RIJKSLANDBOUWPROEFSTATION HOORN.

Over de betrouwbaarheid van voederproeven met melkvee

I

DOOR

E. BROUWER.

(Ingezonden 1 Maart 1929).

Reeds enkele jaren hebben wij ons bezig gehouden met de vraag, welken graad van betrouwbaarheid de voederproeven met melkvee hebben, welke aan ons instituut worden genomen. Deze vraag gaf tot het opstellen en toepassen van een aantal formules aanleiding, welke, voor zoover zij praktisch bruikbaar bleken te zijn, in dit opstel worden medegedeeld; waarmede evenwel nog niet gezegd is, dat onze methoden niet meer voor verbetering vatbaar zouden zijn. In het bijzonder werd onderzocht, òf en zoo ja welke beteekenis aan waargenomen verschillen tusschen de groepen in productie van melk en melkbestanddeelen mag worden toegekend. Bij wiskundige moeilijkheden, welke zich uit den aard der zaak herhaaldelijk voordeden, vond ik prof. M. J. VAN UVEN steeds bereid, mij bij te staan. Ik voel mij dan ook gedrongen hem daarvoor hartelijk dank te zeggen.

I. Opzet der proeven.

De proeven worden steeds in den winter volgens het groepensysteem uitgevoerd, meestal met twee groepen, elk van ongeveer 13 herfstkalvers, welke alle op tuberculine negatief reageeren. Op grond van voorafgaande waarnemingen worden bij elke proef de koeien zoodanig ingedeeld, dat met vrijwel gelijkwaardige groepen kan worden aangevangen.

In de *voorperiode* en ook in de *naperiode*, die elk ongeveer vier weken duren, ontvangen de groepen hetzelfde voedsel. In de daartusschen liggende, ongeveer acht weken durende, *hoofdperiode* (van de beide andere perioden door enkele overgangsdagen gescheiden) is de voedering verschillend: de eene groep ontvangt het controlevoeder, de andere groep het eigenlijke proefvoeder; naast proef- en controlevoeder wordt natuurlijk een grondrantsoen toegediend, dat steeds bij beide groepen gelijk is.

In de laatste jaren worden de koeien individuëel gevoederd (behalve wat het ruwvoeder betreft). In plaats van het voeder per groep af te wegen, wordt het dus per koe afgewogen, waarbij met levend gewicht en opbrengst rekening wordt gehouden. Deze methode van individuëele voedering maakt, dat de onderstaande formules beter kunnen worden toegepast en tevens is het gevolg, dat een proef niet als mislukt behoeft te worden beschouwd, wanneer één of meer dieren in het beloop der proef wegens ziekte of anderszins moeten worden verwijderd. Het experiment kan dan met de overige dieren worden voortgezet, wier voeding daarbij geen verandering ondergaat, behalve misschien wat eenig ruwvoeder aangaat. Maar dit zal, wanneer de totale hoeveelheid ruwvoeder van de betreffende groep met een passende hoeveelheid wordt verminderd, niet van veel beteekenis zijn. Het spreekt echter vanzelf, dat er naar wordt gestreefd, ook bij het ruwvoeder tot de streng individuëele methode over te gaan. Ook bij dit uitvallen van één of meer dieren kunnen de onderstaande formules worden toegepast.

Gedurende twee (niet direct op elkaar volgende) etmalen per week worden bij elke koe de melkproductie, de vetproductie en de vetvrije-droge- stof-productie vastgesteld (het laatste door berekening uit s. g. en vetpercentage). Dit onderzoek heeft bij elke koe in de voorperiode ongeveer 8 maal (dus met betrekking tot 8 etmalen) plaats, evenals in de naperiode; in de hoofdperiode ongeveer 16 maal. Aldus worden een groot aantal *„individuëele (productie)cijfers"* verkregen. Daarnaast wordt gedurende zes etmalen per week hetzelfde onderzoek in de mengmelk per groep uitgevoerd, waardoor de zoogenaamde *„groepcijfers"* worden verkregen. Deze laatste zullen in dit opstel evenwel slechts ééns ter sprake komen.

II. Formules.

Bij het beschrijven van de formules en den gang der berekening zullen wij twee stadia onderscheiden.

In het *eerste stadium* wordt, ter vereenvoudiging van het omvangrijke materiaal der individuëele productiecijfers, dit laatste vervangen door een veel kleiner stel andere cijfers, die bij iedere koe in elk der perioden de voornaamste kenmerken van het beloop der productie zoo goed mogelijk weergeven. In de plaats van de productiecijfers kunnen dus deze afgeleide cijfers treden.

In het *tweede stadium* zullen wij dan ook van deze afgeleide cijfers uitgaan om de eigenlijke vergelijking tusschen de beide groepen te treffen.

A. *Vereenvoudiging van het materiaal der individuëele productiecijfers.*

Zooals gezegd willen wij de voornaamste kenmerken van het beloop der productie bij iedere koe in elke periode in een gering aantal cijfers weergeven. In het onderstaande zij dit nader toegelicht.

Wij beschouwen b. v. de 17 cijfers uit de hoofdperiode, welke betrekking hebben op de vetopbrengst van koe n°. 49 (zie tabel I). In fig. 1 heeft men een diagram van deze tabel gemaakt. Den tijd heeft men op de horizontale as (X-as) uitgezet, de productie (Grammen vet) op de loodrechte as (Y-as). Elke stip stelt dus één waarneming voor. Stel nu het aantal etmalen, waarin de vetproductie werd bepaald, gelijk aan n; de hoeveelheid vet, afgescheiden in het i'de etmaal, noemen wij: y_i en den bijbehoorenden tijd (in etmalen): t_i.

TABEL I.

Productie van koe N°. 49 (1927—1928; Groep II) in de hoofdperiode.

N°. der waarneming.	Datum.	Tijd in etmalen (t_i).	Grammen vet (y_i).
1	24 Jan.	0	625
2	27 Jan.	3	671
3	31 Jan.	7	613
4	3 Febr.	10	618
5	7 Febr.	14	613
6	10 Febr.	17	610
7	14 Febr.	21	567
8	17 Febr.	24	551
9	21 Febr.	28	626
10	24 Febr.	31	547
11	28 Febr.	35	604
12	2 Mrt.	38	549
13	6 Mrt.	42	535
14	9 Mrt.	45	560
15	13 Mrt.	49	569
16	16 Mrt.	52	532
17	20 Mrt.	56	561
n = 17	Som Gemiddeld	$[t] = 472$; $\bar{t} = 27.8$;	$[y] = 9951$ $\bar{y} = 585.4$

Wij stellen nu allereerst belang in de gemiddelde dagelijksche vetgift: $\bar{y}$. Klaarblijkelijk is $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$; $\sum_{i=1}^{n} y_i$ wordt ook wel aldus geschreven: $[y]$; dus $\bar{y} = \frac{1}{n}[y]$. In ons geval luidt dus de uitkomst: $\bar{y} = \frac{1}{17} \times$ $\times\ 9951 = 585.4$ Gr.

Tot zoover hebben wij niets nieuws betoogd; immers het vereenvoudigen van het materiaal der productiecijfers door het berekenen van het gemiddelde per dag in elke periode is een algemeen gebruikelijke bewerking. In de plaats van de 17 cijfers is er dus één getal gekomen en daarmede stelde men zich tot nu toe algemeen tevreden.

Bij de eenvoudigste van de zoo aanstonds te bespreken methoden zullen wij laten zien, hoe men met behulp van deze gemiddelden ook tot een benadering van de betrouwbaarheid der uitkomst kan komen. Ofschoon hiermede reeds veel is gewonnen, kan men, met behulp van een meer minutieuse techniek en door het materiaal der individueele cijfers verder uit te putten, nog wel iets verder komen. Het valt ons namelijk op, dat de y's met het verstrijken van den tijd kleiner worden (zie fig. 1). Er is dus een neiging tot dalen en nu is het duidelijk, dat het evenzeer

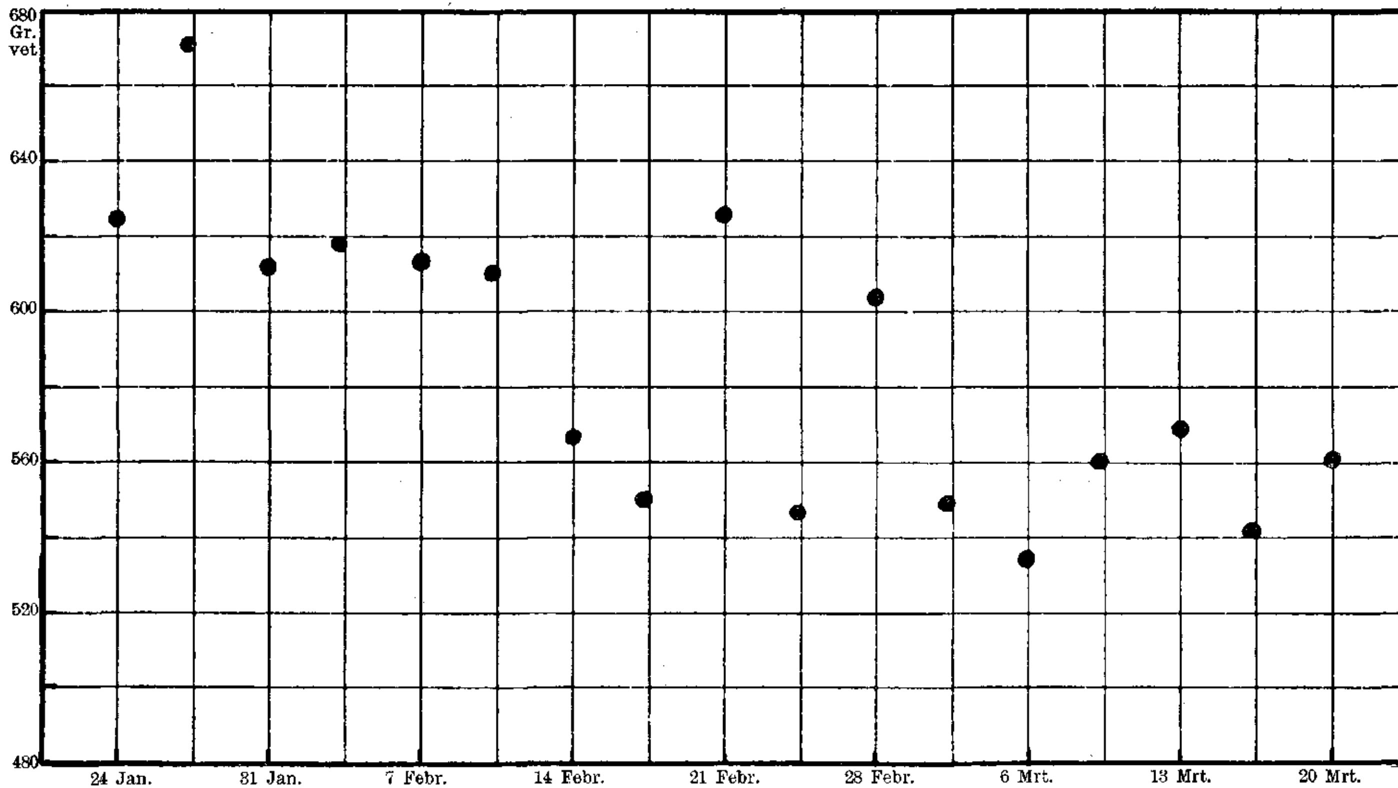

Fig. 1.
Productie (grammen vet) van koe N°. 49 (1927—'28; groep II) in de hoofdperiode.

van belang kan zijn of deze daling in een sneller dan wel in een langzamer tempo plaats vindt. Wij wenschen dus ook deze daling door een getal te kenmerken. Hiervoor zou men bijv. het verschil tusschen het eerste en laatste productiecijfer door het aantal tusschenliggende dagen kunnen deelen, waardoor de daling per dag zou zijn gevonden. Het bezwaar van deze methode is, dat aan de tusschenliggende bepalingen geen beteekenis wordt toegekend, waardoor toevallige schommelingen in de beide eindcijfers hun invloed in veel te sterke mate doen gelden. Wij wenschen daarom een methode, waarbij *alle* bepalingen tot het verkrijgen van de uitkomst het hunne bijdragen, evenzeer als dit bij de berekening van $\bar{y}$ het geval is; kortom, het doel is niet alleen een cijfer te vinden voor de *gemiddelde* dagelijksche vetgift, maar ook één voor de *gemiddelde* dagelijksche daling.

Dit laatste kan met behulp van de methode der kleinste quadraten worden verricht, zooals thans zal worden uiteengezet. Een willekeurige rechte lijn kan men voorstellen door de formule: $Y = a_1 t + a_2$. In deze formule hebben a_1 en a_2 vaste waarden. Geeft men nu aan t achtereenvolgens een aantal verschillende waarden, b.v.: 1, 2, 3, enz., dan kan men, wanneer a_1 en a_2 bekend zijn, telkens de bijbehoorende Y becijferen. Elk stel waarden: t en Y zou men in een diagram (b.v. in fig. 1) door een punt kunnen weergeven; het zal dan blijken, dat deze punten op een rechte lijn liggen. Thans denke men zich de berekening herhaald, maar met een formule, waarin de vaste waarden: a_1 en a_2 zijn vervangen door andere vaste waarden: b_1 en b_2. Na de punten te hebben uitgezet, zal men wederom een rechte lijn vinden. De stand der rechte lijnen ten opzichte van het assenkruis is echter niet gelijk. In het algemeen kan men zeggen: hoe grooter a_2 (resp. b_2) is, hoe hooger de lijn zal liggen en hoe grooter a_1 (resp. b_1) is, hoe meer de lijn zal stijgen; immers de waarde van Y neemt voor elke tijdseenheid (een etmaal) met het bedrag a_1 of b_1 toe. Wordt a_1 of b_1 negatief, dan zal de lijn dalen en wel des te meer naarmate a_1 (resp. b_1) sterker negatief is.

Het is mogelijk voor a_1 en a_2 zoodanige waarden te berekenen, dat de rechte lijn, voorgesteld door de formule: $Y = a_1 t + a_2$, zoo dicht mogelijk bij alle punten van het diagram van fig. 1 aansluit, of m. a. w.: zoo goed mogelijk midden tusschen alle punten van het diagram doorloopt. *De negatieve waarde van de aldus gevonden a_1 zullen wij aannemen als de gemiddelde daling per dag (gemiddelde dagelijksche daling).*

De waarde van a_1 kan men, evenals die van a_2, met behulp van de methode der kleinste kwadraten vinden, waarvoor de onderstaande formules moeten worden gebruikt.

$$\text{Stel: } \frac{1}{n}[t] = \bar{t}; \quad t_i = \bar{t} + u_i; \quad y_i = \bar{y} + v_i.$$

Als nu de vierkante haken weer aanduiden, dat de som over alle n waarnemingen is genomen, dan kan men dus schrijven: $[u^2] = \sum_{i=1}^{n} u_i^2$;

$$[uv] = \sum_{i=1}^{n} u_i v_i.$$

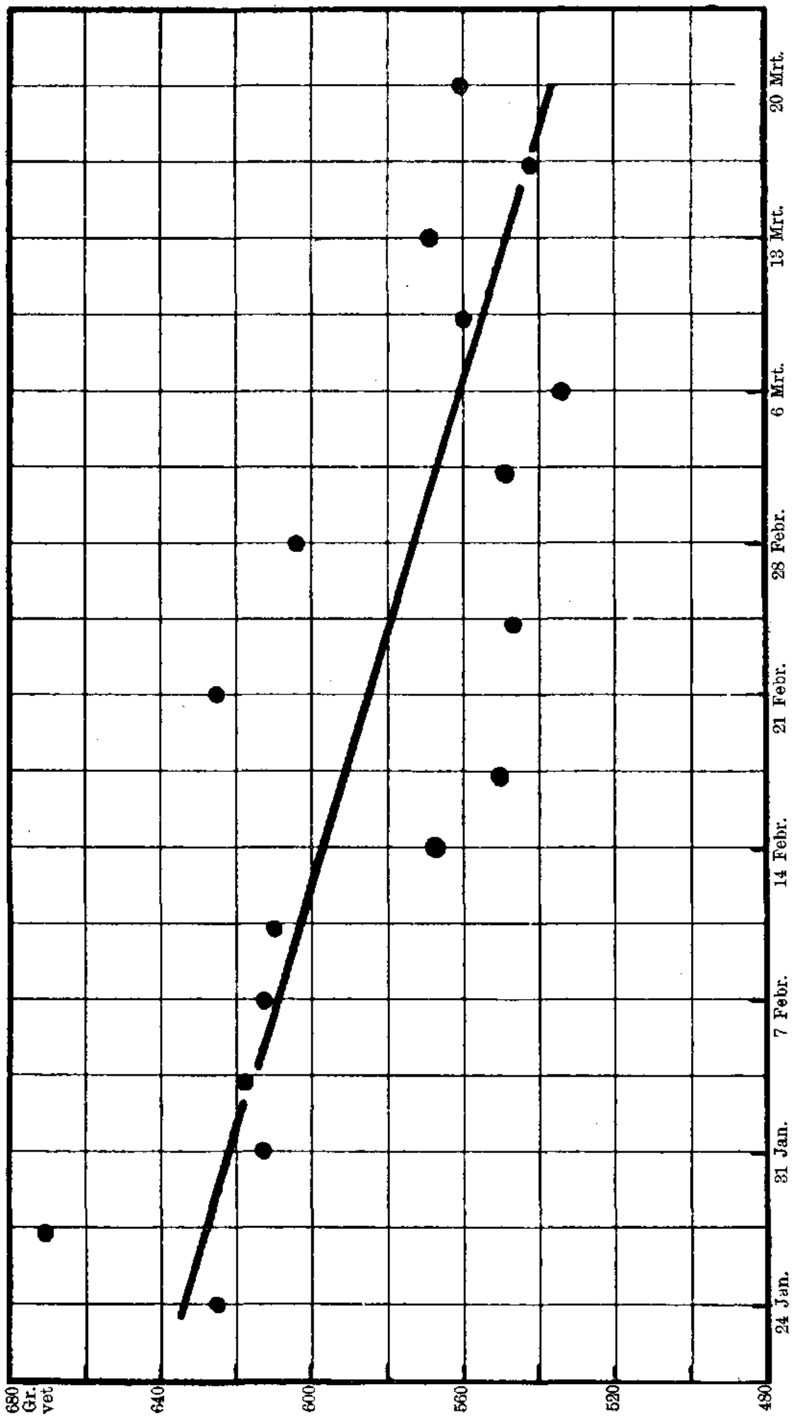

Fig. 2.
Productie (grammen vet) en regressielijn van koe N°. 49 (1927—'28; groep II) in de hoofdperiode.

Wat a_1 betreft kan men aantoonen, dat:

$$a_1 = \frac{[uv]}{[u^2]}.$$ [1])

De waarde, welke men vindt voor a_2, is, in tegenstelling met die voor a_1, afhankelijk van de plaats, waar de oorsprong van het assenkruis wordt gekozen. Kiest men den oorsprong in het punt: $P(\bar{t}, o)$ en duidt men de hierbij behoorende waarde van a_1 en a_2 aan met a'_1 en a'_2, dan zal men vinden:

$$a'_1 = \frac{[uv]}{[u^2]};\ a'_2 = \bar{y};$$

het laatste is dus weer de gemiddelde dagelijksche vetopbrengst.

In het genoemde voorbeeld (tabel I) werd gevonden:

$$[u^2] = 4999 \qquad [y] = 9951$$
$$[uv] = -8602 \qquad n = 17.$$

Dus: $-a'_1 = -\frac{[uv]}{[u^2]} = 1.72;\ a'_2 = \bar{y} = \frac{1}{n}[y] = 585.4.$

De gemiddelde productie is dus: 585.4 Gr. en de gemiddelde daling: 1.72 Gr. per dag. De vergelijking van de gevraagde lijn is:

$$Y = -1.72\,u + 585.4 \text{ of: } Y = -1.72\,(t - \bar{t}) + 585.4.$$

In fig. 2 is de bedoelde lijn getrokken en men ziet, dat zij inderdaad midden tusschen de punten door loopt. Dergelijke lijnen worden wel regressielijnen genoemd.

Soms komt het voor, dat de daling in één periode aanvankelijk vlugger, later langzamer verloopt of omgekeerd; ook kan men soms waarnemen, dat de reeks der productiecijfers een onmiskenbare golving vertoont. Ook dit alles kan zoo noodig, zij het ten koste van veel rekenwerk, in cijfers worden uitgedrukt; maar zulks zal wel zelden de moeite loonen. In het volgende zal dan ook alleen worden gerekend met de gemiddelde productie en de gemiddelde daling van elke koe in elk der perioden; slechts van de gemiddelde daling in de naperiode zal geen gebruik worden gemaakt; daarentegen wèl van de gemiddelde opbrengst in deze periode.

B. *Eigenlijke berekening.*

1. Eenvoudige formules.

Allereerst willen wij enkele eenvoudige methoden aangeven, welke het voordeel hebben weinig wiskundige kennis te vereischen. Bovendien wordt hierbij alléén gebruik gemaakt van de *gemiddelde opbrengst* van elke koe in de verschillende perioden, daarentegen niet (uitgezonderd

[1]) Hoe men tot deze en de eerstvolgende formules kan komen, zal naderhand bij de bespreking van de meer saamgestelde methoden worden aangegeven; a_1, a_2, u en v hebben daar evenwel een andere beteekenis. Zie ook: E. BROUWER, Ned. Tijdschr. v. Geneesk., Bd. 72, 1928, I, bldz. 1454.

bij één formule) van de gemiddelde daling, wier berekening aanmerkelijk méér tijd vordert dan de becijfering der gemiddelde productie.

a. Men laat de gegevens uit vóór- en naperiode onbenut.

Wij noemen z_k: de gemiddelde dagelijksche opbrengst (b.v. vetopbrengst) van de koe met het nummer k in de hoofdperiode. De gemiddelde opbrengst in dezelfde periode van de groep, waartoe koe k behoort, noemen wij $\bar{z}$. Verder stellen wij: $z_k = \bar{z} + w_k$. Is m het aantal dieren in deze groep, dan heeft men dus:

$$\bar{z} = \frac{1}{m} \sum_{k=1}^{m} z_k = \frac{1}{m} [z].$$

De middelbare afwijking: σ_z van één koe vindt men aldus:

$$\sigma_z^2 = \frac{[(z-\bar{z})^2]}{m-1} = \frac{[w^2]}{m-1}.$$

De middelbare afw. van het groepgemiddelde: $\sigma_{\bar{z}}$ volgt gemakkelijk uit de formule:

$$\sigma_{\bar{z}}^2 = \frac{[w^2]}{m(m-1)}.$$

Heeft men twee groepen: I en II, dan vindt men dus voor het verschil in opbrengst ten gunste van groep I, welk verschil $D_{\bar{z}}$ zal worden genoemd:

$$D_{\bar{z}} = \bar{z}_{I} - \bar{z}_{II} \pm \sqrt{\sigma_{I\bar{z}}^2 + \sigma_{II\bar{z}}^2}, \qquad (1)$$

waarbij de indices I en II aanduiden of men met groep I dan wel met groep II te maken heeft.

Bij het toepassen bleek, dat deze formule een zeer onvoldoende uitkomst geeft in dien zin, dat voor de m. afw. in (1) een cijfer wordt gevonden, dat veel te groot is; de formule bevat dus slechts een klein gedeelte van datgene, wat de proef ons omtrent het ware verschil D kan leeren.

b. De voorperiode wordt in rekening gebracht.

Symbolen: z en w blijven dezelfde beteekenis behouden. Verder stellen wij:

x_k = gemiddelde dagelijksche productie van koe k in de voorperiode.

Bereken nu voor elke koe: $x_k - z_k$, die wij f_k zullen noemen; dit is dus de daling van koe k tusschen de middens van vóór- en hoofdperiode (dus niet de *dagelijksche* daling).

Vervolgens becijfert men voor elke groep: $\bar{f} = \frac{1}{m} [f]$; d. i. de gemiddelde daling der groep. Stel nu $f_k = \bar{f} + \omega_k$, dan volgt voor de m. afw. van $\bar{f}$:

$$\sigma_{\bar{f}}^2 = \frac{[\omega^2]}{m(m-1)}.$$

Bij twee groepen: I en II, vindt men voor het verschil $D_{\bar{f}}$ ten gunste van groep I:

$$D_{\bar{f}} = \bar{f}_{II} - \bar{f}_{I} \pm \sqrt{\sigma^2_{I\bar{f}} + \sigma^2_{II\bar{f}}} \quad \dots\dots \quad (2)$$

Deze formule bleek het resultaat der proef tamelijk goed weer te geven.

c. Zoowel voorperiode als naperiode worden in rekening gebracht.

Symbolen: Naast x_k en z_k voeren wij in:

y_k = gemiddelde dagelijksche productie van koe k in de naperiode.

Thans berekenen wij voor elke koe: $z_k - \frac{1}{2}(x_k + y_k)$, welke waarde wij g_k zullen noemen. Verder stellen wij $\bar{g} = \frac{1}{m}[g]$.

Geheel op dezelfde wijze als boven is aangegeven, kan men weer bij elke groep de m. afw. van $\bar{g}$ berekenen en vervolgens het verschil: $D_{\bar{g}}$ ten gunste van groep I:

$$D_{\bar{g}} = \bar{g}_{I} - \bar{g}_{II} \pm \sqrt{\sigma^2_{I\bar{g}} + \sigma^2_{II\bar{g}}} \quad \dots\dots \quad (3)$$

Bij het toepassen van deze laatste formule diene men goed in het oog te houden, dat hierbij wordt verondersteld, dat een *nawerking* van het voedsel zich niet in de naperiode doet gelden. Is dit daarentegen wèl het geval, dan wordt D te groot of te klein gevonden, hetgeen de interpretatie van de uitkomst moeilijk kan maken. Toch loont het zeker de moeite $D_{\bar{g}}$ en zijn m. afw. te becijferen; in de onderzochte gevallen bleek de m. afw. van $D_{\bar{g}}$ aanmerkelijk kleiner te zijn dan die van $D_{\bar{f}}$.

Verder merken wij volledigheidshalve op, dat $D_{\bar{f}}$ en $D_{\bar{g}}$ niet de verschillen tusschen de opbrengstcijfers der hoofdperiode voorstellen, maar gecorrigeerde verschillen zijn, althans indien de groepen in de voorperiode en eventueel in de naperiode niet geheel gelijkwaardig zijn.

Zooals reeds werd opgemerkt, wordt bij onze proeven zes etmalen per week de melk van elke groep onderzocht: aldus ontstaan de „groepcijfers". Ook hieruit kan men voor elke groep en in elke periode een gemiddelde berekenen, dat vrijwel met $\bar{x}$, resp. $\bar{y}$ en $\bar{z}$ moet overeenkomen, maar dat het voordeel heeft iets nauwkeuriger te zijn dan deze laatste, omdat deze slechts op twee etmalen per week betrekking hebben. Het is daarom een klein voordeel om in de bovenstaande formules $\bar{z}_I - \bar{z}_{II}$, $\bar{f}_{II} - \bar{f}_I$ en $\bar{g}_I - \bar{g}_{II}$ te vervangen door overeenkomstige getallen, berekend met behulp van de groepcijfers; de middelb. afwijkingen kunnen natuurlijk slechts uit de individuëele cijfers worden berekend en deze middelb. afwijkingen willen wij onveranderd laten [1]).

De vorenstaande formules zijn zóó eenvoudig, dat zij als zoodanig zeker niet door voorbeelden behoeven te worden toegelicht. Om voorloopig eenigen indruk te geven van de nauwkeurigheid, welke met behulp ervan kan worden bereikt, is hieronder de m. afw. van D vermeld bij één onzer proeven (1927-'28; in totaal 24 koeien); het voorbeeld heeft betrekking op de vetproductie; de eenheden zijn grammen. In een volgend artikel gaan wij hierop dieper in.

Volgens methode *a*, formule (1), werd gevonden:

$$\sqrt{\sigma^2_{I\bar{z}} + \sigma^2_{II\bar{z}}} = 35{,}7.$$

Volgens methode *b*, formule (2), werd gevonden:

$$\sqrt{\sigma^2_{I\bar{f}} + \sigma^2_{II\bar{f}}} = 11{,}7.$$

Volgens methode *c*, formule (3), werd gevonden:

$$\sqrt{\sigma^2_{I\bar{g}} + \sigma^2_{II\bar{g}}} = 7{,}9.$$

Gelijk te verwachten was, geeft de methode *a*, waarbij van de voorperiode geen gebruik werd gemaakt, een zeer slecht resultaat. Zou men er dan ook toe overgaan dergelijke proeven zonder voorperiode op te zetten, dan zou tusschen de beide groepen een verschil van ten minste $2 \times 35{,}7$ Gr. vet, overeenkomende met meer dan 2 K.G. melk per koe en per dag, moeten worden waargenomen om het met zekerheid aan de verschillende voedering te mogen toeschrijven.

Het inschakelen van een *voorperiode* is van oudsher een maatregel geweest om deze proeven betrouwbaarder te maken. Wij kunnen thans zien, dat deze betrouwbaarheid daardoor inderdaad veel grooter wordt (methode *b*); bij de onderhavige proef wordt de m. afw. zelfs tot $^1/_3$ teruggebracht. Dit wil zeggen, dat thans reeds een verschil tusschen de beide

---

[1]) Het is mogelijk voor de m. afw. een iets kleinere waarde te becijferen, overeenkomstig de grootere nauwkeurigheid der groepcijfers en wel door op de als boven berekende m. afw. een correctie aan te brengen. Voorloopige becijferingen leerden, dat deze correctie niet bijzonder groot kan zijn. Ik voeg hieraan toe, dat het *niet* geoorloofd is als volgt te redeneeren: De individueele cijfers hebben betrekking op twee, de groepcijfers op zes, d.i. drie maal zooveel, etmalen per week: de als boven berekende m. afw. mag dus door $\sqrt{3}$ worden gedeeld. De fout, welke hierbij wordt gemaakt, is deze, dat men aanneemt, dat de productiecijfers van één koe onafhankelijk van elkaar zijn. Dit is niet het geval; men weet immers, dat de ééne koe dag aan dag een hooge productie heeft, de andere een lage. Men zou dit als een soort van „intraclass correlation" kunnen beschouwen.

groepen van $2 \times 11{,}7$ Gr. vet per koe en per dag wezenlijk kan worden genoemd en dat men in het geval *a* negen maal zooveel koeien zou noodig hebben om eenzelfden graad van nauwkeurigheid te bereiken.

Evenzeer is van oudsher het gedrag der koeien in de *naperiode* van belang geacht voor het trekken van een conclusie. Inderdaad zien wij thans, dat de m. afw. bij deze proef niet onaanzienlijk kleiner wordt, indien behalve de voorperiode ook de naperiode in rekening wordt gebracht. Wanneer men aanneemt, dat een nawerking mag worden verwaarloosd, is een verschil van $2 \times 7{,}9$ Gr. vet of ongeveer 0,5 K.G. melk reeds van wezenlijke beteekenis. Zonder met vóór- en naperiode rekening te houden zou men dus ongeveer $\frac{35{,}7^2}{7{,}9^2} \times 24$, dus $\pm$ 500 koeien noodig hebben om tot dezelfde nauwkeurigheid te geraken als thans met 24.

Al mag men aan deze cijfers niet veel waarde hechten, omdat zij slechts op één proef betrekking hebben, er blijkt wel voldoende uit, dat de strekking van deze berekeningen iets verder gaat dan alleen het leeren kennen van de einduitkomst; men ervaart er óók door, welke invloed er uit gaat van de verschillende maatregelen, welke men heeft genomen om de proeven zoo betrouwbaar mogelijk te maken, hetgeen voor een juist inzicht, en eventueel ook voor het aanbrengen van verdere verbeteringen in den opzet, van veel beteekenis kan zijn.

Thans nog de *gemiddelde daling.* Hiervoor kan de formule onder *a* worden gebruikt. Noemen wij de gemiddelde dagelijksche daling in de hoofdperiode bij koe *k*: ζ_k, dan vindt men weer:

$$D_{\bar{\zeta}} = \bar{\zeta}_I - \bar{\zeta}_{II} \pm \sqrt{\sigma^2_{I\bar{\zeta}} + \sigma^2_{II\bar{\zeta}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Ofschoon de bouw van deze formule niet verschilt van (1), zal blijken, dat het in rekening brengen van vóór- en naperiode hier slechts weinig voordeel brengt; in sommige gevallen is dit zelfs *na*deelig gebleken. Vermoedelijk zal men bij de beschouwing van ζ voor practische doeleinden wel steeds of bijna steeds met deze formule (4) kunnen volstaan; zij bevat dus bijna alles, wat de proef ons omtrent $D_{\bar{\zeta}}$ kan leeren.

De bovenstaande formules zijn ons reeds herhaaldelijk van nut geweest om de betrouwbaarheid van onze uitkomsten te schatten, hetgeen op deze wijze betrekkelijk weinig rekenwerk vordert. Vermoedelijk kan men nog een ietwat grootere nauwkeurigheid bereiken, indien f_k en g_k worden uitgedrukt in procenten van x_k; een enkele maal heb ik deze handelwijze gevolgd en inderdaad met eenig resultaat. Toch lijkt het mij wenschelijker in twijfelachtige gevallen, wanneer het er op aankomt het cijfermateriaal zoo goed mogelijk uit te putten, de thans volgende methoden in toepassing te brengen.

2. Meer samengestelde formules.

d. Formules met twee veranderlijken.

Overeenkomstig het algemeene gebruik zullen wij deze veranderlijken x en y noemen. Om onze gedachten te bepalen stellen wij:

x_k: gemiddelde dagelijksche productie (b.v. vetproductie) van koe k in de voorperiode.

In tegenstelling met het voorgaande noemen wij thans y_k: de gemiddelde dagelijksche productie van dezelfde stof, als waarop x betrekking heeft, bij koe k in de hoofdperiode.

Verder stellen wij voor elke groep: $x_k = \bar{x} + u_k$; $y_k = \bar{y} + v_k$, . (5)
waarin $\bar{x}$ en $\bar{y}$ weer groepgemiddelden voorstellen.

Is de proef zóóver gevorderd, dat de voorperiode is afgeloopen, dan kan men zich denken, dat x_k wordt uitgezet op de X-as van een diagram. Daarna komt de hoofdperiode; de gemiddelde productie van koe k in deze periode (y_k) zal dus na afloop der hoofdperiode op de Y-as worden uitgezet. Had men een zeer groot aantal (μ) koeien, alle met precies dezelfde x_k, dan zouden deze dieren in de hoofdperiode toch niet alle eenzelfde y_k opleveren, maar de y_k's zouden een zekere groepeering vertoonen ten opzichte van een gemiddelde, dat wij Y_k zullen noemen. Wij nemen aan, dat deze groepeering zoogenaamd „normaal" is. De m. afw. van één koe (σ_k) is de volgende:

$$\sigma_k^2 = \frac{[(y_k - Y_k)^2]}{\mu},$$

waarbij de vierkante haken aanduiden, dat gesummeerd wordt over alle μ koeien.

Bij een willekeurige andere abscis: x_l vinden wij een ander gemiddelde: Y_l. Van de individuëele waarden: y_l nemen wij weer aan, dat zij een normale verdeeling vertoonen. Verder nemen wij aan: $\sigma_k = \sigma_l$. Voor elke abscis nemen wij dus aan, dat de bij deze abscis behoorende y's, mits in een zeer groot aantal aanwezig, een normale verdeeling met constante (d. w. z. van x onafhankelijke) standaardafwijking vertoonen. Tenslotte veronderstellen wij nog, dat de verschillende Y's, dus Y_k, Y_l, enz., op één rechte lijn liggen; of anders gezegd, wij nemen aan, dat praktisch voldaan is aan:

$$Y = a_1 x + a_2 \text{ [1]} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Een dergelijke lijn (al of niet recht) noemt men weer *regressielijn*.

Men ziet, dat het noodig is aan het materiaal zekere eigenschappen toe te kennen. Vanzelfsprekend stellen wij de zaken daardoor eenvoudiger voor dan zij in werkelijkheid zijn; zulks mag nimmer worden vergeten. Het kiezen van deze onderstellingen is een quaestie van ervaring en physiologisch inzicht eenerzijds. Anderzijds is het zaak telkens weer aan de gegevens te controleeren of de onderstellingen tot tegenstrijdigheden voeren. De uitkomsten van de laatste drie proefjaren werden op de

[1] Y, a_1 en a_2 hebben natuurlijk een geheel andere beteekenis dan hiervóór in het hoofdstuk over de vereenvoudiging der productiecijfers.

vermelde eigenschappen nagegaan; tegenstrijdigheden traden daarbij niet aan het licht. Voor een in bijzonderheden tredende contrôle is ons cijfermateriaal echter nog niet groot genoeg, zoodat het over enkele jaren misschien toch nog noodig kan blijken, dat kleine wijzigingen in de formules moeten worden aangebracht.

Thans beschouwen wij *twee* groepen. Allereerst twee denkbeeldige groepen, elk bestaande uit een zeer groot (eigenlijk oneindig groot) aantal individuën, welke individuën in een voorperiode (onder precies gelijke omstandigheden) elk per dag precies x_k vet hebben geproduceerd. Zijn ook in de hoofdperiode de omstandigheden van voeding, verpleging, enz. voor beide groepen precies gelijk, dan is het duidelijk, dat, ondanks de individuëele verschillen in y_k, het groepgemiddelde van groep I precies gelijk is aan dat van groep II; dus $Y_{Ik} = Y_{IIk}$. Wij wijzen er nogmaals nadrukkelijk op, dat dit alleen dàn het geval is, wanneer het aantal koeien per groep zeer groot is. Is dit aantal klein, dan zullen de individuëele verschillen hun invloed op het gemiddelde doen gelden, zoodat in dàt geval wèl verschillen tusschen de groepgemiddelden aan den dag kunnen treden, ook wanneer voeding, verpleging, enz. gelijk zijn.

Voorloopig vasthoudende aan een zeer groot aantal koeien per groep denken wij ons thans het geval, dat, bij gelijk blijven van alle andere factoren, groep I het eigenlijke proefvoeder (in de hoofdperiode) ontvangt, groep II het contrôlevoeder, zoodat thans Y_{Ik} niet meer gelijk behoeft te zijn aan Y_{IIk}. Het eigenlijke doel van de proef is nu vast te stellen, hoeveel de eene groep méér of minder produceert dan de andere. Dat verschil noemen wij D_k; dus $D_k = Y_{Ik} - Y_{IIk}$.

De bovenvermelde proefopstelling is ideaal, maar helaas niet bereikbaar, want het is niet mogelijk telkens twee groepen van een groot aantal koeien tegenover elkaar te plaatsen, waarvan alle individuën in een voorperiode precies dezelfde opbrengst hebben gehad. Wij kunnen slechts tegenover elkaar plaatsen twee groepen van individuën, die in de voorperiode bijna alle een onderling *verschillende* productie hebben gehad. Ten hoogste kunnen wij ervoor zorgen, dat het gemiddelde over alle koeien (groepgemiddelde) van groep I vrijwel gelijk is aan dat van groep II.

In de tweede plaats ondervinden wij het bezwaar, dat het *aantal* koeien klein is, zoodat de individuëele verschillen *niet* geheel worden vereffend. Het gevolg daarvan is, dat zelfs wanneer de groepgemiddelden in de voorperiode gelijk zijn, het productieverschil in de hoofdperiode grooter of kleiner kan uitvallen al naar de toevallige samenstelling der groepen.

Er is dan ook geen denken aan, dat het „ideale" verschil in productie door ons volkomen juist kan worden vastgesteld. Steeds blijft er een zekere graad van onzekerheid over. Het is echter ook hier mogelijk de meest waarschijnlijke waarde van het verschil te berekenen en tevens den graad van onzekerheid aan te geven. Hiertoe kan men eerst van de a's, voorkomende in de achtereenvolgens op groep I en II betrekking hebbende formules:

$$Y_I = a_{I1}\, x + a_{I2} \text{ en } Y_{II} = a_{II1}\, x + a_{II2},$$

de waarschijnlijkste waarden en hun middelbare afwijkingen berekenen. Is dit geschied, dan kent men bij elke willekeurige waarde x de meest waarschijnlijke waarden van Y_I en Y_{II} en tevens kan men een uitdrukking opstellen voor hun middelbare afwijkingen. Daarna is het eveneens gemakkelijk bij elke x de bijbehoorende D uit de formule $D = Y_I - Y_{II}$ te berekenen en de m. afw. van D te bepalen. Om een eventueel verwijt te ontgaan merken wij terloops op, dat de beide regressielijnen Y_I en Y_{II} op bovenstaande wijze iets te vlak worden gevonden (a_1 dus iets te klein), doordat ook de x_k's zelf, als individueele gemiddelden beschouwd, een kleine afwijking hebben. Bij nader onderzoek bleek, dat deze quaestie kon worden verwaarloosd.

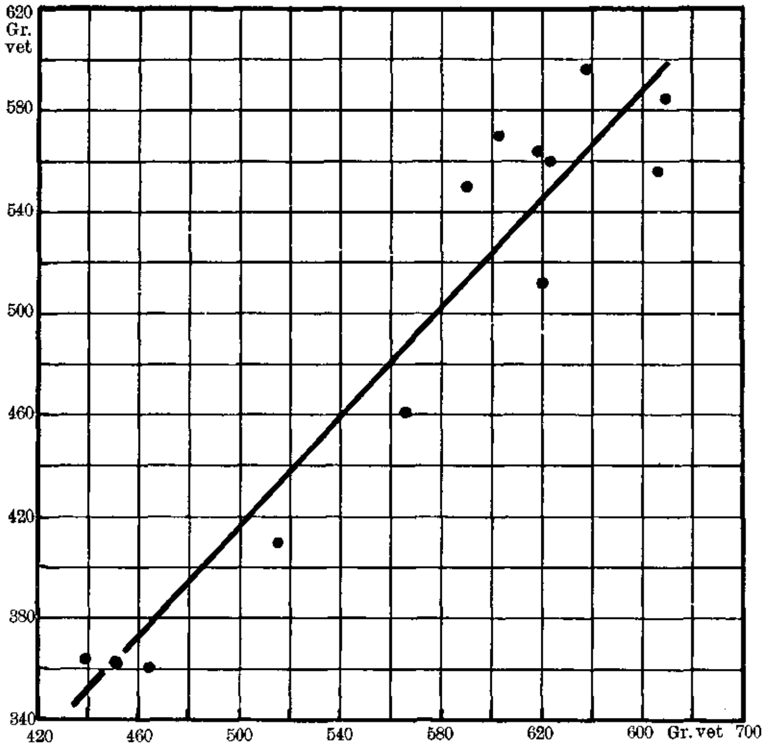

Fig. 3.
Gemiddelde dagelijksche producties (grammen vet) en regressielijn bij groep II (1927—'28).
Abscis: gemiddelde productie van elke koe in de voorperiode; ordinaat: gemiddelde productie van elke koe in de hoofdperiode.

Het berekenen der a's zullen wij nader toelichten. In fig. 3 zijn x_k en y_k van een bepaalde groep (groep II; 1927—'28; 13 koeien) in een diagram uitgezet. a_1 en a_2 kunnen nu weer met behulp van de methode der kleinste quadraten worden gevonden.

Om eenvoudige formules te krijgen gebruiken wij niet de bovenstaande formule (6), maar een ietwat gewijzigde welke ontstaat door den oorsprong tijdelijk in het punt P ($\bar{x}$, $\bar{y}$) te denken, waardoor de coördinaten volgens (5) overgaan in u en v. De bij den nieuwen oorsprong behoorende coëfficienten duiden wij aan met a'_1 en a'_2, zoodat wij (onder weglating van de indices I en II) voor elke groep een formule krijgen van den vorm:

$$V = a'_1 u + a'_2.$$

De normaalvergelijkingen [1]) zijn:

$$[u^2]\, a'_1 + [u]\, a'_2 = [uv]$$
$$[u]\, a'_1 + [1]\, a'_2 = [v].$$

Evenals vroeger duiden de vierkante haken aan, dat over alle m koeien van een groep is gesummeerd; dus ook $[1] = m$.

Men kan nu de onbekenden: a'_1 en a'_2 met behulp van voorloopig onbekende coëfficienten: Q uitdrukken in $[uv]$ en $[v]$, die beide bekend zijn en vindt aldus:

$$a'_1 = Q_{11}\, [uv] + Q_{12}\, [v]$$
$$a'_2 = Q_{21}\, [uv] + Q_{22}\, [v].$$

---

[1]) Deze normaalvergelijkingen volgen uit de voorwaarde, dat de onderstaande quadratensom, die wij S noemen, zoo klein mogelijk moet worden gemaakt door voor de onbekenden: a'_1 en a'_2 de meest doelmatige waarden te berekenen; u en v zijn de waarnemingen en zijn dus bekend.

$$S = \sum_{k=1}^{m} (v_k - V_k)^2 = \sum_{k=1}^{m} (v_k - a'_1 u_k - a'_2)^2 = [(v - a'_1 u - a'_2)^2]$$

De V_k's moeten dus zoo weinig mogelijk van de v_k's verschillen; of anders gezegd: de te berekenen lijn moet zoo goed mogelijk midden tusschen de punten van fig. 3 door loopen.

Uit de minimumvoorwaarde volgen de normaalvergelijkingen direkt door differentiatie naar a'_1 en a'_2. Langs elementairen weg kan men a'_1 en a'_2 als volgt vinden:

$$S = [(v - a'_1 u - a'_2)^2] = [v^2] + a'^2_1 [u^2] + m\, a'^2_2 - 2\, a'_1\, [u\,v] - 2\, a'_2\, [v] + 2 a'_1\, a'_2\, [u],$$

of, daar (zooals uit (5) volgt) $[u] = 0$ en $[v] = 0$:

$$S = a'^2_1\, [u^2] - 2\, a'_1\, [u\,v] + m\, a'^2_2 + [v^2].$$

Het tweede lid wordt in vierkanten gesplitst:

$$S = \frac{1}{[u^2]} (a'_1\, [u^2] - [u\,v])^2 + m\, a'^2_2 + [v^2] - \frac{[u\,v]^2}{[u^2]}.$$

Hierin is: $\frac{1}{[u^2]} > 0$ en $m > 0$. De beide onbekenden bevinden zich in de eerste twee termen van het tweede lid; de laatste twee termen zijn vrij van a'_1 en a'_2 en dus onveranderlijk. Wil men nu S zoo klein mogelijk maken, dan kan men zijn beschouwingen dus ook tot de eerste twee termen beperken en het is duidelijk, dat deze niet kleiner dan nul kunnen zijn.

Stelt men dus $a'_1\, [u^2] - [u\,v] = 0$ en $a'_2 = 0$, dan is men zeker, dat S zoo klein mogelijk is voor de waarden van a'_1 en a'_2, welke aan deze vergelijkingen voldoen.

Men vindt dus: $a'_1 = \frac{[u\,v]}{[u^2]}$; $a'_2 = 0$.

Zoo aanstonds zal blijken, dat deze oplossing ook uit de normaalvergelijkingen volgt.

De getallen Q dienen afzonderlijk te worden berekend uit de zoogenaamde „gewichtsvergelijkingen”. Deze zijn:

$$[u^2]\,Q_{11} + [u]\,Q_{12} = 1$$
$$[u]\,Q_{11} + [1]\,Q_{12} = 0$$
$$[u^2]\,Q_{21} + [u]\,Q_{22} = 0$$
$$[u]\,Q_{21} + [1]\,Q_{22} = 1.$$

(Natuurlijk kan men α'_1 en α'_2 ook vinden zonder de getallen Q afzonderlijk te berekenen; de laatste zijn echter noodig om de m. afw. van α'_1 en α'_2 te leeren kennen.)

Door de keuze van den oorsprong in het punt P $(\bar{x}, \bar{y})$ krijgt de uitkomst een zeer eenvoudigen vorm, daar wegens (5): $[u] = [v] = 0$.

Dus: $$Q_{11} = \frac{1}{[u^2]};\ Q_{12} = Q_{21} = 0;\ Q_{22} = \frac{1}{m};$$
$$\underline{\alpha'_1 = \frac{[uv]}{[u^2]}}\ ^{1)};\quad \underline{\alpha'_2 = 0.}$$

Blijkbaar vindt men dus:

$$V = \alpha'_1\,u + \alpha'_2 = \alpha'_1\,u = \frac{[uv]}{[u^2]}\,u, \quad \ldots\ldots \quad (7)$$

of, teruggaande naar de oorspronkelijke coördinaten:

$$\underline{Y = \alpha'_1\,(x - \bar{x}) + \bar{y}.} \quad \ldots\ldots\ldots \quad (8)$$

Deze regressielijn, welke wij aldus op een zeer eenvoudige wijze kunnen voorstellen, werd voor de bovengenoemde groep koeien berekend en is geteekend in fig. 3.

De middelbare afwijkingen van α'_1 en α'_2 kan men op twee wijzen becijferen. In de eerste plaats kan men de in de voorperiode verkregen cijfers voor de gemiddelde dagelijksche productie bij elk individu als gegeven beschouwen en dus rekenen, dat de eigenlijke proef pas bij het begin der hoofdperiode een aanvang neemt. De verdeelingswet der abscissen kan hierbij in het midden worden gelaten. De genoemde middelbare afwijkingen worden dan op grond van de methode der kleinste quadraten met behulp van de getallen Q gevonden.

Het is ook mogelijk zich op het standpunt der correlatietheorie te plaatsen en in aanmerking te nemen, dat ook de abscissen volgens een toevalsschema zijn gerangschikt. Neemt men hierbij aan, dat deze verdeeling der abscissen „normaal” is, dan komt men toch tot dezelfde uitkomst als die, welke de methode der kleinste quadraten oplevert.

De afwijking van de berekende regressielijn bedraagt blijkbaar voor koe k:

$$v_k - V_k = v_k - \alpha'_1\,u_k.$$

---

1) Schrijft men: $\frac{[u^2]}{m} = \sigma_u^2$ en $\frac{[uv]}{m} = r\,\sigma_u\,\sigma_v$, dan blijkt: $\alpha'_1 = r\frac{\sigma_v}{\sigma_u}$; dit is de formule voor de „regressiecoëfficiënt” uit de leerboeken der statistiek; r is de „correlatiecoëfficiënt”.

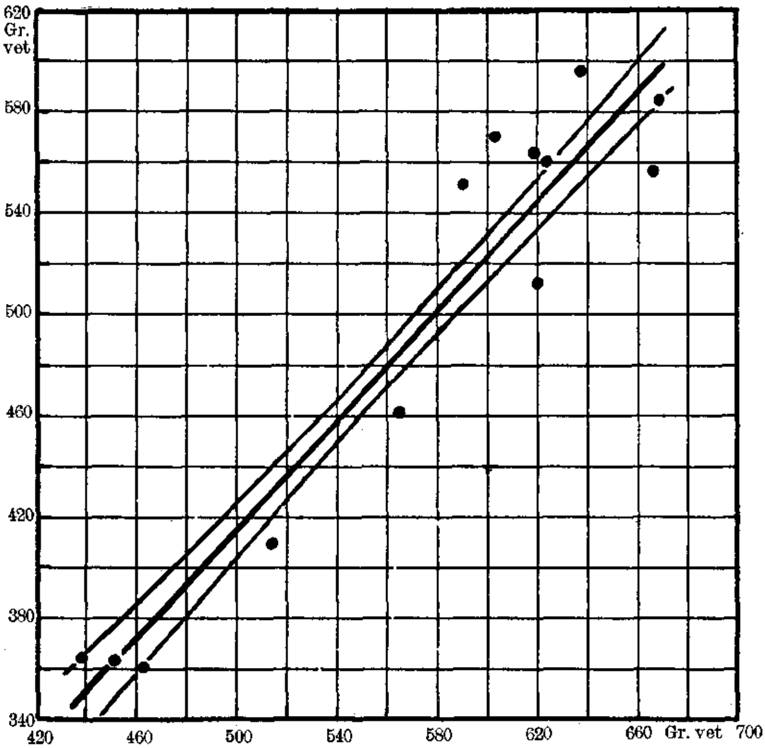

Fig. 4.
Middelbare afwijking van de regressielijn bij groep II
(1927—'28; vetproductie).

Brengt men deze uitdrukking in het kwadraat en middelt men (met noemer: $m - 2$) daarna over alle waarnemingen, dan volgt de m. afw. σ van één koe uit de onderstaande formules:

$$\sigma^2 = \frac{1}{m-2}[(v - a_1' u)^2] = \frac{1}{m-2}\{[v^2] - 2\, a_1' [uv] + a_1'^2 [u^2]\} =$$
$$\frac{1}{m-2}\{[v^2] - a_1' [uv]\}\ ^{1)}.$$

1) Hiervoor kan men ook schrijven: $\sigma^2 = \frac{1}{m-2} \frac{\begin{vmatrix} [u^2] & [uv] \\ [uv] & [v^2] \end{vmatrix}}{[u^2]}$.

De correlatietheorie schrijft deze formule meestal als volgt: Men stelt: $\frac{[v^2]}{m} = \sigma_v^2$; $\frac{[u^2]}{m} = \sigma_u^2$; $\frac{[uv]}{m} = r\, \sigma_u \sigma_v$; dus: $\sigma^2 = \frac{1}{m-2}\left\{[v^2] - \frac{[uv]^2}{[u^2]}\right\} = \frac{m}{m-2}\sigma_v^2 (1 - r^2)$, of bij benadering: $\sigma^2 = \sigma_v^2 (1 - r^2)$ . . . . (9)

Als m. afw. van a'_1 en a'_2 vindt men nu (voor het bewijs zij naar de gebruikelijke leerboeken verwezen):

$$\sigma^2_{a'_1} = Q_{11}\,\sigma^2 = \frac{\sigma^2}{[u^2]}\text{ [1]}; \quad \sigma^2_{a'_2} = Q_{22}\,\sigma^2 = \frac{\sigma^2}{m}.$$

Gemakshalve schrijven wij: $p_1 = \sigma^2_{a'_1}$; $p_2 = \sigma^2_{a'_2}$.

Thans de m. afw. van Y (zie (8)), welke m. afw. gelijk is aan die van V; de laatste denken wij ons in den vorm (7) (eerste twee leden). Geven wij nu aan x een bepaalde vaste waarde, dan vindt men op grond van de methode der kleinste quadraten als m. afw. van de Y, welke bij deze bepaalde x behoort:

$$\sigma^2_Y = \sigma^2_{a'_1} u^2 + \sigma^2_{a'_2} = \sigma^2_{a'_1} (x - \bar{x})^2 + \sigma^2_{a'_2} = p_1 (x - \bar{x})^2 + p_2 \text{ [2]} \quad . \quad . \quad (10)$$

Men kan dus schrijven:

$$\underline{Y = a'_1 (x - \bar{x}) + \bar{y} \pm \sqrt{p_1 (x - \bar{x})^2 + p_2}} \quad . \quad . \quad . \quad (11)$$

Aan (10) ziet men gemakkelijk, dat de m. afw. van Y afhankelijk is van de waarde, welke men aan x geeft. Zij is het kleinst voor $x = \bar{x}$; in dit geval is $\sigma^2_Y = p_2$.

σ^2_Y wordt grooter naarmate x verder van $\bar{x}$ afligt, hetzij in positieve, hetzij in negatieve richting. Zeer duidelijk blijkt dit, wanneer men in een diagram (in fig. 4 is dit voor de bovenvermelde groep geschied) de onderstaande drie lijnen uitzet.

N° 1 . . . . . $Y = a'_1 (x - \bar{x}) + \bar{y}$

N° 2 . . . . . $Y = a'_1 (x - \bar{x}) + \bar{y} + \sqrt{p_1 (x - \bar{x})^2 + p_2}$ [3]

N° 3 . . . . . $Y = a'_1 (x - \bar{x}) + \bar{y} - \sqrt{p_1 (x - \bar{x})^2 + p_2}$.

---

[1] De Engelsche statistici schrijven hiervoor: $\sigma^2_{a'_1} = \frac{1}{m} \frac{\sigma^2}{\sigma^2_u}$, of in verband met (9): $\sigma^2_{a'_1} = \frac{1}{m} \frac{\sigma^2_v (1 - r^2)}{\sigma^2_u}$.

[2] Deze formule volgt dadelijk uit (7) en (8), als men er gebruik van maakt (het volgt uit: $Q_{12} = 0$), dat a'_1 en a'_2 onafhankelijk van elkaar zijn. Plaatst men zich op het standpunt der correlatietheorie, waarbij dus ook de abscissen als onzeker worden beschouwd, dan wordt de zaak ingewikkelder. Prof. VAN UVEN was echter zoo vriendelijk om aan te toonen, dat men ook dàn tot dezelfde uitkomst geraakt.

[3] De lijnen 2 en 3 vormen samen een hyperbool. De vergelijking hiervan is:

$$(Y - \bar{y})^2 - 2\,a'_1 (Y - \bar{y})(x - \bar{x}) + (a'^2_1 - p_1)(x - \bar{x})^2 - p_2 = 0.$$

Men kan zich ervan overtuigen, dat het middelpunt ligt in het punt P $(\bar{x}, \bar{y})$ en dat de lijnen: $y = a'_1 (x - \bar{x}) + \bar{y}$ en $x = \bar{x}$ toegevoegde middellijnen zijn. De

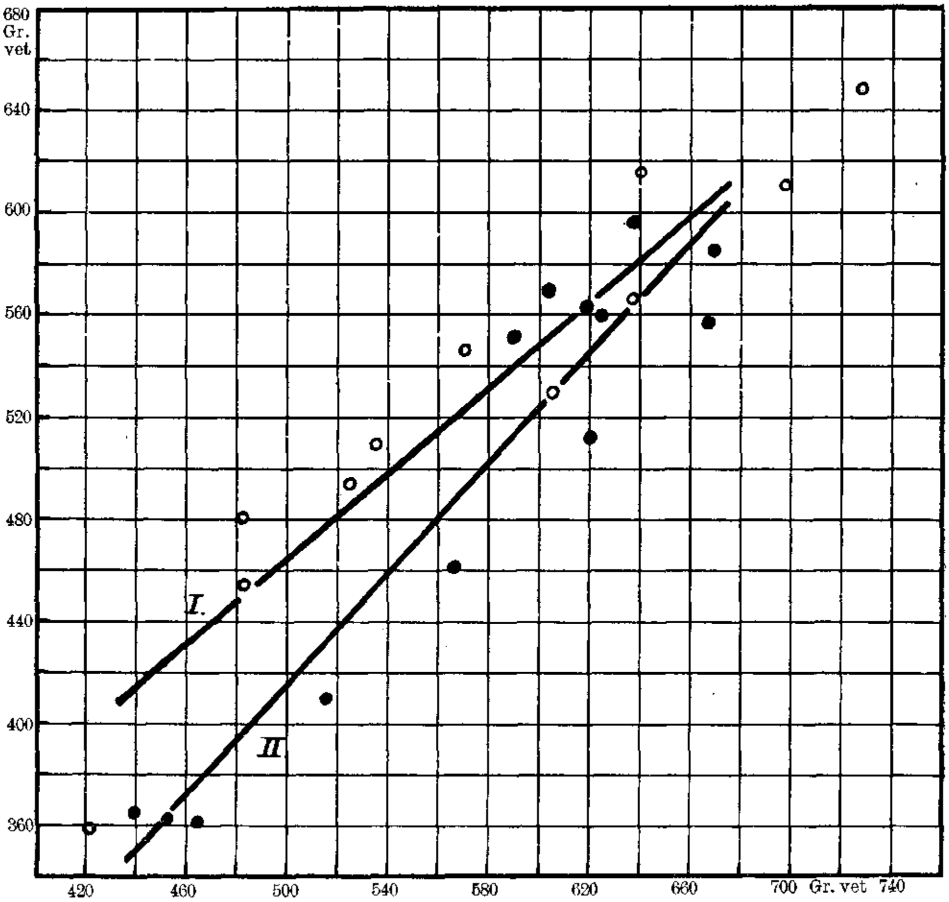

Fig. 5.

Regressielijnen van groep I (nulletjes) en groep II (punten) (1927—'28; vetproductie).

---

vergelijking van de hyperbool op deze toegevoegde middellijnen vindt men door de substitutie:

$$Y - \overline{y} = x' \sin \varphi + y'$$
$$x - \overline{x} = x' \cos \varphi,$$

waarin $tg\ \varphi = a_1'$.

Men vindt dan: $\frac{y'^2}{p_2} - \frac{x'^2}{\frac{p_2}{p_1}(1 + a_1'^2)} = 1$, waarin $\frac{p_2}{p_1} = \sigma_u^2$.

De asymptoten (rechthoekig coördinatenstelsel) zijn:

$$y - \overline{y} = (a_1' \pm \sigma_{a_1'})(x - \overline{x}).$$

Tenslotte het verschil tusschen de groepen. De formules, welke achtereenvolgens op groep I en II betrekking hebben, zie (11), schrijven wij aldus:

$$Y_I = a'_{I1}(x - \bar{x}_I) + \bar{y}_I \pm \sqrt{p_{I1}(x - \bar{x}_I)^2 + p_{I2}}$$

$$Y_{II} = a'_{II1}(x - \bar{x}_{II}) + \bar{y}_{II} \pm \sqrt{p_{II1}(x - \bar{x}_{II})^2 + p_{II2}}.$$

Voor het jaar 1927—'28 zijn deze beide regressielijnen uitgezet in fig. 5.

Voor het verschil vindt men:

$$D = Y_I - Y_{II} = a'_{I1}(x - \bar{x}_I) - a'_{II1}(x - \bar{x}_{II}) + \bar{y}_I - \bar{y}_{II} \pm$$
$$\sqrt{p_{I1}(x - \bar{x}_I)^2 + p_{II1}(x - \bar{x}_{II})^2 + p_{I2} + p_{II2}}.$$

Deze formule geeft bij elke gegeven x het verschil: D en tevens zijn m. afw. aan. Eenige vereenvoudiging krijgt men, wanneer men den oorsprong midden tusschen $\bar{x}_I$ en $\bar{x}_{II}$ neemt. Daarvoor stelt men: $x = \frac{1}{2}(\bar{x}_I + \bar{x}_{II}) + \xi$. Noemt men nu tevens: $\frac{1}{2}(\bar{x}_I - \bar{x}_{II}) = s$, dan krijgt men:

$$D = (a'_{I1} - a'_{II1})\,\xi - (a'_{I1} + a'_{II1})\,s + \bar{y}_I - \bar{y}_{II} \pm$$
$$\pm \sqrt{(p_{I1} + p_{II1})\,\xi^2 - 2\,(p_{I1} - p_{II1})\,s\,\xi + (p_{I1} + p_{II1})\,s^2 + p_{I2} + p_{II2}}.$$

Blijkbaar is D een gecorrigeerd verschil. Het werkelijk in de hoofdperiode gevonden verschil is: $\bar{y}_I - \bar{y}_{II}$. De correctieterm (afgezien van den term met ξ) is: $-(a'_{I1} + a'_{II1})\,s$; deze is nul, indien $\bar{x}_I = \bar{x}_{II}$.

Thans stellen wij:

$$\left.\begin{array}{c} a'_{I1} - a'_{II1} = b_1 \\ -(a'_{I1} + a'_{II1})\,s + \bar{y}_I - \bar{y}_{II} = b_2 \\ p_{I1} + p_{II1} = q_{11} \\ -(p_{I1} - p_{II1})\,s = q_{12} \\ (p_{I1} + p_{II1})\,s^2 + p_{I2} + p_{II2} = q_{22}. \end{array}\right\} \quad \ldots\ldots \quad (12)$$

Hierdoor vindt men D in den onderstaanden vorm:

$$\underline{D = b_1\,\xi + b_2 \pm \sqrt{q_{11}\,\xi^2 + 2\,q_{12}\,\xi + q_{22}}} \quad \ldots \quad (13)$$

Stelt men hierin: $\xi = 0$, dan volgt:

$$\underline{D_0 = b_2 \pm \sqrt{q_{22}}} \quad \ldots\ldots\ldots\ldots \quad (14)$$

Dit is dus het verschil en zijn m. afw. voor $x = \frac{1}{2}(\bar{x}_I + \bar{x}_{II})$.

Er rest nog de m. afw. van b_1:

$$\underline{\sigma^2_{b_1}} = \sigma^2_{a'_{I1}} + \sigma^2_{a'_{II1}} = p_{I1} + p_{II1} = \underline{q_{11}} \quad \ldots\ldots \quad (15)$$

Tot nu toe hebben wij bij onze proeven van de formules (14) en (15) het meest gebruik gemaakt en deze zullen voor praktische doeleinden meestal wel voldoende zijn; af en toe is evenwel ook (13) van nut gebleken.

Thans keeren wij terug naar het voorbeeld. Voor genoemde groep (groep II; 13 koeien) vond men:

$$[u^2] = 78478; \; [uv] = 84783; \; [v^2] = 100677.$$

$$\bar{x}_{II} = 573,8; \; \bar{y}_{II} = 496,5.$$

Hieruit volgt: $a'_{II\,1} = \frac{[uv]}{[u^2]} = 1,0803$

$$\sigma^2 = \frac{1}{m-2} \left\{ [v^2] - a'_{II\,1}\,[uv] \right\} = 826,0$$

$$p_{II\,1} = \frac{1}{[u^2]} \times 826,0 = 10,53 \times 10^{-3}$$

$$p_{II\,2} = \frac{1}{m} \times 826,0 = 63,5.$$

De regressielijn luidt dus:

$$Y_{II} = 1,0803\,(x - 573,8) + 496,5 \pm \sqrt{10,53 \times 10^{-3}\,(x - 573,8)^2 + 63,5}.$$

Deze lijn is uitgezet in fig. 3, 4 en 5.

Voor de tegenover deze groep geplaatste proefgroep (groep I; 11 koeien) vond men:

$[u^2] = 92357$	$\bar{x}_I = 574,5$	$a'_{I\,1} = 0,8286$
$[uv] = 76527$	$\bar{y}_I = 528,5$	$p_{I\,1} = 6,88 \times 10^{-3}$
$[v^2] = 69121$		$p_{I\,2} = 57,7.$

$$Y_I = 0.8286\,(x - 574.5) + 528.5 \pm \sqrt{6.88 \times 10^{-3}\,(x - 574.5)^2 + 57.7}.$$

De beide lijnen: Y_I en Y_{II} zijn geteekend in fig. 5.

Uit bovenstaande gegevens vindt men:

$\frac{1}{2}(\bar{x}_I + \bar{x}_{II}) = 574.15$	$q_{11} = 17.41 \times 10^{-3}$
$b_1 = -0.2517$	$q_{12} = 1.28 \times 10^{-3}$
$b_2 = 31.3$	$q_{22} = 121.2.$

Dus $D = -0.2517\,\xi + 31.3 \pm$

$$\pm \sqrt{17.41 \times 10^{-3}\,\xi^2 + 2 \times 1.28 \times 10^{-3}\,\xi + 121.2} \quad . \quad . \quad (16)$$

Stelt men hierin: $\xi = 0$, dan volgt: $D = 31.3 \pm \sqrt{121.2} = 31.3 \pm 11.0.$

Vonden wij vroeger met behulp van de methode b (formule (2)) een m. afw. van 11.7, thans vinden wij 11.0; deze laatste methode geeft dus inderdaad een betere uitkomst, maar van veel belang is dit bij deze proef niet; in andere gevallen was het voordeel grooter. Het verschil tusschen de beide groepen (31.3) is bijna drie maal zoo groot als de m. afw., zoodat wij wetenschappelijk volkomen verantwoord zijn om het verschil, althans ten deele, op rekening van de verschillende voedering te zetten.

Verder vergelijken wij $\alpha'_{\mathrm{I}\,1}$ en $\alpha'_{\mathrm{II}\,1}$, die de helling van $\mathrm{Y_I}$ en $\mathrm{Y_{II}}$ aangeven (zie fig. 5). De helling der lijnen is niet gelijk. Om nu na te gaan of hier een wezenlijk verschil is, gebruiken wij (12) en (15). Hieruit volgt:

$$b_1 = -0.2517 \pm 0.132.$$

Het verschil is bijna twee maal de m. afw., zoodat het nog niet groot genoeg is om wezenlijk te worden genoemd.

Natuurlijk behoeft deze methode niet uitsluitend te worden gebruikt om de gemiddelde *productie* (melk, vet, vetvrije droge stof) bij de groepen te vergelijken; men kan ook met y_k aanduiden de gemiddelde dagelijksche *daling* van koe k in de hoofdperiode, terwijl men de beteekenis van x_k onveranderd laat. Aldus vindt men bij beide groepen de regressie van de gemiddelde daling in de hoofdperiode t. o. v. de gemiddelde productie in de voorperiode en daarna kan men, ook wat dit punt betreft, weer het verschil tusschen I en II berekenen.

e. Formules met drie veranderlijken.

Al moge de voorafgaande werkwijze veelal betere uitkomsten opleveren dan de vroeger genoemde methode *b*, het materiaal is niet op de meest economische wijze benut, doordat nog niet alle gegevens uit de voorperiode zijn gebruikt. Brachten wij in het voorgaande alléén de gemiddelde *productie* van elke koe in de voorperiode in rekening, het is ook mogelijk om bovendien nog van de gemiddelde dagelijksche *daling* van elke koe in de voorperiode gebruik te maken.

Hierbij zullen wij de volgende symbolen benutten:

x_k noemen wij de gemiddelde dagelijksche opbrengst van koe k in de voorperiode.

Met y_k duiden wij thans (in tegenstelling met hiervóór) aan: de gemiddelde dagelijksche daling van koe k in de voorperiode.

z_k stelt voor: de gemiddelde dagelijksche productie van koe k in de hoofdperiode.

m is wederom het aantal dieren in één groep.

Men denke zich x, y en z voor elke koe in een ruimtelijk schema uitgezet. Wij noemen nu voor elke groep: $\bar{x} = \frac{1}{m}[x]$; $\bar{y} = \frac{1}{m}[y]$; $\bar{z} = \frac{1}{m}[z]$. Verder stellen wij: $x_k = \bar{x} + u_k$; $y_k = \bar{y} + v_k$; $z_k = \bar{z} + w_k$, zoodat $[u] = [v] = [w] = 0$.

Men kan nu weer met behulp van de methode der kleinste quadraten den stand van een plat vlak berekenen, dat zoo dicht mogelijk bij de punten in het bovengenoemde schema aansluit, en wel door de som van de quadraten van de afstanden tusschen punten en vlak (de afstanden gemeten in de richting der Z-as) zoo klein mogelijk te maken; aldus vindt men het zoogenaamde *regressievlak* van z t.o.v. x en y.

Men kan zich ook hier op het standpunt der correlatietheorie plaatsen. In beide gevallen wordt niet alleen hetzelfde vlak gevonden (zooals vanzelf spreekt), maar ook vindt men gelijke uitdrukkingen voor de middelbare afwijkingen der constanten.

Wij kiezen, om tot eenvoudige formules te komen, den oorsprong weer in het punt P $(\bar{x}, \bar{y}, \bar{z})$ en duiden het gevraagde regressievlak aldus aan:

$$W = a_1 u + a_2 v + a_3. \text{ [1]}$$

a_1, a_2 en a_3 zijn dus de onbekenden.

De normaalvergelijkingen hebben onderstaanden vorm:

$$\begin{aligned} [u^2]\ a_1 + [uv]\ a_2 + [u]\ a_3 &= [uw] \\ [uv]\ a_1 + [v^2]\ a_2 + [v]\ a_3 &= [vw] \\ [u]\ a_1 + [v]\ a_2 + [1]\ a_3 &= [w]. \end{aligned}$$

Men kan weer de onbekenden uitdrukken in $[uw]$, $[vw]$ en $[w]$:

$$\begin{aligned} a_1 &= Q_{11}\,[uw] + Q_{12}\,[vw] + Q_{13}\,[w] \\ a_2 &= Q_{21}\,[uw] + Q_{22}\,[vw] + Q_{23}\,[w] \\ a_3 &= Q_{31}\,[uw] + Q_{32}\,[vw] + Q_{33}\,[w]. \end{aligned}$$

Hieruit kunnen dus a_1, a_2 en a_3 direkt worden becijferd, indien de 9 getallen Q bekend zijn. Deze vindt men uit 3 stellen „gewichtsvergelijkingen", waarvan het eerste stel is:

$$\begin{aligned} [u^2]\ Q_{11} + [uv]\ Q_{12} + [u]\ Q_{13} &= 1 \\ [uv]\ Q_{11} + [v^2]\ Q_{12} + [v]\ Q_{13} &= 0 \\ [u]\ Q_{11} + [v]\ Q_{12} + [1]\ Q_{13} &= 0. \end{aligned}$$

Hieruit vindt men: Q_{11}, Q_{12} en Q_{13} en uit twee soortgelijk gebouwde stellen de overige Q's. Bedenkt men, dat $[u] = [v] = [w] = 0$, dan volgt:

$$Q_{11} = \frac{[1]\,[v^2]}{[1]\left\{[u^2]\,[v^2] - [uv]^2\right\}} = \frac{[v^2]}{[u^2]\,[v^2] - [uv]^2}$$

$$Q_{12} = Q_{21} = -\frac{[uv]}{[u^2]\,[v^2] - [uv]^2}$$

$$Q_{22} = \frac{[u^2]}{[u^2]\,[v^2] - [uv]^2}$$

$$Q_{13} = Q_{31} = Q_{23} = Q_{32} = 0$$

$$Q_{33} = \frac{1}{[1]} = \frac{1}{m}$$

$$\underline{\underline{a_1 = \frac{[uw]\,[v^2] - [vw]\,[uv]}{[u^2]\,[v^2] - [uv]^2}}}\,;\quad \underline{\underline{a_2 = \frac{[vw]\,[u^2] - [uw]\,[uv]}{[u^2]\,[v^2] - [uv]^2}}}\,;\quad \underline{\underline{a_3 = 0.}}$$

Deze oplossing voor a_1, a_2 en a_3 wordt natuurlijk ook gevonden, wanneer de getallen Q niet afzonderlijk worden becijferd. Zij zijn echter dienstig voor het berekenen van de m. afw. der onbekenden.

[1]) Deze formule speelt hier dus dezelfde rol als de formule: $V = a'_1\, u + a'_2$ in het schema voor twee veranderlijken; voor het gemak worden thans de accenten weggelaten; het spreekt vanzelf dat de getallenwaarden van a_1 en a'_1 uit deze twee formules in het algemeen niet gelijk zijn, evenmin als die van a_2 en a'_2.

De m. afw. van één koe noemen wij weer σ, welke als volgt wordt berekend:

$$\sigma^2 = \frac{1}{m-3} \sum_{k=1}^{m} (w_k - a_1 u_k - a_2 v_k)^2 =$$

$$= \frac{1}{m-3} \left\{ [w^2] + a_1^2 [u^2] + a_2^2 [v^2] - 2 a_1 [uw] - 2 a_2 [vw] + 2 a_1 a_2 [uv] \right\} =$$

$$= \frac{1}{m-3} \left\{ [w^2] - a_1 [uw] - a_2 [vw] \right\} + \frac{a_1}{m-3} \left\{ a_1 [u^2] - [uw] + a_2 [uv] \right\} +$$

$$+ \frac{a_2}{m-3} \left\{ a_2 [v^2] - [vw] + a_1 [uv] \right\}.$$

De beide laatste termen zijn gelijk aan nul, zoodat:

$$\underline{\sigma^2 = \frac{1}{m-3} \left\{ [w^2] - a_1 [uw] - a_2 [vw] \right\}} \text{ [1]}.$$

Voor de m. afw. van a_1, a_2 en a_3 kan men aantoonen, dat:

$$\underline{\sigma_{a_1}^2 = Q_{11}\,\sigma^2}; \qquad \underline{\sigma_{a_2}^2 = Q_{22}\,\sigma^2}; \qquad \underline{\sigma_{a_3}^2 = Q_{33}\,\sigma^2}.$$

Gemakshalve stellen wij:

$$p_{11} = \sigma_{a_1}^2 = Q_{11}\,\sigma^2; \quad p_{22} = \sigma_{a_2}^2 = Q_{22}\,\sigma^2; \quad p_{33} = \sigma_{a_3}^2 = Q_{33}\,\sigma^2; \quad p_{12} = Q_{12}\,\sigma^2;$$

$$p_{13} = Q_{13}\,\sigma^2 = 0; \quad p_{23} = Q_{23}\,\sigma^2 = 0.$$

De formule voor W en zijn m. afw. kan thans direkt worden opgesteld:

$$W = a_1 u + a_2 v \pm \sqrt{p_{11} u^2 + 2 p_{12} uv + p_{22} v^2 + p_{33}} \text{ [2]}, \quad . \quad . \quad (17)$$

of, teruggaande naar de oorspronkelijke coördinaten, voor groep I:

$$\underline{Z_I = a_{I1} (x - \bar{x}_I) + a_{I2} (y - \bar{y}_I) + \bar{z}_I \pm}$$

$$\underline{\pm \sqrt{p_{I11} (x - \bar{x}_I)^2 + 2 p_{I12} (x - \bar{x}_I)(y - \bar{y}_I) + p_{I22} (y - \bar{y}_I)^2 + p_{I33}}} \quad . \quad (18)$$

en overeenkomstig voor groep II.

In (18) stellen wij nu:

$x = \frac{1}{2} (\bar{x}_I + \bar{x}_{II}) + \xi$; $y = \frac{1}{2} (\bar{y}_I + \bar{y}_{II}) + \eta$; $\frac{1}{2} (\bar{x}_I - \bar{x}_{II}) = s$; $\frac{1}{2} (\bar{y}_I - \bar{y}_{II}) = t$ en berekenen het verschil tusschen de beide groepen: $D = Z_I - Z_{II}$.

Hiervoor vindt men een uitdrukking van onderstaanden vorm:

$$\underline{D = b_1 \xi + b_2 \eta + b_3 \pm}$$

$$\underline{\pm \sqrt{q_{11} \xi^2 + 2 q_{12} \xi \eta + q_{22} \eta^2 + 2 q_{13} \xi + 2 q_{23} \eta + q_{33}}} \quad . \quad . \quad (19)$$

---

[1] In determinantvorm: $\sigma^2 = \dfrac{1}{m-3} \dfrac{\begin{vmatrix} [u^2] & [uv] & [uw] \\ [uv] & [v^2] & [vw] \\ [uw] & [vw] & [w^2] \end{vmatrix}}{\begin{vmatrix} [u^2] & [uv] \\ [uv] & [v^2] \end{vmatrix}}$.

[2] De vergelijking (17) stelt voor een tweebladige hyperboloïde. Blijkbaar ligt het middelpunt in den oorsprong. $W = a_1 u + a_2 v$ is een middelvlak. Verder bewijst men gemakkelijk, dat de W-as de aan dit vlak toegevoegde middellijn is.

Hierin is: $b_1 = a_{\mathrm{I}\,1} - a_{\mathrm{II}\,1}$

$b_2 = a_{\mathrm{I}\,2} - a_{\mathrm{II}\,2}$

$b_3 = -(a_{\mathrm{I}\,1} + a_{\mathrm{II}\,1})\, s - (a_{\mathrm{I}\,2} + a_{\mathrm{II}\,2})\, t + \bar{z}_{\mathrm{I}} - \bar{z}_{\mathrm{II}}$

$q_{11} = p_{\mathrm{I}\,11} + p_{\mathrm{II}\,11}$

$q_{12} = p_{\mathrm{I}\,12} + p_{\mathrm{II}\,12}$

$q_{22} = p_{\mathrm{I}\,22} + p_{\mathrm{II}\,22}$

$q_{13} = -\left\{(p_{\mathrm{I}\,11} - p_{\mathrm{II}\,11})\, s + (p_{\mathrm{I}\,12} - p_{\mathrm{II}\,12})\, t\right\}$

$q_{23} = -\left\{(p_{\mathrm{I}\,12} - p_{\mathrm{II}\,12})\, s + (p_{\mathrm{I}\,22} - p_{\mathrm{II}\,22})\, t\right\}$

$q_{33} = (p_{\mathrm{I}\,11} + p_{\mathrm{II}\,11})\, s^2 + 2\, (p_{\mathrm{I}\,12} + p_{\mathrm{II}\,12})\, st + (p_{\mathrm{I}\,22} + p_{\mathrm{II}\,22})$
$t^2 + p_{\mathrm{I}\,33} + p_{\mathrm{II}\,33}$.

Blijkbaar is D wederom een gecorrigeerd verschil. Het in de hoofdperiode waargenomen verschil is n.l. $\bar{z}_{\mathrm{I}} - \bar{z}_{\mathrm{II}}$; de correctie is, afgezien van ξ en η: $-(a_{\mathrm{I}\,1} + a_{\mathrm{II}\,1})\, s - (a_{\mathrm{I}\,2} + a_{\mathrm{II}\,2})\, t$. Deze uitdrukking is nul, indien de groepen in de voorperiode precies gelijkwaardig zijn wat $\bar{x}$ en $\bar{y}$ betreft.

De formule (19) stelt ons weer in staat om bij elke willekeurige ξ en η, dus ook bij elke x en y, het verschil D en zijn m. afw. te berekenen. Meestal zal het voldoende zijn de formule toe te passen in het bijzondere geval: $\xi = 0$; $\eta = 0$, of ook: $x = \frac{1}{2}(\bar{x}_{\mathrm{I}} + \bar{x}_{\mathrm{II}})$; $y = \frac{1}{2}(\bar{y}_{\mathrm{I}} + \bar{y}_{\mathrm{II}})$.

Men vindt dan:

$$D_0 = b_3 \pm \sqrt{q_{33}} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (20)$$

Bovendien is het dikwijls van belang de m. afw. van b_1 en b_2 te kennen. Blijkbaar is:

$$\left.\begin{aligned} \underline{\sigma^2_{b_1}} &= \sigma^2_{a_{\mathrm{I}\,1}} + \sigma^2_{a_{\mathrm{II}\,1}} = p_{\mathrm{I}\,11} + p_{\mathrm{II}\,11} = \underline{q_{11}} \\ \underline{\sigma^2_{b_2}} &= p_{\mathrm{I}\,22} + p_{\mathrm{II}\,22} = \underline{q_{22}}. \end{aligned}\right\} \quad . \; . \; . \quad (21)$$

Van de formules (19), (20) en (21) hebben wij tot nu toe (20) en (21) het meest gebruikt.

Verder zij vermeld, dat deze formules voor verschillende doeleinden kunnen worden benut. In den aanvang werd, om de gedachten te bepalen, gezegd, dat x, y en z de volgende beteekenis zouden hebben:

x_k = gemiddelde dagelijksche productie van koe k in de voorperiode.
y_k = gemiddelde dagelijksche daling van koe k in de voorperiode. (22)
z_k = gemiddelde dagelijksche productie van koe k in de hoofdperiode.

Meermalen gebruikten wij dezelfde formules, terwijl x, y en z de volgende beteekenis hadden:

x_k = gemiddelde dagelijksche productie van koe k in de voorperiode.
y_k = gemiddelde dagelijksche productie van koe k in de naperiode. (23)
z_k = gemiddelde dagelijksche productie van koe k in de hoofdperiode.

Ook wel aldus:

x_k en y_k als in (22); maar z_k = gemiddelde dagelijksche daling van koe k in de hoofdperiode.

En ten slotte ook aldus:

x_k en y_k als in (23); maar z_k = gemiddelde dagelijksche daling van koe k in hoofdperiode.

Uit het voorgaande is wel voldoende gebleken, dat de formules met drie veranderlijken in beginsel niet van die met twee veranderlijken verschillen, zoodat het wel onnoodig is hier nog een voorbeeld uit te werken.

Kurze Zusammenfassung.

Einige einfache und mehr komplizierte Formeln wurden aufgestellt, mit deren Hilfe die Zuverlässigkeit von Fütterungsversuchen (Gruppenversuche) mit Milchvieh berechnet werden kann.

a. Sie setzen uns in Stand in den einzelnen Versuchen zu beurteilen ob ein gefundener Produktionsunterschied als wesentlich (significant) zu betrachten ist. (Möchte es sein, dasz die Versuchsgruppen in der Vorperiode (eventuell auch in der Nachperiode) nicht ganz gleichwertig sind, so werden die in der Hauptperiode gefundenen Unterschiede zugleicherzeit automatisch korrigiert).

b. Diese und ähnliche Formeln können jedoch auch bei einem mehr theoretischen Studium über die Zweckmässigkeit der Versuchsanordnung überhaupt gebraucht werden. Für die verschiedenen Masznahmen, welche bei der Anordnung dieser Versuche gewöhnlich beachtet werden, kann man nämlich berechnen in welchem Masze sie die Zuverlässigkeit des Endresultates beeinflussen.