

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 197

Statistische analyse van de metingen verricht
bij walvissen gevangen door een walvisvaarder

II. De Speklaagdikte

door

R. Doornbos

en

J.F. van Haastrecht

Mei 1956

1. Inleiding

Voor een beschrijving van het waarnemingsmateriaal verwijzen wij naar de inleiding van het vorige rapport over het walvis-onderzoek van de hand van A. REITSMA (rapport S 191 van de Statistische Afdeling van het Mathematisch Centrum), waarin de betrouwbaarheid van de genomen steekproeven is behandeld.

In het onderhavige rapport wordt getracht een antwoord te geven op de vraag of de dikte van de speklagen van de onderzochte walvissoort systematische verschillen vertoont voor verschillende jaren bij dieren die gelijkwaardig zijn wat de andere factoren (geslacht, lengte, vangstdatum en positie van het vangstschip) betreft. Hierbij zijn alleen de eerste zes vangstjaren beschouwd, daar de berekeningen reeds in een vergevorderd stadium verkeerden op het moment dat de gegevens van het jaar 1954-1955 in ons bezit kwamen.

Uit de conclusies van rapport S 191 blijkt dat in een aantal van de beschouwde jaren de steekproeven bestaande uit de dieren waarvan de speklaagdikte is gemeten niet representatief zijn voor de totale populatie wat betreft de lengte. Zoals in dat rapport werd opgemerkt, is niet na te gaan of dit ook geldt voor de speklaagdikte zelf. Indien dit wel het geval is, zou het uiteraard een grote invloed kunnen hebben op onze conclusies, die wij dan ook uit dien hoofde onder voorbehoud zullen formuleren.

Nadat het onderzoek over de speklaagdikte was afgesloten, kregen wij de waarnemingen tot onze beschikking die in het vangstseizoen 1954-1955 zijn verricht aan boord van een tiental Noorse walvisvaarders. Uit deze waarnemingen bleek, dat tussen de metingen verricht aan boord van de verschillende schepen systematische verschillen optreden, die aanmerkelijk groter zijn dan de verschillen die wij voor de Nederlandse walvisvaarder vinden tussen de verschillende jaren. Deze grote verschillen wijzen op een verschillende waarnemingstechniek bij de waarnemers. Deze meettechniek is dus blijkbaar niet nauwkeurig vastgelegd en het is zeer goed denkbaar dat de verschillen die wij voor de verschillende jaren vinden te wijten zijn aan een geleidelijk veranderende meettechniek. Wij komen op deze punten in onze conclusies nader terug.

2. De gevolgde analyse-methode

Wij zullen de volgende notatie gebruiken:

σ = mannetje,

ϕ = niet-drachtig wijfje,

ϕ^+ = drachtig wijfje,

L = lengte in voeten,

d = speklaagdikte dorsaal in cm,

v = speklaagdikte ventraal in cm,

$D = d + v$.

Als maat voor de speklaagdikte hebben wij $D = d + v$ genomen. Reeds een oppervlakkige beschouwing van het materiaal leert dat de speklaagdikte D samenhangt met lengte, geslacht en vangstdatum. Het lag oorspronkelijk in de bedoeling de dieren te splitsen in groepen van hetzelfde geslacht en, bij benadering, van dezelfde lengte en dezelfde vangstdatum. Het bleek echter niet mogelijk de lengteklassen zo te kiezen, dat enerzijds de groepen niet te klein werden, terwijl aan de andere kant binnen de lengteklassen geen aanwijsbaar verband meer bestond tussen L en D . Daarom is dit verband, de zogenaamde regressie van D op L in de analyse-methode opgenomen. De gevolgde methode, de covariantie-analyse, wordt uiteengezet in het bijgevoegde memorandum S 168 (M 58). De bij deze methode beschouwde regressielijnen geven hier het verband aan tussen de lengte L (de x van het memorandum) en de bijbehorende gemiddelde waarde van D (overeenkomend met het gemiddelde van de stochastische grootheid y uit het memorandum). In hoeverre de voor de geldigheid van deze methode nodige onderstellingen vervuld zijn, is niet nagegaan. Uit theoretische onderzoeken is evenwel bekend, dat de uitkomsten van deze toetsen tamelijk ongevoelig zijn voor geringe afwijkingen van de onderstelling van normaliteit.

Op een groep walvissen van één geslacht en gevangen binnen een zekere periode worden nu achtereenvolgens de volgende statistische toetsen toegepast (men vergelijk memorandum S 47 (M 6), dat aan dit rapport is toegevoegd):

1e. Van de 6 regressielijnen (voor ieder jaar één) wordt de hypothese getoetst dat zij recht zijn. De hierbij gevolgde methode wordt o.a. beschreven in W.J. DIXON en F.J. MASSEY (1951), section 16.5.

2e. Wordt de rechtlijnigheid niet verworpen, dan passen wij vervolgens toets A van memorandum S 168 (M 58) toe om te toetsen of de 6 lijnen voor de verschillende jaren als evenwijdig beschouwd kunnen worden.

3e. Wordt ook de evenwijdigheid niet verworpen, dan passen wij toets B van het memorandum toe om na te gaan of de evenwijdige regressielijnen samenvallen.

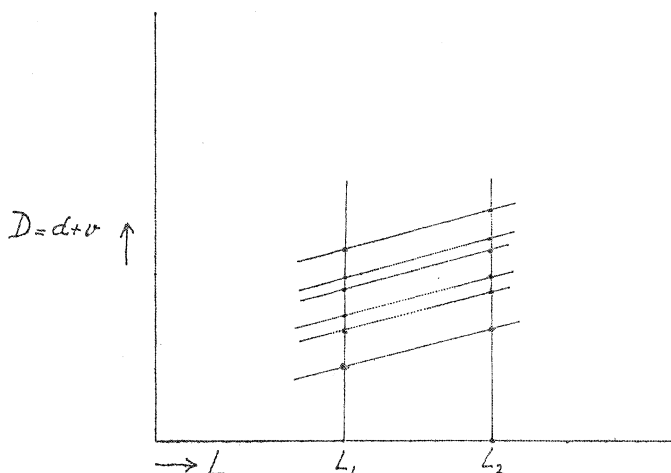


fig. 2.1

Evenwijdige en rechtlijnige regressielijnen

Het geval waarin de toetsen genoemd onder 1e en 2e niet tot verwerping van de getoetste hypothesen leiden, is schematisch in figuur 2.1 aangegeven. Moeten wij uit de laatste toets (onder 3e) concluderen, dat de lijnen niet samenvallen, dan is het dus mogelijk om te schatten welke de volgorde van de jaren is, gerekend naar de dikte van de speklaag. Deze volgorde hangt dan namelijk niet van de lengte af. Dit is in figuur 2.1 aangegeven voor de lengten L_1 en L_2 .

Ter illustratie zijn in figuur 2.2 de waarnemingen getekend van de Q^+Q^+ in de periode tot en met 15 Januari, de jaren '50-'51 en '53-'54. De regressielijnen zijn nagenoeg evenwijdig en bij iedere lengte komen de grootste speklaagdikten voor in het laatste jaar.

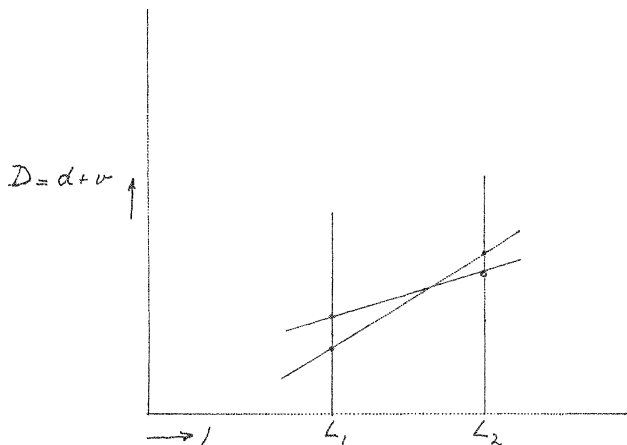
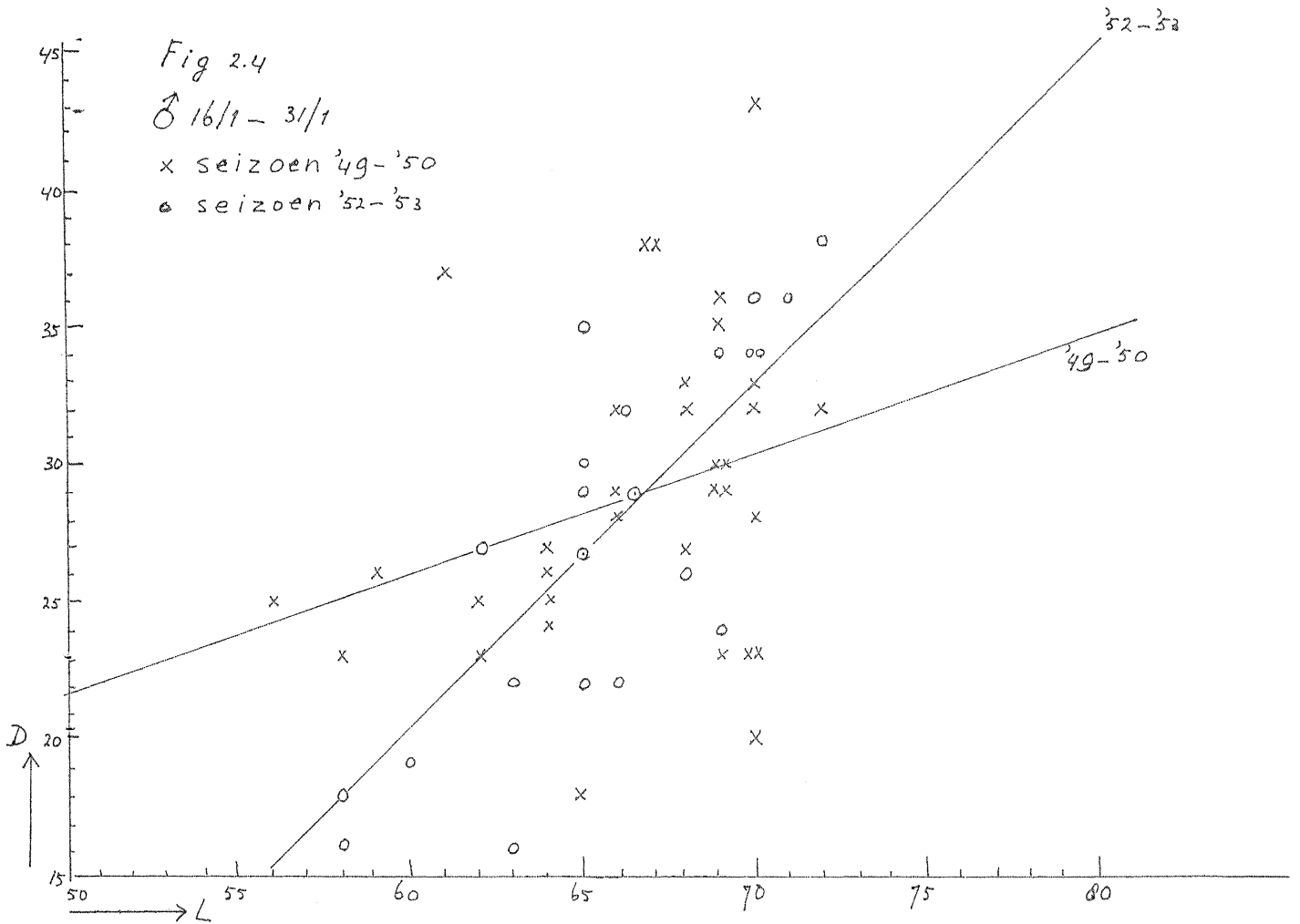
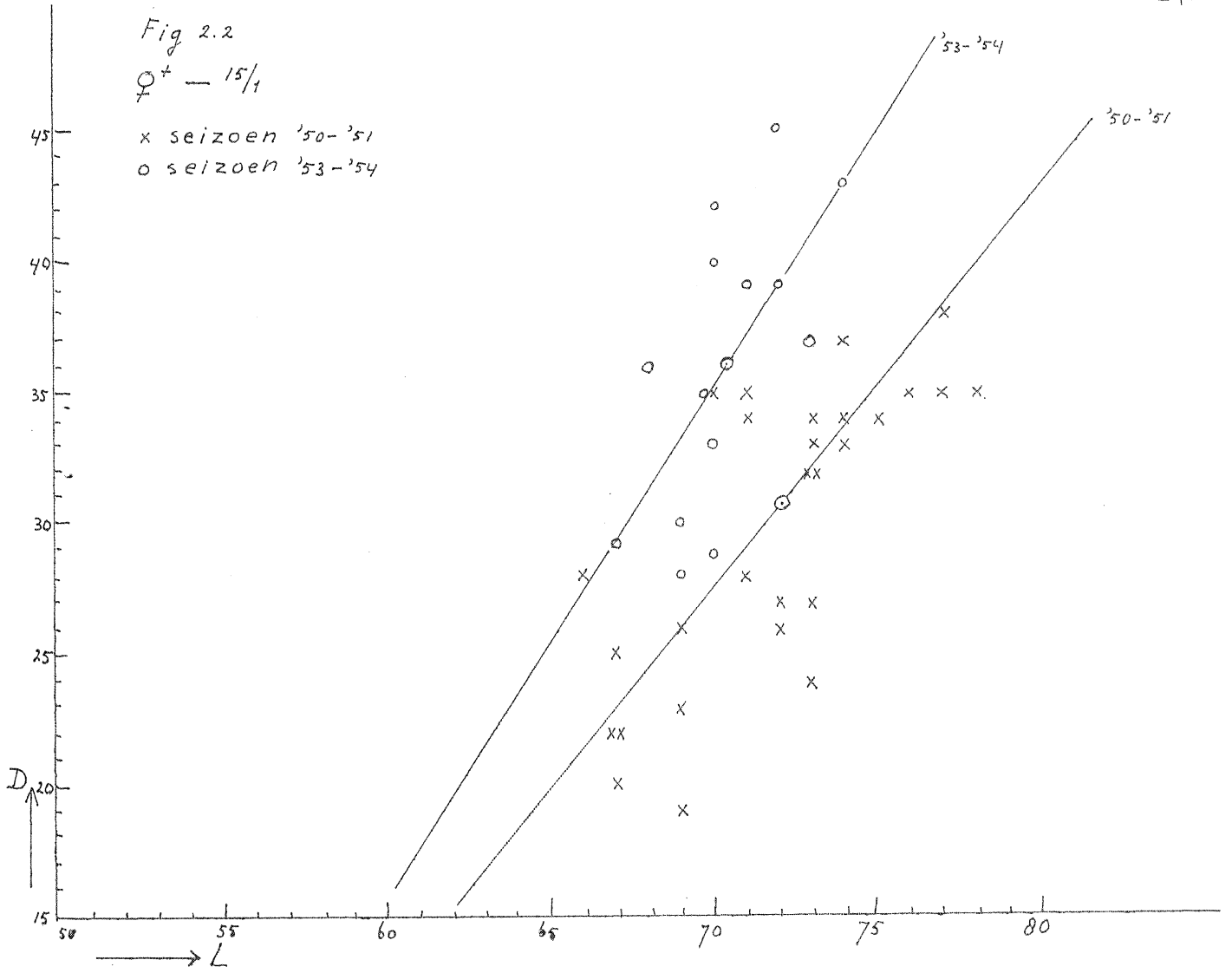


fig. 2.3

Nietevenwijdige en rechtlijnige regressielijnen



Leidt de toets onder 2e tot verwerping van de evenwijdigheid, dan is het soms niet mogelijk te zeggen in welk van een tweetal jaren de walvissen de dikste speklaag hebben, daar dit voor twee verschillende lengtes verschillend kan zijn. In figuur 2.2 is een dergelijk geval geschetst voor 2 lijnen. De volgorde is links van het snijpunt tegengesteld aan die rechts van het snijpunt. In deze gevallen hebben wij dan ook geen conclusie getrokken. Figuur 2.4 geeft een voorbeeld van deze situatie. Bij de ♂♂ in de periode van 16/1 tot en met 31/1 treffen wij voor kleine lengten de grootste speklaagdikten aan in het jaar '49-'50, terwijl voor grote lengten het jaar '52-'53 juist dikkere speklagen oplevert.

Wordt de hypothese dat de lijnen recht zijn verworpen, dan passen wij ook geen verdere analyse toe. In principe is het nl. wel mogelijk om voor kromme maar wel evenwijdige regressielijnen toch tot een conclusie te komen, maar dit vereist veel meer rekenwerk dan bij rechte lijnen. Bovendien werd in de beide gevallen waarin de rechtlijnigheid moest worden verworpen dit resultaat veroorzaakt door één of twee van de regressielijnen, m.a.w. sommige lijnen waren duidelijk gebogen, terwijl dit bij de andere niet het geval was, zodat toch geen conclusie omtrent een volgorde van de jaren getrokken zou kunnen worden.

3. De resultaten van de volledige analyse

In tabel 3.1 zijn de resultaten van de toetsen weergegeven in de vorm van overschrijdingskansen. Kleine overschrijdingskansen leiden daarbij tot verwerping van de getoetste hypothese (vgl. memorandum S 47 (M 6)).

Tabel 3.1

Overschrijdingskansen van de toetsen voor rechtheid, evenwijdigheid en samenvallen van de regressielijnen.¹⁾

		- t/m 15/1	16/1 t/m 31/1	1/2 t/m 15/2	16/2 t/m -
♂	rechtl.	$\approx 0,05$	$> 0,50$	$0,05 < k < 0,10$	$> 0,50$
	evenw.	$0,10 < k < 0,20$	$0,0001 < k < 0,01$	$\approx 0,05$	$\approx 0,10$
	samenv.	$\ll 0,0001$		$\ll 0,0001$	$\ll 0,0001$
♀	rechtl.	$\approx 0,50$	$0,10 < k < 0,20$	$< 0,001$	$0,10 < k < 0,20$
	evenw.	$\approx 0,50$	$0,05 < k < 0,10$		$0,001 < k < 0,01$
	samenv.	$\ll 0,0001$	$\ll 0,0001$		
♀	rechtl.	$> 0,50$	$0,20 < k < 0,50$	$\approx 0,50$	$0,001 < k < 0,01$
	evenw.	$> 0,50$	$> 0,50$	$\approx 0,20$	
	samenv.	$\ll 0,0001$	$\ll 0,0001$	$\ll 0,0001$	

1) De tekens in tabel 3.1 hebben de volgende betekenis: $\approx$: ongeveer gelijk aan, $>$ groter dan, $<$ kleiner dan, $\ll$ veel kleiner dan.

Wij zien dat in twee gevallen de hypothese van de rechtheid van de regressielijnen verworpen moet worden, terwijl bovendien nog in twee gevallen waarin de rechtheid niet verworpen behoeft te worden de lijnen ten duidelijkste niet evenwijdig lopen. In al de andere gevallen, waarin de toets voor samenvallen van de regressierechten dus kon worden toegepast, blijkt zeer duidelijk dat de lijnen verschillend niveau hebben. Wij zien dat bij de drachtige wijfjes de overschrijdingskansen van de toetsen voor rechtheid en evenwijdigheid, afgezien van het laatste deel van het seizoen, groter zijn dan ergens bij de mannetjes en de niet-drachtige wijfjes het geval is. Dit zou b.v. veroorzaakt kunnen worden door zeer jonge dieren bij de $\sigma \sigma$ en de $\phi \phi$, die afwijking van de regressielijnen veroorzaken.

Om een indruk te geven van de verschillen in de dikte van de speklaag tussen de verschillende jaren, geven wij in tabel 3.2 de speklaagdikten D bij een lengte van 70 voet voor die groepen waar de rechtheid en de evenwijdigheid van de regressielijnen niet behoefde te worden verworpen. Hierbij is dus per groep voor ieder jaar een regressielijn geschat, onder de aanname dat ze evenwijdig zijn. Op deze lijnen zijn dan bij een lengte van 70 voet de bijbehorende D -waarden afgelezen. Tevens geven wij in deze tabel de helling van de lijnen aan. Bij deze hellingen zijn eveneens zgn. betrouwbaarheidsgrenzen opgegeven (zie memorandum S 47 (M 18)). De onbetrouwbaarheid is 0,05. Voor de D_{70} waarden hebben wij het opgeven van betrouwbaarheidsgrenzen achterwege gelaten, daar deze voor de 6 jaren, binnen een groep van zelfde vangsttijd en geslacht, sterk van elkaar afhankelijk zijn vanwege de aanname van evenwijdige regressielijnen. De werkelijke volgorde van de jaren wat betreft de speklaagdikte behoeft niet steeds exact overeen te stemmen door die volgorde die uit tabel 3.2 is af te leiden. Deze laatste is slechts als een schatting te beschouwen.

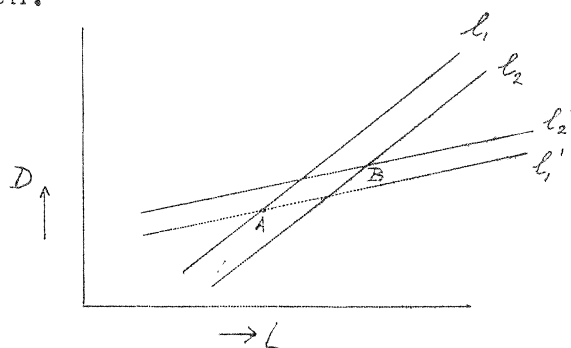


Fig. 3.1

Invloed verandering helling op de onderlinge ligging van twee regressielijnen.

Wij zien dit schematisch aangeduid in figuur 3.1. Als twee regressielijnen dicht bij elkaar liggen, terwijl de gemiddelde $\angle$ waarden vrij veel verschillen, dan kan een kleine verandering in hellingshoek, opgelegd door de voorwaarde van evenwijdigheid aan andere - niet in de figuur geschetste - lijnen, de volgorde doen omkeren, daar de lijnen draaien om de zwaartepunten van de desbetreffende puntenwolken. Deze zijn in de figuur met A resp. B aangegeven.

Tabel 3.2

Geschatte speklaagdikten bij een lengte van 70 voet. Aantallen waarnemingen kleiner dan 10 zijn tussen haakjes vermeld.

		-t/m15/1	16/1t/m31/1	1/2t/m15/2	16/2t/m-
♂	helling	0,66±0,12	.	1,04±0,26	1,13±0,26
	1948-1949	22,9		30,8(3)	28,3(6)
	1949-1950	28,9	niet even- wijdig	37,2	34,7
	1950-1951	26,7		28,8	38,3
	1951-1952	29,2		36,4	34,1
	1952-1953	28,2		34,6(8)	36,2
	1953-1954	30,3		34,4	43,2
♀	helling	0,70±0,16	1,07±0,22	niet recht	niet even- wijdig
	1948-1949	19,5	28,8(4)		
	1949-1950	31,2	34,1		
	1950-1951	28,2	25,0		
	1951-1952	28,9	33,3(3)		
	1952-1953	30,8	34,1		
	1953-1954	32,9	31,1(5)		
♀ ⁺	helling	1,50±0,32	0,73±0,36	0,62±0,63	niet recht
	1948-1949	19,9(6)	27,5(5)	34,2(5)	
	1949-1950	31,7	36,1	41,7	
	1950-1951	27,5	26,0	31,2	
	1951-1952	27,9	29,4(4)	43,7(4)	
	1952-1953	34,3	36,0(2)	45,7(6)	
	1953-1954	35,5	34,8(2)	37,6(8)	

Wij zien dat over het algemeen de $\varphi^* \varphi^*$ van 70 voet een dikkere speklaag hebben dan de dieren van dezelfde lengte behorende tot de andere groepen. Bovendien valt op dat de helling van de regressielijn bij de $\sigma \sigma$ en de $\varphi \varphi$ in de loop van het seizoen steiler wordt, terwijl bij de $\varphi^* \varphi^*$ juist het omgekeerde het geval is.

Zowel de hellingen als de D_{70} waarden die in de tabel zijn opgegeven zijn schattingen van de werkelijke waarden. Vooral wanneer deze schattingen op weinig waarnemingen berusten is de nauwkeurigheid gering. Aantallen kleiner dan 10 hebben wij tussen haakjes aangegeven achter de erop gebaseerde schattingen.

Om na te kunnen gaan of de volgorden in speklaagdikte van de jaren, zoals die binnen de groepen gevonden zijn, overeenstemmen zijn in tabel 3.3 voor elke groep de rangnummers naar opklimmende grootte van de D_{70} waarden aangegeven.

Tabel 3.3
Rangnummers van de D_{70} waarden

		1948- 1949	1949- 1950	1950- 1951	1951- 1952	1952- 1953	1953- 1954
- t/m 15/1		1	4	2	5	3	6
		1	5	2	3	4	6
		1	4	2	3	5	6
16/1 t/m 31/1		niet evenwijdig					
		2	6	1	4	5	3
		2	6	1	3	5	4
1/2 t/m 15/2		2	6	1	5	4	3
		niet recht					
		2	4	1	5	6	3
16/2 t/m -		1	3	5	2	4	6
		niet evenwijdig					
		niet recht					
Som rangnummers		12	38	15	30	36	37

Met behulp van de methode van m -rangschikkingen (memorandum S 47 (M 14)) is onderzocht of er tussen de groepen overeenstemming bestaat wat betreft de volgorde van de jaren. De overschrijdingskans onder de hypothese dat er geen overeenstemming bestaat is kleiner dan 10^{-4} . Wij verwerpen dus deze hypothese en zien aan de kolomtotalen van tabel 3.3 dat het vooral de seizoenen 1948-1949 en 1950-1951 zijn, welke een afwijkende, en wel een lager gelegen, regressielijn bezitten en dus systematisch minder gerder zijn geweest dan andere jaren.

4. Geselecteerde vangstgebieden

Daar wij tot dusver niet hebben gekeken in welke gebieden de walvissen zijn gevangen weten wij nog niet of de geconstateerde verschillen werkelijk het gevolg zijn van het feit dat de speklagen het ene jaar dikker zijn dan in het andere in alle vangstgebieden. Het is namelijk ook mogelijk dat de dieren in het ene gebied steeds een dikkere speklaag hebben dan in een ander gebied, waar de levensomstandigheden wellicht ongunstiger zijn. Vangt men in het ene jaar in het eerste gebied en in het andere jaar in het tweede gebied, dan vindt men ook verschillen als er in werkelijkheid tussen de jaren geen verschil is.

In tabel 4.1 vinden wij een overzicht van de vangstgebieden ingedeeld naar de lengte gedurende de zes beschouwde jaren. In ieder vakje staat het aantal dagen dat er gevist is.

Uit de tabel blijkt dat er geen gebied is waar op ongeveer dezelfde datum in alle zes de beschouwde jaren is gevangen. Wij hebben echter twee gebieden gekozen, nl. 20° W.L. - 10° O.L. en 60° O.L. - 100° O.L. waar resp. vooral in de eerste drie en de laatste drie jaren is gevangen. Wij kunnen dus binnen deze gebieden wel de eerste drie jaren onderling vergelijken en ook het laatste drietal onderling, maar een vergelijking tussen beide drietallen jaren is niet goed te verkrijgen.

De resultaten van deze analyses geven wij in de tabellen 4.2 en 4.3.

Hieruit blijkt dat de grootste verschillen optreden in het gebied 30° W.L. - 10° O.L. In het gebied 60° O.L. - 100° O.L. is er slechts één geval, nl. φ tot 15/1, waarin een aanwijzing voor een systematisch verschil aanwezig is. Kijken wij daar naar de cijfers voor de speklaagdikte, dan blijkt dat het seizoen '53-'54 vetter is dan '51-'52 en '52-'53.

Tabel 4.2

Overschrijdingskansen van de toetsen toegepast op de waarnemingen van de dieren gevangen tussen 30° W.L. en 10° O.L.

		- 15/1	16/1-31/1	1/2-15/2	16/2 -
♂	rechtl.	≈ 0,50	0,05<k<0,10	≈ 0,05	0,10<k<0,20
	evenw.	> 0,50	0,10<k<0,20	≈ 0,25	≈ 0,25
	samenv.	0,01<k<0,05	0,05<k<0,10	0,01<k<0,05	0,001<k<0,01
♀	rechtl.	≈ 0,50	> 0,50	> 0,50	0,05<k<0,10
	evenw.	0,01<k<0,05	≈ 0,25	- 1)	0,05<k<0,10
	samenv.		0,05<k<0,10	0,05<k<0,10	0,01<k<0,05
♀ ⁺	rechtl.	≈ 0,50	> 0,50	≈ 0,25	0,01<k<0,05
	evenw.	> 0,50	> 0,50	0,05<k<0,10	
	samenv.	≈ 0,0001	0,001<k<0,01	0,01<k<0,05	

Tabel 4.3

Overschrijdingskansen van de toetsen toegepast op de waarnemingen van de dieren gevangen in het gebied 60° O.L. - 100° O.L.

		- 15/1	16/1-31/1	1/2-15/2	16/2 -
♂	rechtl.	≈ 0,001	> 0,50	≈ 0,20	> 0,50
	evenw.		0,05<k<0,10	> 0,50	> 0,50
	samenv.		0,20<k<0,50	> 0,50	≈ 0,10
♀	rechtl.	0,20<k<0,50	0,20<k<0,50	0,20<k<0,50	> 0,50
	evenw.	> 0,50	0,10<k<0,20	0,20<k<0,50	0,001<k<0,01
	samenv.	0,05<k<0,10	0,20<k<0,50	≈ 0,10	
♀ ⁺	rechtl.	> 0,50	0,50<k<0,10	0,10<k<0,20	≈ 0,05
	evenw.	> 0,50	> 0,50	0,001<k<0,01	> 0,50
	samenv.	0,001<k<0,01	0,20<k<0,50		> 0,50

Bij de niet drachtige vrouwtjes in dezelfde periode liggen de verhoudingen evenzo. Hoewel daar de overschrijdingskansen voor samenvallen groter dan 0,05 is, levert dit een bevestiging op voor de conclusie dat het laatste jaar in het begin van het vangstseizoen gekenmerkt is door een grotere speklaagdikte bij de wijfjes dan de beide voorgaande jaren.

1) In deze groep komen 7 dieren voor van het jaar '48-'49 en 1 dier van '49-'50. De evenwijdigheid van de regressielijnen valt dus niet te toetsen.

Om een conclusie te kunnen trekken in het gebied 30° W.L. tot 10° O.L. zijn in tabel 4.4 de speklaagdikten bij 70 voet opgegeven berekend in die gevallen waarin evenwijdigheid en rechtheid van de regressielijnen niet kon worden verworpen, maar het samenvallen wel. Evenals in tabel 3.2 geven wij tussen haakjes aantallen kleiner dan 10 aan. Tevens zijn weer de geschatte hellingen opgegeven met een betrouwbaarheidsinterval met onbetrouwbaarheid 0,05.

Tabel 4.4
Geschatte speklaagdikten bij een lengte van 70 voet. Gebied 30° W.L. - 10° O.L.

		-15/1	15/1-31/1	1/2-15/2	16/2-
♂	helling	0,67±0,17	geen duidelijk verschil	1,17± 0,46	0,49±0,57
	1948-1949	24,7(7)		22,0(9)	-
	1949-1950	29,0		32,1	32,6
	1950-1951	28,3		-	37,7
♀	helling	niet even- wijdig	geen duidelijk verschil	geen duidelijk verschil	0,67±0,52
	1948-1949				-
	1949-1950				34,1
	1950-1951				40,3(7)
♀+	helling	1,71±0,53	0,82±0,64	0,49±0,85	niet recht
	1948-1949	19,0(5)	27,2(5)	34,7(5)	
	1949-1950	31,1	35,9	42,1	
	1950-1951	26,6	-	-	

Wij zien dat overal waar het eerste met het tweede jaar vergeleken is het eerste jaar de dunste speklaag laat zien. Dit komt dus goed overeen met de conclusie van paragraaf 3. Waar echter in het laatst van het seizoen het tweede en het derde vangstjaar kunnen worden vergeleken is het derde jaar vetter dan het tweede, terwijl wij in paragraaf 3 vonden dat het derde jaar tot de magere jaren moest worden gerekend. Uit tabel 4.1 blijkt dan ook dat het derde jaar meer westelijk is gevangen dan gedurende enig ander jaar, zodat een vergelijking met andere jaren niet goed mogelijk is.

5. Conclusie

De gemeten speklaagdikten zijn in het eerste jaar (1948-1949) systematisch kleiner dan in het tweede jaar (1949-1950) bij dieren van dezelfde lengte, hetzelfde geslacht, gevangen in hetzelfde gebied en op ongeveer dezelfde datum. In het laatste gedeelte van het seizoen (na 16 Februari) zijn, bij overigens gelijkwaardige walvissen, de gemeten speklaagdikten het derde jaar groter dan het tweede jaar.

Tenslotte is er nog een aanwijzing dat de gemeten speklaag bij de vrouwelijke dieren (wel en niet drachtige) in het begin van het seizoen (tot 15 Januari) het laatste jaar systematisch dikker is dan de beide voorafgaande jaren.

Bovengenoemde conclusies betreffen de metingen. Door een statistisch onderzoek kan niet worden uitgemaakt of de geconstateerde verschillen moeten worden toegeschreven aan een niet voor alle jaren en onder alle omstandigheden gelijke meettechniek. En ook als wij aannemen dat de geconstateerde verschillen reëel zijn, blijft nog de vraag of de steekproef representatief is geweest voor de populatie, gezien de ervaring opgedaan met de lengte.

Het verdient aanbeveling te trachten de wijze waarop de metingen dienen te worden verricht beter vast te leggen. Indien men dan bovendien beschikt over de gegevens van meerdere schepen gedurende een aantal jaren, dan zullen verschillen in speklaagdikte tussen de jaren door al deze schepen worden geconstateerd, zodat de conclusies met groter zekerheid getrokken kunnen worden. Met gegevens van meerdere schepen kan ook veel beter worden onderzocht of er verschillen zijn tussen de vangstgebieden, wat met gegevens van één schip minder goed mogelijk is, omdat dat schip niet ter zelfder tijd in twee gebieden kan vangen, zodat de invloed van het seizoen en eventueel het jaar en tevens veranderingen in meettechniek tot gevolg kunnen hebben, dat verschillen tussen de gebieden niet meer gevonden kunnen worden.

Literatuur

W.J. DIXON and F.J. MASSEY, Introduction to statistical analysis, New York, London, 1951.

Algemene gang van zaken bij het toetsen van een ¹⁾
hypothese.

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

Bij een toets behoort een toetsingsgrootheid u (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \dots, x_n$, een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van u , en wel op zodanige wijze, dat de kans, dat u een in Z gelegen waarde aanneemt, onder de hypothese H_0 , gelijk is aan een gegeven getal α , zodat Z dus van α afhankelijk is ³⁾. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \dots, x_n$, indien de bij deze waarnemingen behorende waarde van u in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α . Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0,05$ resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

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

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans ~~1/2~~ is.

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

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

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

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

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

Litteratuur:

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

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

Methode der m rangschikkingen ¹⁾

Een duidelijke 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 beoordeeling 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 beoordeelingen. De hypothese H_0 , die met deze methode getoetst kan worden, houdt in dat er geen overeenstemming tussen de waarnemers bestaat; precieser gezegd, dat alle rangschikkingen onafhankelijk van elkaar op toevallige wijze zijn ontstaan. Dit is b.v. het geval, als het betrokken kenmerk in werkelijkheid voor alle elementen dezelfde waarde bezit.

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

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

De som van alle rangnummers is $\frac{1}{2}nm(n+1)$. Onder de hypothese H_0 is het theoretische gemiddelde van iedere kolom: $\frac{1}{2}m(n+1)$

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

1	-3	-4	5	-2	3	
---	----	----	---	----	---	--

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

De som der kwadraten van deze afwijkingen noemen wij S .

In ons voorbeeld is $S = 64$.

Als alle m rangschikkingen gelijk zijn wordt het maximum van S bereikt.

Dit maximum is $\frac{1}{12} m^2(n^3-n)$.

We definiëren nu als coëfficiënt van overeenstemming

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

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

W varieert dus tussen 0 en 1.

De verdeling van $\underline{S}$ 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:

$\chi_r^2 = m(n-1)\underline{W} = \frac{12 S}{mn(n+1)}$ heeft voor $m \rightarrow \infty$ 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 $\underline{F} = e^{2z}$

($\underline{F}$ is de $\underline{F}$ van Snedecor, $\underline{z}$ de $\underline{z}$ van Fisher) met

$$v_1 = n-1-\frac{2}{m}$$

$$v_2 = (m-1) v_1 \quad \text{vrijheidsgraden ([1] pg. 84 [2] pg. 33-36)}$$

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 $\underline{S}$ 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 "Parameter vrije Methoden", Hoofdstuk II, Rapport S 59, Mathematisch Centrum (1951), Blz. 18-45.

S47(M18)

Betrouwbaarheidsintervallen (algemeen).¹⁾

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

Een betrouwbaarheidsinterval $\mathcal{I}$ voor θ is een interval, waarvan de grenzen afhankelijk zijn van de waarnemingen $x_1, \dots, x_n$ van 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 $\mathcal{I}$ zo zal uitvallen, dat het θ niet bevat. Hierbij is dus θ constant en het interval $\mathcal{I}$ 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 $\mathcal{I}$ de verzameling van die waarden θ_0 die bij toepassing van $\mathcal{T}$ op grond van de gevonden waarnemingen $x_1, \dots, x_n$ 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.

MATHEMATISCH CENTRUM,
 2e Boerhaavestraat 49,
A m s t e r d a m - O.
 Statistische Afdeling
 S 168 (M 58)

Enige in de covariantieanalyse gebruikte toetsen. 1) 2)

Wij gaan uit van k groepen stochastische grootheden

$$\begin{aligned} & y_{11}, \dots, y_{1n_1}, \\ & y_{21}, \dots, y_{2n_2}, \\ & \vdots \\ & y_{k1}, \dots, y_{kn_k}. \end{aligned}$$

De verwachtingen van deze grootheden voldoen aan de volgende relaties:

$$(1) \quad \begin{cases} E y_{ij} = \alpha_i + \beta_j x_{ij} & (j=1, \dots, n_i), \\ E y_{kj} = \alpha_k + \beta_j x_{kj} & (j=1, \dots, n_k). \end{cases}$$

Hierin stellen α_i en β_j ($i=1, \dots, k$) onbekende parameters voor, terwijl de waarden x_{ij} ($i=1, \dots, k$; $j=1, \dots, n_i$) gegeven zijn. Verder wordt ondersteld, dat alle y 's onderling onafhankelijk en normaal verdeeld zijn met dezelfde spreiding σ . Hierin is dus begrepen het geval waarin x_{ij} een waarneming is van de stochastische grootheid x_{ij} , waarbij y_{ij} en x_{ij} een tweedimensionale *normale* verdeling hebben. Onder de voorwaarde dat x_{ij} de waarde x_{ij} heeft aangenomen, is y_{ij} dan immers normaal verdeeld met als gemiddelde een lineaire functie van x_{ij} . Wij stellen nu toetsen op voor de volgende hypothesen:

$$(A) \quad \beta_1 = \dots = \beta_k$$

met als toegelaten hypothesen willekeurige α 's en β 's.

$$(B) \quad \alpha_1 = \dots = \alpha_k \text{ en } \beta_1 = \dots = \beta_k$$

met als toegelaten hypothesen willekeurige α 's maar gelijke β 's en

$$(C) \quad \alpha_1 = \dots = \alpha_k \text{ en } \beta_1 = \dots = \beta_k$$

met als toegelaten hypothesen willekeurige α 's en β 's.

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

2) De eerste van de in dit memorandum besproken toetsen wordt ook in memorandum S 73 (M 37) behandeld met gebruik van een enigszins andere notatie.

Met (A) toetsen wij dus de evenwijdigheid van de k regressielijnen uitgaande van het model (1). Met (B) toetsen wij of de k lijnen samenvallen, aangenomen dat ze evenwijdig zijn en met (C) toetsen wij direct of wij slechts met één regressielijn te maken hebben, uitgaande van (1).

Wij voeren de volgende afkortingen in:

$$C_{xx_i} = \sum_{j=1}^{n_i} \left(x_{ij} - \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \right)^2 = \sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{j=1}^{n_i} x_{ij} \right)^2}{n_i},$$

$$C_{yy_i} = \sum_{j=1}^{n_i} \left(y_{ij} - \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \right)^2 = \sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{j=1}^{n_i} y_{ij} \right)^2}{n_i},$$

$$\begin{aligned} C_{xy_i} &= \sum_{j=1}^{n_i} \left(x_{ij} - \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \right) \left(y_{ij} - \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \right) = \\ &= \sum_{j=1}^{n_i} x_{ij} y_{ij} - \frac{\sum_{j=1}^{n_i} x_{ij} \sum_{j=1}^{n_i} y_{ij}}{n_i}, \end{aligned}$$

$$C_{xx_W} = \sum_{i=1}^k C_{xx_i},$$

$$C_{yy_W} = \sum_{i=1}^k C_{yy_i},$$

$$C_{xy_W} = \sum_{i=1}^k C_{xy_i},$$

$$C_{xx_T} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} \right)^2}{\sum_{i=1}^k n_i},$$

$$C_{yy_T} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \right)^2}{\sum_{i=1}^k n_i},$$

$$C_{xy_T} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} y_{ij} - \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}}{\sum_{i=1}^k n_i}.$$

De som van de kwadraten van de afwijkingen van de y waarden van de geschatte regressielijnen binnen de afzonderlijke groepen, gesommeerd over de groepen is

$$\underline{S}_1 = \sum_{i=1}^k \left(\underline{C}_{yyi} - \frac{\underline{C}_{xyi}^2}{\underline{C}_{xxi}} \right).$$

Als het model (1) geldt, heeft $\underline{S}_1 / \sigma^2$ een χ^2 -verdeling met $\sum_{i=1}^k n_i - 2k$ vrijheidsgraden.

Wij kunnen ook zo goed mogelijk evenwijdige regressielijnen in de verschillende groepen schatten. De som van de kwadraten van de afwijkingen van deze lijnen is

$$\underline{S}_2 = \sum_{i=1}^k \underline{C}_{yyi} - \frac{\underline{C}_{nyw}^2}{\underline{C}_{xxw}}.$$

Deze som gedeeld door σ^2 heeft een χ^2 -verdeling met $\sum_{i=1}^k n_i - k - 1$ vrijheidsgraden als hypothese (B) of (C) waar is.

Als hypothese (A) geldt, heeft de som van kwadraten van afwijkingen van één lijn

$$\underline{S}_3 = \underline{C}_{yyT} - \frac{\underline{C}_{xyT}^2}{\underline{C}_{xxT}},$$

gedeeld door σ^2 , een χ^2 -verdeling met $\sum_{i=1}^k n_i - 2$ vrijheidsgraden.

Voor de hypothese (A) gebruiken wij de toetsingsgrootte

$$F_A = \frac{\sum_{i=1}^k n_i - 2k}{k-1} \frac{\underline{S}_2 - \underline{S}_1}{\underline{S}_1},$$

die onder de hypothese (A) een F-verdeling bezit met $k-1$ en $\sum_{i=1}^k n_i - 2k$ vrijheidsgraden.

Als toetsingsgrootte voor de hypothese (B) dient

$$F_B = \frac{\sum_{i=1}^k n_i - k - 1}{k-1} \frac{\underline{S}_3 - \underline{S}_2}{\underline{S}_2},$$

welke grootte een F-verdeling met $k-1$ en $\sum_{i=1}^k n_i - k - 1$ vrijheidsgraden heeft, als (B) waar is.

Hypothese (C) toetsen wij tenslotte met

$$F_C = \frac{\sum_{i=1}^k n_i - 2k}{2(k-1)} \frac{\underline{S}_3 - \underline{S}_1}{\underline{S}_1},$$

een F verdeelde grootte met $2(k-1)$ en $\sum_{i=1}^k n_i - 2k$ vrijheidsgraden, onder (C).

Als de betreffende hypothesen niet vervuld zijn, hebben de bijbehorende F 's geen F -verdeling, maar zijn grotere F waarden meer waarschijnlijk. Als kritieke zône gebruikt men daarom $F \geq F_0$, waarin F_0 behorende bij een bepaalde onbetrouwbaarheidsdrempel ε opgezocht kan worden in tabellen van de F -verdeling met het juiste aantal vrijheidsgraden.

Literatuur over de achtergrond van bovengenoemde toetsen kan men vinden in MANN [4] en KENDALL [3]. Het rekenschema is grotendeels ontleend aan DIXON en MASSEY [1]. Al deze boeken bevatten tabellen van de F -verdeling. Deze verdeling is het uitvoerigst getabelleerd in FISHER en YATES [2].

Literatuur

- | | |
|-----------------------------------|---|
| [1] W.J. DIXON en
F.J. MASSEY, | Introduction to Statistical Analysis,
Mc Graw-Hill Book Company, Inc., New York,
Toronto, London, 1951, Chapter 12. |
| [2] R.A. FISHER en
F. YATES, | Statistical Tables for Biological, Agricultural and Medical research, 3d ed.,
Oliver & Boyd, London 1949, Table V. |
| [3] M.G. KENDALL, | The advanced theory of statistics, Vol II,
2 nd ed., Griffin, London, 1948, pp 237-246. |
| [4] H.B. MANN, | Analysis and Design of Experiments, Dover
Publication, Inc., New York 1949. |
