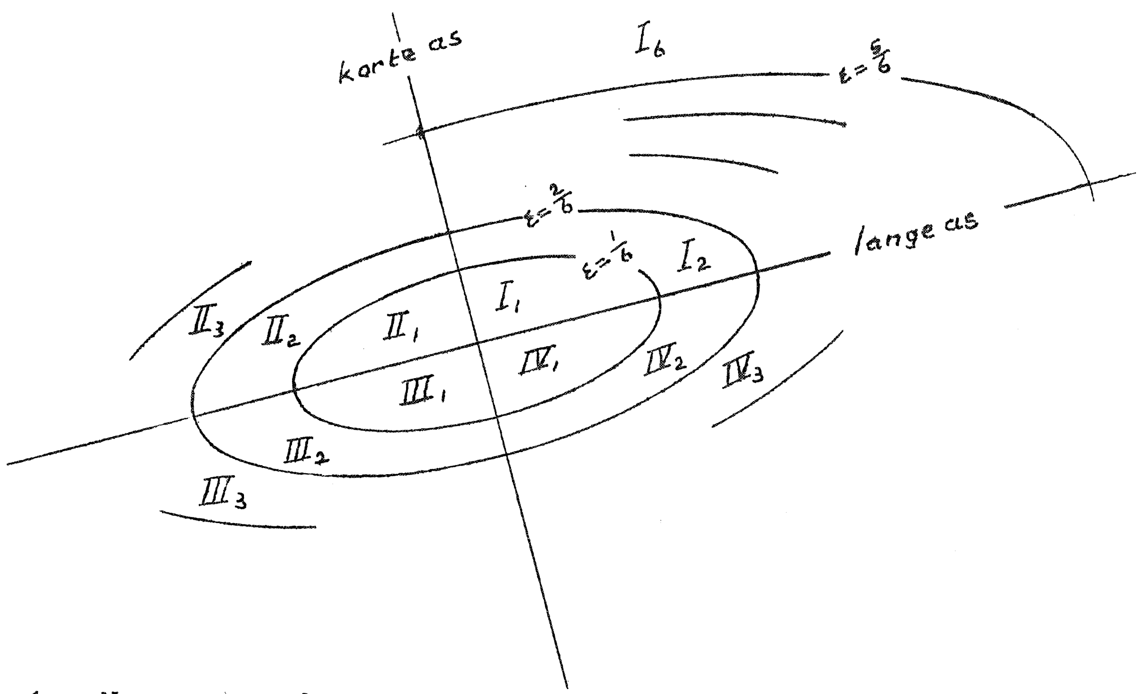


Fig. 1. Nummering der vakken bij de concentratie-ellipsen.

Wij nummerden de vakken als hierboven aangegeven (de vier quadranten ingesloten door de assen zijn aangeduid met I, II, III en IV en de ringen tussen de ellipsen van binnen naar buiten met 1, 2, 3, 4 en 5; het gebied buiten de 5e ellips met 6). De in ieder der vakken gevonden aantallen vindt men in tabel II.

Tabel II

Verdeling van de eieren over de 24 vakken van figuur 1 [3]).

ring → / ↓ quadrant	1	2	3	4	5	6	totaal quadrant
I	$37\frac{1}{2}$	$37\frac{1}{2}$	$35\frac{1}{2}$	46	$35\frac{1}{2}$	36	228
II	26	$29\frac{1}{2}$	$40\frac{1}{2}$	$37\frac{1}{2}$	$35\frac{1}{2}$	$28\frac{1}{2}$	$197\frac{1}{2}$
III	$33\frac{1}{2}$	$40\frac{1}{2}$	41	44	$31\frac{1}{2}$	37	$227\frac{1}{2}$
IV	32	38	32	$28\frac{1}{2}$	27	$44\frac{1}{2}$	202
totaal ring	129	$145\frac{1}{2}$	149	156	$129\frac{1}{2}$	146	855

Het resultaat was een $\chi^2 = 19{,}793$; deze heeft bij 18 vrijheidsgraden een overschrijdingskans $k = 0{,}32$. Onze conclusie is, dat in dit materiaal géén systematische afwijking van normaliteit te bespeuren is.

Op grond van deze uitkomst hebben wij in 3 een toets toegepast, waarbij we gebruik maken van de veronderstelling, dat de simultane verdeling van lengte en breedte van kokmeeuweieren normaal is, ook bij de andere jaren van waarneming. Deze veronderstelling houdt tevens in, dat de lengte en de breedte afzonderlijk beschouwd een één-dimensionale normale verdeling hebben. Strikt genomen is deze veronderstelling met bovenstaande toets alleen niet voldoende op haar juistheid onderzocht. Want het is mogelijk, dat men bij andere jaren wel een afwijking zou constateren en verder zou men de beide afmetingen apart op normaliteit kunnen toetsen met behulp van methoden, die wellicht een groter onderscheidingsvermogen bezitten dan de χ^2-toets, maar die voor het twee-dimensionale geval niet ontwikkeld zijn. In plaats van echter een groot aantal tijdrovende normaliteitstoetsen te gaan uitvoeren, hebben wij, waar enigszins mogelijk, naast of in plaats van toetsen berustendop normaliteit, zogenaamde parametervrije toetsen toegepast, waarbij geen of nagenoeg geen onderstellingen gemaakt behoeven te worden over de verdeling der onderzochte grootheden.

2.2. Onderzoek van de verdeling van b/ℓ bij de eieren uit 1952.

Voor een toets, toe te passen in 3.2 moeten wij onderstellen dat b/ℓ (een grootheid, die wij ook wel "vormindex" zullen noemen) normaal verdeeld is. Tot de normaliteit van de vormindex mag men niet besluiten uit het voorafgaande. In tegendeel, als b en ℓ beide exact normaal verdeeld zijn, is b/ℓ zeker niet exact normaal verdeeld. Het is echter mogelijk, dat de verdelin-

---

3) Waarnemingen juist op de grens van twee gebieden werden half bij het ene, half bij het andere gebied gerekend.

gen van geen der grootheden b , ℓ en b/ℓ voldoende van normaliteit afwijken, om dit met een toets te onderkennen. Wij hebben daarom ook de verdeling van b/ℓ onderzocht en wel voor het materiaal van 1952 met behulp van de normaliteitstoetsen van GEARY en PEARSON (zie memorandum S 47 (M 16), toegevoegd aan dit rapport). Wij hebben deze toets toegepast op de eieren uit nesten van 2 en de eieren uit nesten van 3 afzonderlijk, omdat de mogelijkheid bestaat dat deze systematisch van vorm verschillen (zie 4). De resultaten zijn als volgt:

Eieren uit nesten van twee:

$n = 51$ (26 nesten van 2 eieren; van één ei zijn de afmetingen foutief opgegeven, zie 1.1).

$a = 0,791$, overschrijdingskans $\gg 0,10$,

$\sqrt{b_1} = -0,0948$, " $\gg 0,05$.

Eieren uit nesten van drie:

$n = 24$ $a = 0,809$, overschrijdingskans $\gg 0,10$

$\sqrt{b_1} = 0,0663$, " $\gg 0,05$.

(De letters a en $\sqrt{b_1}$ zijn gekozen in overeenstemming met het memorandum S 47 (M 16).)

Wij concluderen hieruit, dat ook de verdeling van de vormindex niet constateerbaar van een normale verdeling afwijkt.

3. Variantieanalyse om na te gaan of de spreiding van de afmetingen der eieren binnen de nesten afwijkt van de spreiding over alle eieren.

3.1. Voor b en ℓ van de eieren uit groep 1951 A.

Wij pasten variantieanalyse met één classificatie toe; een toets, b.v. beschreven in M.G.KENDALL, The advanced theory of statistics II (Ch. Griffin, London (1948)). De getoetste hypothese is:

De breedten (resp. lengten) van alle eieren zijn onderling onafhankelijk, zowel van de eieren uit een bepaald nest als van eieren uit verschillende nesten, en alle breedten (resp. lengten) komen uit dezelfde normale verdeling.

Onder deze hypothese heeft de toetsingsgrootheid voor de breedte F_b (voor de lengte F_ℓ), gedefinieerd als:

$$F_b = \frac{n-h}{h-1} \frac{Q_{1,b}}{Q_{2,b}}$$

een F-verdeling van SNEDECOR met $h-1$ en $n-h$ vrijheidsgraden. (Men zie tevens het memorandum S 53 (M 24), toegevoegd aan dit rapport.)

Hierbij is:

$$Q_{1,b} = \sum_{j=1}^{h} (\bar{b}_j - \bar{b})^2 m_j \quad ;$$

$$Q_{2,b} = \sum_{j=1}^{h} \sum_{k=1}^{m_j} (b_{j,k} - \bar{b}_j)^2 \quad ;$$

m_j = aantal eieren van nest j ;
h = aantal nesten;
n = totaal aantal eieren = $\sum_{j=1}^{h} m_j$;

$b_{j,k}$ = breedte van ei k van nest j ;
$\bar{b}_j$ = gemiddelde breedte van de eieren uit nest j ;
$\bar{b}$ = gemiddelde breedte van alle eieren.

Geheel analoog zijn deze grootheden voor de lengte gedefinieerd.

Voor voldoend grote waarden van F_b (hetgeen intuïtief betekent dat de spreiding binnen de nesten, die tot $Q_{2,b}$ bijdraagt, kleiner is dan die van het geheel) zal men de getoetste hypothese verwerpen.

De resultaten waren, uitgaande van het materiaal van 1951 A:

$h-1 = 61 \quad n-h = 93$

$F_b = 3{,}45 \rightarrow$ éénzijdige overschrijdingskans $< 10^{-7}$

$F_\ell = 5{,}84 \rightarrow$ " " $\ll 10^{-7}$.

De conclusie is, dat de getoetste hypothese verworpen wordt ten gunste van de volgende:
De spreidingen van breedte en lengte van de eieren binnen de nesten zijn kleiner dan de overeenkomstige spreidingen berekend over alle waarnemingen tezamen. Er is geen reden om aan te nemen dat een effect, zo duidelijk aanwezig in het materiaal van 1951 A, niet zou optreden in het materiaal van 1951 B of van 1952. Wij hebben daarom de toets voor deze gevallen niet toegepast.

3.2. Voor b/ℓ van de eieren uit 1952.

De toets van 3.1 is hier uitgevoerd op de vormindex b/ℓ aan de hand van het waarnemingsmateriaal van 1952.

Voor de toetsingsgrootheid werd hier gevonden:

$F_{b/\ell} = 4{,}40$ met aantallen vrijheidsgraden

$h-1 = 32 \quad n-h = 41$.

De éénzijdige overschrijdingskans van dit resultaat is 10^{-6}.

De conclusie luidt weer dat de spreiding van b/ℓ binnen de nesten kleiner is dan die van de waarnemingen gezamenlijk.

Op dezelfde overwegingen als in 3.1 aangegeven, hebben wij de toets voor 1951 A en 1951 B niet meer uitgevoerd.

3.3. Consequenties van de resultaten van 3.1 en 3.2.

Uit 3.1 en 3.2 blijkt, dat men de lengten (breedten of vormindices) van de eieren uit een nest, niet kan beschouwen als een aselecte steekproef van onderling onafhankelijke waarnemingen uit de collectie van de lengten van alle kokmeeuweieren van hetzelfde jaar en hetzelfde deel van de kolonie. Tussen de eieren van een nest bestaat een zekere afhankelijkheid. Omdat voor de meeste toetsen vereist is, dat wij werken met reeksen onderling onafhankelijke waarnemingen, hebben wij in de rest van dit rapport gewerkt met nestgemiddelden van b, ℓ en b/ℓ en niet met deze grootheden voor de eieren afzonderlijk, zonder dat dit steeds weer vermeld wordt. Wij hebben niet meer apart getoetst of de verdelingen van deze gemiddelden normaal kunnen zijn. Het is nl. onaannemelijk, dat door het middelen van dicht bij elkaar gelegen waarnemingen uit een bij benadering normaal materiaal, systematische afwijkingen van normaliteit gaan optreden.

4. Toetsing van de verschillen in b, ℓ, b/ℓ en volume tussen de eieren uit nesten met 2 en eieren uit nesten van 3 eieren.

Wij hebben met de toets van WILCOXON [4]) nagegaan of de grootheden b, ℓ, b/ℓ en v (het volume) van tweetallen systematisch afwijken van de overeenkomstige grootheden bij drietallen.

Deze toets is voor het volume toegepast op het materiaal van 1952 (de gegevens hierover voor andere jaren ontbraken) en voor de andere grootheden op de waarnemingen van de groepen 1951 A, 1951 B en 1952 afzonderlijk.

Vervolgens hebben wij de toetsen betrekking hebbend op één grootheid en verschillende groepen gecombineerd op twee verschillende manieren, te weten:

I volgens methode 1 beschreven in S 102 (M 17b) [5]) en

II door combinatie van de éénzijdige overschrijdingskansen beschreven in S 73 (M 17a) [5]).

Zoals betoogd in S 102 (M 17b) worden bij de eerste methode de kleine groepen minder zwaar geteld, bij de tweede methode worden alle groepen, ongeacht hun omvang, gelijkwaardig behan-

---

4) Zie memorandum S 32 (M 7) toegevoegd aan dit rapport.

5) Memorandum toegevoegd aan dit rapport.

deld.

De resultaten, zowel van de afzonderlijke als van de gecombineerde toetsen vindt men in tabel III. Ter oriëntatie zijn hier ook de gemiddelden van b , l , b/l en v per groep vermeld.

Tabel III

Vergelijking van b , l , b/l en v van tweetallen met overeenkomstige grootheden van drietallen met behulp van de toets van WILCOXON.

b = breedte,
l = lengte,
b/l = vormindex,
v = volume,
$h_{(2)}$= aantal tweetallen (nesten van twee eieren),
$h_{(3)}$ = aantal drietallen,
k = tweezijdige overschrijdingskans van de enkelvoudige toets,
k_I = tweezijdige overschrijdingskans van de toets, samengesteld volgens I,
k_{II}= tweezijdige overschrijdingskans van de toets, samengesteld volgens II.

Het + resp. - teken van $U-\mu$ komt overeen met grotere resp. kleinere waarde van de desbetreffende grootheid voor drietallen t.o.v. die voor tweetallen.

		$gem_{(2)}$ in mm	$gem_{(3)}$ in mm	$h_{(2)}$	$h_{(3)}$	$U-\mu$	σ	k	k_I	k_{II}
b	1951 A	36,70	36,34	31	31	- 89	71,03	0,21	0,55	0,78
	1951 B	36,93	36,98	33	32	+ 2,5	76,21	0,98		
	1952	36,62	37,03	25	8	+ 22	23,80	0,37		
		in mm	in mm							
l	1951 A	51,26	50,55	31	31	- 83,5	71,03	0,24	0,015	0,032
	1951 B	52,21	51,26	33	32	- 149	76,21	0,051		
	1952	51,72	50,80	25	8	- 29	23,80	0,23		
b/l	1951 A	0,717	0,720	31	31	+ 14,5	71,03	0,84	0,0505	0,016
	1951 B	0,708	0,722	33	32	+ 141,5	76,21	0,06		
	1952	0,709	0,730	25	8	+ 53,5	23,80	0,023		
		in cm^3	in cm^3							
Vol.	1952	33,39	33,29	19[6)]	7	- 4,5	17,30	0,82		

---

6) Hierbij is inbegrepen het nest waarbij van één ei de breedte niet goed is opgegeven (zie 1.1).

We vinden een systematisch verschil bij de lengte en wel dit: de lengte van eieren uit nesten van twee is groter dan de lengte van eieren uit nesten van drie.

We vinden een aanwijzing dat de vormindex b/ℓ van eieren uit nesten van drie systematisch groter is dan die van eieren uit nesten van twee.

Bij de breedte en het volume worden geen verschillen van enige betekenis gevonden.

5. Correlatie tussen lengte en breedte.

Omdat het materiaal noch over de jaren en grondsoorten, noch over de nestgrootten als homogeen kan worden beschouwd, hebben wij het naar deze criteria gesplitst en voor iedere verkregen groep afzonderlijk de (waargenomen) correlatiecoëfficiënten van b en ℓ berekend. In verband met de afhankelijkheid, die er bestaat tussen de breedten en tussen de lengten van de eieren uit een nest, is de berekening gebaseerd op nestgemiddelden (zie 3.3). Voor de berekening van de correlatiecoëfficiënt zie men memorandum S 102 (M 40) par. 1, toegevoegd aan dit rapport.

Vervolgens hebben wij een gecombineerde toets toegepast om na te gaan of de correlatiecoëfficiënten systematisch van nul afwijken. Om een der methoden, beschreven in S 102 (M 17b) toe te kunnen passen, hebben wij de transformatie:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}$$

toegepast op de correlatiecoëfficiënten r . De grootheid z is onder de hypothese dat ℓ en b niet gecorreleerd zijn bij benadering normaal verdeeld met verwachting 0 en spreiding $\sqrt{\frac{1}{n-3}}$ (zie S 102 (M 40), par. 4). Omdat de spreiding afneemt als het aantal waarnemingen toeneemt, hebben wij combinatiemethode II toegepast, aangezien bij methode I in dit geval juist de kleine steekproeven extra zwaar zouden worden gewogen. De resultaten vindt men in tabel IV.

Het blijkt dat alle gevonden correlatiecoëfficiënten positief zijn; deze positieve correlatie is systematisch te achten op grond van de kleine overschrijdingskans (0,002) der gecombineerde toets. Uiteraard kunnen er nog systematische verschillen bestaan tussen de correlaties voor gedeelten van het materiaal. Op dit gebied hebben wij met behulp van de toets beschreven in S 102 (M 40), par. 5 de verschillen tussen de correlatiecoëfficiënten onderzocht van:

A: 1951 A, 1951 B en 1952 twee aan twee, gecombineerd over tweetallen en drietallen volgens methode 2 uit S 102 (M 17b).

B: Tweetallen en drietallen, gecombineerd over 1951 A, 1951 B en 1952 volgens methode 2 uit S 102 (M 17b).

De resultaten vindt men in tabel V.

Het blijkt dat er systematische verschillen tussen de correlaties bestaan. De correlatie is voor 1952 duidelijk hoger dan de correlatie voor 1951 B en er een aanwijzing, dat zij ook hoger is dan de correlatie voor 1951 A; verder is de correlatie voor drietallen duidelijk hoger dan voor tweetallen.

Tabel IV

Correlatiecoëfficiënt van lengte en breedte

i = nummer waarnemingsgroep,

h_i = aantal nesten van groep i ,

r_i = waargenomen correlatiecoëfficiënt van lengte en breedte berekend uit nestgemiddelden van groep i ,

$$z_i = \frac{1}{2} \ln \frac{1+r_i}{1-r_i}$$

jaar en herkomst	nest-grootte	i	h_i	r_i	z_i
1951 A	tweetallen	1	31	0,237	0,242
	drietallen	2	31	0,325	0,337
1951 B	tweetallen	3	33	0,029	0,029
	drietallen	4	32	0,358	0,375
1952	tweetallen	5	25	0,295	0,304
	drietallen	6	8	0,893	1,437

Gecombineerde toetsingsgrootheid (volgens S 102 (M 17b) methode 2):

$$\frac{1}{\sqrt{6}} \sum_{i=1}^{6} z_i \sqrt{h_i - 3} = 3,70$$

Tweezijdige overschrijdingskans: 0,0002.

Tabel V

Verschillen tussen correlatiecoëfficiënten van lengte en breedte.

u = toetsingsgrootheid, berekend volgens S 102 (M 40), par. 5.

vergeleken groepen	verdere indeling	u	gecombineerde toetsingsgrootheid	tweez. overschr. kans
1951 A-1951 B	tweet. driet.	0,810 0,143	$\frac{0,810 + 0,143}{\sqrt{2}} = 0,674$	0,75
1951 A-1952	tweet. driet.	-0,218 -2.268	$\frac{-0,218 - 2,268}{\sqrt{2}} = -1,76$	0,08
1951 B-1952	tweet. driet.	-0,979 -2,194	$\frac{-0,979 - 2,194}{\sqrt{2}} = -2,24$	0,03
tweet.-driet.	1951 A 1951 B 1952	-0,356 -1,331 -2,284	$\frac{-0,356-1,331-2,284}{\sqrt{3}} = 2,29$	0,02

6. Toetsing van de verschillen in b, ℓ en b/ℓ tussen de groepen waarnemingen.

6.1. Verschillen in de afmetingen van eieren uit 1951 A en uit 1951 B.

In 1951 werden de eieren in 2 verschillende gebieden geraapt (deze gebieden worden aangegeven met A en B; zie ook 1.1). Bij de waarnemingen is opgegeven welke eieren van gebied A en welke van gebied B afkomstig zijn en wij onderzoeken nu of tussen de afmetingen van de eieren uit deze 2 groepen systematische verschillen bestaan. Dit onderzoek werd voor de breedte b, de lengte ℓ en de vormindex b/ℓ uitgevoerd met behulp van de toets van WILCOXON. De toets werd apart uitgevoerd voor de tweetallen en de drietallen waarna de resultaten werden gecombineerd volgens de methode 1 van S 102 (M 17b).

De resultaten zijn vermeld in tabel VI (zie blz. 13). Ter oriëntatie zijn hierin de gemiddelden van de genoemde grootheden opgenomen.

Uit deze resultaten concluderen wij: de lengte en breedte van de eieren van 1951 B zijn systematisch groter dan de lengte en breedte van de eieren van 1951 A. Zoals echter uit de opgegeven gemiddelden blijkt zijn de verschillen niet groot. Tussen de vormindices van beide groepen zijn geen systematische verschillen gevonden.

Tabel VI

Vergelijking van de grootheden b , l en b/l van eieren uit 1951 A met die van eieren uit 1951 B met behulp van de toets van WILCOXON.

$h_{(A)}$ = aantal nesten van de desbetreffende soort (twee- of drietallen) in 1951 A

$h_{(B)}$ = idem 1951 B

k = tweezijdige overschrijdingskansen van de toetsen afzonderlijk

k_I = tweezijdige overschrijdingskansen van de gecombineerde toetsen

Het + resp. - teken van $U-\mu$ betekent dat b , l of b/l bij de gemeten eieren van 1951 B groter resp. kleiner is dan bij die van 1951 A.

		gem. 1951 A	gem. 1951 B	$h_{(A)}$	$h_{(B)}$	$U-\mu$	σ	k	k_I
b	tweet.	367,0	369,3	31	33	+ 87,5	74,44	0,24	0,013
	driet.	363,4	369,8	31	32	+ 172	72,74	0,018	
l	tweet.	512,6	522,1	31	33	+ 131	74,44	0,08	0,019
	driet.	505,5	512,6	31	32	+ 114,5	72,74	0,12	
b/l	tweet.	0,717	0,708	31	33	- 89	74,44	0,23	0,72
	driet.	0,720	0,722	31	32	+ 51,5	72,74	0,48	

6.2. Verschillen in de afmetingen van eieren uit 1951 en uit 1952.

Bij dit onderzoek naar eventuele verschillen in de afmetingen van de eieren uit verschillende jaren komt 1950 niet voor. Dit vindt zijn verklaring in het feit dat bij de waarnemingen van 1950 niet onderscheiden is naar de nesten, waardoor de in 3 gevonden afhankelijkheid tussen de eieren uit hetzelfde nest niet kan worden uitgeschakeld. Zoals in 1.1 is opgemaakt is in het materiaal van 1952 geen onderscheid gemaakt naar het gebied A of B, waar de eieren geraapt zijn. Wij konden nu alleen het gehele materiaal van 1952 vergelijken met dat van 1951 A en met dat van 1951 B apart. Dit is gedaan met de toets van WILCOXON voor de b , l en b/l . Op dezelfde wijze als bij 6.1 is de toets apart uitgevoerd voor twee- en drietallen en later gecombineerd. De resultaten vindt men in tabel VII (zie blz. 14).

Tabel VII

Vergelijking van de grootheden b , l en b/l van eieren uit 1951 met die van eieren uit 1952 met behulp van de toets van WILCOXON.

Voor de betekenis van $h_{(A)}$, $h_{(B)}$, k en k_I in deze tabel zie men tabel VI.

$h_{(1952)}$ = aantal nesten van twee resp. 3 in het materiaal van 1952.

Het + resp. - teken van $u-\mu$ betekent dat b , l of b/l bij de gemeten eieren van 1952 groter resp. kleiner is dan bij die van 1951 (A of B).

1951 A - 1952

		gem. 1951 A	gem. 1952	$h_{(A)}$	$h_{(1952)}$	$u-\mu$	σ	k	k_I
b	tweet.	367,0	366,2	31	25	- 8,5	60,67	0,90	0,52
	driet.	363,4	370,3	31	8	+ 52	28,75	0,07	
l	tweet.	512,6	517,2	31	25	+ 56,5	60,67	0,36	0,35
	driet.	505,5	508,0	31	8	+ 7,5	28,75	0,82	
b/l	tweet.	0,717	0,709	31	25	- 76,5	60,67	0,21	0,55
	driet.	0,720	0,730	31	8	+ 36,5	28,75	0,21	

1951 B - 1952

		gem. 1951 B	gem. 1952	$h_{(B)}$	$h_{(1952)}$	$u-\mu$	σ	k	k_I
b	tweet.	369,3	366,2	33	25	- 75	63,69	0,24	0,33
	driet.	369,8	370,3	32	8	+ 7	29,57	0,83	
l	tweet.	522,1	517,2	33	25	- 65,5	63,69	0,31	0,25
	driet.	512,6	508,0	32	8	- 17	29,57	0,58	
b/l	tweet.	0,708	0,709	33	25	+ 6	63,69	0,93	0,67
	driet.	0,722	0,730	32	8	+ 25	29,57	0,41	

Conclusie: We vinden bij de onderzochte grootheden geen systematische afwijkingen vàn het materiaal uit 1952 t.o.v. dat uit 1951 A en 1951 B. Het is echter mogelijk enerzijds door het in 6.1 gevonden verschil in lengte en breedte tussen A en B en anderzijds door de onbekende samenstelling van 1952 naar A en B, dat een eventueel bestaand verschil tussen de jaren niet gevonden wordt.

7. Onderzoek naar de vraag of de vormindex b/l van het eerstgelegde ei uit een nest kleiner is dan de b/l van de volgende eieren.

7.1. Inleiding.

Deze vraag is moeilijk te onderzoeken daar van de eieren niet bekend is welke eieren in de verschillende nesten het eerst gelegd zijn. Alleen bij gegeven volgorde van legging kan deze vraag direct getoetst worden. Wanneer echter ook bij kokmeeuwen het eerstgelegde ei in de regel een kleine b/l heeft (bij kippen is dit een bekend verschijnsel) dan is dit wellicht op indirecte wijze te toetsen aan andere verschijnselen die er door veroorzaakt worden.

Het resultaat van 4 (de vormindex van eieren uit nesten van drie is systematisch groter dan van eieren uit nesten van twee) zou in dit licht bezien kunnen worden, daar een eerste ei met kleine b/l het nestgemiddelde van de b/l meer kleiner maakt bij een tweetal dan bij een drietal.

Ook zou men de in 5 gevonden kleinere correlatie tussen lengte en breedte van tweetallen ten opzichte van drietallen kunnen verklaren doordat een afwijkend eerste ei die correlatie sterker verstoort bij tweetallen dan bij drietallen.

We kunnen verder iets over de gestelde vraag zeggen naar aanleiding van het onderzoek naar de correlatie binnen de nesten van vormindex en volume van de eieren. Daarom wordt dit correlatieonderzoek bij deze paragraaf opgenomen.

7.2 Correlatie b/l -volume.

Getoetst is met behulp van de methode van m rangschikkingen of, binnen de nesten, b/l en volume van de eieren gecorreleerd zijn. In het geval dat tweetallen beschouwd worden komt de methode van m rangschikkingen neer op het uitvoeren van de tekentoets [7].

Resultaten.

Tweetallen

+ resp. - betekent: Het ei met het grootste volume uit een nest heeft ook de grootste resp. kleinste b/l .

aantal nesten = 18
aantal + = 15
aantal - = 3

---

7) De methode van m rangschikkingen en de tekentoets zijn beschreven in de memoranda S 47 (M 14) resp. S 53 (M 22), die toegevoegd zijn aan dit rapport.

We vinden hieruit een tweezijdige overschrijdingskans $k = 0,008$.

Drietallen

m = aantal rijen = aantal nesten = 7.

Op een rij worden de rangnummers van b/l van de eieren uit dat nest (1, 2 of 3 al naar gelang de grootte) geplaatst naar opklimmende grootte van het volume van die eieren.

Daar de nesten uit 3 eieren bestaan is het aantal kolommen dus 3. We bepalen de sommen van de rangnummers uit elke kolom. Wanneer geen correlatie bestaat is de verwachting van het kolomtoaal 7 x de verwachting van het rangnummer = 7 x 2 = = 14.

We vonden:	kl. vol	mi. vol.	gr. vol.
Kolomtotalen b/l :	15	10	17

$S = (15-14)^2 + (10-14)^2 + (17-14)^2 = 26$.

Overschrijdingskans $k = 0,19$.

Conclusies. Bij de tweetallen vinden we een systematische correlatie: Het ei met het kleinste volume uit een nest heeft in het algemeen de kleinste b/l .

Bij de drietallen vinden we geen systematische correlatie en ook de bij de tweetallen gevonden tendens zet zich hier niet voort. Het eerste kolomtotaal was hoger dan het gemiddelde, terwijl wij een lagere uitkomst verwachten wanneer het bij de tweetallen gevonden effect ook hier tot uiting was gekomen.

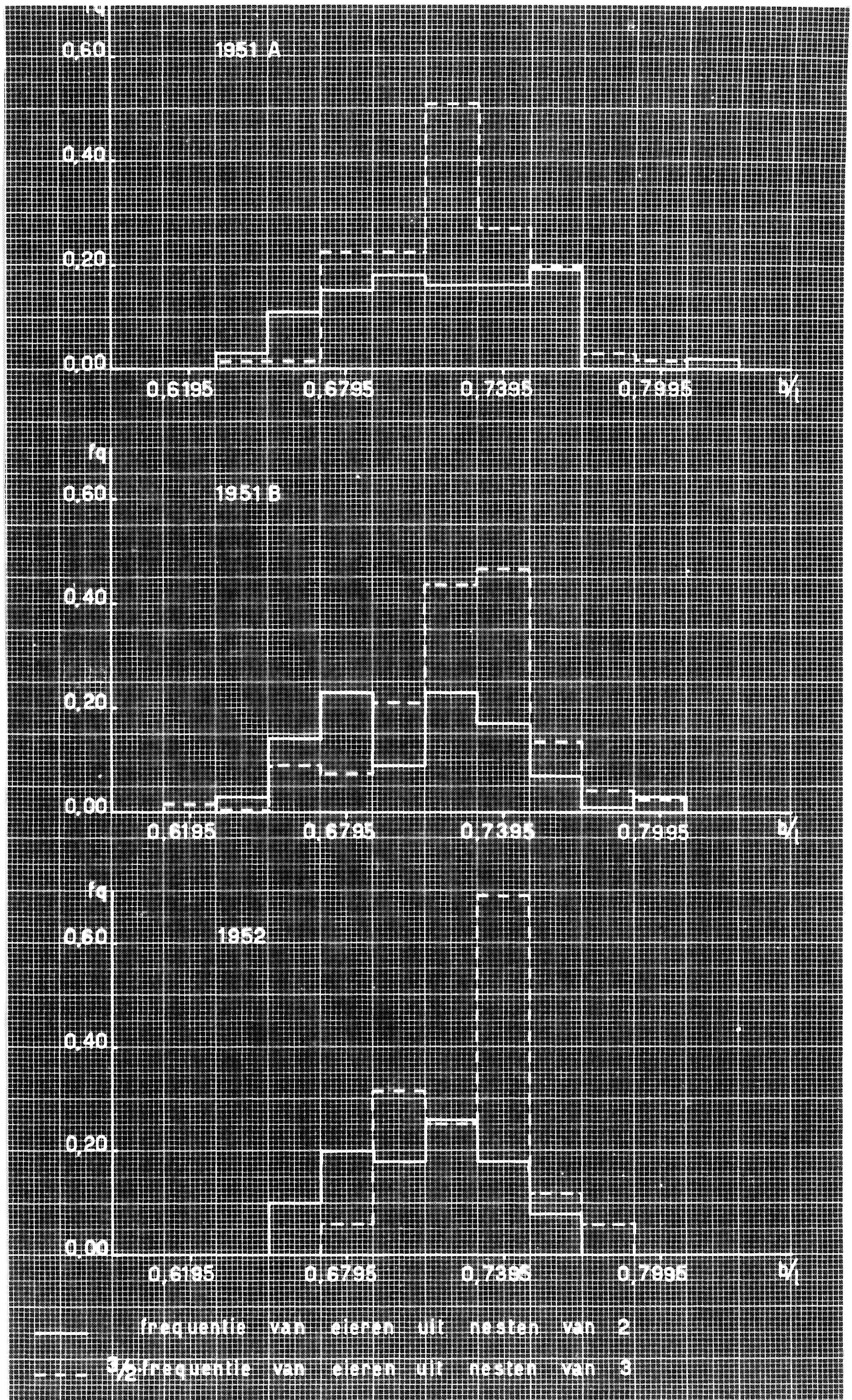
We mogen hieruit echter niet concluderen dat het effect in werkelijkheid bij drietallen niet zou bestaan. Het aantal van slechts 7 drietallen, waarvan wij over gegevens beschikten, is daarvoor veel te gering.

Wanneer wij onderstellen dat het eerstgelegde ei uit een nest in het algemeen de kleinste inhoud heeft kan het bij tweetallen gevonden resultaat een aanwijzing zijn dat het eerste ei in het algemeen ook de kleinste b/l heeft.

7.3. Frequentieverdeling van b/l bij eieren uit nesten van twee en uit nesten van drie.

Om een betere indruk te krijgen van de verdeling van de vormindex zijn histogrammen getekend die de frequentiequotiënten aangeven van de aantallen eieren, waarvan de b/l in een bepaald interval (groot 0,02) ligt. De histogrammen zijn apart getekend voor 1951 A, 1951 B en 1952, waarbij die voor twee- en voor drietallen telkens in één figuur zijn verenigd.

De frequentiequotiënten zijn hier bepaald van de b/l -'s van de eieren apart en niet van de nestgemiddelden, omdat wij


fq
0,60
1951 A
0,40
0,20
0,00
0,6195
0,6795
0,7395
0,7995
b/l
fq
0,60
1951 B
0,40
0,20
0,00
0,6195
0,6795
0,7395
0,7995
b/l
fq
0,60
1952
0,40
0,20
0,00
0,6195
0,6795
0,7395
0,7995
b/l
frequentie van eieren uit nesten van 2
3/2 frequentie van eieren uit nesten van 3

alleen zó eventueel iets kunnen opmerken over de $^b/_\ell$ van de eerstgelegde eieren. De frequentiequotiënten van de eieren uit nesten van drie zijn in de figuren vermenigvuldigd met $\frac{3}{2}$, waardoor een betere vergelijking wordt verkregen tussen de eieren uit nesten van twee en drie. Deze vermenigvuldiging heeft namelijk als gevolg dat de oppervlakten, die door de $^b/_\ell$-'s van de eerste en van de tweede eieren bij de histogrammen van twee- en drietallen worden ingenomen, gelijk zijn. De grotere oppervlakte van het histogram bij de drietallen is dan toe te schrijven aan de derde eieren.

Zoals uit de figuren blijkt ligt het extra oppervlak bij de drietallen in het algemeen bij vrij grote $^b/_\ell$, wat er dus op zou wijzen dat de derde eieren in het algemeen een tamelijk grote $^b/_\ell$ bezitten. Dit geeft enige steun aan de onderstelling dat de eerste eieren uit de nesten in het algemeen een kleine $^b/_\ell$ bezitten.

Verder merkt men op dat kleine $^b/_\ell$-'s bij de tweetallen meer voorkomen dan bij de drietallen.

7.4. Resultaten van 7.1, 7.2 en 7.3.

Voor alle verschijnselen die in deze paragraaf voorkwamen en die hier bezien zijn uit het oogpunt van een afwijkend eerste ei, zijn echter ook andere verklaringen te bedenken (bij voorbeeld een eventueel verschil tussen de eieren van jonge en van oude kokmeeuwen). Op grond van dit onderzoek alleen is een ondubbelzinnige conclusie dan ook niet mogelijk, al is een aanwijzing voor het onderzochte effect wel aanwezig. Voor een ondubbelzinnige conclusie zou bekend moeten zijn in welke volgorde de eieren van elk nest gelegd zijn. Gezien de aanwijzing voor het onderzochte effect die deze paragraaf bevat, lijkt een verder onderzoek in deze richting gerechtvaardigd.

8. Samenvatting der conclusies.

Wij kunnen de conclusies van dit onderzoek op de volgende wijze samenvatten, aansluitend bij de vragen in 1.2 vermeld.

1) Voor de simultane verdeling van b en ℓ werd geen systematische afwijking van een tweedimensionale normale verdeling gevonden; men mag daarom de verdelingen van b en ℓ afzonderlijk bij benadering als normaal beschouwen. Tevens konden wij bij de verdeling van $^b/_\ell$ geen systematische afwijkingen van normaliteit vinden (zie 2).

2) Het blijkt, dat de spreidingen van b , ℓ , en $^b/_\ell$ binnen de nesten systematisch kleiner zijn dan de spreidingen van deze

grootheden over de eierencollecties (1951 A, 1951 B en 1952) in hun geheel. Men kan dus de eieren uit een nest niet als een aselecte steekproef uit de collectie beschouwen (zie 3).

3) De eieren uit nesten van twee zijn systematisch langer dan de eieren uit nesten van drie Mogelijk tengevolge daarvan is de vormindex b/l voor de eieren uit nesten van twee systematisch kleiner dan de vormindex voor eieren uit nesten van drie, want voor de breedten werd geen systematisch verschil gevonden. Evenmin werd een systematisch verschil gevonden tussen de volumina van eieren uit nesten van twee en uit nesten van drie (zie 4).

4) Er is een systematische positieve correlatie tussen de lengte en de breedte van kokmeeuweieren; deze correlatie is niet voor alle deelgroepen even groot. Zij is voor eieren uit nesten van drie duidelijk hoger dan voor eieren uit nesten van twee en voor het jaar 1952 duidelijk hoger dan voor de groep 1951 B (zie 5).

5) Voor het jaar 1951 zijn de eieren van de zandgrond (1951 B) duidelijk groter (d.w.z. zowel langer als breder) dan de eieren van slikgrond (1951 A) (zie 6).

6) Wij konden geen verschil vinden tussen de afmetingen van de eieren uit 1951 en uit 1952 (zie 6).

7) Over de collectie van nesten met twee eieren uit 1952 constateren wij, dat het ei met het grootste volume systematisch ook de grootste vormindex (b/l) heeft. Een dergelijk effect is bij de nesten van drie niet aangetoond; wellicht tengevolge van het kleine aantal waarnemingen daaromtrent (zie 7).

8) Het vermoeden, dat het eerstgelegde ei uit een nest een kleinere vormindex heeft, konden wij door de aard van het materiaal niet direct toetsen Wel werden een aantal verschijnselen gevonden, die onder andere met dit effect kunnen worden verklaard (zie 7).

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

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

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (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 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 H_0, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ 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 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.

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

Maart, 1952.

De toets van Wilcoxon.[1])

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

Voor het toetsen van de hypothese H_o wordt gebruik gemaakt van een toetsingsgrootheid $\underline{U}$ [2]), die als volgt uit de waarnemingen berekend wordt. Onderstellen we, dat de waarnemingen $x_1,\ldots,x_n$ en $y_1,\ldots,y_m$ naar opklimmende grootte gerangschikt zijn, dan bepalen we eerst het aantal waarnemingen uit de tweede steekproef, dat kleiner is dan de kleinste waarneming x_1 uit de eerste steekproef (bij gelijkheid tellen wij $\frac{1}{2}$ in plaats van 1). Noem dit aantal V_1. Vervolgens wordt het aantal waarnemingen uit de tweede steekproef bepaald, dat kleiner is dan de op één na kleinste waarneming x_2 uit de eerste steekproef (bij gelijkheid wordt weer een $\frac{1}{2}$ in plaats van 1 geteld). Dit aantal noemen we V_2. Evenzo worden met betrekking tot $x_3,x_4,\ldots,x_n$ de aantallen $V_3,V_4,\ldots,V_n$ bepaald. De waarde U van de toetsingsgrootheid $\underline{U}$ wordt voor de twee steekproeven dan gegeven door

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

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

---

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

[2]) Stochastische grootheden worden door onderstreping aangeduid.

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

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

en

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

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 eenttabel van de normale verdeling worden gevonden.
Bij eenzijdige toetsing wordt α door 2α vervangen, resp. k gehalveerd.
Een bijzonder geval van het bovenstaande is, dat onder de waarnemingen voor $\underline{x}$ en $\underline{y}$ in 't geheel geen gelijken voorkomen.
In dat geval kan de uitdrukking voor de variantie herleid worden tot

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

---

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

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

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

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

Litteratuur:

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

MATHEMATISCH CENTRUM
2de Boerhaavestraat 49
A M S T E R D A M O.
Statistische Afdeling
S47 (M12)

De χ^2-toets voor aanpassing [1])

De χ^2-toets voor aanpassing dient voor het toetsen van de hypothese H_0, dat een stelsel waarden $x_1, x_2, ---, x_n$ een steekproef is uit een continue waarschijnlijkheidsverdeling met verdelingsdichtheid $f(x)$. Hierbij wordt een grootheid χ^2 waarvan de verdelingsfunctie bekend is, als toetsingsgrootheid gebruikt.
De verdelingsfunctie van χ^2 wordt nog nader gekarakteriseerd door een grootheid ν welke het aantal vrijheidsgraden van χ^2 aangeeft. De grootte van χ^2 en het getal ν worden op de volgende wijze bepaald:

Wanneer $f(x)$ de onderstelde verdelingsdichtheid is (fig.1)

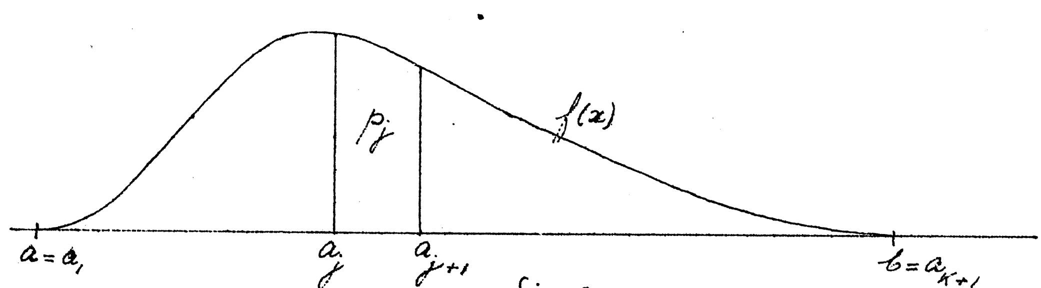


fig.1

wordt het interval $[a, b]$ in K stukken verdeeld, zodanig dat boven ieder interval $[a_j, a_{j+1}]$ een oppervlak p_j ligt $(j = 1, \ldots, K)$ Daar het gehele oppervlak onder de kromme de totale waarschijnlijkheidsmassa voorstelt, dus gelijk aan 1 is, geldt de betrekking $p_1 + p_2 + \cdots + p_K = 1$
Bestaat nu de steekproef uit n waarnemingen $x_1, x_2, \ldots, x_n$ dan zal men onder de hypothese H_0 in het algemeen verwachten, dat ongeveer np_j van de n steekproef-waarnemingen in het interval $[a_j, a_{j+1}]$ komen te liggen. Is het werkelijke aantal waarnemingen uit de steekproef, dat in het interval $[a_j, a_{j+1}]$ valt n_j (dus n_1 waarden in $[a_1, a_2]$ etc.) dan wordt χ^2 als volgt gedefinieerd:

$$\chi^2 = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} + \cdots\cdots + \frac{(n_K - np_K)^2}{np_K}$$

---

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

Deze χ^2 is een maat voor de afwijking van de werkelijke verdelingsdichtheid van x van de ondersteld verdelingsdichtheid $f(x)$. Is de werkelijke verdelingsdichtheid van x inderdaad gelijk aan $f(x)$ zoals volgens H_0 ondersteld wordt, dan zal χ^2 in het algemeen klein zijn; is de verdelingsdichtheid echter verschillend van $f(x)$, dan worden grote waarden van χ^2 waarschijnlijker dan het geval is, als H_0 juist is. De kritieke zône voor deze toets geeft men daarom de gedaante $\chi^2 > \chi_0^2$, waarbij χ_0^2 correspondeert met een onbetrouwbaarheidsdrempel α en uit tabellen voor de χ^2-verdeling opgezocht kan worden, waarbij rekening gehouden moet worden met het in de tabellen aangegeven:

<u>Aantal vrijheidsgraden</u> ν. Dit getal wordt als volgt bepaald: zijn er l parameters van $f(x)$ (b.v. het gemiddelde, de spreiding e.d.) uit de steekproef bepaald, dan is $\nu = k-l-1$, waarin k het aantal intervallen $[a_j, a_{j+1}]$ (zie boven) is. Is $f(x)$ van tevoren geheel gespecificeerd, dan is dus $\nu = k-1$.

<u>Voorbeelden</u>: bij toetsing op bovenstaande wijze van de hypothese, dat de steekproef $x_1, \ldots, x_n$ uit een normale verdeling ("verdeling van Gauss") afkomstig is, worden gemiddelde en spreiding van de steekproef gebruikt om $f(x)$ te bepalen. Door gemiddelde en spreiding is de normale verdeling geheel bepaald. Dan is dus $\nu = k-3$. Wenst men echter de hypothese te toetsen, dat de steekproef afkomstig is uit een volledig gegeven normale verdeling, waarbij dus gemiddelde en spreiding ook gegeven zijn, dan behoeven deze parameters niet uit de steekproef geschat te worden en is $\nu = k-1$.

<u>Opm</u>. In de practijk worden de getallen $p_1, p_2, \ldots p_k$ meestal even groot gekozen, dus alle gelijk $\frac{1}{k}$. Het aantal delen k laat men bovendien nog afhangen van het aantal waarnemingen, waaruit de steekproef bestaat. De benadering, die bij het toepassen van de χ^2-toets gebruikt wordt, is alleen dan behoorlijk indien het aantal k zodanig gekozen wordt, dat het verwachte aantal waarden in de intervallen $[a_j, a_{j+1}]$ d.i. np_j niet al te klein is, b.v. $\geqq 10$.

<u>Litteratuur</u>:

M.G.Kendall, The advanced Theory of Statistics
London 1947, Deel I hfdst. 12.

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 coëfficiënt van overeenstemming

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

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

De verdeling van $\underline{S}$ 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 meeste 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,
Amsterdam-O.

Statistische Afdeling
S 47 (M 16)

Normaliteitstoetsen van Geary en Pearson[1].

Om de hypothese $\mathcal{H}_0$ te toetsen, dat een reeks waarnemingen $x_1, x_2, \ldots x_n$ beschouwd kan worden als een steekproef uit een normale verdeling, d.i. een verdeling met verdelingsdichtheid

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

zijn door Geary en Pearson de verdelingen, onder hypothese $\mathcal{H}_0$, berekend van enkele statistische grootheden, welke betrekking hebben op de vorm van de kromme $f(x)$.

De eerste grootheid is de volgende:

$$(1) \quad \underline{a} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{\sqrt{n \sum_{i=1}^{n} (x_i - \bar{x})^2}} \qquad \text{waarin } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ is.}$$

Deze grootheid a is een maat voor de "kurtosis" ("platheid" of "slankheid" der kromme $f(x)$). In figuur 1 zijn 3 verdelingsdichtheden $f(x)$, $f_1(x)$ en $f_2(x)$ getekend waarvan $f(x)$ een normale is, terwijl $f_1(x)$ te slank is (en te dik in de staarten), en $f_2(x)$ te plat. (en te dun in de staarten). Pearson noemt $f_1(x)$ "leptocurtic" en $f_2(x)$ "platycurtic".

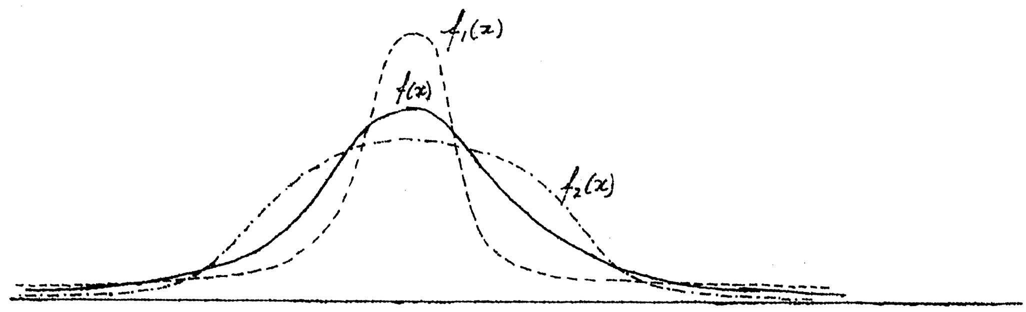


fig. 1

De kritieke zône voor $\underline{a}$ bestaat uit grote en kleine waarden van a.

---

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

Als tweede toetsingsgrootheid wordt gebruikt:

$$(2) \qquad \sqrt{b_1} = \sqrt{n}\,\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{\frac{3}{2}}}$$

Deze grootheid is een maat voor de "scheefheid" (Engels: "skewness") van de verdeling. In figuur 2 zijn 3 verdelingsdichtheden getekend, waarvan $f(x)$ symmetrisch is, terwijl $f_1(x)$ "positief scheef" (omdat de "dikke staart" rechts ligt) en $f_2(x)$ "negatief scheef" is.

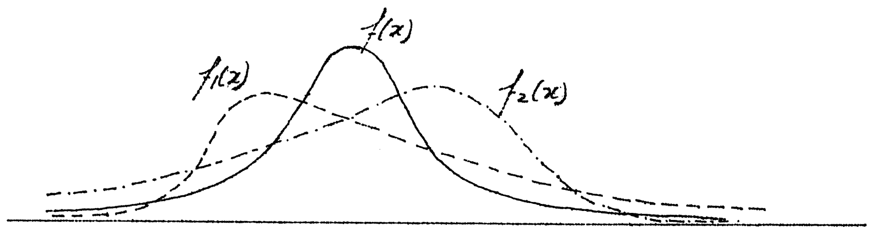


fig. 2

De kritieke zône voor de grootheid $\sqrt{b_1}$, bestaat uit grote en kleine waarden van $\sqrt{b_1}$, (daarbij wordt gewoonlijk ook voor negatieve waarden de notatie $\sqrt{b_1}$ gebruikt).

De derde toetsingsgrootheid

$$(3) \qquad b_2 = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^2}$$

is weer een maat voor de "kurtosis" der kromme en wordt alleen voor zeer grote steekproeven berekend. Voor kleinere steekproeven is de maat a voldoende.

De kritieke zône bestaat uit grote en kleine waarden van b_2.

Nomogrammen der 3 grootheden:

1e. Voor de grootheid a zijn onder de hypothese H_0 nomogrammen en tabellen berekend voor de onbetrouwbaarheidsdrempels 0,01; 0,05 en 0,10, waarbij $10 \leqq n \leqq 1000$.

2e. Voor de grootheid $\sqrt{b_1}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05, $25 \leqq n \leqq 1000$.

3e. Voor de grootheid $\sqrt{b_2}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05. Hierbij is $100 \leqq n \leqq 1000$.

<u>Literatuur:</u>

R.C.Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, Biometrika <u>28</u> (1936) p.295.

R.C.Geary and E.S.Pearson, Tests of Normality, Biometrika Office, London 1938.

Bovenstaande nomogrammen en tabellen komen in deze beide publicaties voor.
Alle daarin genoteerde overschrijdingskansen zijn <u>éénzijdig</u>.

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.
Statistische Afdeling

April 1952.

S 73 (M 17a)

door

Dr J.Hemelrijk Jr

Het combineren van onafhankelijke toetsen[1].

A. Tweezijdige toetsing van éénzelfde of een aantal soortgelijke hypothesen.

1. Eenvoudige beschrijving van de methode.

Wij beschouwen eerst een zeer eenvoudig geval, waarbij één hypothese H_o een aantal (en wel h) malen getoetst wordt door toepassing van één bepaalde toets op de resultaten van h onafhankelijke experimenten. Zij $\underline{t}$ de toetsingsgrootheid[2], waarop de toetsen berusten en laten de bij de h toetsen door $\underline{t}$ aangenomen waarden $t_1, t_2, \ldots, t_h$ zijn.

De linker-éénzijdige overschrijdingskansen, die bij deze uitkomsten behoren, zijn

$$k'_1 = P\left[\underline{t} \leqq t_1 \mid H_o\right], \ldots, k'_1 = P\left[\underline{t} \leqq t_h \mid H_o\right],$$

waarin het symbool H_o, achter de verticale streep, aangeeft, dat deze kansen berekend zijn in de onderstelling, dat H_o juist is. De rechter-éénzijdige overschrijdingskansen zijn

$$k''_1 = P\left[\underline{t} \leqq t_1 \mid H_o\right], \ldots, k''_h = P\left[\underline{t} \leqq t_h \mid H_o\right].$$

Wij berekenen nu de twee producten

$$k'_1 . k'_2 . \ \ldots \ . k'_h \quad \text{en} \quad k''_1 . k''_2 . \ \ldots \ . k''_h ,$$

en nemen van deze twee het kleinste. Dit noemen wij K. Vervolgens berekenen wij

$$Q = - 2 \ln K$$

en zoeken in een tabel of nomogram van de χ^2-verdeling de overschrijdingskans

$$k_{\chi^2} = P\left[\underline{\chi}^2 \geqq Q\right]$$

op, waarbij als aantal vrijheidsgraden 2h genomen

---

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

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

wordt. De tweezijdige overschrijdingskans, behorende bij de gecombineerde toetsingen van H_o, is dan gelijk aan $2k_{\chi^2}$.

2. Opmerkingen.

In plaats van de kleinste van de twee producten van éénzijdige overschrijdingskansen te gebruiken, kan men ook alleen het product van de tweezijdige overschrijdingskansen der oorspronkelijke toetsingen berekenen en dit als K nemen. De overschrijdingskans k_{χ^2} behoeft dan niet meer verdubbeld te worden. Dit is het procédé, dat in de literatuur ([1], [2][3]) wordt aangegeven. De bovenbeschreven methode bezit echter vermoedelijk in vele gevallen een aanzienlijk groter onderscheidingsvermogen en is daarom te prefereren. Een publicatie hierover is in voorbereiding.

Indien de toetsingsgrootheid t een discontinue verdeling bezit, is bovenstaande methode in zoverre nog van toepassing, dat de gevonden waarde van $2k_{\chi^2}$ te groot is. Dat wil zeggen, indien men H_o slechts dan verwerpt, als

$$2k_{\chi^2} \leq \alpha$$

is, is de kans op ten onrechte verwerpen van H_o kleiner dan α (dus de onbetrouwbaarheid van de methode wordt op deze wijze overschat). Dat het verschil aanzienlijk kan zijn blijkt uit [3].

De opheffing van deze moeilijkheid kan o.a. bereikt worden door de verdeling van t continue te maken; een beschrijving daarvan vindt men in [4].

3. Het algemene geval.

In het algemene geval, waarin bovenstaande methode kan worden toegepast, zijn er h hypothesen $H_1, H_2, \ldots, H_h$, die getoetst worden met h toetsingsmethoden, $T_1, T_2, \ldots, T_h$. De hypothesen en de toetsen kunnen nu verschillend zijn; de toetsen berusten echter steeds op h onderling onafhankelijke stelsels van waarnemingen. Van ieder der toetsen wordt ondersteld, dat de kritieke zone één- of tweezijdig kan zijn, dus dat een tweezijdige kritieke zone (bij gegeven onbetrouwbaarheidsdrempel α) in twee delen uiteenvalt, die wij de linker- en rechter-éénzijdige kritieke zone noemen.

----

3) Zie literatuurlijst aan het einde van dit memorandum.

Indien nu de verzameling van alternatieve hypothesen, waartegen het stelsel $H_1, H_2, \ldots, H_h$ getoetst wordt, van dien aard is, dat bij éénzijdige toetsing alle hypothesen links-éénzijdig of alle rechts-éénzijdig zouden worden getoetst, is de boven beschreven methode van toepassing.

4. Tabellen.

Tabellen van de χ^2-verdeling vindt men in vrijwel ieder boek over statistiek. B.v. in M.G.Kendall, The advanced theory of statistics, Griffin and Co., London 1947,
W.J.Dixon and F.J.Massey, Introduction to statistical analysis, Mc Graw-Hill, N.Y.-Toronto-London 1951.

Een speciale tabel voor het combineren van toetsen, die enkele (overigens eenvoudige) berekeningen bespaart, is te vinden in
F.N. David, On the P_{λ_h} test for randomness, Biometrika 26 (1934) pp. 1-11.

Literatuur.

[1] R.A.Fisher, Statistical methods for research workers, Oliver and Boyd, London, 4th ed., 1932.

[2] E.S.Pearson, The probability integral transformation for testing goodness of fit and combining independent tests of significance, Biometrika 30 (1938) pp. 134-148.

[3] W.A.Wallis, Compounding probabilities from independent significance tests, Econometrica 10 (1942) pp. 229-248.

[4] E.S.Pearson, On questions raised by the combination of tests based on discontinuous distributions, Biometrika 37 (1950) pp. 383-398.

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

Statistische Afdeling
Rapport S 102 (M 17b)

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

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

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

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

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

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

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

---

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

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

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

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

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

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

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

Wij geven hier 3 combinatiemethoden van dit type:

Methode 1: $c_1 = c_2 = \cdots\cdots = c_h = 1$

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

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

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

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

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

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

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

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

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

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

Literatuur:

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

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

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

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

Statistische Afdeling
S 53 (M 22)

Tekentoets[1)]

Deze toets dient voor het toetsen van de hypothese H_0, 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_0 juist is:

$$P\left[\underline{n}_1 = n_1 \mid H_0\right] = \binom{m}{n_1} 2^{-m}.$$

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 $\underline{x}_i$ $(i=1,\ldots,n)$ 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 $\underline{z}_i=\underline{x}_i-\underline{y}_i$ $(i=1,\ldots,n)$.

---

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 vóór de helft bij de negatieve te tellen; de door ons gebruikte methode geeft de toets een groter onderscheidingsvermogen.

MATHEMATISCH CENTRUM
2de Boerhaavestr. 49
A M S T E R D A M O.
Statistische Afdeling
S 53 (M 24)

z-toets van Fisher[1)]

Stel we hebben twee onafhankelijke steekproeven

$x_1 \ldots x_{n_1}$ met gemiddelde $\bar{x}$
$y_1 \ldots y_{n_2}$ met gemiddelde $\bar{y}$,

beide uit een normale verdeling.

De te toetsen hypothese is dan, dat de spreidingen van deze normale verdelingen gelijk zijn.

Is

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_i-\bar{x})^2$$

en

$$s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (y_i-\bar{y})^2$$

dan is de toetsingsgrootheid:

$$z = \tfrac{1}{2} \lg \frac{s_1^2}{s_2^2}$$

waarbij van s_1^2 en s_2^2 de grootste in de teller van de breuk genomen wordt.

Als H_0 juist is zal z in het algemeen klein zijn en de kritieke zône bestaat dus uit grote waarden van z. Men kan als toetsinggrootheid ook nemen

$$F = e^{2z} = \frac{s_1^2}{s_2^2} .$$

De kritieke zône bestaat dan uit grote waarden van e^{2z}. De toets wordt dan gewoonlijk de F-toets van Snedecor genoemd.

In de verdelingen van z en van e^{2z} komen twee parameters voor: $\nu_1 = n_1 - 1$ en $\nu_2 = n_2 - 1$, die het aantal vrijheidsgraden in de teller resp. in de noemer genoemd worden.

Litteratuur: M.G.Kendall: The advanced theory of statistics: deel II pag.115.
P.G.Hoel: Introduction to mathematical statistics; pag. 152-154.

---

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

Vervolg litteratuuropg.:

Tabellen: M.G.Kendall: The advanced theory of statistics; deel I pag. 442-443.

P.G.Hoel: Introduction to mathematical statistics; pag. 250-253.

ν_1: 1 tot 500; ν_2: 1 tot 1000; onbetrouwbaarheidsdrempels 0,05 en 0,01.

R.A.Fisher en F.Yates: Statistical tables for biological, agricultural and medical research; pag. 34-43.

ν_1: 1 tot 24; ν_2: 1 tot 120; onbetrouwbaarheidsdrempels:

0,20; 0,10; 0,05; 0,01; 0,001.

MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

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

S 102 (M 40)

Toetsen met betrekking tot tweedimensionale normale verdelingen [1].

Inhoud:

1953.

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

1 Tweedimensionale normale verdeling in het algemeen.

De algemene gedaante van de verdelingsdichtheid van een tweedimensionale normale verdeling is:

$$(1)\quad f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\,e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-2\rho\frac{x_1-\mu_1}{\sigma_1}\cdot\frac{x_2-\mu_2}{\sigma_2}+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}}$$

Hierin komen de volgende vijf parameters voor:

$$(2)\quad\begin{cases}\mu_1=\mathcal{E}\,\underline{x}_1 = \text{mathematische verwachting}^{1)} \text{ van } \underline{x}_1,\\ \sigma_1=\sqrt{\mathcal{E}(\underline{x}_1-\mu_1)^2} = \text{spreiding van } \underline{x}_1,\\ \left.\begin{matrix}\mu_2\\ \sigma_2\end{matrix}\right\} \text{ overeenkomstige grootheden voor } \underline{x}_2,\\ \rho=\dfrac{\mathcal{E}(\underline{x}_1-\mu_1)(\underline{x}_2-\mu_2)}{\sigma_1\sigma_2} = \text{correlatiecoëfficiënt van } \underline{x}_1 \text{ en } \underline{x}_2.\end{cases}$$

Het punt (μ_1,μ_2) noemen wij het middelpunt van de verdeling.

De parameter ρ is een maat voor de afhankelijkheid van de variabelen $\underline{x}_1$ en $\underline{x}_2$. Als $\rho=0$ zijn de twee variabelen stochastisch onafhankelijk, als $|\rho|=1$ zijn de variabelen lineair afhankelijk, d.w.z. dat er een rechte lijn in het (x_1,x_2)-vlak is, waar het stochastische punt $(\underline{x}_1,\underline{x}_2)$ met wh 1 op ligt.

Indien wij beschikken over een steekproef van onafhankelijke waarnemingen (x_{1i},x_{2i}) $(i=1,\ldots,n)$ uit de verdeling, kunnen wij de parameters op de volgende wijze schatten:

$$(3)\quad\begin{cases}\mu_1 \text{ door } \bar{x}_1=\frac{1}{n}\sum_{i=1}^{n}x_{1i} \text{ (gemiddelde van de waarnemingen van } \underline{x}_1),\\ \mu_2 \text{ door } \bar{x}_2=\frac{1}{n}\sum_{i=1}^{n}x_{2i} \text{ (gemiddelde van de waarnemingen van } \underline{x}_2),\\ \sigma_1 \text{ door } s_1=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2} \text{ (spreiding van de waarnemingen van } \underline{x}_1),\\ \sigma_2 \text{ door } s_2=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2} \text{ (spreiding van de waarnemingen van } \underline{x}_2),\\ \rho \text{ door } r=\frac{1}{s_1s_2}\cdot\frac{1}{n-1}\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\end{cases}$$

(correlatiecoëfficiënt van de waarnemingen van $\underline{x}_1$ en $\underline{x}_2$).

---

1) De mathematische verwachting van een grootheid wordt ook wel zijn (theoretische) gemiddelde genoemd.

2 Toets van HOTELLING voor het middelpunt van een tweedimensionale normale verdeling.

Deze toets is een generalisatie van de toets van STUDENT voor het gemiddelde van een normale verdeling (zie Memorandum S 47 (M 8)).

Gegeven zij een steekproef (x_{1i}, x_{2i}) $(i = 1, \ldots, n)$ van onafhankelijke waarnemingen uit een tweedimensionale normale verdeling.

De te toetsen hypothese H_0 is, dat een gegeven punt (μ_1, μ_2) het middelpunt van de verdeling is.

De toetsingsgrootheid is de $\underline{T}^2$ van HOTELLING, gedefinieerd als volgt:

$$(4) \quad \underline{T}^2 = \frac{n}{1-\underline{r}^2}\left\{\left(\frac{\bar{\underline{x}}_1-\mu_1}{\underline{s}_1}\right)^2 - 2\underline{r}\,\frac{(\bar{\underline{x}}_1-\mu_1)(\bar{\underline{x}}_2-\mu_2)}{\underline{s}_1\,\underline{s}_2} + \left(\frac{\bar{\underline{x}}_2-\mu_2}{\underline{s}_2}\right)^2\right\}.$$

Als de nulhypothese juist is, zal

$$(5) \quad \underline{F} = \frac{1}{2}\cdot\frac{n-2}{n-1}\,\underline{T}^2,$$

verdeeld zijn als de $\underline{F}$ van FISHER-SNEDECOR met $\nu_1 = 2$ en $\nu_2 = n-2$ vrijheidsgraden (zie Memorandum S 53 (M 24)).

De kritieke zone is van het type $\underline{T}^2 \geqq T_0^2$ en komt dan overeen met een tweezijdige toets van STUDENT.

Litteratuur: A.HALD, Statistical Theory with engineering applications, New York and London, 1952, p. 607.

3 Toets van HOTELLING voor twee steekproeven.

De toets is een generalisatie van de toets van STUDENT voor twee steekproeven (zie Memorandum S 47 (M 9)). Zij wordt gebruikt voor het toetsen van de hypothese H_0 dat twee steekproeven $(x_{1i}, x_{2i})(i = 1,\ldots,n)$ en $(y_{1j}, y_{2j})(j = 1,\ldots,m)$ beschouwd kunnen worden als steekproeven uit dezelfde tweedimensionale normale verdeling.

Als toetsingsgrootheid wordt hierbij gebruikt de $\underline{T}^2$ van HOTELLING voor twee steekproeven:

$$(5)\quad \underline{T}^2 = \frac{mn}{(m+n)(1-\underline{r}^2)}\left\{\left(\frac{\bar{\underline{x}}_1-\bar{\underline{y}}_1}{\underline{S}_1}\right)^2 - 2\underline{r}\,\frac{(\bar{\underline{x}}_1-\bar{\underline{y}}_1)(\bar{\underline{x}}_2-\bar{\underline{y}}_2)}{\underline{S}_1\,\underline{S}_2} + \left(\frac{\bar{\underline{x}}_2-\bar{\underline{y}}_2}{\underline{S}_2}\right)^2\right\},$$

waarin:

$$\bar{\underline{x}}_1 = \frac{1}{n}\sum_{i=1}^{n}\underline{x}_{1i}\quad,\quad \bar{\underline{x}}_2 = \frac{1}{n}\sum_{i=1}^{n}\underline{x}_{2i}$$

$$\bar{\underline{y}}_1 = \frac{1}{m}\sum_{j=1}^{m}\underline{y}_{1j}\quad,\quad \bar{\underline{y}}_2 = \frac{1}{m}\sum_{j=1}^{m}\underline{y}_{2j}$$

$$(6)\qquad \underline{S}_1 = \sqrt{\frac{\sum_{i=1}^{n}(\underline{x}_{1i}-\bar{\underline{x}}_1)^2 + \sum_{j=1}^{m}(\underline{y}_{1j}-\bar{\underline{y}}_1)^2}{m+n-2}}$$

$$\underline{S}_2 = \sqrt{\frac{\sum_{i=1}^{n}(\underline{x}_{2i}-\bar{\underline{x}}_2)^2 + \sum_{j=1}^{m}(\underline{y}_{2j}-\bar{\underline{y}}_2)^2}{m+n-2}}$$

Als de hypothese H_0 juist is bezit:

$$(7)\quad \underline{F} = \frac{1}{2}\cdot\frac{m+n-3}{m+n-2}\cdot\underline{T}^2$$

de verdeling van de $\underline{F}$ van FISHER-SNEDECOR met $\nu_1 = 2$ en $\nu_2 = m+n-3$ vrijheidsgraden.

De kritieke zone is van het type $\underline{T}^2 \geq T_\alpha^2$ en komt dan overeen met een tweezijdige toets van STUDENT.

Litteratuur: A.HALD, Statistical Theory with engineering applications, New York and London, 1952, p. 616.

4 Toets voor de correlatiecoëfficiënt van een tweedimensionale normale verdeling.

De te toetsen hypothese H_0 is, dat de (theoretische) correlatiecoëfficiënt van een tweedimensionale normale verdeling gelijk is aan een gegeven getal ϱ .

Gegeven zij weer een steekproef (x_{1i}, x_{2i}) $(i = 1, \ldots, n)$ van onafhankelijke waarnemingen uit de verdeling.

De toetsingsgrootheid is de correlatiecoëfficiënt $\underline{r}$ van deze waarnemingen, gedefinieerd in (3).

De verdeling van $\underline{r}$ is voor de waarden van ϱ (0,0(0,1)0,9) en =3(1)25,50,100,200,400 getabelleerd door F.N.DAVID (zie litteratuur).

De kritieke zone (tweezijdig) bij een gegeven onbetrouwbaarheid α, bestaat uit de intervallen $\underline{r} \leq r_1$ respectievelijk $\underline{r} \geq r_2$, bepaald door:

$$P[\underline{r} \leq r_1 \mid \varrho] = \tfrac{1}{2}\alpha \qquad \text{resp.} \qquad P[\underline{r} \geq r_2 \mid \varrho] = \tfrac{1}{2}\alpha$$

Bij eenzijdige toetsing is de kritieke zone van het type $\underline{r} \geq r_0$ (rechtszijdig), r_0 bepaald door:

$$P[\underline{r} \geq r_0 \mid \varrho] = \alpha$$

Voor grote waarden van n maakt men gebruik van het feit, dat

$$(8) \qquad \underline{z} = \frac{1}{2} \ln \frac{1+\underline{r}}{1-\underline{r}} = 1{,}1513 \log \frac{1+\underline{r}}{1-\underline{r}}$$

bij benadering normaal verdeeld is met gemiddelde:

$$(9) \qquad \zeta = \frac{1}{2} \ln \frac{1+\varrho}{1-\varrho}$$

en spreiding

$$(10) \qquad \sigma_{\underline{z}} = \frac{1}{\sqrt{n-3}}$$

Litteratuur:

A.HALD, Statistical Theory with engineering applications, New York and London, 1952, p. 608.

F.N.DAVID, Tables of the Ordinates and Probability integral of the distribution of the correlation coefficient in small samples, London, 1938.

5 Toets voor het verschil van de correlatiecoëfficiënten van twee steekproeven uit tweedimensionale normale verdelingen.

Gegeven: Twee steekproeven $(x_{1i}, x_{2i})(i=1,\dots,n)$ en $(y_{1j}, y_{2j})(j=1,\dots,m)$ uit tweedimensionale normale verdelingen.

De te toetsen hypothese is, dat de correlatiecoëfficiënten van de verdelingen, waaruit de steekproeven genomen zijn, gelijk zijn.

De toetsingsgrootheid is:

$$\underline{u} = \frac{\underline{z}_1 - \underline{z}_2}{\sqrt{\frac{1}{n-3} + \frac{1}{m-3}}}, \tag{11}$$

waarin:

$$\left\{\begin{aligned} \underline{z}_1 &= \frac{1}{2} \ln \frac{1+\underline{r}_1}{1-\underline{r}_1} = 1{,}1513 \log \frac{1+\underline{r}_1}{1-\underline{r}_1} \\ \text{en } \underline{z}_2 &= \frac{1}{2} \ln \frac{1+\underline{r}_2}{1-\underline{r}_2} = 1{,}1513 \log \frac{1+\underline{r}_2}{1-\underline{r}_2} \end{aligned}\right. \tag{12}$$

en $\underline{r}_1$ de correlatiecoëfficiënt van de steekproef $(x_{1i}, x_{2i})(i=1,\dots,n)$ en $\underline{r}_2$ de correlatiecoëfficiënt van $(y_{1j}, y_{2j})(j=1,\dots,m)$ is (zie (3)).

Als de nulhypothese juist is, is $\underline{u}$ voor grote n en m bij benadering normaal verdeeld met gemiddelde 0 en spreiding 1.

De kritieke zone is bij tweezijdige toetsing van het type $|\underline{u}| \geq u_0$, bij eenzijdige toetsing van het type $\underline{u} \geq u_1$ of $\underline{u} \leq u_2$.

Litteratuur: A.HALD, Statistical Theory with engineering appl[illegible]ations, New York and London, 1952, p. 608

6 Toets om na te gaan, of een tweedimensionale verdeling normaal is.

Gegeven is een steekproef $(x_{1i}, x_{2i})(i=1,\dots,n)$. Wij wensen de hypothese H_0 te toetsen, dat dit een steekproef is uit een tweedimensionale normale verdeling. Hiertoe berekenen we de grootheden $\bar{x}_1$, $\bar{x}_2$, s_1, s_2 en r volgens (3).

Beschouw nu bij een λ_ε gegeven door

$$(13) \qquad 1 - e^{-\frac{1}{2}\cdot\frac{\lambda_\varepsilon^2}{1-r^2}} = \varepsilon$$

de ellips:

$$(14) \left(\frac{x_1-\bar{x}_1}{s_1}\right)^2 - 2r\frac{(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{s_1 s_2} + \left(\frac{x_2-\bar{x}_2}{s_2}\right)^2 = \lambda_\varepsilon^2$$

De lange as gaat door het punt $(\bar{x}_1, \bar{x}_2)$ en heeft een helling gegeven door:

$$(15) \qquad h = \frac{2r}{\frac{s_1}{s_2} - \frac{s_2}{s_1} + \sqrt{\left(\frac{s_1}{s_2} - \frac{s_2}{s_1}\right)^2 + 4r^2}}$$

De normale verdeling met $\mu_1 = m_1$, $\mu_2 = m_2$, $\sigma_1 = s_1$, $\sigma_2 = s_2$ en $\rho = r$ noemt men de aangepaste normale verdeling. Voor deze verdeling geldt, dat de kans op een waarnemingen binnen de door (13) en (14) gegeven ellips gelijk aan ε is. Verder is deze verdeling symmetrisch t.o.v. de assen van die ellips.

Wij construeren nu de ellips gegeven door (13) en (14) voor een aantal waarden van ε b.v.: $\varepsilon = \frac{1}{6}, \frac{2}{6}, \dots, \frac{5}{6}$. De ellipsen met hun gemeenschappelijke assen verdelen dan het platte vlak in 24 vakken. De verwachtingen volgens de aangepaste normale verdeling van de aantallen punten binnen deze vakken zijn dan $\frac{1}{24}n$. Wij tellen nu de aantallen waargenomen punten die in ieder der vakken liggen, aan te duiden met $n_1, n_2, \dots, n_{24}$ en berekenen:

$$\chi^2 = \sum_{j=1}^{24} \frac{\left(n_j - \frac{1}{24}n\right)^2}{\frac{1}{24}n}$$

Als H_0 juist is, zal deze grootheid voor grote n bij benadering een χ^2-verdeling hebben met 24-6=18 vrijheidsgraden (er zijn 5 parameters aangepast ; men vergelijke hiervoor het memorandum S 47 (M 17) over de χ^2-toets voor aanpassing). Wij zullen de hypothese H_0 verwerpen als χ^2 groter is dan χ_0^2 gedefinieerd door

$$P[\underline{\chi}^2 \geq \chi_0^2] = \alpha$$

als α de gewenste onbetrouwbaarheidsdrempel is.

Litteratuur: A.HALD, Statistical Theory with engineering applications, New York and London, 1952, p. 602.