

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 121

Onderzoek naar de mogelijkheid van het gebruik van af-  
geknotte normale verdelingen bij een diagnoseprobleem.

door

Gerda Klerk-Grobbe

1953

## 1. Inleiding.

In een biochemisch laboratorium worden dagelijks series gehaltebepalingen verricht van verschillende stoffen in het bloed van patiënten. Een serie bestaat vaak uit 30 tot 50 waarnemingen. Bij elke stof wordt een grenswaarde aangenomen, waarboven het gehalte als abnormaal hoog beschouwd wordt. (Voor sommige stoffen zijn juist de lage waarden abnormaal, eenvoudigheidshalve zullen we steeds denken aan hoge abnormale waarden.) Een dergelijke extreme waarde wordt dan als gevolg van een ziekte gezien. Hierbij moet het volgende goed uit elkaar gehouden worden: Het gehalte van de desbetreffende stof zal over de groep van gezonden zowel als over de groep van zieken een waarschijnlijkheidsverdeling bezitten, waarvan redelijkerwijs verondersteld mag worden dat ze over elkaar vallen, dus b.v. als in figuur 1 is geschetst.

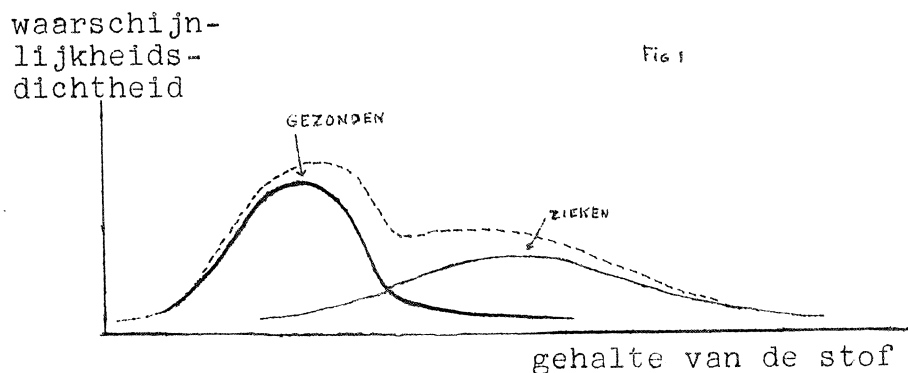


fig. 1 Waarschijnlijkheidsdichtheden der gehalten van een stof bij gezonden en bij zieken.

Over de groep van alle mensen (eventueel alle mensen die in de loop van zekere tijd in het ziekenhuis onderzocht worden) zal het gehalte van de stof een uit deze twee verdelingen samengestelde waarschijnlijkheidsverdeling bezitten, welke in fig. 1, afgezien van de schaal in verticale richting, door de stippellijn is voorgesteld. Nu wordt min of meer willekeurig een scheidingspunt aangebracht (b.v. op ervaringsgronden: boven dit scheidingspunt is een mens meestal ziek, er beneden gezond). De gehalten beneden dit punt worden als normaal en de gehalten boven dit punt worden als abnormaal beschouwd. In het volgende zullen de termen normaal <sup>1)</sup> en abnormaal steeds gebruikt worden voor gehal-

1) Men dient in dit rapport de term normaal (normale waarnemingen), welke een medisch waarde-oordeel inhoudt wel te onderscheiden van de term normaal in statistische zin, die steeds betrekking heeft op de normale waarschijnlijkheidsverdeling of de verdeling van Gauss.

ten boven of beneden het scheidingspunt, terwijl de waarden gezond en ziek op het bij fig. 1 besproken geval betrekking zullen hebben en dus zullen corresponderen met een betrouwbaar ander criterium over de gezondheidstoestand van de proefpersonen.

Wordt van de waarnemingen voor een stof over een lange tijd (b.v. een jaar) een grafiek vervaardigd, dan blijkt men bij de normale waarden van deze waarnemingen vaak goed een normale verdeling te kunnen aanpassen. De grafiek bezit echter aan de kant van de abnormale waarden vaak een te dikke staart in vergelijking met de normale verdelingskromme (zie b.v. fig. 3a, Na gehalte in serum, daarentegen geeft fig. 4a, Ca gehalte in serum, een ander beeld).

De vragen, die naar aanleiding van bovenbeschreven waarnemingen naar voren kwamen kunnen nu als volgt worden samengevat:

1) Is het mogelijk aan de normale waarnemingen een afgeknotte normale waarschijnlijkheidsverdeling aan te passen?

2) Is het mogelijk om op het gemiddelde van de normale waarnemingen voor een stof op één dag een controle op die bepalingen te baseren? En is aan te geven hoe groot het aantal waarnemingen per dag voor een dergelijke controle zou moeten zijn?

3) Indien we aannemen dat de gehalte-bepalingen van de gezonden een normale verdeling bezitten, met de parameters van de aan de normale waarnemingen aangepaste afgeknotte normale verdeling, is er dan iets te zeggen over de kans dat een gevonden waarde van een gezonde of een zieke afkomstig is.

In paragraaf 2 zal de aanpassing van een normale verdeling behandeld worden; paragraaf 3 is gewijd aan de verwachting en spreiding van het gemiddelde van een steekproef uit een afgeknotte normale verdeling, terwijl in paragraaf 4 boven geformuleerde vragen weer ter sprake zullen komen.

Notatie. Een stochastische grootte (d.i. een grootte met een waarschijnlijkheidsverdeling) zal door een onderstreepte letter worden aangegeven. Voor de verwachting (het theoretisch gemiddelde) van een stochastische grootte  $\underline{x}$  wordt het symbool  $\underline{\xi}$  gebruikt; de spreiding van  $\underline{x}$  wordt door  $\sigma(\underline{x})$  of  $\sigma_{\underline{x}}$  aangegeven. Een normale verdeling met gemiddelde  $\xi$  en spreiding  $\sigma$  wordt kort aangeduid met  $N(\xi, \sigma)$ .

2. Aanpassing van een afgeknotte normale verdeling.

a. Formules van A.HALD;  $\chi^2$ -toets voor aanpassing.

De bedoeling is dus een schatting te geven voor gemiddelde  $\xi$  en spreiding  $\sigma$  van de normale verdelingsfunctie uit een steekproef  $x_1, \dots, x_n$ , afkomstig uit slechts een deel van deze verdeling.

We onderstellen dat het afknottingspunt  $\xi_0$  is (rechtszijdig, zie fig. 2, uit de formules bij rechtszijdige afknotting volgen die voor linkszijdige eenvoudig).

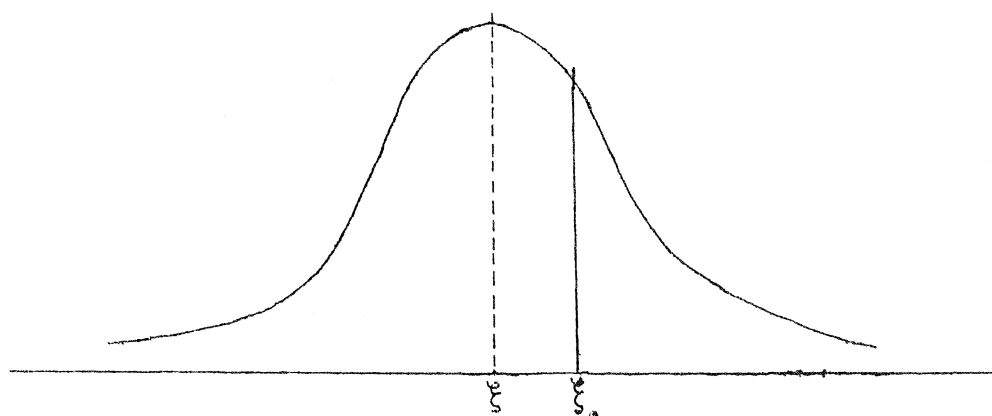


fig. 2 Rechtszijdig afgeknotte normale verdeling.

We noemen:

$$\begin{aligned} x'_1 &\stackrel{\text{DEF}}{=} x_1 - \xi_0 \\ &\vdots \\ x'_n &\stackrel{\text{DEF}}{=} x_n - \xi_0 \end{aligned}$$

De door A.HALD, [1] en [2], afgeleide meest aannemelijke schattingen (Engels: maximum likelihood estimates)  $\hat{\xi}$  en  $\hat{\sigma}$  voor  $\xi$  en  $\sigma$  worden nu als volgt gevonden:

Bereken:

$$Y \stackrel{\text{DEF}}{=} \frac{n \sum x_i'^2}{2(\sum x_i')^2}$$

waarmee, uit door HALD gegeven tabellen, de waarden van bepaalde functies

$$z \stackrel{\text{DEF}}{=} f(Y) \quad \text{en} \quad g(z)$$

kunnen worden afgelezen. De schattingen voor  $\xi$  en  $\sigma$  zijn dan:

$$\hat{\sigma} = \frac{-\sum x_i'}{n} g(z)$$

en

$$\hat{\xi} = \xi_0 + z \cdot \hat{\sigma}$$

Uit de laatste formule volgt, dat

$$\hat{\xi} = -z \hat{\sigma} = \frac{\xi_0 - \hat{\xi}}{\hat{\sigma}}$$

een schatting is voor het genormaliseerde afknottingspunt

$$\xi = \frac{\xi_0 - \xi}{\sigma}$$

Bij linkszijdige afknottung worden de schattingen:

$$\hat{\sigma} = \frac{\sum x_i}{n} g(z)$$

en

$$\hat{\xi} = \xi_0 - z \cdot \hat{\sigma}$$

De schatting voor  $\xi$  wordt in dit geval:

$$\hat{\xi} = z$$

Als asymptotische waarden voor de varianties (spreidingskwadraten) van  $\hat{\xi}$  en  $\hat{\sigma}$  (namelijk als  $n$ , de steekproefgrootte, onbegrensd zou stijgen) vindt HALD:

$$\sigma^2(\hat{\xi}) = \frac{\sigma^2}{n} \mu_{11}(\xi)$$

en

$$\sigma^2(\hat{\sigma}) = \frac{\sigma^2}{n} \mu_{22}(\xi)$$

waarin  $\mu_{11}(\xi)$  en  $\mu_{22}(\xi)$  eveneens door HALD getabelleerde functies zijn. Schattingen van deze varianties krijgen we door voor de onbekende parameters  $\sigma$  en  $\xi$  hun gevonden schattingen  $\hat{\sigma}$  en  $\hat{\xi}$  te substitueren.

Exacte uitdrukkingen voor de varianties van  $\hat{\xi}$  en  $\hat{\sigma}$  bij eindige steekproefgrootte  $n$  zijn ons niet bekend, zodat we met bovenstaande schattingen moeten volstaan.

P.M.GRUNDY, [3], geeft correcties op bovenstaande schattingen voor het geval van gegroepede waarnemingen. Hoewel de ons ter beschikking staande gegevens inderdaad uit gegroepede waarnemingen bestaan, hebben we toch de correcties achterwege gelaten, in verband met het vele rekenwerk en het oriënterende karakter van dit onderzoek.

Bij bovenbeschreven berekeningen wordt uitgegaan van de hypothese dat de waarnemingen afkomstig zijn uit een afgeknotte normale verdeling. Het is mogelijk deze hypothese statistisch te toetsen met de  $\chi^2$ -toets voor aanpassing (zie memorandum S 47 (M 6) en S 47 (M 12) welke aan het einde van dit verslag bijgevoegd zijn). Als intervallen zijn hierbij de intervallen waarin de waarnemingen in de gegevens gegroepeerd zijn gebruikt; aan de staart, waar het aantal waarnemingen per interval te klein wordt, zijn er steeds een aantal samengevoegd tot één dat dan

ongeveer 7 waarnemingen bevat. Het aantal vrijheidsgraden van de  $\chi^2$  is in ons geval het aantal intervallen verminderd met 3 (er zijn nl. 2 parameters uit de waarnemingen geschat).

Verder dient nog opgemerkt te worden dat deze  $\chi^2$ -toets voor aanpassing, vooral bij een klein aantal vrijheidsgraden, niet erg onderscheidend is, d.w.z. bij vrij grove afwijkingen tussen de werkelijkheid en de te toetsen hypothese is de kans dat men een  $\chi^2$  vindt in de kritieke zone <sup>2)</sup> corresponderende met een kleine onbetrouwbaarheid  $\alpha$ , bij toepassing van de toets, nog niet zeer groot.

b. Resultaten der berekeningen.

De in afdeling a. beschreven berekeningen zijn uitgevoerd voor de waarnemingen van een aantal stoffen. Daar ons niet bekend was, wat de, voor de verschillende stoffen gebruikelijke, grenswaarden tussen abnormaal en normaal was, hebben we de waarnemingen steeds afgeknot nà het interval met het hoogste aantal waarnemingen. Dit heeft bovendien het voordeel, dat wij alle stoffen eenzelfde statistische bewerking laten ondergaan zonder dat wellicht enigszins arbitraire medische abnormaliteit criteria hun invloed op de uitkomsten der statistische analyse doen gelden. Bovendien zijn bij elke stof dezelfde berekeningen nog eens uitgevoerd met het afknottingspunt nog één interval opgeschoven om een indruk te krijgen van de invloed hiervan op de schattingen. Deze bewerkingen kunnen desgewenst ook uitgevoerd worden met een op medische gronden bepaald afknottingspunt. Bij enkele stoffen kon alleen een grens opgegeven worden waarboven (of waarbeneden, aangegeven met  $>$  of  $<$ ) de waarden van  $\hat{\mu}$  en  $\hat{\sigma}$  moeten liggen, daar de tabellen van HALD niet verder reikten. Bij een deel van de stoffen is de  $\chi^2$ -toets voor aanpassing uitgevoerd, nl. in die gevallen waarin het aantal intervallen groot genoeg was en de schattingen voor  $\mu$  en  $\sigma$  precies berekend konden worden.

In tabel I zijn de resultaten van een en ander samengevat. De letter  $l$  resp.  $v$  achter het afknottingspunt betekent dat alle waarnemingen kleiner resp. groter dan het afknottingspunt werden weggelaten.

Bij de  $\chi^2$ -toets moest voor 7 van 9 stoffen de nulhypothese (waarnemingen verdeeld volgens een afgeknotte normale verdeling) verworpen worden (onbetrouwbaarheidsdrempel 0,05), wat bij een weinig onderscheidende toets wellicht verwondering kan wekken. Voor we tot een nadere analyse van dit resultaat overgaan, wil-

-----  
2) Zie memorandum S 47 (M 6), toegevoegd aan dit verslag.

Tabel I

Aanpassing van een afgeknotte normale verdeling (volgens A.HALD)

stof	no. fig.	afknottingspunt	$\hat{\xi}$	$\hat{\sigma}\{\hat{\xi}\}$	$\hat{\sigma}$	aanpassings- $\chi^2$	aantal vrijheidsgraden	overschrijdingskans k
eiwit in serum	-	7,0 (1)	7,52	0,135	0,88	4,88	3	0,10-0,20
	3	6,5 (1)	7,38	0,0899	0,92	4,52	4	0,30-0,50
Ca in serum	-	10,50 (r)	>15,7	-	>2,60	-	-	-
	4	11,00 (r)	11,93	0,608	1,88	45,7	7	<<0,001
ureum (< 50 j.)	-	35 (r)	26,40	0,524	6,58	-	-	-
	5	45 (r)	29,21	0,480	8,68	-	-	-
ureum (> 50 j.)	-	35 (r)	27,99	0,630	6,17	-	-	-
	6	45 (r)	32,40	0,566	8,48	-	-	-
ureum (alle leeft.)	-	35 (r)	27,71	0,465	6,19	-	-	-
	7	45 (r)	31,20	0,402	8,29	-	-	-
Na	-	335 (r)	>424	-	>44,59	-	-	-
	8	345 (r)	349	9,543	23,70	13,20	4	$\approx 0,01$
K	-	20,5 (r)	31,6	-	5,54	-	-	-
	9	21,5 (r)	23,9	2,101	3,98	84,9	5	<<0,001
alkali fosfaaten	-	8,5 (r)	6,82	0,368	2,67	12,0	5	0,02-0,05
	10	9,5 (r)	6,90	0,258	2,75	13,8	6	0,02-0,05
fosfaat	-	3,5 (r)	3,65	0,405	0,78	1,39	1	0,20-0,30
	11	4,0 (r)	3,53	0,138	0,75	3,02	2	0,20-0,30
cholesterine	-	175 (r)	173	7,453	33,38	3,91	6	0,50-0,75
	12	225 (r)	184	2,545	39,59	22,9	11	$\approx 0,020$
cholesterine ester	-	115 (r)	95,53	2,14	26,96	15,3	6	$\approx 0,020$
	13	125 (r)	96,50	1,88	27,67	16,3	7	0,01-0,02
alkali reserve	-	55,0 (1)	<16,96	-	>19,02	-	-	-
	14	50,0 (1)	24,10	1,88	18,90	33,9	4	<<0,01

len we er nog op wijzen dat de maximum likelihood schattingen in dit geval niet zodanig behoeven te zijn dat de aanpassings- $\chi^2$  minimaal wordt, zoals bij aanpassing van een niet-afgeknotte normale verdeling wel het geval zou zijn.

Voor verschillende stoffen zijn in de figuren 3a t/m 17a de frequentiekrommen getekend met het meest rechtse afknottingspunt (de figuren werden zo getekend, dat het afgeknotte deel van de frequentiekromme rechts ligt) en de schatting van het ge-

middelde die hieruit volgt. Bij de figuren 4a, 8a, 9a en 14a zien we dat de kleine overschrijdingskansen geweten moeten worden aan de onbevredigende schatting  $\hat{\xi}$  van het gemiddelde, welke bij deze stoffen zelfs rechts van het afknottingspunt komen te liggen. De voornaamste oorzaak voor deze waarden van  $\hat{\xi}$  moet gezocht worden in de te dikke staart van het linker-deel der frequentiekromme. Een bevestiging hiervoor leveren berekeningen bij Na en K (figuur 8a en 9a), waarbij een nieuwe schatting  $\hat{\xi}'$  van  $\xi$  werd bepaald door weglating van de meest links gelegen waarneming (bij Na: 200, bij K: 8); door weglating van deze ene waarneming wordt een veel lagere schatting voor  $\xi$  gevonden.

De overige gevallen met kleine overschrijdingskansen, figuren 10a, 12a, 13a (en ook 9a en 14a enigszins) vertonen in hun frequentiekromme een "deuk" welke voor de slechts aanpassing aansprakelijk gesteld moet worden. In hoeverre deze "deuken" een reële betekenis bezitten, dan wel een gevolg zijn van de waarnemingstechniek (waarnemingsfouten, bij elkaar nemen van verschillende groepen patiënten, enz.) kan door ons niet worden beoordeeld.

Teneinde nog een iets beter inzicht te verkrijgen zijn de waargenomen frequenties uitgezet op waarschijnlijkheidspapier, waarbij als parameters van de normale verdeling de schattingen  $\hat{\xi}$  en  $\hat{\sigma}$  uit tabel I gebruikt zijn.

Op de techniek die bij die uitzetten gebruikt is, gaan we hier niet in; ze is niet moeilijk, wel bewerkelijk. Bij normale verdelingen behoren de uitgezette punten op een rechte lijn te liggen. Vergelijken we nu de gevonden resultaten met het boven besprokene dan valt het op dat voor die stoffen, waarbij de te dikke staart aansprakelijk gesteld moest worden voor de slechte schatting van  $\xi$  en de kleine overschrijdingskans, ook de punten op het waarschijnlijkheidspapier een afwijking van normaliteit aanwijzen (figuren 4b, 8b, 9b en 14b). De stoffen met "deuken" in hun frequentiekromme vertonen daarentegen veel minder afwijkingen van normaliteit volgens het waarschijnlijkheidspapier.

Ureum dat niet met de  $\chi^2$ -toets onderzocht kon worden door het te kleine aantal intervallen, vertoont volgens het wh-papier geen grote afwijkingen van normaliteit.

Ten slotte merken wij nog op dat bij die stoffen, welke geen aanleiding geven tot verwerping van de hypothese van normaliteit het aantal intervallen ook slechts klein is (Ca, ureum, fosfaat), zodat het gevaarlijk blijft zonder verder onderzoek



normaliteit te veronderstellen.

Samenvattend komen we tot de conclusie dat het sterke aanbeveling verdient voor men bij verdere theoriën gebruik gaat maken van normaliteit, dit eerst aan nieuw en zo mogelijk uitgebreider (d.w.z. meer waarnemingen en meer intervallen) materiaal te onderzoeken. Het grotere aantal intervallen is vooral van belang in verband met het kleine onderscheidend vermogen van de  $\chi^2$ -toets bij een laag aantal vrijheidsgraden.

3. Verwachting en spreiding van het steekproefgemiddelde bij een afgeknotte verdeling.

a. Een afgeknotte bekende normale verdeling.

Zijn van de oorspronkelijke normale verdeling gemiddelde  $\xi$  en spreiding  $\sigma$  bekend, dan is het niet moeilijk verwachting en spreiding van het gemiddelde van een steekproef  $x_1, \dots, x_n$  uit een deel van deze verdeling te bepalen.

We onderstellen dat de afknotting rechts van  $\xi_0$  plaats vindt (zie fig. 2) en noemen

$$\xi_0 \stackrel{\text{DEF}}{=} \frac{\xi_0 - \xi}{\sigma}$$

De verdelingsfunctie van de afgeknotte normale verdeling is:

$$F(x) = P[X \leq x] = \frac{\phi\left(\frac{x - \xi}{\sigma}\right)}{\phi(\xi_0)}, \quad x \leq \xi_0,$$

waarin

$$\phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

Voor de verdelingsdichtheid volgt hieruit:

$$\begin{cases} f(x) = \frac{1}{\phi(\xi_0)\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\xi)^2} & , x < \xi_0 \\ f(x) = 0 & \text{als } x > \xi_0 \end{cases}$$

De verwachting van het steekproefgemiddelde is dan:

$$E\left(\frac{1}{n} \sum x_i\right) = E x_i = \int_{-\infty}^{\xi_0} x f(x) dx = -\frac{\sigma \phi(\xi_0)}{\phi(\xi)} + \xi,$$

waarin

$$\phi(\xi_0) = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}\xi_0^2}$$

De variantie van het steekproefgemiddelde is

$$\begin{aligned} \sigma^2\left\{\frac{1}{n}\sum x_i\right\} &= \frac{1}{n}\sigma^2\{x_i\} = \frac{1}{n}\left[\sum x^2 - (\sum x)^2\right] = \\ &= \frac{\sigma^2}{n}\left[1 - \sum \frac{\varphi(\xi)}{\phi(\xi)} - \left(\frac{\varphi(\xi)}{\phi(\xi)}\right)^2\right] \end{aligned}$$

Voor een  $N(0,1)$  -verdeling zijn in tabel II enkele waarden van  $E\bar{x}$  en  $\sigma^2\{\bar{x}\}$  gegeven (we noemen  $\bar{x} \stackrel{\text{DEF}}{=} \frac{1}{n}\sum_{i=1}^n x_i$ ), nl. voor de afknottingspunten  $\xi = 0$ ; 0,5 en 1;  $\sigma^2(\bar{x})$  is afhankelijk van de steekproefgrootte  $n$ , als voorbeelden zijn 30, 40, 50 en 60 waarnemingen genomen.

Tabel II

Verwachting en spreiding van het steekproefgemiddelde bij een afgeknotte  $N(0,1)$  -verdeling

afknotting $\xi$	$E\bar{x}$	$\sigma\{\bar{x}\}$			
		30 waarn.	40 waarn.	50 waarn.	60 waarn.
0	- 0,778	0,1101	0,0953	0,0853	0,0778
0,5	- 0,509	0,1273	0,1102	0,0986	0,0900
1	- 0,288	0,1448	0,1254	0,1122	0,1024

Was de oorspronkelijke verdeling een  $N(\xi, \sigma)$  -verdeling, dan wordt de verwachting  $E\bar{x}'$  van het gemiddelde van een steekproef  $x'_1, \dots, x'_n$  uit de bij  $\xi_0$  afgeknotte verdeling gegeven door

$$E\bar{x}' = \sigma \cdot E\bar{x} + \xi$$

waarin  $E\bar{x}$  de bij  $\xi = \frac{\xi_0 - \xi}{\sigma}$  behorende waarde uit tabel II is. De spreiding van het steekproefgemiddelde is in dit geval:

$$\sigma\{\bar{x}'\} = \sigma \cdot \sigma\{\bar{x}\}$$

b. Een afgeknotte onbekende verdeling met bekend punt van afknotting.

In het bovenbesproken geval is ondersteld dat de oorspronkelijke verdeling normaal was met bekende parameters ( $\xi$  en  $\sigma$ ). Mogen we nu wel onderstellen dat de oorspronkelijke verdeling normaal is, maar kennen we gemiddelde en spreiding niet, dan kan het bovenstaande slechts als benadering of ter oriëntatie dienen, waarbij we voor  $\xi$  en  $\sigma$  de uit groot materiaal volgens paragraaf 2 geschatte waarden nemen. In verband met het stochastische karakter van deze schattingen zullen we slechts aan stochastische afwijkingen onderhevige schattingen van verwachting en spreiding  $E\bar{x}$  en  $\sigma\{\bar{x}\}$  verkrijgen. Met name is er gevaar dat de spreiding  $\sigma\{\bar{x}\}$  van het steekproefgemiddelde onderschat wordt.

Is ook de onderstelling van normaliteit niet te handhaven, dan is er nog een weg om tenminste schattingen voor verwachting

en spreiding te verkrijgen. Namelijk door een serie steekproeven van de gebruikelijke grootte door middel van herhaalde loting uit het waarnemingsmateriaal te nemen en uit de serie zo verkregen steekproefgemiddelden het gemiddelde en de spreiding te schatten. Deze methode heeft alleen dan een goede kans van slagen, indien het waarnemingsmateriaal van niet te kleine omvang is.

Zijn  $\{\bar{x}'\}$  en  $\sigma\{\bar{x}'\}$  bepaald dan kunnen

$$\{\bar{x}' - 2\sigma\{\bar{x}'\}\} \text{ en } \{\bar{x}' + 2\sigma\{\bar{x}'\}\}$$

gebruikt worden als controlegrenzen voor toekomstige steekproefgemiddelden, waarbij we er echter op moeten letten dat de steekproeven steeds dezelfde omvang dienen te hebben. Is

de omvang van de steekproeven waaruit  $\{\bar{x}'\}$  en  $\sigma\{\bar{x}'\}$  bepaald zijn en is  $m$  de omvang van de steekproef die we willen controleren, dan moeten we als grenzen gebruiken:

$$\{\bar{x}' \pm 2\sqrt{\frac{m}{n}} \sigma\{\bar{x}'\}\}$$

#### 4. Bespreking van de vragen (zie paragraaf 1).

1) Uit paragraaf 2 bleek dat de onderstelling dat de normale waarnemingen normaal verdeeld zijn voor vele stoffen weinig aannemelijk is. Ook voor die stoffen waarvoor de  $\chi^2$ -aanpassingstoets geen aanwijzing voor afwijkingen van normaliteit gaf, dient men toch voorzichtig te zijn met het trekken van conclusies hieruit, in verband met het kleine onderscheidingsvermogen van de  $\chi^2$ -toets bij een laag aantal vrijheidsgraden.

Een meer uitgebreid onderzoek zowel wat betreft het aantal waarnemingen als het aantal intervallen waarin de waarnemingen worden ondergebracht is dan ook zeer aan te bevelen.

2) In paragraaf 3 is aangegeven hoe controlegrenzen bepaald kunnen worden zowel wanneer wel als wanneer niet normaliteit ondersteld wordt. Deze grenzen kunnen dus dienen als controle op de per dag binnenkomende waarnemingen. (D.w.z. een controle op bepalingsfouten; hierbij wordt dus ondersteld dat de samenstelling van de groep proefpersonen, wat betreft lijders aan de verschillende ziekten en gezonden, niet van dag tot dag sterk wisselt.) Zijn er geen andere foutenbronnen aanwezig, dan zal in de loop van een lange serie slechts ongeveer 5% van de steekproefgemiddelden buiten deze grenzen liggen. Deze 5% is slechts een ruwe benadering, daar het steekproefgemiddelde slechts bij

ruwe benadering een normale verdeling zal bezitten. De waarde van deze controle-methode t.o.v. de tot nu toe gebruikte (extra bepalingen met paardenserum waarin de gehalten der verschillende stoffen bekend zijn) is te onderzoeken door bij een groot aantal dagen na te gaan of "goede" en "slechte" dagen volgens de oude methode ook corresponderen met daggemiddelden binnen resp. buiten de grenzen bij de nieuwe methode. Het aantal waarnemingen dat nodig is per dag voor bovenbeschreven methode hangt af van het risico dat de onderzoeker wil lopen om een "slechte" dag als "goed" te laten passeren. Hoe groter het aantal waarnemingen, hoe nauwer de grenzen, dus des te kleiner het risico. Schattingen voor deze aantallen met hun bijbehorend risico zijn zeer moeilijk te berekenen. Ze kunnen vermoedelijk beter empirisch bepaald worden, vooral als toch beide controle-systemen enige tijd naast elkaar worden toegepast. Ook kunnen opzettelijk foutieve proeven (dus met opzettelijk aangebrachte fouten bij de bepaling) hierin een aanwijzing geven.

3) Bij de beantwoording van deze vraag dienen we goed het in de inleiding gemaakte terminologisch onderscheid tussen normaal en abnormaal enerzijds en ziek en gezond anderzijds voor ogen te houden.

We onderstellen nu dat de gehalten bij gezonden een normale verdeling bezitten met de parameters van de afgeknotte verdeling der normale waarnemingen (daarbij houden we er dus geen rekening mee dat sommige gehaltebepalingen beneden het scheidingspunt afkomstig kunnen zijn van zieken). We kunnen dan in een tabel van de normale verdeling bij ieder gevonden gehalte de kans opzoeken, dat men bij een gezond persoon dit of een groter gehalte zal vinden. Op grond hiervan zou het dus mogelijk zijn een grens, A, te bepalen welke bij een gegeven kleine kans (b.v. 5%) behoort, zodanig dat men proefpersonen met een gehalte  $> A$  in ieder geval als ziek gaat behandelen.

De hiermee analoge kans dat men bij een zieke het gevonden of een lager gehalte zal vinden kan alleen bepaalde worden indien ook de verdeling der gehalten van zieken bekend zou zijn. Is deze bekend, dan kan men bij een gegeven kans (welke een andere waarde mag hebben dan de boven gebruikte, b.v. 1%), volkomen analoog aan grens A, een tweede grens, B, bepalen, zodanig dat men proefpersonen met een gehalte  $< B$  als gezond beschouwt.

De twee zo gevonden grenzen zullen in het algemeen niet samenvallen (ze zijn immers volkomen onafhankelijk van elkaar bepaald). De twee mogelijke situaties zijn dus:  $B > A$  en  $B < A$

(zie fig. 15)

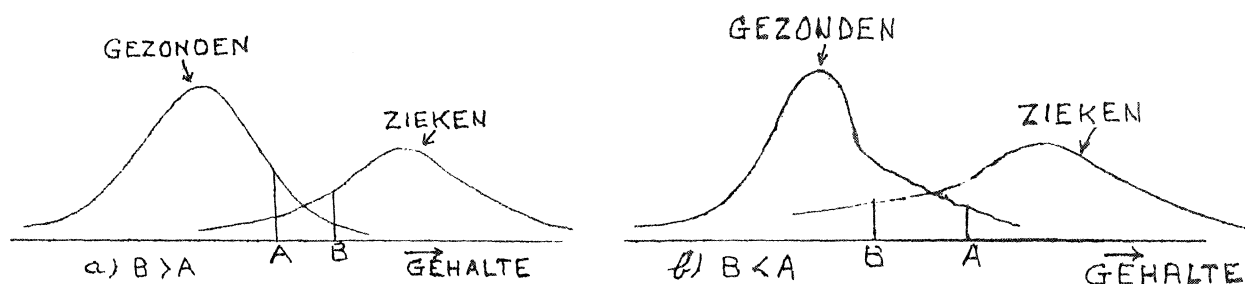


fig. 15 Onderlinge ligging van de grenzen B en A (zie tekst).

$B > A$  Dit zal optreden, indien de verdelingen van gezonden en zieken voldoende ver uit elkaar liggen. Door in plaats van de grenzen A en B één grens tussen beide in te kiezen, worden de boven beschreven kansen nog verkleind en de diagnose vereenvoudigd.

$B < A$  In dit geval zijn de gehaltebepalingen dus in drie groepen verdeeld. Een mogelijke gedragslijn is nu de volgende:

1) gehalten  $\leq B$ ; de onderzochte persoon wordt als gezond beschouwd, daar de kans bij zieken op een gehalte in dit gebied slechts gering is (1%).

2) gehalten  $\geq A$ ; deze groep wordt zonder meer als ziek beschouwd, daar de kans bij gezonden op een gehalte  $\geq A$  eveneens slechts klein is (5%).

3) Gehalten tussen B en A; deze groep moet nader onderzocht worden (nieuwe gehaltebepaling of andere proeven) voor de diagnose gesteld kan worden.

De bovenbeschreven kansen zeggen echter nog niets over de kans dat een gevonden gehalte afkomstig is van een gezonde. Hiervoor zou nl. ook de verhouding van de aantallen gezonden en zieken bekend moeten zijn (ziek en gezond dus bepaald volgens andere criteria dan het al of niet boven een bepaalde grens liggen van het desbetreffende gehalte).

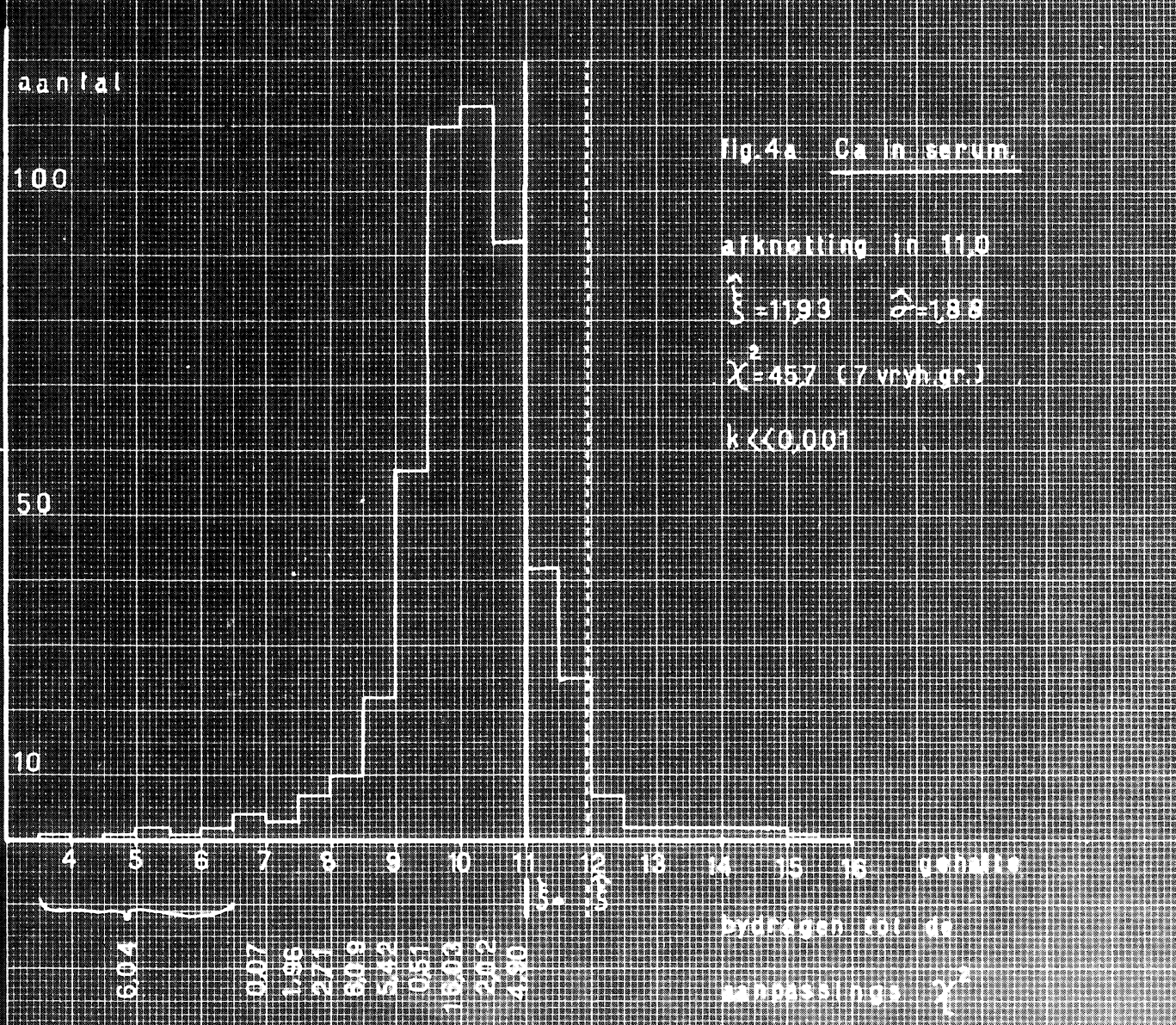
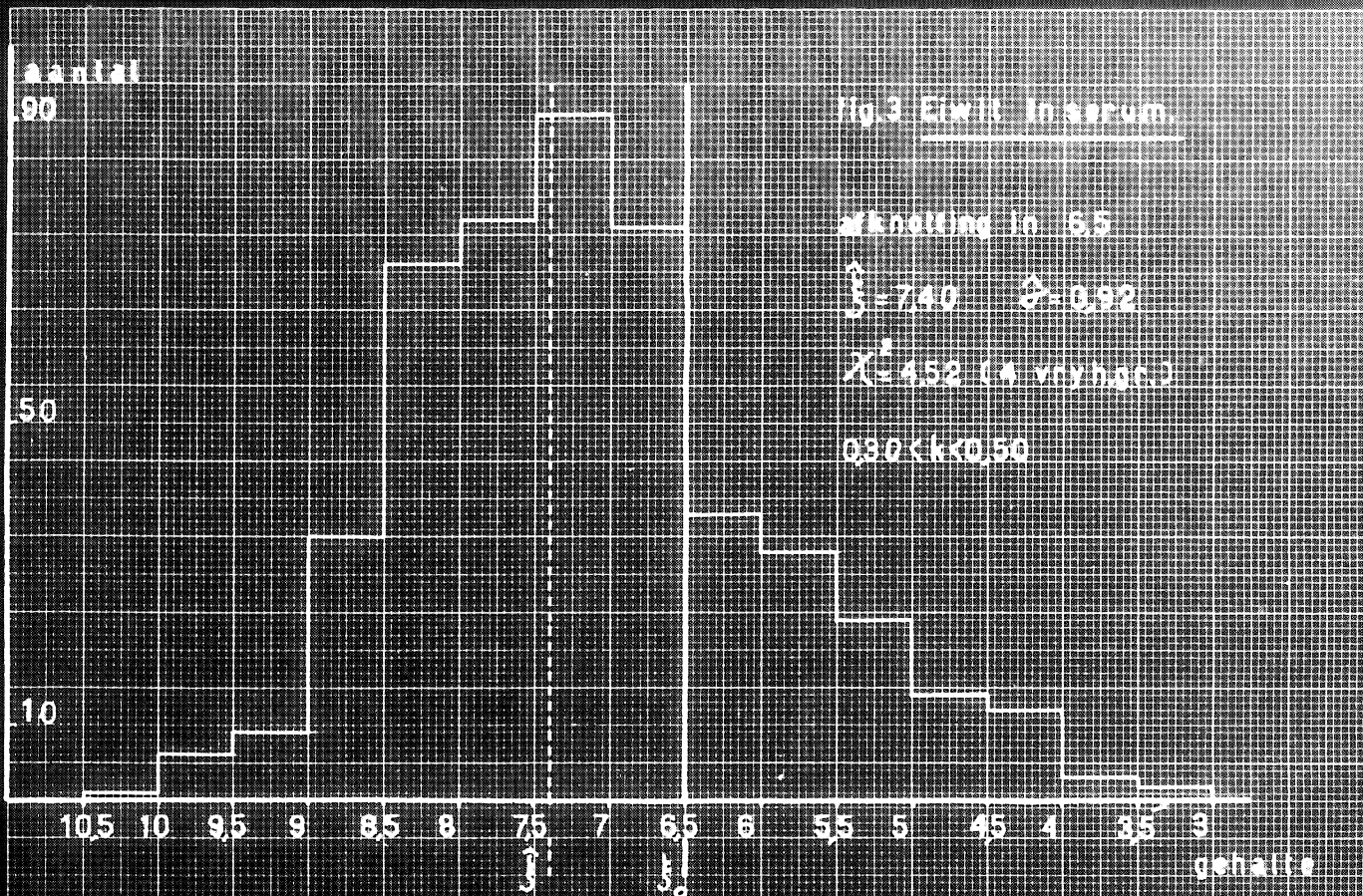
Een opmerking is hierbij nog van groot belang: de verhoudingen van de aantallen gezonden en zieken zullen voor de verschillende onderzoekcentra sterk uiteen kunnen lopen. De in één centrum, op grond van een daar gevonden verhouding, bepaalde "kans op gezond" bij een bepaald gehalte, zou dan ook slechts voor dit centrum geldig zijn.

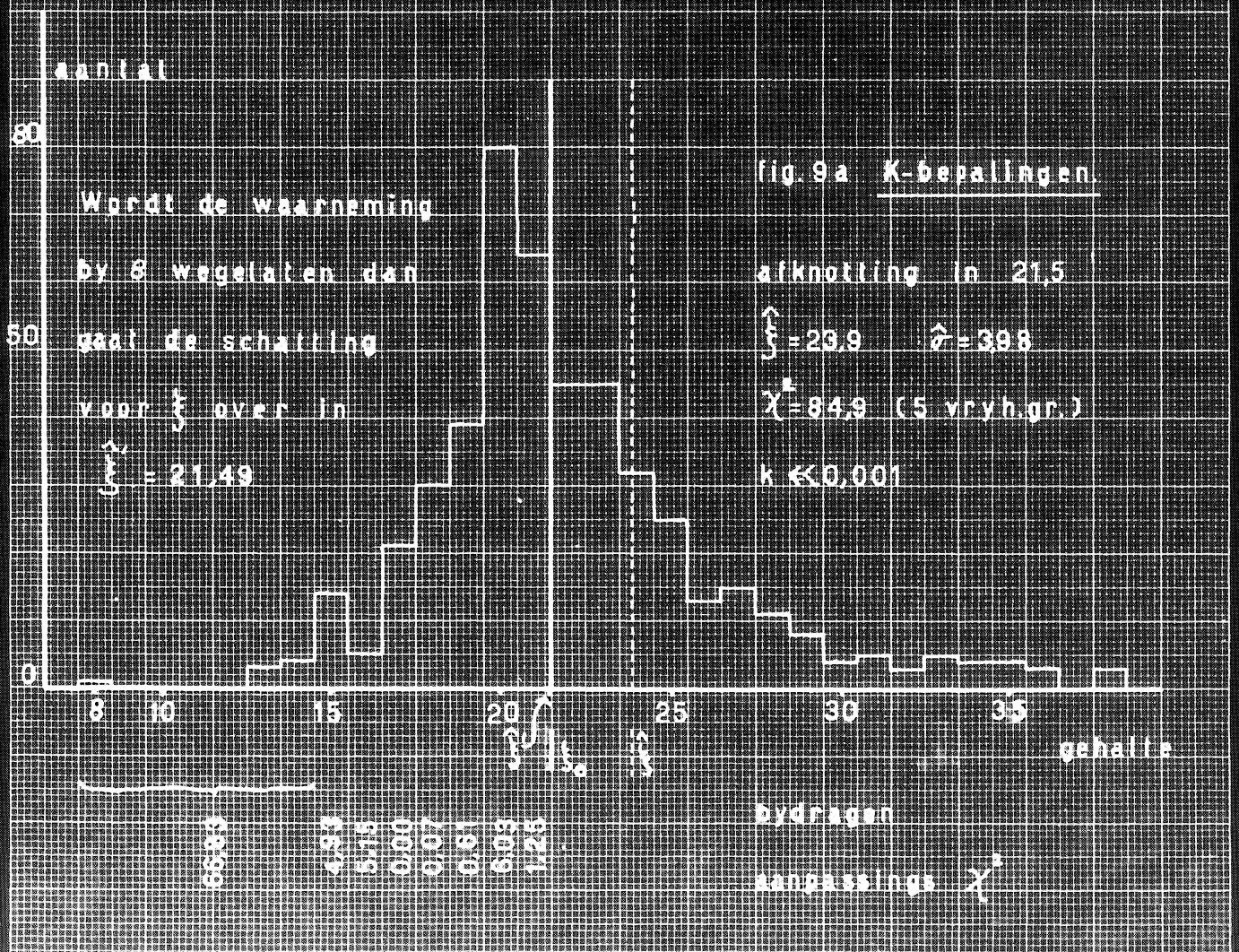
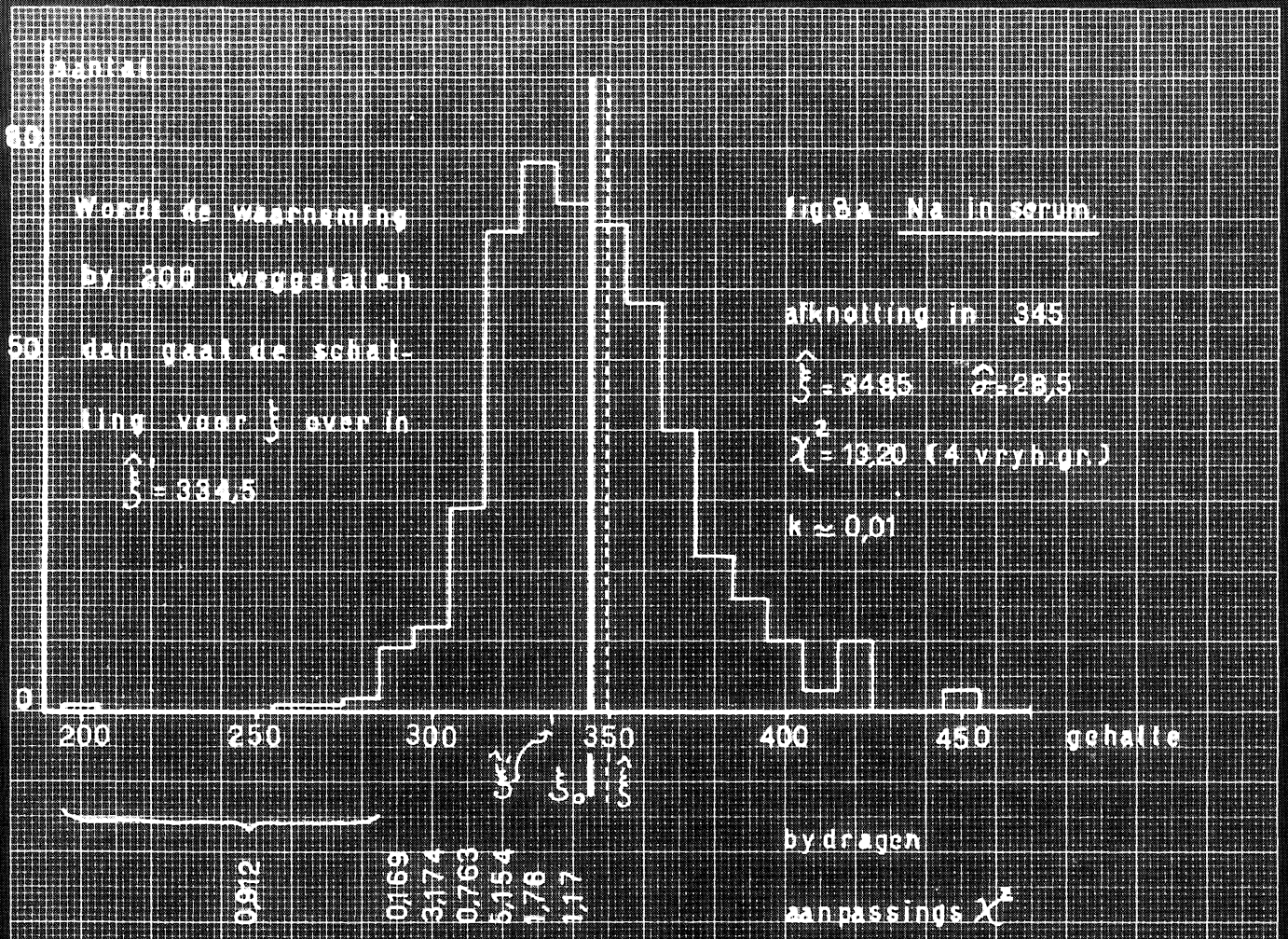
Bovendien zou er rekening gehouden moeten worden met een mogelijke verschuiving van deze verhouding in de loop van zeke-

re tijd bij elk centrum van onderzoek. Het is daarom vermoedelijk beter deze kwestie te laten rusten en het boven beschreven diagnoseschema, voor zover dat voor realisering vatbaar is, te gebruiken.

Literatuur

- [1] A.Hald, Statistical theory with engineering applications, text with tables, Wiley, New York (1952), pp. 144-150, 245, 316.
- [2] A.Hald, Maximum likelihood estimations of the parameters of a normal distribution which is truncated at a known point, Skand. Aktuarietidskrift (1949), pp. 119-134.
- [3] P.M.Grundy, The fitting of grouped truncated and grouped censored normal distributions, Biometrika, 39 (1952), pp. 252-259.







20 gr/min

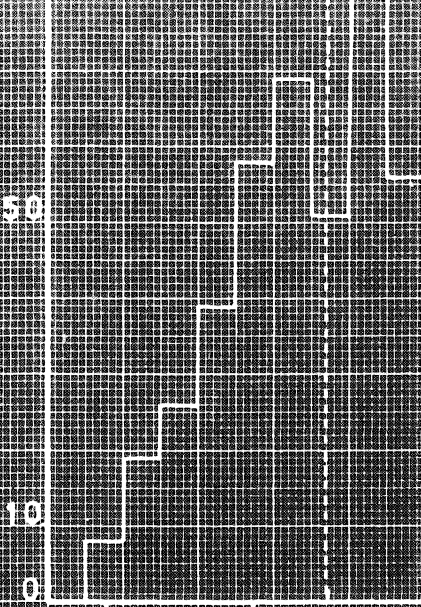


Fig. 10a Arkali in serum

afknotting in 9,5

$$\hat{\xi} = 6,90 \quad \hat{\sigma} = 2,75$$

$$\chi^2 = 13,8 \text{ (6 vrijh.gr.)}$$

$$0,02 < k < 0,05$$

10 15 20 25  $\xi_0$  gehalte

2,18  
1,34  
0,00  
0,13  
0,15  
0,13  
5,97  
3,78  
0,08

bydragen aanpassings  $\chi^2$

aantal

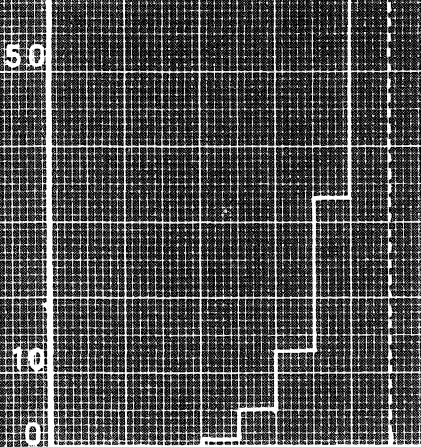


Fig. 11a Fosfaat in serum

afknotting in 4,0

$$\hat{\xi} = 3,53 \quad \hat{\sigma} = 0,75$$

$$\chi^2 = 3,02 \text{ (2 vrijh.gr.)}$$

$$0,20 < k < 0,30$$

1 2 3 4 5 6 7 8  $\xi_0$  gehalte

afknotting

30

10

100 150 200 250 300 350 400 450 gehalte

0,45 0,35 0,25 0,15 0,05 0,01 0,005 0,001 0,0005 0,0001 0,00005 0,00001 0,000005 0,000001 0,0000005 0,0000001 0,00000005 0,00000001

Fig. 12a Cholesteroline

afknotting in 225

$\hat{\xi} = 184 \quad \hat{\sigma} = 39,6$

$\chi^2 = 22,9$  (11 vryh.gr.)

$k < 0,016$

bydragen aanpassings  $\chi^2$

mental

100

50

75 105 155 205 gehalte

0,05 0,02 0,01 0,005 0,002 0,001 0,0005 0,0002 0,0001 0,00005 0,00002 0,00001 0,000005 0,000002 0,000001 0,0000005 0,0000002 0,0000001

Fig. 13a Cholesteroline-ester.

afknotting in 125

$\hat{\xi} = 96,5 \quad \hat{\sigma} = 27,67$

$\chi^2 = 16,3$  (7 vryh.gr.)

$0,01 < k < 0,02$

bydragen aanpassings  $\chi^2$

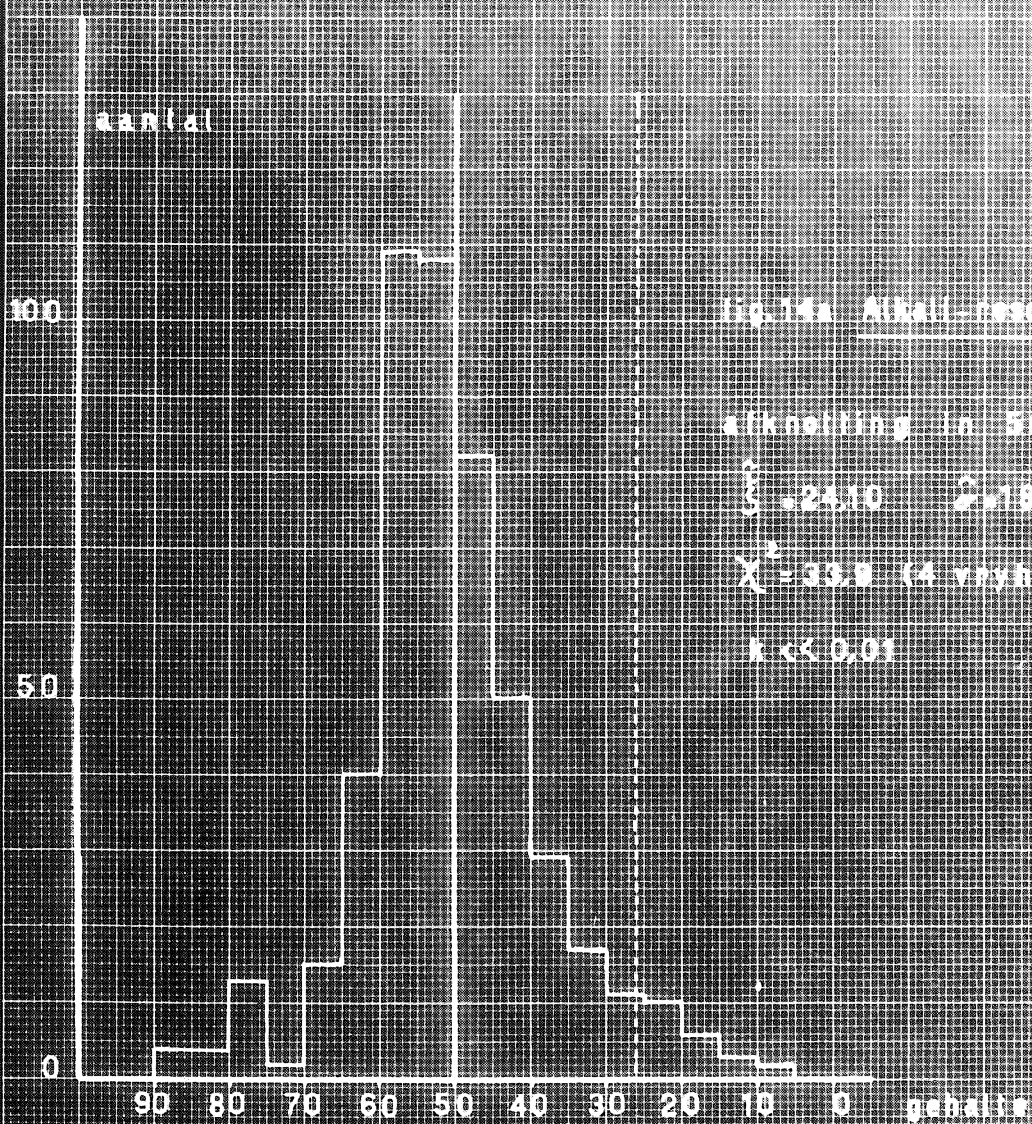
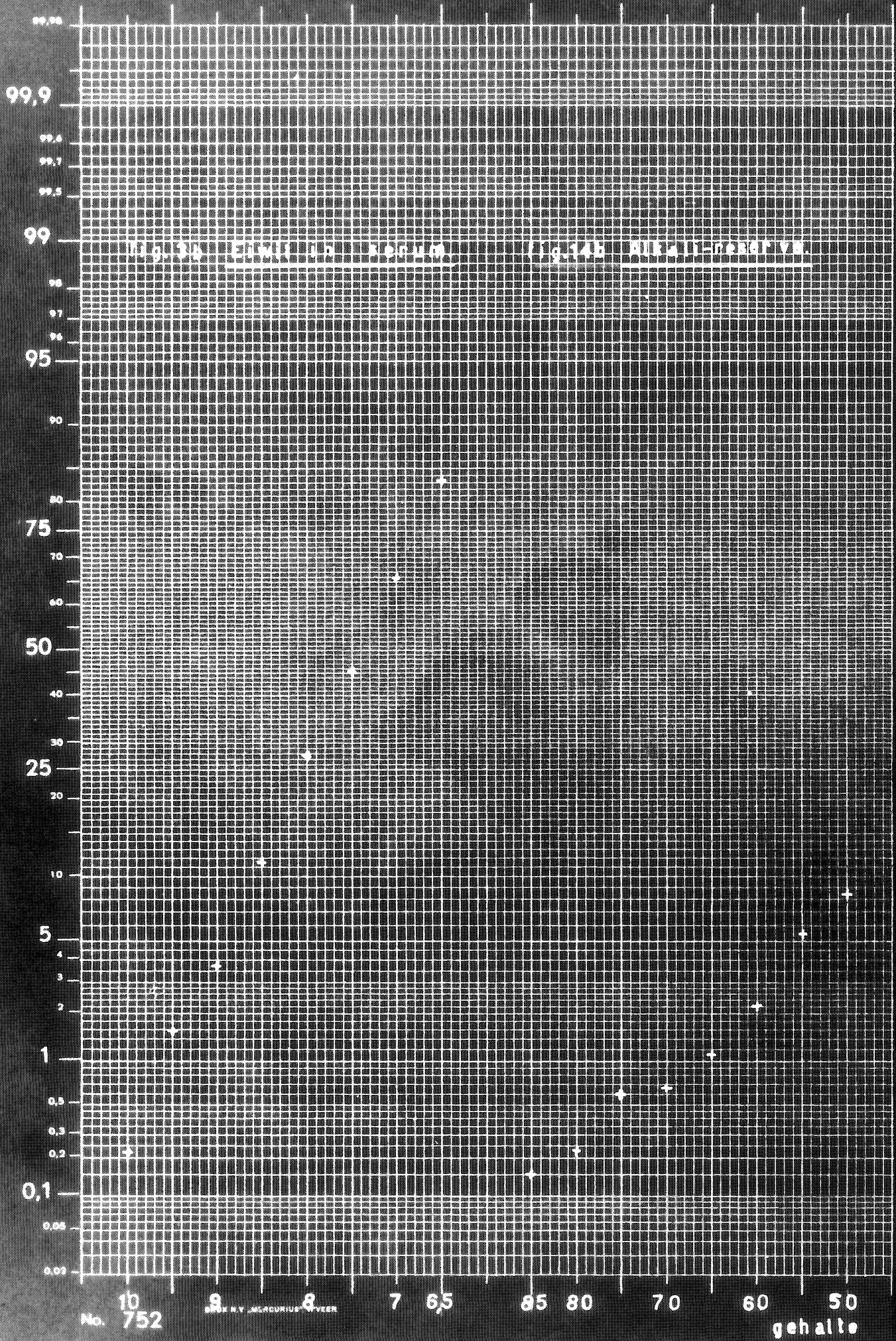


Fig. 104. Alkali reserve

$\xi$  - knowning in  $0,01$   
 $\xi$  -  $0,2480$  -  $2,4801$   
 $X$  -  $49,9$  (at variance)  
 $n < 0,01$

1,36  
 4,28  
 9,86  
 5,64  
 0,79  
 11,34  
 0,62

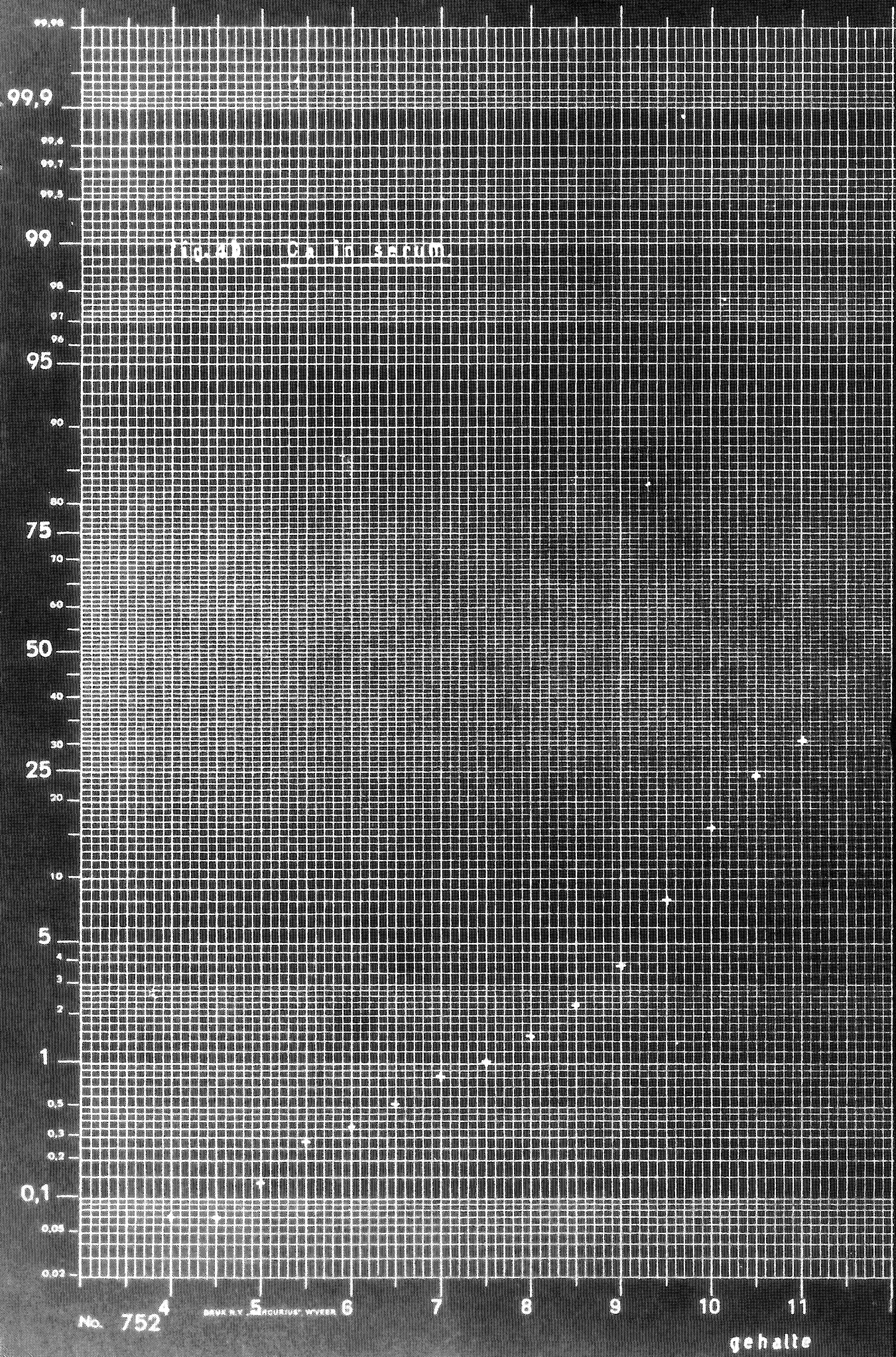
by dragon -  $n$  -  $0,01$   $X$



No. 752

BY N.Y. ALGACURIUS VEER

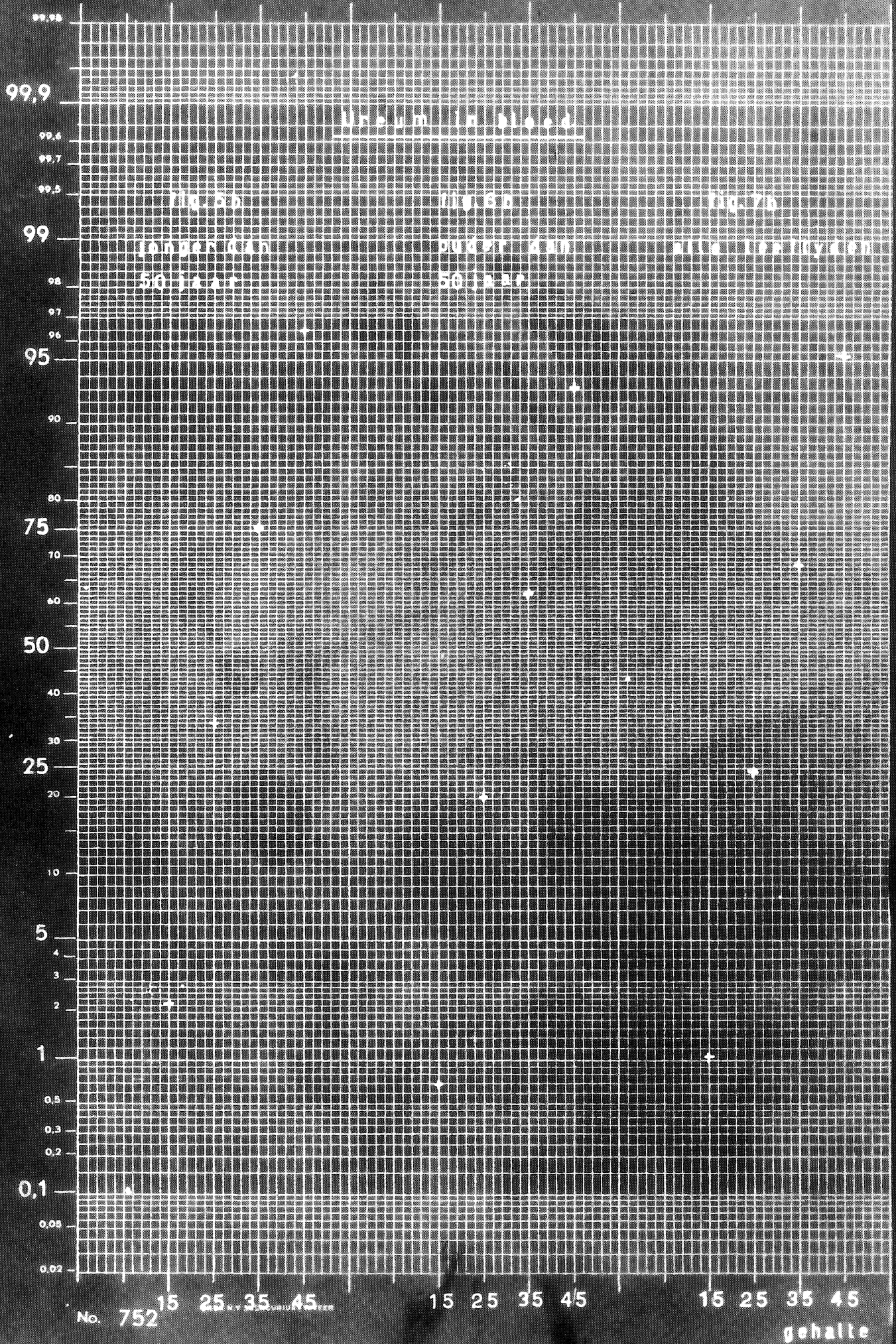
gehalte

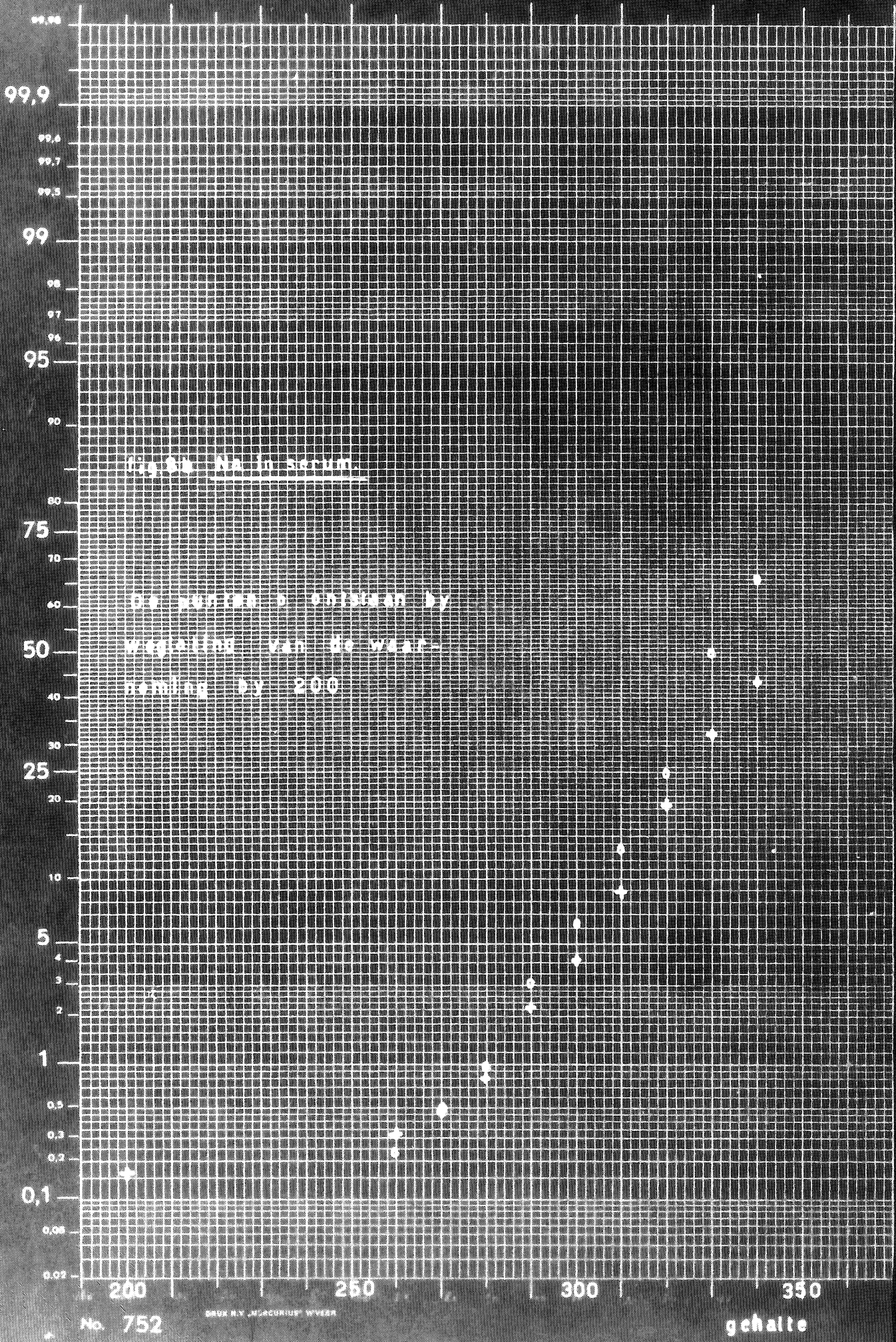


No. 752

BRUNNEN WYKON

gehalte

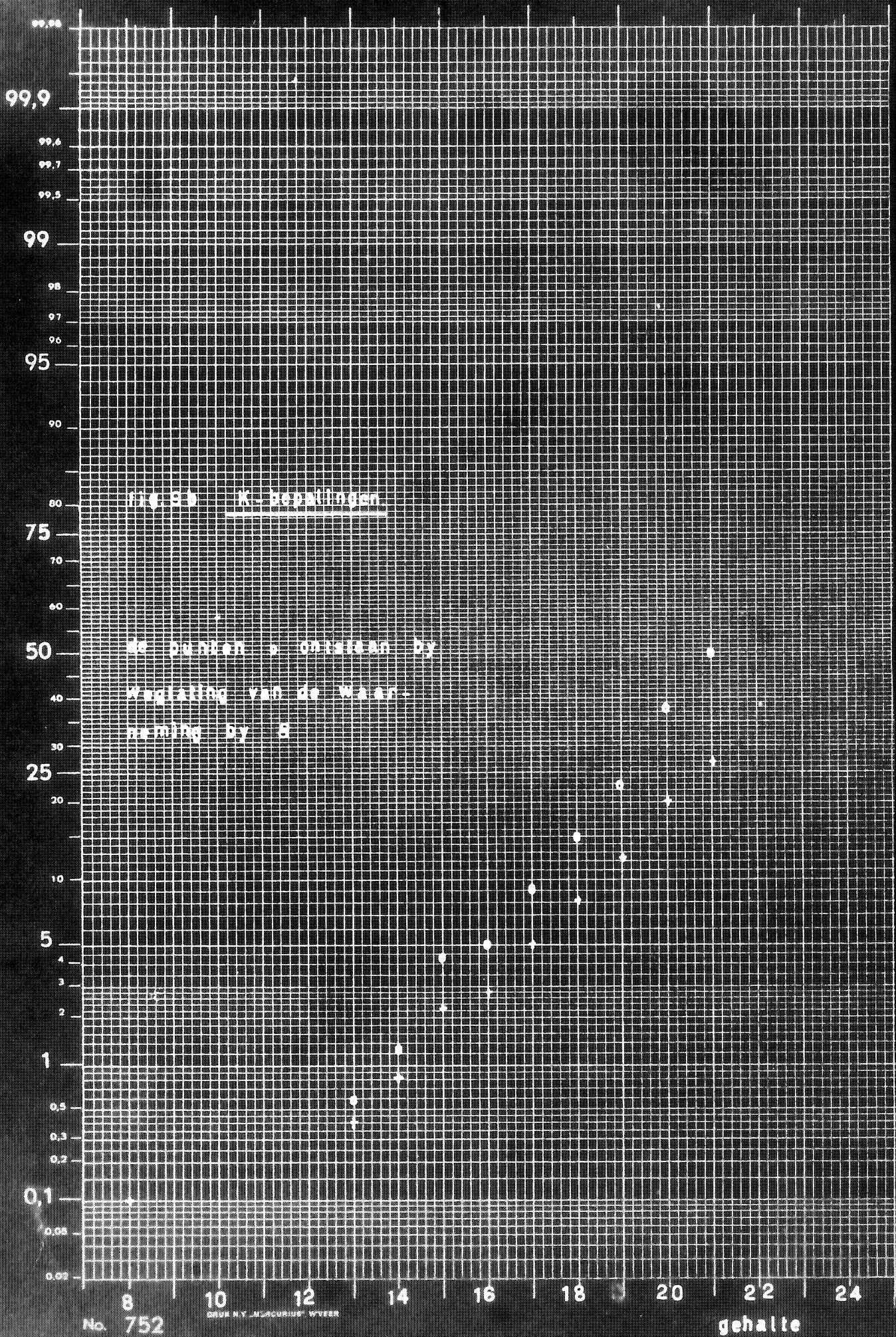




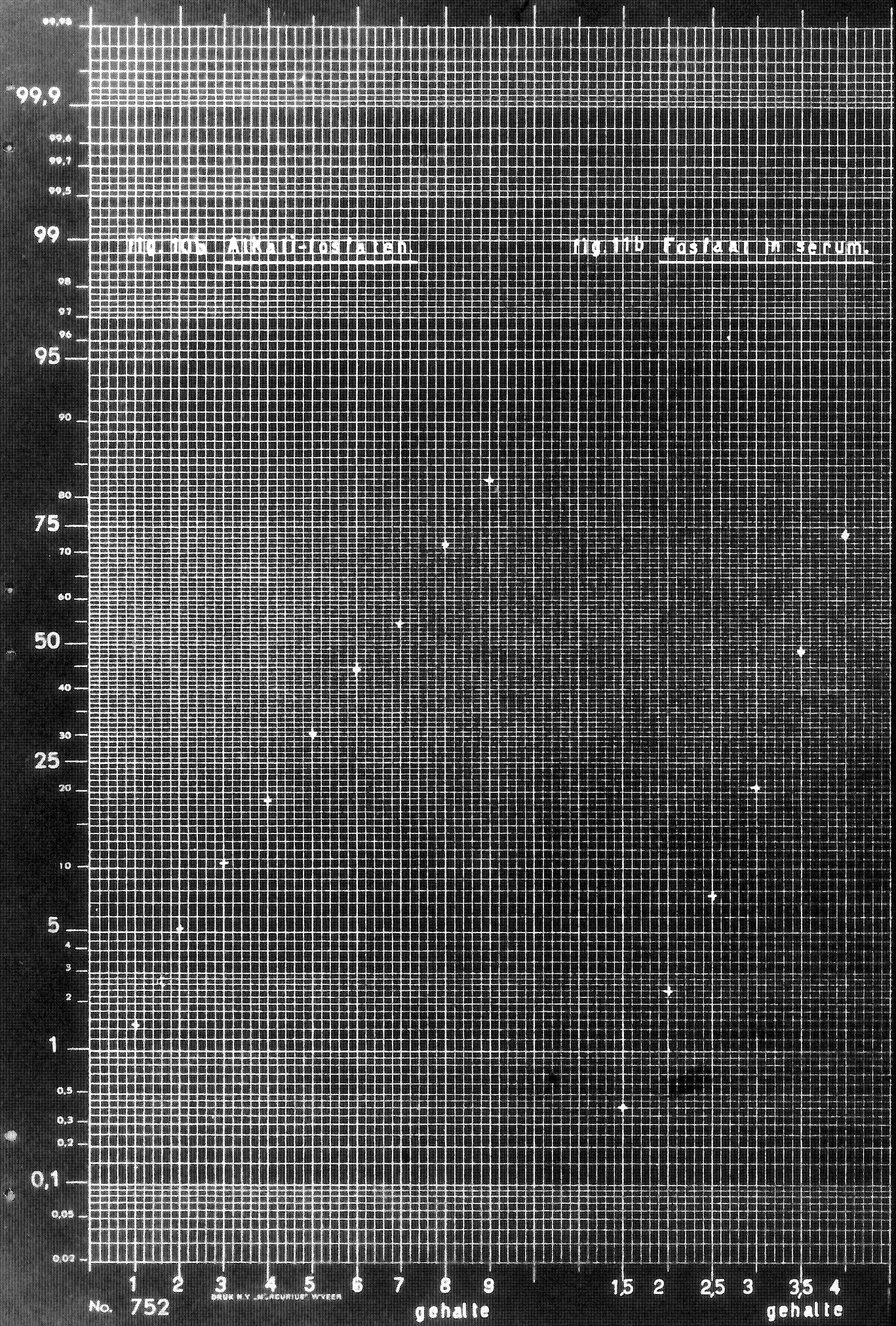
No. 752

BRUNNEN MASCHINEN WERK

gehalte

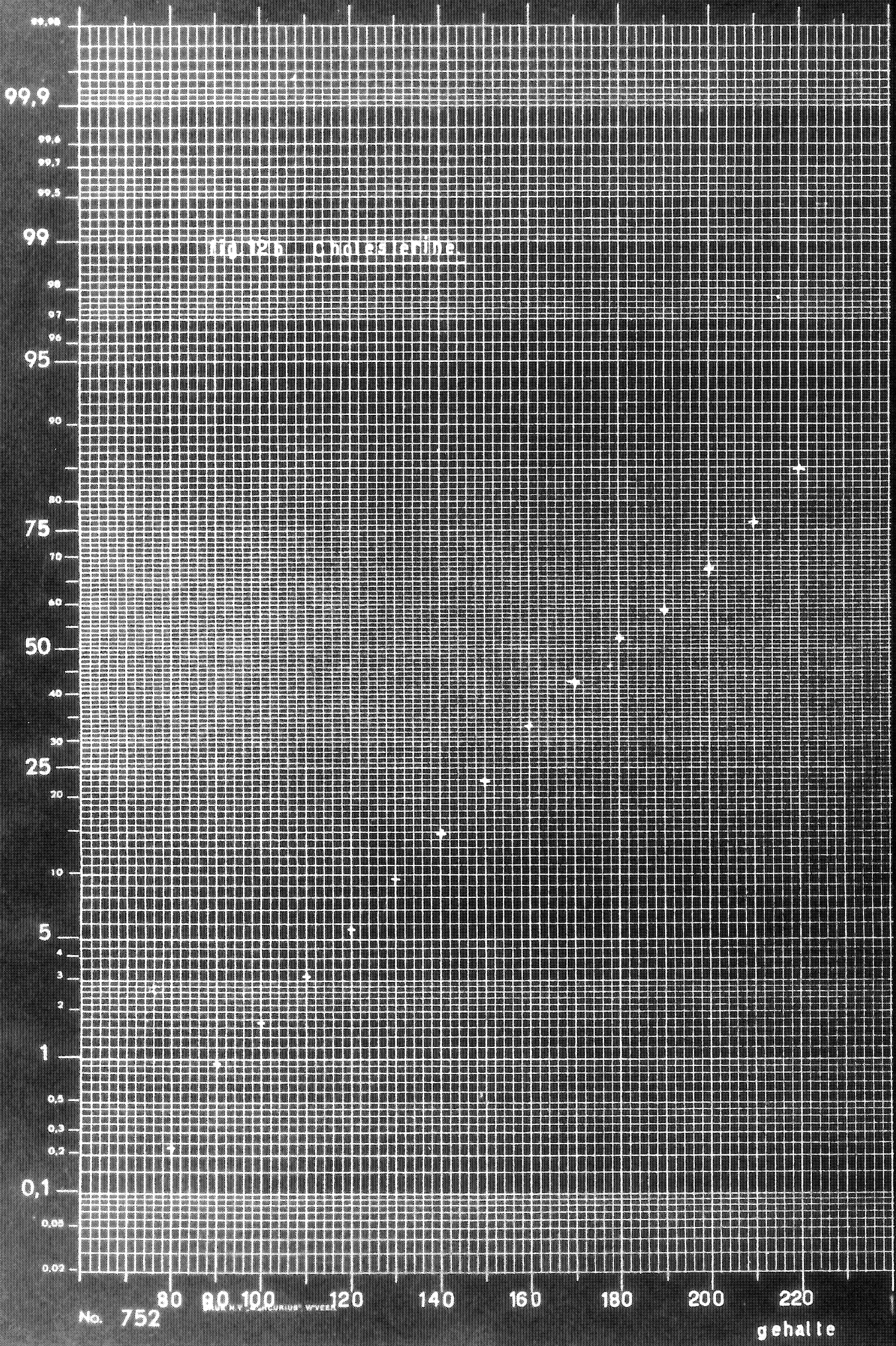






Alkalischer Phosphat

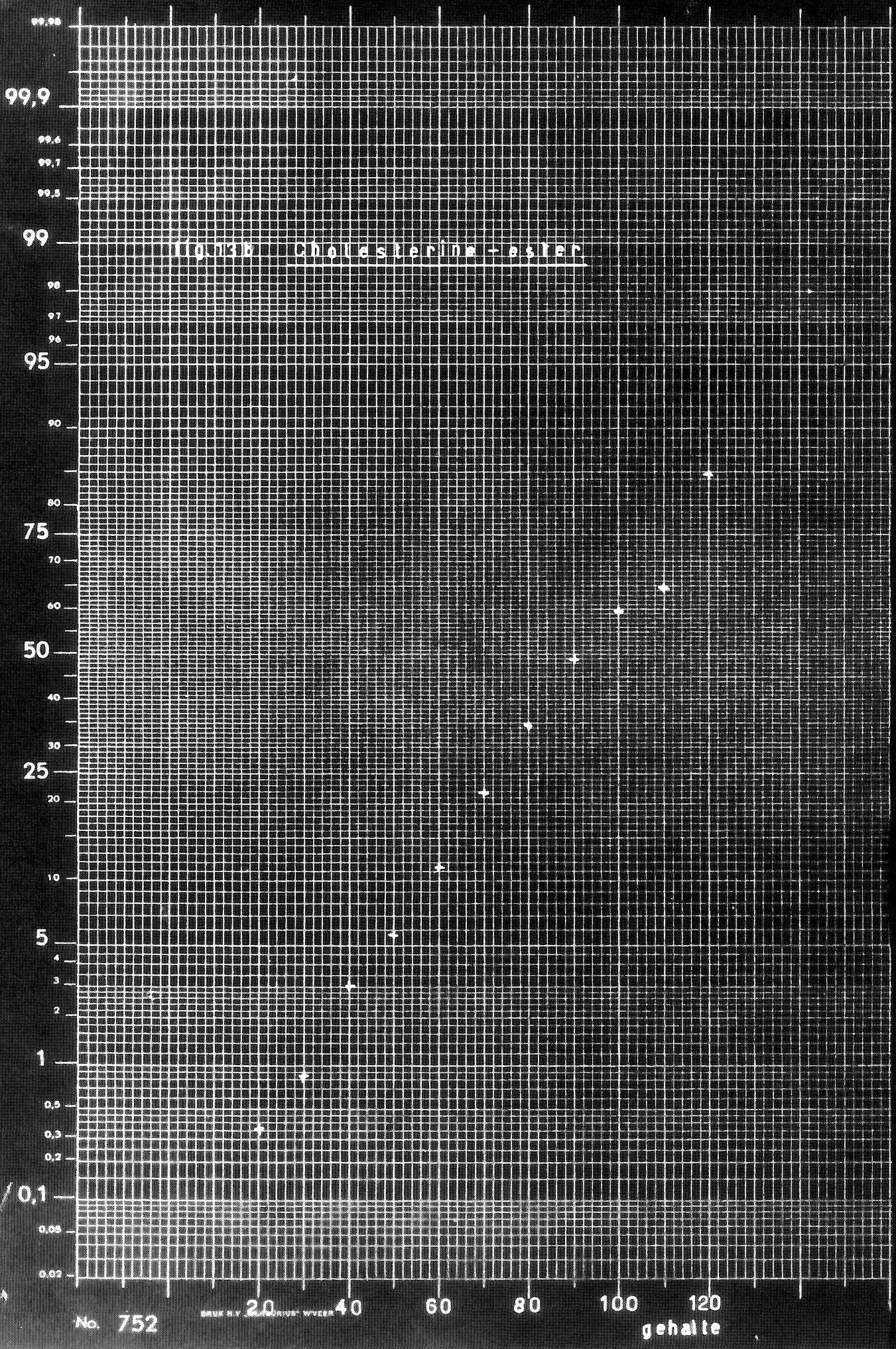
Fastphat in Serum



No. 752

80 80 100 120 140 160 180 200 220

gehalte



No. 752

DRUK N.Y. CAROLUS WVEER

gehalte

Algemene gang van zaken bij het toetsen van een <sup>1)</sup>  
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 <sup>2)</sup>, 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  $\alpha$ , zodat  $Z$  dus van  $\alpha$  afhankelijk is <sup>3)</sup>.  $Z$  heet de kritieke zone van de toets,  $\alpha$  de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor  $\alpha$  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  $\alpha$  moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien  $H_0$  juist is, gelijk aan  $\alpha$ . Derhalve is  $\alpha$  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

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

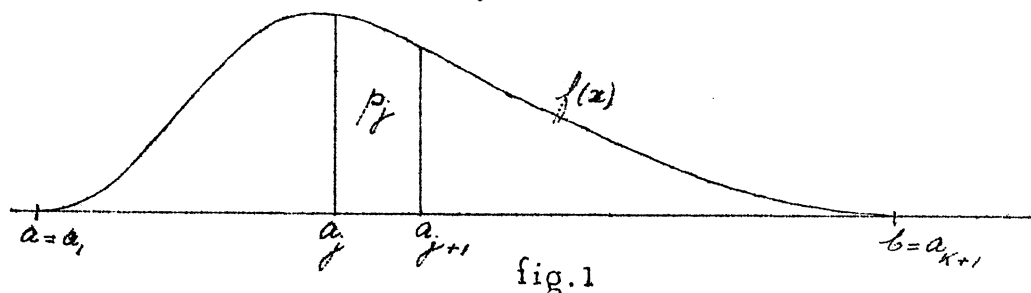
- J.Neyman, First course in probability and statistics, New York, 1950, Chapter 5.  
J.Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.

De  $\chi^2$ -toets voor aanpassing 1)

De  $\chi^2$ -toets voor aanpassing dient voor het toetsen van de hypothese  $\mathcal{H}_0$ , dat een stelsel waarden  $x_1, x_2, \dots, x_n$  een steekproef is uit een continue waarschijnlijkheidsverdeling met verdelingsdichtheid  $f(x)$ . Hierbij wordt een grootte  $\chi^2$  waarvan de verdelingsfunctie bekend is, als toetsingsgrootte gebruikt.

De verdelingsfunctie van  $\chi^2$  wordt nog nader gekarakteriseerd door een grootte  $\nu$  welke het aantal vrijheidsgraden van  $\chi^2$  aangeeft. De grootte van  $\chi^2$  en het getal  $\nu$  worden op de volgende wijze bepaald:

Wanneer  $f(x)$  de onderstelde verdelingsdichtheid is (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 ( $j=1, \dots, K$ ). Daar het gehele oppervlak onder de kromme de totale waarschijnlijkheidsmassa voorstelt, dus gelijk aan 1 is, geldt de betrekking  $p_1 + p_2 + \dots + p_K = 1$

Bestaat nu de steekproef uit  $n$  waarnemingen  $x_1, x_2, \dots, x_n$  dan zal men onder de hypothese  $\mathcal{H}_0$  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  $n_j$  waarden in  $[a_1, a_2]$  etc.) dan wordt  $\chi^2$  als volgt gedefinieerd:

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

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

Deze  $\chi^2$  is een maat voor de afwijking van de werkelijke verdelingsdichtheid van  $\underline{x}$  van de onderstelde verdelingsdichtheid  $f(x)$ . Is de werkelijke verdelingsdichtheid van  $\underline{x}$  inderdaad gelijk aan  $f(x)$  zoals volgens  $\mathcal{H}_0$  ondersteld wordt, dan zal  $\chi^2$  in het algemeen klein zijn; is de verdelingsdichtheid echter verschillend van  $f(x)$ , dan worden grote waarden van  $\chi^2$  waarschijnlijker dan het geval is, als  $\mathcal{H}_0$  juist is. De kritieke zône voor deze toets geeft men daarom de gedaante  $\chi^2 > \chi_0^2$ , waarbij  $\chi_0^2$  correspondeert met een onbetrouwbaarheidsdrempel  $\alpha$  en uit tabellen voor de  $\chi^2$ -verdeling opgezocht kan worden, waarbij rekening gehouden moet worden met het in de tabellen aangegeven:

Aantal vrijheidsgraden  $\nu$ . 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  $\nu = k - 1$ .

Voorbeelden: bij toetsing op bovenstaande wijze van de hypothese, dat de steekproef  $x_1, \dots, x_n$  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  $\nu = k - 1$ .

Opm. In de praktijk worden de getallen  $p_1, p_2, \dots, p_k$  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  $\chi^2$ -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.  $\geq 10$ .

Litteratuur:

M.G.Kendall, The advanced Theory of Statistics  
London 1947, Deel 1 hfdst. 12.