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 191

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

I. De betrouwbaarheid van de steekproef.

door

A. Reitsma

Januari

1956

1. Inleiding

Gedurende de 7 vangstseizoenen 1948-1949 t/m 1954-1955 zijn aan boord van een walvisvaarder van iedere gedode walvis, de volgende gegevens opgetekend.

(1) volgnummer,

(2) tijd van doden: uur en datum,[1)]

(3) soort: Potvissen, Bultruggen, Blauwe vinvissen, gewone vinvissen, e.a.,[2)]

(4) lengte, uitgedrukt in voeten,

(5) geslacht, bij een wijfje wordt bovendien aangegeven of het drachtig of niet drachtig is.

Bovendien wordt iedere dag in de vangstlijsten de

(6) middagpositie van de walvisvaarder genoteerd.

Het vangstseizoen is de periode waarin de walvissen gevangen worden; dit duurt gewoonlijk van midden December van een jaar tot half Maart van het volgende jaar.

Uit de gevangen walvissen werd een steekproef genomen. Van iedere in de steekproef voorkomende walvis werden behalve de bovengenoemde ook de volgende waarnemingen genoteerd:

(7) speklaagdikte in cm, dorsaal (d.w.z. aan de rugzijde), craniaal van (d.w.z. voor) de dorsale vin gemeten,

(8) speklaagdikte in cm, ventraal (d.w.z. aan de buikzijde), craniaal van de anus gemeten.

Met behulp van deze gegevens tracht men enkele vragen te beantwoorden betreffende het verband tussen de speklaagdikte en de datum, tussen speklaagdikte en lengte en omtrent mogelijke verschillen tussen de speklaagdikten in verschillende seizoenen. Op deze vragen hopen wij terug te komen in een volgend rapport. Willen de antwoorden op bovengenoemde vragen betrouwbaar zijn, dan moet de wijze van steekproef nemen aan zekere eisen voldoen, waarvan in dit rapport wordt nagegaan in hoeverre deze vervuld zijn.

----

1) Het uur van de vangst is alleen aangegeven in de originele vangstlijsten en niet in de copieën die wij tot onze beschikking hadden. Slechts van de seizoenen 1952-1953 en 1953-1954 waren later ook de originele vangstlijsten in ons bezit.

2) In dit rapport wordt alleen aandacht geschonken aan de gewone vinvissen, die de grootste groep van het materiaal vormen.

2. De steekproef

Om uit een populatie een steekproef te nemen waaruit men betrouwbare conclusies kan trekken betreffende zekere kenmerken van die populatie, zoals in dit geval lengte, speklaagdikte, geslacht, enz., is het nodig dat de kans van een individu om in de steekproef te komen niet afhangt van deze of hiermee samenhangende kenmerken. Dit is b.v. te bereiken door bij iedere walvis die aan dek komt met een vaste kans te loten of het dier gemeten zal worden of niet. Men bereikt dan dat alle mogelijke steekproeven van dezelfde uitgebreidheid een zelfde kans hebben om gekozen te worden. Zodoende heeft b.v. een vette walvis even veel kans om in de steekproef terecht te komen als een magere en zal, als de steekproef niet te klein is, de fractie walvissen met een lengte van meer dan 70 voet in de steekproef ongeveer even groot zijn als die fractie in de populatie.

Aan boord van de walvisvaarder was, althans gedurende de eerste vier van de beschouwde seizoenen, het voorschrift dat iedere eerste walvis die na een oneven uur gevangen wordt, gemeten moet worden (H.M. VAN DER STELT-STRIJBOS [1] , pagina 3). Deze methode heeft het bezwaar dat op dagen waarop weinig gevangen wordt een veel grotere fractie dieren gemeten wordt dan op dagen met grote vangsten. Van walvissen die in grote scholen zwemmen, zullen dus naar verhouding minder exemplaren in de steekproef voorkomen dan van walvissen die meer solitair optreden en het lijkt a priori niet uitgesloten dat beide groepen walvissen systematisch verschillen wat lengte en speklaagdikte betreft, zodat de steekproef dus wellicht geen juist beeld geeft van de populatie.

Toen wij van de jaren 1952-1953 en 1953-1954 de oorspronkelijke protocollen ter beschikking kregen, bleek evenwel dat, althans in deze jaren, niet volgens bovengenoemd criterium te werk was gegaan. Dit blijkt ook uit de aantallen gemeten walvissen in de verschillende jaren, zoals wij die vinden in tabel I.

Tabel I

Gewone vinvissen gedood in de vangstseizoenen 1948-1949 t/m 1954-1955.

Seizoen	1948-1949	1949-1950	1950-1951	1951-1952	1952-1953	1953-1954	1954-1955
speklaag gemeten	76	419	434	160	184	214 1)	173
speklaag niet gemeten	693	509	395	541	760	831	441
totaal	769	928	829	701	944	1045	614

Het valt op dat in het 2e en 3e vangstseizoen veel meer dieren zijn gemeten dan in de andere jaren. Het blijkt dan ook dat in deze jaren op sommige dagen veel meer walvissen gemeten zijn dan volgens het voorschrift mogelijk was. Soms zijn nl. 25 à 30 dieren gemeten, terwijl het er hoogstens 12 zouden mogen zijn.

In dit rapport zullen wij onderzoeken of de genomen steekproeven representatief zijn wat de lengte betreft. Dit kan, daar de lengte van alle gevangen dieren is opgenomen. Voor de speklaagdikte is dit uiteraard niet na te gaan, omdat deze alleen van de tot de steekproef behorende dieren bekend is.

3. Het statistische onderzoek

Een globaal overzicht van de lengten van alle gevangen dieren en van die uit de steekproef in de verschillende jaren kunnen wij krijgen uit de figuren 1a t/m 7a, waarin voor ieder jaar een histogram is getekend.

Aan de histogrammen voor de jaren 1951-1952, 1952-1953 en 1953-1954 valt op, dat de lengtegroep van 60 voet zeer sterk en die van 59 voet niet of bijna niet vertegenwoordigd is, terwijl ook de dieren van 58 voet zeldzaam zijn. Volgens VAN DER STELT-STRIJBOS [1] is het verboden gewone vinvissen die korter zijn dan 55 voet te vangen. In de protocollen worden evenwel alle gewone vinvissen beneden de 60 voet als onder de maat aangegeven.

----

1) Op de originele protocollen komen 241 gemeten dieren voor. Op de copieën, waarop de in het volgende rapport te vermelden berekeningen zijn gebaseerd, zijn hier echter om ons onbekende redenen 27 van weggelaten. In par. 4 komen wij nog op deze kwestie terug.

Daar wel weer dieren van 50 tot 58 voet gevangen zijn, valt niet aan te nemen, dat de harpoeniers een zo scherp onderscheid tussen lengten van 59 en van 60 voet weten te maken, zodat wij tot de conclusie komen, dat in de drie genoemde jaren systematisch ondermaatse walvissen en wel voornamelijk die van 59 voet en gedeeltelijk die van 58 voet voor 60 genoteerd zijn.

Verder zijn in de figuren 1b t/m 7b de fracties walvissen waarvan de speklaag is gemeten uitgezet voor de verschillende lengteklassen, omdat aan deze figuren beter dan aan de histogrammen te zien is of deze fracties eventueel systematisch van elkaar verschillen. Bij het beschouwen van deze figuren houde men er rekening mee, dat de fracties, indien deze uit kleine aantallen berekend zijn, grote toevallige verschillen kunnen vertonen. Dit is dus in het bijzonder aan de uiteinden het geval. Hieraan is gedeeltelijk tegemoet gekomen door samenvoeging tot klassen. De figuren geven de indruk dat vooral weer in de jaren 1951-1952, 1952-1953 en 1953-1954 de fracties een lichte stijgende tendens vertonen met toenemende lengte. Dit betekent dus dat naar verhouding van de langere dieren van meer exemplaren de speklaagdikte is gemeten dan van de kortere dieren.

Om op een meer exacte wijze na te gaan of de kansen om tot de steekproef te behoren voor walvissen van verschillende lengte gelijk zijn, passen wij een statistische toetsingsmethode toe (memorandum S 47 (M 6))[1]. Wij hebben hiertoe twee toetsen toegepast en wel in de eerste plaats de toets voor de gelijkheid van een aantal kansen met behulp van een 2xk-tabel (memorandum S 145 (M 23a)). Deze toets reageert op alle mogelijke afwijkingen van de nulhypothese, i.e. de gelijkheid van de kansen om tot de steekproef te behoren. Daarnaast bestaat een methode die speciaal gericht is op het ontdekken van een verloop in de genoemde kansen en die dus reageert als de kansen toenemen met toenemende lengte of omgekeerd. Deze toets tegen verloop voor een aantal kansen hebben wij eveneens toegepast (memorandum S 139 (M 48)). Om de verdelingen van de beide toetsingsgrootheden goed te kunnen benaderen met resp. de χ^2 en de normale verdeling, zoals is beschreven in de betreffende memoranda, is het nodig dat de aantallen gevangen dieren binnen een klasse niet te klein zijn voor de eerste en niet te veel verschillen voor de tweede toets. Daarom hebben wij, voor beide toetsen op dezelfde manier, enkele kleine en enkele grote lengteklassen samengenomen.

----

1) De vermelde memoranda zijn aan het eide van dit rapport bijgevoegd.

De resultaten van beide toetsen vinden wij in tabel II.

Tabel II

Overschrijdingskansen van de 2xk-tabel toets
en van de toets voor verloop van kansen.

Vangstseizoen	'48-'49	'49-'50	'50-'51	'51-'52	'52-'53	'53-'53	'54-'55
2xk-tabel	0,74	0,86	0,04	0,05	0,15	0,06	0,45
verlooptoets	0,84(+)	0,50(+)	0,18(+)	0,02(+)	0,06(+)	0,004(+)	0,12(+)

Een + teken achter de overschrijdingskans bij de toets voor verloop betekent dat bij groter wordende lengte de kans om in de steekproef te worden opgenomen toeneemt. Wij zien dat de kleinste overschrijdingskansen voorkomen bij het derde, vierde, vijfde en zesde seizoen. Bij het derde seizoen is er wel een aanwijzing dat de kansen verschillen, maar niet dat ze toe- of afnemen met toenemende lengte. Bij het vierde en het zesde seizoen is er een iets zwakkere aanwijzing dat de kansen verschillen, maar de toets tegen verloop geeft duidelijk aan dat de kansen toenemen met de lengte. Bij het vijfde seizoen tenslotte is er een zwakke aanwijzing voor een verloop, terwijl de χ^2-toets geen kleine overschrijdingskans geeft.

Opmerkelijk is verder dat de toets voor verloop voor alle zeven jaren een afwijking van de verwachte waarde van de toetsingsgrootheid naar dezelfde kant oplevert. Combinatie van deze zeven toetsingen geeft een overschrijdingskans van 0,00013. (Deze combinatie is uitgevoerd door de toetsingsgrootheden der afzonderlijke jaren op te tellen. Deze som heeft weer bij benadering een normale verdeling met als variantie de som van de varianties van de afzonderlijke termen.) Onze conclusie luidt dan ook dat er overtuigende aanwijzingen voor bestaan dat in de steekproef naar verhouding meer lange walvissen voorkomen dan in de populatie van alle gevangen dieren, terwijl dit verschijnsel het sterkst naar voren treedt in de derde tot en met het zesde seizoen.

4. De metingen in het seizoen 1953-1954

Zoals reeds bij tabel I werd opgemerkt zijn in dit seizoen volgens de oorspronkelijke protocollen 241 dieren gemeten. Passen wij op dit aantal de beide toetsen toe, dan vinden wij het volgende resultaat.

Tabel III

Overschrijdingskansen van de χ^2-toets en de verloop-toets toegepast op de werkelijk gemeten dieren in het vangstseizoen 1953-1954 vergeleken met de gedeeltelijke steekproef.

	Totale steekproef (241)	Gedeeltelijke steekproef(214)
χ^2-toets	0,003	0,06
Toets voor verloop	0,00007 (+)	0,004

Wij zien hieruit dat het verloop van de kans om in de steekproef te worden opgenomen bij de volledige steekproef nog sterker optreedt dan bij de steekproef waarbij 27 dieren zijn weggelaten. Deze compensatie door het weglaten van 27 walvissen kan een gevolg zijn van het feit dat ook deze 27 dieren niet aselect uit de steekproef van 241 zijn gekozen, m.a.w. de langste dieren zijn er uit gelaten.

5. Conclusie

De steekproef blijkt, in het bijzonder voor de jaren 1951-1952, 1952-1953 en 1953-1954 wat de lengte betreft niet representatief te zijn. Bij de verdere analyse van de speklaagdikten zou het nog wel mogelijk zijn hier rekening mee te houden. Het is echter gezien deze resultaten ook zeer goed denkbaar dat de steekproef in sommige jaren niet representatief is voor de speklaagdikte, hetgeen echter uit de beschikbare gegevens niet na te gaan is, zodat conclusies betreffende verschillen in dikte van de speklaag tussen de jaren onder voorbehoud gegeven zullen moeten worden. Onderzoekingen omtrent verband tussen lengte en speklaagdikte zullen vermoedelijk door het gebrek aan representativiteit minder sterk beïnvloed worden.

Het verdient aanbeveling, indien mogelijk, de manier waarop de steekproef wordt genomen zo te wijzigen dat een beter resultaat kan worden verwacht. Dit zou kunnen geschieden door b.v. telkens van de 5e, de 10e, de 15e enz. walvis de lengte en de speklaag te meten. Tijdens drukke perioden kan dan overgegaan worden tot het meten van de 10e, de 20e enz. walvis, mits bij de waarnemingen wordt genoteerd wanneer dit gebeurt.

Daar het ook zeer wel mogelijk is, dat verschillende waarnemers een enigszins verschillende waarnemingstechniek gebruiken en daardoor tot systematisch verschillende resultaten komen, zou het voor de statistische analyse tevens van belang zijn bij de metingen van iedere walvis de waarnemer te noteren, b.v. door de metingen te laten paraferen.

Literatuur

[1] H.M. van der Stelt-Strijbos, Voorlopig verslag. Bewerking van waarnemingen, gedaan aan boord van de Willem Barendsz, betreffende speklaagdikte en lengte der walvissen, niet gepubliceerd.

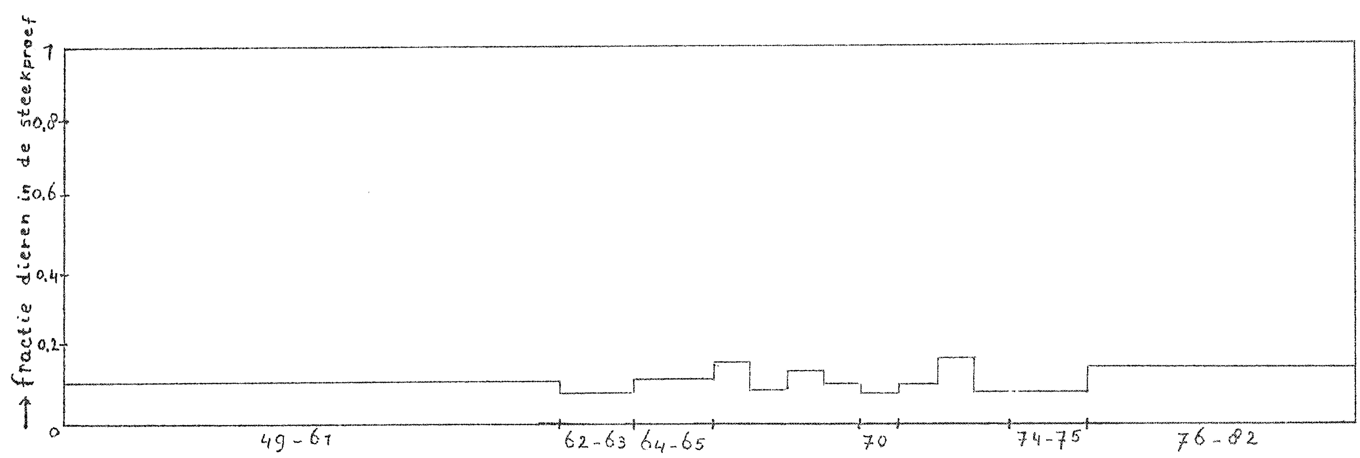

Fig. 1b. Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1948-1949.

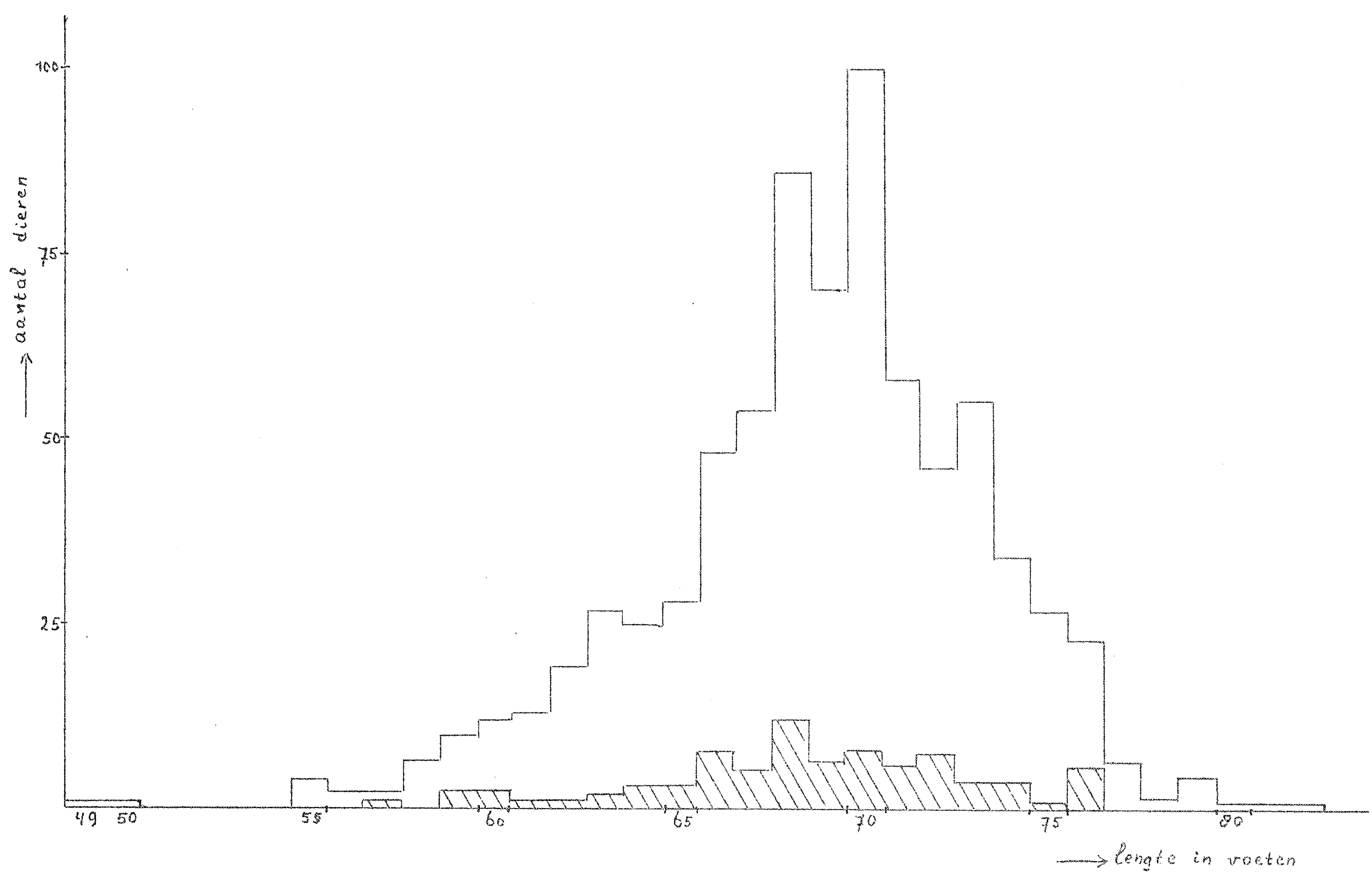

Fig. 1a. Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1948-1949.

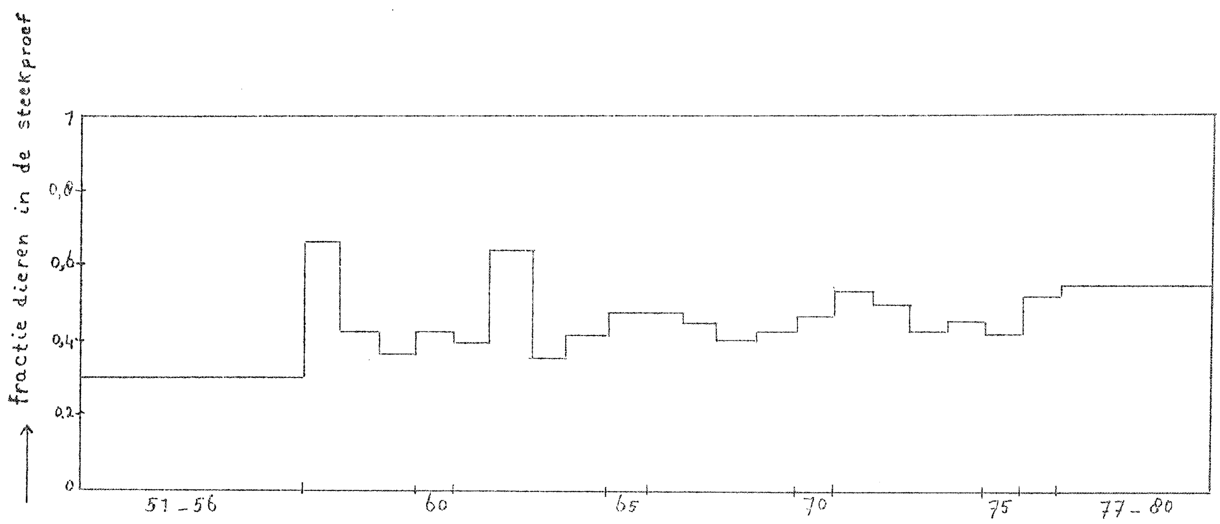

Fig. 2b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1949-1950.

Fig. 2a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1949-1950.

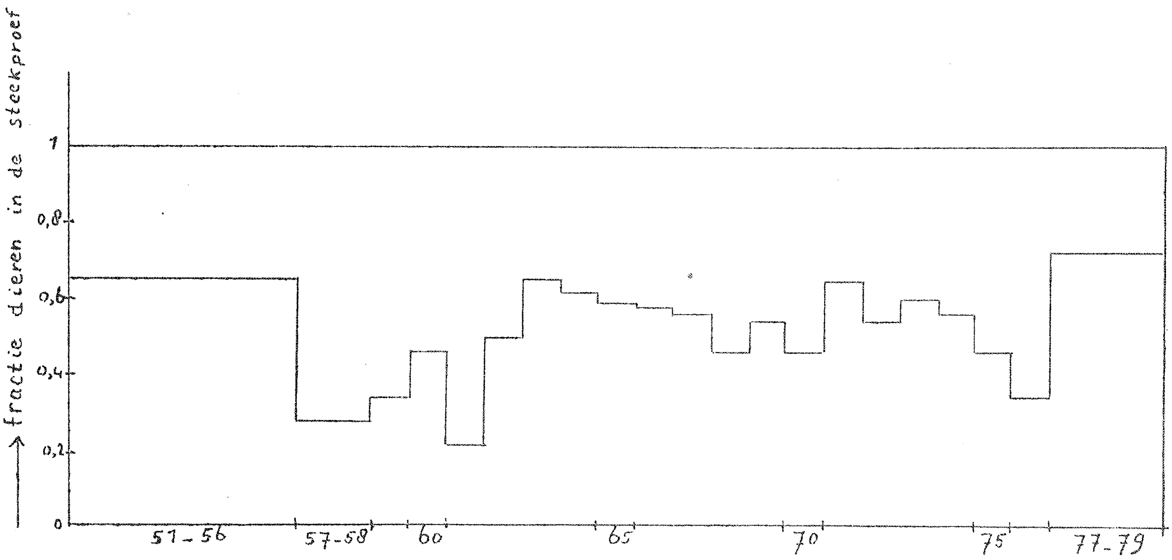

Fig. 3b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1950-1951.

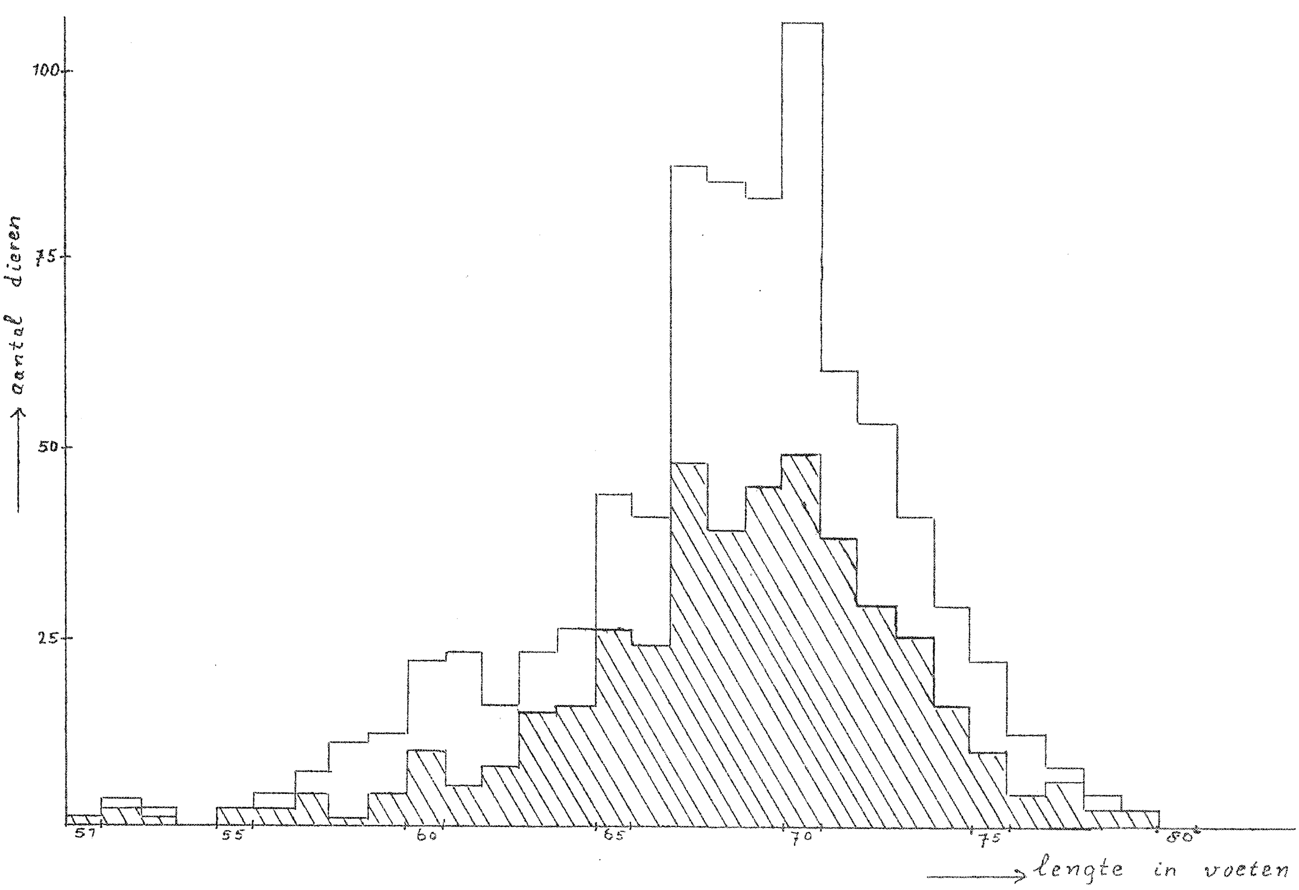

Fig. 3a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1950-1951.

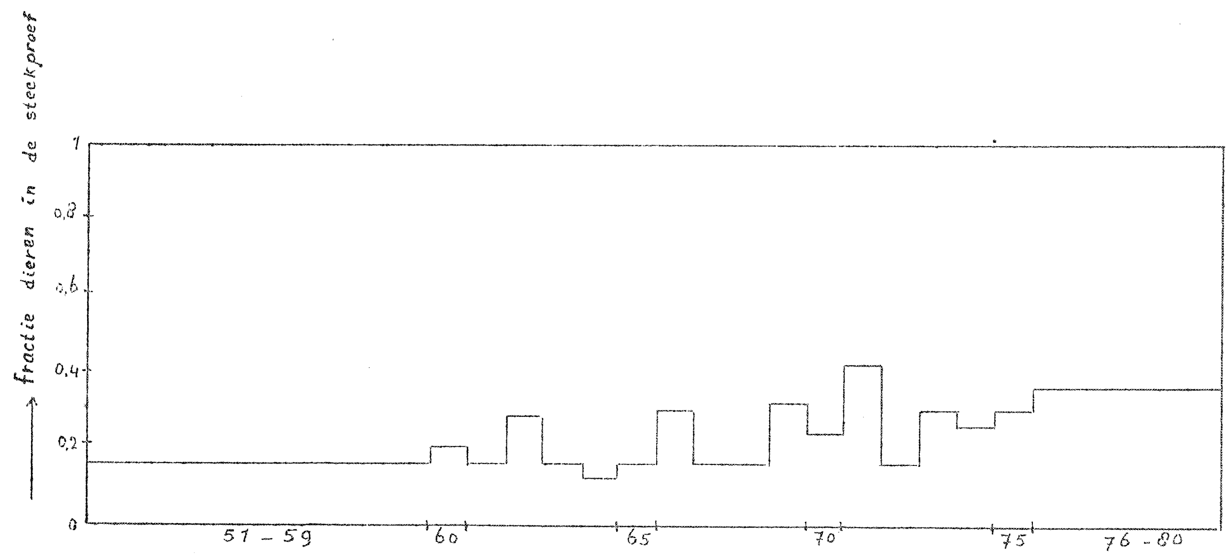

Fig. 4b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1951-1952.

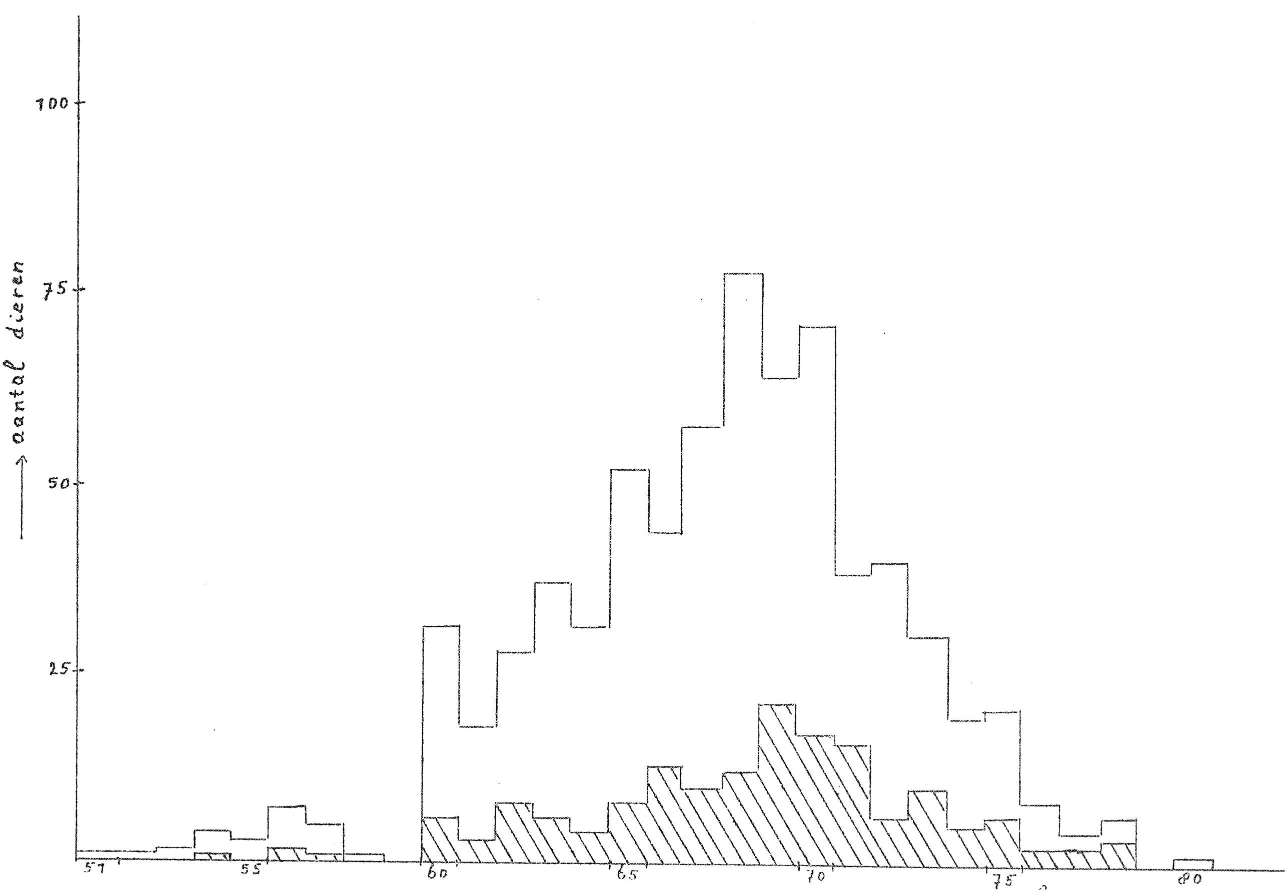

Fig. 4a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1951-1952.

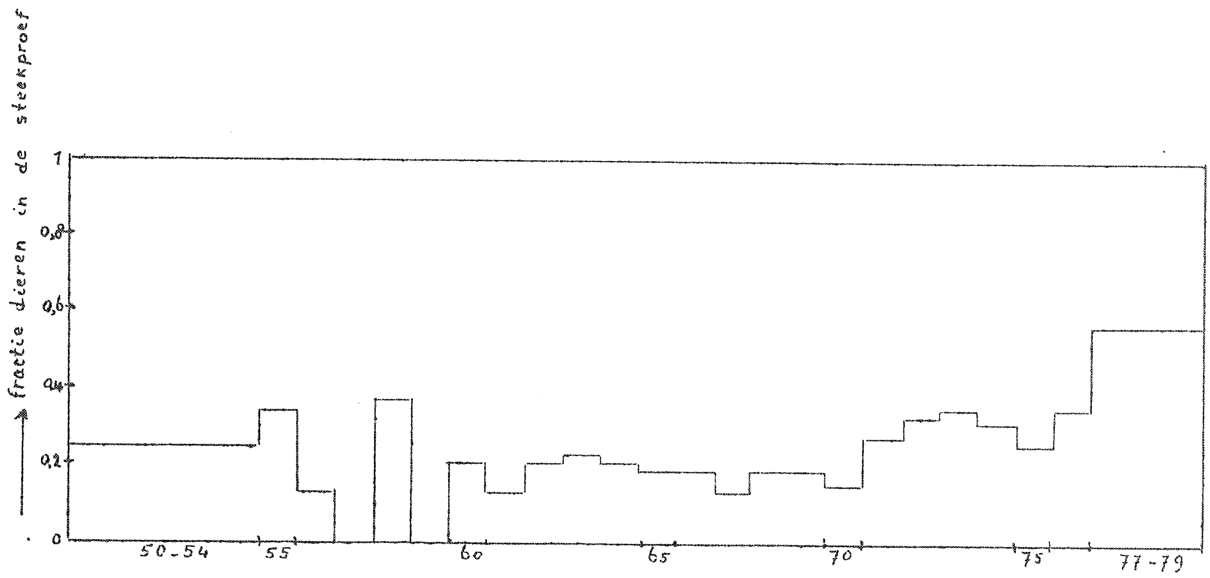

Fig. 5b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1952-1953.

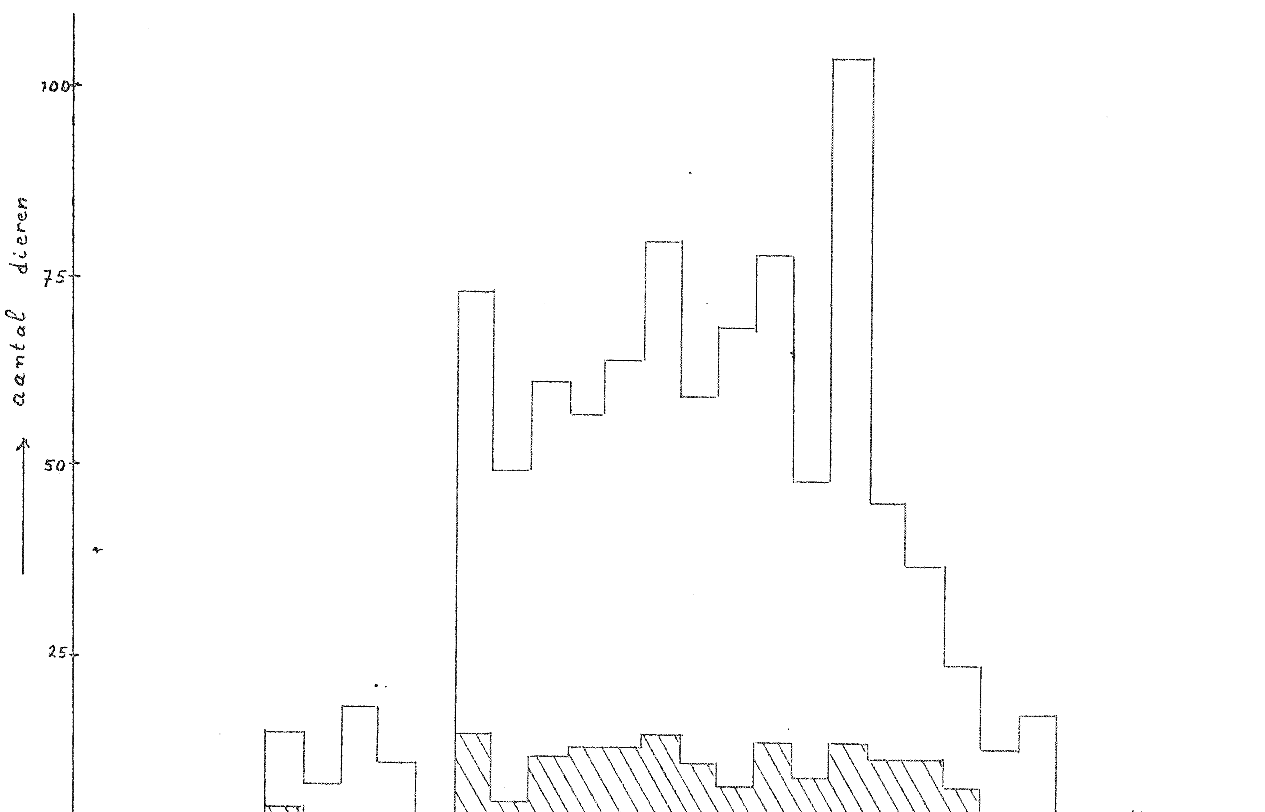

Fig. 5a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1952-1953.

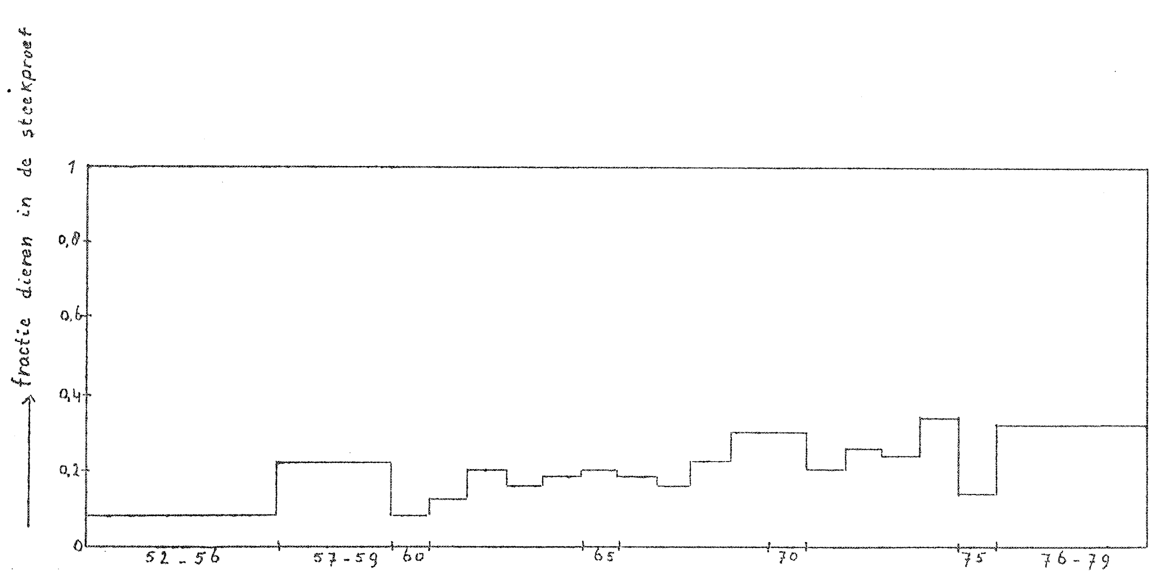

Fig. 6b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1953-1954.

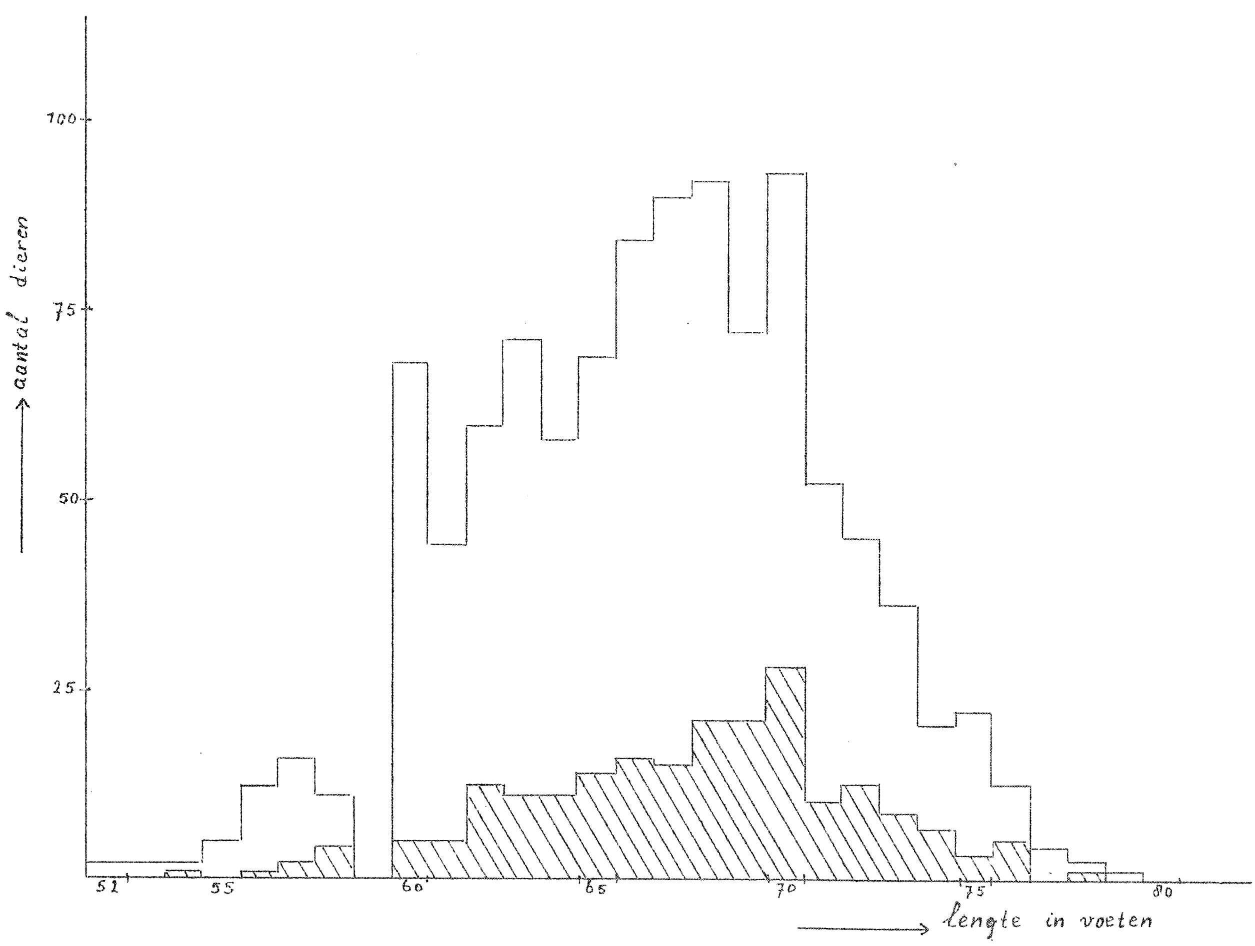

Fig. 6a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1953-1954.

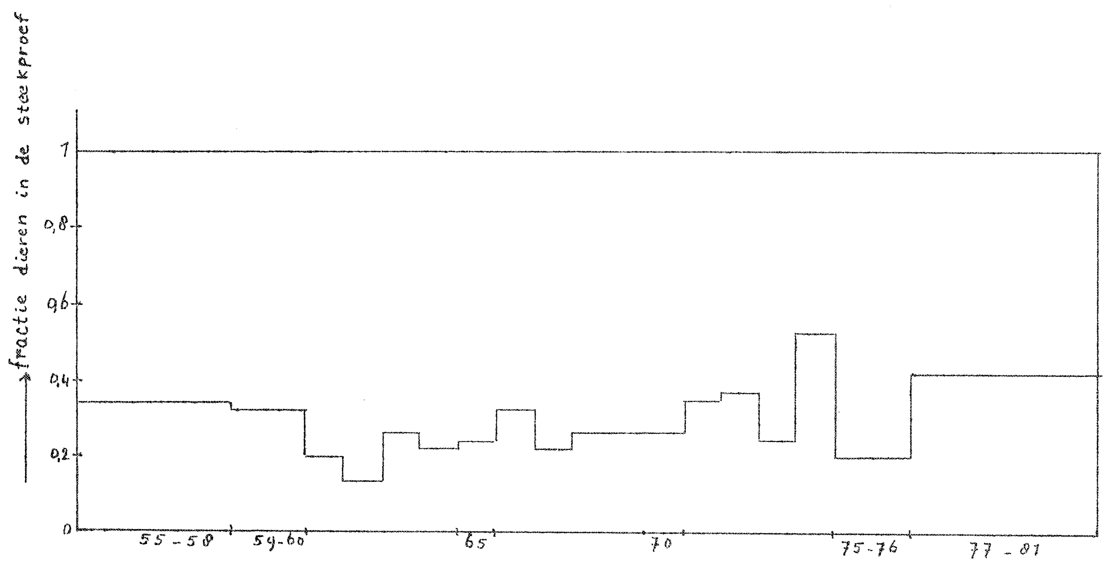

Fig. 7b Walvissen van de steekproef als fractie van alle gevangen dieren per lengtegroep.
Vangstseizoen 1954-1955.

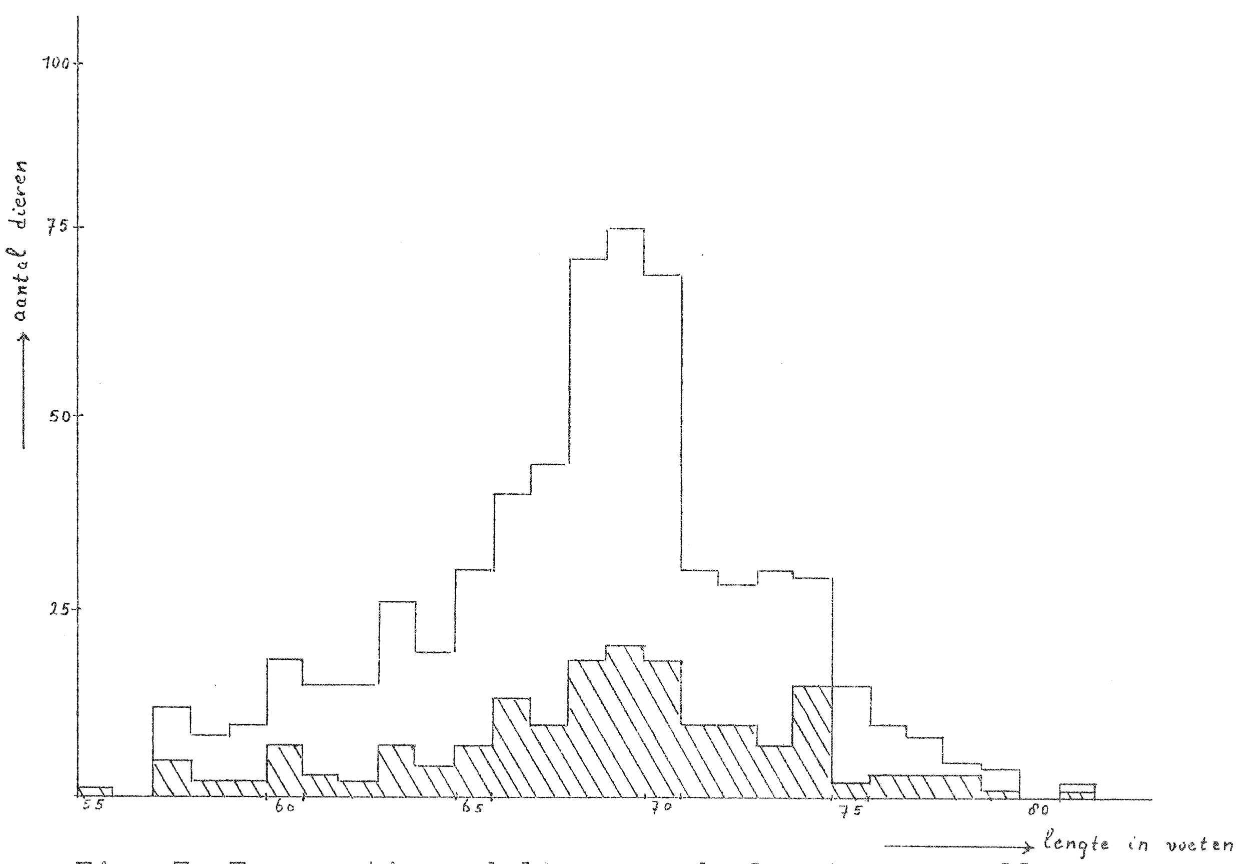

Fig. 7a Frequentieverdeling van de lengten van alle gevangen walvissen en van de tot de steekproef behorende walvissen (gearceerd).
Vangstseizoen 1954-1955.

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 H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

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

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

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

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

Litteratuur:

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

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

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
A m s t e r d a m - O.
Statistische Afdeling
S 145 (M 23a)

Toets voor de hypothese $p_1 = p_2 = \ldots = p_k$ met behulp van een 2 x k tabel 1)

Wij beschouwen k reeksen $R_1, R_2, \ldots, R_k$ van onafhankelijke waarnemingen, waarbij iedere waarneming als resultaat het kenmerk A of het kenmerk $\bar{A}$ (non A) kan geven. De kans op A is binnen ieder der reeksen constant en wel gelijk aan p_i voor de waarnemingen van reeks R_i ($i = 1,2,\ldots,k$).
Laat het aantal waarnemingen van reeks R_i ($i = 1,2,\ldots,k$) gelijk zijn aan n_i en laat hieronder het aantal met kenmerk A, m_i zijn.
Gevraagd wordt dan de hypothese H_0: $p_1=p_2= \ldots =p_k$
te toetsen op grond van deze gegevens.
De gegevens kunnen in een 2 x k -tabel worden samengevat;

	R_1	R_2		R_i		R_k	totaal
A	m_1	m_2	...	m_i	...	m_k	m
$\bar{A}$	n_1-m_1	n_2-m_2	...	n_i-m_i	...	n_k-m_k	$n-m$
totaal	n_1	n_2		n_i		n_k	n

waarin dus $m_1 + m_2 + \ldots + m_k = m$
en $n_1 + n_2 + \ldots + n_k = n$.
De hypothese H_0 wordt getoetst met de grootheid

$$\chi_c^2 = \sum_i \frac{\left(m_i - \frac{m\,n_i}{n}\right)^2}{\frac{m\,n_i}{n}} + \sum_i \frac{\left(n_i - m_i - \frac{(n-m)n_i}{n}\right)^2}{\frac{(n-m)\,n_i}{n}} =$$

$$= \sum_i \frac{(n m_i - m n_i)^2}{m(n-m)n_i} = \frac{n^2}{m(n-m)} \sum_i \frac{m_i^2}{n_i} - \frac{nm}{n-m}$$

Deze grootheid $\underline{\chi}_c^2$ 2) heeft onder de hypothese H_0 bij benadering een χ^2 -verdeling met $k-1$ vrijheidsgraden (zie b.v. [1] p. 445 e.v.).

---

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

2) Als wij grootheden als stochastische grootheden (dit zijn grootheden met een waarschijnlijkheidsverdeling) beschouwen geven wij dit door onderstreping aan. Niet onderstreepte letters geven waarden aan, die door de stochastische grootheden worden aangenomen.

Deze benadering is goed, indien $m\frac{n_i}{n} \geqq 5$ voor iedere i (zie [2]). Indien H_0 onjuist is, dus als er bij verschillende reeksen verschillende kansen op A zijn, zal $\underline{\chi}_c^2$ gewoonlijk grotere waarden aannemen, dan wanneer H_0 juist is.

De kritieke zone bestaat uit die waarden van $\underline{\chi}_c^2$, waarvoor geldt $\underline{\chi}_c^2 \geqq \chi_\alpha^2$. Hierin is χ_α^2 die waarde van $\underline{\chi}^2$, die voldoet aan

$$P[\underline{\chi}^2 \geqq \chi_\alpha^2] = \alpha$$

met α als van te voren vastgelegde onbetrouwbaarheid.

De overschrijdingskans behorende bij een bepaalde gevonden waarde χ_c^2 van $\underline{\chi}_c^2$ is gedefinieerd als

$$P[\underline{\chi}_c^2 \geqq \chi_c^2 | H_0]$$

waarin "$|H_0$" aangeeft, dat deze kans berekend wordt op grond van H_0.

χ_α^2 en de overschrijdingskans kunnen in tabellen of nomogrammen worden opgezicht (zie [3]).

Opmerking. Indien niet voldaan is aan de voorwaarde $m\frac{n_i}{n} \geqq 5$ voor iedere i, kan men een (meer bewerkelijke) exacte toets baseren op de voorwaardelijke waarschijnlijkheidsverdeling van de grootheden $\underline{m}_i$ ($i = 1, \ldots, k$), onder de voorwaarde, dat hun som de waarde m aanneemt:

$$P[\underline{m}_1 = m_1, \underline{m}_2 = m_2, \ldots, \underline{m}_k = m_k | \underline{m}_1 + \underline{m}_2 + \cdots + \underline{m}_k = m\,; H_0] =$$

$$= \binom{n_1}{m_1}\binom{n_2}{m_2} \cdots \binom{n_k}{m_k} \Big/ \binom{n}{m}$$

De geldigheid van deze formule volgt direct uit de waarschijnlijkheidsverdelingen van de $\underline{m}_i$ en van $\underline{m}$ (onder H_0) en uit de definitie van een voorwaardelijke waarschijnlijkheid.

In dit geval definiëren wij de overschrijdingskans behorend bij een gevonden resultaat $(m_1, m_2, \ldots, m_k)$ met $m_1 + m_2 + \ldots + m_k = m$ als de som van alle waarschijnlijkheden van bovengenoemde verdeling (met de gevonden waarde van m), die hoogstens gelijk zijn aan de waarschijnlijkheid van het gevonden resultaat.

Literatuur.

[1] H.Cramér, Mathematical methods of statistics, Princeton University Press, 1946.

[2] P.G.Hoel, On indices of dispersion, Ann. Math. Stat. 14 (1943), p. 155-163.

[3] Tabellen en nomogrammen van de -verdeling.

M.G.Kendall, The advanced theory of statistics, I, 1947, p. 444-446.

H.Cramér, Mathematical methods of statistics, Princeton University Press, 1946, p. 559.

Statistica 1 (1946), p. 109.

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
A m s t e r d a m - O.
Statistische Afdeling
Rapport S 139 (M 48)

Toets tegen verloop voor een aantal kansen [1]

Deze toets kan toegepast worden in het volgende geval: R_i $(i = 1, 2, \ldots, k)$ zijn k onafhankelijke reeksen van onafhankelijke experimenten, waarbij ieder experiment één der mogelijke uitkomsten "succes" of "mislukking" heeft. Bestaat de i^e reeks uit n_i experimenten, treedt hierbij $\underline{a}_i$ [2] maal de uitkomst succes en $\underline{b}_i$ maal de uitkomst mislukking op $(i = 1, 2, \ldots, k)$, dan kunnen wij de resultaten als volgt samenvatten:

reeks no.	aantallen		
	successen	mislukkingen	experimenten
1	$\underline{a}_1$	$\underline{b}_1$	n_1
2	$\underline{a}_2$	$\underline{b}_2$	n_2
.	.	.	.
.	.	.	.
.	.	.	.
k	$\underline{a}_k$	$\underline{b}_k$	n_k
totaal	$\underline{t}_1$	$\underline{t}_2$	n

Is nu de kans op succes bij ieder experiment van de i^e reeks p_i $(i = 1, 2, \ldots, k)$ dan luidt de hypothese die wij willen toetsen:

(1) $$H_0 : p_1 = p_2 = \ldots = p_k ,$$

terwijl de alternatieve hypothesen inhouden dat de kansen p_1, $p_2, \ldots, p_k$ in deze volgorde een stijgens of dalend verloop vertonen, gedefiniëerd door

(2) $$\sum_{i<j}\sum (p_i - p_j) \neq 0 .$$

----

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

2) Stochastische grootheden worden onderscheiden van getallen (b.v. van de waarden die zij bij een experiment aannemen) door hun symbolen te onderstrepen.

De toets wordt voorwaardelijk uitgevoerd onder de voorwaarde, dat $\underline{t}_1$ de bij het experiment gevonden waarde t_1 aanneemt $(\underline{t}_1 + \underline{t}_2 = n)$ en de toetsingsgrootheid is

$$(3) \qquad \underline{W} = \sum_{i<j}\sum \left(\frac{a_i}{n_i} - \frac{a_j}{n_j}\right) = \sum_i (k+1-2i)\frac{a_i}{n_i} .$$

Voor deze grootheid $\underline{W}$ geldt

$$(4) \qquad \mathcal{E}(\underline{W} \mid t_1 ; H_0) = 0,$$

$$(5) \qquad \sigma^2(\underline{W} \mid t_1 ; H_0) = \frac{t_1 t_2}{n(n-1)} \sum_i \frac{(k+1-2i)^2}{n_i} .$$

Voor grote waarden van n en als t_1 en t_2 onderling en $n_1, n_2, \ldots n_k$ onderling niet te veel in grootte verschillen is de verdeling van $\underline{W}$ onder de voorwaarde $\underline{t}_1 = t_1$ en onder de hypothese H_0 bij benadering normaal met gemiddelde en variantie volgens (4) en (5).

Hoe men de exacte verdeling van $\underline{W}$ onder de hypothese H_0 en onder de voorwaarde $\underline{t}_1 = t_1$ berekent vindt men beschreven in [2].

De grootheid $\underline{W}$ zal in het algemeen grote positieve waarden aannemen als de kansen $p_1, p_2, \ldots, p_k$ een dalend verloop vertonemen en grote negatieve waarden bij een stijgend verloop van deze kansen; voor de tweezijdige kritieke zone nemen wij dus grote waarden van $|W|$.

Als $n_i = m$ voor iedere i (dus als de reeksen alle uit m experimenten bestaan) kan de toets beschouwd worden als een toepassing van de toets van WILCOXON (zie b.v. [3] en [4]). Beschouw nl. twee onderling onafhankelijke steekproeven ter grootte t_1 resp. t_2 , die tezamen genomen m maal de waarde i bevatten $(i = 1, 2, \ldots, k)$. Als de eerste steekproef bestaat uit $\underline{a}_i$ maal de waarde i en de tweede uit $\underline{b}_i$ maal de waarde i $(i = 1, 2, \ldots, k)$ en men past op deze twee steekproeven de toets van WILCOXON toe dan geldt

$$(6) \qquad 2\,\underline{U} = -m^2\,\underline{W} + t_1 t_2 ,$$

waarin $\underline{U}$ de toetsingsgrootheid van de toets van WILCOXON is. 3) Verder kan men bewijzen dat de hypothese H_0 onder de voorwaarde $\underline{t}_1 = t_1$ identiek is met de hypothese dat de twee steekproeven afkomstig zijn uit dezelfde verdeling onder de voorwaarde dat deze twee steekproeven tezamen genomen m maal de waarde i bevatten $(i = 1, 2, \ldots, k)$. Dit betekent dus dat de toets voor de hypothese H_0: $p_1 = p_2 = \ldots = p_k$ gebaseerd op de grootheid (3) in dit geval identiek is met de toets van WILCOXON, toegepast op twee steekproeven die, tezamen genomen, m maal de waarde i bevatten en waarvan de eerste steekproef bestaat uit $\underline{a}_i$ maal de waarde i $(i = 1, 2, \ldots, k)$.

Uit (6) volgt nog dat een rechter kritieke zone voor $\underline{W}$ overeenkomt met een linker kritieke zone voor $\underline{U}$ en omgekeerd, d.w.z. ligt $\underline{W}$ in een rechter (resp. linker) kritieke zone, dan ligt $\underline{U} = \frac{-m^2\underline{W} + t_1 t_2}{2}$ in een linker (resp. rechter) kritieke zone. Nu is voor het geval dat $m = 1$, de exacte verdeling van de toetsingsgrootheid van WILCOXON, voor kleine waarden van de steekproefgrootten getabeleerd; in [3] b.v. vindt men tabellen voor t_1 en t_2 beide $\leqq 10$ voor de verdeling van $2\underline{U}$. In dit geval kan men de toets tegen verloop voor kansen dus exact uitvoeren; de rechter overschrijdingskans van een waarde W van $\underline{W}$ vindt men in een tabel van overschrijdingskansen voor de toets van WILCOXON als de linker overschrijdingskans van $U = \frac{-m^2 W + t_1 t_2}{2}$ en omgekeerd.

---

3) De toetsingsgrootheid $\underline{U}$ van de toets van WILCOXON is hierbij als volgt gedefiniëerd: stel twee steekproeven $x_1, x_2, \ldots x_{t_1}$ en $y_1, y_2, \ldots, y_{t_2}$ bestaan, tezamen genomen, uit de waarden $z_1 < z_2 < \ldots < z_k$; n_i is het aantal malen dat z_i optreedt en $\underline{a}_i$ is het aantal malen dat z_i optreedt in de eerste steekproef. Dan geldt, als $\underline{U}$ gedefiniëerd wordt als het aantal paren (x_i, y_j) met $x_i > y_j$ plus de helft van het aantal paren (x_i, y_j) met $x_i = y_j$:

$$2\underline{U} - t_1 t_2 = \sum_{i<j}\sum (\underline{a}_j n_i - \underline{a}_i n_j).$$

In het speciale geval dat $n_i = m$ voor iedere i wordt dit

$$2\underline{U} - t_1 t_2 = m \sum_{i<j}\sum (\underline{a}_j - \underline{a}_i) = -m^2 \underline{W}.$$

Literatuur

[1] van Eeden, Constance and J. Hemelrijk, A test for the equality of probabilties against a class of specified alternative hypotheses, including trend. Proc. Kon. Ned. Ak. v. Wet., A 58 (1955) en Indagationes Mathematicae 17 (1955), 191-198 en 301-308.

[2] van Eeden, Constance, Een toets tegen verloop van een aantal kansen. Zal verschijnen in de 9e jaargang (1955) van Statistica.

[3] Wabeke, Ir Doraline en Constance van Eeden, Handleiding voor de toets van WILCOXON, Rapport S 176 (M 65) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam, 1955.

[4] De toets van WILCOXON, Rapport S 47 (M 7) van de Statistische Afdeling van het Mathematisch Centrum.
