Niet voor publicatie
Nadruk verboden

Interne Nota No 26.

HET GEBRUIK VAN GRAFIEKENPAPIER

INHOUD:

DEEL I. BESCHRIJVING VAN GRAFIEKEN

A. Inleiding

Grafisch papier wordt gebruikt, indien men de samenhang tussen 2 of meer reeksen wenst te beoordelen op grond van een visueel beeld. De keuze van het te gebruiken papier dient er op te zijn gericht, dat de voor het oordeel van belang zijnde kenmerken duidelijk in het oog vallen. Van de veelzijdige mogelijkheden, welke het grafiekenpapier biedt, zijn hieronder enkele belangrijke toepassingen behandeld.

B. Metrisch papier

Op metrisch papier zijn de ordinaten verdeeld in onderling gelijke stukken, gewoonlijk millimeters. Dit heeft het voordeel, dat visueel gelijke afstanden tevens gelijke absolute verschillen voorstellen. Zo is op grafiek I[1] de afstand tussen f.2,80 en f.2,60 dezelfde als tussen f.0,80 en f.0,60. Het verschil tussen de bedragen is in beide gevallen f.0,20. Grafiek I geeft dus een beeld van de absolute seizoenverschillen in de eierprijzen vóór en na de oorlog. Voor de beoordeling van de prijzen is het verder van belang te weten, of de mate van schommeling na de oorlog groter of kleiner is dan vóór de oorlog. Dit is op grafiek I moeilijk te zien, hetgeen een nadeel is van metrisch papier. De grote verschillen tussen toppen en dalen na de oorlog suggereren een grotere mate van schommeling. Dit soort grafieken bergt dus het gevaar in zich, dat men aan dergelijke suggesties zal toegeven.

---

1) Zie Bijlage

Ter verkrijging van inzicht in de relatieve schommelingen zal men de grafiek op logarithmisch papier moeten uitzetten.

C. Enkel logarithmisch papier

Van dit papier is de ene ordinaat metrisch en de andere logarithmisch verdeeld (grafiek II). Op de logarithmische ordinaat zijn van de er naast geschreven getallen de logarithmen afgezet. Dit brengt met zich, dat dezelfde relatieve verschillen overal een zelfde afstand beslaan. Zo zijn er gelijke afstanden tussen 400 en 200, 300 en 150, 200 en 100, 100 en 50, hetgeen betekent, dat de verhouding tussen de genoemde getallen dezelfde is, in dit geval 2 : 1.
Grafiek II geeft, op logarithmische schaal, dezelfde eierprijzen weer als grafiek I. Onmiddellijk is nu te zien, dat de mate van schommeling na de oorlog geringer is dan vóór de oorlog en dat de mate van schommeling van 1934 tot 1939 de neiging heeft, geringer te worden.
De absolute schommeling is niet in één oogopslag te zien, doch wel te berekenen op grond van de getallen op de Y-as.

Een ander voordeel van een logarithmische schaal is de mogelijkheid tot het grafisch weergeven van reeksen, waarvan de getallen in absolute waarde vrij sterk uiteenlopen.
De lijnen op grafiek II beslaan een kleinere ruimte dan op grafiek I, terwijl de curve van 1934 - 1939 bijna 2 maal zo groot is afgebeeld. De schaalverdeling op grafiek I loopt van f.0,20 tot f.3,30, op grafiek II van f.0,10 tot f.5,-. Was de curve van vóór de oorlog op grafiek II ongeveer even groot afgezet als op grafiek I, dan zou de schaal van grafiek II lopen van f.0,10 tot f.51,20.

D. Dubbel logarithmisch papier

Hier zijn beide ordinaten logarithmisch verdeeld. Dit papier wordt toegepast op materiaal waarbij men kan verwachten, dat het verband tussen de reeksen van kromlijnige aard is. In de grafieken III en IV zijn voor het gebied Oostelijk Noordbrabant 1952/'53 de standaarduren per arbeidskracht uitgezet tegenover het aantal arbeidskrachten per bedrijf. Indien het aantal arbeiders precies is afgestemd op de arbeidsbehoefte, zullen alle bedrijven hetzelfde aantal standaarduren per arbeidskracht hebben. Uit de grafieken blijkt, dat deze aanpassing niet ideaal is bij een toenemend aantal arbeidskrachten per bedrijf neemt het aantal standaarduren per arbeidskracht af. In welke mate geschiedt dit? In de grafieken III en IV is dezelfde lijn getekend; de kromme lijn van grafiek III is op grafiek IV een rechte geworden. De relatieve toeneming van de ene factor t.a.v. de relatieve afneming van de andere factor blijkt uit de helling van de lijn. De hellingscoëfficiënt is de tangens van de hoek waaronder deze lijn de X-as snijdt en heeft de waarde - 0,4314. D.w.z. dat, als een bedrijf 1% meer arbeiders heeft dan een ander, het aantal standaarduren per arbeidskracht in het algemeen zal dalen met 0,43%.

De formule van de lijn is:

volgens grafiek IV: $\log y = -0{,}43 \log x + \log 4420$

volgens grafiek III: $y = 4420 \cdot x^{-0{,}43}$

of anders geschreven: $y = \dfrac{4420}{x^{0{,}43}}$

Het voordeel van deze exponentiële vergelijkingen is, dat het verband tussen de relatieve verandering van de ene factor en de relatieve verandering van de andere factor voor alle punten van de lijn dezelfde kwantitatieve inhoud heeft, nl. de exponent van x. Dit voordeel verdient o.m. de aandacht bij het marktonderzoek waarbij men zoekt naar een wiskundig verband tussen de vraag of het aanbod en de daarop van invloed zijnde factoren. Zo kan voor het formuleren van een vraagvergelijking een lineair verband ($y = ax_1 + bx_2 + cx_3 + \ldots\ldots + p.$) bevredigende resultaten opleveren. Indien het verband van kromlijnige aard is, dan zal een logarithmische functie ($y = p.x_1^{\alpha} x_2^{\beta} x_3^{\gamma} \ldots\ldots$ of $\log y = \alpha \log x_1 + \beta \log x_2 + \gamma \log x_3 + \ldots\ldots + \log p.$) betere resultaten kunnen hebben. De coëfficiënt a in de lineaire functie betekent, dat, indien y verandert met een absoluut bedrag 1, x_1 verandert met een absoluut bedrag $\frac{1}{a}$. De elasticiteit van de vraag (de verhouding tussen de relatieve veranderingen in de factoren) is dus voor ieder punt van de curve anders. Ook bij het logarithmisch verband geldt, dat bij een absolute verandering van log y met α, de factor log x_1, met het absolute bedrag $\frac{1}{\alpha}$ verandert. Een absolute verandering van een logarithme echter betekent een relatieve verandering van de variabele (dus x_1). D.w.z. dat de elasticiteit van de vraag bij een logarithmisch verband op alle punten van de curve dezelfde is.

E. Driehoekcoördinaten

Dit papier heeft een driehoekige vorm, waardoor dus 3 variabelen ten opzichte van elkander worden beschouwd. De variabelen worden uitgezet in procenten van hun totaal. De vorm berust op de meetkundige stelling, dat de som der afstanden van een willekeurig punt binnen een gelijkzijdige driehoek gelijk is aan de hoogtelijn, dus aan een constante.

In grafiek V is de bruto-opbrengst in het gebied der zelfkazende bedrijven van Zuidholland in 1952/53 weergegeven in de samenstelling: melk en kaas; omzet en aanwas rundvee; varkenshouderij. De waarde van de variabele wordt gemeten langs de loodlijn. Zo ligt het onderste punt links op: 24% melk en kaas, 6% omzet en aanwas rundvee en 70% varkens. Uit de grafiek blijkt, dat de belangen van de bedrijven bij de opbrengst uit een bepaald bedrijfsonderdeel uiteenlopen. Er is een algemene tendentie, dat bij dalend aandeel van de varkensopbrengst zowel het aandeel van melk en kaas als van omzet en aanwas rundvee stijgt. Volgens de getekende lijn heeft deze tendentie het volgende verloop:

Melk en kaas	Omzet en aanwas rundvee	Varkens	Totale opbrengst
30%	8 %	62 %	100%
40%	10,5%	49,5%	100%
50%	13 %	37 %	100%
60%	15,5%	24,5%	100%
70%	18 %	12 %	100%
75%	19 %	6 %	100%

De bedrijfsgrootte blijkt weinig invloed te hebben op de samenstelling van de bruto-opbrengst. Er is een lichte tendentie dat aan het uiteinde van de lijn aan de varkenskant meer grote dan kleine bedrijven liggen en aan het uiteinde met hoog melk- en kaasaandeel meer kleine dan grote bedrijven.

De grafieken VI en VII geven de samenstelling weer van de bruto-opbrengst in het consulentschap Eindhoven. Uit grafiek VI zien wij, dat de bedrijven van 5-7 ha een groter belang hebben bij de opbrengst van varkens en pluimvee dan de bedrijven van 7-10 ha, die meer in de richting van rundvee gaan. De punten in grafiek VII liggen in het algemeen hoger dan in grafiek VI, zodat de bedrijven boven 10 ha weer iets verder in de richting van rundvee gaan dan die van 7-10 ha.

De verschillen in bedrijfsgrootte van bedrijven boven 10 ha zijn voor de samenstelling van de bruto-opbrengst van weinig belang.

F. Waarschijnlijkheidspapier

Dit papier wordt gebruikt voor de analyse van frequentieverdelingen ten dienste van een doelmatige beschrijving van curven.

Curven kunnen op verschillende manieren worden gekarakteriseerd, b.v.

rekenkundig gemiddelde $= \frac{x_1 + x_2 + \ldots\ldots + x_n}{n}$

modus = het meest voorkomend verschijnsel

mediaan = de waarde van de middelste waarneming

meetkundig gemiddelde $= \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \ldots\ldots x_n}$

De logarithme van het meetkundig gemiddelde is hetzelfde als het rekenkundig gemiddelde van de logarithmen der gegevens:

$\log \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \ldots\ldots x_n} = \frac{1}{n} (\log x_1 + \log x_2 + \ldots\ldots + \log x_n).$

Bij de kromme van Gauss vallen modus, mediaan en rekenkundig gemiddelde samen. Deze curve is klokvormig en ontstaat, indien de waarde van een bepaalde grootheid, behalve door systematische krachten, wordt bepaald door een groot aantal krachten, die elk op zichzelf een kleine afwijking even vaak naar links als naar rechts kunnen veroorzaken, zonder dat de uitwerking van de ene kracht invloed heeft op die van de andere.

De spreiding kan op verschillende manieren worden gemeten. B.v.:

variatiebreedte : $x_{max.} - x_{min.}$

gemiddelde afwijking: $\frac{\Sigma x - \bar{x}}{n}$ 1)

standaardafwijking : $\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$ 2)

variatiecoëfficiënt $= \frac{\sigma}{\bar{x}}$ 3)

kwartielen, dedelen en porcentielen.

Omtrent de vorm van de curve kan men zich oriënteren door de waarde van enkele getallen uit het materiaal aan te geven:

mediaan: de waarde der waarneming, welke het materiaal in 2 gelijke delen verdeelt;

1) Rekenkundig gemiddelde van de absolute afwijkingen.

2) Wortel uit het rekenkundig gemiddelde van de kwadraten der afwijkingen.

3) Standaardafwijking gedeeld door het rekenkundig gemiddelde.

kwartiel : de waarde der 3 waarnemingen, welke het materiaal in 4 gelijke delen verdelen;

deciel : de waarde der 9 waarnemingen, welke het materiaal in 10 gelijke delen verdelen;

percentiel : de waarde der 99 waarnemingen, welke het materiaal in 100 gelijke delen verdelen.

Indien gewenst, kan men het materiaal al naar behoefte indelen.
Bij de kromme van Gauss liggen:

68,27% van de waarnemingen tussen: $(M - \sigma)$ en $(M + \sigma)$,
95,45% van de waarnemingen tussen: $(M - 2\sigma)$ en $(M + 2\sigma)$ en
99,73% van de waarnemingen tussen: $(M - 3\sigma)$ en $(M + 3\sigma)$.

Hierin is M het rekenkundig gemiddelde. Deze kromme is dus te karakteriseren door de waarde van de punten 2%, 16%, 50%, 84% en 98%.

Wenst men de vorm van een curve te onderzoeken door de curve te tekenen, dan rijst het vraagstuk van de klasse-indeling. Kiest men de klassen te klein, zodat men vele klassen heeft, terwijl in een aantal klassen geen waarnemingen zijn, dan krijgt men geen goede indruk van de curve. Dit is evenmin het geval bij te grote klassen, waarbij een zeer groot aantal waarnemingen in slechts 1 klasse valt. Heeft men de juiste klasse-indeling, dan moeten het gemiddelde, de standaardafwijking en de oppervlakte van de figuur worden berekend, waarna met behulp van een bepaalde tabel de best passende "normale" curve wordt bepaald. Men ziet dan onmiddellijk, in hoeverre het onderzochte materiaal met een "normale" verdeling overeenkomt.

Dit onderzoek wordt aanmerkelijk vereenvoudigd door gebruik te maken van waarschijnlijkheidspapier. De horizontale as van dit papier is lineair verdeeld en is bestemd voor het afzetten van de klasse-indeling. De verticale as heeft een waarschijnlijkheidsverdeling in percenten. Deze verdeling is zodanig, dat de frequentieverdeling van een "normale" curve, uitgezet als somcurve een rechte lijn oplevert. De somcurve wordt verkregen door voor elke klasse aan te geven hoeveel percent van de gevallen zich tot en met die klasse heeft voorgedaan. De keuze van de klassegrootte voor het bepalen van de somcurve kan met grotere soepelheid geschieden dan ingeval men onderzoekt zonder gebruik te maken van waarschijnlijkheidspapier. Indien de curve "normaal" is en dus de somcurve op waarschijnlijkheidspapier een rechte is, zijn gemiddelde en standaardafwijking rechtstreeks af te lezen. Het gemiddelde is gelijk aan de mediaan en staat dus op het punt 50%. Zoals boven reeds gezegd omvat het gebied ($M - \sigma$) tot ($M + \sigma$) 68% van de waarnemingen, dus 34% ligt in dit gebied ter weerszijde van M. De standaardafwijking is dus: het verschil van de waarde van het 84% punt met het 16% punt, gedeeld door 2.

In de tabel VIII en de grafieken IX en X is de frequentieverdeling weergegeven van de melkgift per koe op 111 bedrijven in Oostelijk Noordbrabant (Consulentschap Eindhoven) in 1952/53.

Uit grafiek IX zien wij, dat deze frequentieverdeling dicht nadert tot de "normale" verdeling. Op het waarschijnlijkheidspapier (grafiek X) is door de punten een rechte lijn getrokken. De rechte lijn geeft de bij de frequentieverdeling best passende "normale" curve weer. Het gemiddelde van deze "normale" curve is gelijk aan de mediaan en is 3800 kg. De standaardafwijking is $\frac{1000}{2} = 500$ kg. 68% van de waarnemingen ligt dus tussen 3300 en 4100 kg. Het onderste uiteinde van de rechte past moeilijk bij de 2 stippen. Een frequentieverdeling kan door het geringe aantal waarnemingen in de laagste en hoogste klasse afwijkingen vertonen van het "normale" beeld. De middenklassen omvatten een groter aantal waarnemingen, zodat men er naar dient te streven, de rechte lijn zo goed mogelijk bij het middengedeelte te laten passen.

Voor hetzelfde gebied is in tabel XI en de grafieken XII en XIII een frequentieverdeling gemaakt voor het vetgehalte.

De frequentieverdeling vertoont in grafiek XII een duidelijke drie-toppigheid. Er zijn groeperingen om de klassen 3,36 - 3,40%, 3,61 - 3,65% en 3,99 - 3,95%. Dit kan er op wijzen, dat t.a.v. het vetgehalte 3 soorten van bedrijven zijn te onderscheiden.

Binnen het kader van deze nota gaat het te ver de oorzaken van dit verschijnsel te analyseren. Ter toelichting van de curve zullen wij alleen ingaan op de invloed van het verschil in bedrijfsgrootte.

Tabel 1

Frequentieverdeling van bedrijfsgrootte
en vetgehalte

	Aantallen:			
Grootte	<3,51%	3,51-3,85%	>3,85%	totaal
5 - 7 ha	9	8	1	18
7 - 9 "	4	16	2	22
9 -10 "	2	11	3	16
10 -12 "	0	12	3	15
12 -16 "	6	15	3	24
>16 "	4	11	1	16
Totaal	25	73	13	111

Tabel 2

Percentuele verdeling van bedrijfsgrootte en vetgehalte

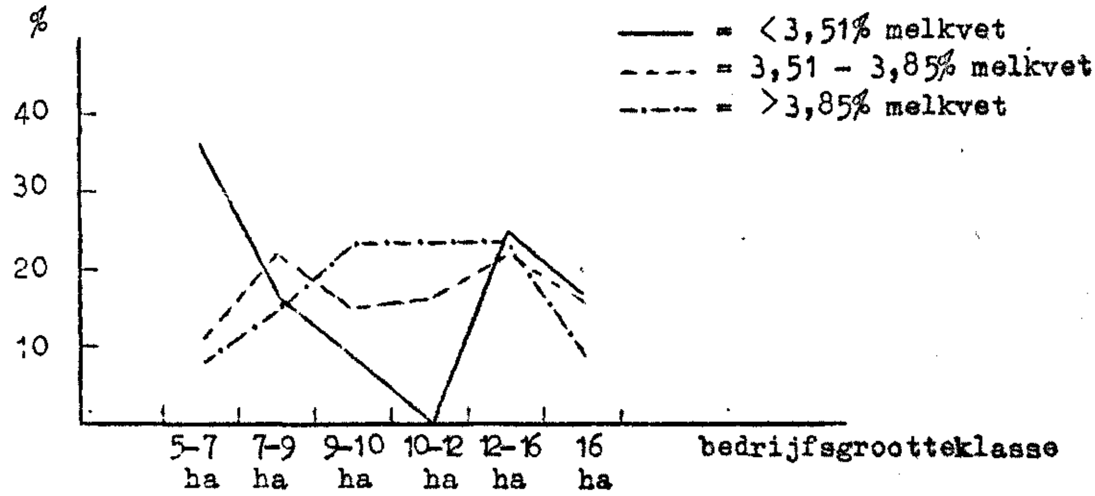
Grootte	A				B		
	<3,51%	3,51 - 3,85%	>3,85%	totaal	<3,51%	3,51-3,85%	> 3.85%
5 - 7 ha	50	45	5	100	36	11	8
7 - 9 "	18	73	9	100	16	22	15
9 -10 "	13	68	19	100	8	15	23
10 -12 "	0	80	20	100	0	16	23
12 -16 "	25	62	13	100	24	21	23
> 16 "	25	68	7	100	16	15	8
Totaal					100	100	100

Uit de frequentieverdeling van de bedrijfsgrootte (A) blijkt, dat 50% van de bedrijven van 5-7 ha een vetgehalte heeft van minder dan 3,51%. Van de overige bedrijfsgrootten bevindt zich van ieder ongeveer 70% in de klasse 3,51-3,85%. Van de bedrijven van 9-12 ha valt een belangrijk aandeel in de klasse boven 3,85%.
Bezien wij de frequentieverdeling van het melkvetgehalte (B), dan wordt de klasse < 3,51% in hoofdzaak gevormd door de bedrijven van 5-9 ha en 12 ha of meer. De klasse 3,51-3,85% omvat van bedrijven boven 7 ha twee toppen, nl. bij 7-9 ha en 12-16 ha. In de klasse boven 3,85% valt vooral het aandeel van de groep 9-16 ha op.

Wij kunnen dit samenvatten in de grafieken 3 en 4 (z.o.z.):

rafiek 3

Frequentieverdeling van de bedrijfsgrootte over 3 vetgehaltegroepen



In de grafieken XIV en volgende zijn enkele voorbeelden uitgewerkt van de wijze waarop curven kunnen zijn opgebouwd. De voorbeelden XIV, XV en XVI hebben betrekking op de superpositie (het gedeeltelijk samenvallen) van "normale" curven; het laatste voorbeeld betreft een scheve curve waarvan de logarithmen der variabelen een normale verdeling vertonen.

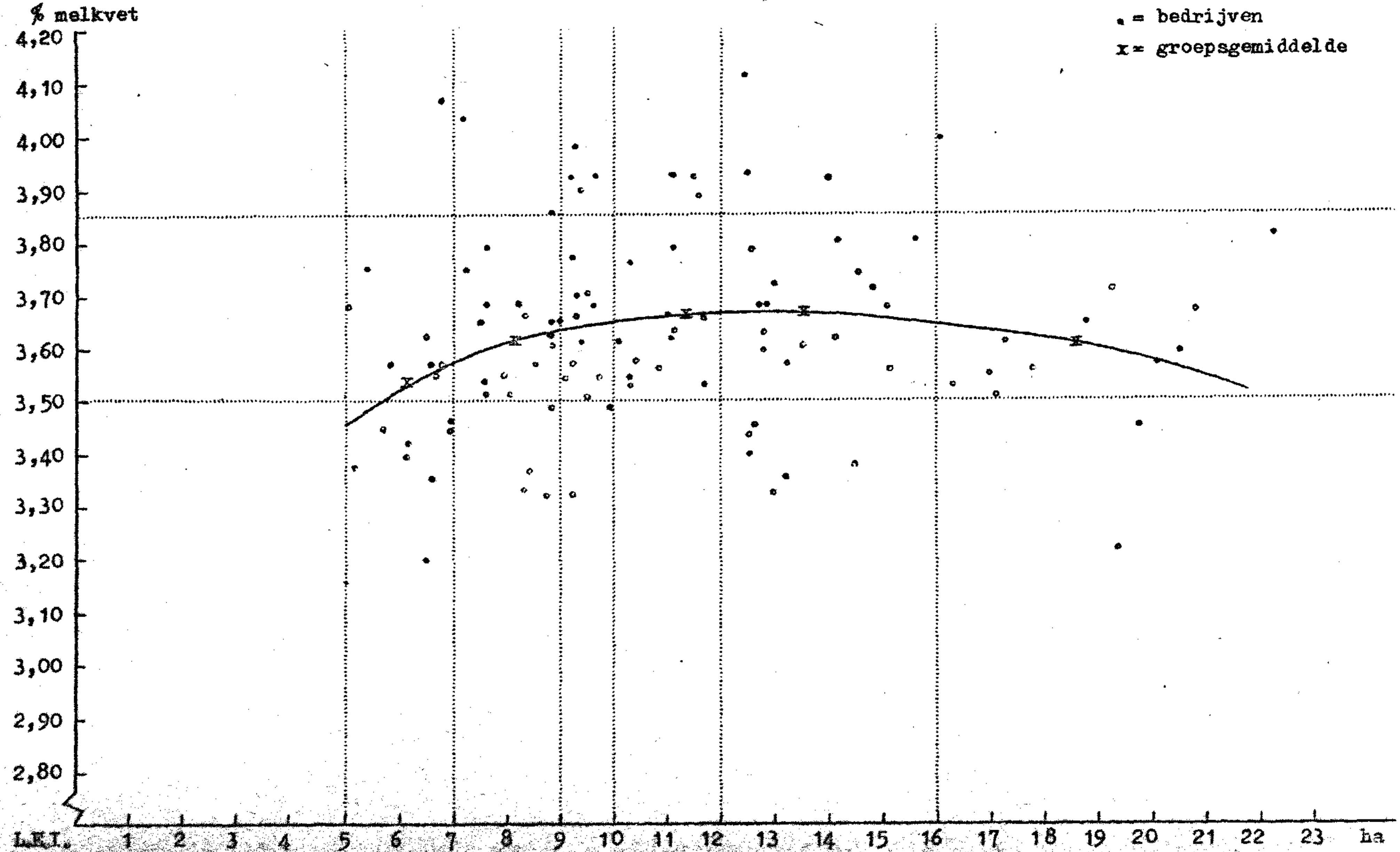
De grafieken XVIII en XIX bevatten een logarithmische verdeling en een waarschijnlijkheidsverdeling voor het zelf tekenen van deze schalen op het gewenste formaat.

's Gravenhage,
6 Februari 1954

A.H.J. Liberg, ec.drs

Grafiek 4

Oostelijk Noordbrabant 1952/53



L.E.I. 84

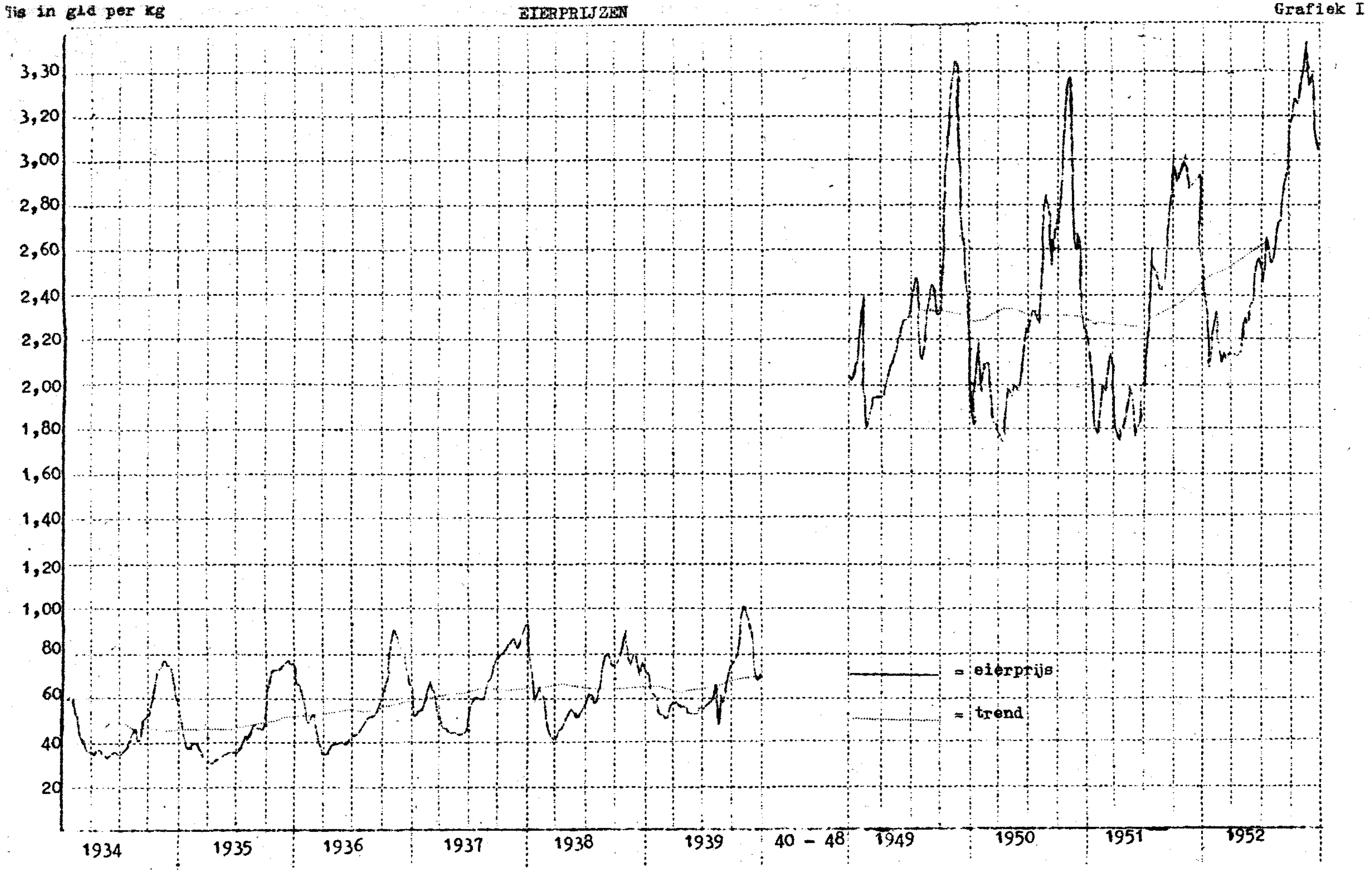
Interne nota No 26

HET GEBRUIK VAN GRAFIEKENPAPIER

Deel II Bijlage

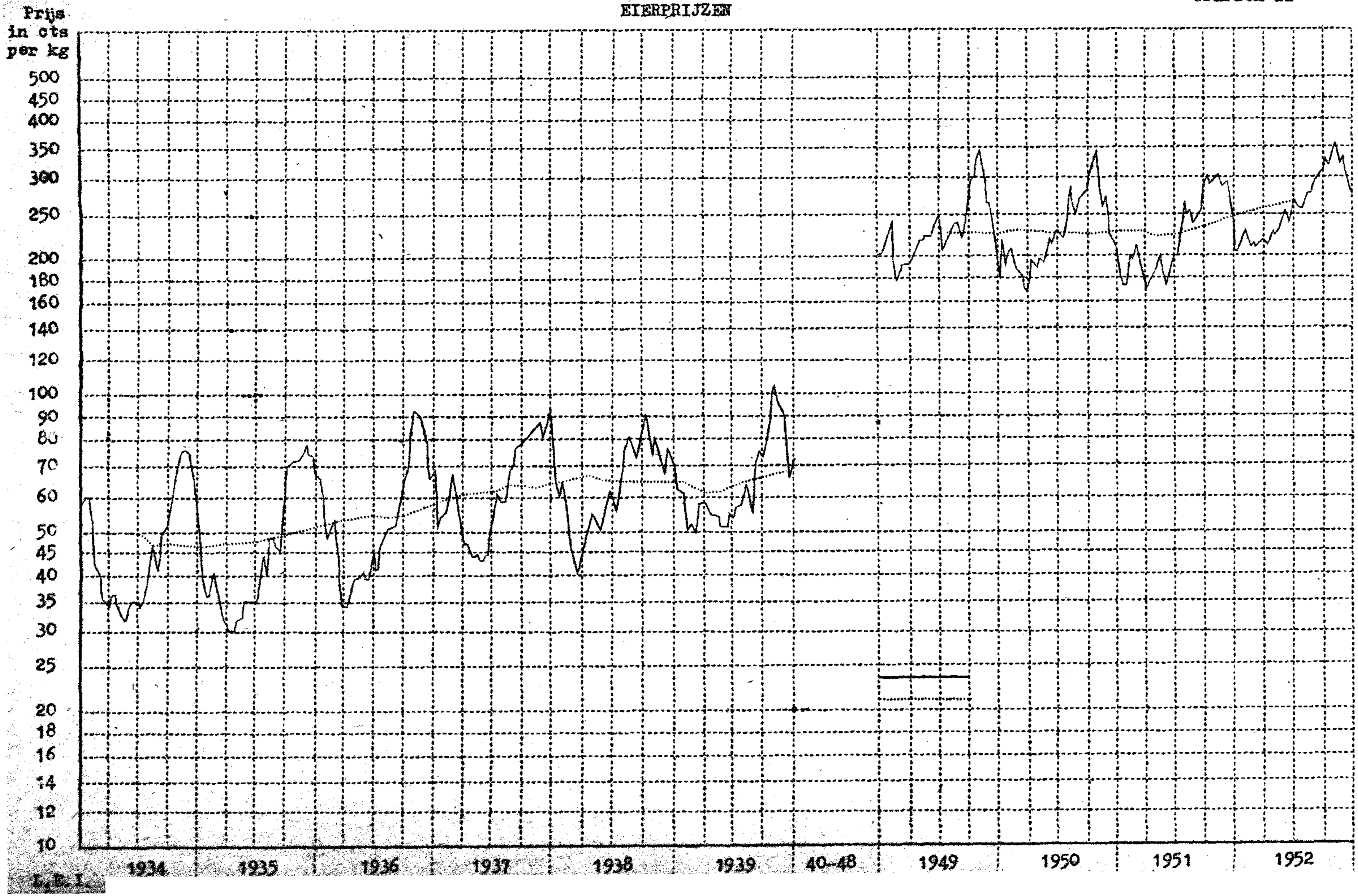
Grafiek I

EIERPRIJZEN



Grafiek II

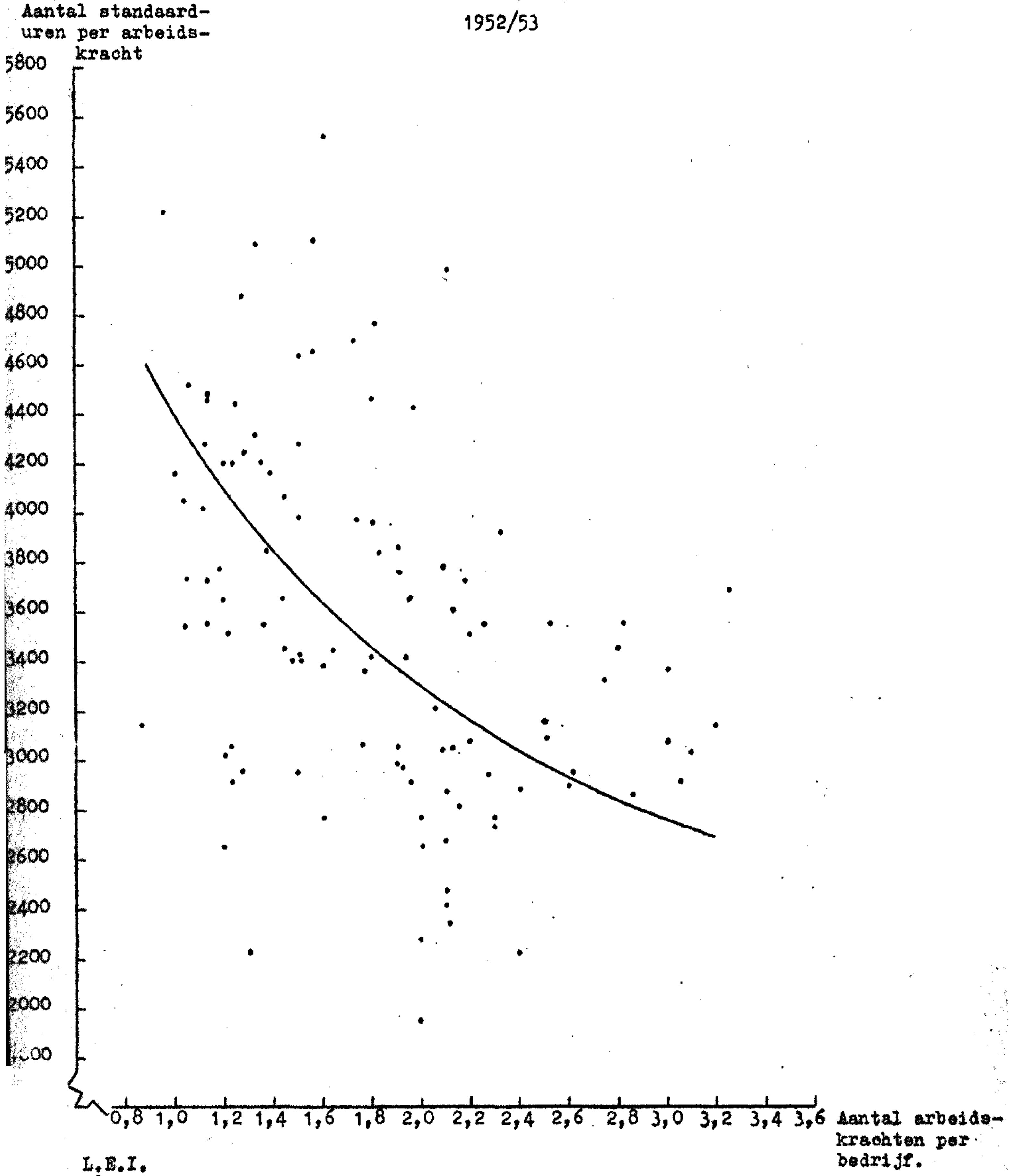
EIERPRIJZEN



Oostelijk Noordbrabant

Consulentschap Eindhoven

1952/53

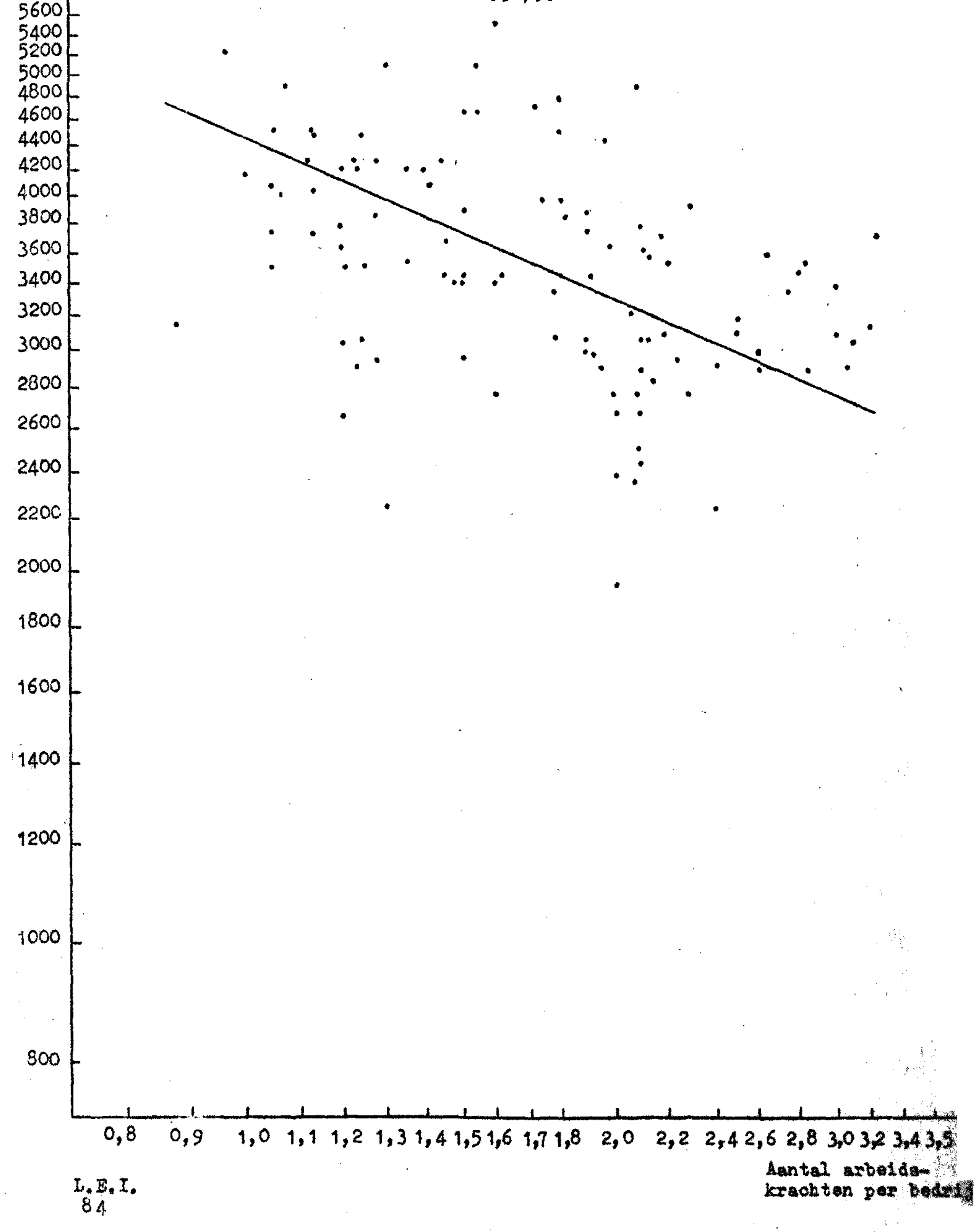


L.E.I.
84

Grafiek IV

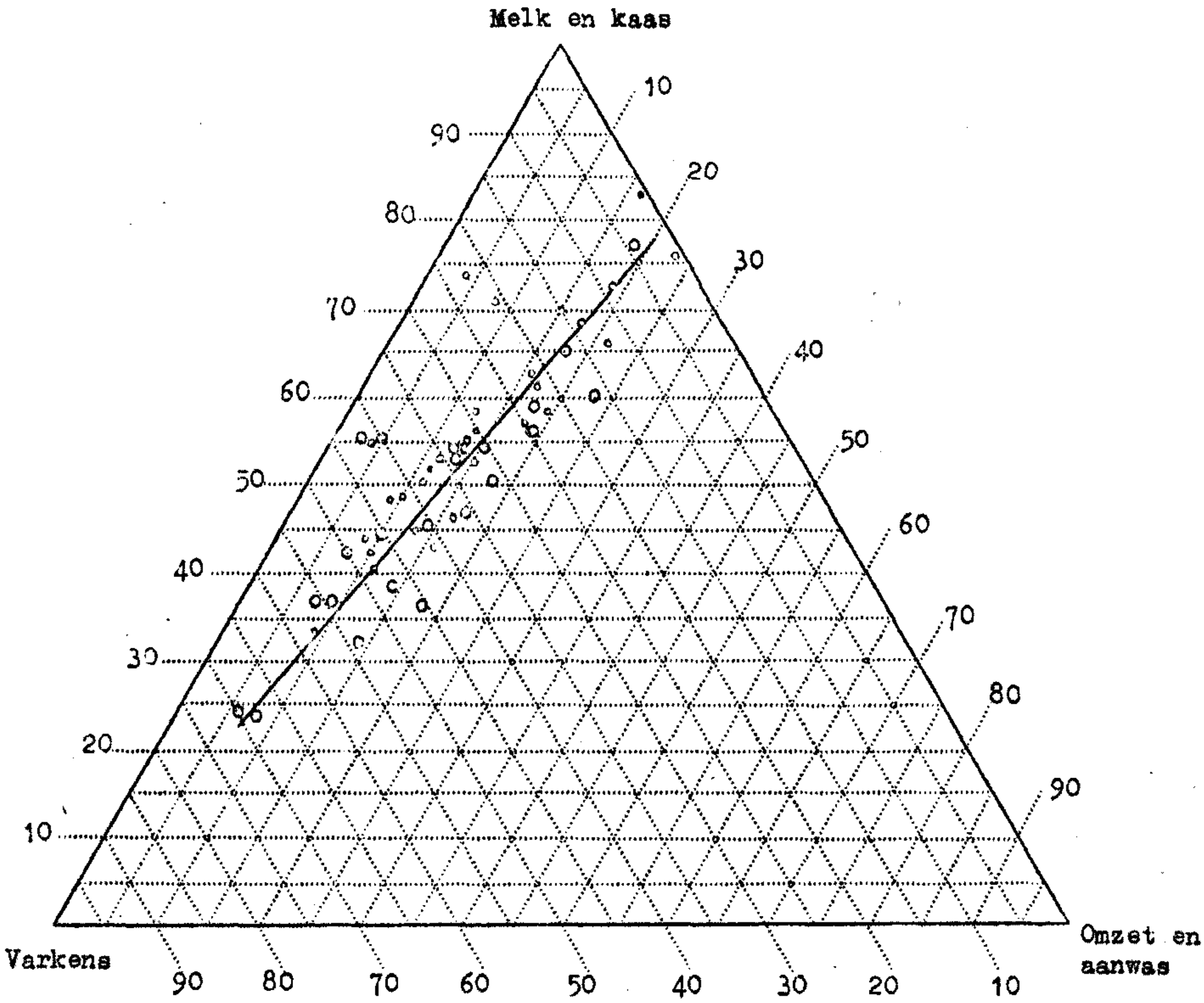
Oostelijk Noordbrabant
Consulentschap Eindhoven
1952/53

Aantal standaarduren
per arbeidskracht



L.E.I.
84

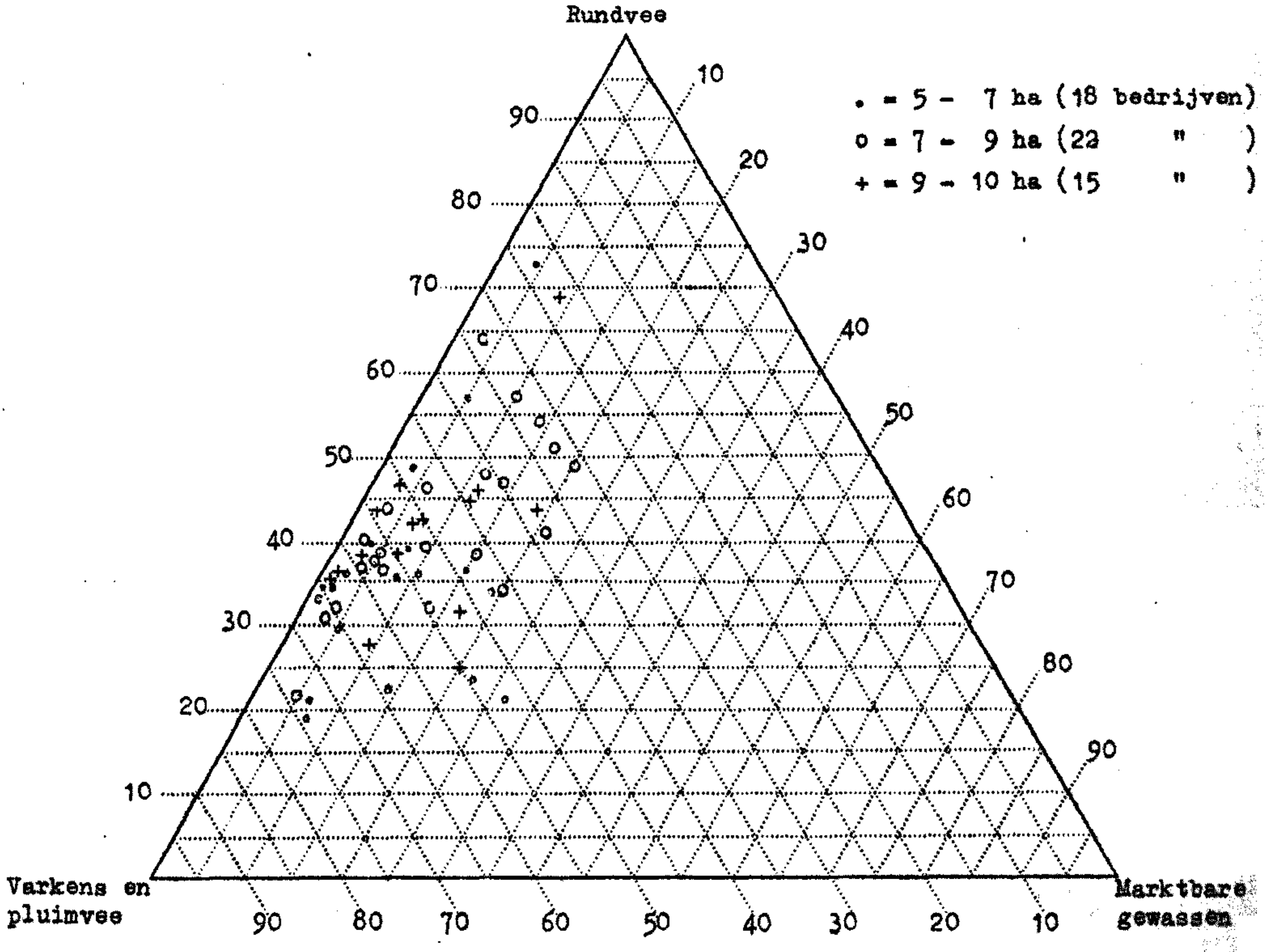
Samenstelling van de bruto-opbrengst
in Zuidholland, Zelfkazend Gebied 1952/53



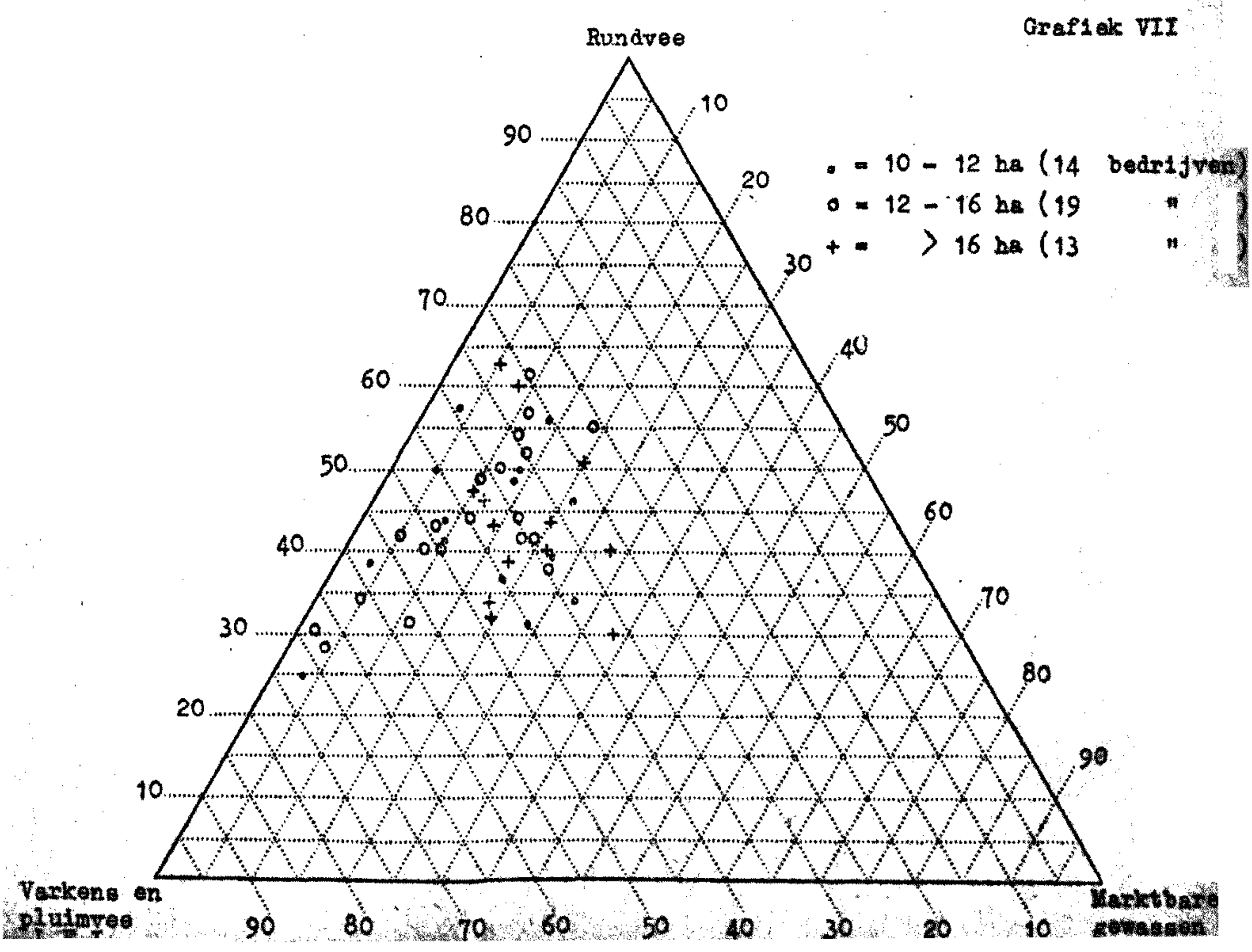
∘ = < 15 ha
o = > 15 ha

Grafiek VI

Samenstelling van de bruto-opbrengst in Oostelijk Noordbrabant, Consulentschap Eindhoven, 1952/53



Grafiek VII

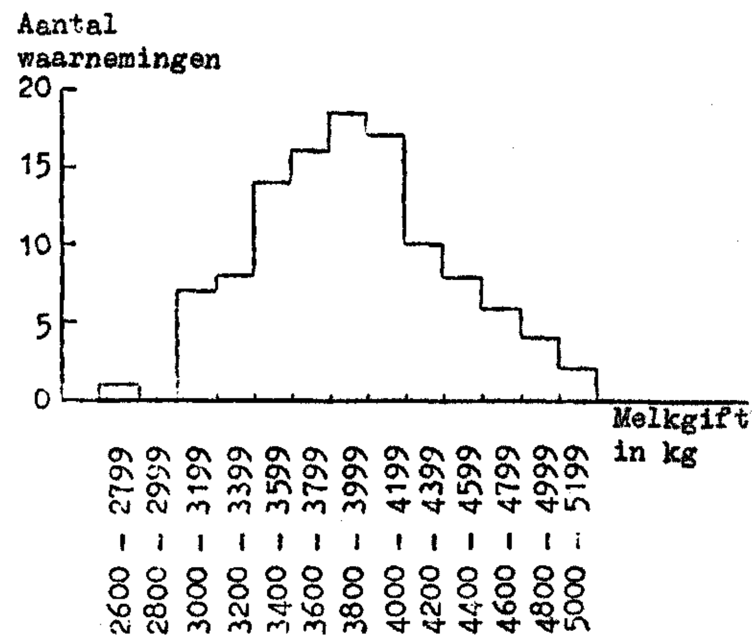


Frequentieverdeling van de melkgift in kg per koe
op 111 bedrijven in Oostelijk Noordbrabant in 1952/53
(Tabel VIII, grafieken IX en X)

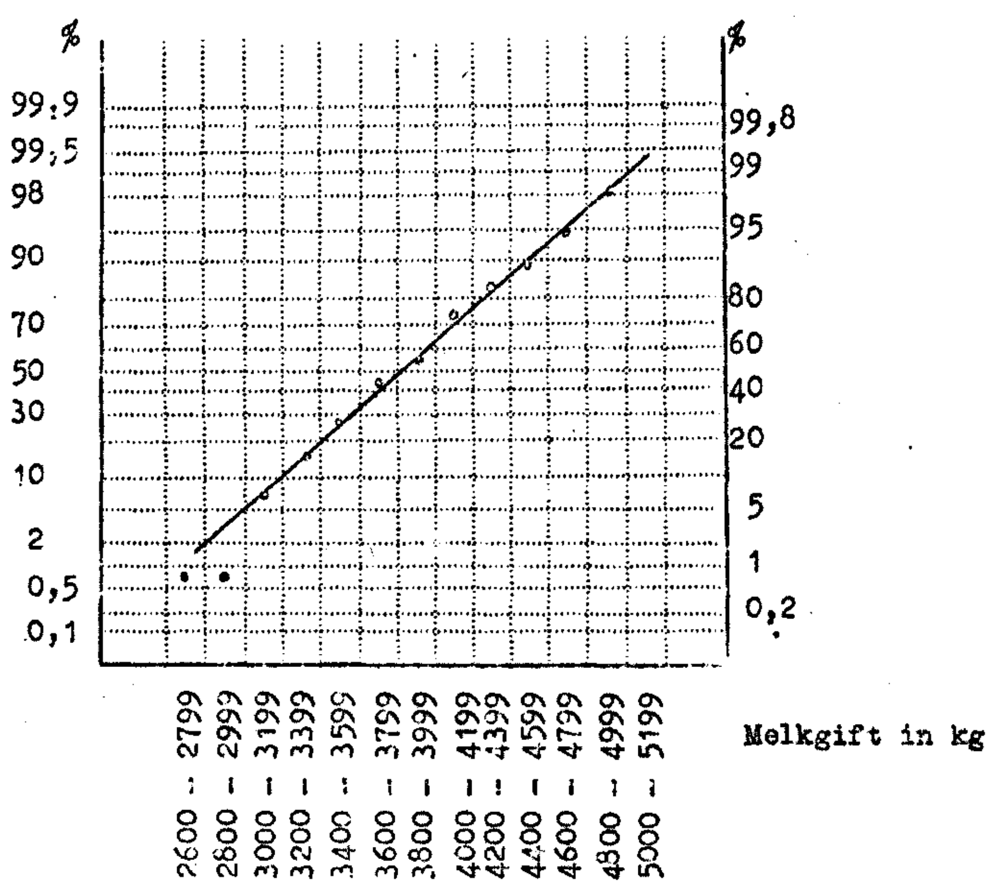
Tabel VIII

Melkgift in kg per koe	Aant. waarnemingen n = 111	Cumulatief percentage
2600 - 2799	1	0,9
2800 - 2999	0	0,9
3000 - 3199	7	7,2
3200 - 3399	8	14,4
3400 - 3599	14	27,0
3600 - 3799	16	41,4
3800 - 3999	18	57,7
4000 - 4199	17	73,0
4200 - 4399	10	82,0
4400 - 4599	8	89,2
4600 - 4799	6	94,6
4800 - 4999	4	98,2
5000 - 5199	2	100,0

Grafiek IX



Grafiek X

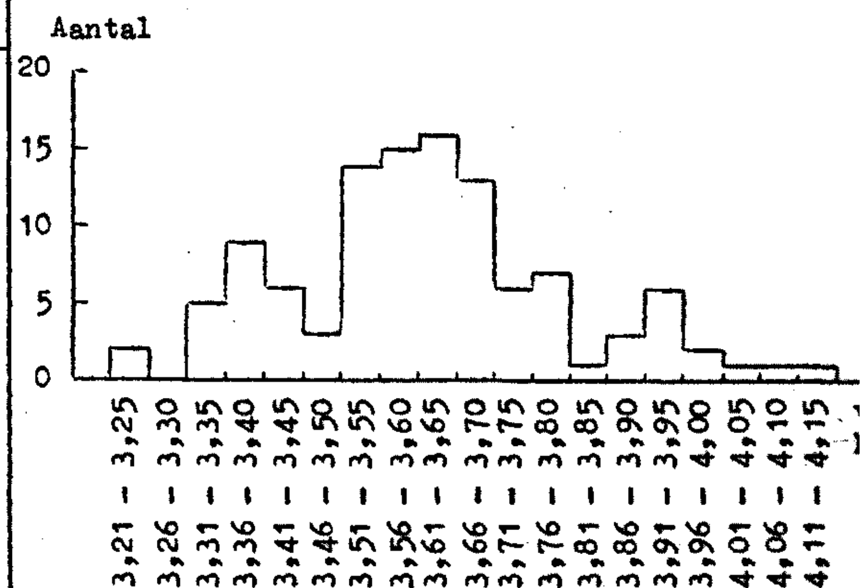


Frequentieverdeling van de melkgift in kg per koe op
111 bedrijven in Oostelijk Noordbrabant in 1952/53
(Tabel XI, grafieken XII en XIII)

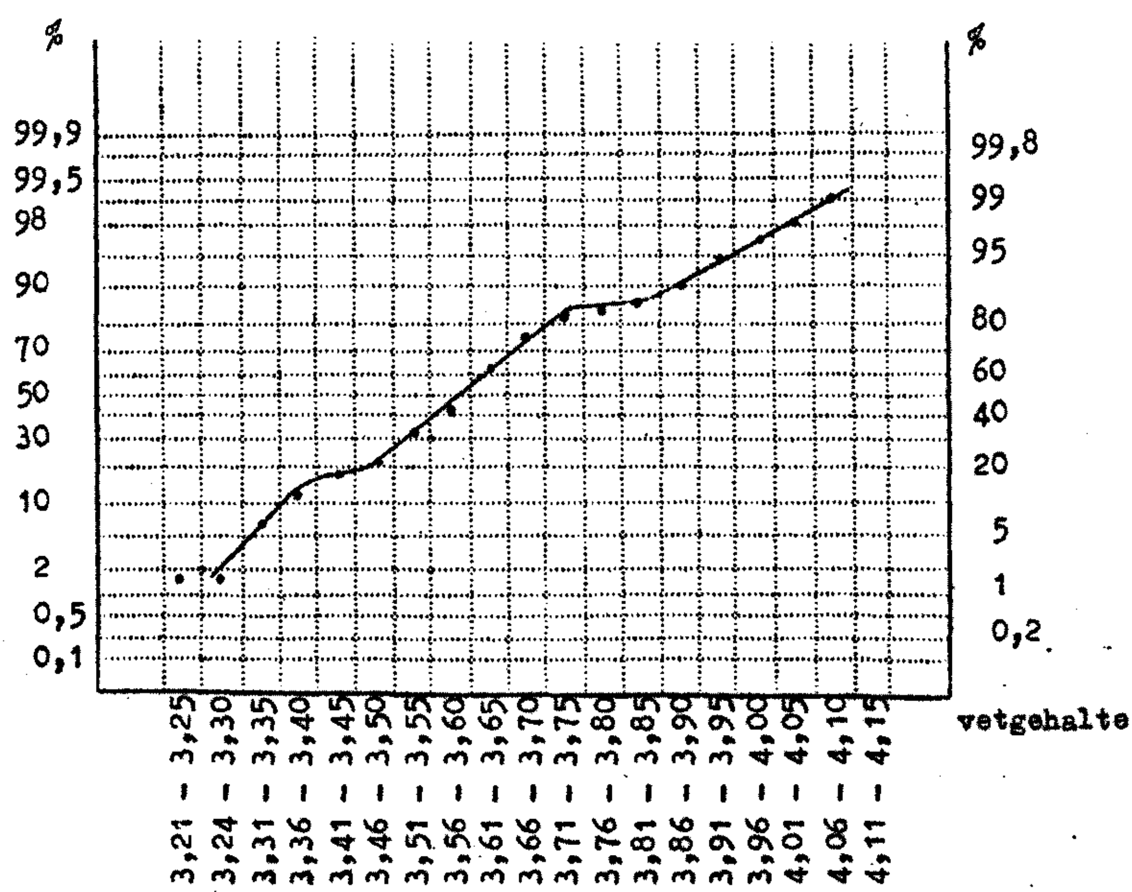
Tabel XI

% melkvet	Aantal n = 111	Cumulatief percentage
3,21 - 3,24	2	1,8
3,24 - 3,30	0	1,8
3,31 - 3,35	5	6,3
3,36 - 3,40	9	14,4
3,41 - 3,45	6	19,8
3,46 - 3,50	3	22,5
3,51 - 3,55	14	35,1
3,56 - 3,60	15	48,6
3,61 - 3,65	16	63,1
3,66 - 3,70	13	74,8
3,71 - 3,75	6	80,2
3,76 - 3,80	7	86,5
3,81 - 3,85	1	87,4
[illegible]6 - 3,90	3	90,1
3,91 - 3,95	6	95,5
3,96 - 4,00	2	97,3
4,01 - 4,05	1	98,2
4,06 - 4,10	1	99,1
4,16 - 4,20	1	100,0

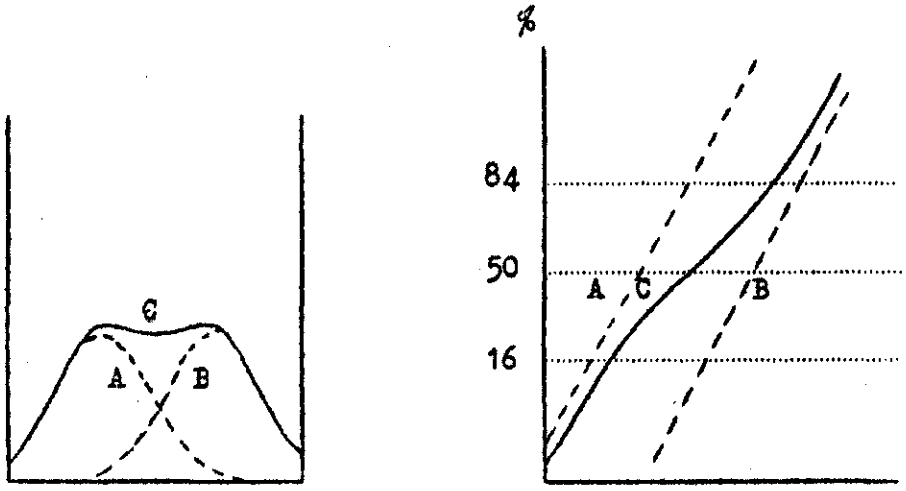
Grafiek XII



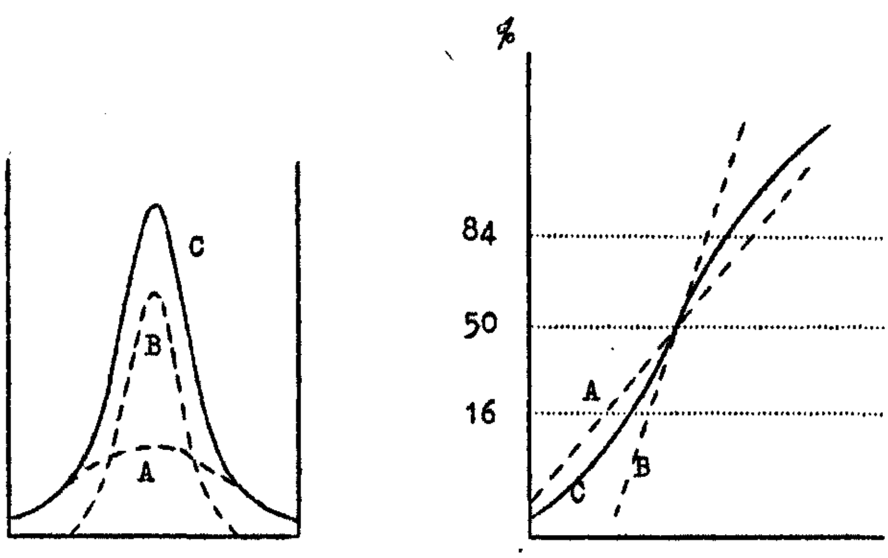
Grafiek XIII



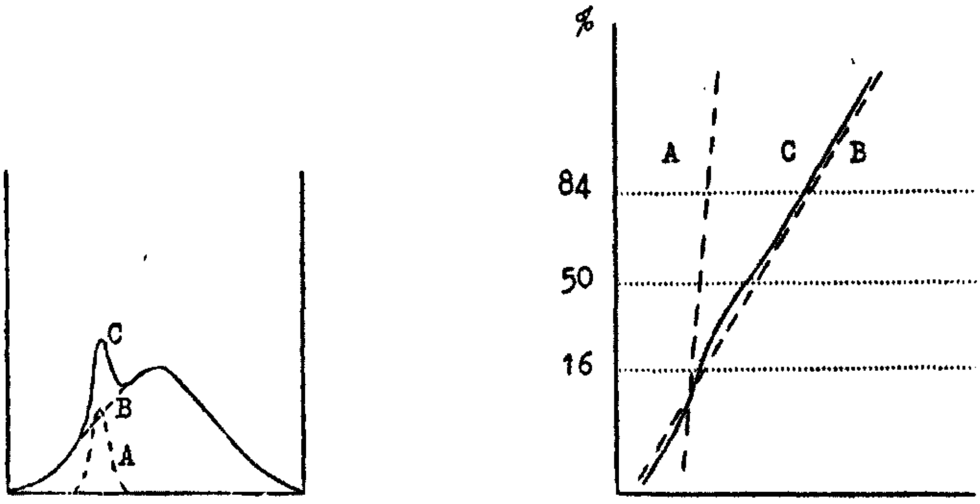
Grafiek XIV. Superpositie van 2 "normale" curven met gelijke standaard-afwijking, dooh verschillend gemiddelde.



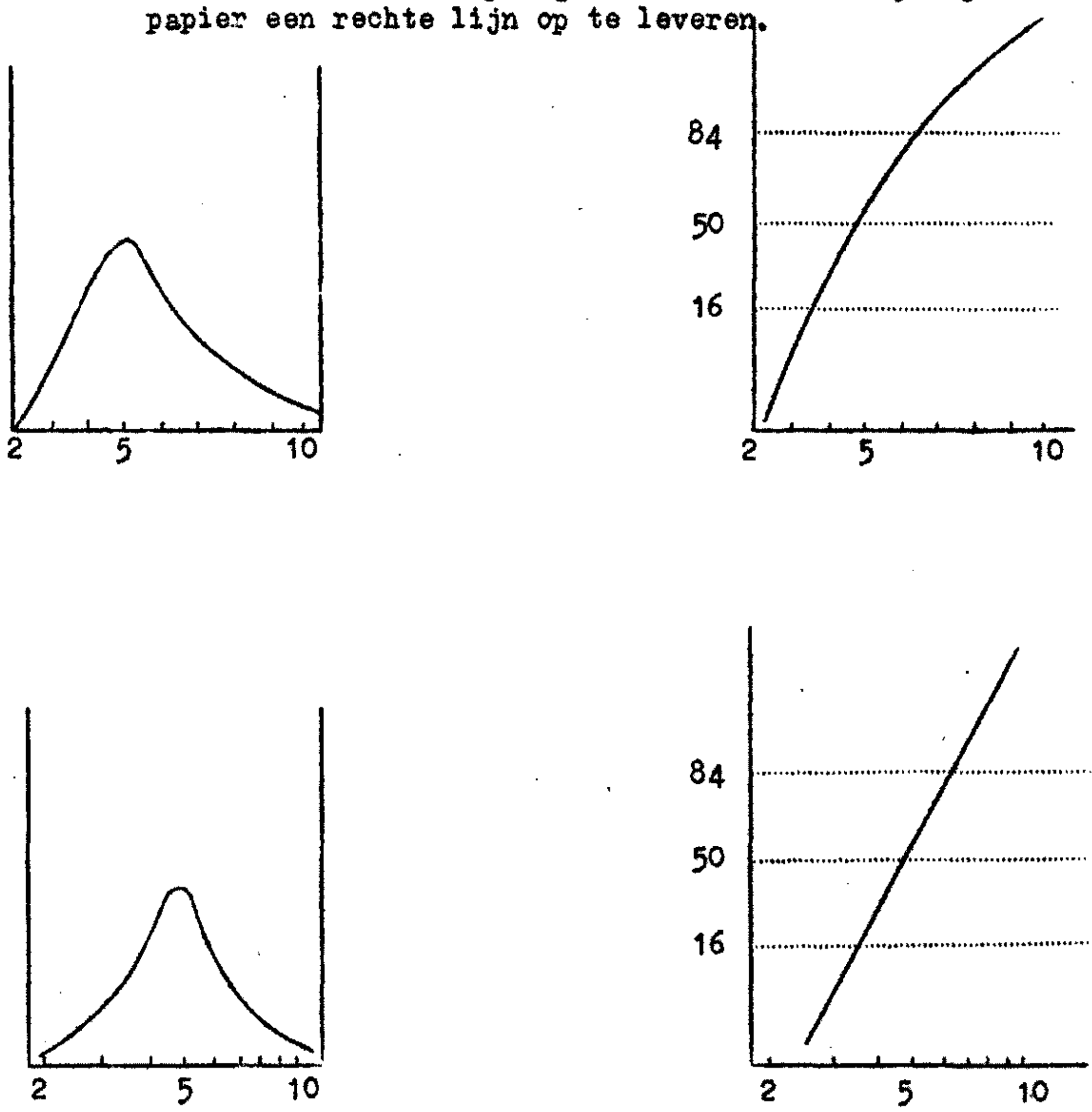
Grafiek XV. Superpositie van 2 "normale" curven met hetzelfde gemiddelde, doch met verschillende standaardafwijking.



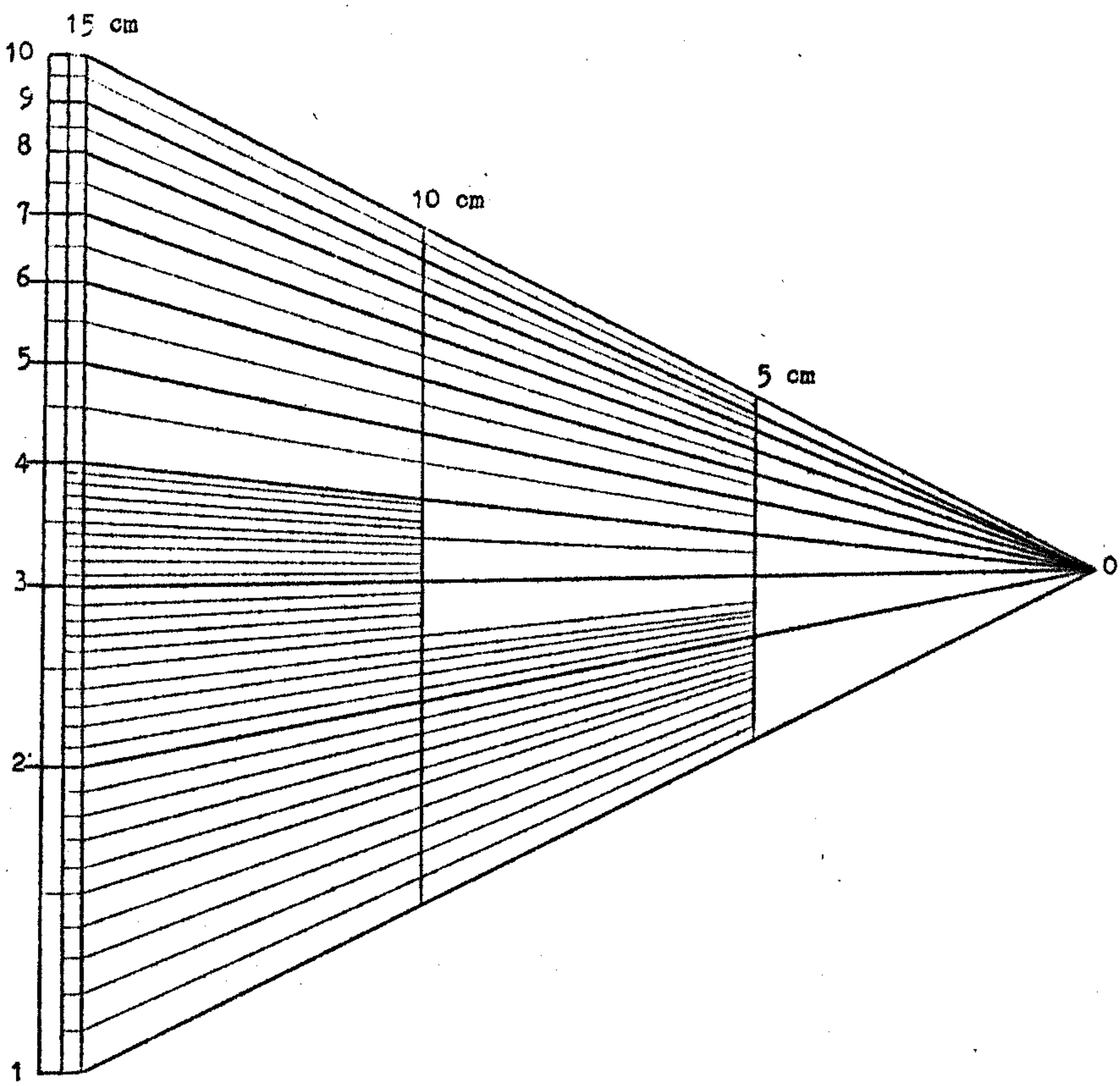
Grafiek XVI. Superpositie van 2 "normale" curven met verschillend gemiddelde en verschillende standaardafwijking.



Grafiek XVII. Een scheve curve op metrisch papier geeft op metrisch waarschijnlijkheidspapier een gebogen lijn. Deze curve blijkt op logarithmisch papier een normale verdeling te vertonen en dus op logarithmisch waarschijnlijkheidspapier een rechte lijn op te leveren.



Grafiek XVIII. Logarithmische schaalverdeling



Grafiek XIX. Waarschijnlijkheidsverdeling