Een mathematisch-statistische studie over de nauwkeurigheid van het organoleptisch onderzoek van boter

G. POSTHUMUS

Dit proefschrift met stellingen van

GERBEN POSTHUMUS,

landbouwkundig ingenieur, geboren te Wolsum (Fr.) den 19den October 1909, is goedgekeurd door de promotoren

Dr M. J. VAN UVEN,

hoogleeraar in de wiskunde,

en

Ir B. VAN DER BURG,

hoogleeraar in de zuivelbereiding en de melkkunde.

De Rector-Magnificus der Landbouwhoogeschool,

MEES.

Wageningen, 9 Juni 1943.

K 2647

Een mathematisch-statistische studie over de nauwkeurigheid van het organoleptisch onderzoek van boter

PROEFSCHRIFT
TER VERKRIJGING VAN DEN GRAAD VAN
DOCTOR IN DE LANDBOUWKUNDE
OP GEZAG VAN DEN RECTOR-MAGNIFICUS
Dr W. C. MEES R.Azn., HOOGLEERAAR IN DE
STAATHUISHOUDKUNDE, DE STATISTIEK EN
HET NEDERLANDSCH AGRARISCH RECHT, TE
VERDEDIGEN TEGEN DE BEDENKINGEN VAN
EEN COMMISSIE UIT DEN SENAAT DER LAND-
BOUWHOOGESCHOOL TE WAGENINGEN, OP
WOENSDAG 14 JULI 1943 TE VIJFTIEN UUR

DOOR

G. POSTHUMUS

ZUTPHEN — W. J. THIEME & CIE — MCMXLIII

Oan 'e neitins fen myn Heit.
Oan myn Mem, Frou en Bern.

Bij het voltooien van dit proefschrift breng ik dank aan allen, die tot mijn wetenschappelijke vorming aan de Landbouwhoogeschool hebben bijgedragen.

Daartoe richt ik mij in het bijzonder tot U, hooggeleerde van Uven en van der Burg, die zoo welwillend zijt geweest als mijn promotoren te willen optreden. Het verheugt mij, dat het onderwerp mijner dissertatie mij opnieuw met U beiden in contact heeft gebracht.

Hooggeleerde van Uven, met veel genoegen denk ik terug aan de lessen, die ik gedurende mijn studietijd van U mocht ontvangen. Ik breng U dank voor het feit, dat ik na het beëindigen van mijn studie nooit tevergeefs met verschillende wiskundig-statistische vragen bij U heb aangeklopt. Dat Uw colleges de richting van mijn wetenschappelijk denken voor een belangrijk deel hebben bepaald, bewijst dit proefschrift.

Van de opleiding, die ik bij U genoot, hooggeleerde van der Burg, mag ik bij mijn werk nog dagelijks profiteeren. Bij het afwerken van dit proefschrift hebben Uw critische opmerkingen mij opgewekt aan verschillende onderdeelen nog eens bijzondere aandacht te besteden.

Den besturen van den Geldersch-Overijselschen Bond van Coöperatieve Zuivelfabrieken te Zutphen en van de andere provinciale zuivelbonden en van den Algemeenen Nederlandschen Zuivelbond te 's Gravenhage ben ik zeer erkentelijk voor het feit, dat zij mij toestemming hebben willen verleenen om gegevens van de door hen georganiseerde boterkeuringen te verwerken.

Verder betuig ik mijn dank aan allen, die mij in eenigerlei opzicht bij het afwerken van dit proefschrift behulpzaam zijn geweest. In het bijzonder een woord van dank aan den heer L. Huts, die een deel van de zeer omvangrijke becijferingen, die voor dit proefschrift noodig waren, met groote accuratesse heeft uitgevoerd.

INLEIDING.

De bedoeling van dit geschrift is een poging te doen om een duidelijker inzicht te krijgen in de nauwkeurigheid, waarmede het organoleptisch onderzoek van boter geschiedt.

De opzet is zoodanig, dat in hoofdstuk I een omschrijving wordt gegeven van de reden en het doel van het onderzoek, waarna de niet ingewijde lezer op de hoogte wordt gebracht met de uitvoering van het organoleptisch onderzoek van boter.

In hoofdstuk II wordt dan een uitvoerige beschouwing gegeven over de nauwkeurigheid, waarmede het keuren van boter geschiedt. In dit hoofdstuk wordt speciale aandacht geschonken aan de nauwkeurigheid, waarmede een combinatie van twee of meer keurmeesters keurt. De nauwkeurigheid van den individueelen keurmeester wordt in hoofdstuk IV nader besproken.

In hoofdstuk III vindt de lezer een nadere rechtvaardiging van de toepassing van enkele mathematisch-statistische wetten, waarvan in hoofdstuk II zonder nader bewijs, of ze in ons geval wel gelden, gebruik werd gemaakt.

Hoofdstuk IV is in hoofdzaak een analyse van de nauwkeurigheid van den individueelen keurmeester. De hierin vervatte beschouwingen zijn in velerlei opzicht eveneens van toepassing op het in hoofdstuk II besprokene en kunnen zoodoende dienen om het inzicht in het geheele probleem te verhelderen en te verdiepen.

HOOFDSTUK I.

DOEL VAN HET ONDERZOEK EN BESCHRIJVING VAN DE WIJZE VAN BOTERKEUREN.

§ 1. Reden en doel van het onderzoek.

Door verschillende organisaties van zuivelfabrieken in Nederland worden periodiek (meestal wekelijks) boterkeuringen georganiseerd, waarbij de aangeslotenen dier organisaties in de gelegenheid worden gesteld monsters boter in te zenden, welke, na eenigen tijd bij een bepaalde temperatuur te zijn bewaard, door twee of meer boterkenners worden gekeurd.

De bedoeling van deze keuringen is den boterproducenten een beeld te geven van de kwaliteit van de door hen geproduceerde boter. In overeenstemming met den opzet van deze keuringen vindt een zeer nauwgezet organoleptisch onderzoek plaats, waarbij ook aan zeer kleine geur- en smaakverschillen aandacht wordt besteed. Het is nl. de bedoeling, dat de boterproducent komt te weten, welke gebreken — hoe klein ook — in zijn product aanwezig zijn, opdat hij kan trachten deze te doen verdwijnen, voordat zij zich in ernstiger mate gaan openbaren.

Deze keuringen zijn dus een onmisbaar hulpmiddel voor den boterproducent bij zijn pogingen de kwaliteit van de boter zoo hoog mogelijk op te voeren. Vandaar, dat het houden ervan van groot belang moet worden geacht.

De kwaliteit van de boter onderscheidt men gewoonlijk in drie onderdeelen:

1°. geur;
2°. smaak;
3°. gehalte-en-bewerking.

Onder gehalte-en-bewerking verstaat men de stevigheid, het uiterlijk aanzien der boter, de aanwezigheid van z.g. los vocht (met het oog waarneembare vochtdruppels, die bij zacht drukken uit de boter worden geperst).

Voor de drie onderdeelen worden volgens een bepaalde schaal punten gegeven, terwijl het eindoordeel zijn uitdrukking vindt in een combinatie van deze drie cijfers. Wij komen hierop in den loop van onze beschouwingen natuurlijk uitvoerig terug. Het is hier in deze inleiding slechts de bedoeling de groote lijnen aan te geven.

Het spreekt vanzelf, dat het toekennen van een bepaald aantal punten voor de drie genoemde onderdeelen tot onnauwkeurigheden leidt. Overal waar beoordeeld, gewogen of gemeten wordt, worden fouten gemaakt. Dit geldt bij het meten van temperaturen, lengten enz., maar dit geldt zeer zeker in veel ergere mate bij het keuren van boter. Immers bij het meten van natuurkundige grootheden zooals temperaturen, lengten enz. hebben wij telkens de beschikking over een objectieven maatstaf, die practisch weinig of niet verandert. Bij het keuren van boter ontbreekt echter een dergelijke maatstaf. Wij zijn hierbij vooralsnog geheel aangewezen op de zeer subjectieve beoordeeling door de keurmeesters. Er zijn geen andere middelen om geur en smaak te beoordeelen dan de reuk- en smaakorganen van den keurmeester. De keurmeester moet de gewaarwordingen, die tijdens het keuren van boter bij hem opkomen, in een cijfer omzetten. Het is te begrijpen, dat de hoogte van dat cijfer ten zeerste kan worden beinvloed door factoren, die geheel onafhankelijk zijn van den werkelijk aanwezigen geur en smaak.

In de eerste plaats zal de gevoeligheid van zijn smaak- en reukorganen van tijd tot tijd belangrijke wijzigingen kunnen ondergaan en sterk afhangen van zijn physischen en psychischen toestand op het moment van het keuren. In de tweede plaats is het allerminst zeker, dat hij een zelfde gewaarwording op elk tijdstip met een zelfde cijfer zal waardeeren.

Het zal dus aanstonds duidelijk zijn, dat het toekennen van een bepaald cijfer voor geur of smaak met vrij groote onnauwkeurigheden gepaard moet gaan.

Wat hierboven gezegd is voor geur en smaak, geldt in bijna even sterke mate voor gehalte-en-bewerking, voor zoover dit onderdeel door den keurmeester op het gevoel en op het gezicht wordt beoordeeld.

(Bij sommige boterkeuringen is het tegenwoordig gebruikelijk, dat een bepaald onderdeel van gehalte en bewerking, nl. de stevigheid, met een objectieven maatstaf wordt bepaald, en wel met een speciaal daarvoor geconstrueerd toestel.)

Het is merkwaardig, dat de nauwkeurigheid van de boterkeuring, die toch een zeer belangrijke factor is in het streven naar kwaliteitsverbetering van de boter, betrekkelijk weinig wetenschappelijk onderzocht is. Het is toch van groot belang de bepaling van de kwaliteit van de boter op haar nauwkeurigheid en reproduceerbaarheid te onderzoeken.

Het is gebruikelijk iedere bepaling, hetzij van een natuurkundige of scheikundige of andere grootheid, op haar reproduceerbaarheid en nauwkeurigheid te toetsen. Pas daarna kan men beoordeelen, welke waarde men aan verschillen moet toekennen.

Wij hebben in de litteratuur vrijwel geen enkele poging in deze richting kunnen ontdekken. Op enkele beschouwingen, die Pasveer[1]) over de nauwkeurigheid van de boterkeuring geeft, komen wij nog nader terug.

Het belang van een dergelijk onderzoek is blijkbaar door verschillende onderzoekers, die zich met het onderzoek van boter bezig hielden, ingezien. Zoo schrijft b.v. Rahn in het Handbuch der Milchwirtschaft (Verlag Julius Springer, Wien 1931), IIer Band, IIer Teil, blz. 96:

> „Eine Zusammenstellung des Umfanges der Abweichungen (nl. van de onderlinge waardeering der keurmeesters) ist dem Verfasser nicht bekannt; Material liegt genügend vor. Eine Übersicht über eine gröszere Zahl

1) A. Pasveer, Een statistisch onderzoek over de factoren, welke invloed uitoefenen op de kwaliteit van de Nederlandsche boter. Proefschrift, Wageningen 1941.

von Butterprüfungen mit einer Berechnung des wahrscheinlichen Fehlers würde bestimmt von Wert sein, denn die Ansichten über die Zuverlässigkeit der Butterbeurteilung gehen sehr weit auseinander, und der zahlenmäszige Beleg würde uns genau zeigen, was von der jetzigen Art des Richtens zu erwarten ist."

Ook indien men wil nagaan, of bepaalde omstandigheden (bijv. de toepassing van een andere werkwijze) invloed hebben op de kwaliteit van de boter en men daartoe twee vergelijkende reeksen proeven doet, is het noodig de waarde van het gevonden verschil te toetsen. Dit kan alleen, indien men weet, met welke mate van nauwkeurigheid de kwaliteit wordt vastgesteld. Zooals wij reeds opmerkten, is hieraan merkwaardigerwijs slechts zelden eenige aandacht besteed.

Nu zou men de reproduceerbaarheid van de boterkeuring op verschillende wijzen kunnen onderzoeken:

a. door na te gaan de nauwkeurigheid, waarmede de keurmeesters het *cijfer*, dat wordt toegekend, vaststellen;

b. door na te gaan, of de aanduiding van een door één of meer keurmeesters bij hetzelfde monster gevonden afwijking na herkeuring van dit monster terugkeert.

Degene, die zich aan de laatste taak zou wijden, zou zich — zeer zeker althans ten deele — moeten begeven op physiologisch en psychologisch terrein.

Wil men het geheele probleem der boterkeuring bespreken, dan zal zeker aan punt *b.* ook volledige aandacht moeten worden gewijd.

De bedoeling van dit geschrift is echter alleen aandacht te schenken aan punt *a.* Ons onderzoek zal dus niet een volledig onderzoek zijn over de boterkeuring in haar geheel. Het zal zich beperken tot mathematisch-statistische beschouwingen van het beschikbare cijfermateriaal.

Alvorens daartoe over te gaan, zal het echter wenschelijk zijn een omschrijving te geven, onder welke omstandigheden en op welke wijze het organoleptisch onderzoek van boter plaats heeft.

Wij moeten hier nog vooropstellen, dat de door ons te bespreken boterkeuringen die keuringen zijn, welke door de bonden van coöperatieve zuivelfabrieken in de verschillende provincies georganiseerd worden.

Naast deze keuringen worden nl. door nog andere instellingen en organisaties in Nederland boterkeuringen gehouden, en wel:

a. door verschillende coöperatieve verkooporganisaties van zuivelfabrieken voor de bij haar aangesloten zuivelfabrieken;

b. door het Zuivel-Kwaliteitscontrôle-Bureau (Z. K. B.).

Het karakter van deze keuringen is echter geheel anders dan dat van de bondskeuringen. Terwijl het — zooals reeds gezegd — bij de laatste de bedoeling is de aangesloten fabrieken aanwijzingen te geven over de kwaliteit van de boter (waarbij men op zeer minieme afwijkingen let), wordt bij de eerste slechts een grove onderscheiding in drieën gemaakt nl. in: 1°. zeer goede en goede boter; 2°. iets afwijkende boter; 3°. zeer afwijkende boter.

Wij zullen ons hier slechts bezighouden met de bondsboterkeuringen, hoewel natuurlijk onze beschouwingen in vele opzichten evenzeer op de andere boterkeuringen van toepassing zijn.

§ 2. Omschrijving van het organoleptisch onderzoek van boter. Wijze van uitvoering der keuring.

De zuivelfabrieken, welke deelnemen aan een boterkeuring, zenden van de productie van een bepaalden — in den regel in het reglement voor de boterkeuring vastgestelden — dag een monster ter grootte van enkele kilogrammen op naar de centrale organisatie, welke de monsters gedurende een bepaalden tijd bij een vastgestelde temperatuur bewaart en ze na afloop van dezen termijn door twee of meer boterkenners laat keuren. Vroeger was het gewoonte de boter na één week bewaring te keuren; thans wordt door verschillende organisaties een bewaringstijd van 12—14 dagen toegepast.

De temperatuur, waarbij de boter wordt bewaard, bedraagt in den regel 13°—14° C.

Eenige uren, voordat de keuring aanvangt, worden de monsters uit de koelruimte genomen en aan gewone kamertemperatuur blootgesteld.

Bij het keuren neemt de keurmeester met een boterboor een boorsel uit het monster en beoordeelt dit op de boor op geur (waarvoor hij de boor met de boter langs zijn neus beweegt), op smaak (door een klein stukje intensief te proeven), op gehalte-en-bewerking (door met den duim over het boormonster te strijken, waardoor hij een indruk krijgt van de stevigheid en de smeerbaarheid van de boter, terwijl hij tegelijkertijd waarneemt, of ook vochtdruppels uit de boter worden geperst en tevens nagaat, of het uiterlijk aanzien van de boter aan de eischen beantwoordt).

Zijn bevindingen drukt hij uit in een cijfer volgens een bepaalde puntenschaal, terwijl hij tevens noteert, welke afwijkingen hij constateert aan den geur, smaak of gehalte-en-bewerking.

Hieronder volgt een lijst met afwijkingen, welke aan de drie onderdeelen opgemerkt kunnen worden.

Geur en smaak

kooksmaak,
metaalsmaak,
verfsmaak,
voedersmaak (ingekuild gras, aardappelen, knollen, bieten),
watersmaak,
vettig, olieachtig, spekkig, visschig, tranig,
zuur, goor, kazig,
geil, bitter, sterk,
zeepig,
branderig,
ranzig,
schimmelig.

Gehalte-en-bewerking

zacht, week, slap,
overwerkt, zalvig, smerig,
brokkelig, kruimelig, hard, kort,
bont, vlammig, strepig,
nat, troebel vocht.

Direct kan hier reeds worden vermeld, dat deze lijst verre van volledig is. Het aantal nuances in de soort afwijkingen is blijkbaar zeer groot. Het spreekt vanzelf, dat het de vraag is, of inderdaad al deze benamingen essentieel verschillende gebreken aangeven. Door velen wordt dit betwijfeld. Hierbij komt, dat het voor den boterproducent verwarrend kan werken, indien de keurmeester een zeer groot aantal benamingen voor afwijkingen bezigt, terwijl de verschillende benamingen in vele gevallen slechts zeer weinig verschillende gewaarwordingen aanduiden. Er is dan ook herhaaldelijk getracht het aantal benamingen te beperken en eenige uniformiteit in de aanduiding der waargenomen afwijkingen te krijgen door verschillende keurmeesters een zelfde gebrek te laten beoordeelen. Hoewel hierdoor wel eenige resultaten zijn bereikt, is het niet gelukt een lijst van benamingen op te stellen, die door alle keurmeesters voldoende wordt geacht om hun gewaarwordingen weer te geven. Dat dit zoo is, is te begrijpen. Het aantal nuances is zoo groot en het geven van de benaming zoo zeer afhankelijk van den oogenblikkelijken indruk van den individueelen keurmeester, dat het ons niet behoeft te verwonderen, dat standaardisatie op dit gebied wel uitermate moeilijk is.

Wij noemen het feit, dat de keurmeester, behalve dat hij een cijfer toekent, zijn gewaarwording uitdrukt in een bepaalde benaming (indien ten minste een afwijking aanwezig is), slechts terloops. Zooals wij reeds aan het slot van § 1 mededeelden, is het niet de bedoeling ons hier verder mee bezig te houden.

Ons interesseert meer het uitdrukken van de beoordeeling in een cijfer.

De toegepaste puntenschaal.

Welke puntenschaal wordt nu meestal gebruikt? Tot voor enkele jaren was het in Nederland bij de door ons bedoelde boterkeuringen algemeen gewoonte voor geur, smaak en gehalte-en-bewerking een tiendeelige puntenschaal te gebruiken, waarbij de punten ongeveer de volgende beteekenis hebben.

1, 2	zeer slecht	6	voldoende
3	zeer onvoldoende	7	goed
4	onvoldoende	8	zeer goed
5	even onvoldoende	9, 10	uitmuntend

In de practijk wordt van deze puntenschaal in hoofdzaak alleen het traject van 3 tot en met 8 gebruikt.

Het is gebruikelijk elk monster steeds door ten minste twee keurmeesters te laten keuren. Het eindoordeel wordt nu gevonden door de cijfers voor geur, die door beide keurmeesters zijn gegeven, op te tellen bij tweemaal de som van de twee cijfers voor smaak en tweemaal de twee cijfers voor gehalte-en-bewerking. [1]

1) De hier beschreven puntenschaal wordt sinds korten tijd bij de bondsboterkeuringen niet meer toegepast.

Bij besluit van 18 Februari 1943 heeft de Overheid bepaald, dat van dien datum af bij alle in Nederland te houden boterkeuringen van de in Duitschland toegepaste puntenschaal gebruik dient te worden gemaakt.

Voor een beschrijving van het laatstgenoemde stelsel verwijzen wij den lezer naar blz. 8.

Bij deze berekeningswijze wegen dus de beoordeeling voor smaak en die voor gehalte-en-bewerking tweemaal zoo zwaar als die voor geur.

Voorbeeld. Aan een bepaald monster is toegekend

	door keurmeester I	door keurmeester II
voor geur	6	7
voor smaak	6	6
voor gehalte-en-bewerking	8	7

Het eindcijfer wordt dan: $(6 + 7) + 2(6 + 6) + 2(8 + 7) = 67$ punten.

Gewoonlijk beperkt de keurmeester zich tot het geven van heele cijfers, hoewel het ook voorkomt, dat hij met $\frac{1}{2}$ punten of zelfs met + en — werkt, teneinde aan te geven, dat hij het gegeven cijfer *iets* te hoog of te laag acht.

Direct moet hierbij worden vermeld, dat de keurmeesters in den regel hun cijfers in eerste instantie zonder onderling overleg vaststellen. Pas als blijkt, dat aanmerkelijke verschillen geconstateerd worden, heeft een *herkeuring* plaats, waarbij soms onderling overleg plaats vindt. Het spreekt vanzelf, dat wij bij onze — straks te bespreken — analyse van de nauwkeurigheid van de gegeven cijfers, zijn uitgegaan van de eerste cijfers, omdat anders een onjuist beeld van de nauwkeurigheid gegeven zou worden.

Iedere Bond volgt bij het beantwoorden van de vraag, wanneer een herkeuring dient plaats te hebben, bepaalde regels. Deze wijken soms vrij wat van elkaar af.

In den laatsten tijd is men bij enkele boterkeuringen iets van bovenstaand schema afgeweken, nl. door gehalte-en-bewerking niet in één cijfer uit te drukken, maar in twee, door zoowel voor stevigheid als voor gehalte-en-bewerking (afgezien van de stevigheid) een cijfer te geven. Het eindcijfer wordt dan gevonden door de som te nemen van de beide cijfers voor geur en tweemaal de som van de beide cijfers voor smaak en de som van de beide cijfers voor stevigheid en die voor gehalte-en-bewerking.

Voorbeeld. Aan een bepaald monster wordt toegekend

	door keurmeester I	door keurmeester II
voor geur	7	7
voor smaak	6	7
voor stevigheid	8	8
voor gehalte-en-bewerking	7	8

Het eindcijfer wordt:

$$(7 + 7) + 2(6 + 7) + (8 + 8) + (7 + 8) = 71 \text{ punten.}$$

De laatste variatie past men toe bij de boterkeuringen, welke door den Bond van Coöperatieve Zuivelfabrieken in Friesland worden georganiseerd.

De Commissie voor Kwaliteitsverbetering van den F.N.Z. heeft een andere puntenschaal aanbevolen (zie Officieel Orgaan van den F.N.Z. van 24 Januari 1940). Volgens deze schaal kunnen de volgende cijfers gegeven worden:

geur	1 — 3
smaak	1 — 6
stevigheid	1 — 3
bewerking	1 — 3

Het eindcijfer verkrijgt men door de punten, die de beide keurmeesters hebben gegeven, te middelen en de gemiddelden op te tellen. Maximaal kunnen 15 punten gegeven worden.

Voorbeeld. Aan een bepaald monster wordt toegekend

	door keurmeester I	door keurmeester II
voor geur	2	2
voor smaak	5	4
voor stevigheid	3	3
voor bewerking	2	3

Eindcijfer: $2 + 4{,}5 + 3 + 2{,}5 = 12$ punten.

De beteekenis der puntenschaal is de volgende:

	voor geur	*voor smaak*	*voor stevigheid*	*voor bewerking*
6		zeer goed		
5		goed		
4		voldoende		
3	goed	matig	stevig	goed
2	matig	onvoldoende	matig	matig
1	afwijkend	slecht	zacht (of te hard)	onvoldoende.

De bedoeling is, dat de keurmeesters van de geheele schaal gebruik maken. Dit in tegenstelling met de zooeven besprokene, waarvan practisch alleen maar het traject 3 tot en met 8 wordt gebruikt.

Het door den F.N.Z. aanbevolen stelsel werd alleen maar toegepast door den Brabantschen Zuivelbond te Breda.

Eenige in andere Europeesche landen gebruikte keuringsschalen.

Zonder hierin ook maar eenigermate naar volledigheid te streven (wij verwijzen den lezer daartoe o.a. naar het werkje „Butterprüfung" van A. Staffe en A. Weich, Wien, 1934) is het toch wenschelijk den lezer van enkele in andere landen toegepaste puntenschalen globaal op de hoogte te stellen.

Wij noemen daartoe:

a. De in Duitschland toegepaste puntenschaal.

Hieronder volgt de puntenschaal voor de verschillende onderdeelen:

	Geruch	Geschmack (Aroma, Reinheit, Salz)	Ausarbeitung (Buttermilch und Wassergehalt)	Gefüge (Härtegrad, Streichbarkeit)	Aussehen (Reinheit, Farbe, Schimmer)
10		hochfein			
9		fein			
8		gut			
7		ausreichend			
6		abfallend			
5					
4		schlecht			
3	hochfein		sehr gut		
2	fein		gut	gut (tadellos)	gut (tadellos)
1	genügend		ausreichend	mangelhaft	mangelhaft
0	schlecht		schlecht	schlecht	schlecht

De smaak speelt bij dit puntenstelsel dus wel een zeer overwegende rol. De cijfers worden vastgesteld door twee personen.

b. Het in Denemarken door de Staatsbotercontrôle toegepaste stelsel.

Hier wordt de kwaliteit niet onderscheiden in verschillende rubrieken. Men geeft dus alleen een cijfer voor den totalen indruk. Als hoogste cijfer kan 15 worden gegeven. Den fabrieken wordt het cijfer niet als zoodanig doorgegeven. Men berekent nl. het gemiddelde aantal punten van alle op een bepaalden dag gekeurde monsters en geeft aan de fabrieken het verschil met dit gemiddelde op. Indien dus bijv. een fabriek 0,3 punten onder het gemiddelde ligt, wordt opgegeven: 0,3 onder K (het gemiddelde wordt aangeduid met de letter K).

Men heeft deze wijze van noteeren ingevoerd, omdat niet iederen dag dezelfde keurmeesters keuren en de éene keurmeester gemiddeld wel eens een iets hooger of lager cijfer geeft dan de andere. Tevens bestaat er aanleiding om te veronderstellen, dat dezelfde keurmeesters ook niet iederen dag gelijk keuren. Men geeft daarom aan de fabrieken niet het gevonden cijfer.

Vermeldenswaard is verder, dat bij de hier bedoelde boterkeuringen de keuring geschiedt door negen keurmeesters. Het gemiddelde van de door deze 9 keurmeesters gegeven cijfers is het aan de fabriek toegekende cijfer.

HOOFDSTUK II.

DE NAUWKEURIGHEID VAN DE BOTERKEURING. (COMBINATIE VAN TWEE KEURMEESTERS).

Nu wij den lezer eenigermate een indruk hebben gegeven van de wijze, waarop de boterkeuring plaats heeft, kunnen wij thans overgaan tot de behandeling van het eigenlijke onderwerp.

§ 1. Het beschikbare materiaal.

Voor het onderzoek stond ons het volgende materiaal ter beschikking:

A. de keuringsuitslagen van de boterkeuring, die wekelijks door den Geldersch-Overijselschen Bond van Coöperatieve Zuivelfabrieken te Zutphen wordt georganiseerd. Wekelijks worden hier door drie keurmeesters 90—120 monsters beoordeeld, op deze wijze, dat bij aanwezigheid van 120 monsters keurmeester 1 de monsters 1—80, keurmeester 2 de monsters van 1—40 en 80—120 en keurmeester 3 de monsters van 40—120 keurt. Elk monster wordt dus door twee keurmeesters beoordeeld.

Daarnaast worden 15 monsters door alle drie keurmeesters gekeurd, teneinde de keurmeesters in staat te stellen hun oordeel onderling te vergelijken.

Wij hebben voor ons onderzoek in hoofdzaak gebruik gemaakt van de uitslagen van de bedoelde 15 monsters, omdat deze door *drie* keurmeesters waren gekeurd, waardoor een vergelijking van de beoordeeling van de drie keurmeesters mogelijk was. Tevens was er bij deze monsters voor gezorgd, dat die cijfers genoteerd bleven, die de keurmeesters in eerste instantie hadden gegeven (en niet die, welke verkregen werden na herkeuring en onderling overleg). Daar voor de monsters, die door twee keurmeesters waren gekeurd de eerste cijfers achterna niet altijd even gemakkelijk te verkrijgen waren en de 15 monsters, die per week door 3 keurmeesters waren beoordeeld, voldoende materiaal opleverden voor ons onderzoek, hebben wij ons in hoofdzaak hiertoe beperkt.

In dit geschrift is een gedeelte van het bedoelde materiaal over de jaren 1937 tot en met 1941 verwerkt en wel:

1. de uitslagen over verschillende kwartalen van de jaren 1937 tot en met 1941.
   (In het vervolg aan te geven als *materiaal A*1).
   Van deze gegevens is gebruik gemaakt in de tabellen 5, 50, 51, 53, 54, 58 en 59.
2. de uitslagen over het geheele jaar 1941.
   (In het volgende aan te duiden als *materiaal A*2).
   Deze gegevens zijn verwerkt in de tabellen 6, 11, 18, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 52, 55, 56 en 60.

B. de keuringsuitslagen van boterkeuringen, die mede op initiatief van den Algemeenen Nederlandschen Zuivelbond (F.N.Z.) te 's-Gravenhage

Keuring te	Datum van keuring	Aantal monsters	Aantal keurmeesters, van wie de gegevens verwerkt zijn:	Aanduiding van het materiaal	Tabel, waarin van het materiaal gebruik is gemaakt
Zutphen	4 Juni 1942	103	Geld. Ov. Bond v. Coöp. Zuivelfabrieken (G.O.Z.): 4 Bond v. Coöp. Zuivelfabrieken in Groningen en Drente (Groningsche en Drentsche Bond): 3	B1	1, 7, 12, 13, 14, 25, 26, 30, 34, 35, 44
Assen	3 Juni 1942	68	G.O.Z.: 3 Gron. en Dr. Bond: 2	B2	2, 8, 15, 27, 31
Leeuwarden	2 Juni 1942	81	Bond v. Coöp. Zuivelfabr. in Friesland (Friesche Bond): 2 Bond v. op coöp. grondslag werkende zuivelfabr. in Noord-Holland (Noord-Holl. Bond): 2	B3	10, 17, 29, 33
Alkmaar	12 Juni 1942	50	Noord-Holl. Bond: 2 Friesche Bond: 3	B4	3, 9, 16, 28, 32
Roermond	9 Juni 1942	34	Zuid-Ned. Zuivelbond (Z.N.Z.): 3 Brabantsche Zuivelbond: 3	B5	4

op verschillende plaatsen in ons land gedurende den zomer van 1942 werden gehouden en waaraan twee groepen van keurmeesters van verschillende provinciale zuivelbonden deelnamen.

Deze keuringen waren — voor zoover er althans in dit geschrift gebruik van is gemaakt — de volgende: zie blz. 11.

Hoewel dit materiaal bij lange na niet zooveel monsters omvat als het onder A genoemde, hebben wij er toch een ruim gebruik van gemaakt, omdat het ons gelegenheid bood na te gaan, of de conclusies, die wij trokken uit het materiaal, afkomstig van de boterkeuringen van den G.O.Z. te Zutphen, ook voor de andere boterkeuringen konden gelden.

Het vraagstuk van de nauwkeurigheid, waarmede de keurmeesters hun cijfers vaststellen, bestaat nu uit twee onderdeelen:

a. Geven de keurmeesters aan dezelfde monsters gemiddeld een gelijk aantal punten?

b. Hoe groot is de nauwkeurigheid van het cijfer, dat door de keurmeesters wordt toegekend?

Wij zullen eerst de vraag onder *a.* behandelen.

§ 2. De overeenstemming in het gemiddelde oordeel van verschillende combinaties van keurmeesters.

Het spreekt vanzelf, dat het voor de beantwoording van deze vraag er veel van afhangt, of de desbetreffende keurmeesters geregeld te zamen keuren of dat zij zulks nooit doen.

Een geschikte gelegenheid om het gemiddelde oordeel der verschillende keurmeesters onderling te vergelijken werd ons geboden, toen op aandrang van den F.N.Z. de keurmeesters van twee verschillende bonden op één plaats al de op dien dag aanwezige monsters beoordeelden. Deze gemeenschappelijke keuringen hadden, zooals onder § 1 reeds is medegedeeld, in den zomer van 1942 plaats. Bij deze keuringen werden — in tegenstelling met den gewonen gang van zaken — de monsters door zoo mogelijk alle keurmeesters (van beide bonden) gekeurd. Hierdoor is een vergelijking mogelijk.

Zoo werden bijv. op de bedoelde keuring te Zutphen 103 monsters door 4 keurmeesters van den Geldersch-Overijselschen Zuivelbond en door 3 van den Groningschen en Drentschen Zuivelbond gekeurd. Hierdoor was het mogelijk voor de Geldersch-Overijselsche keurmeesters 6 combinaties van 2 keurmeesters te vormen en voor de Groningsch-Drentsche 3 combinaties en van deze 9 combinaties het gemiddelde cijfer uit te rekenen, dat de verschillende combinaties voor geur, resp. smaak, gehalte-en-bewerking en „totaal aantal punten" hadden toegekend. Wij verkregen dus de volgende combinaties:

Geldersch-Overijselsche keurmeesters: 1—2, 1—3, 1—4, 2—3, 2—4 en 3—4.

Groningsch-Drentsche keurmeesters: 5—6, 5—7, 6—7.

Wij hebben geen combinaties gemaakt van Geldersch-Overijselsche en Groningsch-Drentsche keurmeesters, omdat dan natuurlijk het verschil in gemiddeld puntenaantal tusschen deze beide groepen van keurmeesters niet meer tot uiting zou komen. Bovendien komt een dergelijke combinatie in de practijk nooit voor.

Wij hadden ook het gemiddelde van ieder der 7 keurmeesters afzonderlijk kunnen uitrekenen, maar achtten dit minder noodzakelijk, omdat wij voor de beoordeeling van het uiteenloopen van het oordeel van iederen keurmeester afzonderlijk wel ander (straks te bespreken) materiaal ter beschikking hadden. Wij berekenden in dit geval het gemiddelde cijfer, dat door *twee* keurmeesters was gegeven (vormden dus telkens een combinatie van twee keurmeesters), omdat het steeds de gewoonte is, dat het uiteindelijke cijfer wordt vastgesteld door twee keurmeesters. Zoodoende sloten wij ons zoo dicht mogelijk aan bij wat in de practijk gewoonte is.

Het resultaat was nu als volgt:

Tabel 1. Gemiddelde cijfer, door een combinatie van twee keurmeesters gegeven.
Materiaal: B1 (*Zutphen*).
Aantal monsters: 103.
4 Geld.-Ov. keurmeesters (1, 2, 3, 4); 3 Gr.-Dr. keurmeesters (5,6,7)

Combinatie keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4	Gemidd. G.-O.keurm.	5—6	5—7	6—7	Gemidd. Gr.-Dr. keurm.
Gemiddeld cijfer voor											
Geur . . .	6,08	6,13	6,15	6,10	6,13	$6{,}17^5$	6,13	6,48	6,54	6,52	6,51
Smaak . . .	5,98	5,92	5,80	6,10	5,97	5,92	5,95	6,05	6,01	6,00	6,02
Geh.-en-Bew. . .	7,43	7,64	7,61	7,56	7,52	7,73	7,58	7,58	7,71	7,35	7,55
Totaal aantal punten . .	65,8	66,5	65,9	66,9	66,2	67,0	66,4	67,5	67,9	66,4	67,3

De combinatie 1—2 heeft dus gemiddeld voor de 103 monsters een cijfer van 6,08 voor geur, van 5,98 voor smaak en van 7,43 voor gehalte-en-bewerking en in totaal van 65,8 punt gegeven, enz.

In de kolom: „Gemiddelde Geldersch-Overijselsche keurmeesters" is het gemiddelde berekend van de cijfers, die door alle G.O.-keurmeesters zijn gegeven (dus de keurmeesters 1, 2, 3 en 4). Evenzoo is in de kolom „Gemiddelde Groningsche en Drentsche keurmeesters" hetzelfde cijfer voor de Gr.-Dr. keurmeesters vastgelegd.

Vanwege de zeer groote berekeningen, die noodig waren — ook voor het verdere deel van dit geschrift — hebben wij voor andere gemeenschappelijke keuringen niet steeds het gemiddelde berekend voor geur, smaak, gehalte-en-bewerking en totaal, maar hebben wij volstaan met de becijferingen uit te voeren voor smaak en totaal aantal punten, temeer daar ander (straks te noemen) materiaal ons voldoende in de gelegenheid stelde aan te toonen, dat de bevindingen ten aanzien van geur en gehalte-en-bewerking niet afweken van die bij smaak en totaal aantal punten.

Wij noemen nog het resultaat van de volgende gemeenschappelijke keuringen:

Tabel 2. Gemiddelde cijfer, door een combinatie van twee keurmeesters gegeven.
Materiaal: B2 (*Assen*).
Aantal monsters: 68.
2 Gr.-Dr. keurmeesters (1, 2); 3 Geld.-Ov. keurm. (3, 4, 5).

Combinatie van keurmeesters	1—2	3—4	3—5	4—5	Gemiddelde Geld.-Ov. keurmeesters
Gemiddeld cijfer voor:					
Smaak	6,13	5,86	5,70	5,90	5,82
Totaal aantal punten . . .	68,1	66,5	65,6	66,8	66,3

Tabel 3; als tabel 2.
Materiaal: B4 (*Alkmaar*).
Aantal monsters: 50.
(Voor „totaal aantal punten" 39 voor de combinaties (3, 4) en (4, 5).)
2 Noord-Holl. keurmeesters (1, 2); 3 Friesche keurmeesters (3, 4, 5).

Combinatie van keurmeesters	1—2	3—4	3—5	4—5	Gemiddelde Friesche keurmeesters
Gemiddeld cijfer voor:					
Smaak	6,04	5,65	5,53	5,66	5,61
Totaal aantal punten . . .	65,8	61,3	61,9	59,9	61,0

Tabel 4; als tabel 2.
Materiaal: B5 (*Roermond*).
Aantal monsters: 34.
3 Z.N.Z.-keurmeesters (1, 2, 3); 3 Brab. keurmeesters (4, 5, 6).

Combinatie van keurmeesters	1—2	1—3	2—3	Gemidd. Z.N.Z.-keurm.	4—5	4—6	5—6	Gemidd. Brab. keurm.
Gemiddeld cijfer voor:								
Smaak . . .	6,81	6,82	6,84	6,82	6,29	6,43	6,31	6,34
Totaal aantal punten . .	72,9	73,0	73,4	73,1	65,4	65,9	66,5	65,9

Uit bovenstaande mag o.i. geconcludeerd worden, dat het gemiddelde oordeel van de keurmeesters, die regelmatig te zamen keuren, zeer weinig verschilt, ten minste voor wat de smaak en het totale aantal punten betreft.

Men zou kunnen opmerken, dat het materiaal niet voldoende is, om een zelfde conclusie voor geur en gehalte-en-bewerking te trekken.

Onderstaand materiaal, dat ontleend is aan de wekelijksche keuringen van den Geldersch-Overijselschen Bond, laat echter zien, dat ook voor geur en gehalte-en-bewerking de conclusie mag gelden. Tevens bood dit materiaal gelegenheid om eenigen indruk te geven van de mate, waarin het gemiddelde cijfer voor *één* keurmeester kan varieeren.

Op de boterkeuring van den Geldersch-Overijselschen Zuivelbond worden (zooals reeds in § 1 werd medegedeeld) elke week 15 monsters door alle drie keurmeesters gekeurd. Dit laatste materiaal nu (15 monsters per week) biedt een welkome gelegenheid het oordeel van de keurmeesters onderling te vergelijken. Daar echter niet elke week dezelfde drie keurmeesters keuren (er zijn in totaal vier keurmeesters, waarvan er telkens slechts drie keuren), kan door optelling van de door hen gegeven punten geen juiste vergelijking worden verkregen van de gemiddelde cijfers, die de drie keurmeesters hebben gegeven. Wij volgden daarom een iets anderen weg.

Wij berekenden voor de 15 monsters het gemiddelde van de cijfers, die de drie keurmeesters hadden gegeven en noteerden het verschil van het cijfer van den individueelen keurmeester met dit gemiddelde, bijvoorbeeld:

	Cijfer voor geur					Afwijking van het gemiddelde			
Keurmeester:	1	2	3	4	Gemidd.	1	2	3	4
30ste keuring									
No. 1	6	7	6		$6\frac{1}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	
No. 2	5	6	6		$5\frac{2}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$-\frac{1}{3}$	
.	.	.	.		.	.	.	.	
.	.	.	.		.	.	.	.	
.	.	.	.		.	.	.	.	
No. 15	7	6	7		$6\frac{2}{3}$	$+\frac{1}{3}$	$-\frac{2}{3}$	$+\frac{1}{3}$	
31ste keuring									
No. 1	6		7	7	$6\frac{2}{3}$	$-\frac{2}{3}$		$+\frac{1}{3}$	$+\frac{1}{3}$
No. 2	7		6	6	$6\frac{1}{3}$	$+\frac{2}{3}$		$-\frac{1}{3}$	$-\frac{1}{3}$
.	.		.	.	.	.		.	.
.	.		.	.	.	.		.	.
.	.		.	.	.	.		.	.
No. 15	4		4	5	$4\frac{1}{3}$	$-\frac{1}{3}$		$-\frac{1}{3}$	$+\frac{2}{3}$
			enz.				enz.		

Wij berekenden dit voor een kwartaal ($\pm$ 13 keuringen) en telden de verschillen, die de cijfers van den individueelen keurmeester met het gemiddelde van de drie keurmeesters hadden vertoond, op en middelden deze. De gemiddelden, die op deze wijze voor ieder der keurmeesters werden berekend, zijn een maat voor de wijze, waarop de keurmeesters zich in hun gemiddeld oordeel aaneensluiten.

Wij noemen hier enkele voorbeelden (zie tabel 5, blz. 17).

Verder berekenden wij voor de monsters, die in 1941 door 3 keurmeesters waren gekeurd, de gemiddelde verschillen tusschen het oordeel van twee keurmeesters.

Het resultaat was het volgende.

Tabel 6. Gemiddeld verschil tusschen het oordeel van twee keurmeesters over 1941.

Materiaal: A2 (Zutphen).
4 Geld.-Ov. keurmeesters (1, 2, 3, 4).

Combinatie keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4
Aantal monsters	405	390	340	405	380	365
Gemidd. verschil voor smaak	$-0,22$	$-0,08$	$+0,09$	$+0,14$	$+0,28$	$+0,15$
Gemidd. verschil voor totaal aantal punten	$-0,54$	$-0,16$	$+0,23$[1]	$+0,34$	$+0,45$	$+0,02$

Hiermede kan wel als bewezen worden beschouwd, dat keurmeesters, die regelmatig te zamen keuren, in hun gemiddelde oordeel zeer dicht bij elkaar liggen.

Zoo straks zullen wij vinden, dat de middelbare fout van de verschillen voor smaak, waarvan de gemiddelden in de laatste tabel zijn aangegeven, ongeveer 0,8 bedraagt. Hieruit volgt, dat de middelbare fout van het *gemiddelde* verschil (berekend uit $\pm$ 400 monsters) bedraagt

$$\frac{0,8}{\sqrt{400}} = 0,04.$$

Voor „totaal aantal punten" vinden wij op dezelfde wijze

$$\frac{5,5}{\sqrt{400}} \cong 0,3.$$

1) Hier waren 355 monsters.

Tabel 5. Gemiddelde afwijking voor iederen keurmeester (1, 2, 3, 4) van het gemiddelde van 3 keurmeesters over een kwartaal.

Materiaal: A1 (Zutphen).

	Geur				Smaak				Geh.-en-Bew.				Totaal aant. punten			
Keurmeester	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
4e kwart. 1939																
Aant. monsters	150	150	150	135	150	150	150	135	150	150	150	135	—	—	—	—[1])
Gem. verschil .	—0,02	+0,04	—0,10	+0,10	—0,12	+0,17	—0,02	—0,08	+0,08	—0,07	—0,02	0,00	—	—	—	—
1e kwart. 1940																
Aant. monsters	135	150	150	150	135	150	150	150	135	150	150	150	135	150	150	150
Gem. verschil .	0,00	—0,14	—0,04	+0,19	—0,11	+0,08	+0,08	—0,06	+0,06	—0,09	+0,03	—0,01	—0,07	—0,29	+0,36	+0,01
3e kwart. 1940																
Aant. monsters	105	100	110	90	105	100	110	90	105	100	110	90	105	100	110	90
Gem. verschil .	—0,08	—0,06	—0,07	+0,21	—0,13	+0,10	+0,09	—0,06	+0,04	—0,14	+0,09	+0,06	—0,50	—0,33	+0,38	+0,50
4e kwart. 1940																
Aant. monsters	65	85	80	85	65	85	80	85	65	85	80	85	65	85	80	85
Gem. verschil .	0,00	—0,11	—0,09	+0,20	—0,05	+0,03	+0,08	—0,07	—0,05	0,00	+0,01	+0,03	—0,42	—0,11	+0,13	+0,27
1e kwart. 1941																
Aant. monsters	130	125	140	130	130	125	140	130	130	125	140	130	130	125	140	130
Gem. verschil .	0,00	—0,05	—0,11	+0,17	0,00	+0,07	+0,04	—0,11	+0,03	+0,04	+0,01	—0,09	+0,12	+0,27	+0,04	—0,43

1) Het cijfer voor „totaal aantal punten" voor het 4de kwartaal 1939 is niet berekend.

Er zijn dus principieele verschillen tusschen het gemiddelde oordeel van sommige keurmeesters, ten minste voor de drie onderdeelen. Voor „totaal aantal punten" zijn de verschillen onbelangrijk, terwijl ook de verschillen voor de onderdeelen, hoewel principieel van beteekenis, practisch zeer gering zijn.

Uit het overzicht op de vorige bladzijde (tabel 5), waarin het gemiddelde verschil van denzelfden keurmeester eenige kwartalen achtereen voorkomt, blijkt, dat de éene keurmeester soms voortdurend neiging heeft om gemiddeld iets lager of hooger te keuren dan het gemiddelde van hun drieën. Het verschil is echter klein en ligt lang niet altijd in dezelfde richting.

De vraag, of bovenbedoelde uitspraak (nl. dat het gemiddelde oordeel der keurmeesters zeer weinig uiteenloopt) ook kan gelden voor keurmeesters, die niet geregeld te zamen keuren, wordt ook door het materiaal, afkomstig van de gemeenschappelijke keuringen, waarover wij op blz. 11 spraken, beantwoord.

Hier is het verschil tusschen het gemiddelde oordeel van de beide — bij deze gelegenheid te zamen beoordeelende — groepen keurmeesters in sommige gevallen zoo groot (bijv. bij de keuring te Alkmaar en te Roermond), dat men in deze gevallen tot een zeer duidelijk werkelijk bestaand verschil moet concludeeren.

Wij komen thans tot de tweede en hoofdvraag. In hoeverre staat het oordeel van de keurmeesters vast? Welke verschillen zijn mogelijk, wanneer een combinatie van twee keurmeesters een zelfde monster vele keeren beoordeelt? Hoe groot is de nauwkeurigheid, waarmede op de keuring het cijfer voor geur, smaak enz. wordt vastgesteld (dus door een combinatie van twee keurmeesters)?

§ 3. De middelbare fout van de boterkeuring.

Om over de nauwkeurigheid, waarmede een combinatie van twee keurmeesters het cijfer voor de verschillende onderdeelen (geur, smaak, gehalte-en-bewerking) en het totale aantal punten vaststelt, een oordeel te krijgen, hebben wij de middelbare fout van dit cijfer bepaald.

Voor het berekenen van deze middelbare fout werd de volgende methode toegepast.

Aangenomen is, dat in totaal N monsters door 2 keurmeesters worden gekeurd en dat keurmeester 1 aan het k^{de} monster (k loopend van 1 tot en met N), het cijfer $x_k^{(1)}$ toekent, keurmeester 2 het cijfer $x_k^{(2)}$.

Het uiteindelijke cijfer, dat wordt toegekend, bedraagt het gemiddelde van deze beide cijfers, dus:

$$\bar{x}_k = \frac{x_k^{(1)} + x_k^{(2)}}{2} \quad \ldots\ldots\ldots\ldots \quad (1)^{1)}$$

1) Zooals reeds eerder medegedeeld, nemen wij voor $x_k^{(1)}$ en $x_k^{(2)}$ steeds het cijfer, dat in eerste instantie is gegeven en niet het cijfer, dat na herkeuring en onderling overleg is verkregen.

(Het gemiddelde wordt steeds aangegeven door een streep boven de betreffende letter te plaatsen.)

Daar wij onder de middelbare fout van de keuring zullen verstaan de middelbare fout van het gemiddelde cijfer, dat door *twee* keurmeesters is gegeven, kunnen wij de middelbare fout van de keuring aangeven door

$\sigma_{\bar{x}}$ (middelbare fout van $\bar{x}$).

$\sigma_{\bar{x}}$ kunnen wij nu als volgt uitdrukken. Wij noemen het „ware" cijfer, dat de ideale keurmeester aan het k^{de} monster zou toekennen $\hat{x}_k$. („Ware" grootheden geven wij steeds aan door een kapje (∧) boven de letter te plaatsen.) Het „ware" cijfer ($\hat{x}_k$) kennen wij natuurlijk niet. Wij zouden het kunnen benaderen door een zeer groot aantal (eigenlijk: oneindig aantal) keurmeesters hetzelfde monster te laten keuren. Het gemiddelde cijfer, dat wij hierbij zouden verkrijgen, zou het „ware" cijfer zijn.

De uitdrukking voor $\sigma_{\bar{x}}$ wordt nu:

$$< (\bar{x}_k - \hat{x}_k)^2 > = \sigma^2_{\bar{x}} \quad \ldots\ldots\ldots\ldots \quad (2)$$

Het gebruikte symbool $< >$ geeft aan, dat wij veronderstellen, dat ons een oneindig aantal waarnemingen ter beschikking staan (dus $N = \infty$). Wij mogen formule (2) dus ook als volgt schrijven:

$$\sigma^2_{\bar{x}} = \lim_{N=\infty} \frac{\sum_{k=1}^{N} (\bar{x}_k - \hat{x}_k)^2}{N} = < (\bar{x}_k - \hat{x}_k)^2 >.$$

De moeilijkheid is, dat wij noch de ware waarden ($\hat{x}_k$) kennen, noch een oneindig aantal waarnemingen tot onze beschikking hebben. Op deze wijze is $\sigma_{\bar{x}}$ dus niet te berekenen.

Het bezwaar, dat wij slechts over een beperkt aantal waarnemingen beschikken, is niet zoo groot. In het volgende zullen wij nl. steeds met vrij groote waarden voor N werken, zoodat wij slechts een zeer kleine fout maken, als wij stellen:

$$\sigma^2_{\bar{x}} = \frac{\sum_{1}^{N} (\bar{x}_k - \hat{x}_k)^2}{N} \quad \ldots\ldots\ldots\ldots \quad (3)$$

Het feit, dat wij $\hat{x}_k$ niet kennen, is veel ernstiger. Dit maakt, dat berekening van $\sigma_{\bar{x}}$ via formule (3) onmogelijk is.

Er is echter een andere weg om de waarde van σ te benaderen en wel als volgt.

Ingevolge (1) is:

$$\bar{x}_k - \hat{x}_k = \frac{x_k^{(1)} + x_k^{(2)}}{2} - \hat{x}_k = \frac{x_k^{(1)} - \hat{x}_k}{2} + \frac{x_k^{(2)} - \hat{x}_k}{2}$$

Stellen wij:

$$x_k^{(1)} - \hat{x}_k = t_k^{(1)}$$
$$x_k^{(2)} - \hat{x}_k = t_k^{(2)}$$

waarbij $t^{(1)}$ en $t^{(2)}$ dus de ware fouten van $x^{(1)}$ en $x^{(2)}$ zijn, dan geldt:

$$\bar{x}_k - \hat{x}_k = \frac{t_k^{(1)} + t_k^{(2)}}{2} \quad \ldots\ldots\ldots\ldots (4)$$

(4) gekwadrateerd, levert:

$$(\bar{x}_k - \hat{x}_k)^2 = \tfrac{1}{4}\,(t_k^{(1)^2} + 2t_k^{(1)}\, t_k^{(2)} + t_k^{(2)^2})$$

en gesommeerd over een oneindig aantal waarnemingen:

$$< (\bar{x}_k - \hat{x}_k)^2 > = \tfrac{1}{4}\{< t_k^{(1)^2} > + 2 < t_k^{(1)}\, t_k^{(2)} > + < t_k^{(2)^2} >\} \quad \ldots (5)$$

Daar $t_k^{(1)}$ en $t_k^{(2)}$ als onafhankelijk beschouwd mogen worden, geldt:

$$< t_k^{(1)}\, t_k^{(2)} > = 0.$$

Stellen wij verder:

$$< t_k^{(1)^2} > = \sigma_1^2 \quad \text{en} \quad < t_k^{(2)^2} > = \sigma_2^2$$

waarbij σ_1 resp. σ_2 de middelbare fout is van het cijfer voor keurmeester 1 resp. 2, dan volgt uit (5) en (2) gemakkelijk:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_1^2 + \sigma_2^2}{4} \quad \ldots\ldots\ldots\ldots (6)$$

Ook σ_1 en σ_2 zijn ons echter niet bekend.

Wanneer wij echter het verschil tusschen de beide beoordeelingen van de keurmeesters nemen (dit verschil noemen we V), dan geldt:

$$V_k = x_k^{(1)} - x_k^{(2)} = (x_k^{(1)} - \hat{x}_k) - (x_k^{(2)} - \hat{x}_k) = t_k^{(1)} - t_k^{(2)} \quad \ldots (7)$$

(7) gekwadrateerd, levert:

$$V_k^2 = t_k^{(1)^2} - 2\, t_k^{(1)}\, t_k^{(2)} + t_k^{(2)^2}.$$

Gesommeerd over een oneindig aantal waarnemingen:

$$< V_k^2 > = < t_k^{(1)^2} > - 2 < t_k^{(1)}\, t_k^{(2)} > + < t_k^{(2)^2} > \quad \ldots (8)$$

Daar $< t_k^{(1)}\, t_k^{(2)} > = 0$, gaat (8) over in:

$$< V_k^2 > = < t_k^{(1)^2} > + < t_k^{(2)^2} > \quad \ldots\ldots\ldots (9)$$

en omdat $< t_k^{(1)^2} > = \sigma_1^2$ en $< t_k^{(2)^2} > = \sigma_2^2$

$$< V_k^2 > = \sigma_1^2 + \sigma_2^2. \quad \ldots\ldots\ldots\ldots (10)$$

Door combinatie van (6) en (10) krijgen wij:

$$\sigma_{\bar{x}}^2 = \frac{< V_k^2 >}{4} \quad \ldots\ldots\ldots\ldots (11)$$

Voor $< V_k^2 >$ mogen wij — bij voldoend groote N — schrijven:

$$< V_k^2 > \simeq \frac{1}{N} \sum_{k=1}^{N} V_k^2 = m_V^2. \quad \ldots\ldots\ldots (12)$$

(Op het feit, dat wij in bovenstaande formule den factor N gebruiken en niet de gebruikelijke $N - 1$, komen wij dadelijk terug.)

Uit (11) en (12) volgt:

$$\sigma_{\bar{x}}^2 = \frac{m_V^2}{4} \quad \text{of} \quad \sigma_{\bar{x}} = \frac{m_V}{2} \quad \ldots\ldots\ldots (13)$$

De waarde van m_V kunnen wij berekenen uit het beschikbare materiaal. $\sigma_{\bar{x}}$ is dan dus ook bekend.

Wij merkten zoo even reeds op, dat wij in formule (12) de waarde van de middelbare fout hebben berekend door de som te nemen van de kwadraten van de afwijkingen, deze som *te deelen door N* en hieruit den wortel te trekken. Wij hebben dus niet den gebruikelijken factor $N - 1$ gebruikt.

Op het resultaat heeft dat geen merkbaren invloed gehad. Wij werken steeds met groote of vrij groote waarden van N, zoodat het practisch geen verschil maakt, of men den factor N of $N - 1$ gebruikt. Overigens voelen wij vrij weinig voor het gebruik van den factor $N - 1$. Wil men een van N afwijkende waarde gebruiken, dan doet men beter $N - 2$ toe te passen.

Deze meening gronden wij op het volgende:

De waarschijnlijkheidsrekening leert, dat de schijnbare middelbare fout s, welke berekend wordt uit N waarnemingen van de grootheid x en welke wij dus kunnen voorstellen door:

$$s^2 = \frac{\sum\limits_{k=1}^{N} (x_k - \bar{x})^2}{N} \quad \ldots\ldots\ldots\ldots\ldots\ldots \text{(A)}$$

de volgende kansformule heeft:

$$dW(s) = f(s)\,.\,ds = \frac{N^{\frac{N-1}{2}}}{\sigma^{N-1} \times 2^{\frac{N-3}{2}} \times \Gamma\left(\frac{N-1}{2}\right)}\, s^{N-2}\, e^{-\frac{N}{2\sigma^2}s^2}\, ds$$

Hierin stelt σ de „ware" middelbare fout voor ($\sigma = \lim\limits_{N=\infty} s$)

De meest waarschijnlijke waarde voor s (s_w) vinden wij, als wij nagaan, voor welke waarde van s $f(s)$ haar maximumwaarde bereikt.

Dan is: $\dfrac{df}{ds} = 0$

Stellen wij: $\dfrac{N^{\frac{N-1}{2}}}{\sigma^{N-1} \times 2^{\frac{N-3}{2}} \times \Gamma\left(\frac{N-1}{2}\right)} = k$, dan wordt:

$$f(s) = k.\, s^{N-2}\, e^{-\frac{N}{2\sigma^2}s^2}$$

$$\frac{df}{ds} = k\left\{s^{N-2} \times -\frac{N}{2\sigma^2} \times 2s \times e^{-\frac{N}{2\sigma^2}s^2} + e^{-\frac{N}{2\sigma^2}s^2} \times (N-2)s^{N-3}\right\} = 0$$

$$\text{of } -\frac{N}{\sigma^2} \times s^{N-1} + s^{N-3} \times (N-2) = 0$$

$$\text{of } -\frac{N}{\sigma^2} \times s^2 + (N-2) = 0.$$

$$\text{of } s_w = \sigma \times \sqrt{\frac{N-2}{N}} \quad \ldots\ldots\ldots\ldots\ldots \text{(B1)}$$

en omgekeerd:

$$\sigma = s_w \times \sqrt{\frac{N}{N-2}} \quad \ldots\ldots\ldots\ldots\ldots \text{(B2)}$$

Wij moeten nu aannemen, dat de door ons waargenomen waarde van s $\left(s^2 = \dfrac{\Sigma\,(x_k - \bar{x})^2}{N}\right)$ de waarschijnlijkste waarde voor s is ($s = s_w$).

Hieruit volgt (door (A) en (B2) te combineeren), dat:

$$\sigma^2 = \frac{\sum_{k=1}^{N} (x_k - \bar{x})^2}{N} \times \frac{N}{N-2} = \frac{\sum_{k=1}^{N} (x_k - \bar{x})^2}{N-2}.$$

Hiermede is aangetoond, dat men de beste schatting van de ware waarde (σ) krijgt, als men den factor: $N-2$ toepast.

Zoo even merkten wij op, dat de waarde van m_V gemakkelijk te berekenen is en daaruit met behulp van formule (13) $\sigma_{\bar{x}}$. Immers wij nemen slechts op, hoeveel keeren een bepaald verschil tusschen twee keurmeesters is voorgekomen, bijv. bij de keuring te Zutphen voor de cijfers voor *smaak*.

Voor smaak.

Tabel 7. Frequentie van een bepaald verschil bij verschillende combinaties van keurmeesters voor *smaak* en de nauwkeurigheid van deze combinaties.

Materiaal: B1 (*Zutphen*).

4 Geld.-Ov. keurm. (1, 2, 3, 4); 3 Gr.-Dr. keurm. (5, 6, 7).

Combinatie van keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4	Totaal	5—6	5—7	6—7	Totaal
Gevonden verschillen											
0	37	43	55	42	44	54	275	46	44	49	139
1	53	53	42	49	50	41	288	47	53	48	148
2	12	7	6	12	7	8	52	10	6	6	22
3	1				2		3				
4											
Totaal	103	103	103	103	103	103	618	103	103	103	309
m^2_V	0,985	0,703	0,558	0,859	0,849	0,626	Gem. 0.763	0,762	0,665	0,617	Gem. 0,681
$\sigma^2_{\bar{x}}$	0,246	0,176	0,140	0,215	0,212	0,157	0,191	0,191	0,166	0,154	0,170
$\sigma_{\bar{x}}$	0,50	0,42	0,37	0,46	0,46	0,39	0,44[1])	0,44	0,41	0,39	0,41[1])

m_V en $\sigma_{\bar{x}}$ zijn uit bovenstaand materiaal als volgt berekend:

Als voorbeeld nemen wij de eerste kolom (combinatie 1 en 2).

---

1) $\sigma_{\bar{x}}$ is voor de kolom: „gemiddeld" berekend door den wortel te trekken uit het gemiddelde van de waarden van $\sigma^2_{\bar{x}}$ voor de verschillende combinaties.

Aantallen		Verschil		
37	$\times$	0^2	=	0
53	$\times$	1^2	=	53
12	$\times$	2^2	=	48
1	$\times$	3^2	=	9

$$\text{Som der kwadraten } \left(\sum_{k=1}^{N} V_k^2\right) = 110$$

$$\text{Gemiddeld kwadraat } \frac{110}{103} = 1{,}068 \left(= \frac{\Sigma V_k^2}{N}\right)$$

$$\text{1e correctie van Sheppard [1]} = \underline{0{,}083}$$

$$m_V^2 = 0{,}985$$

$$\sigma_{\bar{x}}^2 = \frac{m_V^2}{4} = 0{,}246 \quad \text{of} \quad \sigma_{\bar{x}} = 0{,}50.$$

Op volkomen dezelfde wijze zijn ook voor de andere combinaties van keurmeesters de grootheden m_V^2, $\sigma_{\bar{x}}^2$ en $\sigma_{\bar{x}}$ berekend.

Wij noemen hieronder ook de resultaten voor de nauwkeurigheid voor smaak voor de andere gemeenschappelijke keuringen.

Tabel 8. Nauwkeurigheid van combinaties van twee keurmeesters voor *smaak*.
Materiaal: B2 (*Assen*).
Aantal monsters: 68.
2 Gr.-Dr. keurm. (1, 2); 3 Geld.-Ov. keurm. (3, 4, 5).

	Combinatie keurmeesters				Gemiddelde (van de laatste drie kolommen)
	1 — 2	3 — 4	3 — 5	4 — 5	
m_V^2	0,608	0,608	0,711	0,843	0,721
$\sigma_{\bar{x}}^2$	0,152	0,152	0,178	0,211	0,180
$\sigma_{\bar{x}}$	0,39	0,39	0,42	0,46	0,42

1) Deze correctie is noodig, omdat de keurmeesters als cijfer slechts geheele getallen toekennen. Alle verschillen tusschen de keurmeesters worden dus ook aangegeven door geheele getallen. Een verschil, dat in wezen ligt tusschen $1\frac{1}{2}$ en $2\frac{1}{2}$ punten, wordt genoteerd als een verschil van 2. Daar de verschillen bij toenemende grootte minder frequent worden, zijn de verschillen van 2 in werkelijkheid gemiddeld iets geringer. Hetzelfde geldt voor de verschillen van 1, 3 enz.

$\frac{\Sigma V_k^2}{N}$ wordt dus iets te groot berekend.

De correctie, die hiervoor wordt toegepast, bedraagt $\frac{c^2}{12}$, waarbij c het klasse-interval is.

Daar het klasse-interval hier 1 bedraagt, wordt de correctie $\frac{1}{12} = 0.083$.

Deze correctie wordt de 1e correctie van Sheppard genoemd.

Tabel 9. Nauwkeurigheid van combinaties van twee keurmeesters voor *smaak*.
Materiaal: B4 (*Alkmaar*).
Aantal monsters: 50.
2 Noordholl. keurmeesters (1, 2); 3 Friesche keurmeesters (3, 4, 5).

	Combinatie keurmeesters				Gemiddelde (van de laatste drie kolommen)
	1 — 2	3 — 4	3 — 5	4 — 5	
m_V^2	0,557	0,897	0,537	0,477	0,637
$\sigma_{\bar{x}}^2$	0,139	0,224	0,134	0,119	0,159
$\sigma_{\bar{x}}$	0,37	0,47	0,37	0,35	0,40

Tabel 10. Als tabel 8 en 9.
Materiaal: B3 (*Leeuwarden*).
Aantal monsters: 81.
2 Friesche keurmeesters (1, 2).

	Combinatie 1 — 2
m_V^2	0,559
$\sigma_{\bar{x}}^2$	0,140
$\sigma_{\bar{x}}$	0,37

Uit het bovenstaande materiaal mag men concludeeren, dat de keurmeesters (nl. een combinatie van twee) in staat zijn het cijfer voor smaak vast te stellen met een middelbare fout, welke in de buurt van 0,4 punt ligt.

Het is echter wenschelijk deze waarde (0,4) nog met meer bewijsmateriaal te staven. Immers bovenstaand cijfer is ontleend aan de gegevens van slechts enkele keuringen. Wel hadden wij bij deze keuringen het groote voordeel, dat zoovele keurmeesters uit verschillende deelen van het land hun oordeel uitspraken, maar aan den anderen kant bestaat de mogelijkheid, dat de nauwkeurigheid iets is verkleind door de buitengewone omstandigheden, waaronder de keurmeesters hun oordeel moesten vaststellen. Vele van de keurmeesters hadden voor de keuring een lange en vrij vermoeiende reis achter den rug, waardoor de scherpte van het waarnemingsvermogen iets kan zijn afgestompt.

Voor contrôle gebruikten wij de volgende gegevens:

1. de verschillen, die tusschen de vier keurmeesters waren gevonden bij het beoordeelen van de reeds eerder genoemde 15 monsters per week over 1941 (boterkeuringen van den G.O.Z.).

Wij vonden het volgende.

Tabel 11. Nauwkeurigheid van verschillende combinaties van twee keurmeesters voor *smaak* over het jaar 1941.
Materiaal: A2 (*Zutphen*).
4 Geld.-Ov. keurmeesters (1, 2, 3, 4).

Combinatie van keurmeesters	1 — 2	1 — 3	1 — 4	2 — 3	2 — 4	3 — 4	Gemiddeld
Aantal monsters	405	390	340	405	380	365	
m_V^2	0,660	0,504	0,699	0,643	0,891	0,643	0,673
$\sigma_{\bar{x}}^2$	0,165	0,126	0,175	0,161	0,223	0,161	0,168
$\sigma_{\bar{x}}$	0,41	0,36	0,42	0,40	0,47	0,40	0,41

2. materiaal over de jaren 1937 tot en met 1941, dat wij in hoofdstuk IV § 3 verder zullen bespreken. Hoewel dit materiaal niet kan dienen om een nauwkeurige berekening te geven van de grootte van de middelbare fout voor het cijfer voor smaak, konden wij toch daaruit een schatting maken. Wij kwamen daarbij tot een cijfer van 0,44. De bespreking, hoe wij tot dit cijfer kwamen, stellen wij tot later uit (zie blz. 113).

Wij mogen dus besluiten, dat het cijfer voor smaak gemiddeld wordt vastgesteld met een middelbare fout van ongeveer 0,4.

Het spreekt vanzelf, dat dit een middelwaarde is. Vooral het onder 1 genoemde materiaal laat duidelijk zien, dat niet alle combinaties van keurmeesters even groote middelbare fouten vertoonen. Er zijn essentieele verschillen aanwezig. Dit wordt duidelijk, indien wij bedenken, dat de middelbare fout van de onder 1 berekende middelbare fout ongeveer

$$\frac{0,4}{\sqrt{2 \times 400}} \simeq 0,01^5$$

bedraagt.

Verschillen in $\sigma_{\bar{x}}$ van 0,04 moeten, indien deze grootheid berekend is uit 400 waarnemingen, dus zeker als essentieel worden beschouwd, waaruit bijv. volgt, dat de combinatie 1 — 3 absoluut zeker met een kleinere middelbare fout heeft gekeurd dan bijv. de combinatie 1 — 4.

Voor geur.

Bij de berekeningen, die wij zooeven onder 2 bespraken — waarop wij nog nader terugkomen — was ons reeds gebleken, dat de middelbare fout voor geur van dezelfde orde van grootte moest zijn als die voor smaak.

Wij berekenden uit dat materiaal, dat ongeveer 750 monsters omvat, de middelbare fout voor het cijfer voor geur en kwamen op een waarde van 0,43.

Om deze reden hebben wij de middelbare fout voor geur niet voor alle gemeenschappelijke keuringen bepaald.

Wij geven hieronder alleen de cijfers, voortvloeiende uit de reeds meermalen genoemde gemeenschappelijke keuring te Zutphen (103 monsters).

Tabel 12. Nauwkeurigheid van verschillende combinaties van twee keurmeesters voor *geur*.
Materiaal: B1 (Zutphen).
Aantal monsters: 103.

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem
m_V^2	0,635	0.762	0,597	0,684	0,752	0,607	0,673	0,732	0,567	0,539	0,613
$\sigma_{\bar{x}}^2$	0,159	0,191	0,149	0,166	0,188	0,152	0,168	0,183	0,142	0,135	0,153
$\sigma_{\bar{x}}$	0,40	0,44	0,39	0,41	0,43	0,39	0,41	0,43	0,38	0,37	0,39

Conclusie. *De middelbare fout voor het cijfer voor geur (toegekend door een combinatie van twee keurmeesters) mag — evenals voor smaak — op* 0,4 *worden gesteld.*

Voor gehalte-en-bewerking.

Uit het op blz. 113 nader te bespreken materiaal vonden wij een cijfer van ongeveer 0,30 voor de middelbare fout van het cijfer voor gehalte-en-bewerking.

Voor de gemeenschappelijke keuring te Zutphen vonden wij:

Tabel 13. Nauwkeurigheid van verschillende combinaties van twee keurmeesters voor *gehalte-en-bewerking.*
Materiaal: B1 (Zutphen).
Aantal monsters: 103.

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem
m_V^2	0,383	0,344	0,412	0,441	0,490	0,199	0,378	1,004	0,461	0,761	0,742
$\sigma_{\bar{z}}^2$	0,096	0,086	0,103	0,110	0,123	0,050	0,095	0,251	0,115	0,190	0,185
$\sigma_{\bar{x}}$	0,31	0,29	0,32	0,35	0,35	0,23	0,31	0,50	0,34	0,44	0,43

De variaties zijn hier veel grooter. Willen wij een gemiddelde waarde geven voor de middelbare fout voor het cijfer van gehalte-en-bewerking, *dan zou dit* 0,30—0,35 *kunnen zijn.*

Voor „totaal aantal punten".

Wij geven hieronder de berekening van dit cijfer voor de gemeenschappelijke keuring te Zutphen:

Tabel 14. Frequentie van verschillen tusschen twee keurmeesters voor *„totaal aantal punten"* en berekening van de nauwkeurigheid van combinaties van twee keurmeesters.
Materiaal: B1 (*Zutphen*).

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters:	1—2	1—3	1—4	2—3	2—4	3—4	Tot.	5—6	5—7	6—7	Tot.
Verschillen											
0	17	11	17	14	20	19	98	12	18	18	48
2	19	33	20	19	24	29	144	14	27	14	55
4	24	18	28	20	21	19	130	21	14	29	64
6	17	19	18	26	15	18	113	18	17	14	49
8	10	8	5	11	10	6	50	11	10	11	32
10	11	7	9	8	6	9	50	16	14	12	42
12	2	5	3	3	3		16	5	2	1	8
14	3	1	2	1	2	2	11	4	1	3	8
16		1		1	1	1	4	2		1	3
18			1		1		2				
Totaal	103	103	103	103	103	103	618	103	103	103	309
							Gem.				Gem.
m_V^2 1)	35,48	33,54	34,07	35,80	34,78	28,80	33,75	51,70	33,34	37,70	40,91
$\sigma_{\bar{x}}^2$	8,87	8,39	8,52	8,95	8.70	7,20	8,44	12,93	8,34	9,42	10,23
$\sigma_{\bar{x}}$	3,0	2,9	2,9	3,0	2,9	2,7	2,9	3,6	2,9	3,1	3,2

Te Assen werd gevonden:

Tabel 15. Nauwkeurigheid van combinaties van twee keurmeesters voor *„totaal aantal punten"*.
Materiaal: B2 (*Assen*).
Aantal monsters: 68.

	Gr.-Dr. Keurm.	Geld.-Ov. keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
$\sigma_{\bar{x}}^2$	9,60	6,45	6,64	9,42	7,50
$\sigma_{\bar{x}}$	3,1	2,5	2,6	3,1	2,7

1) De op $\frac{\Sigma V_k^2}{N}$ in rekening gebrachte correctie van Sheppard bedraagt hier: $\frac{c^2}{12} = \frac{2^2}{12} = 0,33$

Te Alkmaar:

Tabel 16. Als tabel 15.
Materiaal: B4 (*Alkmaar*).
Aantal monsters: 50 (voor de combinaties (3, 4) en (4, 5) 39).

	N.-Holl. Keurm.	Friesche keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
$\sigma^2_{\bar{x}}$	8,86	14,24	9,82	4,91	9,66
$\sigma_{\bar{x}}$	3,0	3,8	3,1	2,2	3,1

Te Leeuwarden:

Tabel 17. Als tabel 15.
Materiaal: B3 (*Leeuwarden*).
Aantal monsters: 81.

Combinatie van keurmeesters:	Friesche keurmeesters 1 — 2
$\sigma^2_{\bar{x}}$.	10,81
$\sigma_{\bar{x}}$	3,3

Wij zien, dat er in de gevonden cijfers nog wel eenige variatie aanwezig is. Dit kan veroorzaakt zijn door twee omstandigheden.

a. De middelbare fouten zijn berekend uit ten hoogste 103 monsters. Op grond van de waarschijnlijkheidsrekening kan worden aangenomen, dat de uit 103 waarnemingen berekende middelbare fout een middelbare fout heeft ten bedrage van:

$$\frac{\sigma_{\bar{x}}}{\sqrt{2 \times 103}} \simeq \tfrac{3}{14} \simeq 0{,}2.$$

Afwijkingen van 0,4 naar beide kanten mogen dus in dit geval zeker worden verwacht.

Dat het cijfer voor de keuring te Leeuwarden iets hooger ligt dan het gemiddelde bij de andere keuringen, behoeft dus zeker niet te wijzen op een grootere onnauwkeurigheid van de keurmeesters te Leeuwarden.

b. Men mag zeker niet aannemen, dat alle combinaties van twee keurmeesters met dezelfde nauwkeurigheid keuren.

Bovenstaande berekeningen leeren ons, dat de middelbare fout van het cijfer voor „totaal aantal punten" gemiddeld ongeveer 2,9 punt bedraagt.

Om de redenen, die wij reeds bij de beschouwingen over de middelbare fout van het cijfer voor smaak uiteenzetten, doen wij goed dit cijfer te toetsen aan meer gegevens.

Deze gegevens zijn de volgende.

a. Wij berekenden voor alle door drie keurmeesters in 1941 beoordeelde monsters (bij de boterkeuringen van den G.O.Z.) de bij de zes mogelijke combinaties van twee keurmeesters (uit 4 keurmeesters) optredende middelbare fouten.

Wij verkregen het volgende resultaat.

Tabel 18. Nauwkeurigheid van combinatie van twee keurmeesters over het jaar 1941 voor *„totaal aantal punten"*.
Materiaal: A2 (*Zutphen*).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Combinatie van keurmeesters:	1 — 2	1 — 3	1 — 4	2 — 3	2 — 4	3 — 4	Gemiddeld
Aantal monsters	405	390	355	405	380	365	
$\sigma^2_{\bar{x}}$	7,03	5,88	7,29	7,31	9,67	7,73	7,48
$\sigma_{\bar{x}}$	2,65	2,43	2,70	2,70	3,10	2,78	2,73

Uit dit materiaal volgt dus gemiddeld een iets lagere waarde.

Terloops zij hier opgemerkt, dat dit materiaal ook gelegenheid biedt om de verschillen in nauwkeurigheid van de verschillende combinaties wat nader te beschouwen. Wij mogen er van uitgaan, dat de bovenstaande waarden voor $\sigma_{\bar{x}}$ ongeveer een middelbare fout hebben ten bedrage van:

$$\frac{2,7}{\sqrt{2 \times 400}} \simeq 0,1$$

waaruit volgt, dat sommige $\sigma_{\bar{x}}$'s principieel verschillen, m.a.w. de ééne combinatie van keurmeesters keurt nauwkeuriger dan de andere.

De verdere bespreking van het gebruikte materiaal stellen wij uit tot later.

b. Wij berekenden voor verschillende kwartalen van de jaren 1937 tot en met 1941 voor een ander doel bepaalde grootheden en konden daaruit een indruk verkrijgen over de grootte van de middelbare fout voor een combinatie van twee keurmeesters.

Ook met dit uitgebreide materiaal kwamen wij tot een cijfer van ongeveer 2,80 (zie verder blz. 114 tabel 59).

Uiteindelijk mag onze conclusie zijn:

a. het cijfer voor geur wordt door een combinatie van twee keurmeesters vastgesteld met een middelbare fout van ongeveer 0,4;

b. *hetzelfde geldt voor het cijfer voor smaak;*

c. *de middelbare fout van het cijfer voor gehalte-en-bewerking bedraagt 0,3 à 0,35;*

d. *de middelbare fout bij de vaststelling van het eindcijfer mag gemiddeld op 2,8 worden gesteld.*

Men bedenke, dat hier nooit absoluut vaststaande cijfers gegeven kunnen worden. Voor elke combinatie van keurmeesters zijn deze cijfers iets verschillend. De variatie is echter betrekkelijk gering.

Alvorens nu over te gaan tot eèn bespreking van de beteekenis van deze cijfers, willen wij nog het volgende probleem behandelen.

Wij kunnen de vraag stellen: Aangenomen, dat het cijfer voor geur door twee keurmeesters met een middelbare fout van 0,4 kan worden vastgesteld, het cijfer voor smaak van 0,4 en gehalte-en-bewerking van 0,3, welke middelbare fout vloeit hieruit dan voort voor het „totaal aantal punten"?

Stellende

$\bar{g}$ het gemiddelde van de door beide keurmeesters gegeven cijfers voor geur,
$\bar{s}$ „ „ „ „ „ „ „ „ „ „ smaak,
$\bar{j}$ „ „ „ „ „ „ „ „ „ „ geh.-en-bew.
$\bar{K}$ „ „ „ „ „ „ „ „ „ „ totaal,

dan geldt:

$$\bar{K} = 2\bar{g} + 4\bar{s} + 4\bar{j}.$$

Bij verandering van $\bar{g}$ met $\Delta\bar{g}$, $\bar{s}$ met $\Delta\bar{s}$ en $\bar{j}$ met $\Delta\bar{j}$ geldt voor de verandering van $\bar{K}$ ($\Delta\bar{K}$):

$$\Delta\bar{K} = 2\Delta\bar{g} + 4\Delta\bar{s} + 4\Delta\bar{j}.$$

Gekwadrateerd:

$$\Delta\bar{K}^2 = 4\Delta g^2 + 16\Delta\bar{s}^2 + 16\Delta\bar{j}^2 + 16\Delta\bar{g}\,.\,\Delta\bar{s} + 16\Delta\bar{g}\,.\,\Delta\bar{j} + 32\Delta\bar{s}\,.\,\Delta\bar{j}.$$

en gesommeerd over een oneindig aantal waarden van $\Delta\bar{K}$ en gemiddeld (aangegeven door het symbool $< >$, zie blz. 19), geeft:

$$<\Delta\bar{K}^2> = 4<\Delta\bar{g}^2> + 16<\Delta\bar{s}^2> + 16<\Delta\bar{j}^2> + 16<\Delta\bar{g}\,.\,\Delta s> +$$
$$+ 16<\Delta\bar{g}\,.\,\Delta\bar{j}> + 32<\Delta\bar{s}\,.\,\Delta\bar{j}>.$$

Nu is:

$$<\Delta\bar{K}^2> = \sigma^2_{\bar{K}};\quad <\Delta\bar{g}^2> = \sigma^2_{\bar{g}};\quad <\Delta\bar{s}^2> = \sigma^2_{\bar{s}};\quad <\Delta\bar{j}^2> = \sigma^2_{\bar{j}}.$$

Verder:

$$<\Delta\bar{g}\,.\,\Delta\bar{s}> = r_{\bar{g},\bar{s}}\,\sigma_{\bar{g}}\,.\,\sigma_{\bar{s}}$$
$$<\Delta\bar{g}\,.\,\Delta\bar{j}> = r_{\bar{g},\bar{j}}\,\sigma_{\bar{g}}\,.\,\sigma_{\bar{j}}$$
$$<\Delta\bar{s}\,.\,\Delta\bar{j}> = r_{\bar{s},\bar{j}}\,\sigma_{\bar{s}}\,.\,\sigma_{\bar{j}}$$

waarbij $r_{\bar{g},\bar{s}}$, resp. $r_{\bar{g},\bar{j}}$ en $r_{\bar{s},\bar{j}}$ de correlatiecoëfficiënt is tusschen $\bar{g}$ en $\bar{s}$, resp. $\bar{g}$ en $\bar{j}$, en $\bar{s}$ en $\bar{j}$.

Dus:

$$\sigma^2_{\bar{K}} = 4\sigma^2_{\bar{g}} + 16\sigma^2_{\bar{s}} + 16\sigma^2_{\bar{j}} + 16r_{\bar{g},\bar{s}}\,\sigma_{\bar{g}} \,.\, \sigma_{\bar{s}} + 16r_{\bar{g},\bar{j}}\,\sigma_{\bar{g}} \,.\, \sigma_{\bar{j}} + 32r_{\bar{s},\bar{j}}\,\sigma_{\bar{s}} \,.\, \sigma_{\bar{j}}.$$

Wij gingen er van uit, dat

$$\sigma_{\bar{g}} = \sigma_{\bar{s}} = 0{,}4 \quad \text{en} \quad \sigma_{\bar{j}} = 0{,}3.$$

Voor een aantal keuringen berekenden wij de correlatiecoëfficiënten en vonden

$$r_{\bar{g},\bar{s}} = 0{,}8; \quad r_{\bar{g},\bar{j}} = 0{,}3 \quad \text{en} \quad r_{\bar{s},\bar{j}} = 0{,}3$$

waaruit volgt:

$$\sigma^2_{\bar{K}} = 4 \times 0{,}16 + 16 \times 0{,}16 + 16 \times 0{,}09 + 16 \times 0{,}8 \times 0{,}16 +$$
$$+ 16 \times 0{,}3 \times 0{,}12 + 32 \times 0{,}3 \times 0{,}12 =$$
$$0{,}64 + 2{,}56 + 1{,}44 + 2{,}05 + 0{,}58 + 1{,}16 = 8{,}43,$$

waaruit volgt: $\sigma_{\bar{K}} = 2{,}8$ à $2{,}9$,
een resultaat, dat dus goed overeenstemt met wat langs directen weg werd berekend.

§ 4. Bespreking der tot nu toe verkregen resultaten.

Welke beteekenis hebben nu de in de vorige § verkregen waarden voor de middelbare fouten van het cijfer voor smaak, geur, gehalte-en-bewerking en „totaal aantal punten" voor de practijk?

Wanneer op een boterkeuring een botermonster het cijfer 6 voor smaak heeft gekregen, dan staat — zooals ook vanzelf spreekt — absoluut niet vast, dat dit inderdaad het juiste cijfer is. Dit juiste cijfer kennen wij natuurlijk niet. Maar het is toch mogelijk te berekenen, hoeveel kans wij hebben op een bepaalde afwijking van dat „ware" cijfer.

Het begrip: „ware" cijfer zal den lezer wellicht niet geheel duidelijk zijn. Het „ware" cijfer is een cijfer, dat slechts in gedachte bestaat. Men kan het definieeren als het cijfer, dat gemiddeld gegeven zou worden, indien een monster door een zeer groot aantal (eigenlijk: een oneindig aantal) keurmeesters en door iederen keurmeester oneindig vele malen werd gekeurd.

Uit de boven berekende waarde voor de middelbare fout voor het cijfer voor smaak van 0,4 punt kunnen wij afleiden, dat wij de onderstaande kansen hebben op een bepaalde afwijking van het gegeven cijfer met het „ware" cijfer.

Tabel 19. Kans op het voorkomen van een bepaald verschil tusschen een door combinatie van twee keurmeesters gegeven cijfer en het ware cijfer (voor *smaak*).

Verschil tusschen het gegeven en „ware" cijfer	Kans hierop
0 punten	46,8 %
½ punt of minder	93,9 %
1 punt of minder	99,8 %

Dit beteekent dus, dat, indien het cijfer voor smaak, dat aan een bepaald monster diende te worden toegekend (het „ware" cijfer dus) een 6 is, er 46,8 % kans is, dat een combinatie van twee keurmeesters dit monster ook een 6 zal geven (en dus 53,2 % kans, dat ze een ander cijfer geven). Er is 93,9 % kans, dat de keurmeesters een afwijking van $+\frac{1}{2}$, 0 of $-\frac{1}{2}$ van het ware cijfer geven dus $5\frac{1}{2}$, 6 of $6\frac{1}{2}$ geven en 6,1 % kans, dat de afwijking grooter is dan $\frac{1}{2}$, (dat dus een 5 of 7 wordt gegeven).

Het is zeer onwaarschijnlijk, dat een grootere afwijking dan 1 punt wordt gegeven (slechts 0,2 % kans).

Het bovenstaande geeft ons een beeld van de mogelijke afwijkingen t.o.v. het „ware" cijfer. Het „ware" cijfer is echter een begrip, dat in de practijk vrij weinig houvast biedt.

Duidelijker komt de nauwkeurigheid, waarmede twee keurmeesters het cijfer voor smaak vaststellen, tot uiting, wanneer wij de volgende vraag beantwoorden: Indien een zelfde monster door hetzelfde paar keurmeesters twee keeren op smaak wordt beoordeeld, welke kans heb ik dan op bepaalde verschillen?

Dat leert het volgende tabelletje.

Tabel 20. Kans op een bepaald verschil tusschen een eerste en tweede beoordeeling (voor *smaak*).

Verschil tusschen 1e en 2e beoordeeling	Kans hierop
0 punten	34,1 %
$\frac{1}{2}$ punt of minder	81,5 %
1 punt of minder	97,3 %
$1\frac{1}{2}$ punt of minder	99,8 %

Er is dus 34 % kans, dat het monster bij de 2e beoordeeling hetzelfde cijfer voor smaak verkrijgt als bij de 1e beoordeeling, 81,5 % — 34 % = = 47,5 %, dat het verschil een halve punt bedraagt en 97,3 % — 81,5 % = = 15,8 %, dat het verschil een heele punt bedraagt en 99,8 % — 97,3 % = = 2,5 % kans, dat een verschil van anderhalve punt wordt gevonden.

Een vraag, die voor de practijk van belang is, is deze: Hoe groot moet het verschil tusschen de cijfers voor smaak zijn, die op de keuring aan twee monsters zijn toegekend, voor men mag concludeeren, dat er een essentieel verschil in smaak bestaat?

Dit hangt natuurlijk af van de mate van zekerheid, die men wenscht.

In nevenstaande staat is dit nader uitgewerkt.

Tabel 21. Kans op een werkelijk bestaand kwaliteitsverschil, wanneer tusschen twee monsters een bepaald verschil (in punten) is geconstateerd (voor *smaak*).

Verschil tusschen twee monsters van:	Kans, dat het monster met de hoogste punten inderdaad het best is van smaak
0 punten	50,0 %
½ punt	81,2 %
1 punt	96,2 %
1½ punt	99,6 %

Wij merken op, dat wij deze kanstabel op iets andere wijze hebben berekend dan de beide voorgaande. Bij de beide voorgaande tabellen hebben wij als klassegrens steeds genomen het genoemde verschil + ¼ punt (dus de bovenste klassegrens), omdat in wezen een genoteerd verschil van 1 punt in werkelijkheid een verschil van 1¼ punt kan zijn. Bij de laatste tabel mogen wij dit niet doen. Hier is het genoteerde verschil als grens genomen.

Uit bovenstaand staatje volgt, dat men bij een verschil van 1 punt met 96 % kans tot een essentieel verschil in kwaliteit mag concludeeren. Wenscht men een nog grootere zekerheid, dan zou het verschil 1½ punt moeten zijn. Men heeft dan slechts 0,8 % kans, dat in werkelijkheid geen verschil aanwezig is, of zelfs, dat het monster met de hoogste punten in werkelijkheid slechter is geweest dan het monster met de lage punten.

Voor *geur* en *gehalte-en-bewerking* zijn natuurlijk analoge berekeningen op te stellen. Wij zullen dit echter niet doen. Bij de nauwkeurigheid van het cijfer voor *„totaal aantal punten"* dienen wij echter nog wel even stil te staan.

Tabel 22. Kans op het voorkomen van een bepaald verschil tusschen een door een combinatie van twee keurmeesters gegeven cijfer voor *„totaal aantal punten"* en het ware cijfer.

Verschil tusschen het gegeven cijfer voor „totaal aantal punten" en het „ware" cijfer	Kans hierop (middelbare fout 2,8 punt)	Kans hierop (middelbare fout 3,2 punt)
0 punten	14,1 %	12,4 %
1 punt of minder	40,8 %	36,1 %
2 punten of minder	62,8 %	56,5 %
3 „ „ „	78,9 %	72,6 %
4 „ „ „	89,2 %	84,1 %
5 „ „ „	95,1 %	91,4 %
6 „ „ „	98,0 %	95,8 %
7 „ „ „	99,3 %	98,1 %
8 „ „ „	99,8 %	99,2 %
9 „ „ „		99,7 %

Wij geven hier dezelfde tabellen als bij smaak, echter met twee kolommen voor de kans op een bepaald verschil, nl. één, die berekend is met een middelbare fout van 2,8 punt en de andere, die afgeleid is uit een middelbare fout van 3,2 punt.

Tabel 23. Kans op een bepaald verschil tusschen een eerste en tweede beoordeeling (voor „*totaal aantal punten*").

Verschil tusschen 1e en 2e beoordeeling	Kans hierop (middelbare fout 2,8 punt)	Kans hierop (middelbare fout 3,2 punt)
0 punten	10,0 %	8,8 %
1 punt of minder	29,5 %	25,9 %
2 punten of minder	47,2 %	41,9 %
3 „ „ „	62,3 %	56,1 %
4 „ „ „	74,4 %	68,0 %
5 „ „ „	83,5 %	77,6 %
6 „ „ „	89,9 %	84,9 %
7 „ „ „	94,2 %	90,3 %
8 „ „ „	96,9 %	94,0 %
9 „ „ „	98,4 %	96,4 %
10 „ „ „	99,2 %	98,0 %
11 „ „ „	99,6 %	98,9 %

Tabel 24. Kans op een werkelijk bestaand kwaliteitsverschil, wanneer tusschen twee monsters een bepaald verschil is geconstateerd (voor „*totaal aantal punten*").

Verschil tusschen twee monsters van:	Kans, dat het monster met de hoogste punten inderdaad het beste is	
	(middelbare fout 2,8 punt)	(middelbare fout 3,2 punt)
0 punten	50,0 %	50,0 %
1 punt	60,0 %	58,7 %
2 punten	69,3 %	67,1 %
3 „	77,6 %	74,6 %
4 „	84,4 %	80,2 %
5 „	89,7 %	86,5 %
6 „	93,5 %	90,8 %
7 „	96,2 %	94,0 %
8 „	97,8 %	96,2 %
9 „	98,8 %	97,7 %
10 „	99,4 %	98,6 %

Indien wij aannemen, dat een monster met 68 punten als goede boter wordt beschouwd, een monster van 74 punten als zeer goede boter en een monster van 62 punten als minder goede boter, dan blijkt uit tabel 22, dat de kans, dat een monster in werkelijkheid goede boter (68 punten) òf als minder goed òf als zeer goed wordt beoordeeld 2—4 % bedraagt (naar gelang de middelbare fout, waarmede de combinatie van keurmeesters keurt).

Verder is uit de tweede tabel (tabel 23) af te lezen, dat een monster, dat bij een eerste beoordeeling 68 punten heeft verkregen, bij een tweede beoordeeling met 10—15 % kans hetzij 74 punten of hooger, hetzij 62 punten of lager zal verkrijgen.

Uit de laatste tabel volgt ten slotte, dat een verschil in totaal aantal punten pas als essentieel is te beschouwen (dus te wijten is aan een werkelijk bestaand kwaliteitsverschil), indien het verschil ten minste 7 tot 8 punten bedraagt (men heeft dan ongeveer 95 % kans).

Uit het bovenstaande volgt, dat aan een verschil van bijv. 5 à 6 punten in den keuringsuitslag van week tot week geen *essentieele* waarde mag worden toegekend. Ook blijkt, dat in den keuringsuitslag van de boter van één fabriek — ook al blijft in werkelijkheid de kwaliteit van de boter constant — van week tot week aanmerkelijke verschillen kunnen optreden. Met verschillen van 7 tot 8 punten moet in enkele gevallen zeker rekening worden gehouden.

Men krijgt in het algemeen pas een indruk, of een fabriek betere boter maakt dan een andere, indien men verscheidene keuringen met elkaar vergelijkt. Een gemiddelde van 4 keuringen is bijv. 2 × zoo nauwkeurig als het cijfer van een enkele keuring. Bij maandgemiddelden is dus een verschil van 3,5 punt reeds voldoende om te concludeeren tot een werkelijk bestaand verschil (met ± 95 % zekerheid!).

Het spreekt vanzelf, dat het jaargemiddelde een nog grootere nauwkeurigheid bezit. Men mag hier rekenen met een middelbare fout van

$$\frac{2,8}{\sqrt{52}} \simeq 0,4$$

De middelbare fout van het verschil tusschen twee jaargemiddelden bedraagt $0,4 \times \sqrt{2} = 0,55$.

Verschillen van 0,9 punt en kleiner mogen dus bij jaargemiddelden niet als essentieel worden beschouwd, indien men een zekerheid van ten minste 95 % wenscht.

Terloops zij hier aan toegevoegd, dat het bestaan van een onzekerheid van bijna 1 punt van het jaargemiddelde maakt, dat de ranglijst, die door verschillende zuivelbonden wordt gepubliceerd van de fabrieken, die gedurende het jaar gemiddeld het hoogste aantal punten verkregen, vrij onzeker is. De verschillen tusschen opeenvolgende fabrieken in de ranglijst zijn soms niet grooter dan 0,1 à 0,2. Het is dan ook te begrijpen, dat het rangcijfer, dat een bepaalde fabriek verkrijgt, van jaar tot jaar nogal aan veranderingen onderhevig is. Natuurlijk wordt dit mede veroorzaakt door werkelijke kwaliteitsveranderingen van de boter, maar zelfs indien de kwaliteit van de boter van de fabrieken van jaar tot jaar gelijk bleef, zou men rekening moeten houden met de mogelijkheid van een vrij belangrijke variatie.

Als eindconclusie van deze paragraaf mogen wij verklaren, dat een combinatie van twee keurmeesters in staat is verschillen in kwaliteit van 7 tot 8 punten met voldoenden graad van zekerheid te constateeren. Bij kleinere verschillen heeft natuurlijk het monster met de hoogste punten ook de grootste kans het beste te zijn, maar indien men een behoorlijke betrouwbaarheid wenscht, mag men het verschil niet als essentieel beschouwen.

Wij kunnen onze beschouwingen van deze paragraaf ook aldus samenvatten. Een combinatie van twee keurmeesters is in staat drie tot vier groepen boter te onderscheiden nl.:

zeer goede en goede boter (eventueel is deze groep nog te splitsen in twee groepen: zeer goede boter en goede boter),
matige boter,
slechte boter.

Bij keuring door twee keurmeesters en bij één keer beoordeelen wordt een verdergaande splitsing onzeker.

Men zou hieruit de gevolgtrekking kunnen maken, dat de wijze van uitdrukken van het oordeel, zooals die plaats heeft bij de keuringen van het Z.K.B. en de verkoopvereenigingen (zie blz. 4) de voorkeur verdient boven het gebruik van het door de bonden van coöperatieve zuivelfabrieken toegepaste puntenstelsel.

In § 8 van dit hoofdstuk zullen wij nader deze vraag bespreken.

Aan het slot van deze paragraaf merken wij op, dat alle in deze paragraaf uitgevoerde kansberekeningen gebaseerd zijn op een normale verdeeling der fouten. In hoeverre deze veronderstelling juist is, zal een nader onderzoek in hoofdstuk III leeren.

§ 5. Bespreking van eenige beschouwingen uit de litteratuur.

Het is hier ook de plaats aandacht te schenken aan de eenige beschouwingen, die wij in de litteratuur konden vinden betreffende de nauwkeurigheid van de boterkeuring. Wij bedoelen hier hetgeen Pasveer in zijn proefschrift „Een statistisch onderzoek over de factoren, welke invloed uitoefenen op de kwaliteit van de Nederlandsche boter" (Wageningen, 1941) hierover mededeelt.

Na gewezen te hebben op de wenschelijkheid, dat iedere onderzoeker,die zich met de kwaliteit van de boter wil bezighouden, zich op de hoogte stelt van de nauwkeurigheid van de boterkeuringen, gaat deze schrijver op blz. 11 als volgt verder:

> „Van de schaal 1 tot 10 wordt in hoofdzaak (95 % of meer van de gevallen) slechts het gedeelte van 4 tot 7 gebruikt. De cijfers 4 en 5

worden voor geur en smaak gegeven als een grooter of kleiner euvel wordt opgemerkt. De cijfers 6 en 7 verkrijgen boters, die goed of beter worden gewaardeerd. Voor gehalte en bewerking ligt dit iets anders, daarbij wordt voor goede normale boter een 8 gegeven, voor iets minder goede, bijvoorbeeld minder stevige boter, een 7 en voor boter, waar een wat grooter gebrek valt op te merken een 6, zelden minder.

De toestand is nu zoo, dat het aantal keeren, waarin door beide keurmeesters een monster voor geur met 7 wordt gewaardeerd kleiner is, dan dat waarin de eene keurmeester een 6 en de andere een 7 geeft. Dit laatste aantal komt in orde van grootte nabij het aantal keeren, dat beide keurmeesters met een 6 waardeeren. Dit geldt evenzoo voor den smaak. Beperken wij ons nu tot geur en smaak, dan zien wij, dat bij het vaststellen van de waardeering voor één bepaald monster het toeval een zeer groote rol kan spelen. Veronderstellen wij : een monster wordt door 2 keurmeesters voor geur gewaardeerd met 6 en 7, gemiddeld dus 6½. Vervangen wij nu in gedachten een van de keurmeesters door een derden, dan zal deze een cijfer geven, dat in het grootste aantal gevallen hetzij een 6 dan wel een 7 is. In een enkel geval kan dit cijfer wel een 5 of een 8 zijn. Eenvoudigheidshalve beperken wij ons tot het meest waarschijnlijke, dat dus weer 6 of 7 wordt gegeven. Hetzelfde monster zou dus alleen tengevolge van het wisselen van de keurmeesters voor geur gewaardeerd kunnen zijn met 6, 6½ of 7. Eenzelfde redeneering geldt ook voor de waardeering voor smaak. Het bovenstaande zou nu niet zoo belangrijk zijn, indien daar nog niet het volgende bijkwam. De practijk beschouwt boter met 2 zessen voor geur en smaak (na vermenigvuldiging dus „12—24"), als boter weliswaar zonder gebrek, maar toch zeer neutraal van smaak, zonder opvallend goede eigenschappen. Boter daarentegen met 2 zevens (14—28) wordt als zeer goede boter aangezien.

Uit het voorafgaande is voldoende duidelijk gebleken, dat men zeer voorzichtig moet zijn met de keuringsresultaten als waardemeter. Zooals nog nader zal blijken hebben verschillende onderzoekers, in het bijzonder diegenen, die niet uit de zuivelbereiding voortkwamen, zich van dit feit niet voldoende rekenschap gegeven.

Evenzoo, hoewel in mindere mate is het bij de monsters met een grooter of kleiner gebrek gesteld. Het aantal monsters, dat voor geur een 6 en een 5 ontvangt is van dezelfde orde van grootte als het aantal, dat 2 zessen verkrijgt. Hetzelfde monster kan door wisselen van de keurmeesters met „12—24" (geen gebrek!) en met „10—20" worden gewaardeerd. Bij deze laatste waardeering wordt het monster als „slecht" beschouwd.

Na het kennisnemen van het voorgaande zal men vragen, heeft bij een dergelijken stand van zaken een keuring nog wel waarde? Inderdaad is dit een vraag, die een ieder, die zich met de boterkeuringen nader gaat bezig houden, zich in den beginne wel eens stelt. Het antwoord hierop is, dat de boterkeuring zeker groote waarde heeft, indien men de resultaten maar op de juiste manier beziet. Een fabriek waar regelmatig een goed product wordt gemaakt, zal wanneer eens een enkele keer een kleine afwijking wordt geconstateerd, niet dan onder bijzondere omstandigheden, daaraan aandacht behoeven te schenken. Verder zal het product van een fabriek, dat regelmatig iedere keuring weer, door beide keurmeesters vlot met zevens wordt gewaardeerd, met groote waarschijnlijkheid beter zijn, dan het product, dat nu eens met 12—24, dan weer met 13—26 wordt bedacht. Het gemiddelde van de keuring over een geheel jaar geeft van de kwaliteit van het door een bepaalde fabriek bereide product wel een goed beeld. Echter, en dit is zeer belangrijk, wanneer men twee monsters boter heeft en het eene is met 6, 6 het andere met 7, 7 voor geur en smaak gewaardeerd, dan heeft men wel eenige waarschijnlijkheid, dat het laatste monster beter is dan het eerste, maar beslist zeker is men hiervan in geen geval. Wanneer het betreft 2 monsters met een kleiner verschil in waardeering, is de kans, dat het gevonden verschil ook inderdaad het juiste is, nog veel kleiner."

Uit wat wij in de vorige § mededeelden, blijkt, dat wij gedeeltelijk de conclusies van Pasveer kunnen onderschrijven. Het zal echter een ieder, die onze beschouwingen tot nu toe heeft gevolgd, duidelijk zijn, dat wij zijn bewijsvoeringen voor het meerendeel als onvoldoende streng moeten beschouwen. De zeer vaag gehouden redeneering lijkt op het eerste gezicht wel eenigszins aannemelijk, maar heeft een te onzekere basis om een zoo beslist oordeel over de betrouwbaarheid der keuring te geven. Letten we bijv. eens op de redeneering, die Pasveer er toe leidt om de uitspraak te doen, dat een overgang van 6, 6 voor geur *en* smaak naar 7, 7 voor geur en smaak zeer gemakkelijk mogelijk is. Bij toepassing van de door ons verkregen gegevens blijkt, dat deze overgang minder gemakkelijk is. Voor het totaal cijfer voor geur + smaak (d.i. dus: $T = 2\bar{g} + 4\bar{s}$) kan de volgende middelbare fout worden berekend:

$$\sigma_T^2 = 4\sigma_{\bar{g}}^2 + 16\sigma_{\bar{s}}^2 + 16 r_{\bar{g},\bar{s}}\, \sigma_{\bar{g}} \,.\, \sigma_{\bar{s}}.$$

Vullen wij in bovenstaande vergelijking voor $\sigma_{\bar{g}} = \sigma_{\bar{s}} = 0,4$ en voor $r_{\bar{g},\bar{s}} = 0,8$ in, dan verkrijgen we:

$$\sigma_T^2 = 0,64 + 2,56 + 2,05 = 5,25$$

waaruit volgt: $\sigma_T = 2,3.$

Hoeveel kans is er nu, dat in werkelijkheid „neutrale" boter (met 12—24) als „zeer goede" boter (14—28) wordt beoordeeld?

Wij stellen dus de vraag, welke kans aanwezig is, dat $T - \hat{T} \geqq 6$ is ($T = 42$, $\hat{T} = 36$).

Voeren wij de grootheid: $z = \dfrac{T - \hat{T}}{\sigma_T\sqrt{2}}$ in, dan mogen wij schrijven:

$$\text{kans op } T - \hat{T} \geqq 6 = \text{kans op } z \geqq \frac{6}{\sigma_T\sqrt{2}} = 1 - \left(\text{kans op } z \leqq \frac{6}{\sigma_T\sqrt{2}}\right).$$

De kans op $z \leqq \frac{6}{\sigma_T \sqrt{2}}$ wordt nu uitgedrukt door de functie:

$$\Theta\left(\frac{6}{\sigma_T \sqrt{2}}\right) = \Theta\left(\frac{6}{2{,}3\sqrt{2}}\right) = \Theta\,(1{,}8) = 0{,}995.$$

Hierbij stelt Θ de volgende integraal voor:

$$\Theta\,(\zeta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\zeta} e^{-z^2}\, dz.$$

Uit het bovenstaande volgt, dat de kans op $T - \hat{T} \geqq 6$ gelijk is aan: $1 - 0{,}995 = 0{,}005$ of **0,5 %**.

Wij hebben dus 0,5 % kans, dat in werkelijkheid „neutrale" boter als „zeer goede" wordt beoordeeld.

Precies dezelfde redeneering kan worden gevolgd om aan te toonen, dat de kans, dat de keurmeesters „12—24"-boter (geen gebreken!) zullen keuren als „10—20"-boter (slecht!) lang zoo groot niet is, als Pasveer doet voorkomen.

De kans, dat twee boters, waarvan de éene met 6 — 6 en de andere met 7 — 7 is beoordeeld, in werkelijkheid niet in kwaliteit verschillen of zelfs, dat de 6 — 6-boter beter is geweest dan de 7 — 7-boter, bedraagt:

$$1 - \Theta\left(\frac{6}{4{,}6}\right) = 1 - \Theta(1{,}3) = 1 - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+1{,}3} e^{-z^2}\, dz = 1 - 0{,}967 = \mathbf{3{,}3}\ \%.$$

Inderdaad is dus het verschil in kwaliteit niet absoluut zeker, maar met de uitdrukking „wel eenige waarschijnlijkheid" stelt Pasveer de zaak veel te pessimistisch voor.

Wij gaan accoord met zijn opmerking, dat aan de boterkeuring een vrij groote mate van onnauwkeurigheid verbonden is. De constateering van dit feit mag er echter niet toe leiden, dat dit te overdreven wordt voorgesteld. Daarmede zou men ten onrechte de keurmeesters een brevet van onvermogen toekennen, dat zij in het geheel niet verdienen.

Het kan misschien sommigen lezers toeschijnen, dat de hierboven genoemde 0,5 % kans in strijd is met wat wij op blz. 33 (tabel 22) berekenden. Daar vonden wij immers, dat een monster (in werkelijkheid) neutrale boter (68 punten) 2—4 % kans heeft om als minder goed (62 punten) of als zeer goed (74 punten) beoordeeld te worden.

Men bedenke echter, dat wij toen de kans op twee mogelijkheden nagingen (minder goed en zeer goed), terwijl wij zoo pas alleen maar de kans op minder goed (of zeer goed) berekenden. Bovendien spraken wij toen over het cijfer voor „totaal aantal punten" (met een middelbare fout van 2,8 punt (of 3,2 punt), terwijl we nu het totaal cijfer voor geur en smaak (middelbare fout 2,3 punt) in oogenschouw nemen.

Dat de bovengenoemde onderzoeker geen juiste voorstelling heeft van de nauwkeurigheid, waarmede de keurmeesters hun cijfers vaststellen, blijkt ook uit hetgeen hij op blz. 86 van zijn bovenaangehaald proefschrift mededeelt. Hij schrijft hier het volgende:

> „Reeds eerder is uiteengezet, dat het gebied van waardeeringsmogelijkheden voor keuringsboter betrekkelijk klein is. In de practijk komt het er op neer, dat voor den geur 10, 11, 12, 13 of 14 punten worden gegeven en voor den smaak 20, 22, 24, 26 of 28 punten, met enkele uitzonderingen naar beneden en naar boven.
>
> Boter met „12—24" voor geur en smaak geldt als neutrale boter, zonder goed of kwaad.
>
> Boter met „10—20" voor geur en smaak geldt als slechte boter en boter met „14—28" voor geur en smaak geldt als goede boter.
>
> Dit kleine traject brengt nu met zich mede, dat minder goede boter heel wat meer kans heeft om „te gunstig" te worden beoordeeld dan „te slecht", terwijl voor goede boter de kans, dat ze „te laag" wordt gewaardeerd heel wat grooter is, dan dat ze „te hoog" wordt gewaardeerd.
>
> Dit heeft tot gevolg, dat wanneer wij den invloed nagaan, die een of andere factor op de kwaliteit van de boter heeft, wij als regel een resultaat zullen vinden, dat kleiner is dan den invloed, dien die factor in werkelijkheid op de kwaliteit van de boter uitoefent."

In hoofdstuk III zullen wij nog nader bewijzen, dat de fouten, die de keurmeesters bij het keuren maken, normaal verdeeld zijn, m.a.w. dat de kans, dat boter te laag of te hoog wordt gewaardeerd, volkomen gelijk is.

§ 6. De correlatiecoëfficiënt tusschen het door twee keurmeesters toegekende cijfer en het ware cijfer.

Tot nu toe hebben wij de nauwkeurigheid, waarmede een combinatie van twee keurmeesters haar cijfer vaststelt, uitgedrukt in de middelbare fout, welke aan de vaststelling van dat cijfer is verbonden. Door deze middelbare fout te berekenen zijn wij in staat te beoordeelen, in welke mate het gegeven cijfer betrouwbaar is.

Toch is met het opgeven van de middelbare fout de nauwkeurigheid van de keuring nog niet voldoende tot uitdrukking gebracht. Om deze laatste uitspraak te verduidelijken, stellen wij tegenover elkaar twee combinaties van twee keurmeesters, waarvan de éene combinatie keurt met een middelbare fout van 3 punten, terwijl zij bij de keuring gebruik maakt van het puntentraject van 62—74 punten. De andere combinatie — stellen wij — vertoont ook een middelbare fout van 3 punten, maar gebruikt voor dezelfde monsters het traject van 52—84 punten. Het spreekt vanzelf, dat wij de laatste combinatie als in wezen de meest nauwkeurige moeten beschouwen, hoewel zulks in de grootte van de middelbare fout in het geheel niet tot uiting komt. Absoluut genomen (d.w.z. wat de grootte van de middelbare fout betreft) zijn beide combinaties even nauwkeurig, relatief beschouwd (d.i. de grootte van de middelbare fout naast het gebruikte puntentraject genomen) is de tweede combinatie nauwkeuriger.

Wij moeten zoodoende onderscheiden in twee soorten nauwkeurigheid:

a. de subjectieve nauwkeurigheid, welke wordt uitgedrukt door een bepaalde waarde van de middelbare fout;

b. de objectieve nauwkeurigheid, welke wij zullen uitdrukken in den correlatiecoëfficiënt tusschen het door een combinatie van twee keurmeesters gegeven cijfer en het ware cijfer.

Wij komen in het volgende verschillende malen op deze onderscheiding terug en zullen dan zien, welke beteekenis aan beide soorten van nauwkeurigheid moet worden toegekend.

Men kan het best de objectieve nauwkeurigheid van de keuring (d.i. een combinatie van twee keurmeesters) tot uiting brengen door de correlatie op te geven tusschen het door de twee keurmeesters gegeven cijfer en het werkelijke cijfer. Dit lijkt een onmogelijke taak. Immers het werkelijke cijfer, het cijfer, dat gegeven zou worden door den idealen keurmeester, is niet bekend. Langs een omweg gelukt het echter de bovenbedoelde correlatie, uitgedrukt in den correlatiecoëfficiënt, te vinden, althans te benaderen.

Indien wij nl. uit gaan rekenen de middelbare afwijking van de cijfers, die een combinatie van twee keurmeesters aan de verschillende monsters heeft gegeven, t.o.v. het cijfer, dat gemiddeld aan *alle* monsters is toegekend, dan kan dit gelden als een maat voor de wijze, waarop de combinatie van beide keurmeesters de schaal heeft gebruikt.

Noemen wij het cijfer, dat door beide keurmeesters gezamenlijk aan het k^{de} monster is gegeven (k loopend van 1 tot en met N): $\bar{x}_k$.

Stellen wij verder

$$\frac{\sum\limits_{k=1}^{N} \bar{x}_k}{N} = \bar{\bar{x}} \quad \ldots\ldots\ldots\ldots\ldots \quad (1)$$

dan is $\bar{\bar{x}}$ het gemiddelde cijfer voor alle gekeurde monsters.

De bovenbedoelde middelbare afwijking, die wij $m_{\text{tot.}}$ zullen noemen (de index „tot." gebruikende, omdat wij van alle beschikbare monsters gebruik maken), kan nu op de volgende wijze worden uitgedrukt:

$$m^2_{\text{tot.}} = \frac{\sum\limits_{1}^{N} (\bar{x}_k - \bar{\bar{x}})^2}{N} \quad \ldots\ldots\ldots\ldots \quad (2)$$

Een kleine middelbare afwijking ($m_{\text{tot.}}$) wijst op het gebruik van een nauwe schaal (of op weinig variatie in kwaliteit van de onderzochte monsters), een groote middelbare afwijking geeft het omgekeerde aan.

Wij geven hieronder een voorbeeld.

Door de combinatie 1 — 2 van de gemeenschappelijke keuring te Zutphen is gegeven:

Tabel 25. Frequentieverdeeling van „totaal aantal punten" voor de combinatie (1, 2).
Materiaal: B1 (Zutphen).

Aantal punten	Aantal keeren	Berekening van het gemiddelde	Berekening van de middelbare afwijking
50	1	$1 \times -17 = -17$	$1 \times 17^2 = 289$
55	1	$1 \times -12 = -12$	$1 \times 12^2 = 144$
56	2	$2 \times -11 = -22$	$2 \times 11^2 = 242$
57	4	$4 \times -10 = -40$	$4 \times 10^2 = 400$
58	1	$1 \times -9 = -9$	$1 \times 9^2 = 81$
59	1	$1 \times -8 = -8$	$1 \times 8^2 = 64$
60	4	$4 \times -7 = -28$	$4 \times 7^2 = 196$
61	5	$5 \times -6 = -30$	$5 \times 6^2 = 180$
62	4	$4 \times -5 = -20$	$4 \times 5^2 = 100$
63	7	$7 \times -4 = -28$	$7 \times 4^2 = 112$
64	8	$8 \times -3 = -24$	$8 \times 3^2 = 72$
65	6	$6 \times -2 = -12$	$6 \times 2^2 = 24$
66	8	$8 \times -1 = -8$	$8 \times 1^2 = 8$
67	7	$7 \times 0 = 0$	$7 \times 0^2 = 0$
68	8	$8 \times +1 = +8$	$8 \times 1^2 = 8$
69	11	$11 \times +2 = +22$	$11 \times 2^2 = 44$
70	10	$10 \times +3 = +30$	$10 \times 3^2 = 90$
71	6	$6 \times +4 = +24$	$6 \times 4^2 = 96$
72	6	$6 \times +5 = +30$	$6 \times 5^2 = 150$
73	1	$1 \times +6 = +6$	$1 \times 6^2 = 36$
74	1	$1 \times +7 = +7$	$1 \times 7^2 = 49$
75	1	$1 \times +8 = +8$	$1 \times 8^2 = 64$
Totaal	103	-123	2449

Voorloopig is voor de berekening van het gemiddelde van het totaal aantal gegeven punten als hulpnulpunt 67 genomen.

Het blijkt echter, dat dit $\frac{123}{103} = 1{,}194$ te hoog is geweest. Het werkelijke gemiddelde is dus $67 - 1{,}194 \cong 65{,}8$.

De gemiddelde kwadratische afwijking ten opzichte van het hulpnulpunt is: $\frac{2449}{103} = 23{,}777$. Ter berekening van de middelbare afwijking moet hiervan worden afgetrokken:

a. het kwadraat van het verschil tusschen hulpnulpunt en werkelijk gemiddelde, dus $1{,}194^2 = 1{,}426$ [1]);

b. de correctie van Sheppard (zie blz. 23) ten bedrage van 0,083. Het kwadraat van de middelbare afwijking wordt dus: $23{,}777 - 1{,}426 - 0{,}083 = 22{,}268$. De middelbare afwijking is dus $\sqrt{22{,}268} = 4{,}72$.

Het zal thans duidelijk zijn geworden, dat het uitgaan van een hulpnulpunt slechts een middel is om de berekeningen iets te vereenvoudigen.

Het berekenen van het kwadraat van $m_{\text{tot.}}$ is de eerste stap om te komen tot een schatting van den bovenbedoelden correlatiecoëfficiënt.

Een stelling uit de waarschijnlijkheidsrekening zegt nl., dat de correlatiecoëfficiënt (r) met de partieele middelbare fout ($\sigma_{\text{part.}}$) en met de totale middelbare fout ($\sigma_{\text{tot.}}$) als volgt samenhangt:

$$\frac{\sigma^2_{\text{part.}}}{\sigma^2_{\text{tot.}}} = 1 - r^2 \quad \ldots\ldots\ldots\ldots \quad (3)$$

$\sigma_{\bar{x}}$ kunnen wij opvatten als een partieele middelbare fout. Immers de partieele middelbare fout geeft de spreiding aan van de éène grootheid bij een vaste waarde van de andere grootheid. Dezelfde functie verricht $\sigma_{\bar{x}}$. Deze is de middelbare fout van x bij een vaste waarde van $\bar{x}_k$. De totale spreiding van $\bar{x}_k$ (zonder op $\hat{x}_k$ te letten) wordt aangegeven door $m_{\text{tot.}}$. Deze is dus de totale middelbare fout.

M.a.w. we mogen formule (3) schrijven als:

$$\frac{\sigma^2_{\bar{x}}}{m^2_{\text{tot.}}} = 1 - r^2 \quad \ldots\ldots\ldots\ldots \quad (3a)$$

Daar $\sigma_{\bar{x}}$ en $m_{\text{tot.}}$ uit het materiaal te berekenen zijn, kan r met behulp van formule (3*a*) gemakkelijk gevonden worden.

Op blz. 27 vonden wij, dat $\sigma^2_{\bar{x}}$ voor de combinatie 1 — 2 (keuring Zutphen) gelijk was aan 8,87; boven vonden wij, dat $m^2_{\text{tot.}} = 22{,}268$, dus:

$$1 - r^2 = \frac{8{,}87}{22{,}268} = 0.398$$

$$\text{of:} \quad r^2 = 0{,}602$$

$$\text{of} \quad r = 0{,}776.$$

1) Indien $\bar{x}_k - \bar{\bar{x}} = u_k$ wordt genoemd en $\bar{x}_k - X = U_k$ (X = aangenomen gemiddelde), terwijl $X - \bar{\bar{x}} = d$ wordt gesteld, kan gemakkelijk worden bewezen, dat: $\overline{u^2} = \overline{U^2} - d^2$.
$\overline{u^2}$ is het gemiddelde van het kwadraat van het verschil t.o.v. het werkelijk gemiddelde.
$\overline{U^2}$ is het gemiddelde van het kwadraat van het verschil t.o.v. het aangenomen nulpunt.

Op volkomen dezelfde wijze verkregen wij voor de andere combinaties van keurmeesters (keuring te Zutphen) de volgende uitkomsten.

Tabel 26. Berekening van den correlatiecoëfficiënt voor verschillende combinaties van keurmeesters (voor „totaal aantal punten"). Materiaal: B1 (Zutphen).

	Geld.-Overijselsche keurmeesters							Gron.-Dr. keurmeesters			
Combinatie van keurmeesters	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem.
aant. monsters	103	103	103	103	103	103		103	103	103	
$m^2_{tot.}$	22,27	15,45	18,42	16,92	20,66	14,18	17,98	28,44	17,84	23,01	23,10
$m_{tot.}$	4,72	3,93	4,30	4,10	4,54	3,77	4,24	5,33	4,22	4,80	4,81
$\sigma^2_{\bar{x}}$	8,87	8,39	8,52	8,95	8,70	7,20	8,44	12,93	8,34	9,42	10,23
$1-r^2\left(=\frac{\sigma^2_{\bar{x}}}{m^2_{tot.}}\right)$	0,398	0,543	0,463	0,529	0,421	0,508	0,469	0,455	0,468	0,410	0,443
r^2	0,602	0,457	0,537	0,471	0,579	0,492	0,531	0,545	0,532	0,590	0,557
r	0,776	0,676	0,733	0,686	0,761	0,701	0,729	0,738	0,729	0,768	0,746

In tabel 26 zijn de waarden van $m^2_{tot.}$, $m_{tot.}$, $\sigma^2_{\bar{x}}$, $1 - r^2$, r^2 en r *in de kolom:* „gemiddeld" als volgt gevonden:

a. $m^2_{tot.}$ is het gemiddelde van de waarden van $m^2_{tot.}$ voor de verschillende combinaties van keurmeesters.

b. $m_{tot.}$ is de wortel uit de onder *a* gevonden waarde van $m^2_{tot.}$

$m_{tot.}$ is dus *niet* het gemiddelde van de waarden van $m_{tot.}$ voor de verschillende combinaties van keurmeesters. De laatste berekeningswijze zou principieel niet juist zijn, omdat middelbare afwijkingen niet arithmetrisch gemiddeld mogen worden. Men dient steeds de kwadraten te middelen. De wortel uit dit gemiddelde levert de juiste waarde voor de gemiddelde middelbare fout. Bij geringe variatie in $m_{tot.}$ voor de verschillende combinaties van keurmeesters wijkt de uitkomst, die men verkrijgt, als men de waarden van $m_{tot.}$ middelt, echter slechts weinig hiervan af.

c. $\sigma^2_{\bar{x}}$ is het gemiddelde van de waarden van $\sigma^2_{\bar{x}}$ voor de verschillende combinaties van keurmeesters.

d. $1 - r^2$

Geheel juiste waarden verkrijgt men slechts, indien men het quotient $\frac{\sigma^2_{\bar{x}}}{m^2_{tot.}}$ $\left(\text{hierna te noemen: } \frac{y}{x}\right)$ geometrisch vereffent. Wij geven hieronder de wijze, waarop dit kan geschieden.

Als voorbeeld nemen wij de berekening van $1 - r^2$ *voor de kolom „Geld.-Ov. keurmeesters: Gemiddeld."* Als wij de vereffende waarde voor $m = \frac{y}{x}$ $\bar{m}$ $(= \text{tg}\ \bar{\varphi})$ noemen, dan geldt (zie

(*M. J. van Uven*, Mathemathical Treatment of the Results of Agricultural and other Experiments, Groningen-Batavia, 1935, Chapter XI):

$$\text{tg } 2\,\overline{\varphi} = \frac{2\sum x_k y_k}{\sum x_k^2 - \sum y_k^2}.$$

De grootheden x_k^2, $x_k\,y_k$ en y_k^2 worden als volgt berekend:

x	y	x^2	y^2	xy
22,27	8,87	495,953	78,677	197,535
15,45	8,39	238,703	70,392	129,626
18,42	8,52	339,296	72,590	156,938
16,92	8,95	286,286	80,103	151,434
20,66	8,70	426,836	75,690	179,742
14,18	7,20	201,072	51,840	102,096
		1988,146	429,292	917,371

$$\text{tg } 2\overline{\varphi} = \frac{2 \times 917{,}371}{1988{,}146 - 429{,}292} = \frac{1834{,}742}{1558{,}854}$$

log 1834,742 = 3,26358
log 1558,854 = 3,19281
log tg $2\overline{\varphi}$ = 10,07077 — 10

$$2\overline{\varphi} = 49°38'53''$$
$$\overline{\varphi} = 24°49'27''$$
$$\log \text{tg } \overline{\varphi} = \log \overline{m} = 0{,}66518 - 1$$
$$\overline{m} = 0{,}463$$

Rekenen wij $1 - r^2$ uit door $m^2_{\text{tot.}}$ (gemiddelde van de waarden van $m^2_{\text{tot.}}$ voor de verschillende combinaties van keurmeesters) te deelen op $\sigma^2_{\overline{x}}$ (eveneens gemiddelde van de 6 waarden voor $\sigma^2_{\overline{x}}$), dan vinden wij:

$$1 - r^2 = \frac{\sigma^2_{\overline{x}}}{m^2_{\text{tot.}}} = \frac{8{,}44}{17{,}98} = 0{,}469.$$

Er is dus eenig verschil.

Daar wij de in de kolom „gemiddeld" berekende waarden voor $1 - r^2$ (resp. r) verder niet gebruiken, levert de laatste, meer eenvoudige methode voor ons een voldoend nauwkeurige uitkomst. In de volgende tabellen hebben wij daarom deze methode steeds gebruikt.

e. r^2 wordt berekend uit de onder *d* gevonden waarde van $1 - r^2$.

f. r verkrijgt men door den wortel te trekken uit de onder *e* berekende waarde van r^2.

Verschillen de r's van tabel 26 dusdanig, dat tot een essentieel verschil in objectieve nauwkeurigheid tusschen de verschillende combinaties van keurmeesters moet worden geconcludeerd?

Het is bekend, dat de onzekerheid van r kan worden uitgedrukt door de formule voor de middelbare fout van den correlatiecoëfficiënt:

$$\sigma_r^2 = \frac{(1 - r^2)^2}{N} \quad \ldots\ldots\ldots\ldots \quad (4)$$

(Veelal wordt bij deze formule als noemer de factor $N - 1$ gebruikt. Daar wij echter steeds met vrij groote waarden voor N werken, maakt het practisch geen verschil, of men den factor N dan wel $N-1$ gebruikt.)

Voor ons geval (tabel 26), waar $N = 103$ en $r^2 \simeq 0{,}5$ geeft formule (4) voor de waarde van de middelbare fout van r:

$$\sigma_r^2 = \frac{1-r^2}{\sqrt{103}} = \frac{0{,}5}{\sqrt{103}} \simeq 0{,}05.$$

Hieruit blijkt, dat geen der combinaties van keurmeesters een correlatiecoëfficiënt vertoont, waaruit men met zekerheid zou kunnen opmaken, dat de éène combinatie veel nauwkeuriger of onnauwkeuriger is geweest dan de andere.

Wij moeten hierbij opmerken, dat de berekening van de middelbare fout van r slechts een zeer beperkte beteekenis heeft, doordat de verdeeling van r (vooral indien wij werken met waarden van r, die vrij sterk van 0 afwijken) slechts zeer ten naaste bij normaal is.

De berekening van r voor de gemeenschappelijke keuring te Assen levert het volgende resultaat.

Tabel 27. Berekening van den correlatiecoëfficiënt voor verschillende combinaties van twee keurmeesters (voor „*totaal aantal punten*").

Materiaal: B2 (Assen).

	Gron.-Dr. keurmeesters	Geld.-Overijselsche keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
Aantal monsters	68	68	68	68	
$m^2_{tot.}$	15,04	30,58	24,09	31,55	28,74
$m_{tot.}$	3,88	5,53	4,91	5,62	5,36
$\sigma^2_{\bar{x}}$	9,60	6,45	6,64	9,42	7,50
r	0,602	0,888	0,851	0,837	0,859

De waarde van r voor de combinatie (1 — 2) loopt er hier iets uit, hetgeen vooral wordt veroorzaakt door de kleine $m_{tot.}$ (gebruik van een kleine schaal!).

Het aantal gekeurde monsters bedroeg op deze keuring slechts 68, zoodat men uit bovenstaande niet te vergaande conclusies mag trekken.

Voor de keuring te Alkmaar, waar slechts weinig monsters (50 of 39) werden gekeurd, waren de cijfers als volgt.

Tabel 28. Berekening van den correlatiecoëfficiënt voor verschillende combinaties van twee keurmeesters (voor „*totaal aantal punten*").

Materiaal: B4 (*Alkmaar*).

	N.-Holl. keurmeesters	Friesche keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
Aantal monsters	50	39	50	39	
$m^2_{tot.}$	43,82	40,55	47,02	42,01	43,19
$m_{tot.}$	6,62	6,37	6,86	6,48	6,57
$\sigma^2_{\bar{x}}$	8,86	14,24	9,82	4,91	9,66
r	0,893	0,805	0,890	0,940	0,880

De variaties zijn hier vrij groot (vooral in $\sigma^2_{\bar{x}}$), wat verklaarbaar is bij een zoo gering aantal monsters. $m_{tot.}$ is zeer groot in verband met het feit, dat hier de kwaliteit van de boter sterk uiteenloopt. In verband daarmede is de r ook groot.

Keuring te Leeuwarden:

Tabel 29. Als tabel 27 en 28.

Materiaal: B3 (*Leeuwarden*).

	Fr. keurmeesters
Aantal monsters	81
$m^2_{tot.}$	21,97
$m_{tot.}$	4,69
$\sigma^2_{\bar{x}}$	10,81
r	0,712

Voor smaak (en in één geval voor geur en gehalte-en-bewerking) zijn analoge berekeningen gemaakt. Wij volstaan met de mededeeling van enkele resultaten.

Voor de keuring te Zutphen:

Smaak.

Tabel 30. Berekening van den correlatiecoëfficiënt voor verschillende combinaties van twee keurmeesters (voor *smaak*).

Materiaal: B1 (*Zutphen*).

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters:	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem.
Aant. monsters	103	103	103	103	103	103		103	103	103	
$m^2_{tot.}$	0,450	0,308	0,370	0,303	0,358	0,218	0,335	0,479	0,300	0,349	0,376
$m_{tot.}$	0,671	0,555	0,608	0,550	0,598	0,467	0,579	0,692	0,548	0,591	0,613
$\sigma^2_{\bar{x}}$	0,246	0,176	0,140	0,215	0,212	0,157	0,191	0,191	0,166	0,154	0,170
r	0,673	0,655	0,789	0,539	0,639	0,529	0,656	0,775	0,669	0,748	0,740

Keuring te Assen:

Tabel 31. Als tabel 30 (dus ook voor *smaak*).

Materiaal: B2 (*Assen*).

	Gron.-Dr. keurmeesters	Geld.-Overijselsche keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
Aantal monsters	68	68	68	68	
$m^2_{tot.}$	0,313	0,780	0,577	0,731	0,696
$m_{tot.}$	0,559	0,883	0,760	0,855	0,834
$\sigma^2_{\bar{x}}$	0,152	0,152	0,178	0,211	0,180
r	0,717	0,897	0,832	0,843	0,860

Keuring te Alkmaar:

Tabel 32. Als tabel 31.

Materiaal: B4 (*Alkmaar*).

	N.-Holl. keurmeesters	Friesche keurmeesters			
Combinatie van keurmeesters:	1 — 2	3 — 4	3 — 5	4 — 5	Gemiddeld
Aantal monsters	50	50	50	50	
$m^2_{tot.}$	0,480	0,703	0,854	0,784	0,780
$m_{tot.}$	0,693	0,838	0,924	0,885	0,883
$\sigma^2_{\bar{x}}$	0,139	0,224	0,134	0,119	0,159
r	0,843	0,825	0,918	0,921	0,892

Keuring te Leeuwarden:

Tabel 33. Als tabel 31.

Materiaal: B3 (*Leeuwarden*).

	Fr. keurm.
Combinatie van keurmeesters:	1 — 2
Aantal monsters	81
$m^2_{tot.}$	0,510
$m_{tot.}$	0,714
$\sigma^2_{\bar{x}}$	0,140
r	0,852

Voor geur.

Keuring te Zutphen:

Tabel 34. Berekening van den correlatiecoëfficiënt voor verschillende combinaties van twee keurmeesters (voor *geur*).
Materiaal: B1 (*Zutphen*).

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters:	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem.
Aant. monsters	103	103	103	103	103	103		103	103	103	
$m^2_{tot.}$	0,377	0,330	0,388	0,278	0,280	0,290	0,324	0,481	0,327	0,330	0,379
$m_{tot.}$	0,614	0,574	0,623	0,527	0,529	0,538	0,569	0,694	0,572	0,574	0,616
$\sigma^2_{\bar{x}}$	0,159	0,191	0,149	0,166	0,188	0,152	0,168	0,183	0,142	0,135	0,153
r	0,760	0,649	0,785	0,632	0,574	0,691	0,694	0,788	0,752	0,770	0,772

Voor gehalte-en-bewerking.

(eveneens keuring te Zutphen).

Tabel 35. Als tabel 34, maar voor *gehalte-en-bewerking*.
Materiaal: B1 (*Zutphen*).

	Geld.-Ov. keurmeesters							Gr.-Dr. keurmeesters			
Combinatie van keurmeesters:	1—2	1—3	1—4	2—3	2—4	3—4	Gem.	5—6	5—7	6—7	Gem.
Aant. monsters	103	103	103	103	103	103		103	103	103	
$m^2_{tot.}$	0,225	0,142	0,173	0,192	0,223	0,164	0,187	0,396	0,195	0,344	0,312
$m_{tot.}$	0,474	0,377	0,416	0,438	0,472	0,405	0,432	0,629	0,442	0,586	0,559
$\sigma^2_{\bar{x}}$	0,096	0,086	0,103	0,110	0,123	0,050	0,095	0,251	0,115	0,190	0,185
r	0,757	0,628	0,636	0,653	0,670	0,834	0,701	0,605	0,642	0,669	0,638

Evenals voor „totaal aantal punten" is de waarde van $m_{tot.}$ voor smaak gemiddeld weer het grootst op de keuringen te Alkmaar en Assen.

Blijkbaar vertoonen — althans op de bedoelde keuringsdagen — de botermonsters een grootere variatie in kwaliteit dan te Leeuwarden en Zutphen.

De $m_{tot.}$ voor geur en voor smaak zijn grooter dan die voor gehalte-en-bewerking. De correlatiecoëfficiënt is gemiddeld ook wel iets lager, maar toch is het verschil niet groot.

Dit wordt veroorzaakt door de kleinere middelbare fout voor gehalte-en-bewerking.

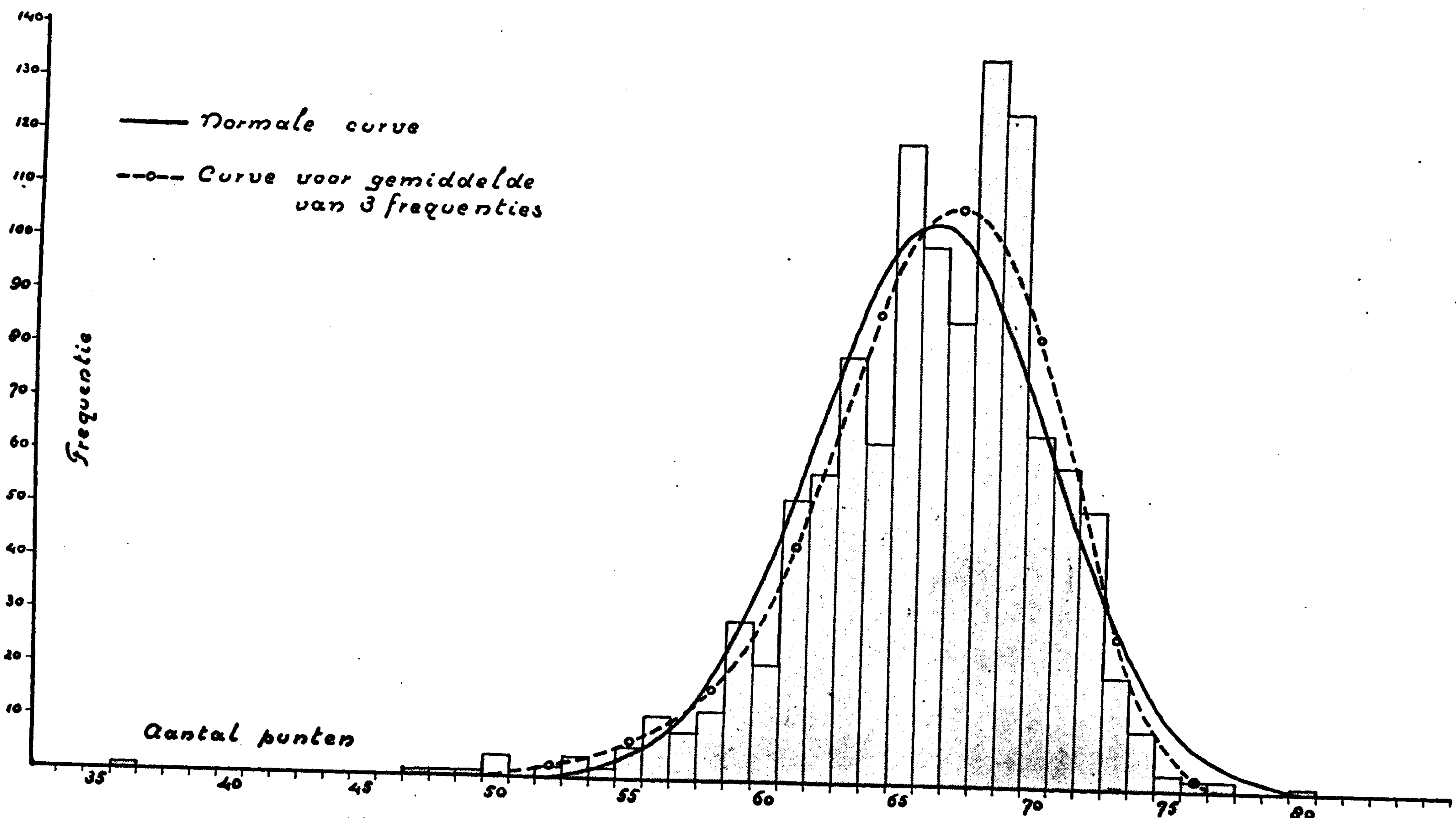

Fig. 1. Frequentieverdeeling voor „totaal aantal punten" bij 1151 monsters.

Alvorens nu over te gaan tot een bespreking van de resultaten, moeten wij opmerken, dat in het gegeven geval — theoretisch althans — aan het berekenen van een correlatiecoëfficiënt enkele bezwaren verbonden zijn. Voorwaarde voor de juistheid van de berekening is, dat $\sigma_{\bar{x}}$ en $m_{tot.}$ worden berekend uit grootheden, die een normale verdeeling vertoonen. Dit is voor $\sigma_{\bar{x}}$, zooals wij in hoofdstuk III nog nader zullen bewijzen, wel het geval, voor $m_{tot.}$ echter niet. Immers bekijken wij de frequentie van de cijfers voor „totaal aantal punten" (zie bijv. tabel 25 blz. 42), dan blijkt, dat de punten niet volgens een normale toevalscurve om het gemiddelde 65,8 verdeeld liggen. De curve klimt van de lage waarden af langzaam op en stijgt na het passeeren van het gemiddelde nog tot 69 en daalt daarna zeer snel.

In figuur 1 hebben wij het verloop van de frequentielijn afgebeeld voor het totaal aantal punten bij 1151 monsters. In deze figuur is ook de normale toevalscurve geteekend (berekend uit de middelbare fout, zooals het aanwezige materiaal die geeft).

Twee bijzonderheden vallen ons hierbij op:

a. in de empirische curve komen „diepe dalen en hooge bergen" voor, die er op wijzen, dat de keurmeesters sommige cijfers betrekkelijk weinig toekennen (bijv. 67) en andere in een meer dan normaal aantal gevallen;

b. de frequentielijn wijkt hier en daar belangrijk van de normale toevalscurve af.

Deze afwijking wordt vooral veroorzaakt door de dalen en de bergen in de empirische frequentielijn. Vatten wij echter een aantal klassen samen, dan blijken de afwijkingen veel geringer te zijn. In vorenstaande figuur hebben wij drie klassen te zamen gevoegd (66, 67 en 68 enz.) en hiervan het gemiddelde genomen en door de aldus berekende punten een vloeiende lijn (stippellijn) getrokken.

Wij geven toe, dat hierin iets willekeurigs zit. Wij hadden evengoed de klassen op andere wijze kunnen combineeren (bijv. 65, 66, 67 enz.). Maar steeds zullen wij vinden, dat nu de frequentielijn veel minder afwijkt van de normale. De neiging blijft wel duidelijk bestaan, dat de empirische curve van de lage waarden af minder snel stijgt en blijft stijgen tot boven het gemiddelde en daárna sneller daalt dan de toevalscurve.

Practisch zijn de afwijkingen echter vrij gering, zoodat wij met gerustheid mogen concludeeren, dat de door ons uitgevoerde correlatieberekeningen door de geringe afwijking van de normale curve weinig of niets van hun waarde verliezen.

Voor het cijfer voor smaak, geur en gehalte-en-bewerking hebben de frequentiecurven principieel hetzelfde — zooeven geschetste — verloop als voor „totaal aantal punten". De dalen en bergen komen hier echter niet voor. Dat deze bij „totaal aantal punten" wel voorkomen, moet verklaard worden door het feit, dat er correlatie bestaat tusschen het cijfer voor smaak, voor geur en voor gehalte-en-bewerking. Deze correlatie openbaart zich in een bovennormale resp. benedennormale frequentie voor bepaalde waarden van „totaal aantal punten".

§ 7. Bespreking van de resultaten van het voorgaande.

Wij vonden in het voorgaande waarden voor den correlatiecoëfficiënt, waarvan de meeste in de buurt van 0,8 liggen.

Het zal niet allen lezers direct duidelijk zijn, welke waarde nu moet worden toegekend aan de berekende correlatiecoëfficiënten. Teneinde het besprokene te verduidelijken, geven wij hieronder een afbeelding van twee grootheden, die een correlatiecoëfficiënt vertoonen van 0,8 (fig. 2).

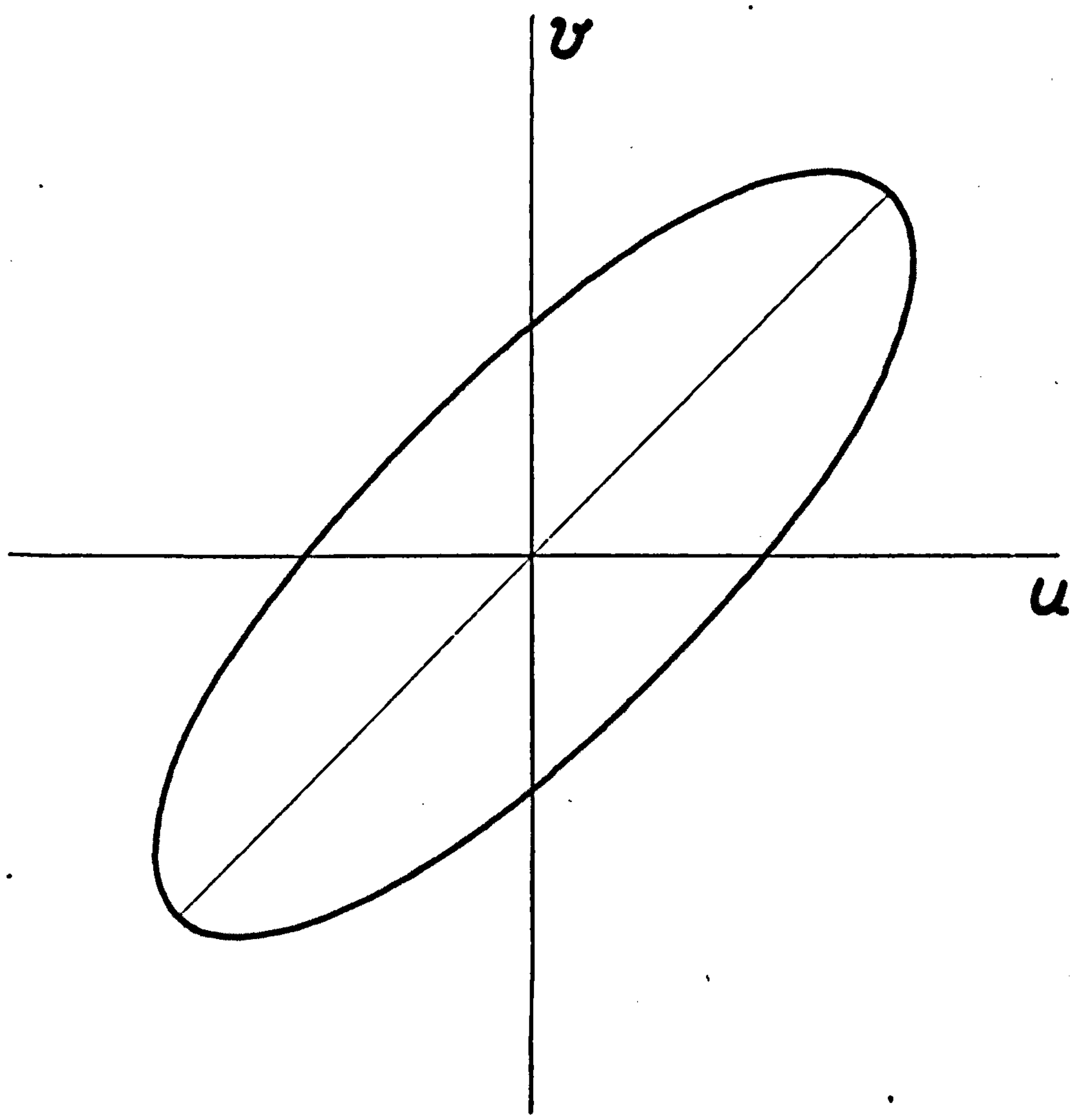

Fig. 2. Kansellips, waarbinnen bij $r = 0{,}8$ 98 % der waarnemingen vallen. (De afstand van den oorsprong tot de snijpunten van de ellips met de assen is ongeveer gelijk aan 2,8 × de partieele middelbare fout).

Ter toelichting van de figuur het volgende.

Men stelle zich voor, dat van N individuën twee grootheden (te noemen: $g^{(1)}$ en $g^{(2)}$) bekend zijn. Bijv. van N monsters boter: 1°. de „ware" cijfers van deze monsters ($\hat{x}_k$); 2°. de cijfers, die een combinatie van twee keurmeesters aan deze monsters hebben toegekend ($\bar{x}_k$).

Voor het eerste monster hebben wij gevonden de waarden $g_1^{(1)}$ en $g_1^{(2)}$, voor het $k$de monster: $g_k^{(1)}$ en $g_k^{(2)}$ enz. Wij beschikken dus in totaal over N combinaties van bepaalde waarden van $g^{(1)}$ en $g^{(2)}$.

Zooals steeds gebruikelijk is, werken wij ook hier met de afwijkingen ten opzichte van het gemiddelde. Deze zullen wij u en v noemen, dus:

$$u_k = g_k^{(1)} - \overline{g^{(1)}},$$
$$v_k = g_k^{(2)} - \overline{g^{(2)}}.$$

Elke waarneming is gekarakteriseerd door de bij haar gevonden waarden van u en v (u_k, v_k). Wij kunnen ze afbeelden als punten in een vlak door middel van een coördinatenstelsel. De waarneming (u_k, v_k) wordt weergegeven door een punt, dat als abscis u_k heeft en als ordinaat v_k. Er zullen dus, over het vlak verspreid, N punten voorkomen.

Bij het bestaan van correlatie tusschen beide grootheden zullen de punten gerangschikt kunnen worden in gelijkvormige en gelijkvormig gelegen ellipsen van gelijke kansdichtheid om hun gemeenschappelijk middelpunt. In de buurt van het gemeenschappelijke middelpunt van de ellipsen zullen verreweg de meeste punten voorkomen. Hoe verder men zich van het middelpunt verwijdert, hoe dunner de punten gezaaid zijn.

In de figuur hebben wij de punten niet geteekend, maar wel één der ellipsen nl. die, *waarbinnen (bij* $r = 0.80$) 98 % *van de punten vallen.*

Hoe langgerekter de ellips, hoe sterker de correlatie tusschen de grootheden is. Bij een waarde van $r = 1$ zou de ellips zijn samengedrukt tot een lijn. Bij $r = 0$ is de ellips overgegaan in een cirkel.

De figuur laat zien, hoe groot de spreidingsmogelijkheid van de ééne grootheid is bij een vaste waarde van de andere m.a.w. hoe sterk de correlatie tusschen beide grootheden is.

Op blz. 63 bij de bespreking van fig. 4 komen wij op deze voorstelling van den correlatiecoëfficiënt uitvoerig terug. Daar hebben wij ook ellipsen aangegeven, die de beteekenis van andere waarden van den correlatiecoëfficiënt dan 0,80 verduidelijken.

Uit het bovenstaande en ook uit hetgeen vermeld is bij de bespreking van de berekende middelbare fouten (§ 4) zal het den lezer duidelijk zijn, dat de keuring niet met een *groote* mate van nauwkeurigheid plaats heeft. Zeer zeker zijn de keurmeesters in staat *belangrijke* verschillen aan te toonen, maar aan kleine verschillen (bijv. aan 3 à 4 punten) kan in het algemeen toch geen *essentieele* waarde worden toegekend.

Er is zeker reden om na te gaan, op welke wijze de nauwkeurigheid van het toegekende puntenaantal kan worden vergroot.

Dit kan op twee manieren worden bereikt:

I. door meer dan twee keurmeesters aan de vaststelling van de punten voor een monster te laten deelnemen.

De waarde van $\sigma_{\bar{x}}$ (d.i. de middelbare fout van de keuring) wordt hierdoor verkleind. Indien nl. eerst n keurmeesters keuren en men bereikt een nauwkeurigheid van $\sigma_{\bar{x}}$, dan kan worden verwacht, dat bij het keuren van n' keurmeesters een $\sigma'_{\bar{x}}$ wordt bereikt van

$$\sigma'_{\bar{x}} = \sigma_{\bar{x}} \times \sqrt{\frac{n}{n'}};$$

m.a.w.

als een combinatie van twee keurmeesters keurt met een middelbare fout van 3,0 punten, keurt een combinatie van drie keurmeesters met een middelbare fout van $3{,}0 \times \sqrt{\frac{2}{3}} \simeq 2{,}5$ punt, een combinatie van vier keurmeesters met een middelbare fout van $3{,}0 \times \sqrt{\frac{2}{4}} \simeq 2{,}0$ punt.

De „winst" is dus niet evenredig aan de vermeerdering van het aantal keurmeesters, maar toch is dit een effectief middel om de waarde van de keuring op te voeren.

In Denemarken heeft men dit middel toegepast om de zekerheid van het toegekende cijfer te vergrooten. Zooals reeds op blz. 9 werd medegedeeld worden de monsters, ingezonden voor de Staatsboterkeuring, beoordeeld door negen keurmeesters.

De vaststelling van het cijfer wordt hierdoor $\sqrt{\frac{9}{2}} = 2{,}1$ maal zoo nauwkeurig als bij de vaststelling door twee keurmeesters.

Toch is de vooruitgang in objectieve nauwkeurigheid door vergrooting van het aantal keurmeesters niet zoo groot als boven werd berekend voor de subjectieve nauwkeurigheid (middelbare fout). Immers te verwachten is, dat de variatie in punten, die de keurmeesters toekennen, door het combineeren van vele keurmeesters kleiner zal worden (de uiterste cijfers — de zeer lage en zeer hooge — zullen naar het gemiddelde worden toegetrokken).

Men zou kunnen opmerken, dat bij vergrooting van het aantal keurmeesters ook wel eens een grootere variatie in punten gevonden zou kunnen worden, nl. als de toegevoegde keurmeesters een veel wijdere schaal gebruiken en/of hun middelwaarde ver afligt van die van de reeds aanwezige keurmeesters.

Inderdaad is dit theoretisch juist. Zou een dergelijk geval zich voordoen, dan zou de mogelijkheid niet uitgesloten zijn, dat bij vergrooting van het aantal keurmeesters een grootere variatie optrad.

In de practijk speelt echter de afwijking van de middelwaarde geen rol, doordat — zooals wij in hoofdstuk I § 2 bewezen — het gemiddelde oordeel van de keurmeesters zeer weinig uiteenloopt. Ook met het andere punt — toevoeging van vele keurmeesters met groote dispersie — behoeven wij practisch niet te rekenen, omdat de kans, dat steeds keurmeesters met groote dispersie worden toegevoegd, zeer klein is. Te verwachten is, dat er bij de toegevoegde keurmeesters ook zijn, die een kleinere dispersie dan de gemiddelde hebben.

In het algemeen mogen wij bij stijging van het aantal keurmeesters een kleinere dispersie verwachten. Wij zullen dit straks ook bewijzen.

In welke mate het aantal keurmeesters invloed uitoefent op de objectieve nauwkeurigheid der keuring, leert onderstaande berekening.

Berekening van de afhankelijkheid van de objectieve nauwkeurigheid van het aantal keurmeesters.

Stel, dat p keurmeesters N monsters keuren en dat aan het gemiddelde van deze monsters het cijfer X toekomt.

Keurmeester 1 geeft aan het k^{de} monster het cijfer: $x_k^{(1)}$, keurmeester l aan hetzelfde monster: $x_k^{(l)}$ en ten slotte keurmeester p: $x_k^{(p)}$.

De waarde van $m_{\text{tot.}}$ wordt in dit geval:

$$m_{\text{tot.}}^2 = \frac{\sum\limits_{k=1}^{N}\left(\frac{x_k^{(1)} + \ldots + x_k^{(l)} + \ldots + x_k^{(p)}}{p} - X\right)^2}{N} =$$

$$= \frac{\Sigma\left\{\frac{(x_k^{(1)} - X) + \ldots + (x_k^{(l)} - X) + \ldots + (x_k^{(p)} - X)}{p}\right\}^2}{N} =$$

$$= \frac{1}{p^2}\left\{\Sigma\frac{(x_k^{(1)} - X)^2}{N} + \ldots + \Sigma\frac{(x_k^{(l)} - X)^2}{N} + \ldots + \Sigma\frac{(x_k^{(p)} - X)^2}{N} + \right.$$

$$+ 2\Sigma\frac{(x_k^{(1)} - X)(x_k^{(l)} - X)}{N} + \ldots + 2\Sigma\frac{(x_k^{(1)} - X)(x_k^{(p)} - X)}{N} + \ldots +$$

$$\left. + 2\Sigma\frac{x_k^{(l)} - X)(x_k^{(p)} - X)}{N}\right\} \quad \ldots \ldots \ldots \ldots (1)$$

Nu geldt:

$$\Sigma\frac{(x_k^{(1)} - X)(x_k^{(l)} - X)}{N} = \Sigma\frac{(x_k^{(1)} - X)(x_k^{(1)} - X + x_k^{(l)} - x_k^{(1)})}{N} =$$

$$= \Sigma\frac{(x_k^{(1)} - X)^2}{N} + \Sigma\frac{(x_k^{(1)} - X)(x_k^{(l)} - x_k^{(1)})}{N} \quad \ldots \ldots (2)$$

Teneinde den laatsten term van vergelijking (2) verder te gaan bewerken, stellen wij

$$x_k^{(1)} - X = w_k + t_k^{(1)},$$

en

$$x_k^{(l)} - X = w_k + t_k^{(l)};$$

m.a.w. wij beschouwen het verschil, dat keurmeester 1 resp. l bij monster k constateert ten opzichte van de middelwaarde X, als te zijn opgebouwd uit twee deelen nl. als eerste deel: de ware afwijking (w_k) d.i. de afwijking, die de ideale keurmeester zou vinden (dus $w_k = \hat{x}_k - X$) en als tweede deel: de fout t_k.

Dus geldt:

$$\Sigma\frac{(x_k^{(1)} - X)(x_k^{(l)} - x_k^{(1)})}{N} = \Sigma\frac{(w_k + t_k^{(1)})(t_k^{(l)} - t_k^{(1)})}{N} =$$

$$= \Sigma\frac{w_k t_k^{(l)}}{N} - \Sigma\frac{w_k t_k^{(1)}}{N} + \Sigma\frac{t_k^{(1)} t_k^{(l)}}{N} - \Sigma\frac{t_k^{(1)^2}}{N} \quad \ldots \ldots (3)$$

Bij groote waarden van N mogen wij aannemen, dat

$$\Sigma\frac{w_k t_k^{(l)}}{N} \simeq < w_k t_k^{(l)} >, \quad \Sigma\frac{w_k t_k^{(1)}}{N} \sim < w_k t_k^{(1)} >$$

$$\text{en } \Sigma\frac{t_k^{(1)} t_k^{(l)}}{N} \simeq < t_k^{(1)} t_k^{(l)} >.$$

Daar w_k en $t_k^{(1)}$ als onafhankelijk van $t_k^{(l)}$ mogen worden beschouwd, terwijl hetzelfde geldt voor $t_k^{(1)}$ en $t_k^{(l)}$, is

$$< w_k \, t_k^{(l)} > = 0, \quad < w_k \, t_k^{(1)} > = 0 \quad \text{en} \quad < t_k^{(1)} \, t_k^{(l)} > = 0.$$

Hierdoor gaat (3) over in

$$\Sigma \frac{(x_k^{(1)} - X) \, (x_k^{(l)} - x_k^{(1)})}{N} = - \Sigma \frac{t_k^{(1)^2}}{N}.$$

Teneinde de berekening niet ingewikkelder te maken dan noodig is, nemen wij aan, dat alle p keurmeesters met dezelfde middelbare fout (σ) keuren en dat zij ook dezelfde schaal gebruiken, m.a.w. dat wij mogen stellen:

$$\frac{\Sigma \, t_k^{(1)^2}}{N} = \ldots . = \frac{\Sigma \, t_k^{(l)^2}}{N} = \ldots . = \frac{\Sigma \, t_k^{(p)^2}}{N} = \sigma^2$$

en: $$\Sigma \frac{(x_k^{(1)} - X)^2}{N} = \ldots . = \Sigma \frac{(x_k^{(l)} - X)^2}{N} = \ldots . = \Sigma \frac{(x_k^{(p)} - X)^2}{N} = m^2.$$

(Beide vergelijkingen gelden slechts bij voldoend groote N.)

De veelterm van de rechterzijde van vergelijking (1) bestaat nu uit p^2 termen nl.

p termen van den vorm: $\Sigma \dfrac{(x_k^{(l)} - X)^2}{N}$, elk in waarde gelijk aan m^2,

$2 \times \dfrac{p(p-1)}{1 \times 2} = p^2 - p$ termen van den vorm: $\Sigma \dfrac{(x_k^{(l)} - X) \, (x_k^{(m)} - X)}{N}$

met een waarde gelijk aan: $m^2 - \sigma^2$.

Wij vinden dus uit (1)

$$m_{\text{tot.}}^2 = \frac{1}{p^2} \{pm^2 + (p^2 - p) \, (m^2 - \sigma^2)\} = m^2 - \frac{p-1}{p} \sigma^2 \, . \quad . \quad . \quad . \quad . \quad (4)$$

$$\left(= m^2 - \sigma^2 + \frac{1}{p} \sigma^2 \right).$$

Hieruit blijkt dus al, dat de totale variatie ($m_{\text{tot.}}$) kleiner wordt bij vergrooting van het aantal keurmeesters (p).

In welke mate $m_{\text{tot.}}^2$ verandert bij het stijgen van het aantal keurmeesters is gemakkelijk af te leiden uit formule (4). Onderstaande tabel geeft de waarden voor $m_{\text{tot.}}^2$, als wij aannemen, dat $m^2 = \dfrac{1{,}36}{0{,}72} \sigma^2$ (een vergelijking, die wij op blz. 59 zullen afleiden).

Tabel 36. Waarde van $m^2_{tot.}$ bij een bepaald aantal keurmeesters, afgeleid uit formule (4), waarbij aangenomen is, dat $m^2 = \frac{1,36}{0,72} \sigma^2$.

Aantal keurmeesters	Waarde van $m^2_{tot.}$
1	$\sigma^2 \times 1,89$
2	$\sigma^2 \times 1,39$
3	$\sigma^2 \times 1,22$
4	$\sigma^2 \times 1,14$
5	$\sigma^2 \times 1,09$
6	$\sigma^2 \times 1,06$
7	$\sigma^2 \times 1,03$
8	$\sigma^2 \times 1,01^5$
∞	$\sigma^2 \times 0,89$

Bij grooter wordende p wordt dus de verkleining hoe langer hoe geringer. Dit volgt ook oogenblikkelijk uit (4) door differentiatie:

$$\Delta\,(m^2_{tot.}) = -\sigma^2 \times \frac{1}{p^2} \Delta\, p \quad \ldots\ldots\ldots\ldots \quad (5)$$

waarbij $\Delta(m^2_{tot.})$ de verandering voorstelt van $m^2_{tot.}$, indien p toeneemt met Δp. De verkleining van $m^2_{tot.}$ is omgekeerd evenredig met het kwadraat van het aantal keurmeesters.

Wij dienen thans terug te keeren tot ons eigenlijke vraagstuk nl. in welke mate de objectieve nauwkeurigheid (r) stijgt bij het toenemen van het aantal keurmeesters.

Daartoe gaan we uit van formule (3a) uit § 6 (blz. 43):

$$\frac{\sigma^2_{\bar{x}}}{m^2_{tot.}} = 1 - r^2 \quad \ldots\ldots\ldots\ldots \quad (6)$$

$\sigma_{\bar{x}}$ is in dit geval gelijk aan de middelbare fout bij de beoordeeling door een combinatie van p keurmeesters. Daar σ de middelbare fout van één keurmeester aanduidt, bedraagt de middelbare fout van een combinatie van p keurmeesters: $\frac{\sigma}{\sqrt{p}}$.

Vullen wij nu in (6) in:

$$\sigma^2_{\bar{x}} = \frac{\sigma^2}{p}$$

en

$$m^2_{tot.} = m^2 - \frac{p-1}{p}\sigma^2$$

dan gaat (6) over in

$$\frac{\sigma^2}{pm^2 - (p-1)\sigma^2} = 1 - r^2 \quad \ldots\ldots\ldots\ldots \quad (7)$$

Deze vergelijking geeft dus de afhankelijkheid van de objectieve nauwkeurigheid van het aantal keurmeesters.

De afhankelijkheid wordt echter ook beinvloed door de verhouding, die er bestaat tusschen σ^2 en m^2. Welke waarde deze verhouding ongeveer heeft, kunnen we afleiden uit vroeger verkregen gegevens. Wij vonden immers, dat bij een combinatie van twee keurmeesters ($p = 2$) r een waarde heeft van ongeveer 0,80 (zie blz. 53). Wij gaan nu verder van deze waarde uit. Door $p = 2$ en $r = 0{,}80$ in te vullen, krijgen wij

$$m^2 = \frac{1{,}36}{0{,}72}\sigma^2 = \frac{17}{9}\sigma^2.$$

Substitutie van $m^2 = \frac{17}{9}\sigma^2$ in (7) geeft

$$\frac{1}{1 + \frac{8}{9}p} = 1 - r^2 \quad \ldots\ldots\ldots\ldots \quad (8)$$

Vergelijking (8) geeft dus de afhankelijkheid van r van het aantal keurmeesters, *waarbij wij er van zijn uitgegaan, dat voor $p = 2$ $r = 0{,}80$ is.*

Wij kunnen uit deze vergelijking de waarde van r berekenen bij het keuren door een combinatie van een bepaald aantal keurmeesters.

Onderstaande tabel geeft een overzicht.

Tabel 37. Waarde van den correlatiecoëfficiënt (r) bij een bepaald aantal keurmeesters.

Aantal keurmeesters	Waarde van $1-r^2$	Waarde van r	Waarde van $\sqrt{1-r^2}$
1	0,53	$0{,}68^{6}$	$0{,}72^{8}$
2	0,36	0,80	0,60
3	$0{,}27^{3}$	$0{,}85^{3}$	$0{,}52^{1}$
4	0,22	$0{,}88^{3}$	$0{,}46^{9}$
5	$0{,}18^{4}$	$0{,}90^{3}$	$0{,}42^{9}$
6	$0{,}15^{8}$	$0{,}91^{8}$	$0{,}39^{8}$
7	$0{,}13^{9}$	$0{,}92^{8}$	$0{,}37^{3}$
8	$0{,}12^{3}$	$0{,}93^{6}$	$0{,}35^{0}$
9	$0{,}11^{1}$	$0{,}94^{3}$	$0{,}33^{3}$

In achterstaande grafiek is dit in beeld gebracht (zie blz. 60).

Wij hebben in bovenstaande tabel en in de grafiek naast de waarde van r de waarde van $\sqrt{1-r^2}$ gegeven. Wij hebben dit gedaan, omdat voor het nagaan van de correlatie de waarde van $\sqrt{1-r^2}$ een ten minste even goed, zoo niet beter beeld geeft dan r zelf, hoewel deze laatste de meer gebruikelijke maat is.

In het verloop van beide waarden ($\sqrt{1-r^2}$ en r) komt beide sterk tot uiting, dat het vergrooten van het aantal keurmeesters de nauwkeurigheid verhoogt.

Het effect wordt echter bij een groot aantal keurmeesters relatief veel minder.

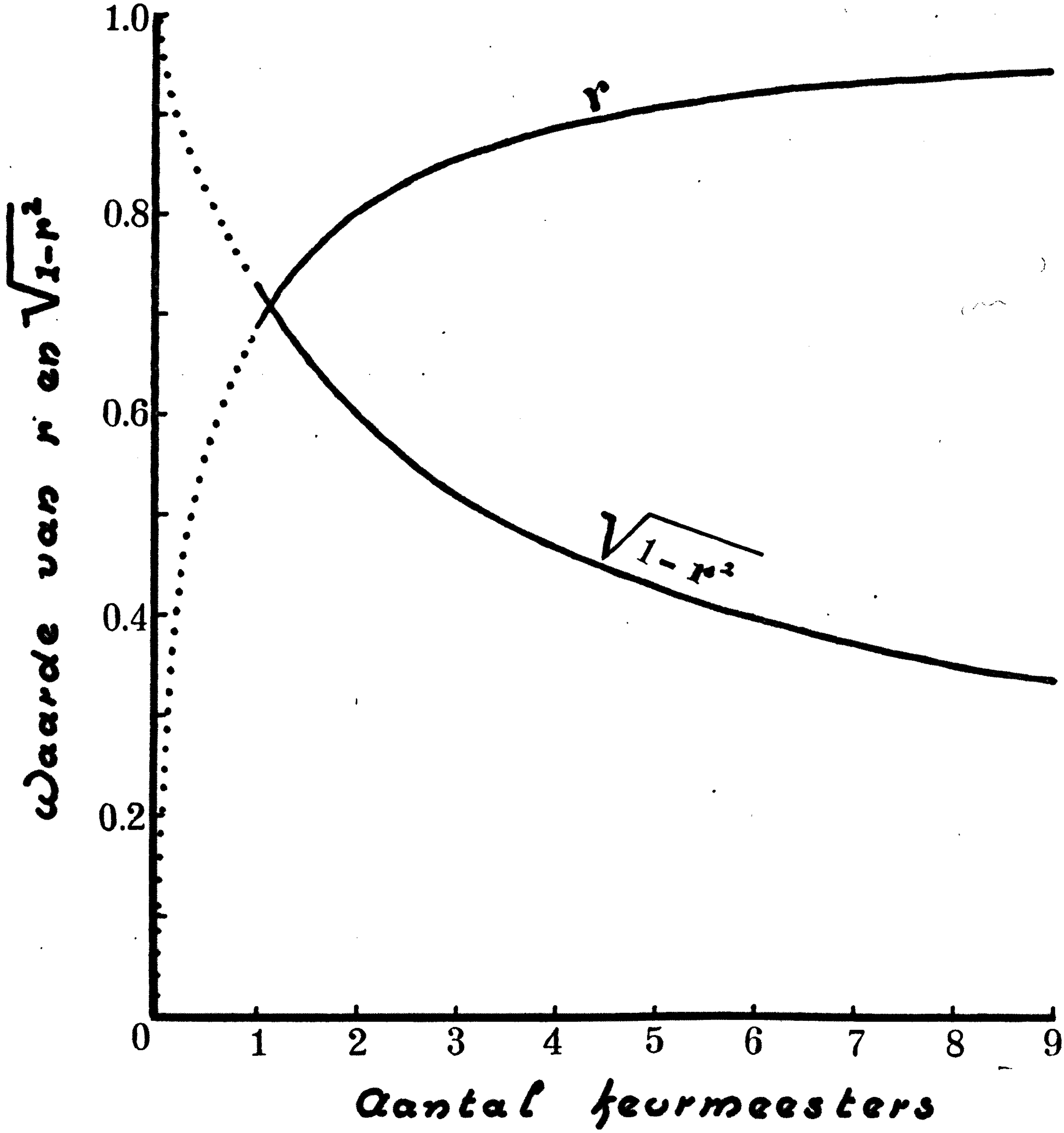

Fig. 3. De afhankelijkheid van de waarde van r resp. $\sqrt{1-r^2}$ van het aantal keurmeesters.

Teneinde den lezer een indruk te geven, in hoeverre de verkleining van $m_{tot.}$ bij keuren door meer keurmeesters de vergrooting van de objectieve nauwkeurigheid heeft tegengewerkt, geven wij hieronder de waarde van $\sqrt{1-r^2}$ resp. r, berekend in de veronderstelling, dat $m_{tot.}$ bij vergrooting van het aantal keurmeesters niet verandert, maar gelijk blijft aan de waarde van $m_{tot.}$ voor twee keurmeesters.

Tabel 38. Vergelijking van de grootte van den correlatiecoëfficiënt (en van de waarde van $\sqrt{1-r^2}$) 1°. bij gelijkblijvende $m_{tot.}$, 2°. bij veranderde $m_{tot.}$.

Aantal keurmeesters	Waarde van $\sqrt{1-r^2}$ resp. r bij gelijkblijvende $m_{tot.}$		Waarde van $\sqrt{1-r^2}$ resp. r bij veranderde $m_{tot.}$	
	$\sqrt{1-r^2}$	r	$\sqrt{1-r^2}$	r
1	0,849	0,529	0,728	0,686
2	0,60	0,80	0,60	0,80
3	0,490	0,872	0,521	0,853
4	0,424	0,905	0,469	0,883
5	0,379	0,925	0,429	0,903
6	0,346	0,938	0,398	0,918
7	0,321	0,947	0,373	0,928
8	0,30	0,954	0,350	0,936
9	0,284	0,959	0,330	0,943

Het verschil is dus niet zoo belangrijk.

Meetkundige voorstelling van de afhankelijkheid van de objectieve nauwkeurigheid van het aantal keurmeesters.

Wij kunnen de mate, waarin de nauwkeurigheid bij het keuren door een grooter aantal keurmeesters wordt verhoogd, op de volgende wijze aanschouwelijk voorstellen (zie ook blz. 53).

Wij teekenen daartoe een rechthoekig coördinatenstelsel met als horizontale as de u_1-as, als verticale de u_2-as. Hierbij is

$$u_1 = \hat{x} - \hat{\bar{x}},$$

$$u_2 = \bar{x} - \bar{\bar{x}};$$

u_1 geeft dus voor een bepaald monster de afwijking van het ware cijfer t.o.v. de middelwaarde, u_2 de afwijking van het door p keurmeesters gezamenlijk gegeven cijfer van de middelwaarde.

Elk monster wordt nu afgebeeld als een punt, welks coördinaten u_1 en u_2 zijn. De zoo verkregen punten groepeeren zich in ellipsen (van gelijke kansdichtheid) om een gemeenschappelijk middelpunt. De punten op een bepaalde ellips stellen dus waarnemingen voor, die alle even waarschijnlijk zijn.

Deze ellipsen worden weergegeven door de volgende vergelijking:

$$h_1^2 u_1^2 - 2h_1 h_2 r u_1 u_2 + h_2^2 u_2^2 = \Omega \quad \ldots\ldots\ldots \quad (9)$$

Hierin is volgens de bekende formule uit de correlatierekening:

$$h_1^2 = \frac{1}{2\sigma_1^2 (1 - r^2)} \quad \text{en} \quad h_2^2 = \frac{1}{2\sigma_2^2 (1 - r^2)}.$$

$$(\sigma_1^2 = \overline{u_1^2}) \qquad\qquad (\sigma_2^2 = \overline{u_2^2})$$

Op de beteekenis van Ω komen wij straks terug (blz. 63).

Voor ons geval is

$$\sigma_2^2\,(=m^2_{\mathrm{tot.}}) = m^2 - \frac{p-1}{p}\sigma^2 \quad \ldots . \text{(zie formule(4) blz. 57)}$$

en

$$\sigma_1^2\,(= \underset{\lim p=\infty}{m^2_{\mathrm{tot.}}}) = m^2 - \sigma^2.$$

Wij krijgen dus

$$h_1^2 = \frac{1}{2(m^2-\sigma^2)\,(1-r^2)} \quad \ldots\ldots\ldots\ldots (10a)$$

$$h_2^2 = \frac{1}{2\left(m^2 - \frac{p-1}{p}\,\sigma^2\right)(1-r^2)} \quad \ldots\ldots (10b)$$

Daar

$$1 - r^2 = \frac{\frac{\sigma^2}{p}}{m^2 - \frac{p-1}{p}\,\sigma^2} \qquad \text{(zie formule (7) blz 58)}$$

is

$$r^2 = \frac{m^2-\sigma^2}{m^2 - \frac{p-1}{p}\,\sigma^2} = \frac{h_2^2}{h_1^2}$$

of

$$r = \frac{h_2}{h_1} \quad \ldots\ldots\ldots\ldots\ldots (11)$$

(Deze formule zullen wij op blz. 65 gebruiken.)

Formule (9) gaat hierdoor over in

$$h_1^2u_1^2 - 2h_2^2u_1u_2 + h_2^2u_2^2 = \Omega \quad \ldots\ldots\ldots (12)$$

Vullen wij nu in $m^2 = \frac{17}{9}\sigma^2$ (zie blz. 59), dan vinden wij

$$1 - r^2 = \frac{1}{1+\frac{9}{8}p} \qquad \text{(formule (8) blz. 59)}$$

en volgens (10a): $h_1^2 = \dfrac{1}{2\sigma^2 \times \frac{8}{9} \times (1-r^2)} = \dfrac{\frac{9}{8}\,(1+\frac{8}{9}p)}{2\,\sigma^2},$

volgens (10b): $h_2^2 = \dfrac{1}{2\sigma^2 \times \left(\frac{8}{9}+\frac{1}{p}\right)(1-r^2)} = \dfrac{p}{2\sigma^2}.$

Vergelijking (12) gaat dus over in:

$$(\tfrac{9}{8}+p)u_1^2 - 2pu_1u_2 + pu_2^2 = \Omega \times 2\sigma^2 \quad \ldots\ldots\ldots (13)$$

Vergelijking (13) geeft, als Ω en σ constant worden gehouden en het aantal keurmeesters (p) als parameter wordt opgevat, een *ellipsenbundel*.

Schrijven wij (13) als

$$p(u_1-u_2)^2 + \tfrac{9}{8}\,u_1^2 = \Omega \times 2\sigma^2,$$

dan blijkt, dat de volgende lijnen tot den bundel behooren:

a. $p = 0$: $u_1 = \pm \sqrt{\frac{16}{9}\Omega \times \sigma^2}$ (dus twee lijnen $\perp$ op de u_1-as);

b. $p = \infty$: $u_1 = u_2$ (dus een lijn door den oorsprong, die een hoek van 45° maakt met de u_1- en u_2-as).

De beide snijpunten van de lijn: $u_1 = u_2$ en de lijnen: $u_1 = \pm \sqrt{\frac{16}{9}\Omega \times \sigma^2}$ zijn gemeenschappelijk eigendom van alle ellipsen van den bundel d.w.z. alle ellipsen raken elkaar in deze beide punten. De helling van de groote as van de ellipsen ten opzichte van de u_1-as (φ) kan worden berekend uit:

$$\mathrm{tg}\, 2\varphi = \frac{-2p}{(\frac{9}{8} + p) - p} = -\tfrac{16}{9}\, p.$$

Wij moeten ons thans nog bezighouden met de beteekenis van den factor Ω. Uit de waarschijnlijkheidsrekening is bekend, dat Ω samenhangt met de kans (P), dat de waarnemingen zich binnen de ellips bevinden. Wenschen wij nl. een ellips te construeeren, waarbinnen de waarnemingen met een kans P liggen, dan geldt:

$$\Omega = {}^e\!\log \frac{1}{1 - P}.$$

Nemen wij nu: $P = 98\,\%$, dan wordt Ω:

$$\Omega = {}^e\!\log \frac{1}{1 - 0{,}98} = {}^e\!\log 50 = 3{,}9121.$$

Bij onze verdere beschouwingen zijn wij nu uitgegaan van een waarde van $\Omega = 4$ (overeenkomende met $P = 98{,}17\,\%$).

Voor de waarde van σ hebben wij genomen: $2{,}8\sqrt{2}$. Voor een combinatie van twee keurmeesters vonden wij immers: $\sigma_{\bar{x}} = 2{,}8$ (zie blz. 30).

Vullen wij in vergelijking (13) in: $\Omega = 4$ en $\sigma^2 = (2{,}8\sqrt{2})^2 = 15{,}68$, dan krijgen wij

$$(\tfrac{9}{8} + p)u_1^2 - 2pu_1u_2 + pu_2^2 = 4 \times 2 \times 15{,}68 \quad . \; . \; . \; . \; . \; (14)$$

Deze vergelijking geeft de ellipsen weer, waarbinnen de waarnemingen met 98 % kans (eigenlijk 98,17 %) liggen.

In fig. 4 zijn een aantal ellipsen geteekend (nl. voor $p = 0$, $p = 1$, $p = 2$, $p = 3$, $p = 9$ en $p = \infty$).

Hierbij vallen de volgende bijzonderheden op:

a. bij grooter wordende p wordt vooral de kleine as der ellipsen hoe langer hoe kleiner.

De teekening laat duidelijk zien, in welke mate de afwijkingen van het door de keurmeesters gevonden cijfer van de ware waarde kleiner worden bij het toenemen van het aantal keurmeesters.

Men krijgt uit de teekening ook een goeden indruk, welke afwijkingen nog mogelijk zijn bij het bestaan van een bepaalde correlatie (welke beteekenis dus moet worden gehecht aan een bepaalde waarde van den correlatiecoëfficiënt).

b. alle ellipsen raken twee lijnen ($u_1 = \pm \sqrt{\frac{16}{9}\Omega \times \sigma^2}$), loodrecht op de

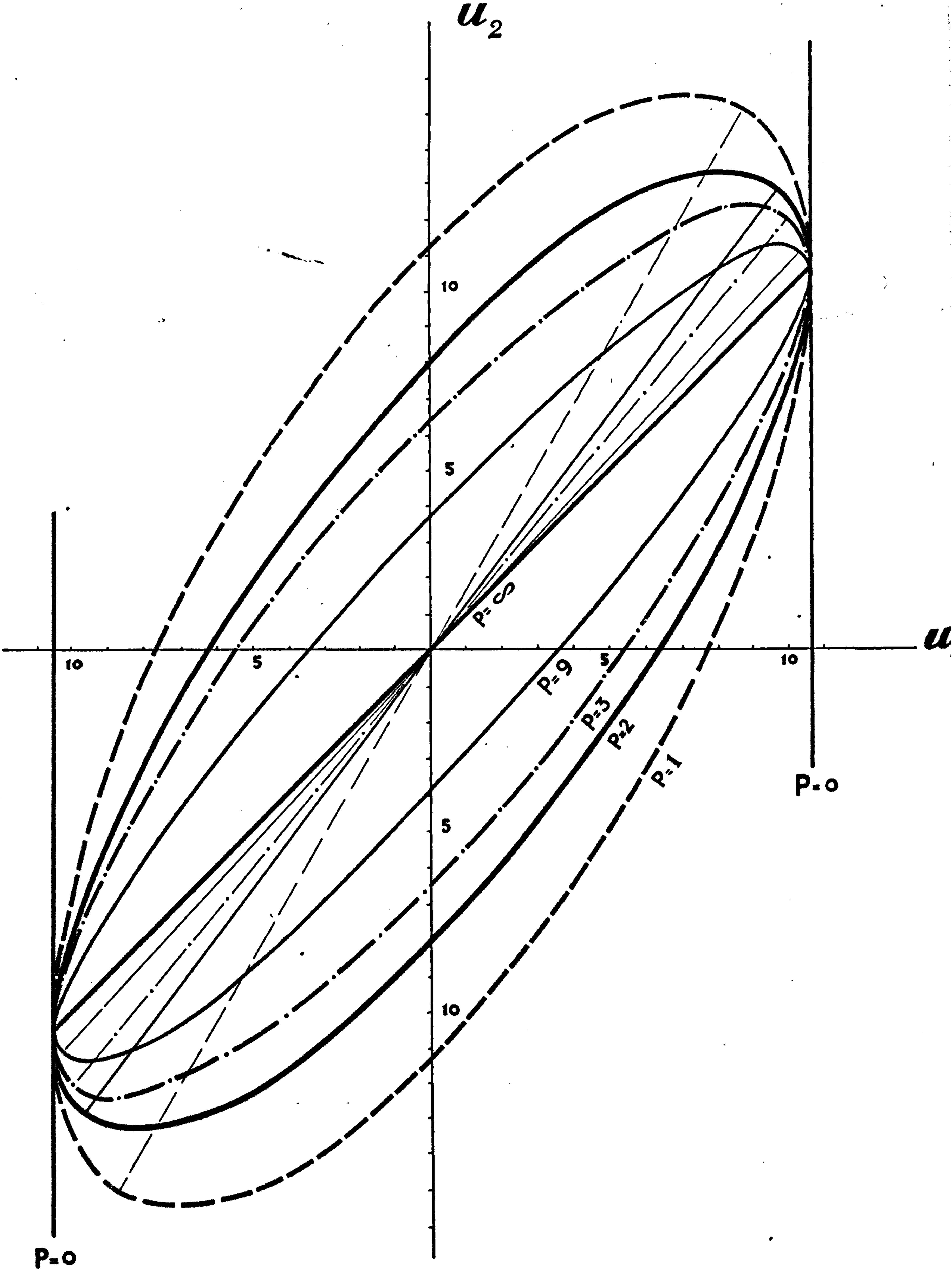

Fig. 4. Kansellipsen, waarbinnen 98 % der waarnemingen vallen.

u_1-as m.a.w. de variatie van u_1 (d.i. dus de afwijking van de ware waarde van de middelwaarde) is altijd gelijk, ongeacht hoeveel keurmeesters keuren. Het resultaat is volkomen vanzelfsprekend: de variatie van de ware cijfers is natuurlijk onafhankelijk van het aantal keurmeesters.

De variatie van u_2 ($m_{tot.}$) wordt bij het stijgen van het aantal keurmeesters geringer (zie blz. 58). Dit is ook in de figuur te zien. De twee lijnen, loodrecht op de u_2-as, die aan de ellipsen raken, schuiven bij grooter worden van p dichter naar den oorsprong. Beide lijnen bereiken haar limietstand voor $p = \infty$. Dan gaan zij door de beide raakpunten, die gemeenschappelijk zijn aan alle ellipsen.

c. de regressievergelijking d.i. de vergelijking, die aangeeft, welk cijfer een combinatie van p keurmeesters gemiddeld zal geven aan een monster met een bepaald „waar" cijfer — is onafhankelijk van het aantal keurmeesters. Immers de regressievergelijking luidt in het algemeen:

$$\bar{u}_2 = r \times \frac{h_1}{h_2} \times u_1.$$

Volgens (11) is echter: $r = \frac{h_2}{h_1}$.

De vergelijking van de regressielijn wordt dus:

$$\bar{u}_2 = u_1.$$

Hieruit blijkt, dat deze onafhankelijk is van p.

Gemiddeld zullen dus — ongeacht, hoeveel keurmeesters keuren — de keurmeesters het „ware" cijfer aan een monster toekennen.

d. het stuk van de lijn $u_1 = u_2$, besloten tusschen de beide raakpunten, moet ook beschouwd worden als een ellips nl. één, waarvan de korte as de lengte 0 heeft ($p = \infty$).

De beide lijnen: $u_1 = \pm \sqrt{\frac{16}{9} \Omega \times \sigma^2}$ dienen eveneens te worden opgevat als een ellips, en wel als één, die zich pas in het oneindige sluit.

Conclusie uit het bovenstaande (blz. 55 tot 65).

Vergrooting van het aantal keurmeesters is een effectief middel om de zekerheid van het oordeel der keurmeesters op te voeren.

De „winst" wordt bij het toenemen van het aantal keurmeesters wel hoe langer hoe geringer (zie ook tabel 37 op blz. 59 en figuur 3 op blz. 60), zoodat het de vraag is, of de grootere nauwkeurigheid, die men verkrijgt, wel opweegt tegen de hoogere kosten, die men moet maken, als men de boter door *vele* keurmeesters laat keuren.

Ervan uitgaande, dat het van groot belang is, dat de fabriek een zoo juist en nauwkeurig mogelijk oordeel over haar boter tot haar beschikking krijgt, zonder dat dit te groote financieele offers vraagt, mogen wij uit het bovenstaande wel concludeeren, dat *verhooging van het aantal keurmeesters van het gebruikelijke tweetal op drie of vier keurmeesters zeer wenschelijk is.*

Zijn er aan het gegeven oordeel voor de fabrieken ingrijpende consequenties verbonden, dan moet men dit laatste aantal als eisch stellen, wil men niet gevaar loopen onbillijkheden te begaan.

Wij merken op, dat men bij vergrooting van het aantal keurmeesters het systeem van de herkeuring — met onderling overleg — geheel kan laten vervallen.

Het is zelfs de vraag, of bij het keuren door een combinatie van twee keurmeesters de herkeuring — met onderling overleg — wel een betrouwbaarder eindoordeel geeft. De kans is nl. vrij groot, dat bij onderling overleg de keurmeester, die gewend is zijn oordeel zeer positief uit te spreken en groote overredingskracht bezit, op het eindoordeel een te grooten invloed heeft. Te overwegen ware, om bij het constateeren van te groote verschillen tusschen twee keurmeesters bij de beoordeeling van bepaalde monsters de keurmeesters zonder onderling overleg te laten herkeuren en dan als eindoordeel te nemen het gemiddelde van de vier beoordeelingen (van iederen keurmeester twee beoordeelingen).

Het spreekt vanzelf, dat men zorg dient te dragen, dat de keurmeesters bij de herkeuring volkomen onbekend zijn met het cijfer, dat zij in eerste instantie hebben gegeven. Ook moeten de keurmeesters *geheel* onafhankelijk van elkaar het cijfer vaststellen. Dit geldt trouwens evenzeer voor het cijfer, dat in eerste instantie wordt toegekend. (Het is van belang, dat hier in de practijk nauwlettend op wordt toegezien, daar anders de betrouwbaarheid van het oordeel wordt verkleind.)

Wij gaan thans over tot de bespreking van een tweede middel om de nauwkeurigheid van de keuring te verhoogen.

II. door grootere verschillen in de kwaliteit der monsters te scheppen.

Deze methode is reeds door verschillende bonden in toepassing gebracht nl. door de monsters niet meer te keuren na 7 dagen bewaring, maar na 10 of 14 dagen bewaring. Hierdoor worden de kwaliteitsverschillen tusschen de verschillende monsters duidelijker, waardoor $m_{tot.}$ grooter zal worden. Daar niet verwacht behoeft te worden, dat $\sigma_{\bar{x}}$ zal veranderen, zal dus de r stijgen m.a.w. de objectieve nauwkeurigheid zal toenemen.

Naast een langeren duur van bewaring kan men ook een hoogere bewaringstemperatuur te baat nemen om grootere verschillen te creëeren.

De cijfers, in dit proefschrift vermeld, toonen aan, dat het zeer zeker wenschelijk is, beide middelen ernstig te overwegen.

§ 8. Beoordeeling van de gebruikte puntenschaal.

Wij kunnen ons ten slotte afvragen, of de puntenschaal, die bij de hier besproken boterkeuringen toegepast wordt, voldoet aan de eischen, die in het algemeen aan een puntenschaal moeten worden gesteld. Het spreekt vanzelf, dat wij geen oordeel zullen uitspreken, of er een beoordeeling dient te geschieden van gehalte en bewerking in één cijfer of dat een splitsing moet plaatsvinden. Ook zullen wij ons niet bezighouden met de vraag, of het juist is, dat ter berekening van het totaal aantal punten, het cijfer voor geur vermenigvuldigd wordt met twee, het cijfer voor smaak en voor gehalte-en-bewerking met vier. Dit zijn natuurlijk zeer gewichtige punten om de juistheid van de wijze van uitdrukken van het totale oordeel nader te beoordeelen, maar ze liggen buiten het door ons gestelde doel nl. de boterkeuring te bekijken uit het oogpunt van de mathematische statistiek.

Tegen de gebruikte puntenschaal worden dikwijls de volgende bezwaren aangevoerd:

a. de puntenschaal wordt slechts voor een deel gebruikt. (In hoofdzaak het traject 4 tot en met 8). Men zou beter een puntenschaal kunnen gebruiken, waarbij alle „schaaldeelen" gebruikt worden.

b. het „totaal aantal punten" vertoont een vrij groote middelbare fout. Daarom zou men beter doen, steeds bijv. op vijftallen af te ronden, of de boter te onderscheiden in 3 of 4 groepen (bijv. 1°. zeer goede en goede boter, 2°. matige boter en 3°. slechte boter).

Wij achten geen van beide bezwaren gegrond. Bespreken wij eerst het onder *a.* genoemde bezwaar, nl. dat niet de geheele puntenschaal wordt gebruikt. Het is, uit mathematisch-statistisch oogpunt beschouwd, juist een voordeel, dat niet de geheele puntenschaal van 1—10 gebruikt wordt. Indien wel de geheele schaal werd gebezigd, zou dit geschrift nooit tot stand zijn gekomen en zouden beschouwingen over de nauwkeurigheid der keuring practisch niet of slechts zeer onvoldoende mogelijk zijn. Het spreekt vanzelf, dat ook wij het wenschelijk achten, dat de keurmeester een ruime schaal gebruikt bijv. 2 tot en met 9, desnoods 1 tot en met 10, indien maar wordt gezorgd, dat de cijfers 1 (voor zéér slechte boter) en 10 (voor buitengewoon goede boter) slechts uiterst zelden worden gegeven. Zoo bezien, moet het onjuist worden geacht, dat een zoodanige schaal wordt gebezigd, dat voor een goede en uitermate goede boter hetzelfde, — hoogste — cijfer wordt gegeven. Precies hetzelfde geldt voor de zeer slechte en uitermate slechte boters.

Er is zeer weinig tegen, dat de keurmeester van een schaal de uitersten practisch niet gebruikt.

Men veronderstelt dikwijls, dat het gebruik van een ruimere schaal de nauwkeurigheid verhoogt. Dit is echter zeer onwaarschijnlijk. Aangenomen mag worden, dat bij ruimer worden van de schaal de middelbare fout van de keuring in dezelfde mate wordt vergroot als de schaal ruimer is geworden, m.a.w. door de ruimere schaal wordt in wezen geen vergrooting der nauwkeurigheid verkregen. (Immers $m_{tot.}$ en $\sigma_{\bar{x}}$ worden in dezelfde verhouding vergroot).

Bekijken wij uit bovenstaand oogpunt bijv. de schaal, die in Duitschland voor de beoordeeling van den smaak wordt toegepast (en van 18 Februari 1943 af ook in Nederland), dan blijkt deze niet geheel aan onzen eisch te voldoen. Immers aan goede boter dient een 9 voor smaak te worden gegeven. Naar boven is dus zeer weinig spreiding mogelijk. Naar beneden is voldoende „ruimte".

De mathematisch-statistische wetten, waarvan dit geschrift toch zeer duidelijk de geldigheid bij het keuren van boter heeft aangetoond, zullen hier niet zoo goed tot hun recht kunnen komen.

Ten slotte de vraag, of men niet beter doet het geven van punten achterwege te laten en het eindoordeel te geven in drie of vier aanduidingen bijv. zeer goede en goede boter, matige boter en slechte boter (zoo men wil, voege men hier nog een groep: zeer slechte boter bij).

Inderdaad heeft de analyse van de nauwkeurigheid van een combinatie van twee keurmeesters ons in § 4 van dit hoofdstuk tot de conclusie gebracht, dat een dergelijke combinatie tot een verdergaande splitsing in „kwaliteiten" over het algemeen niet in staat is.

Toch zijn wij van meening, dat desondanks aan het opgeven van een cijfer, hoewel dit blijkens onze berekeningen een vrij groote onnauwkeurigheid bezit, voordeelen zijn verbonden. Het uitdrukken van het oordeel in een benaming (goed, matig, slecht) maakt, dat het verwerken (uitrekenen van gemiddelden enz.) minder gemakkelijk wordt. Tevens zij er met nadruk op gewezen, dat — al behoeft er geen *essentieel* verschil in kwaliteit te zijn, tusschen een monster met 68 en 65 punten — de kans toch vrij groot is, dat inderdaad het monster van 68 punten beter is dan dat van 65 punten. Daarbij komt, dat aan een systeem van uitdrukking, waarbij men slechts enkele klassen onderscheidt, het groote bezwaar verbonden is, dat men grensgevallen krijgt, waardoor een monster, dat op de grens van 2 klassen ligt, gevaar loopt veel te hoog of te laag te worden gewaardeerd.

In de practijk is tegen het 100-puntenstelsel, zooals dat bij de bondsboterkeuringen gebruikelijk is, wel eens het bezwaar aangevoerd, dat de belanghebbenden (directeuren en personeel van de zuivelfabrieken) over het algemeen geneigd zijn reeds aan een verschil van 2 punten in de beoordeeling beteekenis toe te kennen. Het feit, dat hun bij sommige gelegenheden (bijv. bij het insturen van duplo monsters) bleek, dat aanmerkelijk grootere verschillen konden worden gevonden, heeft dan wel eens aanleiding gegeven tot de meening, dat aan de boterkeuring in het geheel geen waarde kan worden toegekend. Dat deze gevallen werkelijk in de practijk zijn voorgekomen, hoeft echter geen reden te zijn van het toegepaste puntenstelsel af te stappen. Er is in het geheel geen bezwaar tegen aan de fabriek op te geven, dat aan de door haar ingezonden boter 65 of 68 punten zijn toegekend, mits de ontvanger er maar mede op de hoogte is, dat het „werkelijke" puntenaantal er eenige punten van kan afwijken.

HOOFDSTUK III.

DE VERDEELINGSWET DER FOUTEN.

Wij zijn bij onze beschouwingen in het vorige hoofdstuk ervan uitgegaan, dat de fouten van de cijfers, welke door den keurmeester worden toegekend, een normale verdeeling vertoonen. Onze kanstabellen op de blz. 31 tot 35 zijn bijv. geheel op deze veronderstelling gebaseerd.

Tot nadere rechtvaardiging van veel, wat in het vorige hoofdstuk is medegedeeld, is dus een onderzoek naar de verdeeling en zoo mogelijk een bewijs, dat deze verdeeling inderdaad normaal of practisch normaal is, dringend gewenscht.

De vraag, die wij ons stellen, is dus de volgende.

Vertoont de grootheid $t^{(1)}$ resp. $t^{(2)}$, $t^{(3)}$, enz., welke gelijk is aan:

$$t^{(1)} = x^{(1)} - \hat{x}, \quad \text{enz.,}$$

dus de ware fout, een normale verdeeling?

Rechtstreeks kunnen wij dit vraagstuk nooit oplossen, omdat de ware waarden ($\hat{x}$) niet bekend zijn.

Wij hebben nu langs indirecten weg op twee manieren getracht een antwoord op deze vraag te geven.

a. Indien wij de verdeeling nagaan van de verschillen tusschen het aantal toegekende punten van twee keurmeesters (V) en wij kunnen bewijzen, dat deze V's normaal zijn verdeeld, dan is het zeer waarschijnlijk, dat de fouten der beide keurmeesters ook normaal verdeeld zijn.

Immers wij mogen verwachten, dat de fouten van de beide keurmeesters een zelfde verdeeling vertoonen. Nemen wij dit aan, dan moeten de verdeelingen van de fouten ook normaal zijn, als het verschil tusschen beide een normale verdeeling vertoont.

(Zie voor de tweede manier blz. 88).

§ 1. De verdeelingswet der fouten voor het cijfer voor smaak, geur en gehalte-en-bewerking.

Voor het nagaan van de verdeeling van de V's hebben wij de verschillen opgenomen, die de 4 keurmeesters van den G.O.Z. in 1941 voor het cijfer voor smaak vertoonden bij de keuring van de 15 — vroeger reeds meermalen genoemde — monsters per week. (Hetzelfde materiaal dus, dat o.a. ook in tabel 11 werd gebruikt).

Er werden de volgende verschillen gevonden.

Tabel 39. Frequenties van de verschillen tusschen twee keurmeesters (voor *smaak*).
Materiaal: A2 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

	Aantal keeren, dat er een bepaald verschil werd gevonden.					
Combinatie van keurmeesters	1 — 2	1 — 3	1 — 4	2 — 3	2 — 4	3 — 4
Verschillen						
— 4						
— 3	1				1	
— 2	21	10	9	7	9	5
— 1	119	95	69	80	63	65
0	196	209	162	183	150	188
+ 1	61	70	83	118	127	90
+ 2	7	6	15	17	27	14
+ 3			2		3	2
+ 4						1
Totaal aantal monsters	405	390	340	405	380	365

De contrôle, of deze verdeelingen van de bovengenoemde verschillen voor de verschillende combinaties normaal zijn, hebben wij op drie manieren verricht.

A. *Door vergelijking van de gevonden verdeeling met een berekende normale verdeeling, welke een zelfde gemiddelde en een zelfde middelbare fout vertoont als de gevonden verdeeling.*

De redeneering is de volgende. Indien de grootheid $u = V - \overline{V}$ een normale verdeeling vertoont, dan wordt de kans op een bepaalde u (grenzen u_1 en u_2) weergegeven door de formule:

$$W = \frac{h}{\sqrt{\pi}} \int_{u_1}^{u_2} e^{-h^2u^2}\, du,$$

waarbij $h = \dfrac{1}{\sigma_u \sqrt{2}}$.

Het is nu mogelijk de berekende kans op een bepaalde u (en V) te vergelijken met de waargenomen kans en op deze wijze een indruk te verkrijgen, of de verdeeling normaal mag worden geacht.

Als voorbeeld geven wij de berekening voor de verschillen van de combinatie 1 — 2.

Uit de gegevens, voor deze combinatie genoemd, kan men berekenen, dat

$$\sigma_u^2 = \frac{\Sigma V_k^2}{N} - \overline{V}^2 - \tfrac{1}{12} = 0{,}743 - (0{,}22)^2 - 0{,}083 = 0{,}612$$

of
$$h = \frac{1}{\sqrt{1{,}224}} = 0{,}9039.$$

De verdere berekening is nu in tabel 40 gegeven.

In kolom (1) zijn de voorgekomen verschillen (V) genoteerd; in kolom (2) de daarmede overeenkomende u's ($u = V - \overline{V}$); in kolom (3) de bovenste klassegrens ($u + \frac{1}{2}$) van de in kolom (2) genoemde u; in kolom (4) de waarde van $\zeta = h(u + \frac{1}{2})$, dat is kolom (3), gedeeld door $\sqrt{1{,}224}$. In kolom (5) is ten slotte de theoretische kans berekend, die aanwezig is op het voorkomen van afwijkingen van de grootte van $-\infty$ tot de in kolom (4) genoemde ζ.

Deze kans wordt weergegeven door

$$\Theta(\zeta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\zeta} e^{-z^2}\, dz.$$

De waarde van $\Theta(\zeta)$ (kolom 5) is voor de verschillende waarden van ζ ontleend aan een tabel. Het verschil tusschen twee opeenvolgende waarden van (5) levert ons kolom (6): de theoretische kans op het voorkomen van in de kolom (1) genoemde verschillen. In kolom (7) is deze theoretische kans vermenigvuldigd met het totale aantal (N, hier $N = 405$). Deze kolom geeft ons dus de theoretisch te verwachten frequentie (A_k).

In kolom (8) is de gevonden frequentie (Y_k) vermeld. Kolom (9) geeft de verschillen tusschen kolom (8) en kolom (7). ($\Delta_k = Y_k - A_k$).

De beteekenis van de kolommen (10) en (11) zullen wij op blz. 73 behandelen.

Tabel 40. Vergelijking van de waargenomen met de berekende frequenties van de verschillen voor *smaak* voor de combinatie 1 — 2. Materiaal: eerste kolom tabel 39.

V	u (klasse-middel-punt)	$u+\frac{1}{2}$ (klasse-grens)	$\zeta = h(u+\frac{1}{2})$	$\Theta(\zeta)$	$\Theta(\zeta_n) - \Theta(\zeta_{n-1})$	A_k	Y_k	Δ_k	Δ_k	$\frac{\Delta_k^2}{A_k}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
—3 (en lager)	—2,78	—2,28	—2,061	0,0018	0,0018	0,7	1	+0,3		
—2	—1,78	—1,28	—1,157	0,0509	0,0491	19,9	21	+1,1	+1,1	0,061
—1	—0,78	—0,28	—0,253	0,3602	0,3093	125,3	119	—6,3	—6,3	0,317
0	+0,22	+0,72	+0,651	0,8214	0,4612	186,8	196	+9,2	+9,2	0,453
+1	+1,22	+1,72	+1,555	0,9861	0,1644	66,7	61	—5,7	—5,7	0,487
+2	+2,22	+2,72	+2,46	0,9998	0,0137	5,5	7	+1,5	+1,7[1]	0,459
+3 (en hooger)	+3,22	+3,72		1,0000	0,0002	0,1	0	—0,1		
					1,0000	405,0	405	0	0	1,777

dus: $\chi^2 = 1{,}777$.

1) combinatie van $\Delta_k = +0{,}3$ (eerste rij), $+1{,}5$ en $-0{,}1$.

Voor de andere combinaties voerden wij dezelfde berekening uit met het volgende resultaat.

Tabel 40*b*. Vergelijking van de waargenomen met de berekende frequenties voor de overige combinaties.
Materiaal: A2 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Cmb van keurm :	1 — 3		1 — 4		2 — 3		2 — 4		3 — 4	
Verschillen	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.
—3							0,4	1		
—2	8,7	10	9,2	9	7,6	7	8,8	9	6,8	5
—1	99,6	95	70,7	69	76,5	80	64,2	63	68,7	65
0	202,1	209	153,1	162	189,1	183	153,0	150	170,4	188
+1	74,7	70	91,3	83	114,5	118	120,0	127	103,5	90
+2	4,8	6	15,0	15	16,7	17	30,9	27	15,1	14
+3	—		0,7	2	0,6	—	2,6	3	0,5	2
+4							0,1	0		1
Totaal	389,9	390	340,0	340	405,0	405	380,0	380	365,0	365
Kans, dat de verdeeling normaal is	52 %		49 %		73 %		63 %		12 %	

Als wij de gevonden en de berekende waarden voor de verschillende combinaties met elkaar vergelijken, dan zien wij, dat de aansluiting van deze beide waarden over het algemeen zeer goed is.

Er komen echter verschillen voor (vooral bij de combinatie 3 — 4). Het is de vraag, of deze zoodanig zijn, dat van een principieele afwijking van de normale verdeeling moet worden gesproken. Om dit na te gaan, past men veelal de „χ^2-Test for Goodness of Fit" van Pearson toe (zie bijv. R. A. Fisher, Statistical Methods for Research Workers, 3th Edition, London 1930, Chapter III).

In ons geval is zij echter minder geschikt. Dit zal duidelijk worden, wanneer wij bijv. de gegevens voor combinatie 1 — 2 (vermeld in tabel 40) met de χ^2-Test op normaliteit gaan toetsen.

Zooals reeds medegedeeld, zijn in kolom (9) van tabel 40 de verschillen (Δ_k) genoteerd tusschen Y_k (empirische frequentie) en A_k (theoretische frequentie).

Het symbool χ^2 stelt nu de volgende som voor:

$$\chi^2 = \sum_{k=1}^{n} \frac{\Delta_k^2}{A_k} \qquad (n \text{ is het aantal klassen}).$$

Terloops zij opgemerkt, dat wij in kolom (10) voor de kleine frequenties Δ_k en ook A_k van kolom (9) te zamen hebben gevoegd, omdat deze anders de betrouwbaarheid van de toepassing van de χ^2-Test in gevaar zouden brengen.

Voor χ^2 vinden wij (zie tabel 40) voor de combinatie 1 — 2 een waarde van 1,777.

De kans, dat χ^2 deze waarde heeft bij een normale verdeeling, geeft R. A. Fisher op blz. 96 en 97 (table III) in „Statistical Methods for Research-workers" (third edition).

Om deze kans te bepalen, dienen wij echter ook het aantal vrijheidsgraden van χ^2 te weten (n').

χ^2 is in ons geval berekend door 5 grootheden van den vorm: $\frac{\Delta_k^2}{A_k}$ op te tellen. Dit geeft 5 vrijheidsgraden. Dit aantal wordt echter verkleind, doordat A_k berekend is uit het beschikbare materiaal door uit te gaan van drie veronderstellingen:

a. dat het totale aantal waarnemingen bij de theoretisch berekende frequentieverdeeling gelijk is aan dat bij de empirisch gevonden verdeeling, dus:

$$\Sigma Y_k = N = \Sigma A_k.$$

b. dat het gemiddelde van de verschillen ($\overline{V}$) dezelfde waarde voor beide frequentiecurven heeft, dus:

$$\overline{V} = \frac{\Sigma\, Y_k V_k}{N} = \frac{\Sigma\, A_k V_k}{N}.$$

c. dat de middelbare afwijkingen van de beide verdeelingen gelijk zijn, dus:

$$\sigma_u^2 = \frac{\Sigma\, Y_k u_k^2}{N} = \frac{\Sigma\, A_k u_k^2}{N}.$$

Hierdoor verliest χ^2 drie van haar vrijheidsgraden. Er blijven dus slechts $5 - 3 = 2$ over.

De zooeven genoemde tabel van Fisher geeft voor $\chi^2 = 1,777$ en $n' = 2$ een kans van iets meer dan 40 %. De conclusie zou dus moeten zijn, dat de frequentieverdeeling van de verschillen voor combinatie 1 — 2 niet principieel afwijkt van de normale, maar dat toch de overeenstemming niet zoo goed is.

Wij moeten echter voorzichtig zijn met de toepassing van de χ^2-Test, omdat bij zulke kleine waarden voor n' als wij hier hebben de te vinden waarde voor χ^2 vrij sterk afhangt van het feit, hoeveel en welke van de kleine frequenties men combineert.

Combineert men vele kleine frequenties, dan wordt n' in dit geval zeer klein. Hierdoor wordt het bepalen van de kans uit de genoemde tabel vrij onzeker. Combineert men niet, dan heeft men het voordeel, dat n' grooter blijft. Daartegenover staat echter, dat bij de kleine frequenties de waarde $\frac{\Delta_k^2}{A_k}$ gemakkelijk vrij groot kan worden. De waarde van χ^2 wordt dan dus vrij onzeker.

Het cijfer van 40 % voor de kans, dat de frequentieverdeeling voor de combinatie 1 — 2 als normaal kan worden beschouwd, moeten wij dus wel met de noodige voorzichtigheid hanteeren.

Voor de frequentieverdeelingen van de andere combinaties hebben wij eveneens de kans berekend op het normaal zijn. Deze kansen zijn aangegeven in tabel 40*b*.

Ook uit deze cijfers zullen wij geen absoluut vaststaande conclusies mogen trekken.

Wij mogen er echter wel uit afleiden, dat geen der verdeelingen sterk van de normale afwijkt.

Alleen voor de combinatie 3 — 4 geeft de χ^2-Test een abnormaal lage waarde voor de kans, dat de verdeeling normaal genoemd mag worden.

B. *Wij hebben de frequentieverdeelingen op nog een andere wijze onderzocht om na te gaan, of zij al dan niet normaal zijn.*

Wij hebben dit op de volgende wijze gedaan.

Als voorbeeld nemen wij de gegevens voor de combinatie 1 — 2 (zie tabel 39).

Tabel 41. Vergelijking van de gevonden frequentieverdeeling voor de combinatie 1 — 2 met de normale.

(1)	(2)	(3)	(4)
Waarde van V	Frequentie van $-\infty$ tot en met V	Waarde uit kolom (2), gedeeld door N	ζ
-3	1	0,0025	$-1,98$
-2	22	0,0543	$-1,135$
-1	141	0,3481	$-0,276$
0	337	0,8321	$+0,680$
$+1$	398	0,9827	$+1,495$
$+2$	405	1,0000	$+\infty$

Wij berekenden in kolom (3) de empirische kans op het voorkomen van verschillen van $-\infty$ tot en met het in kolom (1) genoemde verschil (V). Deze kans wordt gelijk gesteld aan $\Theta(\zeta)$.

Door middel van de vergelijking:

$$\Theta(\zeta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\zeta} e^{-z^2}\, dz$$

kunnen wij voor iedere in kolom (1) genoteerde waarde van V een hierbij behoorende waarde van ζ vinden (uit een tabel).

Als nu de verdeeling normaal is, moet de lijn, die het verband aangeeft tusschen V en ζ, recht zijn. Inderdaad blijkt dit het geval te zijn (zie fig. 5*a*).

De bewuste verdeeling mag dus als normaal worden beschouwd.

Ook voor de andere combinaties hebben wij de puntenreeks geteekend. De punten liggen hier eveneens op een rechte lijn. Alleen voor de combinatie 3 — 4 vertoont de lijn een vrij duidelijke kromming, hetgeen er op wijst, dat deze frequentiecurve (= beeld van de frequentieverdeeling) iets van de normale afwijkt (zie fig. 5*b*).

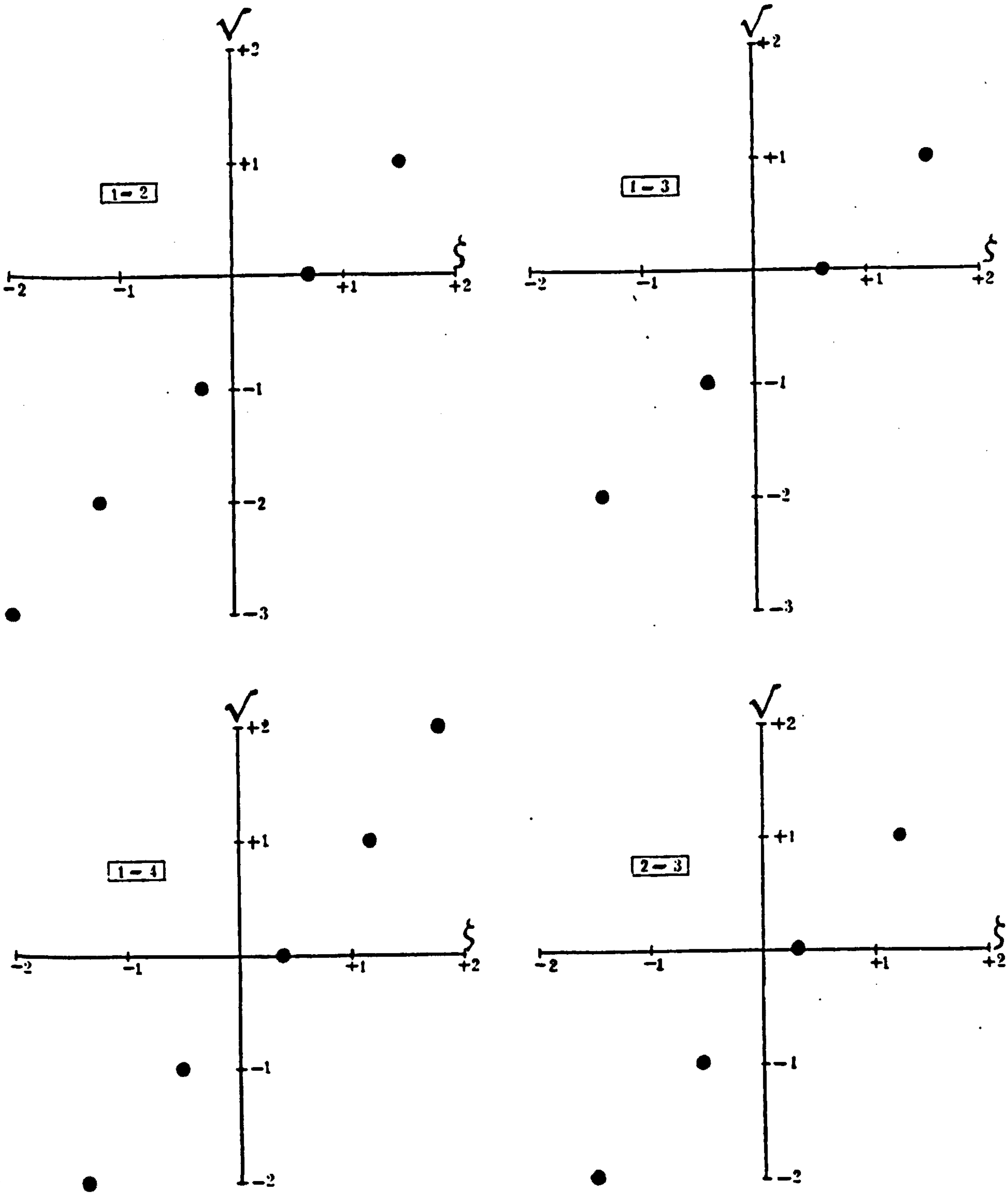

Fig. 5*a*. Verband tusschen *V* en ζ bij smaak voor de combinaties 1—2, 1—3, 1—4, en 2—3.

Wij mogen niet onvermeld laten, dat ook de zoo juist besproken contrôle op het al of niet normaal zijn der curven in dit geval vrij ruw is. Voor het bepalen van het verband tusschen *V* en ζ kan aan de punten aan de uiteinden van de lijn slechts weinig waarde worden toegekend. Hieruit volgt, dat wij over het algemeen slechts drie punten overhouden om de lijn op zijn lineariteit te onderzoeken.

Uit het onder A. en B. medegedeelde mogen wij besluiten, *dat alle verdeelingen practisch normaal zijn. Alleen de curve voor de combinatie* 3 — 4 *wijkt iets af.*

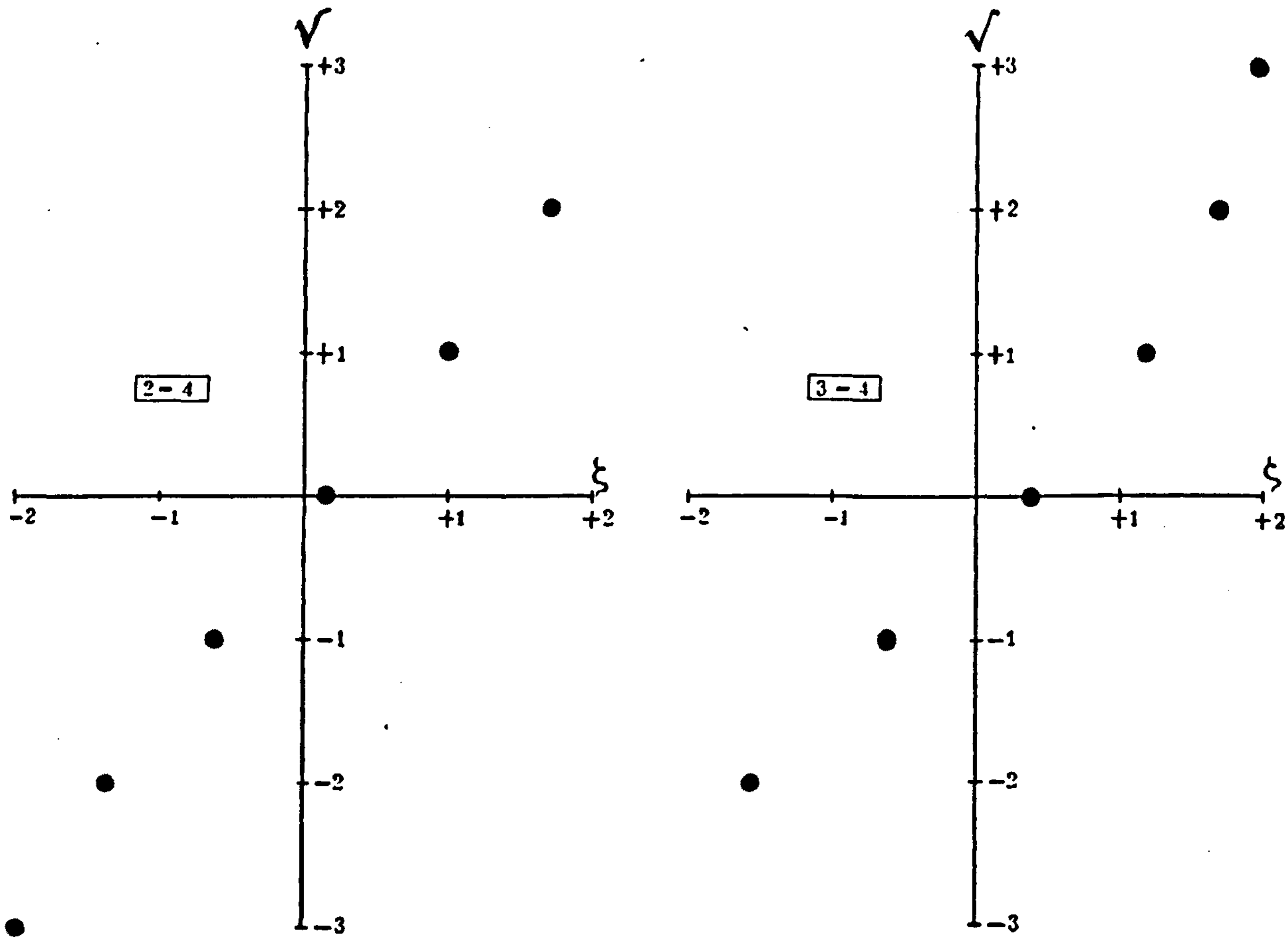

Fig. 5*b*. Verband tusschen V en ζ bij smaak voor de combinaties 2—4 en 3—4.

Voor de contrôle, of een verdeeling normaal is, wordt ook dikwijls gebruik gemaakt van de *scheefheid* en het *exces*. Deze grootheden kunnen tevens een aanwijzing geven, waarin de afwijkingen van de normale verdeeling bestaan.

In het volgende hebben wij deze voor de 6 frequentieverdeelingen berekend.

C. *Berekening van scheefheid en exces.*

De scheefheid wordt aangegeven door de grootheid:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}},$$

het exces door:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

De γ_1 meet de scheefheid van de curve. Bij positieve γ_1 stijgt de curve — gerekend van de negatieve waarden van V — sneller dan normaal en daalt voorbij de top langzamer. Bij negatieve γ_1 doet zich het omgekeerde voor.

De γ_2 geeft meestal een aanwijzing, of de top van de curve hooger (bij positieve γ_2) of lager (bij negatieve γ_2) ligt dan bij de normale verdeeling.

In bovenstaande uitdrukkingen voor γ_1 en γ_2 beteekent μ_2 het voor „groepeering" en „afwijkend middelpunt gecorrigeerde tweedegraadsmoment" (dus het kwadraat van de middelbare fout), μ_3 het gecorrigeerde derdegraadsmoment en μ_4 het gecorrigeerde vierdegraadsmoment.

Om deze gecorrigeerde momenten uit te rekenen, wordt eerst berekend:

$$\nu'_1 = \frac{\sum\limits_{k=1}^{n} Y_k V_k}{N}\,; \quad \nu'_2 = \frac{\sum\limits_{k=1}^{n} Y_k V_k^2}{N}\,; \quad \nu'_3 = \frac{\sum\limits_{k=1}^{n} Y_k V_k^3}{N}\,; \quad \nu'_4 = \frac{\sum\limits_{k=1}^{n} Y_k V_k^4}{N}$$

waarbij Y_k de frequentie van het geconstateerde verschil (V_k) is en N het totaal aantal monsters en n het aantal klassen.

Deze „ruwe" momenten (ν') worden gecorrigeerd:

a. voor de fout, die gemaakt wordt, omdat het verschil gemiddeld niet 0 is ($\overline{V} = \nu'_1 \neq 0$) door middel van de volgende formules:

$$\nu_2 = \nu'_2 - \nu'^2_1$$
$$\nu_3 = \nu'_3 - 3\nu'_1\nu'_2 + 2\nu'^3_1$$
$$\nu_4 = \nu'_4 - 4\nu'_1\nu'_3 + 6\nu'^2_1\nu'_2 - 3\nu'^4_1$$

De grootheden ν_2, ν_3 en ν_4 geven dus de voor „afwijkend middelpunt" gecorrigeerde momenten, derhalve de momenten t.o.v. het gemiddelde.

b. voor de fout, die gemaakt wordt, doordat de klasse-intervallen, vergeleken met de geheele schaal, zoo groot zijn:

$$\mu_2 = \nu_2 - \frac{c^2}{12}$$
$$\mu_3 = \nu_3$$
$$\mu_4 = \nu_4 - \frac{c^2}{2}\mu_2 - \frac{c^2}{80}.$$

(c = klasse-interval)

Deze — onder *b.* genoemde — correcties heeten de *correcties van Sheppard*.

De momenten (μ), gecorrigeerd voor „groepeering" en „afwijkend middelpunt", kunnen nu gebruikt worden om γ_1 en γ_2 uit te rekenen.

De hierachter volgende berekening van γ_1 en γ_2 voor de verschillen van de combinatie 1 — 2 zal den lezer na bovenstaande toelichting wel duidelijk zijn.

Tabel 42. Berekening van γ_1 en γ_2 voor de verschillen voor de combinatie 1 — 2.

Materiaal: A2 (Zutphen).

	Y_k	V_k	Y_kV_k	$Y_kV_k^2$	$Y_kV_k^3$	$Y_kV_k^4$
	1	— 3	— 3	9	— 27	81
	21	— 2	— 42	84	— 168	336
	119	— 1	— 119	119	— 119	119
	196	0	0	0	0	0
	61	+ 1	+ 61	61	+ 61	61
	7	+ 2	+ 14	28	+ 56	112
			— 164 + 75 + 75		— 314 + 117 + 117	
Totaal	405		— 89	301	— 197	709

Hieruit volgt:

$$\nu_1' = -\frac{89}{405} = -0{,}22$$

$$\nu_2' = \frac{301}{405} = 0{,}743$$
$$\nu_1'^2 = (-0{,}22)^2 = 0{,}048$$
$$\nu_2 = 0{,}695$$
$$\frac{c^2}{12} = 0{,}083$$
$$\mu_2 = 0{,}612$$

$$\nu_3' = -\frac{197}{405} = -0{,}486$$
$$-3\nu_1'\nu_2' = +0{,}490$$
$$+2\nu_1'^3 = -0{,}021$$
$$\nu_3' = \mu_3 = -0{,}017$$

$$\nu_4' = \frac{709}{405} = 1{,}750$$
$$-4\nu_1'\nu_3' = -0{,}428$$
$$+6\nu_1'^2\nu_2' = +0{,}216$$
$$-3\nu_1'^4 = -0{,}007$$
$$\nu_4 = 1{,}531$$
$$-\frac{c^2}{2}\mu_2 = -0{,}306$$
$$-\frac{c^2}{80} = -0{,}013$$
$$\mu_4 = 1{,}212$$

dus: $$\gamma_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{-0{,}017}{0{,}479} = -0{,}035.$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = +0{,}236.$$

Wij hebben dezelfde berekening gemaakt voor de andere combinaties en verkregen de volgende uitkomsten.

Tabel 43. Waarden van γ_1 en γ_2 voor alle combinaties.
Materiaal: A2 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Combinatie	γ_1		γ_2	
1 — 2	— 0,035	(± 0,122)	+ 0,236	(± 0,244)
1 — 3	— 0,005	(± 0,124)	+ 0,185	(± 0,248)
1 — 4	+ 0,215	(± 0,133)	+ 0,380	(± 0,266)
2 — 3	— 0,028	(± 0,122)	— 0,333	(± 0,244)
2 — 4	— 0,044	(± 0,126)	+ 0,200	(± 0,251)
3 — 4	+ 0,569	(± 0,128)	+ 1,730	(± 0,256)

In bovenstaande tabel zijn ook aangegeven de middelbare fouten van γ_1 en γ_2. Deze zijn gelijk aan $\sqrt{\frac{6}{N}}$ resp. $\sqrt{\frac{24}{N}}$.

Deze middelbare fouten geven een indruk, hoever γ_1 en γ_2 zich van 0 mogen verwijderen, voordat geconcludeerd behoeft te worden, dat de verdeeling principieel van de normale afwijkt.

De in tabel 43 gegeven waarden voor γ_1 en γ_2 zijn geheel in overeenstemming met hetgeen wij onder A en B vonden nl. dat de verdeelingen practisch normaal zijn. Alleen de combinatie 3 — 4 loopt er, wat γ_1 en γ_2 betreft, iets uit. Ook dit komt overeen met wat het onderzoek onder A en B ons had geleerd.

Conclusie uit A, B en C. *De fouten van de keurmeesters voor het cijfer voor smaak mogen practisch als normaal verdeeld worden beschouwd.*

Voor *geur* en *gehalte-en-bewerking* hebben wij de verdeeling niet zoo serieus nagegaan. De verdeeling van de verschillen, die wij voor de gemeenschappelijke keuring te Zutphen vonden, maakt het echter waarschijnlijk, dat de fouten voor geur en voor gehalte-en-bewerking zich niet anders gedragen dan die voor het cijfer voor smaak. Wij zullen deze dus ook als normaal verdeeld mogen beschouwen (zie tabel 44).

Tabel 44. Frequentie van de verschillen tusschen twee keurmeesters voor *geur* en *gehalte-en-bewerking*.

Materiaal: B1 (Zutphen).

4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

	1 — 2		1 — 3		1 — 4	
Verschillen	geur	geh. en bew.	geur	geh. en bew.	geur	geh. en bew.
0	47	58	37	62	50	58
— 1 en + 1	50	44	59	40	49	43
— 2 en + 2	6	1	7	1	3	2
— 3 en + 3					1	
Totaal	103	103	103	103	103	103

§ 2. De verdeelingswet der fouten voor „totaal aantal punten". Thans dienen wij de verdeeling van de fouten bij het bepalen van het totaal aantal punten te bekijken. Aangezien wij voor smaak de normale verdeeling hebben bewezen en wij aangenomen hebben, dat voor geur en gehalte-en-bewerking hetzelfde geldt, is te verwachten, dat ook de verdeeling van de fouten voor „totaal aantal punten" normaal is. Immers de som van drie normaal verdeelde grootheden vertoont eveneens een normale verdeeling.

Toch is het wenschelijk, dat wij de verdeeling van de fouten voor „totaal aantal punten" nog nader beschouwen. Zijn deze normaal verdeeld, dan hebben wij meteen onze veronderstelling, dat de fouten voor geur en voor gehalte-en-bewerking normaal verdeeld zijn, veel aannemelijker gemaakt.

Voor het nagaan van de verdeeling van de fouten van den keurmeester bij de vaststelling van het totaal aantal punten hebben wij de verschillen in totaal aantal punten opgenomen, die wij voor de 4 keurmeesters gedurende 1941 vonden bij de keuring van de 15 reeds eerder aangeduide monsters per week (dus hetzelfde als zoo pas voor smaak).

Er werden de volgende verschillen gevonden.

Tabel 45. Frequentie van de bij verschillende combinaties van twee keurmeesters voorgekomen verschillen (voor „*totaal aantal punten*").

Materiaal: A2 (Zutphen).

4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Verschillen	1 — 2	1 — 3	1 — 4	2 — 3	2 — 4	3 — 4	Totaal
— 16	2	2	1	2	2	2	11
— 14	2	2		2	2		8
— 12	8	1	4	2	11	4	30
— 10	16	9	8	14	13	14	74
— 8	8	7	12	19	13	11	70
— 6	39	41	37	31	36	40	224
— 4	55	52	30	39	27	24	227
— 2	51	36	51	35	44	53	270
0	82	97	76	90	62	81	488
+ 2	42	50	33	31	33	33	222
+ 4	37	37	44	57	54	43	272
+ 6	36	34	22	49	34	28	203
+ 8	11	6	11	10	18	9	65
+ 10	11	8	16	15	18	16	84
+ 12	2	6	3	8	3	1	23
+ 14	1	2	4	1	5	2	15
+ 16	2		1		2	1	6
+ 18			2		1	1	4
+ 20					1		1
+ 22					1	1	2
+ 24						1	1
Totaal	405	390	355	405	380	365	2300

Wij berekenden, op dezelfde wijze als voor smaak aangegeven, de verdeeling, zoals die moest zijn, indien de verdeeling geheel normaal was geweest.

Tabel 46. Vergelijking van de waargenomen en berekende frequenties van de in tabel 45 genoemde verschillen.

Verschillen	1 — 2		1 — 3		1 — 4		2 — 3		2 — 4		3 — 4	
	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.
— 16	1,3	2	0,4	2	0,9	1	0,9	2	2,5	2	1,2	2
— 14	2,4	2	1,1	2	1,6	0	1,8	2	3,3	2	2,3	0
— 12	5,9	8	3,4	1	4,1	4	4,5	2	6,6	11	5,1	4
— 10	12,4	16	8,4	9	8,8	8	9,7	14	11,9	13	10,5	14
— 8	22,8	8	17,5	7	16,6	12	18,3	19	19.5	13	18,7	11
— 6	35,9	39	31,3	41	27,0	37	30,1	31	28,6	36	29,2	40
— 4	49,2	55	46,9	52	38,6	30	43,4	39	37,7	27	40,2	24
— 2	58,6	51	59,4	36	47,9	51	54,2	35	45,1	44	48,9	53
0	60,3	82	63,4	97	51,9	76	59,4	90	48,3	62	51,8	81
+ 2	54,5	42	57,8	50	49,5	33	56,8	31	47,2	33	48,9	33
+ 4	42,2	37	44,4	37	41,1	44	47,4	47	41,3	54	40,5	43
+ 6	28,5	36	28,7	34	29,6	22	34,6	49	32,8	34	29,5	28
+ 8	16,6	11	15,8	6	18,7	11	22,0	10	23,3	18	18,9	9
+ 10	8,5	11	7,3	8	10,3	16	12,2	15	15,0	18	10,6	16
+ 12	3,7	2	2,9	6	5,0	3	6,0	8	8,7	3	5,3	1
+ 14	1,5	1	0,9	2	2,1	4	2,5	1	4,6	5	2,3	2
+ 16	0,5	2	0,3	0	0,8	1	1,2	0	2,1	2	0,9	1
+ 18	0,1	0			0,3	2			0,9	1	0,2	1
+ 20									0,3	1	—	0
+ 22									0,1	1	—	1
+ 24											—	1
Totaal	404,9	405	389,9	390	354,8	355	405,0	405	379,8	380	365,0	365

Wanneer wij voor iedere combinatie de berekende waarde ($\varDelta_k$) vergelijken met de gevonden waarde (Y_k) en weer $\chi^2 = \Sigma \frac{\Delta_k^2}{\varDelta_k}$ gaan berekenen, blijkt bij alle combinaties χ^2 zoo groot te worden, dat er practisch absoluut geen kans is, dat deze verdeelingen als zuiver normaal verdeeld mogen worden beschouwd. Dit ligt aan de zeer groote afwijking tusschen de theoretische en de gevonden waarden.

Een eenvoudige contrôle toont reeds, dat deze afwijkingen zeer groot zijn. Bijv. theoretisch is te verwachten (bij veronderstelde normale verdeeling), dat voor de combinatie 1 — 2 de afwijking 0 60,3 maal zal voorkomen.

De middelbare fout van deze theoretische frequentie ($\varDelta_k$) is:

$$\sqrt{\frac{\varDelta_k(N - \varDelta_k)}{N}} = \sqrt{\frac{60{,}3 \times (405 - 60{,}3)}{405}} \cong 7.$$

Hieruit volgt, dat een gevonden waarde van 82 onmogelijk als een van de theoretische waarde (60,3) toevallig afwijkende waarde kan worden opgevat. Voor vele andere verschillen geldt precies hetzelfde. Ook dit toont reeds aan, dat de verdeeling niet als normaal mag worden beschouwd.

Dit lijkt voor de waarde van onze vroegere beschouwingen een ernstige bedreiging. Immers indien geen normale of geen bijna normale verdeeling aanwezig is, — ten minste als zich ernstige afwijkingen voordoen — dan komt de waarde van onze beschouwingen van blz. 33 tot blz. 37 over de nauwkeurigheid der keuring en de op die bladzijden voorkomende staatjes over bepaalde kansen op bepaalde afwijkingen ten zeerste in gevaar.

Toch zal blijken, dat dit in werkelijkheid niet zoo is.

Indien wij uit de gevonden middelbare fout (via $h = \frac{1}{\sqrt{2\mu_2}}$) de kans willen berekenen op het voorkomen van een afwijking van een bepaalde grootte, dan kunnen wij — zooals het bovenstaande leert — een cijfer krijgen, dat geheel strijdig is met wat het experiment geeft. Op deze wijze wordt het vraagstuk echter in de practijk bijna nooit gesteld. In verreweg de meeste gevallen wordt men voor de vraag gesteld: welke kans heb ik, dat de afwijking niet grooter is dan een bepaald bedrag?

Gaan wij de theoretische en de gevonden waarden aldus naast elkaar stellen, dan wordt de toestand geheel anders.

Tabel 47. Vergelijking van de waargenomen en berekende frequenties van absolute verschillen, ter grootte van of kleiner dan een bepaald getal. Materiaal: als tabel 46.

Absolute verschillen van	1 — 2		1 — 3		1 — 4		2 — 3		2 — 4		3 — 4	
	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.	Ber.	Waargen.
0	60,3	82	63,4	97	51,9	76	59,4	90	48,3	62	51,8	81
2 en minder	173,4	175	180,6	183	149,3	160	170,4	156	140,6	139	149,6	167
4 en minder	264,8	267	271,9	272	229,0	234	261,2	252	219,6	220	230,3	234
6 en minder	329,8	342	331,9	347	285,6	293	325,9	332	281,0	290	289,0	302
8 en minder	368,6	361	365,2	360	320,9	316	366,2	361	323,8	321	326,6	322
10 en minder	389,5	388	380,9	377	340,0	340	388,1	390	350,7	352	347,7	352
12 en minder	399,1	398	387,2	384	349,1	347	398,6	400	366,0	366	358,1	357
14 en minder	403,0	401	389,2	388	352,8	351	402,9	403	373,9	373	362,7	359
16 en minder	404,8	405	389,9	390	354,5	353	405,0	405	378,5	377	364,8	362
18 en minder					354,8	355			379,4	378	365,0	363
20 en minder									379,7	379	365,0	363
22 en minder									379,8	380	365,0	364
24 en minder											365,0	365

Hieruit blijkt, dat alleen de frequenties voor de verschillen van 0 belangrijke afwijkingen vertoonen, maar dat reeds voor de afwijkingen van 2 en minder de overeenkomst zeer goed is, terwijl bij het grooter worden der afwijkingen de theoretische berekende waarde hoe langer hoe beter gaat overeenstemmen met de gevonden waarde.

Wij mogen dus besluiten, dat ook voor „totaal aantal punten" onze in het vorige hoofdstuk met de middelbare fout uitgevoerde manipulaties niet tot onjuiste uitkomsten leiden.

Dat de afwijking van de normale verdeeling niet zeer ernstig is, volgt ook uit de scheefheid en het exces van de genoemde verdeelingen.

Tabel 48. Waarde van γ_1 en γ_2 voor de verschillende combinaties van twee keurmeesters (voor *„totaal aantal punten"*).
Materiaal: A2 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Wij vonden:

Combinatie	γ_1		γ_2	
1 — 2	+ 0,003	(± 0,122)	+ 0,232	(± 0,244)
1 — 3	+ 0,026	(± 0,124)	+ 0,507	(± 0,248)
1 — 4	+ 0,335	(± 0,127)	+ 0,422	(± 0,254)
2 — 3	— 0,176	(± 0,122)	— 0,163	(± 0,244)
2 — 4	+ 0,085	(± 0,126)	+ 0,215	(± 0,251)
3 — 4	+ 0,401	(± 0,128)	+ 1,33	(± 0,256)

De afwijkingen van γ_1 en γ_2 van 0 zijn dus inderdaad over het algemeen weinig beteekenend; alleen de combinatie 3 — 4 vertoont een onmiskenbare scheefheid en exces.

Het berekende exces klopt in het algemeen wel ongeveer met de geaardheid van de curve in het topgebied. Indien wij voor het topgebied de waarden — 2, 0 en + 2 nemen, dan leert tabel 46 op blz. 82, dat de volgende verschillen bestaan.

Tabel 49. Vergelijking tusschen de waarde van γ_2 en het verschil van waargenomen en theoretische frequentie gebied in het „topgebied".
Materiaal: als tabel 48.

Combinatie	Verschil in topgebied tusschen waargenomen en theoretische frequentie	γ_2
1 — 2	+ 1,6	+ 0,232
1 — 3	+ 2,4	+ 0,507
1 — 4	+ 10,7	+ 0,422
2 — 3	— 14,4	— 0,163
2 — 4	— 1,6	+ 0,215
3 — 4	+ 17,4	+ 1,33

Ook als wij de normaliteit van de verdeeling nagaan op de wijze, zooals in § 1 onder B voor smaak beschreven is, dan blijkt, dat de afwijkingen weinig beteekenen. De figuren 6*a*, 6*b* en 6*c*, die het verband tusschen V en ζ aangeven voor de verschillende combinaties van keurmeesters, laten dat duidelijk zien.

Alleen voor de combinatie 3 — 4 is van een vrij duidelijke afwijking van de rechte lijn sprake.

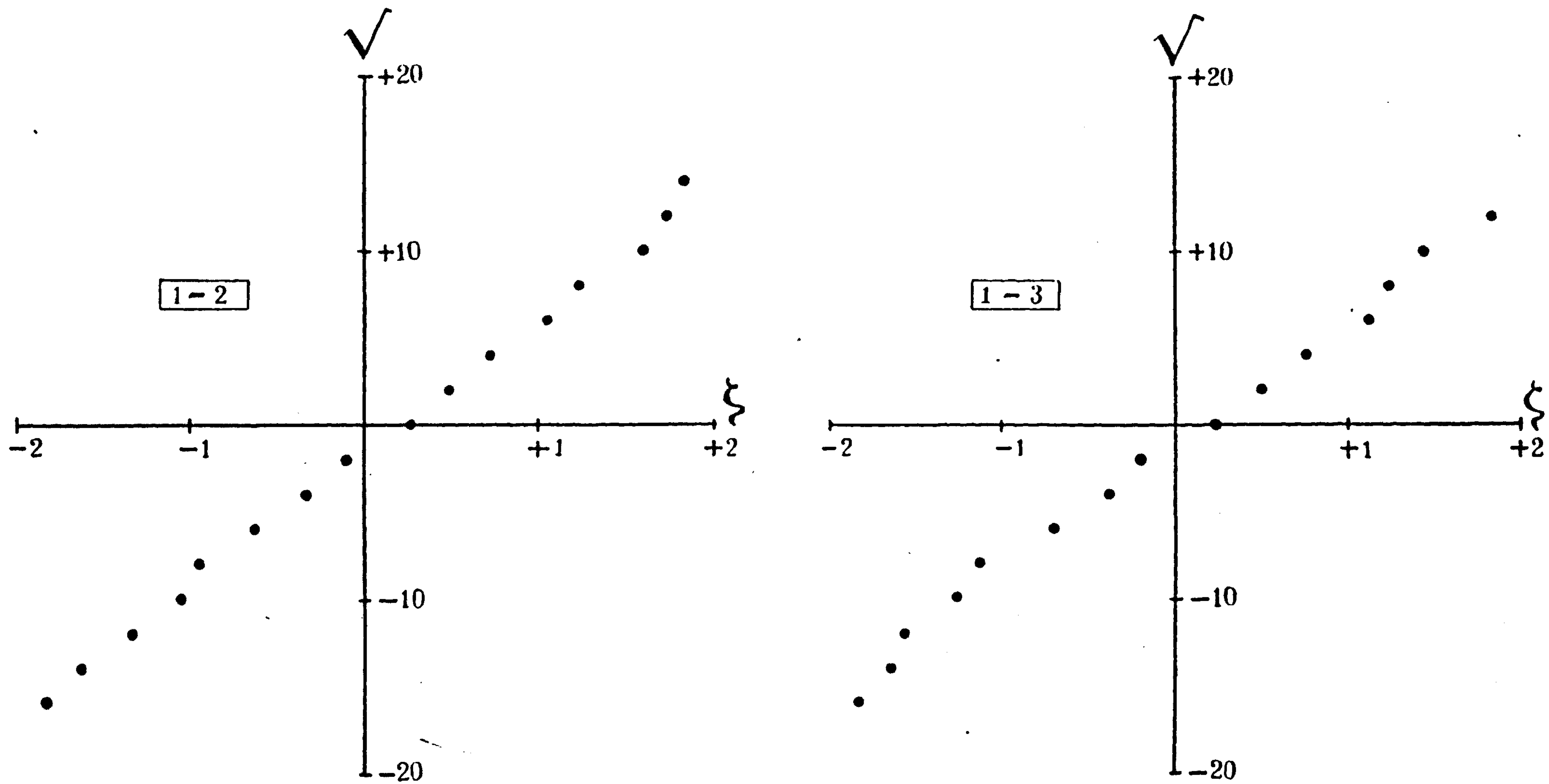

Fig. 6*a*. Verband tusschen V en ζ bij „totaal aantal punten" voor de combinaties 1 — 2 en 1 — 3.

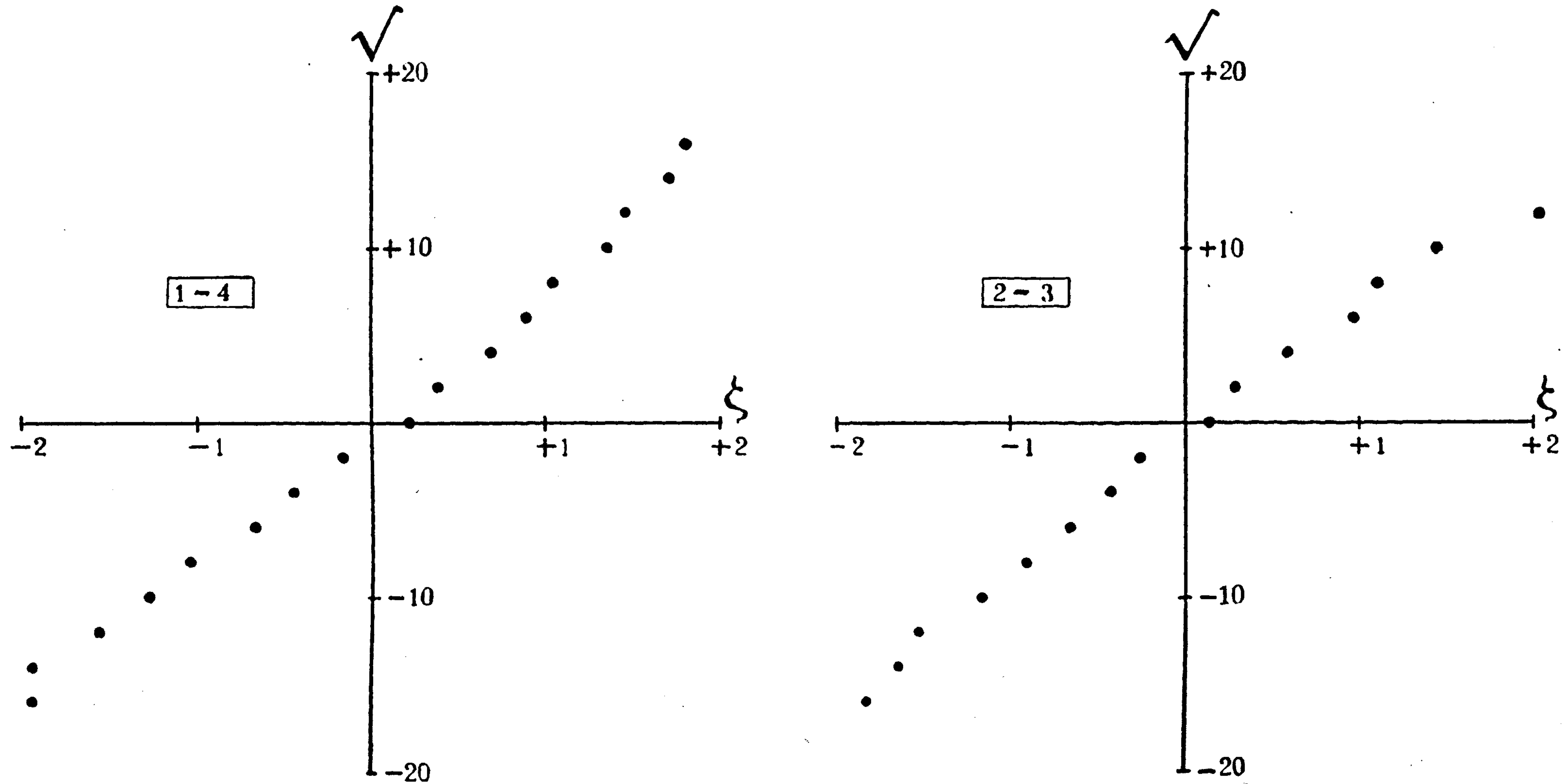

Fig. 6*b*. Verband tusschen V en ζ bij „totaal aantal punten" voor de combinaties 1 — 4 en 2 — 3.

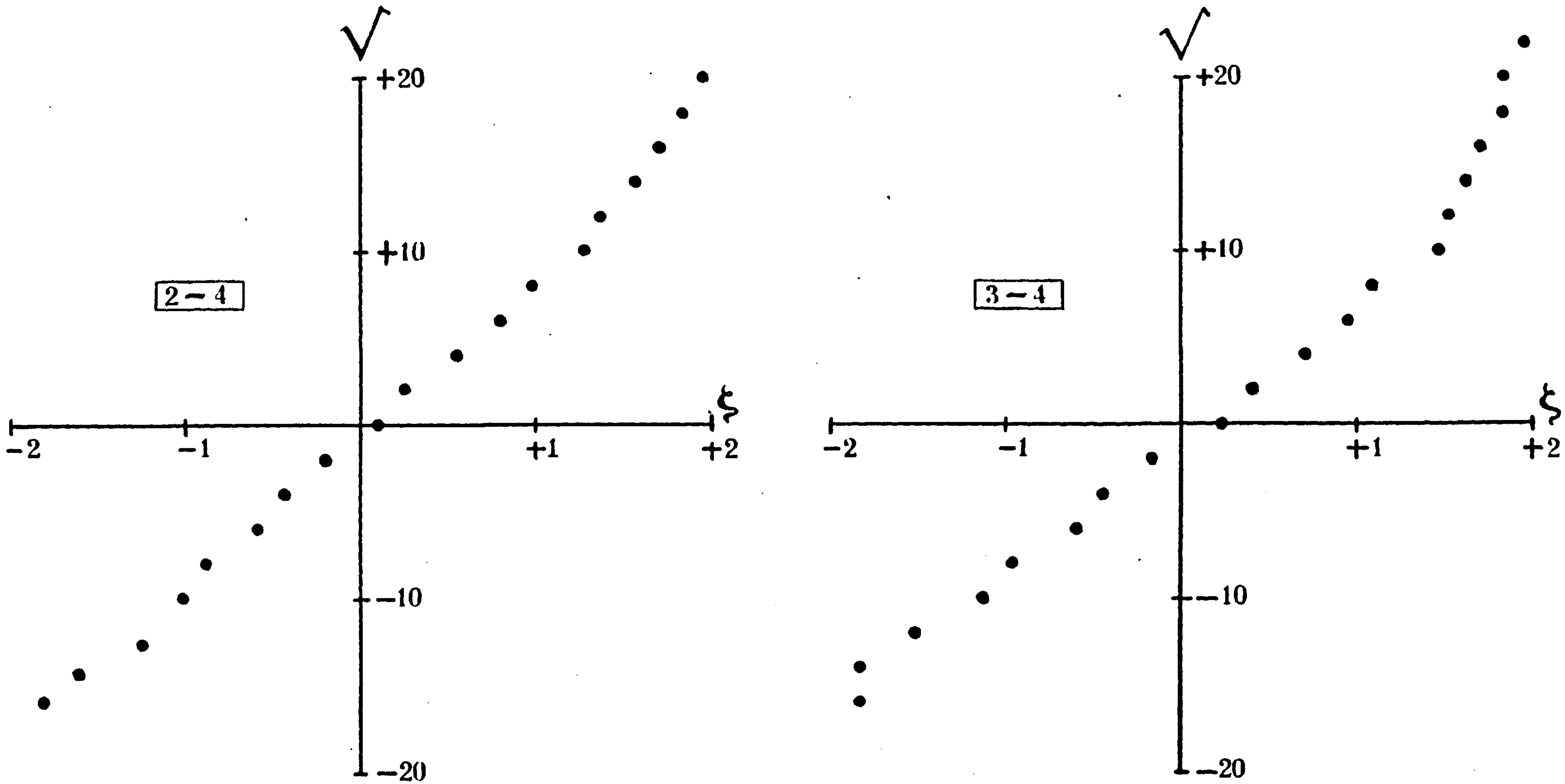

Fig. 6c. Verband tusschen V en ζ bij „totaal aantal punten" voor de combinaties 2 — 4 en 3 — 4.

In het algemeen is dus de verdeeling normaal. Alleen voor de frequenties van de afzonderlijke klassen komen verschillen voor.

In het begin van dit hoofdstuk spraken wij er over, dat eigenlijk de verdeeling moest worden nagegaan van de ware fout d.i. dus de afwijking van het gegeven cijfer met het ware cijfer. Daar de ware waarde niet bekend is, moesten wij onze toevlucht nemen tot het onderzoek van de verdeeling van de verschillen tusschen de beoordeelingen van de keurmeesters om daaruit een conclusie te trekken ten opzichte van de verdeeling der ware fouten. Toch bestaat er een benaderingsmethode, waardoor wij — zij het eenigszins ruw — een indruk kunnen krijgen omtrent de verdeeling der ware fouten. Indien wij nl. de cijfers, welke door vele keurmeesters aan hetzelfde monster zijn toegekend, middelen, dan verkrijgen wij een cijfer, dat het ware cijfer gaat benaderen. Nu staan ons slechts gegevens ten dienste, waarbij een groot aantal monsters door *drie* keurmeesters zijn gekeurd. Dit aantal keurmeesters is eigenlijk te gering om aan hun gemiddeld oordeel de kwalificatie „waar" toe te kennen, maar toch is dit in dit geval de beste benadering van het „ware" cijfer.

Noemen wij weer het cijfer, gegeven door keurmeester 1 aan het k^{de} monster $x_k^{(1)}$, het cijfer van keurmeester 2 $x_k^{(2)}$, van keurmeester 3 $x_k^{(3)}$, en hun gemiddelde $\bar{x}_k$ en stellen wij:

$$u_k^{(1)} = x_k^{(1)} - \bar{x}_k$$
$$u_k^{(2)} = x_k^{(2)} - \bar{x}_k$$
$$u_k^{(3)} = x_k^{(3)} - \bar{x}_k$$

dan geldt, indien wij aannemen, dat $\bar{x}_k \simeq \hat{x}_k$ is ($\hat{x}_k$ is het „ware" cijfer voor monster k), dat

$$u_k^{(1)} \simeq t_k^{(1)}$$
$$u_k^{(2)} \simeq t_k^{(2)}$$
$$u_k^{(3)} \simeq t_k^{(3)},$$

waarbij $t_k^{(1)}$, $t_k^{(2)}$ en $t_k^{(3)}$ de ware fouten van de drie keurmeesters zijn.

In tabel 50 (blz. 89) zijn de geconstateerde t's aangegeven met daarnaast de voorgekomen frequenties. Hierop komen 4 keurmeesters voor, omdat — zooals reeds eerder medegedeeld werd — in totaal 4 keurmeesters bij den Geldersch-Overijselschen Zuivelbond regelmatig keuren.

Wij beschouwen alleen de totaaltelling van alle keurmeesters, omdat voor één keurmeester het aantal wat klein is om de verdeeling na te gaan. Wij meenen hiertoe gerechtigd te zijn, omdat, wanneer bewezen is, dat de totaaltelling een normale verdeeling vertoont, ook de cijfers, die betrekking hebben op de individueele keurmeesters als normaal verdeeld mogen worden beschouwd. Immers er is in het geheel geen aanleiding om aan te nemen, dat de eene keurmeester zich in dit opzicht anders zou gedragen dan de andere.

Wij controleeren de overeenstemming tusschen waargenomen en te verwachten frequentie thans op dezelfde wijze, als onder § 1 sub B beschreven voor smaak en op blz. 85 tot 88 voor „totaal aantal punten".

Tabel 50. Frequentie van de verschillen, die geconstateerd zijn tusschen het oordeel van één keurmeester en het gemiddelde van drie keurmeesters over een kwartaal (voor „totaal aantal punten"). Materiaal: A1 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Keurmeesters	1	2	3	4	Totaal
Verschillen (t)					
$-\frac{42}{3}$		1			1
$-\frac{30}{3}$					
$-\frac{28}{3}$		2			2
$-\frac{26}{3}$		3	1		4
$-\frac{24}{3}$					
$-\frac{22}{3}$		1	1	2	4
$-\frac{20}{3}$	2	4	1		7
$-\frac{18}{3}$	2			3	5
$-\frac{16}{3}$	1	5	1		7
$-\frac{14}{3}$	9	4	3	4	20
$-\frac{12}{3}$	3	6	7	4	20
$-\frac{10}{3}$	3	8	2	9	22
$-\frac{8}{3}$	9	3	8	9	29
$-\frac{6}{3}$	9	11	11	11	42
$-\frac{4}{3}$	9	8	10	6	33
$-\frac{2}{3}$	13	10	13	17	53
0	14	16	17	19	66
$+\frac{2}{3}$	10	10	15	12	47
$+\frac{4}{3}$	12	15	17	18	62
$+\frac{6}{3}$	9	12	7	6	34
$+\frac{8}{3}$	14	3	6	9	32
$+\frac{10}{3}$	6	10	13	8	37
$+\frac{12}{3}$	4	5	4	6	19
$+\frac{14}{3}$	3	2	6	2	13
$+\frac{16}{3}$		5	1	2	8
$+\frac{18}{3}$	1	2	2	1	6
$+\frac{20}{3}$	2		1	1	4
$+\frac{22}{3}$		1	3		4
$+\frac{24}{3}$		2			2
$+\frac{26}{3}$					
$+\frac{28}{3}$		1			1
$+\frac{30}{3}$				1	1
Totaal	135	150	150	150	585

Tabel 51. Berekening van het verband tusschen t en ζ.
Materiaal: Totaal-kolom van tabel 50.

Waarde van t	Waargenomen frequentie van $-\infty$ tot t	Kolom (2), gedeeld door totaal aantal	ζ
$-\frac{28}{3}$	3	0,005	$-1{,}82$
$-\frac{26}{3}$	7	0,012	$-1{,}60$
$-\frac{24}{3}$	7	0,012	$-1{,}60$
$-\frac{22}{3}$	11	0,019	$-1{,}47$
$-\frac{20}{3}$	18	0,031	$-1{,}32$
$-\frac{18}{3}$	23	0,039	$-1{,}245$
$-\frac{16}{3}$	30	0,051	$-1{,}155$
$-\frac{14}{3}$	50	0,086	$-0{,}965$
$-\frac{12}{3}$	70	0,120	$-0{,}83$
$-\frac{10}{3}$	92	0,157	$-0{,}71$
$-\frac{8}{3}$	121	0,207	$-0{,}58$
$-\frac{6}{3}$	163	0,279	$-0{,}415$
$-\frac{4}{3}$	196	0,335	$-0{,}30$
$-\frac{2}{3}$	249	0,426	$-0{,}13$
0	315	0,538	$+0{,}07$
$+\frac{2}{3}$	362	0,619	$+0{,}214$
$+\frac{4}{3}$	424	0,725	$+0{,}422$
$+\frac{6}{3}$	458	0,783	$+0{,}55$
$+\frac{8}{3}$	490	0,838	$+0{,}70$
$+\frac{10}{3}$	527	0,901	$+0{,}91$
$+\frac{12}{3}$	546	0,933	$+1{,}06$
$+\frac{14}{3}$	559	0,956	$+1{,}21$
$+\frac{16}{3}$	567	0,969	$+1{,}32$
$+\frac{18}{3}$	573	0,979	$+1{,}44$
$+\frac{20}{3}$	577	0,986	$+1{,}55$
$+\frac{22}{3}$	581	0,993	$+1{,}75$
$+\frac{24}{3}$	583	0,997	$+1{,}95$
$+\frac{26}{3}$	583	0,997	$+1{,}95$
$+\frac{28}{3}$	584	0,998	$+2{,}05$
$+\frac{30}{3}$	585	1,000	$+\infty$

Figuur 7 toont aan, dat de punten vrij wel op een rechte lijn liggen. Het verloop van de lijn laat zien, dat de top van de frequentiecurve iets hooger ligt dan bij een zuiver normale verdeeling te verwachten is.

De afwijking is echter gering.

Slotconclusie van dit hoofdstuk.

De ware fouten van de keurmeesters, zoowel voor de onderdeelen smaak, geur, gehalte-en-bewerking als voor „totaal aantal punten" mogen voor practische berekeningen als normaal verdeeld worden beschouwd.

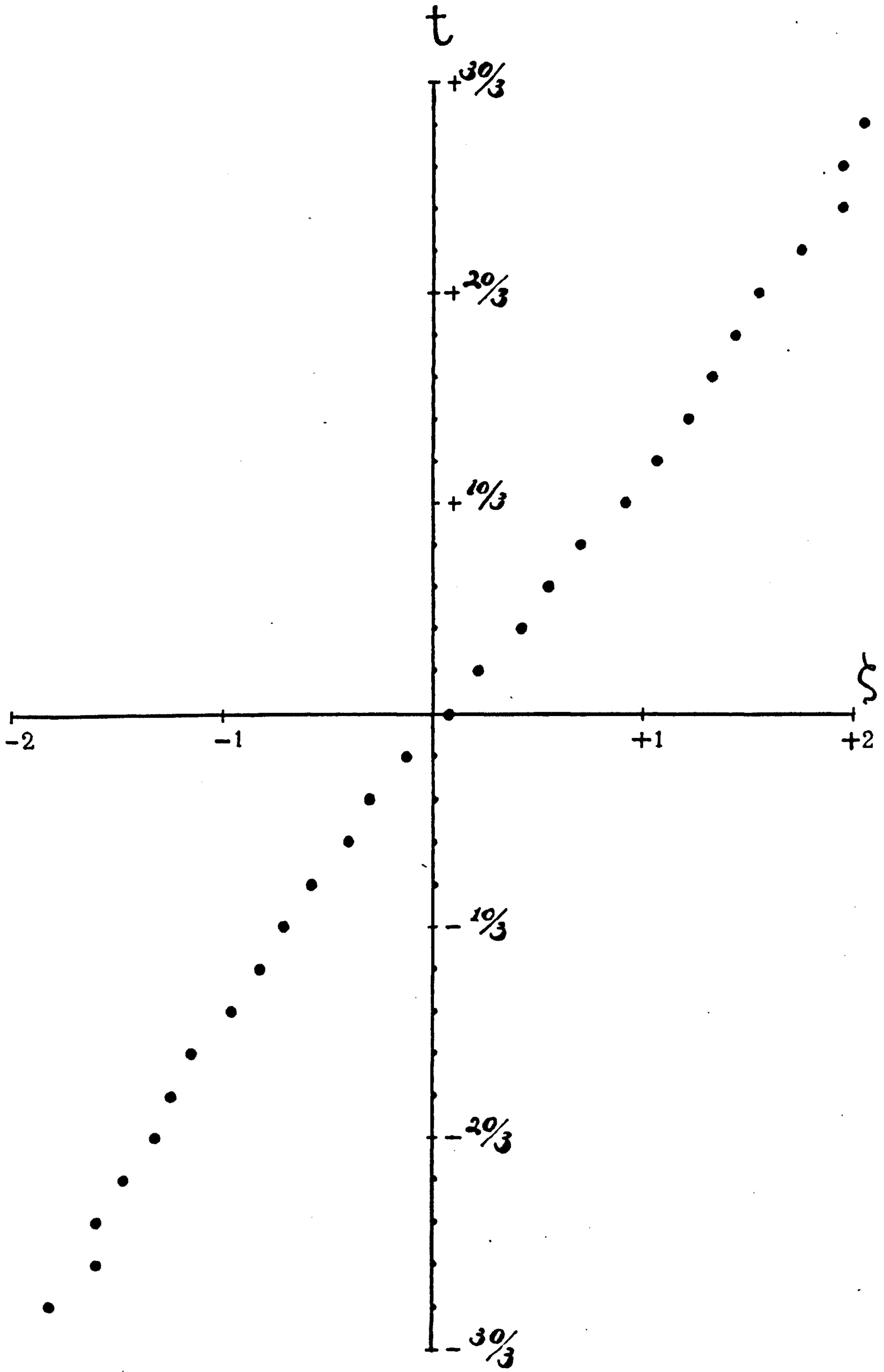

Fig. 7. Verband tusschen t en ζ bij „totaal aantal punten".

HOOFDSTUK IV.

DE NAUWKEURIGHEID VAN DEN INDIVIDUEELEN KEURMEESTER.

In hoofdstuk II hebben wij ruime aandacht besteed aan de nauwkeurigheid, waarmede een combinatie van twee keurmeesters haar oordeel (uitgedrukt in een cijfer) over de kwaliteit van de boter en over onderdeelen daarvan vaststelt. Wij beschouwden toen steeds een combinatie van twee keurmeesters, omdat het in de practijk gewoonte is de boter door ten minste twee keurmeesters te laten beoordeelen. De cijfers, die wij verkregen, gaven dus een beeld van de betrouwbaarheid van het uiteindelijke cijfer, waarin het gemiddelde oordeel der *beide* keurmeesters wordt uitgedrukt.

Er is echter alle reden om ook een onderzoek in te stellen naar de nauwkeurigheid van den individueelen keurmeester.

Wij zullen daarbij gelegenheid krijgen op verschillende vraagpunten dieper in te gaan, die ook van belang zijn voor het beoordeelen van de nauwkeurigheid van een combinatie van twee of meer keurmeesters. Onze bespreking over de nauwkeurigheid van één keurmeester zal dus het besprokene in hoofdstuk II nog nader kunnen verduidelijken.

Wij onderscheiden de nauwkeurigheid van den individueelen keurmeester — evenals wij dit deden bij onze beschouwingen over de nauwkeurigheid van een combinatie van keurmeesters — in twee soorten nauwkeurigheid:

a. *de subjectieve nauwkeurigheid*, welke haar uitdrukking vindt in de middelbare fout, waarmede de keurmeester zijn cijfer vaststelt;

b. *de objectieve nauwkeurigheid*, die wordt uitgedrukt in den correlatiecoëfficiënt tusschen het door den keurmeester toegekende cijfer en het ware cijfer.

Wij zullen hierbij ook gelegenheid krijgen op de beteekenis van beide iets dieper in te gaan.

§ 1. De subjectieve nauwkeurigheid van een keurmeester. (I)

Om hiervan een indruk te krijgen dienen wij dus de middelbare fout te berekenen, waarmede één keurmeester zijn cijfer vaststelt.

Om deze te vinden, volgden wij de onderstaande methode.

Noem het cijfer, dat door keurmeester 1 aan monster k (k loopend van 1 tot en met N) is toegekend $x_k^{(1)}$, het cijfer, dat door keurmeester 2 aan hetzelfde monster is gegeven $x_k^{(2)}$. Het verschil tusschen beide cijfers wordt aangeduid met V_k, dus:

$$V_k = x_k^{(1)} - x_k^{(2)} \quad \ldots\ldots\ldots\ldots \quad (1)$$

Indien wij nu het verschil tusschen de beoordeeling ($x_k^{(1)}$ en $x_k^{(2)}$) en de ware waarden ($\hat{x}_k^{(1)}$ en $\hat{x}_k^{(2)}$) t noemen (t is dus de ware fout), dan krijgen wij:

$$\left.\begin{array}{l} x_k^{(1)} - \hat{x}_k^{(1)} = t_k^{(1)} \\ x_k^{(2)} - \hat{x}_k^{(2)} = t_k^{(2)} \end{array}\right\} \quad \ldots\ldots\ldots\ldots \quad (2)$$

Wij zullen hier twee ware waarden (van elken keurmeester één) moeten onderscheiden. Immers bij het beschouwen van de nauwkeurigheid van één keurmeester zullen wij het feit, dat de ééne keurmeester gemiddeld een iets lager of iets hooger cijfer geeft dan de andere, niet als een fout mogen aanmerken.

Uit (1) en (2) volgt:

$$V_k = x_k^{(1)} - x_k^{(2)} = (x_k^{(1)} - \hat{x}_k^{(1)}) - (x_k^{(2)} - \hat{x}_k^{(2)}) + (\hat{x}_k^{(1)} - \hat{x}_k^{(2)}) =$$
$$= t_k^{(1)} - t_k^{(2)} + (\hat{x}_k^{(1)} - \hat{x}_k^{(2)}).$$

Daar wij aannemen, dat $\hat{x}_k^{(1)} - \hat{x}_k^{(2)}$ voor iedere waarde van k dezelfde waarde heeft, mogen wij stellen:

$$\hat{x}_k^{(1)} - \hat{x}_k^{(2)} = \hat{x}^{(1)} - \hat{x}^{(2)} = \delta.$$

In hoeverre deze veronderstelling juist is, zullen wij in § 2 van dit hoofdstuk nader bekijken.

Het blijkt dus, dat

$$V_k = t_k^{(1)} - t_k^{(2)} + \delta \quad \ldots\ldots\ldots\ldots \quad (3)$$

Kwadrateering van (3) geeft

$$V_k^2 = t_k^{(1)^2} - 2t_k^{(1)}t_k^{(2)} + t_k^{(2)^2} + 2\delta.t_k^{(1)} - 2\delta.t_k^{(2)} + \delta^2$$

of gesommeerd over een oneindig aantal waarnemingen:

$$<V_k^2> = <t_k^{(1)^2}> - 2<t_k^{(1)}\, t_k^{(2)}> + <t_k^{(2)^2}> + 2\delta.<t_k^{(1)}> -$$
$$- 2\delta <t_k^{(2)}> + <\delta^2> \quad \ldots\ldots\ldots\ldots \quad (4)$$

Hierbij geeft het teeken $< >$ aan, dat — ten minste in gedachte — gesommeerd wordt over een oneindig aantal waarnemingen. De vorm $<t_k^{(2)}>$ is dus synoniem met de uitdrukking: $\sum\limits_{\substack{k=1 \\ \lim N=\infty}}^{N} \frac{t_k^{(2)}}{N}$.

Daar $t_k^{(1)}$ en $t_k^{(2)}$ als onafhankelijk beschouwd mogen worden, is

$$<t_k^{(1)}\, t_k^{(2)}> = 0.$$

Verder is

$$<t_k^{(1)}> = 0 \text{ en } <t_k^{(2)}> = 0 \text{ en } <\delta^2> = \delta^2$$

(4) gaat dus over in

$$<V_k^2> = <t_k^{(1)^2}> + <t_k^{(2)^2}> + \delta^2 \quad \ldots\ldots \quad (5)$$

Nu is $<V_k^2>$ d.i. $\frac{1}{N} \sum\limits_{\substack{k=1 \\ \lim N=\infty}}^{N} V_k^2$ ons niet bekend, doordat ons slechts N waarnemingen ten dienste staan.

Wij kunnen $< V_k^2 >$ dus slechts benaderen. Indien N echter een vrij groot getal is, mogen wij — zonder een merkbare fout te begaan — stellen

$$< V_k^2 > = \frac{1}{N} \sum_{k=1}^{N} V_k^2 = m_V^2.$$

Noemen wij verder:

$$< t_k^{(1)2} > = \sigma_1^2 \quad \text{en} \quad < t_k^{(2)2} > = \sigma_2^2$$

(waarbij σ_1 en σ_2 de middelbare fouten zijn, waarmede keurmeester 1 en keurmeester 2 keuren), dan volgt uit (5):

$$m_V^2 = \sigma_1^2 + \sigma_2^2 + \delta^2 \quad \ldots\ldots\ldots\ldots \quad (6)$$

Behalve σ_1 en σ_2 is in deze vergelijking ook δ onbekend, daar de ware waarden ($\hat{x}_k^{(1)}$ en $\hat{x}_k^{(2)}$) niet bekend zijn. Indien wij echter een groot aantal monsters ter beschikking hebben, mogen wij aannemen, dat

$$\delta \simeq \overline{x^{(1)}} - \overline{x^{(2)}}$$

waaruit dus δ is te berekenen.

Voor de combinatie (1, 2) kunnen we dus vinden: $\sigma_1^2 + \sigma_2^2$ nl.:

$$\sigma_1^2 + \sigma_2^2 = m_{V_{1,2}}^2 - \delta_{1,2}^2 \quad \ldots\ldots\ldots\ldots \quad (7^1)$$

Evenzoo geldt voor:

combinatie (1, 3): $\sigma_1^2 + \sigma_3^2 = m_{V_{1,3}}^2 - \delta_{1,3}^2 \quad \ldots\ldots\ldots\ldots \quad (7^2)$

combinatie (1, 4): $\sigma_1^2 + \sigma_4^2 = m_{V_{1,4}}^2 - \delta_{1,4}^2 \quad \ldots\ldots\ldots\ldots \quad (7^3)$

combinatie (2, 3): $\sigma_2^2 + \sigma_3^2 = m_{V_{2,3}}^2 - \delta_{2,3}^2 \quad \ldots\ldots\ldots\ldots \quad (7^4)$

combinatie (2, 4): $\sigma_2^2 + \sigma_4^2 = m_{V_{2,4}}^2 - \delta_{2,4}^2 \quad \ldots\ldots\ldots\ldots \quad (7^5)$

combinatie (3, 4): $\sigma_3^2 + \sigma_4^2 = m_{V_{3,4}}^2 - \delta_{3,4}^2 \quad \ldots\ldots\ldots\ldots \quad (7^6)$

Uit deze 6 vergelijkingen met 4 onbekenden (σ_1, σ_2, σ_3 en σ_4) kunnen de onbekenden worden opgelost.

Berekening van σ voor het cijfer voor smaak.

Voor smaak werden over 1941 de volgende verschillen gevonden (bij de 15 monsters, die iedere week bij de keuring te Zutphen door drie keurmeesters worden beoordeeld).

Tabel 52. Frequentie van de verschillen, geconstateerd bij verschillende combinaties van twee keurmeesters voor *smaak*.

Materiaal: A2 (Zutphen).

4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

Combinatie van keurmeesters:	1 — 2	1 — 3	1 — 4	2 — 3	2 — 4	3 — 4	Totaal
Aantal monsters	405	390	340	405	380	365	2285
Verschillen van:							
+ 4						1	1
+ 3			2		3	2	7
+ 2	7	6	15	17	27	14	86
+ 1	61	70	83	118	127	90	549
0	196	209	162	183	150	188	1088
— 1	119	95	69	80	63	65	491
— 2	21	10	9	7	9	5	61
— 3	1				1		2
Gem. verschil (δ)	—0,22	—0,08	+0,09	+0,14	+0,28	+0,14	
$\frac{\Sigma V_k^2}{N}$	0,743	0,587	0,782	0,726	0,974	0,726	
Correctie van Sheppard	0,083	0,083	0,083	0,083	0,083	0,083	
m_V^2	0,660	0,504	0,699	0,643	0,891	0,643	
δ^2	0,048	0,006	0,008	0,020	0,078	0,029	
$\sigma^2 + \sigma^2$	0,612	0,498	0,691	0,623	0,813	0,622	

Indien wij nu: $\sigma_1^2 = p_1$, $\sigma_2^2 = p_2$, $\sigma_3^2 = p_3$ en $\sigma_4^2 = p_4$ stellen, worden de volgende z.g. *waarnemingsvergelijkingen* verkregen:

$$
\begin{aligned}
p_1 + p_2 &= 0{,}612 \\
p_1 + p_3 &= 0{,}498. \\
p_1 + p_4 &= 0{,}691 \\
p_2 + p_3 &= 0{,}623 \\
p_2 + p_4 &= 0{,}813 \\
p_3 + p_4 &= 0{,}622
\end{aligned}
$$

(Terloops zij hier opgemerkt, dat tegen bovenstaande behandeling, streng theoretisch beschouwd, een klein bezwaar bestaat. Immers de bovengenoemde $V_{1,2}$, $V_{1,3}$ enz. zijn niet alle uit de waarnemingen van dezelfde monsters berekend. De combinatie (1, 2) bijv. heeft gedeeltelijk andere monsters gekeurd dan de combinatie (1, 4). Practisch heeft dit feit echter in het geheel geen beteekenis.)

Bekend is, dat bij het beschikbaar hebben van de volgende n waarnemingsvergelijkingen:

$$\begin{array}{l}
a_1p_1 + b_1p_2 + c_1p_3 + d_1p_4 = q_1 \\
a_2p_1 + b_2p_2 + c_2p_3 + d_2p_4 = q_2 \\
\vdots \\
a_kp_1 + b_kp_2 + c_kp_3 + d_kp_4 = q_k \\
\vdots \\
a_np_1 + b_np_2 + c_np_3 + d_np_4 = q_n
\end{array}$$

de meest waarschijnlijke waarden voor p_1, p_2, p_3 en p_4 (te noemen $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ en p_4) kunnen worden opgelost uit de volgende normaalvergelijkingen:

$$\begin{array}{l}
\bar{p}_1 . \Sigma a_k^2 + \bar{p}_2 . \Sigma a_kb_k + \bar{p}_3 . \Sigma a_kc_k + \bar{p}_4 . \Sigma a_kd_k = \Sigma a_kq_k \\
\bar{p}_1 . \Sigma b_ka_k + \bar{p}_2 . \Sigma b_k^2 + \bar{p}_3 . \Sigma b_kc_k + \bar{p}_4 . \Sigma b_kd_k = \Sigma b_kq_k \\
\bar{p}_1 . \Sigma c_ka_k + \bar{p}_2 . \Sigma c_kb_k + \bar{p}_3 . \Sigma c_k^2 + \bar{p}_4 . \Sigma c_kd_k = \Sigma c_kq_k \\
\bar{p}_1 . \Sigma d_ka_k + \bar{p}_2 . \Sigma d_kb_k + \bar{p}_3 . \Sigma d_kc_k + \bar{p}_4 . \Sigma d_k^2 = \Sigma d_kq_k
\end{array}$$

Op dezelfde wijze kunnen voor ons geval de normaalvergelijkingen worden berekend:

	a	ab	ac	ad	b^2	bc	bd	c^2	cd	d^2	aq	bq	cq	dq
$p_1 + p_2 = 0{,}612$	1	1	0	0	1	0	0	0	0	0	0,612	0,612	0	0
$p_1 + p_3 = 0{,}498$	1	0	1	0	0	0	0	1	0	0	0,498	0	0,498	0
$p_1 + p_4 = 0{,}691$	1	0	0	1	0	0	0	0	0	1	0,691	0	0	0,691
$p_2 + p_3 = 0{,}623$	0	0	0	0	1	1	0	1	0	0	0	0,623	0,623	0
$p_2 + p_4 = 0{,}813$	0	0	0	0	1	0	1	0	0	1	0	0,813	0	0,813
$p_3 + p_4 = 0{,}622$	0	0	0	0	0	0	0	1	1	1	0	0	0,622	0,622
Totaal	3	1	1	1	3	1	1	3	1	3	1,801	2,048	1,743	2,126

Hieruit volgen de *normaalvergelijkingen*:

$$\begin{array}{l}
3\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + \bar{p}_4 = 1{,}801 \\
\bar{p}_1 + 3\bar{p}_2 + \bar{p}_3 + \bar{p}_4 = 2{,}048 \\
\bar{p}_1 + \bar{p}_2 + 3\bar{p}_3 + \bar{p}_4 = 1{,}743 \\
\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + 3\bar{p}_4 = 2{,}126
\end{array}$$

Als men deze 4 vergelijkingen optelt, de som door 6 deelt en daarna de uitkomst aftrekt van vergelijking 1, resp. 2, 3 en 4, krijgt men gemakkelijk de waarden voor de $\bar{p}$'s nl.

Waarden van $\bar{p}$ voor de 4 keurmeesters (voor *smaak*)

$$\begin{array}{l}
\bar{p}_1 = 0{,}257 \\
\bar{p}_2 = 0{,}381 \\
\bar{p}_3 = 0{,}228 \\
\bar{p}_4 = 0{,}420
\end{array}$$

waaruit de volgende waarden voor de middelbare fouten van de individueele keurmeesters volgen:

$$\sigma_1 = 0{,}51$$
$$\sigma_2 = 0{,}62$$
$$\sigma_3 = 0{,}48$$
$$\sigma_4 = 0{,}65$$

(Waarden afgerond op 0,01).

Wanneer wij bovenstaande waarden voor $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ en $\bar{p}_4$ invullen in de waarnemingsvergelijkingen, waarvan wij zijn uitgegaan, dan blijken de waarden voor $\bar{p}_1 + \bar{p}_2$, $\bar{p}_1 + \bar{p}_3$, enz. vrij dicht bij de gevonden waarden te liggen. Dit wordt bevestigd, wanneer wij de onzekerheid van $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ en $\bar{p}_4$ gaan berekenen, zoals die voortvloeit uit de strijdigheid van de zes vergelijkingen. Die onzekerheid wordt uitgedrukt door: $\sigma_{\bar{p}_1}$, $\sigma_{\bar{p}_2}$, $\sigma_{\bar{p}_3}$ en $\sigma_{\bar{p}_4}$.

Voor deze grootheden geldt:

$$\sigma^2_{\bar{p}_1} = P.\sigma^2; \quad \sigma^2_{\bar{p}_2} = Q.\sigma^2; \quad \sigma^2_{\bar{p}_3} = R.\sigma^2 \quad \text{en} \quad \sigma^2_{\bar{p}_4} = S.\sigma^2.$$

waarbij:

$$\sigma^2 = \frac{\sum\limits_{k=1}^{n} \{q_k - (a_k\bar{p}_1 + b_k\bar{p}_2 + c_k\bar{p}_3 + d_k\bar{p}_4)\}^2}{n - l}$$

waarin: n is het aantal waarnemingsvergelijkingen (hier 6),
l is het aantal onbekenden (hier 4).

en:
$$P = \frac{N_a}{N}; \quad Q = \frac{N_b}{N}; \quad R = \frac{N_c}{N} \quad \text{en} \quad S = \frac{N_d}{N}.$$

waarbij N gelijk is aan de waarde van den volgenden determinant:

$$N = \begin{vmatrix} \Sigma a_k^2 & \Sigma a_k b_k & \Sigma a_k c_k & \Sigma a_k d_k \\ \Sigma b_k a_k & \Sigma b_k & \Sigma b_k c_k & \Sigma b_k d_k \\ \Sigma c_k a_k & \Sigma c_k b_k & \Sigma c_k^2 & \Sigma c_k d_k \\ \Sigma d_k a_k & \Sigma d_k b_k & \Sigma d_k c_k & \Sigma d_k^2 \end{vmatrix}$$

N_a resp. N_b, N_c en N_d zijn onderdeterminanten van N naar Σa_k^2 resp. Σb_k^2, Σc_k^2 en Σd_k^2.

Voor ons geval krijgen wij het volgende:

$$\sigma^2 \simeq \frac{2000 \times 10^{-6}}{6 - 4} = 1000 \times 10^{-6}$$

$$N = \begin{vmatrix} 3 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 3 \end{vmatrix} = 48; \qquad N_a = N_b = N_c = N_d = \begin{vmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{vmatrix} = 20$$

dus:
$$P = Q = R = S = \tfrac{20}{48}$$

waaruit volgt, dat $\sigma^2_{\bar{p}_1} = \sigma^2_{\bar{p}_2} = \sigma^2_{\bar{p}_3} = \sigma^2_{\bar{p}_4} = \frac{20}{48} \times 1000 \times 10^{-6}$

of:
$$\sigma_{\bar{p}_1} = \sigma_{\bar{p}_2} = \sigma_{\bar{p}_3} = \sigma_{\bar{p}_4} \simeq 0{,}02.$$

De strijdigheid der waarnemingsvergelijkingen is dus betrekkelijk gering.

Een schatting van de nauwkeurigheid van de berekende σ_1, σ_2, σ_3 en σ_4 krijgen we als volgt:

$$\sigma_1^2 = \bar{p}_1$$

of gedifferentieerd:

$$2\sigma_1 \Delta\sigma_1 = \Delta\bar{p}_1$$

of

$$\Delta\sigma_1 = \frac{\Delta\bar{p}_1}{2\sigma_1}$$

dus

$$\Delta\sigma_1^2 = \frac{\Delta\bar{p}_1^2}{4\sigma_1^2}$$

en hierna gesommeerd over een oneindig aantal waarnemingen en gemiddeld:

$$< \Delta\sigma_1^2 > \simeq \frac{< \Delta\bar{p}_1^2 >}{4\sigma_1^2}$$

of:

$$\sqrt{< \Delta\sigma_1^2 >} \simeq \frac{\sqrt{< \Delta\bar{p}_1^2 >}}{2\sigma_1} \quad \ldots\ldots\ldots\ldots (8)$$

$\sqrt{< \Delta\bar{p}_1^2 >}$ is de middelbare fout van p_1 (een andere schrijfwijze dus voor $\sigma_{\bar{p}_1}$), terwijl $\sqrt{< \Delta\sigma_1^2 >}$ de middelbare fout van σ voorstelt.

Formule (8) toegepast op ons geval geeft, daar $2\sigma_1 \simeq 1$ en $\sqrt{< \Delta\bar{p}_1^2 >}$ $(= \sigma_{\bar{p}_1}) = 0{,}02$, de volgende waarde voor $\sqrt{< \Delta\sigma_1^2 >}$:

$$\sqrt{< \Delta\sigma_1^2 >} \simeq 0{,}02.$$

Ongeveer een zelfde waarde vinden wij voor $\sqrt{< \Delta\sigma_2^2 >}$, $\sqrt{< \Delta\sigma_3^2 >}$ en $\sqrt{< \Delta\sigma_4^2 >}$. Deze waarde voor de onzekerheid van de σ's klopt met die, welke bij onderlinge onafhankelijkheid zou voortvloeien uit het feit, dat σ_1, σ_2, σ_3 en σ_4 bepaald zijn uit de waarnemingen bij $\pm$ 400 monsters. Deze onzekerheid wordt ten naaste bij uitgedrukt door de volgende middelbare fout:

$$\frac{\sigma}{\sqrt{2 \times 400}} \simeq \frac{0{,}5}{27} \simeq 0{,}02.$$

De conclusie, die wij mogen trekken uit de hierboven uitgevoerde berekeningen van de middelbare fouten van de individueele keurmeesters (voor het cijfer voor smaak), is dus, dat σ_1 en σ_3 niet als essentieel verschillend mogen worden beschouwd, evenmin σ_2 en σ_4, maar dat een essentieel verschil bestaat tusschen σ_1 en σ_3 eenerzijds en σ_2 en σ_4 anderzijds.

Berekening van de middelbare fout van den individueelen keurmeester voor „totaal aantal punten".

Voor „totaal aantal punten" hebben wij — ook voor de cijfers, verkregen in 1941 bij het keuren van de 15 meer genoemde monsters per week op de boterkeuring van den G.O.Z. te Zutphen — een zelfde berekening gemaakt als hierboven voor het cijfer voor smaak is geschied.

We vonden hier de volgende *waarnemingsvergelijkingen*:

$$
\begin{aligned}
p_1 + p_2 &= 27{,}81\\
p_1 + p_3 &= 23{,}50\\
p_1 + p_4 &= 29{,}11\\
p_2 + p_3 &= 29{,}13\\
p_2 + p_4 &= 38{,}49\\
p_3 + p_4 &= 30{,}93
\end{aligned}
$$

waaruit volgen de *normaalvergelijkingen*:

$$
\begin{aligned}
3\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + \bar{p}_4 &= 80{,}42\\
\bar{p}_1 + 3\bar{p}_2 + \bar{p}_3 + \bar{p}_4 &= 95{,}43\\
\bar{p}_1 + \bar{p}_2 + 3\bar{p}_3 + \bar{p}_4 &= 83{,}56\\
\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + 3\bar{p}_4 &= 98{,}53
\end{aligned}
$$

Oplossing geeft de volgende waarden.

Waarden van $\bar{p}$ en σ voor de 4 keurmeesters (voor „*totaal aantal punten*")

$$
\begin{aligned}
\bar{p}_1 &= 10{,}38 & \text{dus:}\quad \sigma_1 &= 3{,}2\\
\bar{p}_2 &= 17{,}89 & \sigma_2 &= 4{,}2\\
\bar{p}_3 &= 11{,}95 & \sigma_3 &= 3{,}4\\
\bar{p}_4 &= 19{,}44 & \sigma_4 &= 4{,}4
\end{aligned}
$$

De onzekerheid van $\bar{p}_1$, enz., voortvloeiende uit de strijdigheid der waarnemingsvergelijkingen, bedraagt hier ongeveer 0,9, waaruit een onzekerheid van σ_1, enz. volgt van

$$\sqrt{< \Delta\sigma_1^2 >} \simeq \frac{\sqrt{< \Delta\bar{p}_1^2 >}}{2\sigma_1} \simeq \frac{0{,}9}{6{,}4} \simeq 0{,}14.$$

De middelbare fout, te wijten aan het feit, dat $\pm$ 400 monsters zijn gebruikt voor de berekening, bedraagt ongeveer:

$$\frac{\sigma_1}{\sqrt{2 \times 400}} \simeq \frac{3{,}2}{2{,}7} \simeq 0{,}12.$$

De overeenstemming tusschen de beide schattingen voor de middelbare fout is dus ook hier goed.

Conclusie. *Voor „totaal aantal punten" geldt — evenals voor smaak — dat keurmeester 1 en 3 hun cijfer ongeveer met een zelfde middelbare fout vaststellen, terwijl keurmeesters 2 en 4 een wat grootere middelbare fout vertoonen.*

Hoewel wij in § 4 van dit hoofdstuk dit punt nog nader zullen bespreken, zij nu reeds medegedeeld, dat deze uitspraak niet beteekent, dat de keurmeesters met de grootste middelbare fout minder scherpe beoordeelaars zijn dan keurmeesters, die met een kleinere middelbare fout keuren.

§ 2. De invloed van het gebruik van een nauwe of wijde schaal op de grootte van de berekende middelbare fout.

Bij het lezen van deze verhandeling kan bij den lezer de vraag zijn opgekomen, of bij de berekeningen van de middelbare fout — zoowel voor een combinatie van keurmeesters in hoofdstuk II als voor den individueelen keurmeester in § 1 van dit hoofdstuk — wel voldoende aandacht besteed is aan het feit, dat de éène keurmeester bij het keuren een veel wijdere schaal gebruikt dan de andere. Immers wij hebben hiermede noch bij de berekening van de middelbare fout van de keuring (combinatie van twee keurmeesters) noch bij de berekening van de middelbare fout van één keurmeester rekening gehouden. De afwijkingen tusschen de beoordeelingen van twee keurmeesters, welke worden veroorzaakt, doordat de keurmeesters een verschillende schaalwijdte gebruiken, zijn tot nu toe als fouten aangemerkt.

Is dit juist?

Bij het keuren door een combinatie van twee keurmeesters is er veel voor te zeggen zulke afwijkingen als fouten te beschouwen. Immers aan een ideale keuring moet men den eisch stellen, dat beide keurmeesters een zelfde schaalwijdte gebruiken. Welke schaal dat zal zijn (een wijde of een nauwe) is ten slotte een kwestie van de tweede orde, al moet natuurlijk de voorkeur worden gegeven aan het gebruik van een wijde schaal. Een eerste vereischte is echter, dat beide keurmeesters een gelijke schaalwijdte gebruiken.

Bij het beschouwen van de nauwkeurigheid van één keurmeester ligt de zaak geheel anders. Dan mag zeer zeker een afwijking als gevolg van het gebruik van een nauwere of wijdere schaal *niet* als een fout worden aangemerkt. Deze uitspraak schijnt op het eerste gezicht een ernstige bedreiging van onze — in § 1 uitgevoerde — berekeningen. Om een voorbeeld te noemen: op blz. 93 namen wij aan, dat $\hat{x}_k^{(1)} - \hat{x}_k^{(2)}$ voor iedere waarde van k dezelfde waarde heeft. Indien de keurmeesters 1 en 2 belangrijk verschillende schaalwijdtes gebruiken, zal dit niet juist zijn. Gebruikt b.v. keurmeester 1 een wijde schaal, keurmeester 2 een nauwe schaal, dan zal in het algemeen gelden:

$$\hat{x}_k \text{ zeer laag, dan } \hat{x}_k^{(1)} - \hat{x}_k^{(2)} < 0$$

$$\hat{x}_k \text{ zeer hoog, dan } \hat{x}_k^{(1)} - \hat{x}_k^{(2)} > 0.$$

Het zal echter blijken, dat — ook al zijn de schaalwijdtes voor de verschillende keurmeesters vrij sterk verschillend — hiermede bij de berekening van de subjectieve nauwkeurigheid (middelbare fout) practisch geen rekening behoeft te worden gehouden.

Wij geven hieronder twee calculaties, waaruit blijkt, dat deze conclusie getrokken mag worden.

Bewijs, dat met de schaalwijdte bij de berekening van σ practisch geen rekening behoeft te worden gehouden (I).

Ons bewijsmateriaal ontleenen wij aan de 15 monsters per week, die door drie keurmeesters worden gekeurd.

Onze redeneering is de volgende.

Stel, dat keurmeester 1 van een bepaald monster een totaal aantal punten: $x_k^{(1)}$ toekent, waarbij (1) aangeeft, dat keurmeester 1 het monster heeft beoordeeld en (k), dat die punten zijn toegekend aan het k^{de} monster van de N monsters, die door drie keurmeesters worden gekeurd.

Evenzoo geeft keurmeester 2 aan het l^{de} monster: $x_l^{(2)}$,

keurmeester 3 aan het m^{de} monster: $x_m^{(3)}$.

Bij het keuren van de N monsters worden dus de volgende punten toegekend:

	Gegeven cijfers				Afwijkingen van het gemiddelde		
	1	2	3	Gemiddeld	1	2	3
1e monster:	$x_1^{(1)}$	$x_1^{(2)}$	$x_1^{(3)}$	$\bar{x}_1$	$u_1^{(1)}$	$u_1^{(2)}$	$u_1^{(3)}$
2e monster:	$x_2^{(1)}$	$x_2^{(2)}$	$x_2^{(3)}$	$\bar{x}_2$	$u_2^{(1)}$	$u_2^{(2)}$	$u_2^{(3)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k^{de} monster:	$x_k^{(1)}$	$x_k^{(2)}$	$x_k^{(3)}$	$\bar{x}_k$	$u_k^{(1)}$	$u_k^{(2)}$	$u_k^{(3)}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N^{de} monster:	$x_N^{(1)}$	$x_N^{(2)}$	$x_N^{(3)}$	$\bar{x}_N$	$u_N^{(1)}$	$u_N^{(2)}$	$u_N^{(3)}$

In bovenstaand overzicht is ook het gemiddelde van de door de drie keurmeesters aan een bepaald monster gegeven cijfers ($\bar{x}_k$) aangegeven en tevens de afwijking van de beoordeeling van de keurmeesters met dit gemiddelde opgenomen (u). Er geldt dus:

$$\bar{x}_k = \frac{x_k^{(1)} + x_k^{(2)} + x_k^{(3)}}{3} \quad \ldots\ldots\ldots\ldots (1)$$

en

$$u_k^{(1)} = x_k^{(1)} - \bar{x}_k \quad \ldots\ldots\ldots\ldots (2)$$

en overeenkomstige uitdrukkingen voor $u_k^{(2)}$ en $u_k^{(3)}$.

Wij nemen nu aan, dat alle $x^{(1)}$'s schommelen om een middelwaarde X. Daar de gemiddelde cijfers, die de keurmeesters geven, practisch gelijk zijn, mogen wij aannemen, dat $x^{(2)}$ en $x^{(3)}$ ook om die waarde schommelen.

Noem nu de afwijking, die een door keurmeester 1 gekeurd monster k van die middelwaarde heeft, $v_k^{(1)}$, dus:

$$v_k^{(1)} = x_k^{(1)} - X \quad \ldots\ldots\ldots\ldots (3)$$

Deze $v_k^{(1)}$ is nu gelijk aan w_k (een waarde, welke de ideale keurmeester voor v_k zou vinden), vermenigvuldigd met een factor $\rho^{(1)}$, welke aangeeft, met welke schaalwijdte keurmeester 1 keurt. Daarbij zal echter keurmeester 1 een fout maken: $t_k^{(1)}$, dus:

$$v_k^{(1)} = x_k^{(1)} - X = w_k^{(1)} + t_k^{(1)} = \rho^{(1)}.w_k + t_k^{(1)} \quad \ldots\ldots (4)$$

Terloops zij medegedeeld, dat wij

$$X + \rho^{(1)}w_k = \hat{x}_k^{(1)}$$ kunnen noemen.

De middelbare fout van keurmeester 1 (waaruit thans de fouten als gevolg van een afwijking in ρ zijn geëlimineerd) kan worden aangegeven door:

$$\sigma_1^2 = \langle t_k^{(1)^2} \rangle \simeq \frac{\sum_{k=1}^{N} t_k^{(1)^2}}{N}$$

Evenzoo voor keurmeester 2:

$$\sigma_2^2 = \langle t_k^{(2)^2} \rangle \simeq \frac{\sum_{k=1}^{N} t_k^{(2)^2}}{N}$$ en voor keurmeester 3 : $$\sigma_3^2 = \langle t_k^{(3)^2} \rangle \simeq \frac{\sum_{k=}^{N} t_k^{(3)^2}}{N}.$$

Het gaat er nu om σ_1, resp. σ_2 en σ_3 te bepalen.

Daartoe gaan wij uit van (2) en (4):

$$u_k^{(1)} = x_k^{(1)} - \bar{x}_k = (x_k^{(1)} - X) - (\bar{x}_k - X) = \rho^{(1)} . w_k + t_k^{(1)} - (\bar{x}_k - X) \quad . . (5)$$

Nu is:
$$\bar{x}_k - X = \bar{v}_k = \bar{\rho} . w_k + \bar{t}_k \quad . . . . (6)$$

waarbij
$$\bar{\rho} = \frac{\rho^{(1)} + \rho^{(2)} + \rho^{(3)}}{3} \text{ [1]} \quad \text{en} \quad \bar{t}_k = \frac{t_k^{(1)} + t_k^{(2)} + t_k^{(3)}}{3}.$$

Nu stellen wij $\bar{\rho} = 1$; m.a.w. wij nemen aan, dat de gemiddelde keurmeester de juiste schaalwijdte gebruikt.

Hierdoor gaat (6) over in:

$$\bar{v}_k = \bar{x}_k - X = w_k + \bar{t}_k \quad . . . . (7)$$

Dit ingevuld in (5) geeft:

$$u_k^{(1)} = (\rho^{(1)} - 1)\, w_k + (t_k^{(1)} - \bar{t}_k) \quad . . . . (8)$$

(8) gekwadrateerd levert:

$$u_k^{(1)2} = (\rho^{(1)} - 1)^2 . w_k^2 + (t_k^{(1)} - \bar{t}_k)^2 + 2(\rho^{(1)} - 1) . w_k . (t_k^{(1)} - \bar{t}_k)$$

en gesommeerd over alle waarden:

$$\sum_{k=1}^{N} u_k^{(1)2} = (\rho^{(1)} - 1)^2 . \sum_{k=1}^{N} w_k^2 + \sum_{k=1}^{N} (t_k^{(1)} - \bar{t}_k)^2 + 2(\rho^{(1)} - 1) \sum_{k=1}^{N} w_k (t_k^{(1)} - \bar{t}_k) \quad (9)$$

Nu geldt voor den tweeden term uit het rechterlid van (9):

$$\Sigma\, (t_k^{(1)} - \bar{t}_k)^2 = \Sigma t_k^{(1)2} + \Sigma \bar{t}_k^2 - 2\Sigma\, t_k^{(1)} \bar{t}_k.$$

Verder:
$$\Sigma \bar{t}_k^2 = \Sigma \left(\frac{t_k^{(1)} + t_k^{(2)} + t_k^{(3)}}{3} \right)^2 =$$

$$= \tfrac{1}{9} \Sigma\, (t_k^{(1)2} + t_k^{(2)2} + t_k^{(3)2} + 2t_k^{(1)} t_k^{(2)} + 2t_k^{(1)} t_k^{(3)} + 2t_k^{(2)} t_k^{(3)}) =$$
$$= \tfrac{1}{9}\, \Sigma (t_k^{(1)2} + t_k^{(2)2} + t_k^{(3)2}).$$

Immers daar $t_k^{(1)}$ onafhankelijk is van $t_k^{(2)}$ en $t_k^{(3)}$ en tevens $t_k^{(2)}$ onafhankelijk van $t_k^{(3)}$, geldt (bij voldoend groote N):

$$\Sigma\, t_k^{(1)}\, t_k^{(2)} \simeq N < t_k^{(1)}\, t_k^{(2)} > = 0, \quad \Sigma\, t_k^{(1)}\, t_k^{(3)} \simeq N < t_k^{(1)}\, t_k^{(3)} > = 0$$

en
$$\Sigma\, t_k^{(2)}\, t_k^{(3)} \simeq N < t_k^{(2)}\, t_k^{(3)} > = 0.$$

Voorts is:

$$\Sigma\, t_k^{(1)} \bar{t}_k = \tfrac{1}{3}\, \Sigma\, t_k^{(1)} (t_k^{(1)} + t_k^{(2)} + t_k^{(3)}) = \tfrac{1}{3}\, \Sigma\, t_k^{(1)2}$$

m.a.w. de tweede term uit het rechterlid van (9) is vereenvoudigd tot:

$$\Sigma\, (t_k^{(1)} - \bar{t}_k)^2 = \Sigma\, t_k^{(1)2} + \tfrac{1}{9}\, (\Sigma t_k^{(1)2} + \Sigma\, t_k^{(2)2} + \Sigma\, t_k^{(3)2}) - \tfrac{2}{3}\, \Sigma\, t_k^{(1)2} =$$
$$= \tfrac{4}{9}\, \Sigma\, t_k^{(1)2} + \tfrac{1}{9}\, \Sigma\, (t_k^{(2)2} + t_k^{(3)2}) \quad . . . . . . (10)$$

Thans de derde term uit het rechterlid van (9):

$$\Sigma\, w_k (t_k^{(1)} - \bar{t}_k) = \Sigma\, w_k\, t_k^{(1)} - \tfrac{1}{3}\, (\Sigma\, w_k\, t_k^{(1)} + \Sigma\, w_k\, t_k^{(2)} + \Sigma\, w_k\, t_k^{(3)}) = 0$$

daar w_k onafhankelijk is van $t_k^{(1)}$, $t_k^{(2)}$ en $t_k^{(3)}$.

---

1) De lezer kan zich afvragen, of het rekenkundig middelen van getallen, die een verhouding aangeven, wel juist is. Wij zullen dit op blz. 105 nader bespreken.

(9) gaat dus over in:

$$\sum_{k=1}^{N} u_k^{(1)^2} = (\rho^{(1)} - 1) \cdot \sum_{k=1}^{N} w_k^2 + \tfrac{4}{9} \sum_{k=1}^{N} t_k^{(1)^2} + \tfrac{1}{9} \sum_{k=1}^{N} (t_k^{(2)^2} + t_k^{(3)^2}) \quad . (11^1)$$

en soortgelijke uitdrukkingen voor $\Sigma\, u_k^{(2)^2}$ en $\Sigma\, u_k^{(3)^2}$ nl.

$$\sum_{k=1}^{N} u_k^{(2)^2} = (\rho^{(2)} - 1)^2 \cdot \sum_{k=1}^{N} w_k^2 + \tfrac{4}{9} \sum_{k=1}^{N} t_k^{(2)^2} + \tfrac{1}{9} \sum_{k=1}^{N} (t_k^{(1)^2} + t_k^{(3)^2}). \quad . (11^2)$$

$$\sum_{k=1}^{N} u_k^{(3)^2} = (\rho^{(3)} - 1)^2 \cdot \sum_{k=1}^{N} w_k^2 + \tfrac{4}{9} \sum_{k=1}^{N} t_k^{(3)^2} + \tfrac{1}{9} \sum_{k=1}^{N} (t_k^{(1)^2} + t_k^{(2)^2}). \quad . (11^3)$$

Uit deze drie vergelijkingen kunnen $\Sigma\, t_k^{(1)^2}$, $\Sigma t_k^{(2)^2}$ en $\Sigma t_k^{(3)^2}$ (en dus σ_1, σ_2 en σ_3) berekend worden, indien Σw_k^2 en $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$ bekend zijn. $\Sigma\, u_k^{(1)^2}$, $\Sigma\, u_k^{(2)^2}$ en $\Sigma\, u_k^{(3)^2}$ zijn immers gemakkelijk uit de keuringsuitslagen te berekenen en zijn dus bekend. Wij moeten dus Σw_k^2 en $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$ nog nader trachten te vinden.

Waarde van Σw_k^2. Absoluut nauwkeurig is deze nooit te vinden, omdat wij het oordeel van den idealen keurmeester niet kennen. Het is echter mogelijk Σw_k^2 te benaderen en wel op de volgende wijze.

Formule (6) luidde:

$$\bar{v}_k = w_k + \bar{t}_k$$

of

$$w_k = \bar{v}_k - \bar{t}_k$$

of

$$w_k^2 = \bar{v}_k^2 + \bar{t}_k^2 - 2\bar{v}_k \cdot \bar{t}_k$$

en

$$\Sigma w_k^2 = \Sigma \bar{v}_k^2 + \Sigma \bar{t}_k^2 - 2\Sigma \bar{v}_k \cdot t_k \quad \ldots\ldots\ldots (12)$$

Nu is:

$$\Sigma \bar{v}_k \cdot t_k = \Sigma (w_k + \bar{t}_k)\bar{t}_k = \Sigma w_k \bar{t}_k + \Sigma \bar{t}_k^2 =$$
$$= \tfrac{1}{3} \Sigma w_k (t_k^{(1)} + t_k^{(2)} + t_k^{(3)}) + \Sigma \bar{t}_k^2 = \Sigma \bar{t}_k^2,$$

waardoor (12) wordt vereenvoudigd tot

$$\Sigma w_k^2 = \Sigma \bar{v}_k^2 - \Sigma \bar{t}_k^2 \quad \ldots\ldots\ldots\ldots (13)$$ [1]

of ook

$$\Sigma w_k^2 = \Sigma \bar{v}_k^2 - \tfrac{1}{9} \Sigma (t_k^{(1)^2} + t_k^{(2)^2} + t_k^{(3)^2}) \quad \ldots\ldots (14)$$

---

1) Deze formule is in overeenstemming met die, welke wij op blz. 57 vonden (formule (4).

Daar vonden wij:

$$m_{\mathrm{tot.}}^2 = m^2 - \frac{p-1}{p}\sigma^2.$$

Nu is: $m_{\mathrm{tot.}}^2 = \dfrac{\Sigma \bar{v}_k^2}{N}$ en $\Sigma \bar{t}_k^2 = \dfrac{\sigma^2}{p}$.

Verder geldt: $\dfrac{\Sigma w_k^2}{N} = \lim\limits_{p=\infty} m_{\mathrm{tot.}}^2$.

Daar $\lim\limits_{p=\infty} m_{\mathrm{tot.}}^2 = m^2 - \sigma^2$ geldt tevens: $\dfrac{\Sigma w_k^2}{N} = m^2 - \sigma^2$

dus: $\dfrac{\Sigma w_k^2}{N} = m^2 - \sigma^2 = m^2 - \dfrac{p-1}{p}\sigma^2 - \dfrac{\sigma^2}{p} = m_{\mathrm{tot.}}^2 - \dfrac{\sigma^2}{p} = \dfrac{\Sigma \bar{v}_k^2}{N} - \dfrac{\Sigma \bar{t}_k^2}{N}$.

of $\Sigma w_k^2 = \Sigma \bar{v}_k^2 - \Sigma\, \bar{t}_k^2$

waarmede formule (13) uit deze paragraaf weer is teruggevonden.

Vullen wij deze waarde voor Σw_k^2 in in vergelijkingen (11), dan krijgen wij:

$$\Sigma u_k^{(1)^2} = (\rho^{(1)}-1)^2.\Sigma\bar{v}_k^2 - \tfrac{1}{9}(\rho^{(1)^2}-2\rho^{(1)}-3)\Sigma t_k^{(1)^2} - \tfrac{1}{9}(\rho^{(1)^2}-2\rho^{(1)})\Sigma(t_k^{(2)^2}+t_k^{(3)^2}).$$

Deelen wij deze laatste vergelijking door N en vullen wij in:

$$\overline{u_k^{(1)^2}} = \frac{\sum_{k=1}^{N} u_k^{(1)^2}}{N},\ \overline{\bar{v}_k^2} = \frac{\sum_{k=1}^{N} \bar{v}_k^2}{N} \text{ en } \frac{\sum_{k=1}^{N} t_k^{(1)^2}}{N} \simeq \langle t_k^{(1)^2}\rangle = \sigma_1^2 \text{ (evenzoo } \sigma_2 \text{ en } \sigma_3),$$

dan krijgen wij na eenig omwerken:

$$9\overline{u_k^{(1)^2}} = 9(\rho^{(1)}-1)^2.\overline{\bar{v}_k^2} - (\rho^{(1)^2}-2\rho^{(1)}-3)\sigma_1^2 - (\rho^{(1)^2}-2\rho^{(1)})(\sigma_2^2+\sigma_3^2) \quad (14^1)$$

evenzoo:

$$9\overline{u_k^{(2)^2}} = 9(\rho^{(2)}-1)^2\overline{\bar{v}_k^2} - (\rho^{(2)^2}-2\rho^{(2)}-3)\sigma_2^2 - (\rho^{(2)^2}-2\rho^{(2)})(\sigma_1^2+\sigma_3^2) \quad (14^2)$$

$$9\overline{u_k^{(3)^2}} = 9(\rho^{(3)}-1)^2\overline{\bar{v}_k^2} - (\rho^{(3)^2}-2\rho^{(3)}-3)\sigma_3^2 - (\rho^{(3)^2}-2\rho^{(3)})(\sigma_1^2+\sigma_2^2) \quad (14^3)$$

Uit deze vergelijkingen zijn σ_1, σ_2 en σ_3 op te lossen, indien $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$ bekend zijn.

Waarde van $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$.

Deze hebben wij op de volgende wijze bepaald. Ga voor iederen keurmeester na: het gemiddelde van het absolute verschil (dus zonder rekening te houden met + of — teekens) van het door dien keurmeester toegekende totaal aantal punten en de middelwaarde X en stel deze gemiddelden evenredig met de ρ's, dus:

$$\rho^{(1)} = k\,\overline{|v^{(1)}|}$$
$$\rho^{(2)} = k\,\overline{|v^{(2)}|}$$
$$\rho^{(3)} = k\,\overline{|v^{(3)}|}, \text{ waarbij } |v^{(1)}| = |x_k^{(1)} - X|.$$

(Dat niet op het teeken wordt gelet, wordt aangegeven door het symbool tusschen twee strepen te plaatsen.)

Door optelling wordt verkregen:

$$3\bar{\rho} = k\,\{\overline{|v^{(1)}|} + \overline{|v^{(2)}|} + \overline{|v^{(3)}|}\}.$$

Aangezien wij aannemen, dat $\bar{\rho} = 1$, volgt hieruit, dat

$$k = \frac{3}{\overline{|v^{(1)}|} + \overline{|v^{(2)}|} + \overline{|v^{(3)}|}}$$

of:

$$\rho^{(1)} = \frac{3\,\overline{|v^{(1)}|}}{\overline{|v^{(1)}|} + \overline{|v^{(2)}|} + \overline{|v^{(3)}|}}$$

en soortgelijke uitdrukkingen voor $\rho^{(2)}$ en $\rho^{(3)}$.

Wij hadden ρ ook kunnen bepalen uit:

$$\rho^{(1)} = k\,.\sqrt{\overline{v^{(1)^2}}};\ \rho^{(2)} = k\sqrt{\overline{v^{(2)^2}}} \text{ en } \rho^{(3)} = k\sqrt{\overline{v^{(3)^2}}}$$

m.a.w. door ρ evenredig te stellen aan den wortel uit het gemiddelde kwadraat van het verschil tusschen waargenomen punten aantal en de middelwaarde. Wij hadden dan echter de schaal van den „gemiddelden" keurmeester niet mogen karakteriseeren met:

$$\bar{\rho} = \frac{\rho^{(1)} + \rho^{(2)} + \rho^{(3)}}{3},$$

maar hadden voor dit doel een andere grootheid moeten gebruiken nl.

$$\rho = \sqrt{\frac{\rho^{(1)^2} + \rho^{(1)^2} + \rho^{(3)^2}}{3}}.$$

Deze ρ (d.i. de factor, die de schaal karakteriseert van den „gemiddelden" keurmeester) hadden wij — evenals wij dit deden voor $\bar{\rho}$ — gelijk aan 1 kunnen stellen. Het zou echter veel lastiger geweest zijn deze ρ in de formules te verwerken. Om deze reden hebben wij ρ bepaald door haar evenredig te stellen aan het gemiddelde absolute verschil tusschen waargenomen puntenaantal en de middelwaarde.

Een andere vraag, die wij in de voetnoot op blz. 103 stelden, is, of het wel juist is grootheden als ρ, die een verhouding weergeven, rekenkundig te middelen. Men zou kunnen denken, dat geometrische vereffening (op de wijze, zooals op blz. 45 voor $1 - r^2$ aangegeven) principieel juister is. Dit is hier echter niet het geval. Immers $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$ zijn verkregen door $\overline{|v^{(1)}|}$, $\overline{|v^{(2)}|}$ en $\overline{|v^{(3)}|}$ te deelen door denzelfden vorm nl.

$$\frac{3}{\overline{|v^{(1)}|} + \overline{|v^{(2)}|} + \overline{|v^{(3)}|}}.$$

Hierdoor verliezen $\rho^{(1)}$, $\rho^{(2)}$ en $\rho^{(3)}$ ten opzichte van elkaar het karakter van een verhouding. Daarom mogen deze factoren voor het berekenen van een representatieve middelwaarde arithmetisch gemiddeld worden.

Hiermede zijn dus alle gegevens bekend om de vergelijkingen (14) op te lossen.

Deze vergelijkingen geven de waarden van de middelbare fouten, waarbij de invloed van de gebruikte schaalwijdte opgeheven is. Wenscht men de aldus verkregen waarden te vergelijken met die, welke verkregen zouden zijn, indien met den invloed van de gebruikte schaalwijdte geen rekening was gehouden (waarbij dus afwijkingen in schaalwijdte als fouten zijn opgevat), dan kan men in vergelijkingen (14) eenvoudig $\rho^{(1)} = \rho^{(2)} = \rho^{(3)} = 1$ stellen.

Men verkrijgt dan de volgende drie vergelijkingen:

$$\overline{u^{(1)^2}} = \tfrac{4}{9}\,\sigma_1^2 + \tfrac{1}{9}\,(\sigma_2^2 + \sigma_3^2) \quad \ldots\ldots\ldots \quad (15^1)$$

$$\overline{u^{(2)^2}} = \tfrac{4}{9}\,\sigma_2^2 + \tfrac{1}{9}\,(\sigma_1^2 + \sigma_3^2) \quad \ldots\ldots\ldots \quad (15^2)$$

$$\overline{u^{(3)^2}} = \tfrac{4}{9}\,\sigma_3^2 + \tfrac{1}{9}\,(\sigma_1^2 + \sigma_2^2) \quad \ldots\ldots\ldots \quad (15^3)$$

Hieruit zijn σ_1, σ_2 en σ_3 op te lossen.

Wanneer men nu de ρ's voor de verschillende keurmeesters gaat uitrekenen, blijkt, dat zij soms vrij sterk kunnen verschillen. Wij vonden bijv. voor vier keurmeesters (telkens combinaties van 3):

Tabel 53. Waarden van ρ voor de verschillende keurmeesters.
Materiaal: A1 (Zutphen).

	$\rho^{(1)}$	$\rho^{(2)}$	$\rho^{(3)}$	$\rho^{(4)}$
1e serie, waarbij de combinatie 1, 2 en 3 keurde:	0,877	1,119	1,004	—
2e serie, „ „ „ 1, 2 en 4 „ :	0,961	1,067	—	0,972

terwijl verder gevonden werd:

Tabel 54. Waarden voor $\overline{u^2}$.
Materiaal: als tabel 53.

	Voor de 1e serie	Voor de 2e serie
$\overline{u^{(1)2}}$	8,87	13,58
$\overline{u^{(2)2}}$	8,78	13,16
$\overline{u^{(3)2}}$	8,22	—
$\overline{u^{(4)2}}$	—	13,80
$\overline{v^2}$	21,05	25,99

Stellen wij de vergelijkingen (14) op voor deze beide series, dan krijgen wij het volgende ($\sigma_1^2 = p_1$, $\sigma_2^2 = p_2$, $\sigma_3^2 = p_3$ en $\sigma_4^2 = p_4$ stellende):

$$\left.\begin{aligned} 8{,}87 &= 0{,}015 \times 21{,}05 + \tfrac{1}{9} \times 3{,}985 \times p_1 + \tfrac{1}{9} \times 0{,}985(p_2 + p_3) \\ 8{,}78 &= 0{,}014 \times 21{,}05 + \tfrac{1}{9} \times 3{,}986 \times p_2 + \tfrac{1}{9} \times 0{,}986(p_1 + p_3) \\ 8{,}22 &= \tfrac{1}{9} \times 4 \times p_3 + \tfrac{1}{9} \times (p_1 + p_2) \end{aligned}\right\} \quad . \; . \; . \; (14^A)$$

en

$$\left.\begin{aligned} 13{,}58 &= 0{,}0015 \times 25{,}99 + \tfrac{1}{9} \times 3{,}9985 \times p_1 + \tfrac{1}{9} \times 0{,}9985(p_2 + p_4) \\ 13{,}16 &= 0{,}0045 \times 25{,}99 + \tfrac{1}{9} \times 3{,}9955 \times p_2 + \tfrac{1}{9} \times 0{,}9955(p_1 + p_4) \\ 13{,}80 &= 0{,}0008 \times 25{,}99 + \tfrac{1}{9} \times 3{,}9992 \times p_4 + \tfrac{1}{9} \times 0{,}9992(p_1 + p_2) \end{aligned}\right\} \quad (14^B)$$

Stellen wij de vergelijkingen op voor het geval, dat $\rho^{(1)} = \rho^{(2)} = \rho^{(3)} = \rho^{(4)} = 1$ gesteld wordt, dan worden de volgende vergelijkingen verkregen:

$$\left.\begin{aligned} 8{,}87 &= \tfrac{4}{9} p_1 + \tfrac{1}{9} \times (p_2 + p_3) \\ 8{,}78 &= \tfrac{4}{9} p_2 + \tfrac{1}{9} \times (p_1 + p_3) \\ 8{,}22 &= \tfrac{4}{9} p_3 + \tfrac{1}{9} \times (p_1 + p_2) \end{aligned}\right\} \; . \; . \; . \; . \; . \; . \; . \; . \; (15^A)$$

en

$$\left.\begin{aligned} 13{,}58 &= \tfrac{4}{9} p_1 + \tfrac{1}{9} \times (p_2 + p_4) \\ 13{,}16 &= \tfrac{4}{9} p_2 + \tfrac{1}{9} \times (p_1 + p_4) \\ 13{,}80 &= \tfrac{4}{9} p_4 + \tfrac{1}{9} \times (p_1 + p_2) \end{aligned}\right\} \; . \; . \; . \; . \; . \; . \; . \; (15^B)$$

Oplossing van de bovenstaande vergelijkingen levert de volgende waarden voor p_1, p_2, p_3 en p_4:

	Opgelost uit (14A)	Opgelost uit (15A)		Opgelost uit (14B)	Opgelost uit (15B)
p_1	13,16	13,67	p_1	20,47	20,47
p_2	12,94	13,40	p_2	19,25	19,21
p_3	11,97	11,72	p_4	21,14	21,13

Ondanks de vrij groote verschillen in ρ zijn dus de uitkomsten — berekend met en zonder uitschakeling van den invloed van de schaalwijdte — weinig verschillend. Wij zullen dus — ook indien men de afwijkingen in het gebruik van de schaal niet als fouten wenscht te zien aangemerkt — toch bij de berekeningen ervan mogen uitgaan, dat de schaalwijdtes nooit dusdanig verschillen, dat zij een merkbaren invloed uitoefenen op de waarde van de berekende middelbare fout.

Opvallend is, dat de waarden voor p_1, p_2 en p_3 bij de vergelijkingen A (1e serie) en die voor p_1 en p_3 en p_4 bij de vergelijkingen B (2e serie) zeer dicht bij elkaar liggen. Direct worde vermeld, dat lang niet altijd een dergelijke overeenstemming kan worden bereikt. Dit blijkt reeds daaruit, dat uit vergelijkingen A wel ongeveer gelijke waarden voor p_1, p_2 en p_3 worden verkregen, maar dat deze waarden vrij sterk verschillen van die uit vergelijkingen B. Dit laatste wordt verklaard door het feit, dat voor vergelijkingen A slechts 45 monsters als grondslag hebben gediend, bij vergelijkingen B slechts 35 monsters. Uit A volgt, dat $\sigma_1 \backsimeq \sigma_2 \backsimeq \sigma_3 \backsimeq 3{,}50$ en uit B: $\sigma_1 \backsimeq \sigma_2 \backsimeq \sigma_4 \backsimeq 4{,}50$, waaruit voldoende het groote verschil blijkt. Ons voorbeeld had echter ook niet de bedoeling een nauwkeurige berekening te geven van de middelbare fouten van de individueele keurmeesters. Daarvoor is het aantal monsters veel te gering.

Het door ons gestelde doel — nl. te bewijzen, dat met de schaalwijdte voor de berekening van de middelbare fout practisch geen rekening behoeft te worden gehouden — hebben wij er echter wel mee bereikt.

Bewijs, dat met de schaalwijdte bij de berekening van σ practisch geen rekening behoeft te worden gehouden (II).

Bij deze tweede calculatie zijn wij niet uitgegaan van de verschillen van het cijfer van één keurmeester ten opzichte van het gemiddelde van drie keurmeesters, maar hier hebben wij als uitgangspunt gekozen het verschil tusschen het oordeel van twee keurmeesters.

De redeneering was de volgende.

Het cijfer, dat keurmeester 1 toekent, ($x_k^{(1)}$) kan ingevolge (4) van blz. 101 worden voorgesteld door:

$$x_k^{(1)} = X + \rho^{(1)} . w_k + t_k^{(1)}$$

evenzoo:

$$x_k^{(2)} = X + \rho^{(2)} . w_k + t_k^{(2)} .$$

Wij merken terloops op, dat wij hierbij — evenals op blz. 101 — in tegenstelling met blz. 93 ervan uitgaan, dat beide keurmeesters gemiddeld, over een groot aantal monsters gerekend, een zelfde cijfer geven (nl. beide: X). Wij zouden, evenals wij dat deden op blz. 93, ook hier moeten uitgaan van twee middelwaarden, voor iederen keurmeester één (dus $X^{(1)}$ en $X^{(2)}$). De calculaties zouden daardoor echter veel ingewikkelder worden, terwijl het resultaat door onze vereenvoudiging practisch niet wordt beinvloed. Immers wij hebben in hoofdstuk II § 2 reeds bewezen, dat het gemiddelde oordeel van de keurmeesters zeer weinig uiteenloopt. Wij mogen dus wel aannemen, dat

$$X^{(1)} \backsimeq X^{(2)} \backsimeq X.$$

Wij blijven dus uitgaan van slechts één middelwaarde X voor beide keurmeesters. De berekening gaat nu als volgt verder:

$$V_k = x_k^{(1)} - x_k^{(2)} = (\rho^{(1)} - \rho^{(2)})\, w_k + t_k^{(1)} - t_k^{(2)}$$

hetgeen volgt uit bovenstaande vergelijkingen.

Uit deze laatste vergelijking volgt gemakkelijk:

$$\Sigma V_k^2 = (\rho^{(1)} - \rho^{(2)})^2\, \Sigma w_k^2 + \Sigma t_k^{(1)2} + \Sigma t_k^{(2)2}$$

en

$$m_V^2 = (\rho^{(1)} - \rho^{(2)})^2 \,.\, \overline{w_k^2} + \sigma_1^2 + \sigma_1^2 \quad \ldots\ldots\ldots \quad (16)$$

Dergelijke uitdrukkingen krijgen wij voor de combinatie 1, 3; 1, 4; 2, 3; 2, 4 en 3, 4, dus 6 vergelijkingen met 5 onbekenden (nl. σ_1, σ_2, σ_3, σ_4 en $\overline{w_k^2}$).

Wanneer wij het bovenstaande toepassen op de verschillen, verkregen tusschen de keurmeesters gedurende 1941 bij de keuring van de 15 monsters per week (materiaal dus, dat wij reeds eerder gebruikten), dan krijgen wij onderstaande zes vergelijkingen. Hierin is: $\sigma_1^2 = p_1$; $\sigma_2^2 = p_2$, $\sigma_3^2 = p_3$ en $\sigma_4^2 = p_4$ en $\overline{w_k^2} = p_5$ gesteld.

$$p_1 + p_2 + (\rho^{(1)} - \rho^{(2)})^2 \times p_5 = 28{,}10\,^{1)} \quad \ldots\ldots \quad (17^1)$$
$$p_1 + p_3 + (\rho^{(1)} - \rho^{(3)})^2 \times p_5 = 23{,}53 \quad \ldots\ldots \quad (17^2)$$
$$p_1 + p_4 + (\rho^{(1)} - \rho^{(4)})^2 \times p_5 = 29{,}16 \quad \ldots\ldots \quad (17^3)$$
$$p^2 + p_3 + (\rho^{(2)} - \rho^{(3)})^2 \times p_5 = 29{,}24 \quad \ldots\ldots \quad (17^4)$$
$$p_2 + p_4 + (\rho^{(2)} - \rho^{(4)})^2 \times p_5 = 38{,}69 \quad \ldots\ldots \quad (17^5)$$
$$p_3 + p_4 + (\rho^{(3)} - \rho^{(4)})^2 \times p_5 = 30{,}93 \quad \ldots\ldots \quad (17^6)$$

Om de ρ's te bepalen hebben wij voor iederen keurmeester het gemiddelde van de absolute verschillen, die alle door den betreffenden keurmeester gekeurde monsters vertoonden ten opzichte van het gemiddelde van door hem gegeven cijfers, berekend.

Aangenomen, dat door keurmeester 1 in totaal N monsters werden gekeurd (waarvan Y_k monsters een puntenaantal verkregen, aangegeven door klasse k, dus $\sum\limits_{k=1}^{n} Y_k = N$, indien n het aantal klassen voorstelt), terwijl het verschil tusschen voorloopig aangenomen gemiddelde (X) en het cijfer voor het bepaalde monster v_k wordt genoemd, werd eerst berekend:

$$\sum_1^n Y_k \,|v_k|$$

en hieruit:

$$\sum_1^n Y_k \,|u_k|\,,$$

waarbij $|u_k|$ aangeeft het absolute verschil tusschen het toegekende cijfer en het werkelijke gemiddelde $\left(\frac{1}{N}\sum\limits_1^n Y_k\, x_k^{(1)} = \overline{x^{(1)}}\right)$.

De omrekening van $\sum\limits_1^n Y_k \,|\, v_k \,|$ in $\sum\limits_1^n Y_k \,|\, u_k \,|$ had plaats met behulp van de volgende formule:

$$\sum_1^n Y_k \,|\, u_k \,| = \sum_1^n Y_k \,|\, v_k \,| - \left(\sum_{j+1}^{n} Y_k - \sum_1^{j} Y_k\right)\delta.$$

1) Hierin is voor de berekening van m_V^2 uit $\frac{\Sigma V_k^2}{N}$ de correctie van Sheppard toegepast.

Hierbij is:

$$\delta = \bar{v} \left(= \frac{\sum_{1}^{n} Y_k \, v_k}{N} \right)$$

$\sum_{j+1}^{n} Y_k$: het totaal aantal monsters, dat ten opzichte van het werkelijke gemiddelde ($\overline{x^{(1)}}$) een *positief* verschil vertoonde.

$\sum_{1}^{j} Y_k$: het totaal aantal monsters, dat ten opzichte van het werkelijke gemiddelde een *negatief* verschil vertoonde.

Deze formule mag slechts worden toegepast, als men zorg draagt, dat X zoodanig wordt gekozen, dat

$$\delta < c \quad (c = \text{klasse-interval}).$$

Men zal begrijpen, dat de grootheid X (en v_k) is ingevoerd om de berekeningen te vereenvoudigen.

Het gemiddelde absolute verschil, waarover wij boven spraken en waar het om begonnen is, bedraagt nu:

$$\overline{|u|} = \frac{\sum_{1}^{n} Y_k \, | u_k |}{N}.$$

Voor $\overline{|u|}$ voor de vier keurmeesters vonden wij uit het beschikbare materiaal de volgende waarden.

Tabel 55. Waarde van $\overline{|u|}$ voor de verschillende keurmeesters.
Materiaal: A2 (Zutphen).

	Aantal gekeurde monsters	$\overline{\|u\|}$
Keurmeester 1	585	4,03
,, 2	600	4,77
,, 3	585	3,77
,, 4	550	4,16
Gemiddeld		4,18

De ρ werd nu berekend door de gemiddelde waarde van $\overline{|u|}$ voor de vier keurmeesters te deelen op de waarde van $\overline{|u|}$ voor den betreffenden keurmeester.

We vonden zoodoende het volgende (zie tabel 56).

Tabel 56. Waarde van ρ voor de verschillende keurmeesters.
Materiaal: als tabel 55.

	ρ
Keurmeester 1	0,964
„ 2	1,141
„ 3	0,901
„ 4	0,995

Vullen wij deze waarden voor ρ in in de op blz. 108 genoemde waarnemingsvergelijkingen (17), dan krijgen wij:

$$\left.\begin{aligned} p_1 + p_2 + 0{,}177^2 \times p_5 &= 28{,}10 \\ p_1 + p_3 + 0{,}063^2 \times p_5 &= 23{,}53 \\ p_1 + p_4 + 0{,}031^2 \times p_5 &= 29{,}16 \\ p_2 + p_3 + 0{,}240^2 \times p_5 &= 29{,}24 \\ p_2 + p_4 + 0{,}146^2 \times p_5 &= 38{,}69 \\ p_3 + p_4 + 0{,}094^2 \times p_5 &= 30{,}93 \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (18)$$

Uit deze 6 vergelijkingen zijn p_1, p_2, p_3, p_4 en p_5 op te lossen. Aangezien echter de factoren van p_5 zeer klein zijn (ten hoogste $0{,}240^2 = 0{,}0576$) en p_5 wel ten naaste bij bekend is, begaan wij geen merkbare fout en vereenvoudigen wij de berekening aanmerkelijk, indien wij deze waarde voor p_5 hier invullen. Een behoorlijke schatting van p_5 kunnen wij verkrijgen, indien wij bedenken, dat

$$p_5 = \overline{w_k^2} = m^2 - \sigma^2 \qquad \text{(zie voetnoot blz. 103)}$$

De in deze formule genoemde m^2 geeft de middelbare afwijking aan van het toegekende cijfer ten opzichte van het gemiddelde cijfer, dat door één keurmeester aan alle monsters is gegeven, m.a.w.

$$m^2 = \overline{u^2}$$

m^2 heeft nu in ons geval de volgende waarden.

Tabel 57. Waarden van m^2 voor de verschillende keurmeesters, berekend op twee verschillende manieren.
Materiaal: als tabel 55.

	m^2 (bepaald uit $\overline{u^2}$)	m^2 $\left(\text{berekend uit } \overline{u^2} = \frac{\pi}{2}\,\overline{\lvert u \rvert}^2\right)$
Keurmeester 1	25,57	25,50
„ 2	36,19	35,72
„ 3	22,30	22,31
„ 4	29,43	27,18
Gemiddeld	28,37	27,68

In kolom 1 zijn de berekende waarden genoteerd. Wij merken terloops op, dat deze waarden goed kloppen met die, welke wij vinden, indien m^2 wordt berekend uit $\overline{|u|}$. Immers bij een normale verdeeling geldt:

$$\overline{u^2} = \frac{\pi}{2}\,\overline{|u|}^2.$$

Hoewel de verdeeling van de door één keurmeester toegekende cijfers rondom het gemiddelde van deze cijfers niet geheel normaal is, is de overeenstemming toch heel goed, hetgeen de tweede kolom, waarin de aldus berekende waarden genoteerd zijn, duidelijk laat zien.

Uit tabel 57 volgt, dat wij voor m^2 gemiddeld mogen nemen: 28.

Een globale schatting van de waarde van σ^2 krijgen wij als volgt.

Op blz. 29 vonden wij, dat een combinatie van twee keurmeesters keurt met een middelbare fout van iets minder dan 2,8. Hieruit volgt, dat één keurmeester keurt met een middelbare fout van $2{,}8\sqrt{2}$ m.a.w.

$$\sigma^2 = (2{,}8\sqrt{2})^2 = 15{,}76$$

of afgerond naar beneden: $\sigma^2 \simeq 15$.

Bij een waarde van $m^2 = 28$ en $\sigma^2 = 15$ vinden wij dus voor p_5:

$$p_5 = 28 - 15 = 13.$$

Op ongeveer dezelfde waarde voor p_5 komen wij, als wij de gegevens van tabel 36 van blz. 58 gebruiken. Daar staat vermeld, dat $m^2_{\text{tot.}}$ bij een oneindig aantal keurmeesters een waarde heeft van:

$$\underset{p=\infty}{m^2_{\text{tot.}}} \; (= \overline{w^2_k}) = 0{,}89 \times \sigma^2.$$

Verder volgt uit dezelfde tabel, dat $\underset{(p=1)}{m^2} = 1{,}89 \times \sigma^2$ m.a.w.

$$\overline{w^2_k} = 0{,}89 \times \frac{m^2}{1{,}89} = 0{,}89 \times \frac{28}{1{,}89} = 13{,}2.$$

Wij gaan nu verder werken met de waarde $\overline{w^2_k}\; (= p_5) = 13$.

De waarde van $p_5 = 13$ ingevuld in de waarnemingsvergelijkingen (18) levert:

$$\begin{aligned}
p_1 + p_2 &= 27{,}69\\
p_1 + p_3 &= 23{,}48\\
p_1 + p_4 &= 29{,}15\\
p_2 + p_3 &= 28{,}49\\
p_2 + p_4 &= 38{,}41\\
p_3 + p_4 &= 30{,}82
\end{aligned}$$

De hieruit afgeleide normaalvergelijkingen zijn:

$$\begin{aligned}
3\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + \bar{p}_4 &= 80{,}32\\
\bar{p}_1 + 3\bar{p}_2 + \bar{p}_3 + \bar{p}_4 &= 94{,}59\\
\bar{p}_1 + \bar{p}_2 + 3\bar{p}_3 + \bar{p}_4 &= 82{,}79\\
\bar{p}_1 + \bar{p}_2 + \bar{p}_3 + 3\bar{p}_4 &= 98{,}38
\end{aligned}$$

waaruit volgt: $\bar{p}_1 = 10{,}49 \quad (10{,}38)$
$\bar{p}_2 = 17{,}62 \quad (17{,}89)$
$\bar{p}_3 = 11{,}72 \quad (11{,}95)$
$\bar{p}_4 = 19{,}52 \quad (19{,}44)$

Vergelijken wij deze waarden met die, verkregen op blz. 99 (hier achter de andere waarden tusschen haakjes geplaatst) — waarbij met een verschil in ρ geen rekening is gehouden —, dan blijken de verschillen gering te zijn, niettegenstaande ρ voor de verschillende keurmeesters vrij sterk uiteenloopt.

Onze conclusie uit de beide calculaties mag dus zijn, dat voor de berekening van de subjectieve nauwkeurigheid van één keurmeester met een verschil in schaalwijdte geen rekening behoeft te worden gehouden (tenzij de schaalwijdtes zeer ver uit elkaar liggen). Dat deze conclusie ook kan gelden voor een combinatie van twee keurmeesters spreekt vanzelf, daar hier de schaalwijdtes minder uiteenloopen dan bij de keurmeesters afzonderlijk.

Voor het berekenen van de middelbare fout van één keurmeester zullen wij dus steeds de in § 1 besproken eenvoudige methode mogen toepassen. Tevens kan het in deze paragraaf besprokene duidelijk maken, dat wij in hoofdstuk II bij de bespreking van de middelbare fout van een combinatie van keurmeesters terecht over de schaalwijdte niet hebben gesproken.

§ 3. De subjectieve nauwkeurigheid van een keurmeester (II).

In § 1 bespraken wij een methode om de subjectieve nauwkeurigheid van den individueelen keurmeester vast te stellen, waarbij wij uitgingen van de verschillen tusschen twee keurmeesters.

In § 2 noemden wij een tweede methode, waarbij werd uitgegaan van de verschillen, die het cijfer van één keurmeester vertoonde ten opzichte van het gemiddelde van drie. De vergelijkingen (15) maakten het mogelijk de middelbare fouten van de drie keurmeesters uit deze verschillen te berekenen. Deze methode hebben wij echter nooit voor groote aantallen toegepast. Wel hebben wij over verscheidene kwartalen van de afgeloopen vijf jaren voor geur, smaak, gehalte-en-bewerking, en „totaal aantal punten" uitgerekend:

$$\sqrt{\overline{u^{(1)^2}}},\quad \sqrt{\overline{u^{(2)^2}}},\quad \sqrt{\overline{u^{(3)^2}}} \text{ en } \sqrt{\overline{u^{(4)^2}}} \qquad \text{(zie ook blz. 104).}$$

Deze grootheden kunnen bij benadering worden opgevat als een maat vóor de grootte van de middelbare fout van de individueele keurmeesters. Zooals uit vergelijkingen (15) van blz. 106 volgt, is dit slechts ten naaste bij juist. Als σ_1, σ_2, σ_3 en σ_4 veel verschillen, wordt de maat minder deugdelijk.

Wij geven hieronder een overzicht van een deel der verkregen resultaten:

Tabel 58. Schatting van de middelbare fout voor de keurmeesters afzonderlijk.
Materiaal: A1 (Zutphen).
4 Geld.-Overijselsche keurmeesters (1, 2, 3, 4).

	geur				smaak				gehalte-en-bew.				tot. aant. punten			
Keurmeesters	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Zomer 1937 (135—150 monsters)	0,46	0,45	0,48	0,55	0,54	0,65	0,45	0,58	—	—	—	—	—	—	—	—
4e kwartaal 1939 (135—150 monsters)	0,48	0,50	0,47	0,61	0,51	0,61	0,49	0,49	0,39	0,42	0,35	0,36	—	—	—	—
1e kwartaal 1940 (135—150 monsters)	0,48	0,53	0,49	0,50	0,48	0,55	0,49	0,49	0,28	0,34	0,26	0,26	2,80	3,78	2,92	2,80
3e kwartaal 1940 (90—110 monsters)	0,46	0,55	0,47	0,57	0,51	0,61	0,49	0,49	0,39	0,42	0,35	0,36	3,02	3,76	3,20	3,34
4e kwartaal 1940 (65—85 monsters)	0,41	0,50	0,51	0,48	0,44	0,50	0,45	0,53	0,32	0,36	0,28	0,31	2,77	3,72	3,13	3,53
1e kwartaal 1941 (125—140 monsters)	0,47	0,49	0,47	0,55	0,49	0,50	0,43	0,56	0,30	0,36	0,32	0,34	2,98	3,38	2,80	3,61
Gemiddeld	0,46	0,50	0,48	0,54	$0{,}49^5$	0,57	0,47	0,52	$0{,}33^5$	0,38	0,31	$0{,}32^5$	2,89	3,66	3,01	3,32

Deze staat geeft zeker een beeld van de orde van grootte van de middelbare fout der individueele keurmeesters. Zij bevestigt de conclusie, die wij ook reeds vroeger trokken uit ander materiaal, dat keurmeesters 1 en 3 over het algemeen de grootste subjectieve nauwkeurigheid bereiken. De verschillen zijn niet zeer uitgesproken en lang niet altijd constant, hetgeen te begrijpen is, daar de grootheden uit aantallen zijn berekend, die om de 100—150 liggen.

Wij geven bovenstaande cijfers ook, omdat zij gelegenheid bieden, zij het slechts ruw, de waarden voor de middelbare fouten te toetsen, die wij in hoofdstuk II voor geur, smaak, gehalte-en-bewerking en „totaal aantal punten" opgaven.

Indien wij nl. in vergelijking (15) blz. 106 stellen $\sigma_1 = \sigma_2 = \sigma_3 = \sigma$, dan blijkt, dat

$$\overline{u^2} = \tfrac{6}{9}\sigma^2$$

of

$$\sigma^2 = \tfrac{9}{6}\,\overline{u^2}$$

of daar de nauwkeurigheid van een combinatie van twee keurmeesters ($\sigma_{\bar{x}}$) gelijk is aan:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{2}$$

mag men besluiten:

$$\sigma_{\bar{x}}^2 = \tfrac{9}{12}\,\overline{u^2} = \tfrac{3}{4}\,\overline{u^2}$$

of

$$\sigma_{\bar{x}} \simeq 0{,}866 \times \sqrt{\overline{u^2}}.$$

Vullen wij de gemiddelde waarden voor $\sqrt{\overline{u^2}}$ uit bovenstaanden staat in, dan krijgen wij de volgende schattingen voor de middelbare fouten.

Tabel 59. Globale schatting van de middelbare fout voor een combinatie van twee keurmeesters.

voor geur:	$\sigma_{\bar{x}} \simeq 0{,}866 \times 0{,}49^5 \simeq 0{,}43$
voor smaak:	$\sigma_{\bar{x}} \simeq 0{,}866 \times 0{,}51 \simeq 0{,}44$
voor gehalte-en-bewerking:	$\sigma_{\bar{x}} \simeq 0{,}866 \times 0{,}34 \simeq 0{,}30$
voor „totaal aantal punten":	$\sigma_{\bar{x}} \simeq 0{,}866 \times 3{,}22 \simeq 2{,}79$

Deze waarden sluiten goed aan bij die, welke wij in hoofdstuk II vonden (blz. 30).

§ 4. De objectieve nauwkeurigheid van één keurmeester.

In het voorgaande is reeds uiteengezet, dat wij onderscheid moeten maken tusschen de middelbare fout van den individueelen keurmeester (d.i. zijn subjectieve nauwkeurigheid) en de objectieve nauwkeurigheid.

In § 2 van dit hoofdstuk zagen wij, dat voor de berekening van de subjectieve nauwkeurigheid met het gebruik van een afwijkende schaalwijdte (ρ) practisch geen rekening behoefde te worden gehouden. Bij het berekenen van de objectieve nauwkeurigheid speelt echter de schaalwijdte een overwegende rol, zooals wij in hoofdstuk II ook reeds hebben gezien.

Evenals wij dat in bovengenoemd hoofdstuk deden, zullen wij ook hier de objectieve nauwkeurigheid uitdrukken in den correlatiecoëfficiënt, die het verband aangeeft tusschen het door den betreffenden keurmeester toegekende cijfer en het — voor dien keurmeester geldende — ware cijfer.

Den correlatiecoëfficiënt berekenden wij weer met de bekende formule:

$$\frac{\sigma_i^2}{m_{\text{tot.}}^2} = 1 - r_i^2.$$

Uit tabel 57 en uit het medegedeelde op blz. 112 blijkt, dat deze grootheden (σ^2 en $m_{\text{tot.}}^2$) voor de 4 keurmeesters de volgende waarden hebben.

Tabel 60. Berekening van den correlatiecoëfficiënt voor de 4 keurmeesters. Materiaal: A2 (Zutphen).

	σ_i^2	$m_{\text{tot.}}^2$	$1 - r_i^2 \left(= \frac{\sigma_i^2}{m_{\text{tot.}}^2}\right)$	r_i^2	r_i
Keurm. $i =$ 1	10,49	25,57	0,410	0,590	0,768
„ 2	17,62	36,19	0,487	0,513	0,716
„ 3	11,72	22,30	0,521	0,479	0,692
„ 4	19,52	29,43	0,664	0,336	0,580

Wij zien nu een geheel ander beeld optreden dan bij de subjectieve nauwkeurigheid. Hadden keurmeesters 2 en 4 een zeer groote middelbare fout (kleine subjectieve nauwkeurigheid), thans blijkt, dat — wat objectieve nauwkeurigheid betreft — keurmeester 2 niet voor keurmeester 1 en 3 behoeft onder te doen. Keurmeester 4 blijft ook in objectieve nauwkeurigheid iets beneden de andere drie.

In hoeverre zijn de verschillen in berekende r essentieel?

Volgens blz. 45 formule (4), bedraagt de middelbare fout van de berekende r:

$$\sigma_r = \frac{1 - r^2}{\sqrt{N}}.$$

Daar in ons geval: $r \simeq 0{,}7$ en $1 - r^2 \simeq 0{,}50$ en $N \simeq 600$, geldt

$$\sigma_r = \frac{0{,}5}{\sqrt{600}} \simeq 0{,}02,$$

waaruit volgt, dat de waarden van r voor keurmeester 1, 2 en 3 practisch principieel niet sterk verschillen. Alleen de waarde van r voor keurmeester 4 is belangrijk lager.

Terloops merken wij op, dat formule (7) van § 7 hoofdstuk II blz. 58 ons in staat stelt te berekenen, hoeveel malen de éene keurmeester nauwkeuriger is dan de andere. Aangenomen, dat voor een bepaalden keurmeester geldt: $\sigma^2 = 24$ en $m^2_{tot.} = 32$ (dus $1 - r^2 = 0{,}75$ en $r = 0{,}50$), terwijl voor een anderen keurmeester een correlatiecoëfficiënt wordt gevonden van $r = 0{,}63$, dan kan uit bedoelde formule worden afgeleid, dat twee van de eerstgenoemde keurmeesters met een gelijke objectieve nauwkeurigheid keuren als één keurmeester, wiens objectieve nauwkeurigheid wordt uitgedrukt door $r = 0{,}63$.

Immers bovenbedoelde formule luidt:

$$\frac{\sigma^2}{pm^2 - (p-1)\sigma^2} = 1 - r^2$$

Voor $p = 2$, $\sigma^2 = 24$ en $m^2 = 32$ wordt dit:

$$\frac{24}{2 \times 32 - 24} = \frac{24}{40} = 0{,}6 \; (= 1 - r^2)$$

dus: $r = 0{,}63$.

§ 5. Over de beteekenis van de subjectieve en de objectieve nauwkeurigheid en van de toegepaste schaalwijdte.

Welke van beide — de subjectieve of de objectieve nauwkeurigheid — moet nu als de belangrijkste worden beschouwd?

Voor het nagaan van de mate, waarin een keurmeester in staat is een vaststaand oordeel uit te spreken over de kwaliteit van de boter, heeft alleen de objectieve nauwkeurigheid waarde. De keurmeester met een groote objectieve nauwkeurigheid zal bij herhaling van de beoordeeling steeds een zelfde — ten minste een slechts weinig afwijkend — oordeel geven, terwijl bij een kleinere objectieve nauwkeurigheid het oordeel nog al eens verschillend zal zijn.

Zou het oordeel der keurmeesters steeds worden uitgedrukt in een aanduiding (bijv. goed, matig, slecht), dan was alleen de objectieve nauwkeurigheid van waarde. Zoodra echter bij een keuring het oordeel wordt uitgedrukt in een cijfer, moet de subjectieve nauwkeurigheid ook van groote beteekenis worden geacht. Deze immers geeft aan, in welke mate het gegeven *cijfer* nauwkeurig is. We hebben in hoofdstuk II deze aangelegenheid uitvoerig behandeld.

In vorige paragrafen hebben wij gezien, dat een groote middelbare fout soms gepaard gaat met de toepassing van een groote schaalwijdte, zoodat de subjectieve nauwkeurigheid van de keurmeesters hoogstwaarschijnlijk veel minder uiteen zou loopen, indien ze een zelfde schaalwijdte toepasten. Bij de practische keuring heeft het gebruik van een gelijke schaalwijdte bovendien nog andere voordeelen. Het is van belang, dat bij de keuring een bepaald cijfer ook een bepaalde kwaliteit uitdrukt. Anders zou bij den éénen keurmeester (of bij de ééne combinatie) een bepaald cijfer voor bijv. „totaal aantal punten" heel iets anders kunnen beteekenen dan bij den anderen keurmeester. Dit zou voor een keuring, waar niet iedere week dezelfde keurmeesters keuren, zeer ongewenscht zijn.

Ons onderzoek heeft geleerd, dat de door de individueele keurmeesters toegepaste schaalwijdtes soms aanmerkelijke verschillen vertoonen. Toch wordt het verschil bij een combinatie van twee keurmeesters (waarmede men in de practijk te maken heeft) sterk genivelleerd. Nemen wij bijv. de combinatie (1, 3) tegenover de combinatie (2, 4) (zie blz. 109), dan bedraagt $\overline{|u|}$

$$\text{voor de combinatie } (1, 3) \quad \frac{4{,}03 + 3{,}77}{2} = 3{,}90,$$

$$\text{voor de combinatie } (2, 4) \quad \frac{4{,}77 + 4{,}16}{2} = 4{,}47$$

hetgeen dus zal beteekenen, dat, indien combinatie (1, 3) een cijfer geeft van 57 voor totaal aantal punten, combinatie (2, 4) een cijfer zal geven van $67 - \frac{4{,}47}{3{,}90} \times (67 - 57) = 67 - 11{,}5 = 55{,}5$ (waarbij 67 het gemiddelde van alle monsters voorstelt). Het verschil is dus, zelfs indien het cijfer ver van het algemeen gemiddelde ($\bar{x}$) ligt, betrekkelijk gering. Over het algemeen is het verschil nog kleiner, in de eerste plaats, omdat wij in bovenstaand overzicht juist die beide combinaties tegenover elkaar stelden, die het meest in schaalwijdte verschilden, in de tweede plaats, omdat wij een cijfer (57) namen, dat zoo ver van het algemeen gemiddelde (67) ligt, dat de frequentie ervan slechts zeer klein is. Niettegenstaande in de practijk dus het verschil in schaalwijdte niet een belangrijken invloed uitoefent op het uiteindelijk te geven cijfer, is het toch wenschelijk, dat de keurmeesters zich in het gebruik van de schaalwijdte zooveel mogelijk bij elkaar trachten aan te sluiten.

SLOTBESCHOUWING.

Nu wij aan het einde van de behandeling van ons onderzoek zijn gekomen, is het wenschelijk de conclusies, waartoe wij in de afzonderlijke hoofdstukken kwamen, nogmaals in vogelvlucht te overzien.

Uit mathematisch-statistisch oogpunt beschouwd, is de conclusie, die wij in hoofdstuk II trokken, wel de meest belangrijke. Het feit, dat daarin bewezen kon worden, dat de fouten, die bij het organoleptisch onderzoek door de keurmeesters worden gemaakt, een normale verdeeling vertoonen, maakte het mogelijk, dat wij de betrekkelijk eenvoudige wetten, die de normale verdeeling beheerschen, konden toepassen en resultaten konden verkrijgen, waardoor een exact inzicht werd verkregen in de nauwkeurigheid zoowel van een combinatie van keurmeesters als van den individueelen keurmeester.

Wij konden een beeld geven van de verschillen, die bij de beoordeeling van boter als gevolg van fouten van de keurmeesters kunnen optreden. De vraag, die in de practijk zoo dikwijls wordt gesteld, nl. in hoeverre nu het organoleptisch onderzoek van boter als betrouwbaar kan worden beschouwd, kon in groote lijnen volledig tot een oplossing worden gebracht.

Door het geheele vraagstuk der nauwkeurigheid van het boterkeuren op verschillende punten zoo grondig en zoo exact mogelijk te behandelen, kon het inzicht in de problemen, die zich hierbij voordoen, worden verdiept.

SAMENVATTING.

Dit proefschrift omvat een onderzoek naar de nauwkeurigheid, waarmede het toekennen van een cijfer bij het keuren van boter geschiedt.

Het onderzoek is uitgevoerd aan de hand van materiaal, dat de wekelijksche boterkeuringen van den Geldersch-Overijselschen Bond van Coöperatieve Zuivelfabrieken te Zutphen opleverde, terwijl tevens gegevens zijn verwerkt van boterkeuringen van andere zuivelbonden.

Bij de verwerking van de gegevens hebben wij in ruime mate gebruik gemaakt van de wiskundige statistiek. In deze samenvatting zullen wij op de gebruikte methoden en de mathematische problemen, die zich bij het onderzoek voordeden en welker behandeling een vrij belangrijk deel van het proefschrift uitmaakt, niet ingaan. Wij verwijzen den lezer daartoe naar het eigenlijke geschrift. In dit overzicht geven wij alleen een kort resumé der verkregen practische resultaten.

Wij stellen voorop, dat bij de hier behandelde boterkeuringen de kwaliteit van de boter wordt onderscheiden in drie onderdeelen: 1°. geur; 2°. smaak; 3°. gehalte-en-bewerking. De beoordeeling geschiedt door *twee* keurmeesters.

De keurmeesters waardeeren deze onderdeelen volgens een puntenschaal van 0—10 punten (waarvan in de practijk het traject 3 tot en met 8 wordt gebruikt). Het eindoordeel („totaal aantal punten") wordt gevonden door de som van de door beide keurmeesters gegeven cijfers voor geur op te tellen bij *twee*maal de beide cijfers voor smaak + *twee*maal de beide cijfers voor gehalte-en-bewerking. [1])

Het onderzoek omvat een analyse van de nauwkeurigheid, waarmede één keurmeester of een combinatie van twee of meer keurmeesters de cijfers voor geur, voor smaak, voor gehalte-en-bewerking en voor „totaal aantal punten" vaststellen.

Allereerst toonden wij aan, dat keurmeesters, die regelmatig te zamen keuren en dus gelegenheid hebben zich op elkaar in te stellen, *gemiddeld* in hun oordeel slechts geringe verschillen vertoonen.

Verder werd berekend, in hoeverre de keurmeesters bij het keuren van één monster in hun oordeel overeenkomen. Daartoe bepaalden wij de *middelbare fout*, waarmede één keurmeester of een combinatie van keurmeesters hun cijfers vaststellen.

---

1) Sinds korten tijd wordt deze puntenschaal bij de door ons bedoelde boterkeuringen niet meer toegepast. De Overheid heeft nl. bepaald, dat van 18 Februari 1943 af bij alle in Nederland te houden boterkeuringen moet worden gebruik gemaakt van het in Duitschland gebruikelijke stelsel.

Voor *een combinatie van twee keurmeesters* vonden wij het volgende:

a. het cijfer voor geur wordt vastgesteld met een middelbare fout van ongeveer 0,4 punt;
b. hetzelfde geldt voor het cijfer voor smaak;
c. de middelbare fout van het cijfer voor gehalte-en-bewerking bedraagt 0,3 à 0,35 punt;
d. de middelbare fout bij de vaststelling van het cijfer voor „totaal aantal punten" kan men op 2,8 punt stellen.

Bovenstaande cijfers mag men niet opvatten als algemeen geldende waarden. Voor elke combinatie van keurmeesters zijn deze cijfers iets verschillend. De variatie is echter betrekkelijk gering.

In het proefschrift is op de beteekenis van deze waarden uitvoerig ingegaan. Zij leeren ons o.a.:

a. hoeveel kans er is, dat de keurmeesters aan een monster een cijfer geven, dat in bepaalde mate afwijkt van het „ware" cijfer (d.i. het cijfer, dat de ideale keurmeester aan het monster zou toekennen);
b. hoeveel kans aanwezig is, dat bepaalde verschillen bij een eerste en tweede beoordeeling van *hetzelfde* monster worden gevonden;
c. welke zekerheid er bestaat, dat een monster, dat hoogere punten verkrijgt dan een ander, inderdaad beter in kwaliteit is.

Wij geven hier als voorbeeld een tabel (betrekking hebbende op vraag *c.*) voor smaak (middelbare fout 0,4 punt) en „totaal aantal punten" (middelbare fout 2,8 punt).

Kans op een werkelijk bestaand kwaliteitsverschil, als tusschen twee monsters een bepaald verschil (in punten) is geconstateerd.

Voor smaak		Voor totaal aantal punten	
Verschil tusschen twee botermonsters	Kans, dat het monster met de hoogste punten inderdaad het best van smaak is.	Verschil tusschen twee botermonsters	Kans, dat het monster met de hoogste punten inderdaad het beste van kwaliteit is.
0 punten	50,0 %	0 punten	50,0 %
½ „	81,2 %	1 punt	60,0 %
1 „	96,2 %	2 punten	69,3 %
1½ „	99,6 %	3 „	77,6 %
		4 „	84,4 %
		5 „	89,7 %
		6 „	93,5 %
		7 „	96,2 %
		8 „	97,8 %
		9 „	98,8 %
		10 „	99,4 %

Men bedenke, dat in het bovenstaande steeds is verondersteld, dat het cijfer wordt vastgesteld door een combinatie van *twee* keurmeesters.

De conclusie uit het meegedeelde (en in het proefschrift uitvoerig behandelde) is, dat een combinatie van twee keurmeesters in staat is verschillen in kwaliteit van 7 tot 8 punten (voor „totaal aantal punten") met voldoenden graad van zekerheid te constateeren. Bij kleinere verschillen heeft natuurlijk het monster met de hoogste punten ook de grootste kans het beste te zijn, maar, als men een behoorlijke zekerheid wenscht, mag men een verschil van bijv. 3 à 4 punten niet als essentieel beschouwen.

Men kan deze conclusie ook aldus uitdrukken. Een combinatie van keurmeesters kan drie tot vier groepen boter onderscheiden:

a. zeer goede en goede boter (eventueel is deze groep nog te splitsen in twee groepen: zeer goede en goede boter);

b. matige boter;

c. slechte boter.

Een verdergaande splitsing wordt bij de keuring door twee keurmeesters en één keer beoordeelen onzeker.

Het feit, dat een combinatie van twee keurmeesters 3 à 4 groepen boter kan onderscheiden, houdt niet in, dat de keurmeesters bij de keuring zouden moeten volstaan met het plaatsen van de te keuren boter in één dezer drie of vier groepen en dat men het geven van een cijfer beter achterwege zou kunnen laten. In het proefschrift wordt aangetoond, dat het aangeven van de kwaliteit door middel van cijfers, ook al is aan deze cijfers een onzekerheid van een bepaald aantal punten verbonden, toch belangrijke voordeelen biedt boven een eenvoudige groepeering in drie of vier klassen.

Behalve van de middelbare fout hebben wij voor het tot uitdrukking brengen van de nauwkeurigheid, waarmede de keurmeesters hun cijfers vaststellen, gebruik gemaakt van den *correlatiecoëfficiënt* tusschen het door twee keurmeesters toegekende cijfer en het „ware" cijfer.

De middelbare fout brengt nl. de nauwkeurigheid nog niet voldoende tot uitdrukking, daar naast de middelbare fout ook de schaalwijdte, die de keurmeesters toepassen, een rol speelt. (De mogelijkheid bestaat, dat een combinatie van twee keurmeesters met een *kleine* middelbare fout, die slechts een zeer klein traject van de schaal gebruikt, minder scherp keurt dan een andere combinatie met een *groote* middelbare fout en een groote schaalwijdte).

Wij hebben zoodoende onderscheiden:

a. de *subjectieve nauwkeurigheid*: uitgedrukt in de middelbare fout;

b. de *objectieve nauwkeurigheid*: uitgedrukt in den correlatiecoëfficiënt.

In het proefschrift toonden wij aan, dat de correlatiecoëfficiënt tusschen het cijfer, dat een combinatie van twee keurmeesters aan een monster toekent en het „ware" cijfer ongeveer 0,80 bedraagt.

Om duidelijk te maken, welke beteekenis men aan een correlatiecoëfficiënt van 0,80 moet toekennen, zijn in nevenstaande figuur twee grootheden afgebeeld, die een correlatie van 0,80 vertoonen. Bij volkomen afhankelijkheid

($r = 1$) zou het verband tusschen beide grootheden weergegeven worden door een rechte lijn. Als de correlatiecoëfficiënt kleiner dan 1 is, liggen de punten gerangschikt in ellipsen van gelijke kansdichtheid om hun gemeenschappelijk middelpunt. Hoe sterker de correlatie, hoe lang-gerekter de ellipsen. Wij hebben in de figuur de ellips geteekend, waarbinnen, bij een correlatiecoëfficiënt van 0,80, 98 % van de punten vallen.

Voor een meer uitvoerige toelichting moeten wij naar het proefschrift verwijzen.

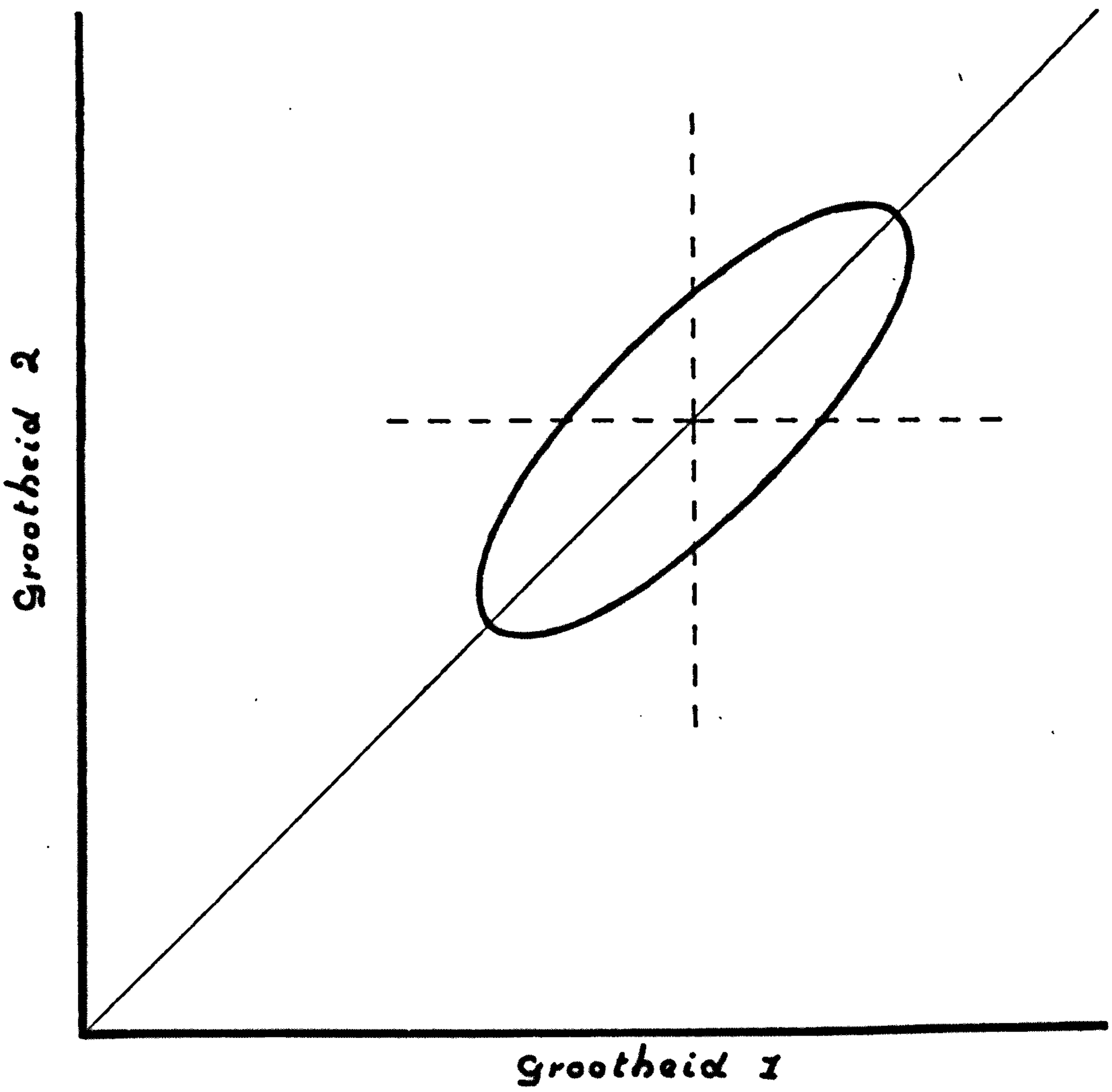

Fig. 8. Kansellips, waarbinnen bij een correlatiecoëfficiënt van 0,80 98 % der waarnemingen vallen.
(De afstand van den oorsprong van het gestippelde assenstelsel tot de snijpunten van de ellips met deze assen is gelijk aan 2,8 × de middelbare fout).

Uit de cijfers, die boven voor de middelbare fout en voor den correlatiecoëfficiënt gegeven zijn, is af te leiden, dat de keurmeesters in staat zijn vrij belangrijke verschillen tusschen botermonsters te constateeren, maar dat kleinere verschillen niet met *zekerheid* kunnen worden aangetoond.

Om de nauwkeurigheid van het vastgestelde cijfer te verhoogen, worden twee middelen aanbevolen:

1. *vergrooting van het aantal keurmeesters.*

Langs wiskundigen weg wordt aangetoond, in welke mate de subjectieve en de objectieve nauwkeurigheid door vergrooting van het aantal keurmeesters toenemen.

Op grond van deze uitkomsten wordt verhooging van het aantal keurmeesters van twee op drie tot vier aanbevolen.

2. *het scheppen van grootere verschillen in de kwaliteit der monsters.*

En wel: door de monsters te keuren na een langeren tijd van bewaring en/of door de bewaringstemperatuur te verhoogen.

Wij hebben ook enkele beschouwingen gewijd aan de herkeuring van een monster, die wordt gehouden, als het oordeel van de beide keurmeesters te veel verschilt. Onze conclusie was, dat een herkeuring, waarbij de keurmeesters onderling overleg plegen, geen zekerheid geeft, dat het eindoordeel dichter bij het ware cijfer komt te liggen. In overweging werd gegeven om bij het constateeren van te groote verschillen tusschen twee keurmeesters bij de beoordeeling van bepaalde monsters de keurmeesters zonder overleg te laten herkeuren en als eindoordeel het gemiddelde van de vier beoordeelingen te nemen. Bij keuring door drie of vier keurmeesters kan men de herkeuring gevoegelijk laten vervallen.

Bij de bespreking van de geschiktheid van de bij de hier besproken boterkeuringen gebruikte schaal werd er op gewezen, dat deze schaal voordeelen biedt boven sommige in andere landen gebruikte.

Het verdere deel van het proefschrift geeft een uitvoerige beschouwing over de mathematische grondslagen, waarvan in het voorgaande is gebruik gemaakt, terwijl ook de nauwkeurigheid van den individueelen keurmeester aan een onderzoek wordt onderworpen.

Deze onderdeelen leenen zich echter minder goed voor een korte samenvatting, zoodat wij den lezer daarvoor naar het proefschrift verwijzen.

INHOUDSOPGAVE.

STELLINGEN.

I.

Door het onderzoek van Kruisheer c.s. is niet bewezen, dat het aantal in de boter aanwezige kiemen invloed uitoefent op de kwaliteit hiervan.

C. I. Kruisheer, P. C. den Herder, W. C. Smit en A. de Haan.
Het bacteriologisch-chemisch kwaliteitsonderzoek der Nederlandsche keuringsboter, 's-Gravenhage 1940.

II.

Toepassing van de reductaseproef (vooral in de uitvoering-Wilson) kan de zuivelfabriek met eenvoudige middelen een vrij betrouwbaren indruk verschaffen over het aantal bacteriën in de aangevoerde melk.

III.

Bij de opsomming van de verschillende enzymen, die in de melk voorkomen, mag het enzym reductase niet worden genoemd.

IV.

De methode van R. A. Fisher ter beoordeeling van de overeenstemming tusschen twee bepalingen van de middelbare fout is, zelfs bij kleine waarden van den correlatiecoëfficiënt, niet bruikbaar om de betrouwbaarheid van den correlatiecoëfficiënt te toetsen.

V.

Aan een verslag van een wetenschappelijk onderzoek, waarbij statistisch materiaal is verwerkt, dient de eisch te worden gesteld, dat hierin aan een wiskundige verwerking van de resultaten voldoende aandacht is besteed.

VI.

Hoewel aan de uitvoering bezwaren verbonden zijn, is het gewenscht bij de melkproductiecontrôle meer aandacht te schenken aan de verbruikte hoeveelheid voeder.

VII.

Het is wenschelijk, dat het gebod tot standaardisatie van het vetgehalte van de consumptiemelk in de toekomst — zij het op een hooger niveau dan thans — gehandhaafd blijft.

VIII.

De classificatie van kunstmatig gedroogd gras dient niet alleen gebaseerd te zijn op het eiwitgehalte, maar tevens op het gehalte aan ruwe celstof in het product.

E. Brouwer en N. D. Dijkstra, Versl. landbk. onderz. 45(5)C, 1939, blz. 139.

IX.

Het is niet onmogelijk, dat er bij de verklaring van het karnproces en van het oproomingsproces tot nu toe te weinig aandacht is geschonken aan de electrische lading der vetbolletjes.

X.

Het staat niet vast, dat het gebruik van caseïne-agar voor de vaststelling van het aantal schadelijke bacteriën in boter voordeelen biedt boven dat van standaard-agar.

K. J. Demeter en F. X. Maier, Milchwirtsch. Forschungen 1931, blz. 418 en volgende.

XI.

Het feit, dat de boter van practisch alle fabrieken door het Z.K.B. wordt gekeurd, houdt niet in, dat de keuringen vanwege de zuivelbonden min of meer overbodig worden. Het is van groot belang, dat deze keuringen in de toekomst blijven bestaan.

XII.

Het is van groote beteekenis, dat de zuivelfabrieken de beschikking hebben over gegevens betreffende de chemische samenstelling en den bacteriologischen toestand van producten en van de hulpstoffen.

De gewestelijke organisaties van de fabrieken zijn voor het verrichten van deze onderzoekingen en voor het verstrekken van deze gegevens de aangewezen lichamen.