WISKUNDIGE BESCHOUWINGEN BIJ EENIGE HOOFDSTUKKEN UIT DE PHILOSOPHIE DER LANDBOUWWETENSCHAP

J. P. Sijpkens

I.

Het streven van den Gouvernementslandbouwvoorlichtingsdienst, om bij de hernieuwde pogingen tot oprichting eener katoencultuur op Java, deze als bevolkingscultuur te propageeren, moet onjuist geacht worden.

II.

De tusschenbeplanting met Acacia decurrens in de theecultuur ter bestrijding van de nachtvorstschade verergert juist de schade.

III.

Het verdient geen aanbeveling bij zelfbestuivers een nieuw ras te vormen uit een mengsel van zuivere lijnen.

IV.

De nieuwe rassen moeten gekweekt worden voor een bepaalde grondsoort, waarvoor op de voornaamste grondsoorten een veredelingsbedrijf aanwezig moet zijn.

V.

Met behulp van kunstmatige genmutaties is het mogelijk pleiotropie te onderscheiden van absoluut c.q. intensief gekoppelde factoren.

VI.

De theorie der kalifixatie van Domingo verdient de voorkeur boven die van Hauser.

VII.

De z.g. single-value kaarten krijgen pas beteekenis naast een bodemkundige kaart, die de morphologie van het profiel als basis heeft.

VIII.

De tegenwoordige rassenproefvelden zijn ongeschikt om een economische bemesting te bevorderen.

NN08201,0130

WISKUNDIGE BESCHOUWINGEN BIJ ENIGE HOOFDSTUKKEN UIT DE PHILOSOPHIE DER LANDBOUWWETENSCHAP.

104289

Dit proefschrift met stellingen van
JAKOB PIETER SIJPKENS
geboren te Sneek den 2en Augustus 1916 is
goedgekeurd door de promotoren:
Dr. M.J. van Uven, hoogleeraar in de wiskunde
en
J.E. van der Stok, hoogleeraar in den tropischen landbouw.

De Rector-Magnificus der Landbouwhoogeschool
S.C.J.Olivier.

Wageningen, 23 November 1945.

281.A20

WISKUNDIGE BESCHOUWINGEN BIJ EENIGE HOOFDSTUKKEN UIT DE PHILOSOPHIE DER LANDBOUWWETENSCHAP.

Proefschrift
ter verkrijging van den graad van
DOCTOR IN DE LANDBOUWKUNDE
op gezag van den Rector Magnificus
Dr Ir S.C.J. Olivier,
hoogleeraar in de scheikunde
te verdedigen tegen de bedenkingen van
een Commissie uit den Senaat der
Landbouwhoogeschool te Wageningen op
Maandag 10 December 1945 te 16 uur

door

J.P. SIJPKENS

Bibliotheek
der
Landbouwhoogeschool
Wageningen

Bij het voltooien van dit proefschrift herdenk ik met genoegen en dankbaarheid de jaren 1935 tot 1945 welke ik in Wageningen doorbracht.

In het bijzonder gaat mijn dank uit aan hen, wier onderwijs ik heb mogen genieten, in het bijzonder aan U, Hooggeleerde van Uven en U, Hooggeleerde van der Stok. Dat Gij beiden als Promotor hebt willen optreden verdient mijn bijzondere erkentelijkheid; het proefschrift is toch niet op de gebruikelijke wijze tot stand gekomen, hetgeen Uw taak ten zeerste verzwaard moet hebben. Door de oorlogsomstandigheden is er geen enkele bespreking met U aan vooraf gegaan; bovendien werd een groot deel, dat verloren ging, herschreven tijdens de evacuatieperiode van Wageningen, wat de duidelijkheid wel niet ten goede gekomen zal zijn. Daar Indië roept, moest tenslotte in een te snel tempo afgewerkt worden.

Hooggeleerde Honing, het cytologisch practicum was mij het aangenaamste practicum.

Hooggeleerde Hudig, Uw ruim opgevatte colleges heb ik steeds gewaardeerd.

Zeer geachte Heer Cleveringa, Uw excursies in mijn practijktijd zijn een waardevolle schat in mijn herinneringen.

Zeer geachte Heer Wind, dit proefschrift, dat zijn ontstaan dankt, aan mijn werk aan het Centraal Instituut voor Landbouwkundig Onderzoek, had zonder de vrijheid, die U mij steeds gelaten hebt in mijn werk, niet tot stand kunnen komen. Ook de hulp ondervonden bij de publicatie er van, verdient blijvend dank. Hoewel ontstaan uit vraagstukken uit de practijk, zoeke men hierin geen bepaalde recepten, doch in de eerste plaats de algemeene ideeën, of, indien men deze aanvaardt, principes. Wie die kent; kan voor alle practische vraagstukken altijd de oplossing vinden.

Tenslotte breng ik een woord van dank aan allen, die mij behulpzaam zijn geweest, in het bijzonder aan den Heer IJwema, bibliotheekbeambte, wiens groote hulpvaardigheid ik vele jaren heb leeren kennen, en aan de tikkamer van het C.I.L.O., in het bijzonder aan Mej. Noordhuis.

INHOUD

pag.

Inleiding 8

HOOFDSTUK I

Krzymowski's philosophie der landbouwwetenschap.

§ 1. Inleiding 9
§ 2. Krzymowski's philosophie der landbouwwetenschap 9
§ 3. Krzymowski's philosophie als uitgangspunt 11
§ 4. Eenige hoofdstukken uit de philosophie der landbouwwetenschap. 11

HOOFDSTUK II

Inleidende Probleemstelling

§ 1. Het spel der factoren als centraal probleem 18
§ 2. Omschrijving der methode. 18

HOOFDTUK III

Ceteris paribus en factorial design

§ 1. Het factorial design als efficient en volledig ontwerp en als inductieve basis voor generalisatie 21
§ 2. Afbeelding van interacties 23
§ 3. Mathematische formuleering der main-effects en interacties 26
§ 4. Geidealiseerde modellen van het factoren mechanisme . . . 29
§ 5. Twee nieuwe begrippen: main-action en interplay. 36

HOOFDSTUK IV

Voortgezette probleemstelling 40

HOOFDSTUK V

Interpretatie van het factorial design

§ 1. Fisher's analysis of variance 42
§ 2. Ezekiel's "additive and joint functions". 45
§ 3. De φ-regressiefuncties 46
§ 4. De betrekkingen tusschen de analysis of variance en de φ regressiefuncties bij 2 factoren 48
§ 5. Variances uitgedrukt in de constanten der regressievergelijking 51
§ 6. Substitutie van de reeksen van hoogere orde 52
§ 7. Het gedrag der variances 54
§ 8. De onderlinge verhouding der variances 54
§ 9. Voorbeeld 1 56
§ 10. Herhaling bij hoogere graads regressievergelijking . . . 63
§ 11. Voorbeeld 2 70
§ 12. De betrekkingen bij 3 factoren 74
§ 13. Voorbeeld 3 82
§ 14. De toevalsvariance als interactievariance 86
§ 15. Bewijsgang voor niet-aequidistante ordinaten 87
§ 16. Drie korte voorbeelden over het afleiden van de φ-regressievergelijking uit de analysis of variance en omgekeerd. 88

HOOFDSTUK VI

pag.

De regressiefunctie in de natuur

§ 1. Algemeene types 93
§ 2. Historische ontwikkeling van de theorie der opbrengstcurven 93
§ 3. Formuleering voor één factor volgens Pearl 95
§ 4. Formuleering voor twee of meer factoren 97
§ 5. Over het optreden van onbekende types van opbrengstcurven. 98
§ 6. De tragiek van het spel der factoren. 100

HOOFDSTUK VII

Het pan parallelle experiment

§ 1. Over de grenzen van het natuurwetenschappelijk kenvermogen 103
§ 2. Het incomplete factorial design en het incomplete complex experiment 104
§ 3. Uitvoering van een incomplete complex experiment met 3 factoren 105
§ 4. Bespreking van het incomplete complex experiment bij 4 en 5 factoren. 115
§ 5. Tabellen 121
§ 6. Splitsen in groepen onafhankelijke factoren met het 2^n experiment. 128
Samenvatting en overzicht 132

INLEIDING.

In de landbouwwetenschap, zooals wij die heden ten dage kennen, vormen philosophie en wiskunde twee bijzondere componenten, waarvan de eerste, de philosophie, een bijkans onbekende, de tweede, de wiskunde, een algemeen bekende, doch eenigszins betwiste component is.

Eenige uitleg betreffende de eerst genoemde component is derhalve noodzakelijk, wil het geheel voor een grootere lezerskring aanvaardbaar zijn. Eenige gedeelten van de eerste hoofdstukken zijn dienovereenkomstig daaraan gewijd. Hierbij is gebruik gemaakt van min of meer gangbare opvattingen, die niet boven de traditioneele zienswijzen uitgaan. Zij vormen niet de quintessens of het discutabele gedeelte van dit geschrift, in tegenstelling tot de volgende hoofdstukken, die zij inleiden en over genoemde tweede component handelen.

Daar de probleemstelling in het volgende zich daardoor slechts langzaam ontwikkelt, doet men goed - om de gedachten beter te bepalen - eerst de samenvatting en het overzicht aan het slot van dit geschrift te lezen.

Er is aldus getracht den lezerskring zoo groot mogelijk te doen zijn.

HOOFDSTUK I

KRZYMOWSKI'S PHILOSOPHIE DER LANDBOUWWETENSCHAP.

§ 1. Inleiding.

De philosophie der landbouwwetenschap duidt een begrip aan, waarvoor men tot dusver weinig aandacht heeft gehad. Het is daarom misschien niet ongewenscht eerst even stil te staan bij de philosophie in het algemeen.

Hiervoor zijn tal van definities in omloop, die echter zeer sterk van elkaar verschillen, hetgeen evenwel geen verwondering kan wekken. Bij alle oude volken ging nl. overal de godsdienst aan de wijsbegeerte vooraf; pas later ontwikkelde zich hieruit de wijsbegeerte, terwijl de hiermede samenhangende zich ontwikkelende probleemstelling het aanzien gaf, aan steeds wisselende definities van de wijsbegeerte. Doch niet alleen in den loop der historie, ook tijdens een bepaald tijdperk zijn vele definities naast elkaar in omloop in overeenstemming met de naast elkaar staande philosofische stelsels. Het is hier niet de plaats, daarop nader in te gaan. Ieder kieze zijn eigen standpunt.

Bij het beoordeelen der afzonderlijke wetenschappen, o.a. het critisch onderzoek van hun methodologische principes, kan men steeds uitgaan van, en zich laten leiden door de philosophie, en zoo voor elke wetenschap een zgn. vakphilosophie (waarover in de volgende § meer) opbouwen. Deze vakphilosophie kan verschillend uitvallen al naar gelang waar men het zwaartepunt legt bij de philosophie. In de oudere philosophie (vooral in Europa) was het hoofdprobleem steeds weer de verhouding tusschen het denken en de werkelijkheid. In de nieuwere philosophie (vooral in Amerika: James, Dewey) legt men den nadruk op het probleem van de verhouding tusschen het denken en het handelen. Een zeer globaal onderscheid, doch zeer typeerend. Bij Krzymowski overheerscht nog het eerstgenoemde standpunt. Van de landbouwwetenschap zegt hij o.a.: "Het hoofddoel van elke wetenschap is het zuivere kennen. Wij willen alles wat in de wereld voorkomt en gebeurt, kennen en begrijpen; wij willen onze wetensdrang bevredigen en daartoe dient de wetenschap. Maar of een leer ons practisch ten nutte is of niet, is daarbij een vraag van secondair belang. Wie door een wetenschap werkelijk eenmaal diep aangegrepen is, die wil slechts de schoonheid er van genieten, dat is alleen het bevredigende en verhevene. Of daarnaast de wetenschap nog practische beteekenis heeft, is een andere vraag, een vraag die met het aesthetisch genot, dat de echte geleerde het naaste aan het hart ligt, niets te maken heeft. Ook wanneer de landbouwwetenschap in het geheel geen practische beteekenis had, wanneer de theorie in het geheel niet in staat was de practijk te verbeteren, dan was er toch een landbouwwetenschap. Er zouden altijd menschen zijn, die de verschijnselen in den landbouw, even goed als alle andere dingen, verklaart zouden willen zien."

Krzymowski gaat uit van een tweetal (eenvoudige en beperkte) met het eerstgenoemde standpunt in overeenstemming zijnde definities van de philosophie. Bezien wij dat nader. Aan Krz. inleiding zij daartoe het volgende ontleend:

§ 2. Krzymowski's philosophie der landbouwwetenschap. [1])

Men kan de philosophie definieeren als de wetenschap der principes (Ueberweg-Heinze, Grundrisz der Geschichte der Philosophie) en daar het volgende over de principieele grondslagen van de landbouwwetenschap handelt, in het bijzonder methodologische principes, kan men spreken van een philosophie van deze wetenschap. Men kan ook philosophie definieeren als kenniscritiek (Mauthner, Wörterbuch der Philosophie) en ook daarmee is het begrip philosophie der landbouwwetenschap in overeenstemming, want hierin wordt in het bijzonder gehandeld over de critiek van het kenvermogen der landbouwwetenschap.

Zooals bekend, wisselen in den loop der tijden de wetenschappen en kunsten herhaaldelijk van standpunt, nieuwe gezichtspunten, nieuwe smaakrichtingen doemen op en het veranderde standpunt - in de wetenschap vaak een veranderde philosophische instelling - heeft niet zelden geheel andere perspectieven tengevolge. In het bijzonder worden de beschouwingswijzen menigvuldiger; wat men vroeger van uit een gezichtspunt beschouwde werd later gelijktijdig vanuit verschillende gezichtspunten beschouwd. Hoe voordeelig dit ook voor de

---

[1]) Prof.Dr.R.Krzymowski. Philosophie der Landwirtschaftslehre. Duitsche uitgave 1919 (Russische uitgave 1927, Japansche uitgave 1932).

menschelijke kennis mag zijn, de wetenschap wordt daardoor veel gecompliceerder en het begrip er voor wordt veel moeilijker. Elke richting krijgt zijn eigen taal en vaak zullen onderzoekers en kunstenaars, die in een bepaalde richting goed ingeleefd zijn, geen begrip voor een evenzeer gerechtigde andere richting hebben.

Het is daarom noodzakelijk en gebruikelijk de verschillende richtingen in kunst en wetenschap in de geschiedenis genetisch weer te geven. Ook in een philosophie der landbouwwetenschap zal men verplicht zijn, dezen weg in te slaan; er moet een historische inleiding op genetischen grondslag in opgenomen worden (1e Hoofdstuk van Krzym.: Geschiedenis van het ontstaan der landbouwwetenschap).

Het is begrijpelijk, dat bij de onoverzienbare hoeveelheid landbouwkundige literatuur, daarbij philosophische grondslagen reeds herhaaldelijk behandeld werden, ook wel in inleidende hoofdstukken van algemeene leerboeken. Desondanks zijn er weinig geschriften, die zich uitsluitend hiermede bezig houden en er dieper op in gaan. In andere wetenschappen daarentegen is men er reeds lang toe overgegaan de methodologische grondslagen zelf weer methodisch te bewerken; men denke slechts aan de begrippen: kunstphilosophie, philosophie van den gondsdienst, philosophie der geschiedenis, kultuurphilosophie, philosophie van het recht, philosophie van de wiskunde, van de natuurwetenschappen (men denke bij dit laatste aan de uiteenzetting van de begrippen ruimte en tijd, materie en energie, beweging, selectie, ontwikkeling, enz.). Des te opvallender is het dat een dergelijke Vakphilosophie met betrekking tot de landbouwwetenschap, nog zoo weinig ontwikkeld is.

De landbouwwetenschap heeft nu daadwerkelijk zekere philosophische grondslagen, welke de geheele leer beheerschen en die, zij het soms onbewust, bij de meeste vertegenwoordigers leven.

Tot nog toe worden slechts enkele methodologische principes in de landbouwwetenschap wel als zoodanig beschouwd; zoo bv. dat het doel van de landbouwwetenschap de verbetering van de landbouwkundige practijk is; dat de bron van landbouwkundige kennis in het opstellen en verwerken van experimenten ligt; dat de landbouwer, die de resultaten van deze experimenten en hun causale samenhang kent, den op het gevoel, den empirisch werkenden landbouwer verre de baas is en dat alleen hij rationeel zijn bedrijf leidt.

Dergelijke uitspraken leest men in landbouwkundige verhandelingen bij herhaling. Zij schijnen het gezonde menschenverstand zoo vanzelfsprekend en evident, dat ze vaak als axioma's gesteld worden, waaraan in principe niet te tornen valt.

Toch zullen we zien, dat bovengenoemde uitspraken, hoe zeer ze ook op theorie en practijk bevruchtend mogen werken, toch niet zoo vanzelfsprekend en algemeen geldig zijn, als men gewoonlijk aanneemt en dat men zich niet eenzijdig slechts aan deze dogma's vast mag houden. Eenerzijds dienen bovenstaande stellingen aanzienlijk beperkt te worden, anderzijds kan men de landbouwwetenschap nog van geheel andere gezichtspunten behandelen, die een geheel ander beeld geven. De tegenwoordige landbouwwetenschap die te zeer op bovengenoemde en gelijksoortige grondstellingen ingesteld is, doet niet alle feiten en verschijnselen van de werkelijkheid recht wedervaren. Tot zoover Krzymowski's inleiding.

Na dit uitvoerig citeeren van Krzymowski's inleiding, ter rechtvaardiging van het begrip philosophie der landbouwwetenschap, dienen we ons thans te beperken tot de weergave van de inhoudsopgave van Krz. boekwerk.

Slechts van die hoofdstukken, die hier nader uitgewerkt zullen worden, zal een verkorte inhoud weergegeven worden, voor het overige wordt verwezen naar het werk van Krz.

1. Inleiding (hierboven uitvoerig geciteerd).
2. Geschiedenis van het ontstaan van de landbouwwetenschap. (Van de oudste tijden tot Thaer; Thaer; het rationalisme; von Schwerz; von Thünen; de onderzoekers uit den tijd tusschen Thaer en Liebig; Liebig).
3. De landbouw als bedrijf en als natuurverschijnsel. Onontbeerlijkheid van den landbouw.
4. Plaats en doel van de landbouwwetenschap.
5. Het rationalisme in de landbouwwetenschap en het factorengewirwar (Faktorengewirr).
6. Het experimenteele onderzoek in den landbouw.
7. Het niet-experimenteele onderzoek in den landbouw.
8. Het principe van de selectie en de historische ontwikkeling in den landbouw.

9. Aanvulling der systematisch-experimenteele landbouwwetenschap door landbouwgeschiedenis en landbouwgeographie.
10. De practische ervaring.
11. Aesthetisch genot van den landbouw en de landbouwwetenschap.
12. Enkele misstanden in de landbouwwetenschap.
13. Aanhangsel: de landbouwgeographie.

§ 3. Krzymowski's philosophie der landbouwwetenschap als uitgangspunt.

Zooals reeds opgemerkt werd in § 1, ging Krzymowski in philosophisch opzicht vooral van het oudere Europeesche standpunt uit. Het modernere (Amerikaansche) standpunt, dat het hoofdprobleem legt bij de verhouding tusschen het denken en handelen, zou, indien op den voorgrond gebracht in een vakphilosophie als de philosophie der landbouwwetenschap, tegenwoordig van zeer groot belang zijn. Immers er wordt thans op zoo menigvuldige wijze in den landbouw ingegrepen, waarbij men zich meestal slechts baseert op politieke drijfveeren en idealen - die uit den aard der zaak slechts een tijdelijke waarde hebben - waardoor zonder twijfel de onzekere toestanden gecontinueerd blijven. Een verschijnsel, dat niet met overeenkomstige uit vroegere tijden te vergelijken is, daar men toen naast de even veelvuldige politieke drijfveeren, slechts één ideaal, een Christelijk ideaal had. Waar verwacht mag worden dat het Christelijk Geloof steeds meer terrein verliest, klemt deze groote lacune des te meer. Hoofdzaak zij echter, dat men zich bewust is, dat genoemde philosophie der landbouwwetenschap van Krzymowski lang niet het onderwerp uitput, in hoofdzaak treedt het aesthetische en het logische, kentheoretische op den voorgrond. Ook metaphysische beschouwingen ontbreken geheel. [1]) Nochtans moet opgemerkt worden, dat het aantal geschriften dat over de philosophie der landbouwwetenschap handelt, zeer gering is. Het werk van Krz. is het meest omvattende en breedst opgezette, dat schrijver bekend is. Bovendien kan een groote eenheid het werk niet ontzegd worden. Men moet echter indachtig blijven, dat voor de philosophie der landbouwwetenschap wel zeer sterk geldt, dat de problematiek zich nog ontwikkelen moet. Schrijver heeft evenwel niet geaarzeld Krzymowski's werk als uitgangspunt te nemen; dat wil niet zeggen, dat een ieder die zoo doet, het in alles eens behoeft te zijn met Krz. Zoo zal elke lezer die Krz.'s werk bestudeert wel tot de conclusie komen, dat aan de hoofdstukken 5, 6 en 7, die min of meer bij elkaar behooren, iets ontbreekt; wel geven zij den stand van zaken helder weer, doch tot een oplossing van de gestelde problemen komt het niet.

De taak die schrijver zich gesteld heeft, is te trachten deze leemte aan te vullen. Om tot een goed begrijpen der gestelde problemen te komen, moge hier eerst een korte weergave van de genoemde hoofdstukken volgen.

§ 4. Eenige hoofdstukken uit de philosophie der landbouwwetenschap.

(Hoofdstuk 5:) De ten tijde van Albrecht Thaer (1752-1828) zich ontwikkelende landbouwwetenschap moet beschouwd worden als een product van het tijdperk der verlichting en speciaal van het in deze periode opkomende rationalisme. Het hoofdwerk van Thaer draagt dan ook den titel "Grondbeginselen van den rationeelen landbouw" en de door Thaer gegrondveste school wordt in de geschiedenis de "school der rationalisten" genoemd.

De voor de hand liggende verwantschap tusschen het philosophisch rationalisme en de rationalistische richting der landbouwwetenschap is echter zelden naar voren gebracht. Waarom, is niet moeilijk te verklaren: de bijna geheele hedendaagsche landbouwwetenschap is nog steeds rationalistisch; het overgroote deel der landbouwkundige onderzoekers denkt rationalistisch, zonder het zelf te weten. In het tijdperk der Verlichting trachtte men het verstand van de boeien der dogma's te bevrijden; men trachtte tot de natuur en het gezonde menschenverstand terug te keeren. In de philosophie en de theologie bestreed men de dogma's (Voltaire). Men wilde de dingen zoo veel mogelijk natuurlijk verklaren (natuurlijke religie). De opvoeding, zooals die tot nog toe gold, diende verworpen te worden (Rousseau). In de theologie keerde men zich tegen de openbaring en de wonderen. De religie werd meer en meer een questie van het verstand (Schleiermacher). De rechtswetenschap verkondigde het zoogenaamde natuurrecht in tegenstelling met het geldende, historisch ontstane, positief recht ("droits de l'homme"). In de staathuishoudkunde verkondigden de economen de leer van een natuurlijke regeling van de economische processen (laisser faire, laisser passer).

---

[1]) Dit zeer zeldzame onderwerp kan men bv. vinden bij Graf Hermann Keyserling: "Das Reisetagebuch eines Philosophen".

In een natuurwetenschap als de physiologie kwam echter het rationalisme veel later tot bloei, pas ongeveer in het midden van de 19e eeuw. Het rationalisme is onhistorisch, het werkte met het zgn. tijdelooze verstand, ze wilde de historische fundeering van de kultuur door een rationeele vervangen. Voor de bonte verscheidenheid der gebeurtenissen, voor de noodzakelijkheid van de geschiedkundige ontwikkeling, de doelmatige aanpassing en de bijzondere schoonheid van het historisch gegroeide, had het rationalisme geen begrip. Dit had een groote verarming van de geestenswereld tengevolge. Het rationalisme geloofde echter aan een geweldigen opbloei van alle krachten, die tot nog toe in de windselen van de dogma's gesluimerd hadden. Hierin lag ook daadwerkelijk de geschiedkundige missie van het rationalisme. Zeker heeft ze veel goeds voortgebracht, maar ze heeft toch haar eigen kenvermogen overschat, ze was te optimistisch, te weinig realistisch. De philosophie van het rationalisme had geen begrip voor het schoone, het levensvolle, ze was nuchter en plat, wilde te veel verklaren, was in zijn streven te utilitaristisch, te onhistorisch.

Zooals we reeds zagen kwam in de natuurwetenschappen het rationalisme veel later tot bloei en in de landbouwwetenschap "bloeit" deze richting nog steeds en hoewel ze veel goed gedaan heeft, laat zich toch niet ontkennen, dat de richting eenzijdig is. De aangeduide ongunstige eigenschappen zijn duidelijk bemerkbaar in de landbouwwetenschap.

In de reeds even aangestipte theologie, philosophie, rechtswetenschap enz. is deze richting reeds lang voorbij en wordt nog slechts beschouwd als een [illegible] ontwikkelingsfase.

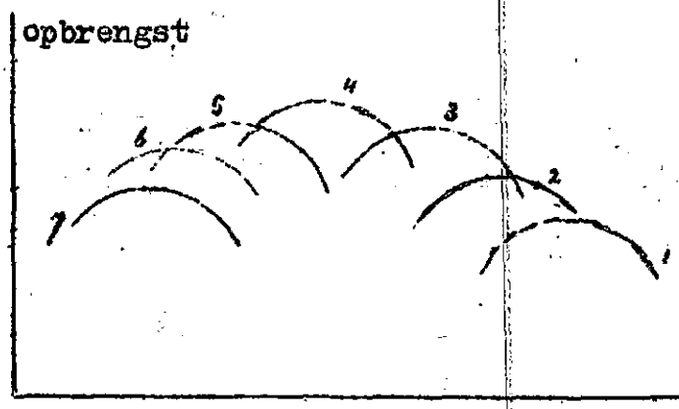
Houden we ons thans bezig met de toepassing van de rationalistische wereldbeschouwing op de landbouwwetenschap, dan kunnen we ons dadelijk met recht afvragen of het menschelijk verstand werkelijk het vermogen bezit, alle dingen zoo juist te doorschouwen, dat het nieuwe beter is dan het oude, het historisch ontstane. Thaer maakte zich hierover geen zorgen: hij verlangt dat de landbouwer de causaliteit der verschijnselen zoo ver kan en moet beheerschen, dat hij in elk individueel geval in staat is, de regel te vinden die hij volgen moet en het resultaat van zijn maatregelen te voorzien en te berekenen. Nu zijn echter de landbouwkundige verschijnselen van een zeer groot aantal verschillende factoren afhankelijk, waar men bovendien vaak slechts een deel van kent; wil men dan als boven te werk gaan, dan moet men al deze factoren en hun werking kennen, terwijl om een resultaat vooruit te berekenen, men al deze factoren ook tot op vrij groote nauwkeurigheid moet kunnen meten. Dergelijke bepalingen zijn dermate ingewikkeld, dat ze in werkelijkheid niet uit te voeren zijn, terwijl bovendien voor vele factoren (nog) geen bruikbare meetwijze bestaat. Geen landbouwer berekent volgens wiskundige vergelijkingen hoe hij de verschillende factoren moet groepeeren en moet laten werken om het grootste effect te bereiken.

Om een voorbeeld te nemen, men wil de optimale zaaidiepte van een gewas leeren kennen. Voor het kiemen van het zaad zijn hoofdzakelijk drie factoren van belang: warmte, lucht en water. Voor voldoende warmte moet het zaad ondiep ondergebracht worden, voor voldoende vocht juist weer dieper. Hoe moeten deze factoren, waarop de landbouwer bovendien slechts een zeer beperkten invloed kan uitoefenen, gemeten worden en hoe moet hieruit de optimale diepte bepaald worden? Bovendien zijn er nog andere kiemfactoren, die maar zeer ten deele bekend zijn: licht, groeistoffen. Bovendien komt het niet alleen op de kieming zelve aan, ook de architectuur van het kiemplantje is van het hoogste belang en die wordt in belangrijke mate beinvloed door de zaaidiepte, hoe moet dit berekend worden? Bovendien laat zich de vraag herhalen voor een verschillende bodem, klimaat, intensiteitsgraad van het bedrijf. Bovendien gaat het bij het zaaien niet alleen om de juiste zaaidiepte, doch ook om de zaaizaadhoeveelheid, zaaitijd, zaaimethode, zaaizaadontsmetting, bemesting voor en na het zaaien; kortom een gewirwar van factoren, die zich op geenerlei wijze berekenen laat.

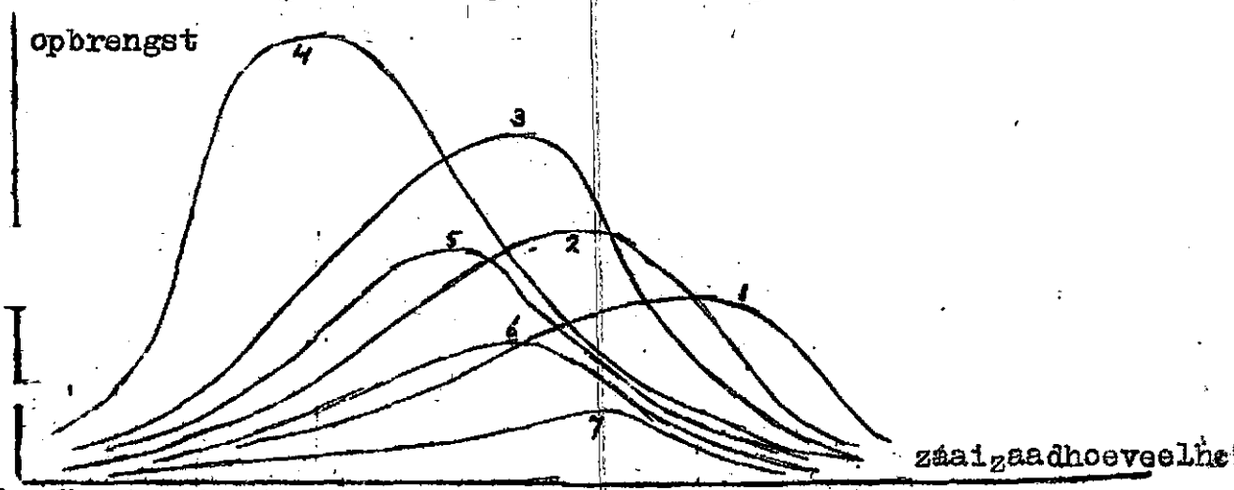
In werkelijkheid handelt de landbouwer geheel anders, wel probeert hij zich over de werking van enkele factoren te orienteeren, om bij tijden zekere veranderingen aan te brengen, doch over het geheel blijft hem niets anders over, dan zich aan het plaatselijk gebruik te houden, aan de overlevering, de traditie. In deze eeuwenlange ervaring is reeds lang de juiste zaaidiepte voor dat gewas gevonden, die over een zeker aantal jaren gemiddeld de beste resultaten geeft. Van een theoretisch vooruit be-

palen van de regels waaraan de rationeele landbouw gelooft, is geen sprake. Al kent de landbouwer de factoren theoretisch nog zoo goed, zoo heeft hij toch niet het geringste houvast, hoe nu in een speciaal geval deze factoren naar hun gradueele waarde te bepalen.

Wij zullen dit voorbeeld nog even aanhouden en nader analyseeren. Van groot belang bij het zaaien is bv. de zaaizaadhoeveelheid. Houdt men alle overige factoren constant, dan verkrijgt men bij een zekere zaaizaadhoeveelheid, de optimale zaaizaadhoeveelheid, een maximum opbrengst. Graphisch voorgesteld krijgt men in dit "ceteris paribus" geval een zgn. optimum curve. Nemen wij nu een tweeden factor in onze beschouwingen op bv. bemesting, dan zal met een stijgende bemesting, de opbrengst toenemen, doch slechts tot een bepaald bedrag, om daarna weer te dalen. Om dit experimenteel na te gaan bepale men de optimum zaaizaadhoeveelheid bij bv. 7 mestgiften. Bij elke mestgift ontstaat een optimum curve, elk met het maximum op een andere plaats en van een verschillende grootte, graphisch voorgesteld aldus:



Bij toenemende mestgiften dus een optimum verschuiving naar links. Opmerking: Deze figuur, die overgenomen is van Krz., moge voldoende zijn, om de gedachten te bepalen, indien de factoren onbenoemd waren gebleven, bv. factor A en factor B. Waar hier echter uitdrukkelijk gesproken wordt van zaaizaadhoeveelheid en mestgift, moet opgemaakt worden dat onderstaande figuur beter geweest zou zijn.



De landbouwer bezit geen formules die het hem mogelijk maken deze optima te berekenen. Bij uitvoering van een experiment wordt het aantal objecten ook te groot. Bovendien hebben wij hier nog pas de werking van 2 factoren behandeld; de factor mestgift moet echter gesplitst worden in een N, P en K-gift en misschien nog andere, waarnaast nog komt de wijze waarin ze toegediend moeten worden, N als NH_3 of NO_3, de tijd van aanwending, enz.. Verder, hoe moeten deze factoren gecombineerd worden met verschillende zaaidiepte, zaaimethode, voorvrucht, enz.? Hoe moet dit berekend worden? Het verlangen van den rationeelen landbouw, dat de landbouwer op grond van zijn kennis van den causalen samenhang der factoren in staat moet zijn, de toegepaste methode te doorzien en eventueel vooruit te berekenen is een utopie. In werkelijkheid doorziet de landbouwer het factorenspel niet, maar volgt de plaatselijke gewoonten. De landbouw in een streek is niet iets willekeurigs, iets wat zich berekenen laat, zooals de rationalistische beschouwingswijze wil doen laten voorkomen, maar is historisch gegroeid; door de ervaring van talrijke generaties zijn de

doelmatigste methoden gevonden en door overlevering aan de huidige generatie bekend. De landbouwer gaat zeer veel op zijn gevoel af; wat hij met het verstand niet bevatten kan - geen enkele landbouwer kan het spel der factoren doorschouwen - vat hij instinctief veel juister met het gevoel, zooals Schiller zegt: "Die Empfindung der meisten Menschen ist richtiger als ihr Räsonnement. Erst mit der Reflexion fängt der Irrtum an." De landbouw is niet zoo doorzichtig als de rationeele richting wel aanneemt.

De rationeele richting voert er toe dat vele onderzoekers in een bepaalde streek, den daar heerschenden afwijkenden landbouw dadelijk critiseeren, en als achterlijk bestempelen en een gemoderniseerden volmaakten landbouw voor die streek uit de mouw schudden, maar al hun maatregelen op niets zien uitloopen.

Aan den anderen kant moet men niet denken, dat theoretische overwegingen geen rol spelen in den landbouw, men zou dan in een tegengesteld extremum vervallen. Zoo goed als het rationalisme in andere wetenschappen zijn taak verricht heeft, we wezen daar reeds op, zoo is ook zijn taak in den landbouw ten einde.

Tot op zekere hoogte is de landbouwer wel in staat op grond van theoretische overwegingen practische maatregelen te nemen; ze zijn echter niet voldoende, ze moeten door de ervaring, de overlevering aangevuld worden. Keeren we in verband hiermede naar ons voorbeeld terug. Een landbouwer bezaait een veld dat in een hoogen vruchtbaarheidstoestand verkeert; om het legeren te voorkomen zal hij een kleinere zaaizaadhoeveelheid nemen, hoeveel minder laat zich niet berekenen, maar hij zal zich moeten laten leiden door een door de ervaring geschoold gevoel in dezen. Op een veld zaait hij vroeger dan op een ander; voor het eerste moet weer een kleinere, voor het tweede een grootere zaaizaadhoeveelheid gekozen worden, maar het eerste veld heeft een voldoende N-voorraad van een voorvrucht, het tweede moet echter een N-bemesting hebben. De landbouwer overweegt nu hoeveel, ook in verband met het vroegere en latere zaaien. In deze overwegingen komt de rationalistische opvatting van den landbouw naar boven, evenwel nooit zuiver, maar altijd afgewisseld door empirisch-practische ervaring.

Een landbouwer die zich alleen door rationalistische overwegingen zou laten leiden, zou ten onder gaan, geestelijk door het getob met het factorenspel, economisch door den ondergang van zijn bedrijf. De abstracte ratieneele synthese is aan de werkelijke synthese der verschijnselen nooit congruent, daar het spel der factoren niet te doorzien is; altijd heerscht de tegenstelling tusschen de subjectieve logica van het denken en de objectieve logica van de gebeurtenissen.

(Hoofdstuk 6:) In het voorgaande werden de impulsen, die de zich ontwikkelende landbouwwetenschap van het rationalisme ontving behandeld. Er zijn echter nog andere invloeden werkzaam geweest. De groote beteekenis van het experiment in de natuurwetenschappen, werd aanleiding ook de landbouwwetenschap op een experimenteelen grondslag te plaatsen. Thaer wees reeds op de belangrijkheid van het experiment. Bij het grondvesten van de landbouwchemie door Liebig, kwam het experiment als een machtig impuls opzetten. Heden ten dage zijn alle landbouwstaten met een netwerk van proefstations, landbouwkundige instituten, proefvelden, proefboerderijen e.d. overtrokken. In het bijzonder maakt men gebruik van zgn. vergelijkende proeven. Men houdt alle factoren constant op een na (ceteris paribus) en onderzoekt den factor waarom het gaat als een alternatief of in meerdere trappen.

Hoe zegenrijk deze methode ook geweest mag zijn en nog is, zekere grenzen zijn toch aan deze methode gesteld. Waar men nauwelijks nog andere methoden toepaste, ontstond een groote eenzijdigheid. Ten eerste is de mogelijkheid van kennisverkrijging met het experiment beperkt en ten tweede is de onderzoeking van den plaatselijk historisch gegroeiden landbouw volgens empirischen weg noodzakelijk en onontbeerlijk, al was het alleen maar om de resultaten van de op de experimenteer steunende theorie te controleeren.

De overschatting van den experimenteelen landbouw is wederom een fout van de rationeele wereldbeschouwing. Bezien we dit nader. Onderzoeken we een factor (ceteris paribus) zooals dat gewoonte is, dan worden we slechts ingelicht over de werking van dezen eenen factor. In het

voorbeeld van het zaaien hebben we reeds gezien hoe groot het aantal factoren inderdaad is, toestand van het zaad, bodem en klimaat, vaak factoren die zelf weer uit tal van factoren bestaan. Wanneer men nu werkelijk volgens de rationeele wijze de juiste zaaidiepte wil bepalen, moet men experimenten opstellen voor al deze factoren. Maar dan is men nog niet klaar, we moeten dan nog de factoren combineeren en elke combinatie onderzoeken! Die bepaling, de theoretische reconstructie van zooiets is echter een onmogelijkheid. Het werkelijke resultaat van het gewirrwar der factoren laat zich alleen empirisch in de practijk vaststellen.

Een andere moeilijkheid van het experimenteele onderzoek is de juiste interpretatie van de experimenten. Experimenten juist uitvoeren kan men leeren, maar de experimenten juist interpreteeren is een gave, een kunst. Het is niet moeilijk dit te demonstreeren. Nemen we varieteiten proeven als voorbeeld: hoeveel jaren moeten de proeven voortgezet worden. Vaak neemt men aan 3 jaren, doch meestal is dit te weinig: de toevallige afwijkingen zullen groot blijven [1]). Wil men zeker zijn, zoo zegt men vaak, men neme tal van proeven op verschillende plaatsen en in verschillende jaren en werke met de gemiddelden. Dit is onjuist, neemt men gemiddelden, dan worden de verschillen tusschen de varieteiten weggevaagd:

In streek I brengt ras A meer op dan ras B, in streek II brengt ras B meer op dan ras A. Middelen we over beide streken, dan gaat een groot deel van het onderscheid tusschen A en B verloren. Bij veel proeven en bij veel rassen houdt men dan een nietszeggend resultaat over, waarmee men niet adviseeren kan, voor geen enkele streek geldt het gemiddelde resultaat. Gaat men de proeven weer scheiden naar grondsoort, klimaat enz., dan krijgt men binnen de groepeeringen weer kleine aantallen en dus groote toevallige afwijkingen.

Dat de interpretatie van varieteitenproeven zeer moeilijk is, kan ook nog uit het volgende blijken. Men onderzoekt 2 varieteiten, een oud ras A en een nieuw ras B. Oogenschijnlijk blijken op het proefveld beide rassen hetzelfde op te brengen. Ras A moet dun gezaaid worden wegens gevaar voor legeren. Men heeft nu ras B in deze vergelijkende proef waarin een factor gevarieerd wordt, even dun gezaaid, hoewel om maximale opbrengsten van B te krijgen, dicht gezaaid moet worden. Bijgevolg is deze proef waardeloos.

Veel landbouwkundige proeven hebben geen ander resultaat, dan dat ze de regels van de practijk bevestigen. De practijk heeft in generaties lange ervaring in duizenderlei variaties door alle landbouwers, reeds lang de beste werkwijze gevonden (Een ongedwongen verklaring vindt men door toepassing van de selectiehypothese, men zie hoofdstuk 8 van Krz.; het principe van de selectie en de historische ontwikkeling in den landbouw).

Anders wordt het echter, wanneer de omstandigheden in een streek, hetzij plotseling, hetzij vrij snel, zich wijzigen; dan zijn experimenten doelmatig en noodzakelijk. Wanneer de landbouw in een streek den laatsten tijd intensiever geworden is, bv. na een verbeterde afwatering, waardoor beter bemest kan worden, dan moet ook de tot dusver gebruikelijke zaaizaadhoeveelheid veranderen. Maar zulke veranderingen brengt de landbouwer op grond van ervaring langzamerhand aan en in gelijken tred met de hoogere intensiveering, zonder daartoe berekeningen uit te voeren of proeven te nemen.

Degenen die zich op een eng standpunt stellen en de landbouwwetenschap uitsluitend als een experimenteele wetenschap bepleiten, bedienen zich vaak van het epitheton "exact". Dit is een bedriegelijke voorstelling van zaken. Men kan het uit historisch oogpunt echter wel begrijpen, dat men de natuurwetenschappen exacte wetenschap gaat noemen. Vaak noemt men die wetenschap exact, die hun problemen mathematisch nauwkeurig probeeren op te lossen (physica, chemie, astronomie, mechanica). In de natuurwetenschappen gaat men echter vaak verder, men noemt ook biologische wetenschappen als genetica, bacteriologie a.a. exact en tenslotte ook de landbouwwetenschap. Exact is feitelijk alleen de wiskunde, als men afziet van enkele begrippen als bv. onein-

[1]) De verschillen tusschen de jaren, hoofdzakelijk veroorzaakt door weersomstandigheden, hebben we niet in onze macht. Het is nu veelal gewoonte de oncontroleerbare variatie-oorzaken als toevallige te rekenen.

dig, dat als transcendent moeilijk te vatten is. De physica die het dichtst bij de wiskunde staat, moet echter reeds veel terrein prijs geven aan hypothesen en is feitelijk alleen maar exact in zooverre de wiskunde toegepast wordt. Gaan we vervolgens over naar de moderne astronomie, de chemie, naar de biologische wetenschappen physiologie, anthropologie, enz. dan verdwijnt het exacte karakter steeds meer. Het werken met steeds wisselende hypothesen behoort juist tot het karakter der natuurwetenschappen.

Het epitheton exact geeft den leek het geloof dat de natuurwetenschappen hooger zouden staan dan de andere wetenschappen. Ook in de landbouwwetenschap is het woordt niet op zijn plaats. Hoe problematisch, hoe kwellend onzeker zijn nog vele resultaten van die wetenschap; hoe duister en twijfelachtig zijn nog de bemestingsleer en de bodemkunde enz. ondanks alle exacte experimenten! De kortzichtige beperking van de landbouwwetenschap als experimenteele wetenschap alleen, uit zich juist in dit woord exact. Het is zinloos, dat vele landbouwkundige proeven geen ander resultaat hebben, dan dat ze de regels van de practijk exact bevestigen!

(Hoofdstuk 7:) In het voorgaande zagen we, dat hoewel de experimenteele methode van groot belang is, deze alleen toch niet in staat is, alle problemen op te lossen. Het gewirwar van de factoren is zoo groot, dat het niet mogelijk is, met experimenten alle verschijnselen in hun factoren te analyseeren en vervolgens met deze factoren, door synthese tot oplossing te komen. Het werkelijke resultaat van het factorengewirwar laat zich niet berekenen, laat zich alleen uit empirische waarnemingen in de practijk vaststellen. Hieruit volgt, dat een preciese empirische beschrijving van de landbouwkundige practijk en zijn resultaten, een voornaam deel van de landbouwwetenschap uitmaakt. Alleen al voor controle, verificatie en correctie op de experimenteele resultaten, is een zuivere beschrijving onontbeerlijk. Deze beschrijving is eenerzijds basis, anderzijds toetssteen der Landbouwkundige theorieën.

En dit geldt niet alleen voor het experimenteele (natuurwetenschappelijke) deel van de landbouwkunde. In de economie hebben we een zelfde methode, maar nu als abstractie, niet in de concrete werkelijkheid. In gedachten houdt men alle factoren op één na constant, laat dezen factor varieeren en tracht door een abstracte, logische constructie de werking van dien factor te leeren kennen. Men kan dit voor meer factoren herhalen en trachten het effect van combinaties van deze factoren op te sporen. Deze bloote abstractie en deductie is echter van beperkte waarde, de geschiedenis toont het ons ten overvloede: de klasieke school (Smith, Ricardo, Malthus, e.a.) en de physiocraten werkten met deze methode, doch kwamen in tegenspraak met de werkelijkheid zoodat men omzwaaide naar een historisch realistische school.

Zoo komen voor de landbouwwetenschap, behalve de experimenteele nog andere methoden in aanmerking: op de beteekenis van een zuivere beschrijving wezen we reeds, daarnaast kunnen we gaan vergelijken met behulp van de landbouwgeographie (zie aanhangsel Krz.), verder de statistiek, de landbouwboekhouding. En zeer belangrijk is de speculatieve methode. Tegelijkertijd zeer gevaarlijk en nuttig. Theorieën en hypothesen opstellen kan iedereen, ook in de landbouwwetenschap zijn er menschen, die voor elk feit zonder moeite een verklaring hebben. Men moet echter den juisten weg weten te kiezen, ook hier blijkt de afstand tusschen kunst en wetenschap niet groot te zijn.

Hoe vruchtbaar is de aanname van een aether in de physica geweest, de selectiehypothese van Darwin in de afstammingstheorie, de abstractie van de "geisoleerde staat" van v.Thünen in de landhuishoudkunde, enz., enz.. Zonder phantasie brengt men het niet ver. De wetenschappelijke onderzoeker heeft evenveel phantasie noodig als een goede componist of dichter. Een universeel genie als Goethe, moge hier als voorbeeld gelden. Een vooruitziende,

phantasievolle geest, zal door eenige gedurfde hypothesen de geheele chaos van feiten tot een groot organisch geheel aaneensmeden.

De beroemde mineraaltheorie van Justus von Liebig is niet op experimenteele wijze gevonden, zooals men algemeen denkt, maar geheel speculatief. Dat planten zich door anorganische stoffen kunnen voeden is pas later aangetoond door Boussingault e.a.. Slechts een scherpzinnig combinatievermogen en een rijke phantasie leidden Liebig tot het opstellen van de mineraaltheorie, die voor de toekomst zoo gewichtige, theoretische en practische gevolgen had. Al het heil is niet afkomstig van de experimenten.

Tot zoover deze hoofdstukken van Krzymowski.

HOOFDSTUK II.

INLEIDENDE PROBLEEMSTELLING.

§ 1. Het spel der factoren als centraal probleem.

Overzien wij de beschouwingen van Krzymowski, zooals zij in het voorgaande hoofdstuk weergegeven zijn, dan valt op, ten eerste dat
a. een zeer overheerschende plaats ingenomen wordt, door het vraagstuk van het factoren gewirwar (Faktorengewirr); dat
b. dit spel der factoren, zooals wij dit liever zullen noemen een onopgelost probleem gebleven is; dat
c. vele moeilijkheden die hier mee samenhangen door Krz. teruggevoerd worden op een door hem scherp bestreden rationalisme;
en verder ten tweede, dat de bestrijding van de experimenteele methode gezien de omstandigheden waarin de landbouw tegenwoordig verkeert, te ver gaat; thans is tooh dat juist van toepassing, waarop slechts even gewezen werd, nl. dat de experimenteele methode wel van groot belang is, als de omstandigheden zich wijzigen. Welnu, dat is juist in deze eeuw, in tegenstelling met de voorgaande 19e eeuw, het geval: nieuwe varieteiten, zelfs nieuwe gewassen, werden in nooit zoo grooten getale aangeboden, een stroom van nieuwe kunstmeststoffen komt op de markt enz. Daarnaast doet de uitvoering van groote cultuurtechnische werken de omstandigheden sterk wijzigen. Overheidsverordeningen zijn van sterken invloed op den intensiteitsgraad van het bedrijf en scheppen nieuwe omstandigheden. Tenslotte zij de aanwinst van groote stukken nieuw land (droogmakerijen, ontginningen) vermeld, waarop immers alle ervaring zelfs ontbreekt.

Nochtans moet toegegeven worden, dat het van belang is, dat er eens duidelijk gewezen wordt op de tekortkomingen van de experimenteele methode en dat het zeker niet aangaat, om de landbouwwetenschap een experimenteele wetenschap te noemen.

Keeren we terug tot het eerste standpunt, dan zien we, dat de beheersching van het spel der factoren het onopgelost probleem gebleven is, waar feitelijk alle andere punten omheendraaien. Toch meent schrijver gerechten twijfel te mogen koesteren, dat dit probleem geheel onoplosbaar zou zijn. Een van de voornaamste punten waarop Krz. strandt bij dit probleem, is de wederkeerige invloed van de factoren op elkaar. Evenwel heeft R.A.Fisher hier met het zgn. "factorial design" een weg gewezen, om bij weinige factoren deze wisselwerkingen te analyseeren. In ieder geval geeft Krz. geen overtuigend bewijs van de onoplosbaarheid. Deze onzekerheid over de onoplosbaarheid van het probleem tot beheersching van het spel der factoren is zeer teleurstellend:
1e. komt daardoor de critiek tegen het rationalisme minder tot haar recht, men blijft onzeker over de werkelijke plaats, die het rationalisme in de landbouwwetenschap toekomt, waardoor het ook moeilijker wordt de beteekenis van nieuwe richtingen in de philosophie voor de landbouwwetenschap te beoordeelen;
2e. de waarde der experimenteele methode zou zeer sterk stijgen, indien dit probleem opgelost kon worden of zoo men toteen meer beperkte oplossing kwam, was in ieder geval de waarde der experimenteele methode in de landbouwwetenschap beter te bepalen.

De hoofdopgave van dit geschrift zij derhalve gewijd aan de oplossing van het probleem, dat het spel der factoren ons stelt, en indien mocht blijken dat het slechts gedeeltelijk op te lossen is, de grenzen aan te geven tot hoever een eventueele methode kan rijken.

§ 2. Omschrijving der methode.

We zagen reeds, dat Krzymowski zich liet leiden door het meer Europeesche standpunt, waarin de verhouding tusschen het denken en de werkelijkheid steeds weer het hoofdprobleem was. Een belangrijke rol hierin spelen de kentheoretische beschouwingen (Krz. sprak van methodologische principes en kenniscritiek). We willen op de vele theorieën[1]) hier niet ingaan, slechts een enkel punt belichten, dat ons een indeeling der wetenschappen volgens de methoden aan de hand doet. Waar wij een methode zoeken, kan op deze manier haar toepasselijkheid in de landbouwwetenschap, die veelal als een conglomeraat van allerlei weten-

[1]) Dr.L.de Raeymaecker. Inleiding tot de wijsbegeerte en het Thomisme.

schappen wordt voorgesteld (men spreekt van het encyclopaedisch karakter der landbouwwetenschap), beter begrepen worden. Vragen we naar den oorsprong van onze kennis (wetenschap), dan dient opgemerkt te worden, dat de strijd tusschen het rationalisme en het empirisme in zeker opzicht beslecht werd door Kant, met het invoeren van de aprierische beginselen, nl. de vormen der aanschouwing en de categorieën van het verstand. Daarbij heeft iedere wetenschap haar specifiek eigen categorieën. Als na 1870 de philosophie herleeft, begint men met het kenprobleem, wat ons hier in het bijzonder interesseert: er ontwikkelen zich tal van zgn. neo-kantiaansche scholen, waarin men echter de categorieën op verschillende manieren gaat interpreteeren. Bovendien zijn er natuurlijk nog wel andere gezichtspunten bijgekomen. Zoo is bv. van de reeds eerder aangehaalde Amerikaansche philosophen het zgn. radicaal-empirisme afkomstig, waarover in Hoofdstuk VI meer. Bij het eerstgenoemde standpunt is kennen onderscheiden met behulp van categorieën, waardoor de verschillende wetenschappen ontwikkeld kunnen worden en waardoor aan iedere wetenschap haar eigen functie, een methodische bijzonderheid, gegeven wordt. Voor de natuurwetenschappen zijn de causaliteit en de meetbaarheid de centrale categorieën; zij vragen naar de wijze waarop iets (regelmatig) geschiedt, op grond der ervaring. Zij voeren de verschijnselen terug tot een wetmatig causaal samenhangend meetbaar stelsel. In deze meetbaarheid ligt de ontwikkeling tot mathematische formuleering besloten (Geheel andere categorieën ontwikkelen en bepalen de geesteswetenschappen, de psychologie, enz.). Het is duidelijk dat in velerlei wetenschappen in zeker opzicht van een spel der factoren gesproken kan worden, gezien het universeele gebruik van het woord factor, doch het door Krzymowski gebruikte factorbegrip laat geen twijfel over (zooals uit de aangehaalde hoofdstukken kan blijken) dat de te zoeken methode, die het spel der factoren beheerschen moet, een natuurwetenschappelijke moet zijn. Dit houdt dus ook in, dat de straks te geven methode alleen toegepast kan worden in de landbouwwetenschap op natuurwetenschappelijke basis, in theorie een zeer groote beperking dus, in feite valt er evenwel bijna de geheele landbouwwetenschap onder. Zooals reeds gezegd, spreekt men nl. veelal van het encyclopaedisch karakter van de landbouwwetenschap, met als hoofdschotel een conglomeraat van natuurwetenschappelijke vakken. (Ook in dit opzicht laat zich een hiaat in de philosophie der landbouwwetenschap constateeren. Men stelt meestal het encyclopaedisch karakter a priori voorop en maakt vervolgens verschillende onderverdeelingen in de landbouwwetenschap, ieder naar eigen inzicht. Nooit is men van het meer philosophische, hierboven geïndiceerde standpunt uitgegaan, nl. de landbouwwetenschap als geheel ontwikkelen met behulp van bepaalde categorieën, waarbij iedere bijzondere wetenschap een "onbegrijpelijk restant overlaat", dat voor haar bijzondere methode niet toegankelijk is; in meer abstracte termen: ieder bijzonder begrip, drijft het andere voor zich zelf uit, waarbij een voortschrijden van het enkelvoudige onverschillige, naar het kleurrijke en kwalitatieve, van het schematische naar het rijkere: voor natuurkundigen blijft een begrip als het "leven" der natuur het onbegrijpelijke, zij geeft dit begrip door aan de biologen, welke nu haar reeks categorieën tot ontplooiing brengen, en zoo gaat het voort.)

Tegen de hegemonie van de natuurwetenschappelijke beschouwingswijze in de 19e eeuw, die met deze categorieën causaliteit en meetbaarheid, alles meende te kunnen verklaren, waarbij de natuurwetenschap opgevat werd als een op alles toepasselijk deterministisch mechanisme werd heftig geageerd, men denke slechts aan Bergson en vele anderen; in de biologie ageerde Driesch hiertegen met het neo-vitalisme. We kunnen deze, met elkaar strijdende standpunten ieder hun eigen waarde geven als we de reeds genoemde opvatting huldigen, dat de causaliteit alleen een natuurwetenschappelijke categorie is, die alleen in dat gebied heerscht, geldigheid heeft, waarmede een gegeven onderwerp echter nog niet uigeput is, doch waarop thans met hetzelfde recht andere categorieën tot ontwikkeling gebracht kunnen worden.

Overigens zij opgemerkt, dat het recht van een bestaan van een kennistheorie door sommigen ook wel bestreden wordt (positivisten). Bekend in dit opzicht tegen den strijd tegen het neo-kantianisme zijn Mach en Avenarius.

Maar ook binnen het terrein der natuurwetenschappen wordt de causaliteit zeer verschillend geïnterpreteerd: Vaihinger spreekt van on-

ontbeerlijke, doch niet te wettigen grondprincipes, van louter fictie; Mach denkt de causaliteit een economische functie toe, ze levert een vereenvoudigd beeld der ervaring en is een middel om onze krachten te sparen en doelmatig handelen mogelijk te maken; Henri Poincaré ziet in de natuurwetenschappelijke grondprincipes conventioneele stellingen, waarvan dan logisch de gevolgtrekkingen afgeleid worden.

Tegenwoordig spreekt men ook wel van een statistische causaliteit. De taak der natuurwetenschappen is dan de verschijnselen met de onderling afhankelijke causale factoren te beschrijven, welke beschrijving inzicht geeft, terwijl men gerechtigd is onder de gegeven omstandigheden een herhaling van de verschijnselen met zekere waarschijnlijkheid te verwachten. Wij zullen ons in het bijzonder nog bezig houden met het onderzoek, tot welke grenzen men met deze beschrijving kan gaan (Hoofdstuk VII).

Het is hier niet de plaats op dit causaliteitsbeginsel in het bijzonder verder in te gaan, waar het door alle eeuwen heen de menschheid heeft beziggehouden, waarbij men tot zeer uiteenloopende opvattingen kwam.

Doch als natuurwetenschappelijke methode is de te geven oplossing nog niet voldoende scherp afgebakend. Wel hebben we reeds het begrip causaliteit ingevoerd, doch we willen ook de meetbaarheid in het bijzonder op den voorgrond brengen. Zonder dit laatste kan ook een natuurwetenschappelijke oplossing mogelijk geacht worden, nl. door een zgn. synthetische methode, die kort gezegd op het volgende neerkomt: de werkelijkheid wordt vervalscht door te trachten deze uit de proefstelsels, die slechts één factor onderzoeken, alsof er geen pluricausaliteit en wisselwerkingen bestonden, op te bouwen. Waar men geen kans zag de wisselwerkingen te beheerschen, zocht men een synthetische methode. Vooral Prof. Jordan trachtte dit in de biologie te bereiken door een geordende veeloorzakelijkheid voorop te stellen en dit te ontwikkelen met begrippen als amboceptorverband, polyceptorverband en panteceptorverband voor verschillende causale structuren (Hij onderscheidt feitelijk een factorencausaliteit en een relatiecausaliteit). De hier te geven oplossing voor het factorenspel legt het zwaartepunt juist op het onderzoek en meting dezer wisselwerkingen zelve, waarmede a.h.w. getracht wordt de analytische methode op een hooger plan te brengen. In het natuurwetenschappelijke (causaliteit en meetbaarheid) ligt de aanleiding tot een formeel wiskundige behandeling, en wel met behulp van het zgn. "factorial design".

HOOFDSTUK III

CETERIS PARIBUS EN FACTORIAL DESIGN.

§ 1. Het factorial design als efficient en volledig ontwerp en als inductieve basis voor generalisatie.

Het factorial design is in zekere opzichten een principe dat tegenover het bekende "ceteris paribus" gesteld moet worden, aan den anderen kant als een verdere ontwikkeling van het ceteris paribus gezien kan worden. Om dit duidelijk te maken zij het volgende aan Fisher [1]) ontleend:

In de experimenteele wetenschap in het algemeen vindt men vaak, dat groote nadruk gelegd wordt op het feit, dat men bij het onderzoek alle factoren op een na, constant gehouden heeft. M.a.w. dat men het ceteris paribus principe gevolgd heeft. De onderzoekergeinteresseerd in de oorzaken, die tot een zeker verschijnsel bijdragen, isoleert door abstractie deze oorzaken in een aantal elementaire factoren, waarbij men dan vervolgens vaak meent, dat het onderzoek van een dezer factoren volgens het bovengenoemde principe van ceteris paribus, de essentieele wetenschappelijke methode van een experimenteel onderzoek is.

Voor demonstratiedoeleinden kan deze methode zijn voordeelen hebben, in een gevordard onderzoek echter is deze simpele voorstelling onvoldoende. We weten in het algemeen niet of een factor onafhankelijk van alle andere factoren zal werken of dat een eventueele afhankelijkheid eenvoudige regels volgt; integendeel, als we enkele factoren kiezen voor onderzoek, doen we dit niet, omdat we denken, dat de natuurwetten geformuleerd kunnen worden op eenvoudige wijze uit deze paar variabelen, maar omdat dit variabelen zijn, die we in onze macht hebben of die op eenvoudige wijze gemeten kunnen worden. De onderzoeker die zich dan beperkt tot een factor, is het ongelukkige slachtoffer van een doctrinaire theorie. Laten we dit aan een voorbeeld duidelijk maken. We willen experimenteel het effect onderzoek veroorzaakt door variatie van 4 bestanddeelen, die een zekere alliage vormen. Voor ons voorbeeld doet het er in principe niet toe, hoe groot deze variatie is; laten we voor elk bestanddeel volstaan met een kleine en met een groote hoeveelheid, die voor elk ingredient nog een verschillende grootte mogen hebben; ook zouden we eventueel voor één of meer der ingredienten het al of niet toevoegen kunnen aannemen. Hoe het ook zij, de vraag welke combinatie de beste alliage geeft kan beantwoord worden door alle 16 mogelijke combinaties te vormen. Laat elke combinatie 6x getoetst worden op zijn bruikbaarheid als alliage; dit geeft dus 96 tests totaal.

Ten opzichte van elk der 4 ingredienten kunnen we dan opmerken dat elk ingredient 48x in een groote hoeveelheid en 48x in een kleine hoeveelheid getest wordt. Deze 48 gevallen zijn niet alle aan elkaar gelijk, zij zijn gelijk in reeksen van 6, maar deze 8 reeksen van 6 verschillen onderling. Correspondeerend met elke reeks van 6 waarin een groote hoeveelheid van de eerste factor gebruikt is, is er een reeks van 6 met de kleine hoeveelheid van de eerste factor en verder precies gelijk t.o.v. de overige factoren. Het verschil in deze twee reeksen moet, afgezien van toevallige afwijkingen, toegeschreven worden aan het ingredient waarin zij verschillen. Nu zijn er totaal 8 paren van dergelijke reeksen, waarin dit effect onderzocht kan worden!

Voor elke factor, hebben we derhalve een directe vergelijking van de gemiddelden (of totalen) van 2 reeksen van 48 proeven. En deze vergelijking heeft dezelfde precisie, alsof alle 96 proeven gewijd waren aan het onderzoek van alleen deze ene factor. In werkelijkheid worden echter in deze 96 tests 4 factoren met deze zelfde precisie onderzocht! Het ontwerp, dat alle combinaties van de trappen waarin de factoren onderzocht worden omvat, door Fisher het Factorial design genoemd, blijkt dus reeds een van groote efficiency te zijn. Genoemde 4 verschillen noemt Fisher de main-effects. [2])

De voordeelen van het factorial design boven een groep experimenten, elk ontworpen om slechts één enkele factor te onderzoeken, zijn echter nog veel grooter. Want met deze afzonderlijke experimenten hadden we geen inlichtingen

---

[1]) Prof.R.A.Fisher. The Design of Experiments. Chapter VI Edinburgh 1935.

[2]) Voor de meeste uitdrukkingen die hier gebruikt zullen worden, bestaan nog geen vaststaande Nederlandsche woorden, zoodat meestal het Engelsche woord is aangehouden in de tekst.

verkregen over de eventueele interacties tusschen de verschillende factoren, ook niet, al hadden we elke factor in 96 tests onderzocht, zoodat we dezelfde precisie verkregen hadden, als volgens het vermeld factorial design. Als bv. een stijging in ingredient A van voordeel was, in tegenwoordigheid van B maar een nadeel of ineffectief bij afwezigheid van B c.q. bij een groote dan wel een kleine hoeveelheid van B, dan kunnen we dit alleen te weten komen door het onderzoek op A werkelijk uit te voeren bij aan- of afwezigheid van B en dat is juist wat het factorial design doet! Maar bovendien zoo grondig doet, dat alle mogelijke interacties onderzocht worden! Om dit effect van B op A te onderzoeken vergelijken we het verschil in factor A bij 24 paren onder aanwezigheid van factor B met het verschil in factor A bij de overige 24 paren, waarin factor B afwezig is (of in kleine hoeveelheid aanwezig is). Het verschil van deze 2 verschillen drukt den invloed van factor B op factor A uit, of wat we ook mogen zeggen de invloed van factor A op factor B. Deze invloed noemen we de interactie tusschen A en B. Deze interactie wordt weer met dezelfde precisie vastgesteld als de main-effects der 4 factoren als boven beschreven. We komen nl. tot dezelfde uitdrukking als we het verschil beschouwen tusschen 48 objecten waarin A en B beide in een groote hoeveelheid, dan wel beide in een kleine hoeveelheid voorkomen en 48 objecten waarin een groote hoeveelheid van A samengaat met een kleine hoeveelheid van B of omgekeerd.

Er zijn totaal 6 van dege interacties tusschen de 4 factoren en al deze 6 worden met dezelfde precisie vastgesteld, als het experiment volgens het factorial design ontworpen is!

Deze 4 main-effects en 6 interacties putten het materiaal echter nog niet uit. Het is mogelijk dat de waarde van de interactie tusschen de factoren A en B verandert wordt door den 3en factor C. Deze interactie stellen we vast door de interactie van A en B te vergelijken bij de objecten met factor C in een groote hoeveelheid met de interactie van A en B bij een kleine hoeveelheid van C. In feite is dit hetzelfde als het verschil tusschen alle 48 objecten die 3 of 1 der ingredienten A,B en C in de groote hoeveelheid hebben en alle 48 objecten die 2 of geen dezer hoeveelheden hebben, zoodat ook deze interactie van de 2e orde met dezelfde precisie vastgesteld wordt. Even goed mogen we zeggen dat factor B invloed heeft op de interactie tusschen A en C of dat de factor A invloed heeft op de interactie tusschen B en C. Het is duidelijk dat er in ons voorbeeld 4 dergelijke interacties, of interacties van de 2e orde zijn. Tenslotte kan de factor D nog invloed op de interactie van de 2e orde tusschen de factoren A,B en C hebben. Er is natuurlijk slechts 1 dergelijke interactie van de 3e orde. De 15 onafhankelijke verschillen die we dus tusschen de 16 mogelijke combinaties kunnen maken, vallen uiteen in de volgende 15 componenten: main-effects 4

interacties van de 1e orde 6
" " " 2e " 4
" " " 3e " 1

Tot nog toe werd op twee groote voordeelen van het factorial design gewezen: 1e groote efficiency: het effect van elken factor wordt thans bij eenzelfde precisie met slechts $\frac{1}{4}$ der waarnemingen bereikt, die anders noodig waren, als elke factor apart onderzocht werd volgens het principe van ceteris paribus, 2e het is veel omvattender, het geeft ons veel meer inlichtingen, nl. de interacties (in ons voorbeeld 11 stuks) het put het gegeven waarnemingsmateriaal volledig uit. Er is echter nog een 3e voordeel, dat vooral van betekenis is, als men de resultaten van het experiment wil uitdragen in de practijk, nl. dat de conclusie zooals b.v. dat het voordeelig is, de hoeveelheid van een der ingredienten te verhoogen, een breedere inductieve basis heeft, wanneer het geconcludeerd wordt uit een experiment, waarin de andere factoren gevarieerd zijn, dan wanneer het geconcludeerd wordt uit een experiment, waarin de andere factoren strikt constant gehouden zijn. Een standarisatie van de omstandigheden waaronder een proef genomen wordt beperkt de geldigheid van de gevolgtrekkingen tot een nauw gebied van bijzondere omstandigheden.

Aldus Fisher. De uitdrukking ceteris paribus is door schrijver toegevoegd in het bovenstaande; Fisher laat dat principe onbenoemd. Schrijver hoopt zoodoende de tegenstelling met de nieuwere exact geformuleerde werkwijze van Fisher duidelijker te doen uitkomen. Men moet zich er echter wel van bewust zijn, dat in het factorial design, waarbij vele factoren kunnen varieeren, de niet onderzochte factoren voor elk object echter constant gehouden worden; het factorial design zou dus op te vatten zijn als een ceteris paribus van hoogere

orde, we willen hier echter de uitdrukking factorial design van Fisher hiervoor behouden.[1])

Volgen wij Fisher inmiddels nog weer een oogenblik: Er is nog een ander punt, waarin het factorial design de onderzoeker helpen kan: onderzocht hij slechts een factor, en houdt hij de overige constant, dan is dit constant zijn vaak slechts schijn: nemen we b.v. een proef met 2 varieteiten van een gewas. Op het eerste gezicht lijkt het eenvoudig om alle overige factoren voor beide varieteiten gelijk te maken, doch dit is niet zoo, zoo moet b.v. de hoeveelheid zaaizaad gelijk zijn. Maar als deze hoeveelheid gemeten wordt in hl/ha, dan zal het gewicht aan zaaizaad, een evenzeer gebruikelijke maat, verschillend zijn. Een varieeren van de hoeveelheid, m.a.w. het opnemen van de hoeveelheid zaaizaad als een 2e factor in de proef brengt hier uitkomst. Tot zoover volgens Fisher.

Hieraan kunnen we toevoegen, dat ook de oude vraag, hoe 2 varieteiten te vergelijken, terwijl elke varieteit niet gelijk behandeld wordt, doch de voor hem meest gunstige behandeling ontvangt, alleen rationeel opgelost kan worden, door toepassing van het factorial design.

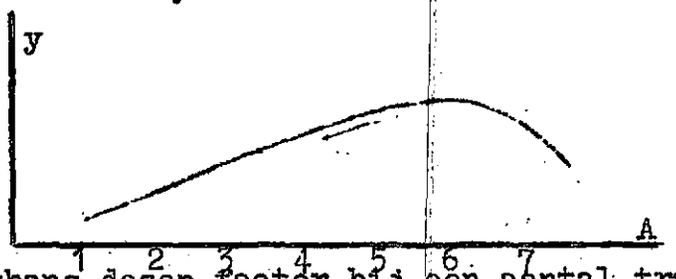
Zoo komt men op logische wijze van zelf tot proeven in enkelvoud, waarin alle herhalingen vervangen zijn, door andere factoren in het onderzoek op te nemen. We spreken in het factorial design dan ook van z.g. verborgen herhalingen (hidden replication). Om goed vertrouwd te geraken met het factorial design, volgen hier nog eenige voorbeelden en andere details in de volgende §.

§ 2. Afbeelding van interacties

De landbouwkundige, die gewend is, het begrip factor te zien tegen den achtergrond van een reactiecurve of opbrengstcurve, zal misschien niet dadelijk de beteekenis van bovenvermelde abstracte behandeling van interacties vatten, voor de reactiecurven.

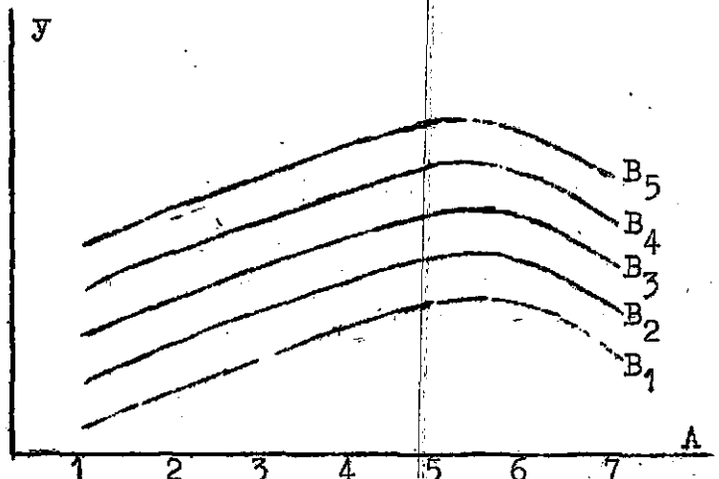
Gesteld we onderzoeken een factor A in 7 trappen bij een gegeven trap van den factor B, we krijgen b.v. de volgende opbrengstcurve:

fig.1



We onderzoeken thans dezen factor bij een aantal trappen, b.v. 5 trappen van B, dus we passen een factorial design toe, totaal bestaande uit 7 x 5 = 35 combinaties. We krijgen dan het volgende beeld:

fig.2



[1]) Het idee van het factorial design leefde natuurlijk reeds veel langer, doch het was Fisher, die dit idee in het volle licht bracht, maar pas in het licht kon brengen, nadat hij in de analysis of variance het instrument gevonden had, om de op die wijze verkregen uitkomsten exact te kunnen interpreteeren, hoeveel factoren er ook in opgenomen mochten zijn, waarna als vanzelfsprekend de noodzakelijkheid ontstond, dit idee door een naam, het factorial design, scherp te kenmerken en zoo tot uitdrukking te brengen, wat reeds lang aan velen voorzweefde.

De vorm der reactiecurve is bij elken trap van B dezelfde gebleven; het eenige verschil is dat alle waarden die de reactiecurve ons geeft, met een constant bedrag vermeerderd zijn. De werkingen van factor A en B zijn additief; er bestaat geen wisselwerking of interactie tusschen deze 2 factoren, de factoren zijn onafhankelijk van elkaar (Natuurlijk zoolang we alle andere factoren constant houden).

We hadden ook de volgende beelden kunnen verkrijgen:

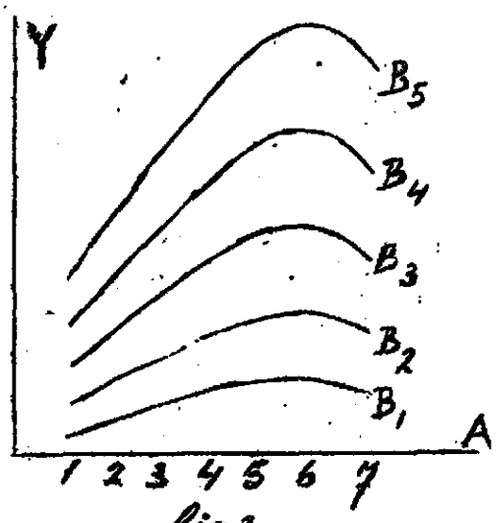


fig 3

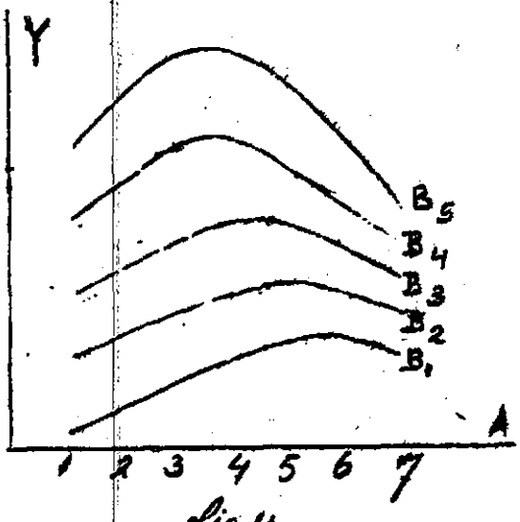
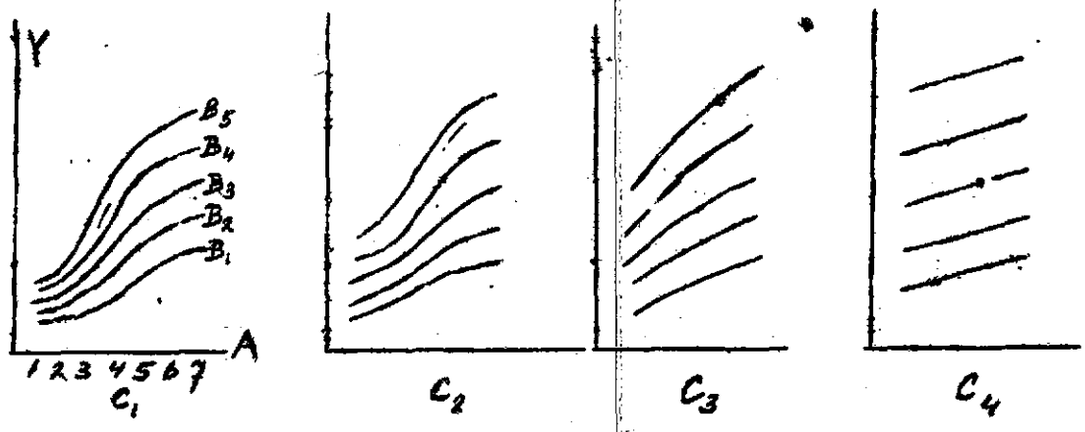


fig 4

In fig. 3 blijft het optimum van factor A op dezelfde plaats, maar voor elken trap van B is de toevoeging niet meer constant. Deze aangroeiing is voor de optimumwaarden het grootst en deze toename neemt toe voor hoogere waarden van factor B: er is een wisselwerking tusschen de factoren A en B: de toename door B veroorzaakt is afhankelijk van de waarde van factor A, of omgekeerd de toename door factor A is afhankelijk van de waarde van factor B.

In fig. 4 hebben we hetzelfde geval, maar hier verandert bovendien nog het optimum van plaats. De wisselwerking is wat ingewikkelder.

Voegen we nog een 3en factor C toe in b.v. 4 trappen. Het factorial design bestaat dan uit 7 x 5 x 4 = 140 combinaties. Er zijn 3 interacties van de 1e orde, en een interactie van de 2e orde.



Voor 3 van de 4 trappen van den factor C zien we een duidelijke interactie tusschen de factoren A en B, maar deze interactie verandert blijkbaar als C verandert: er is een interactie van de 2e orde!

De interacties tusschen de factoren A en C eh de factoren B en C laten zich niet zoo gemakkelijk uit deze figuur aflezen. We kunnen hierin slagen, door de graphische voorstelling te vervormen door b.v. elke graphiek de factoren B en C te laten weergeven, zoodat we dan 7 van dergelijke voorstellingen, n.l. één voor elken trap van A, naast elkaar verkrijgen.

Spreken we voorloopig af, dat we het woord wisselwerking door interactie zullen vervangen, dan vindt men hier dus een iets minder abstracte voorstelling van de beteekenis der interacties.

Het volgende voorbeeld, dat de belangrijkheid van het factorial design demonstreert, zij ontleend aan Yates[1]). Er wordt verondersteld dat we een nieuw gewas in een land of streek willen invoeren en dat daarom natuurlijk weinig of niets bekend is van de meest gunstige varieteiten, soort en hoeveelheid van bemesting, cultuurmethodes etc. De klassieke methode zou zijn te werken volgens het principe van ceteris paribus, d.w.z. aparte proeven, waarin de beste varieteit, de beste zaaitijd, de beste N-bemesting enz. gevonden wordt. Maar we kunnen geen bemestingsproef nemen, zonder eerst een varieteit te kiezen, waarmee we deze proef zullen uitvoeren, noch kunnen we een varieteiten proef nemen, zonder eerst een bepaalde bemesting te kiezen! Nu is het mogelijk en in den regel is dit ook werkelijk zoo, dat het effect van de bemesting op de diverse varieteiten verschillend is. Dus de conclusie die we na veel tijd en met veel geld verkregen hebben over de optimale bemesting voor de varieteit die we gekozen hebben in de proef, is misschien in het geheel niet geldig voor de varieteit die we uiteindelijk kiezen; of deze varieteit is ten onrechte gekozen, omdat we niet gelet hebben op de mogelijkheid om de opbrengst van de andere varieteiten te kunnen verhoogen, door een andere bemesting te kiezen. Alleen het gebruik van het factorial design geeft hier den eenigen juisten weg aan.

Gingen we in het vorige voorbeeld uit van een experiment, ook in de verwerking van waarnemingen is het idee van het factorial design van groot nut. Zoo b.v. het volgende geval.

Met een organisme werden eenige assimilatieproeven genomen. Nadat de cultures in evenwicht gekomen waren, werd een waarde, die de omvang van de cultures c.q. van de gezamenlijke organismen op een of andere wijze weergaf afgezet als onafhankelijke x, tegen de hoeveelheid dissimilatie producten, als afhankelijk veranderlijke y. Gewerkt werd volgens het ceteris paribus principe, waarbij o.a. de temperatuur constant gehouden werd door de waterleidingstroom. Deze proef werd meerdere keeren na elkaar herhaald, omdat men een groot traject van de onafhankelijk veranderlijke x wenschte en deze moeilijk kon verkrijgen, aangezien men afhankelijk was van den evenwichtstoestand die intrad. In totaal werden 4 proeven achter elkaar genomen, waarbij de volgende uitkomsten verkregen werden:

1e proef		2e proef		3e proef		4e proef	
x	y	x	y	x	y	x	y
1	6	2.4	15	3.3	22	5.3	41
2.8	15	3.4	19	6.9	40	5.8	45
4.5	23	5.7	30			7.7	55
		8.3	37			9.4	59

De onderzoeker zette deze gegevens in een graphiek uit en trok hierdoor een kromlijnige regressiecurve.

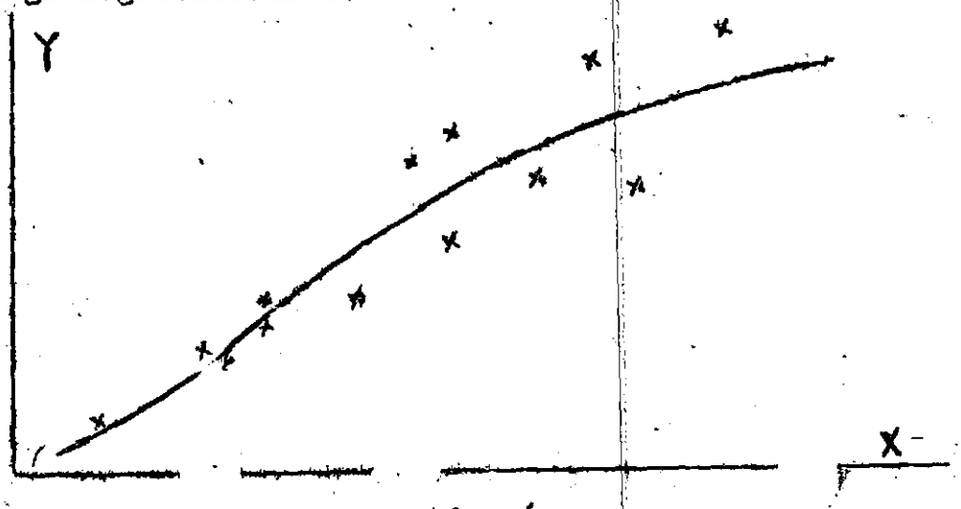


fig 6.

[1]) F.Yates. The design and analysis of factorial experiments. Imperial Bureau Soil Science Tech. Comm. No. 35 1937.

De hierbij gevonden groote spreiding werd als teleurstellend ondervonden en op allerlei wijzen werd getracht de inrichting der proef te verbeteren om zoodoende de spreiding om de regressielijn te verkleinen.

Doch deze spreiding is slechts schijn. Dit blijkt als men de 4 proeven gescheiden houdt en alleen die punten met elkaar verbindt, die een proef vormen. We krijgen dan de volgende voorstelling.

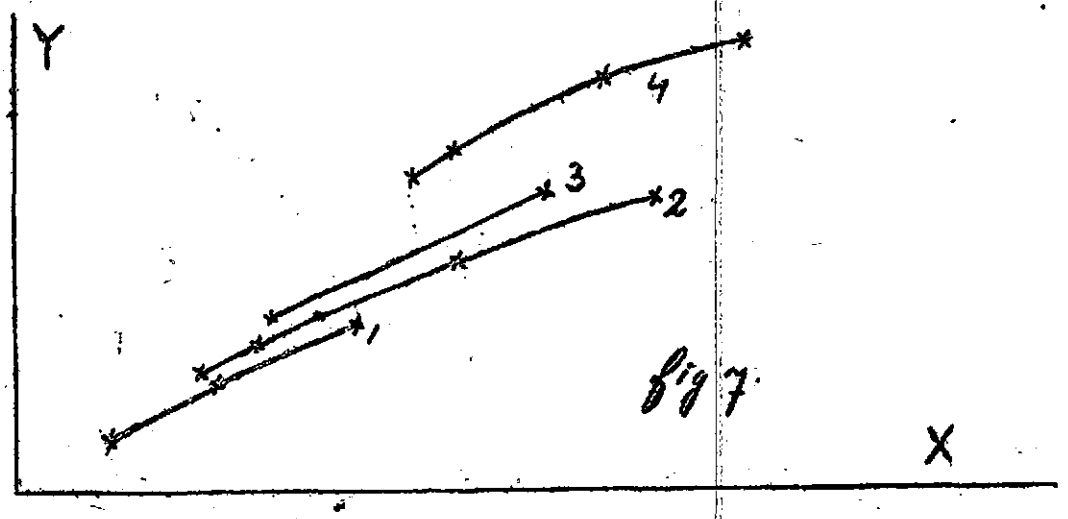


fig 7.

Het is niet moeilijk deze figuur op de volgende manier, als hypothese, eens uit te breiden.

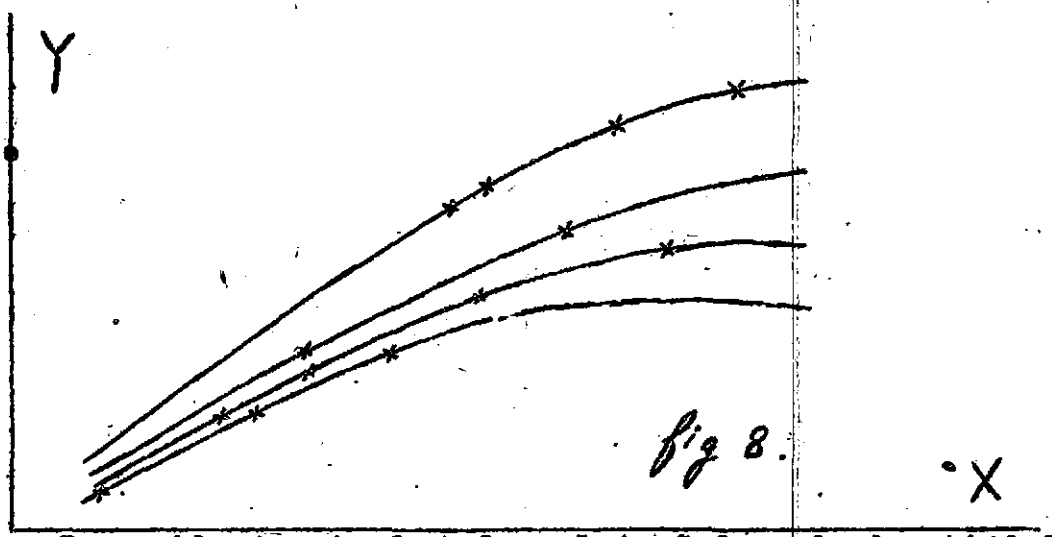


fig 8.

De verklaring is dan als volgt: Gedurende den tijd dat de 4 proeven genomen werden, is de temperatuur van de waterleiding gestegen. Zooals bekend is de invloed van de temperatuur op assismilatieprocessen zeer groot. Met het factorial design voor oogen interpreteeren we thans de proef alsof hij genomen was bij 4 verschillende temperaturen en komen zoo tot de laatste figuur. De spreiding is thans nihil, en zou men de proef nog eens willen herhalen, dan moet men doelbewust de temperatuur als 2e factor opnemen, in plaats van de proef b.v. 4 x te herhalen. Kortom men moet het factorial design toepassen en het oude principe ceteris paribus laten schieten. (Was er nog wel een belangrijke spreiding overgebleven dan moet men nog een 3e of een 4e factor opnemen en alle combinaties consequent uitvoeren).

§ 3. Mathematische formuleering der main-effects en interacties.

Hoe meer factoren men in zijn experimenten opneemt, of tracht te isoleeren uit het waarnemingsmeteriaal, hoe moeilijker het wordt ons dit voor te stellen. We zagen bij het voorbeeld der reactiecurven, dat bij 3 factoren we reeds een aantal graphieken naast elkaar krijgen (Het resultaat van 2 factoren kunnen we nog op continue wijze in é n graphiek, een 3-dimensionale figuur aanschouwelijk voorstellen). Hoe meer factoren des te sneller ook het aantal interacties stijgt, die zeer moeilijk voor te stellen zijn. Bij veel factoren zegt ons de graphische voorstelling van diagrammen naast elkaar bij 3 factoren, van series van diagrammen naast elkaar bij 4 factoren, van complexen naast elkaar bij 5 factoren enz. enz. ook niets meer, en het eenige

wat ons overblijft is een wiskundige formuleering!

Dit kan ons trouwens niet verwonderen, het begrip weetbaarheid in de natuurwetenschap voert vanzelf tot het functiebegrip, terwijl het functiebegrip de mogelijkheid schept van een doelmatige mathematiseering. Deze opvatting is toch ook algemeen verspreid en geaccepteerd. Henri Bergson[1]) zegt: ..... de natuurwetenschappen wenden zich af van het individuele en werkelijke in een wereld van toenemende wiskundige abstracties .....

Waar ons een factorial design voor ogen staat in het bijzonder met zeer vele factoren, waarmee steeds een progressief stijgend aantal interacties gepaard gaat, is het noodzakelijk deze interacties op ondubbelzinnige wijze tot in de kleinste finesses wiskundig vast te leggen, om zich een indruk te kunnen vormen over hun beteekenis en werking, zoodat dan ook de geheele rest van dit hoofdstuk hieraan gewijd is.

Het eenvoudigst denkbare geval is dat van 2 factoren, waarbij elke factor zich slechts op 2 wijze kan uiten, een alternatief, of anders gezegd: elke factor komt in 2 trappen voor.[11])

Noemen de we factoren A en B en de trappen 1 en 2, dan krijgen we dus de volgende 4 objecten A_1B_1, A_2B_1, A_1B_2 en A_2B_2. Nu zijn er dus 2 reacties op den factor A, een bij B_1 n.l. $A_2B_1 - A_1B_1$ en een bij B_2 n.l. $A_2B_2 - A_1B_2$. Deze 2 reacties kunnen van elkaar verschillen, maar soms is dit verschil gering, en dan niet te onderscheiden van toevallige fouten; in zulke gevallen is het voldoende de gemiddelde reactie te beschouwen. Deze gemiddelde reactie, main-effect dus, is $M(A) = \frac{1}{2}\{(A_2B_2-A_1B_2) + (A_2B_1-A_1B_1)\}$
ovenzoo is het main-effect van factor B

$$M(B) = \tfrac{1}{2}\{(A_2B_2-A_2B_1) + (A_1B_2-A_1B_1)\}$$

We kunnen nu ook de interactie definieeren, deze is het verschil tusschen beide reacties op factor A, bij B_1 en B_2 dus

$$I(AB) = \tfrac{1}{2}\{(A_2B_2-A_1B_2) - (A_2B_1-A_1B_1)\}$$

We kunnen de interactie ook definieeren als het verschil tusschen beide reacties op B bij A_1 en A_2 we vinden dan

$$I(AB) = \tfrac{1}{2}\{(A_2B_2-A_2B_1) - (A_1B_2-A_1B_1)\}$$

wat gelijk is aan den vorigen vorm, er is dus slechts een interactie tusschen de factoren A en B, conform onze vroegere beschrijving van het experiment met de 4 ingredienten.

(Voor het jaar 1935 werd de interactie gedifinieerd zonder de coefficient $\frac{1}{2}$, de invoering van deze coefficient brengt tot stand 1e dat de standaardafwijking van alle main-effects en interacties van 2 x 2 x 2 x ..... experimenten nu gelijk zijn (aangezien elke waarneming een zelfde middelbare fout toegekend wordt); 2e dat de reactie op een factor met betrekking tot een willekeurige combinatie der andere factoren, thans uit te drukken is als een som of verschil van de verschillende main-effects en interacties, zonder eenige coefficienten in te moeten voeren. Zoo gelden in een 2 x 2 experiment de volgende relaties:
reactie op A bij B_1 = $A_2B_1 - A_1B_1 = M(A) - I(AB)$
reactie op A bij B_2 = $A_2B_2 - A_1B_2 = M(A) + I(AB)$
de reactie $A_2B_2 - A_1B_1 = M(A) + M(B)$.).

Bij 3 factoren, dus bij een 2 x 2 x 2 experiment, hebben we de volgende 8 combinaties: $A_1B_1C_1$ $A_2B_1C_1$ $A_1B_2C_1$ $A_2B_2C_1$ $A_1B_1C_2$ $A_2B_1C_2$ $A_1B_2C_2$ en $A_2B_2C_2$.

Voor de main-effects vinden we dan:

$$M(A)=\tfrac{1}{2}\left[\tfrac{1}{2}\{(A_2B_1C_1-A_1B_1C_1)+(A_2B_2C_1-A_1B_2C_1)\}+\tfrac{1}{2}\{(A_2B_1C_2-A_1B_1C_2)+(A_2B_2C_2-A_1B_2C_2)\}\right]$$
$$M(B)=\tfrac{1}{2}\left[\tfrac{1}{2}\{(A_1B_2C_1-A_1B_1C_1)+(A_2B_2C_1-A_2B_1C_1)\}+\tfrac{1}{2}\{(A_1B_2C_2-A_1B_1C_2)+(A_2B_2C_2-A_2B_1C_2)\}\right]$$
$$M(C)=\tfrac{1}{2}\left[\tfrac{1}{2}\{(A_1B_1C_2-A_1B_1C_1)+(A_2B_1C_2-A_2B_1C_1)\}+\tfrac{1}{2}\{(A_1B_2C_2-A_1B_2C_1)+(A_2B_2C_2-A_2B_2C_1)\}\right]$$

respectievelijk

$$M(A)=\tfrac{1}{4}\{(A_2B_1C_1+A_2B_2C_1+A_2B_1C_2+A_2B_2C_2)-(A_1B_1C_1+A_1B_2C_1+A_1B_1C_2+A_1B_2C_2)\}$$
$$M(B)=\tfrac{1}{4}\{(A_1B_2C_1+A_2B_2C_1+A_1B_2C_2+A_2B_2C_2)-(A_1B_1C_1+A_2B_1C_1+A_1B_1C_2+A_2B_1C_2)\}$$
$$M(C)=\tfrac{1}{4}\{(A_1B_1C_2+A_2B_1C_2+A_1B_2C_2+A_2B_2C_2)-(A_1B_1C_1+A_2B_1C_1+A_1B_2C_1+A_2B_2C_1)\}$$

dus het gemiddelde van de 4 reacties op A bij de combinaties B_1C_1 B_2C_1 B_1C_2 en B_2C_2, respectievelijk het gemiddelde van de reacties op B en C.

De interacties van de 1e orde wordt thans gedefinieerd als het gemiddelde van de 2 interacties AB bij C_1 en C_2 evenzoo de andere interacties van de 1e orde:

---

[1]) Henri Bergson: L'évolution créatice. Paris 1917.

[11]) Notes on the construction and Use of the Summary Tables. Rothamsted. Experimental Station. Report 1936.

$I(AB)=\frac{1}{2}[\frac{1}{2}\{(A_2B_1C_1-A_1B_1C_1)-(A_2B_2C_1-A_1B_2C_1)\}+\frac{1}{2}\{(A_2B_1C_2-A_1B_1C_2)-(A_2B_2C_2-A_1B_2C_2)\}$
$I(AC)=\frac{1}{2}[\frac{1}{2}\{(A_1B_1C_2-A_1B_1C_1)-(A_2B_1C_2-A_2B_1C_1)\}+\frac{1}{2}\{(A_1B_2C_2-A_1B_2C_1)-(A_2B_2C_2-A_2B_2C_1)\}$
$I(BC)=\frac{1}{2}[\frac{1}{2}\{(A_1B_2C_1-A_1B_1C_1)-(A_1B_2C_2-A_1B_1C_2)\}+\frac{1}{2}\{(A_2B_2C_1-A_2B_1C_1)-(A_2B_2C_2-A_2B_1C_2)\}$
respectievelijk:
$I(AB)=\frac{1}{4}\{(A_2B_1C_1+A_1B_2C_1+A_2B_1C_2+A_1B_2C_2)-(A_1B_1C_1+A_2B_2C_1+A_1B_1C_2+A_2B_2C_2)\}$
$I(AC)=\frac{1}{4}\{(A_1B_1C_2+A_2B_1C_1+A_1B_2C_2+A_2B_2C_1)-(A_1B_1C_1+A_2B_1C_2+A_1B_2C_1+A_2B_2C_2)\}$
$I(BC)=\frac{1}{4}\{(A_1B_2C_1+A_1B_1C_2+A_2B_2C_1+A_2B_1C_2)-(A_1B_1C_1+A_1B_2C_2+A_2B_1C_1+A_2B_2C_2)\}$

Terwijl de interactie van de 2e orde I(ABC) gedefinieerd wordt als het verschil tusschen de 2 interacties I(AB) van de 1e orde bij C_1 en C_2, of het verschil van de 2 interacties I(AC) bij B_1 en B_2, resp. I(BC) bij A_1 en A_2. Klaarblijkelijk is er slechts een interactie van de 2e orde:
$I(ABC)=\frac{1}{2}[\frac{1}{2}\{(A_2B_1C_1-A_1B_1C_1)-(A_2B_2C_1-A_1B_2C_1)\}-\frac{1}{2}\{(A_2B_1C_2-A_1B_1C_2)-(A_2B_2C_2-A_1B_2C_2)$
respectievelijk:
$I(ABC)=\frac{1}{4}\{(A_2B_1C_1+A_1B_2C_1+A_1B_1C_2+A_2B_2C_2)-(A_1B_1C_1+A_2B_2C_1+A_2B_1C_2+A_1B_2C_2)\}$

Deze laatste interactie is dus geen gemiddelde, evenzoo was bij 2 factoren de eenige interactie geen gemiddelde.

(Ook hier kunnen we de verschillende reacties weer uitdrukken in de main-effects en interacties b.v. reactie op A bij B_2 en C_2 is $A_2B_2C_2-A_1B_2C_2= M(A)-I(AB)-I(AC)+I(ABC)$ enz.)

Verder zien we, dat de 3 main-effects en de 4 interacties ook geschreven kunnen worden als het verschil tusschen 2 sommen van telkens 4 combinaties. Dit is conform de beschrijving van de proef met de vier ingredienten: Bij 4 factoren kunnen we de 4 main-effects en de 11 interacties schrijven als het verschil tusschen 2 sommen van telkens 8 combinaties; we werkten daar in 6-voud, dus als het verschil tusschen de som van telkens 48 combinaties.

Met bovenstaande formules kan nu elk geval quantitatief en ondubbelzinnig beschreven worden; de formules laten zich voor elk gewenscht aantal factoren uitbreiden.

Echter werken we meestal met meer dan 2 trappen. We moeten de formules nu dus uitbreiden voor meer trappen:

b.v. factor A wordt onderzocht in 3 trappen A_1, A_2 en A_3, dan noemen we $A_3-A_1=A_L$ linear response, terwijl het exces van de reactie A_3-A_2 over A_2-A_1 curvature genoemd wordt, dus $A_C=(A_3-A_2)-(A_2-A_1)=A_3-2A_2+A_1$. Bestaat tusschen A_3 A_2 en A_1 een lineair verband dan is dus de curvature $A_C=0$.

Met deze 2 uitdrukkingen, die de 2 vrijheidsgraden van factor A kunnen vertegenwoordigen, zijn de reacties thans als volgt te beschrijven $A_2-A_1=\frac{1}{2}(A_L-A_C)$, $A_3-A_2=\frac{1}{2}(A_L+A_C)$.

Bij 2 factoren, elk in 3 trappen zijn de linear response en de curvature de gemiddelden over alle 3 trappen van den anderen factor.

De interactie tusschen de linear responses kan precies als in het voorgaande gedefinieerd worden dus $I(A_LB_L)=\frac{1}{2}\{(A_3B_3-A_1B_3)-(A_3B_1-A_1B_1)\}$.

De andere 3 componenten van de interactie, die immers 4 vrijheidsgraden heeft, kunnen op analoge wijze gedefinieerd worden, $I(A_CB_L)$, $I(A_LB_C)$ en $I(A_CB_C)$.

Bij 3 factoren elk in 3 trappen bestaat de interactie van de 2e orde thans uit 8 componenten inplaats van 1, zooals in het 2 x 2 x 2 experiment. De belangrijkste daarvan is weer de interactie van de 2e orde tusschen de linear responses $I(A_LB_LC_L)$.

Worden de factoren in nog meer dan 3 trappen onderzocht, wat vaak het geval is, dan kunnen op een dergelijke wijze de vrijheidsgraden gesplitst worden, waarbij ook verschillende mogelijkheden zich voordoen, die verdeelingen moeten dan correspondeeren met een logische simplificatie van de reacties b.v. onder aanname, dat de reactiecurve van den 2en graad is, enz.

Evenwel neemt bij veel factoren, elk in meerdere trappen, deze symboliek een dergelijke omvang aan, dat we genoodzaakt zijn, naar eenvoudiger uitdrukkingen te zoeken. Deze vinden we, door de reacties zich te laten vertegenwoordigen door hun variances. Bij een factor in n trappen wordt de variance weergegeven door $\frac{\sum(x_i-\bar{x})^2}{n-1}$

Bij 2 factoren, in respectievelijk n en m trappen, wordt de totale variance weergegeven door $\frac{\sum\sum(x_{ij}-\bar{\bar{x}})^2}{nm-1}$

Gaan we thans weer op dergelijke wijze als boven te werk n.l. karakteriseeren we den invloed van factor A door de waarde voor factor A over alle trappen van den factor B te middelen, dan kunnen we voor deze gemiddelden eveneens de variance bepalen; evenzoo vinden we een variance voor factor B

door over alle waarden van factor A te middelen. De dan nog niet verantwoorde variance wordt veroorzaakt door de interactie tusschen de factoren A en B, waarmee deze interactie dus meteen gekarakteriseerd is.
Volgens Fisher mogen we schrijven:

$$\sum_{k}^{n}\sum_{\ell}^{m}(x_{k\ell}-\bar{\bar{x}})^2 = \sum_{k}^{n}\sum_{\ell}^{m}(\bar{x}_{k.}-\bar{\bar{x}})^2 + \sum_{k}^{n}\sum_{\ell}^{m}(\bar{x}_{.\ell}-\bar{\bar{x}})^2 + \sum_{k}^{n}\sum_{\ell}^{m}(x_{k\ell}-\bar{x}_{k.}-\bar{x}_{.\ell}+\bar{\bar{x}})^2$$

met behulp waarvan we de genoemde variances kunnen vinden, door deze quadraatsommen respectievelijk te deelen door nm-1, n-1, m-1 en (n-1)(m-1). Voor een uitvoerige afleiding en beteekenis van deze z.g. analysis of variance moet hier verwezen worden naar de bekende leerboeken!).

Bij 3 factoren is de splitsing als volgt:

$$\sum_{k}^{n}\sum_{\ell}^{m}\sum_{i}^{c}(x_{k\ell i}-\bar{\bar{\bar{x}}})^2 = \sum\sum\sum(\bar{\bar{x}}_{k..}-\bar{\bar{\bar{x}}})^2 + \sum\sum\sum(\bar{\bar{x}}_{.\ell.}-\bar{\bar{\bar{x}}})^2 + \sum\sum\sum(\bar{\bar{x}}_{..i}-\bar{\bar{\bar{x}}})^2 +$$
$$+\sum\sum\sum(\bar{x}_{.\ell i}-\bar{\bar{x}}_{.\ell.}-\bar{\bar{x}}_{..i}+\bar{\bar{\bar{x}}})^2 + \sum\sum\sum(\bar{x}_{k.i}-\bar{\bar{x}}_{k..}-\bar{\bar{x}}_{..i}+\bar{\bar{\bar{x}}})^2 +$$
$$+\sum\sum\sum(\bar{x}_{k\ell.}-\bar{\bar{x}}_{k..}-\bar{\bar{x}}_{.\ell.}+\bar{\bar{\bar{x}}})^2 +$$
$$+\sum\sum\sum(x_{k\ell i}-\bar{x}_{.\ell i}-\bar{x}_{k.i}-\bar{x}_{k\ell.}+\bar{\bar{x}}_{k..}+\bar{\bar{x}}_{.\ell.}+\bar{\bar{x}}_{..i}-\bar{\bar{\bar{x}}})^2$$

respectievelijk voorstellende de tellers van de variances voor de factoren A, B en C, voor de interacties BC, AC en AB en voor de interactie van de 2e orde ABC.

Vervolgens kunnen we met de z.g. z-functie van Fisher nagaan welke interacties enz. reeel zijn, d.w.z. niet veroorzaakt door toevallige afwijkingen.

(De berekeningen nemen hierbij een veel geringeren omvang aan: we kunnen $\sum_{k}^{n}(x_k-\bar{x})^2$ nl. ook schrijven als $\sum x^2-\frac{(\sum x)^2}{n}$ enz., een vorm die aangepast is aan de electrische rekenmachines, die bij het aanslaan van een reeks getallen, gelijktijdig deze getallen en de quadraten van deze getallen optellen.

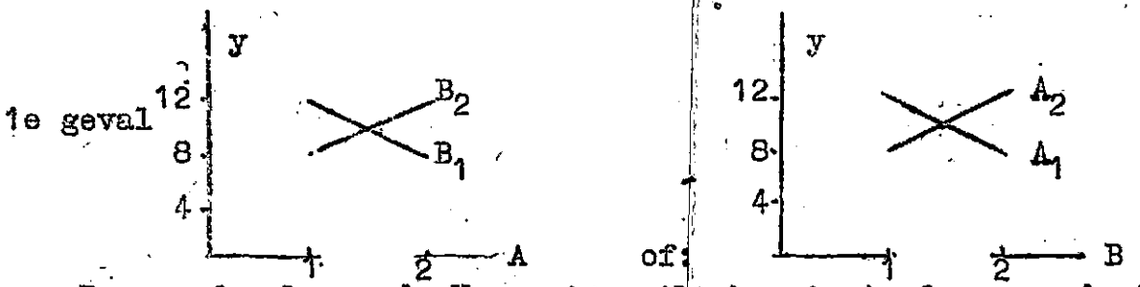
Voor korte rekenschema's zie men de bekende "receptenboeken" !!))

§4. Geïdealiseerde gevallen van het factorenmechanisme.

In beide gevallen van mathematische weergave van de reacties zijn dus bij 2 factoren de main-effects gemiddelden, de interactie niet. Bij 3 factoren zijn de main-effects en de interacties van de 1e orde gemiddelden, de interactie van de 2e orde niet enz. Dat het gemiddelden zijn, is de reden waarom b.v. in een analysis of variance sommige interacties en of main-effects =0 kunnen zijn, hetgeen op het eerste gezicht voor oningewijden vreemd aan doet. Voor velen is het moeilijk zich het spel der factoren voor te stellen, waarvan het resultaat is, dat b.v. bij 3 factoren de variances van de main-effects en van alle interacties van de 1e orde =0 zijn, en alleen de variance van de interactie van de 2e orde $\neq 0$ is. Daarom worden de volgende voorbeelden gegeven, daar toch ons doel is vertrouwd te geraken met het spel der factoren. Hierbij beschouwen we alleen extreme gevallen, waarbij alle of een gedeelte der variances =0 of $\neq 0$ is; alle andere mogelijkheden zijn dan overgangen tusschen deze extreme gevallen. (We nemen aan dat de toevalsvariance = 0, dat dus behalve de genoemde factoren geen andere factoren werken, het zijn dus geïdealiseerde gevallen).

Beschouwen we eerst slechts 2 factoren A en B elk in 2 trappen, dus weer 4 objecten A_1B_1 A_2B_1 A_1B_2 en A_2B_2.

We krijgen dan graphisch voorgesteld:



De waarden langs de Y as zijn willekeurig in deze voorbeelden.

!) Prof.R.A.Fisher. Statistical Methods for Research workers 1937.
Prof.Dr.M.J.v.Uven: Mathematical Treatment of the Results of Agricultural and other Experiments. 1935.

!!) b.v. J.Paterson. Statistical Technique in Agriculture 1939.

De main-effects en de interacties zijn:

$M(A)=\frac{1}{2}\{(A_2B_2-A_1B_2)+(A_2B_1-A_1B_1)\}=\frac{1}{2}\{(12-8)+(8-12)\}=0$

$M(B)=\frac{1}{2}\{(A_2B_2-A_2B_1)+(A_1B_2-A_1B_1)\}=\frac{1}{2}\{(12-8)+(8-12)\}=0$

$I(AB)=\frac{1}{2}\{(A_2B_2-A_1B_2)-(A_2B_1-A_1B_1)\}=\frac{1}{2}\{(12-8)-(8-12)\}=4$

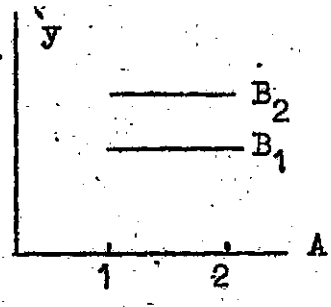
We kunnen de waarden der objecten ook kort als volgt weergeven:

	B_1	B_2	
A_1	12	8	20
A_2	8	12	20
	20	20	40

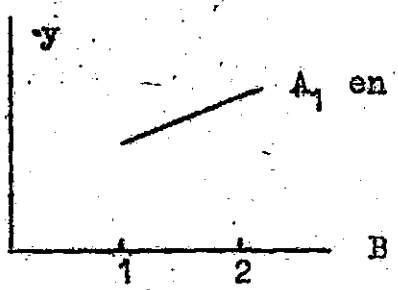
De gemiddelden c.q. de sommen voor de factoren A én B zijn gelijk, de variances voor de factoren A en B zullen dus =0 zijn, terwijl de variance voor de interactie ≠ 0 zal zijn.

Doordat we hieronder telkens voor elk geval twee graphische voorstellingen zullen geven, kunnen we bij dit eenvoudige factorenspel van 2 factoren de resultaten van een analysis of variance zoo uit de figuur aflezen, welk resultaat er naast aan toegevoegd staat.

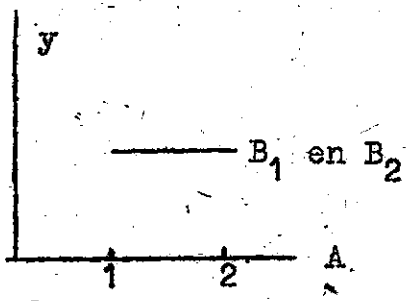
2e geval
V(A) =0
V(B) ≠0
V(AB)=0



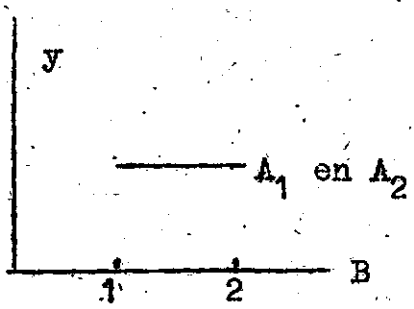
of:



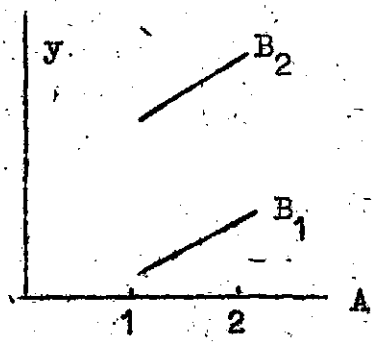
3e geval
V(A)=0
V(B)=0
V(AB)=0



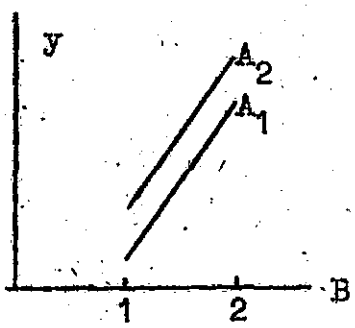
of



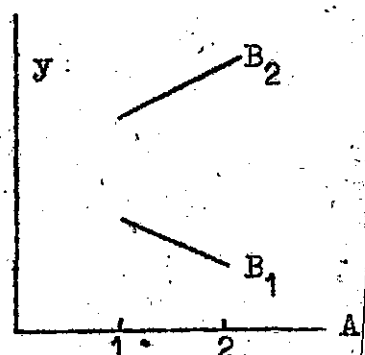
4e geval
V(A) ≠ 0
V(B) ≠ 0
V(AB)≠ 0

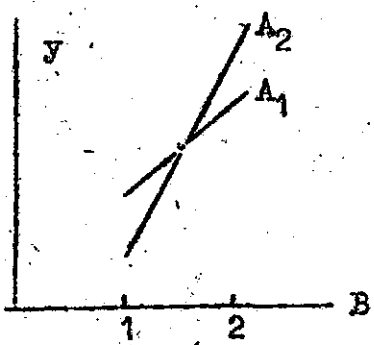


of:

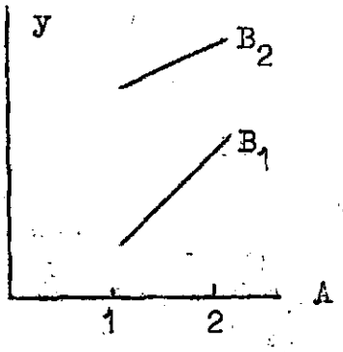


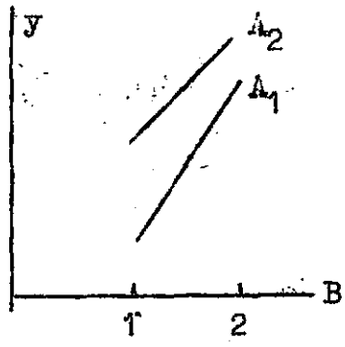
5e geval
V(A) = 0
V(B) ≠ 0
V(AB)≠ 0





6e geval
V(A) ≠ 0
V(B) ≠ 0
V(AB) ≠ 0





Bij het 2e geval behoort nog een analoog geval waarbij V(A) ≠ 0 V(B) = 0 V(AB) = 0 en eveneens behoort bij het 5e geval een analoog geval, waarbij V(A) ≠ 0 V(B) = 0 V(AB) ≠ 0.

Hieronder zullen thans nog kort een aantal extreme gevallen, die zich bij 3 factoren kunnen voordoen, kort weergegeven worden. Door hierbij telkens de graphische voorstelling op 3 wijzen te geven, krijgt men elk geval op de meest aanschouwelijke wijze voorgesteld, terwijl het resultaat van een eventuele analysis of variance ook meteen zonder berekening uit de graphische voorstelling kan worden afgelezen, welk resultaat nevens de figuur is toegevoegd.

Bij 3 factoren zijn er 32 extreme gevallen mogelijk, zonder meerekenen van de analoge gevallen; worden deze meegerekend, dan zijn er 128. Geen wonder, dat men een willekeurig geval, hetwelk in een 4-dimensionale ruimte moet worden afgebeeld, niet dadelijk kan doorgronden, en men met recht van het spel der factoren kan spreken!

De 32 extreme gevallen zijn vereenigd in het volgende overzicht (hierbij beteekent een - dat de variance = 0 en een x dat de variance ≠ 0).

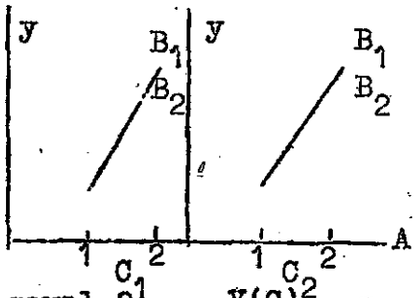
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
V(A)	-	x	x	x	-	-	-	-	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	-	-	-
V(B)	-	-	x	x	-	-	-	-	-	x	x	-	x	x	-	x	x	-	x	x	-	x	x	-	x	x	-	x	x	-	-	-
V(C)	-	-	-	x	-	-	-	-	-	-	x	-	-	x	-	-	x	-	-	x	-	-	x	-	-	x	-	-	x	-	-	-
V(AB)	-	-	-	-	x	x	x	-	x	x	x	x	x	x	x	x	x	-	-	-	x	x	x	x	x	x	x	x	x	x	x	x
V(AC)	-	-	-	-	-	x	x	-	-	-	-	x	x	x	x	x	x	-	-	-	-	-	-	x	x	x	x	x	x	-	x	x
V(BC)	-	-	-	-	-	-	x	-	-	-	-	-	-	-	x	x	x	-	-	-	-	-	-	-	-	-	x	x	x	-	-	x
V(ABC)	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Van de No. 1, 4, 7, 8, 17, 20, 29 en 32 kunnen zich geen analoge gevallen voordoen.

Van enkele der andere No. volgen hier eenige voorbeelden van analoge gevallen (altijd 2 of 8 analoge gevallen bij 3 factoren).

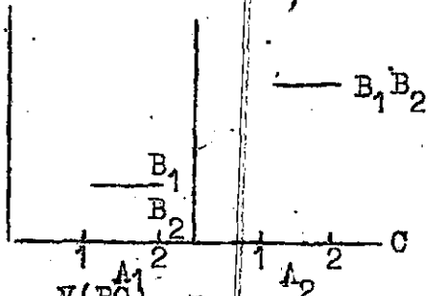
No.	2^1	2^2	3^1	3^2	5^1	5^2	6^1	6^2	9^1	9^2	9^3	9^4	9^5	9^6	9^7	9^8	14^1	14^2
V(A)	-	-	x	-	-	-	-	-	x	x	-	-	-	-	-	-	x	x
V(B)	x	-	-	x	-	-	-	-	-	-	x	x	x	-	-	-	x	x
V(C)	-	x	x	x	-	-	-	-	-	-	-	-	-	x	x	x	x	x
V(AB)	-	-	-	-	-	-	x	-	-	-	x	-	-	x	-	-	x	-
V(AC)	-	-	-	-	x	-	-	x	x	-	-	x	-	-	x	-	-	x
V(BC)	-	-	-	-	-	x	x	x	-	x	-	-	x	-	-	x	x	x
V(ABC)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Stellen we enkele gevallen weer graphisch voor:

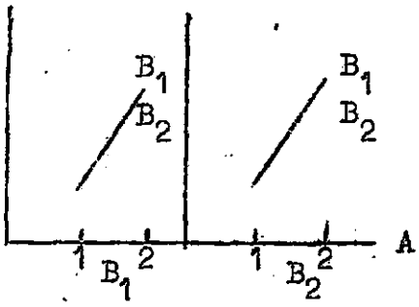


geval 2^1
V(A) x
V(B) -

$V(C)^2$ -
V(AB) -
V(AC) -



V(BC) -
V(ABC) -

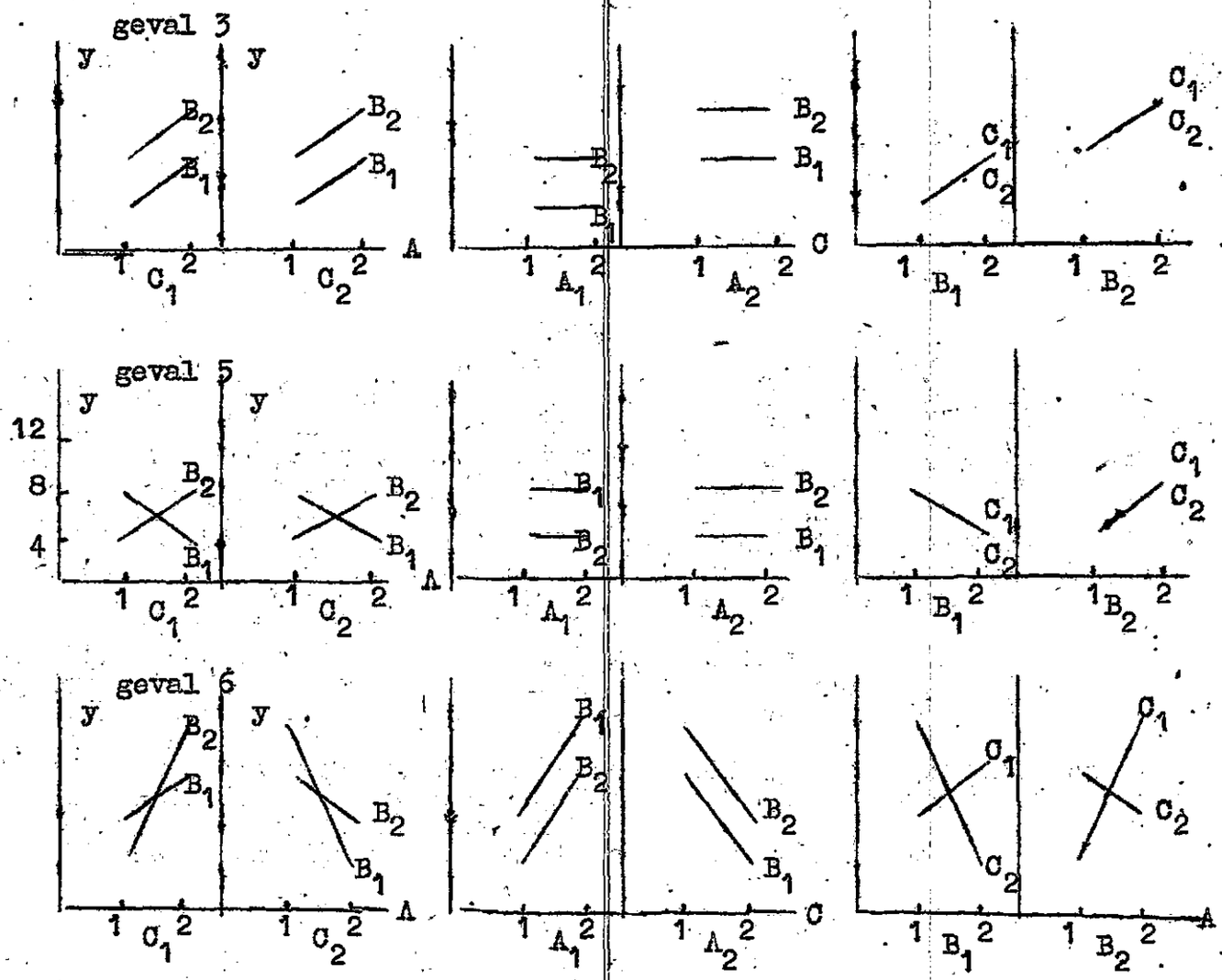


Uit de linker figuur lezen we dadelijk af: geen interactie tusschen de factoren A en B, noch bij C_1 noch bij C_2, dus ook de interactie van de 2e orde =0. Bij C_1 is er een reactie op A, bij C_2 eenzelfde reactie, dus moet er een variance voor den factor A zijn. Ook zien we dat het gemiddelde van de 4 objecten bij C_1, gelijk is aan het gemiddelde van de 4 objecten bij C_2, dus ook de variance voor den factor C =0.

Uit de middelste figuur lezen we af, geen interactie tusschen de factoren B en C noch bij A_1, noch bij A_2, dus ook geen interactie van de 2e orde, welk laatste we ook reeds uit de linker figuur wisten. In deze middelste figuur zien we duidelijker, dat er geen reactie op factor C is, noch bij A_1 noch bij A_2. Verder is het gemiddelde van de 4 objecten bij A_1, lager dan het gemiddelde van de objecten bij A_2; dus gemiddeld genomen is er een reactie op A, een variance voor den factor A in overeenstemming met de linker figuur.

Uit de rechter figuur lezen we onmiddellijk af; geen interactie tusschen A en C, wat we verder nog kunnen aflezen is in overeenstemming met de linker en middelste figuur.

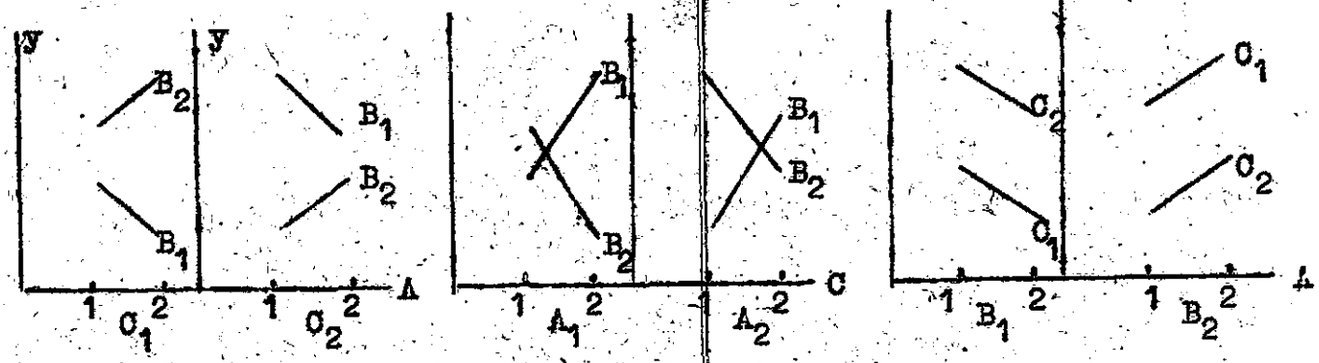
Evenzoo voor alle andere graphische voorstellingen:



geval 3	
V(A)	x
V(B)	x
V(C)	-
V(AB)	-
V(AC)	-
V(BC)	-
V(ABC)	-

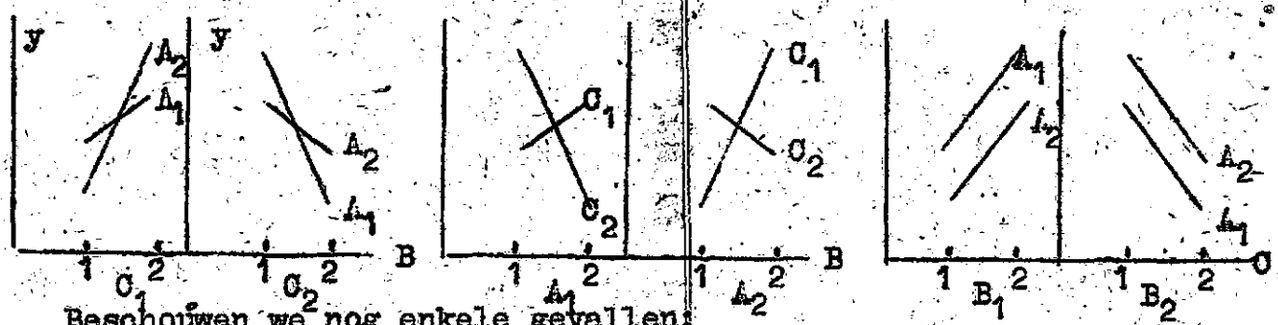
geval 5	
V(A)	-
V(B)	-
V(C)	-
V(AB)	x
V(AC)	-
V(BC)	-
V(ABC)	-

geval 6	
V(A)	-
V(B)	-
V(C)	-
V(AB)	x
V(AC)	x
V(BC)	-
V(ABC)	-

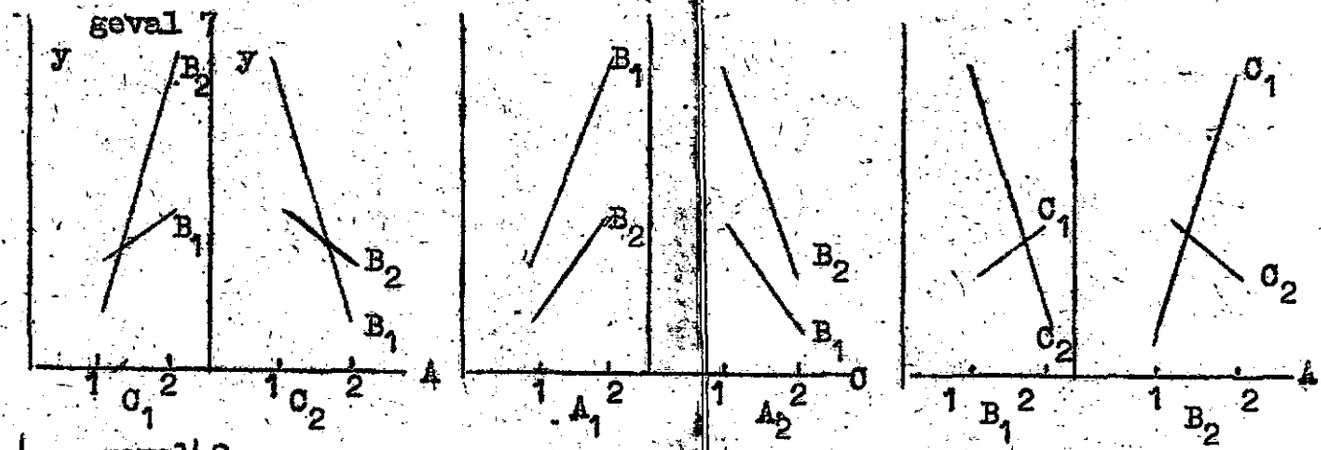


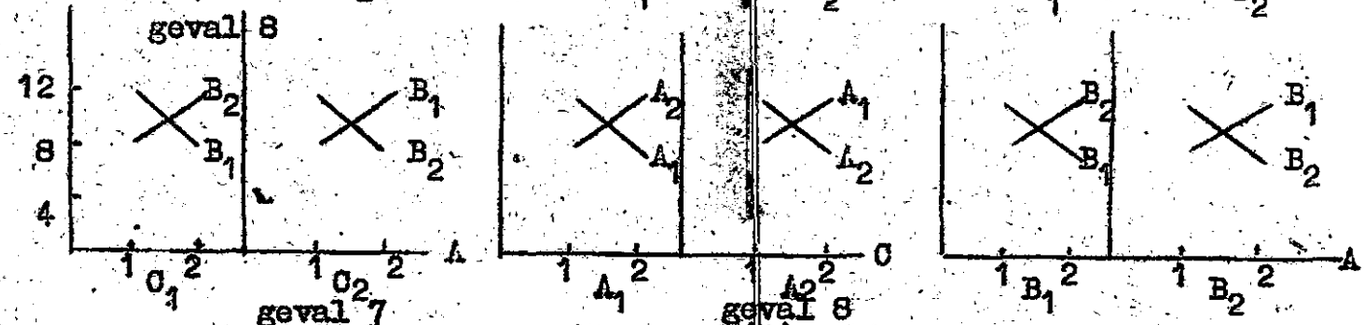
geval 6^1
V(A) -
V(B) -
V(C) -
V(AB) x
V(AC) -
V(BC) x
V(ABC) -

Oogenschijnlijk zijn 6 en 6^1 niet analoog. Vervormen wij echter de figuren van no. 6^1, dan zien we analoge figuren als in No. 6 ontstaan, alleen de letters zijn verwisseld:



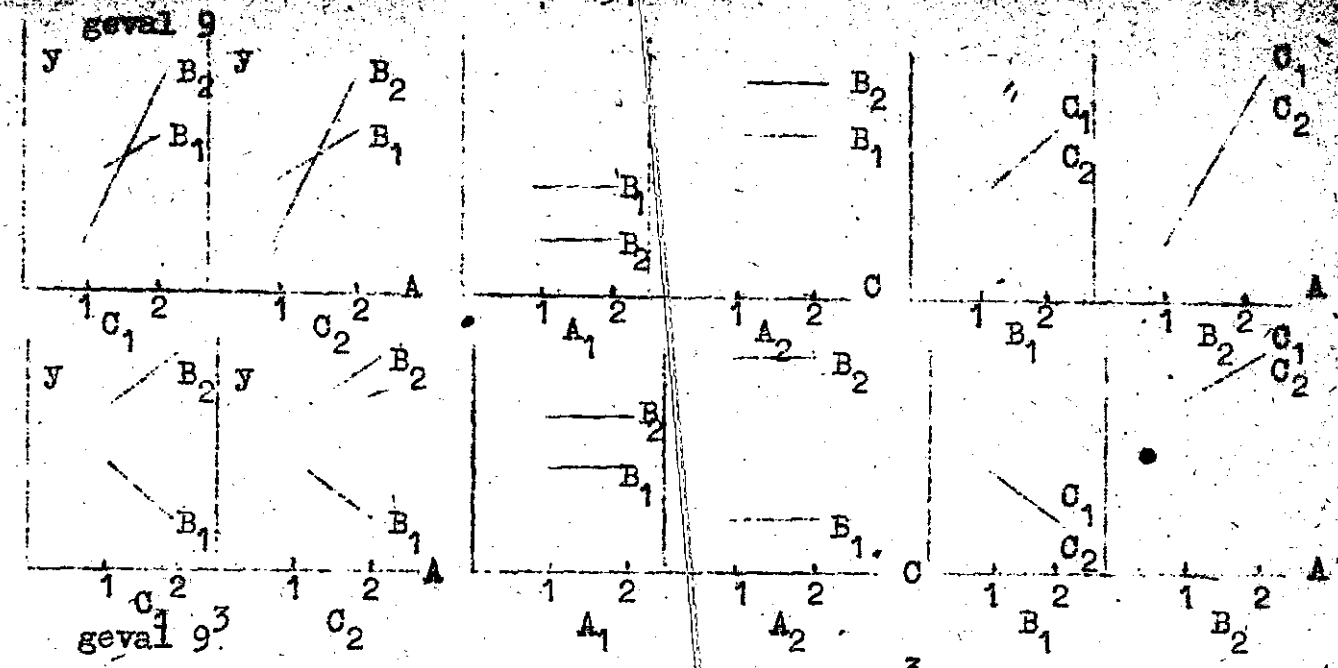
Beschouwen we nog enkele gevallen:





geval 7
V(A) -
V(B) -
V(C) -
V(AB) x
V(AC) x
V(BC) x
V(ABC) -

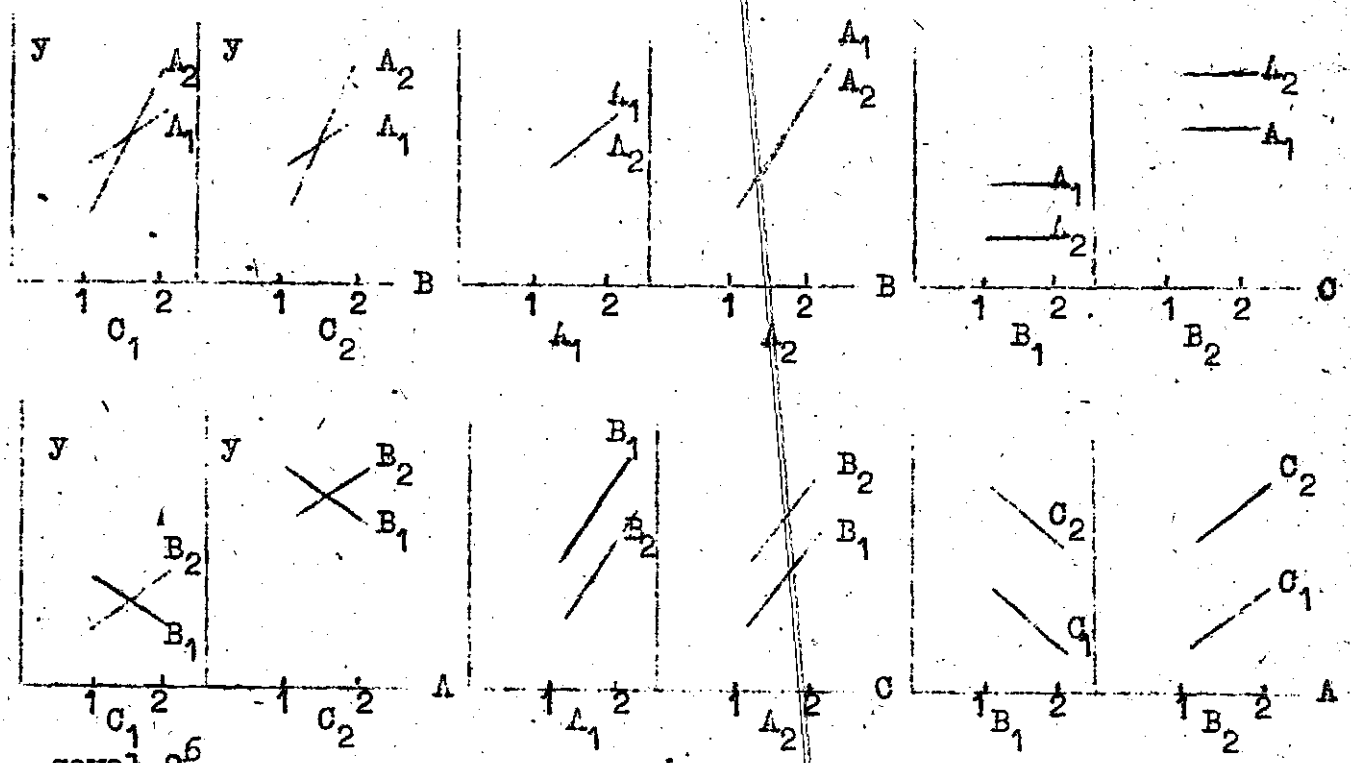
geval 8
V(A) -
V(B) -
V(C) -
V(AB) -
V(AC) -
V(BC) -
V(ABC) x



geval 9	
V(A)	x
V(B)	-
V(C)	-
V(AB)	x
V(AC)	-
V(BC)	-
V(ABC)	-

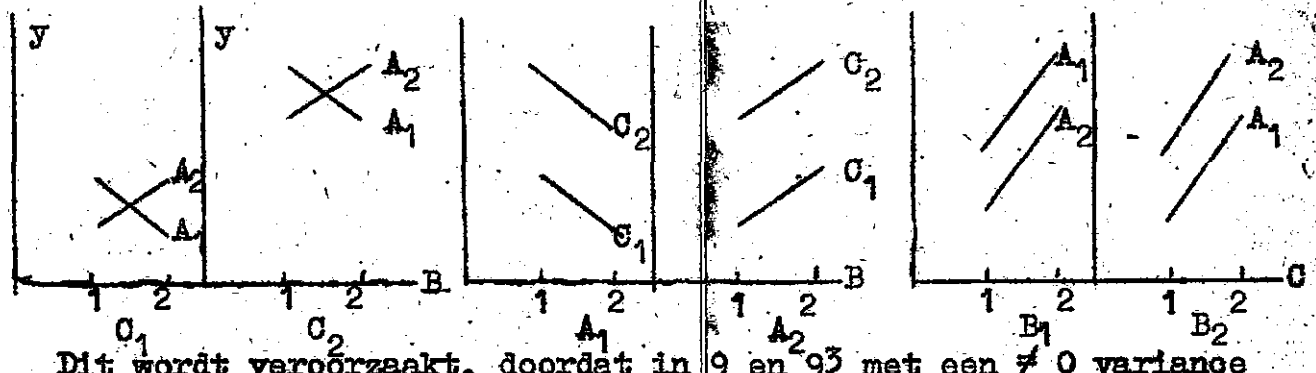
geval 9³	
V(A)	-
V(B)	x
V(C)	-
V(AB)	x
V(AC)	-
V(BC)	-
V(ABC)	-

Oogenschijnlijk ontbreekt weer de analogie tusschen 9 en 9^3. Vervorming van de figuren van 9^3 doen echter analoge figuren met 9 ontstaan:



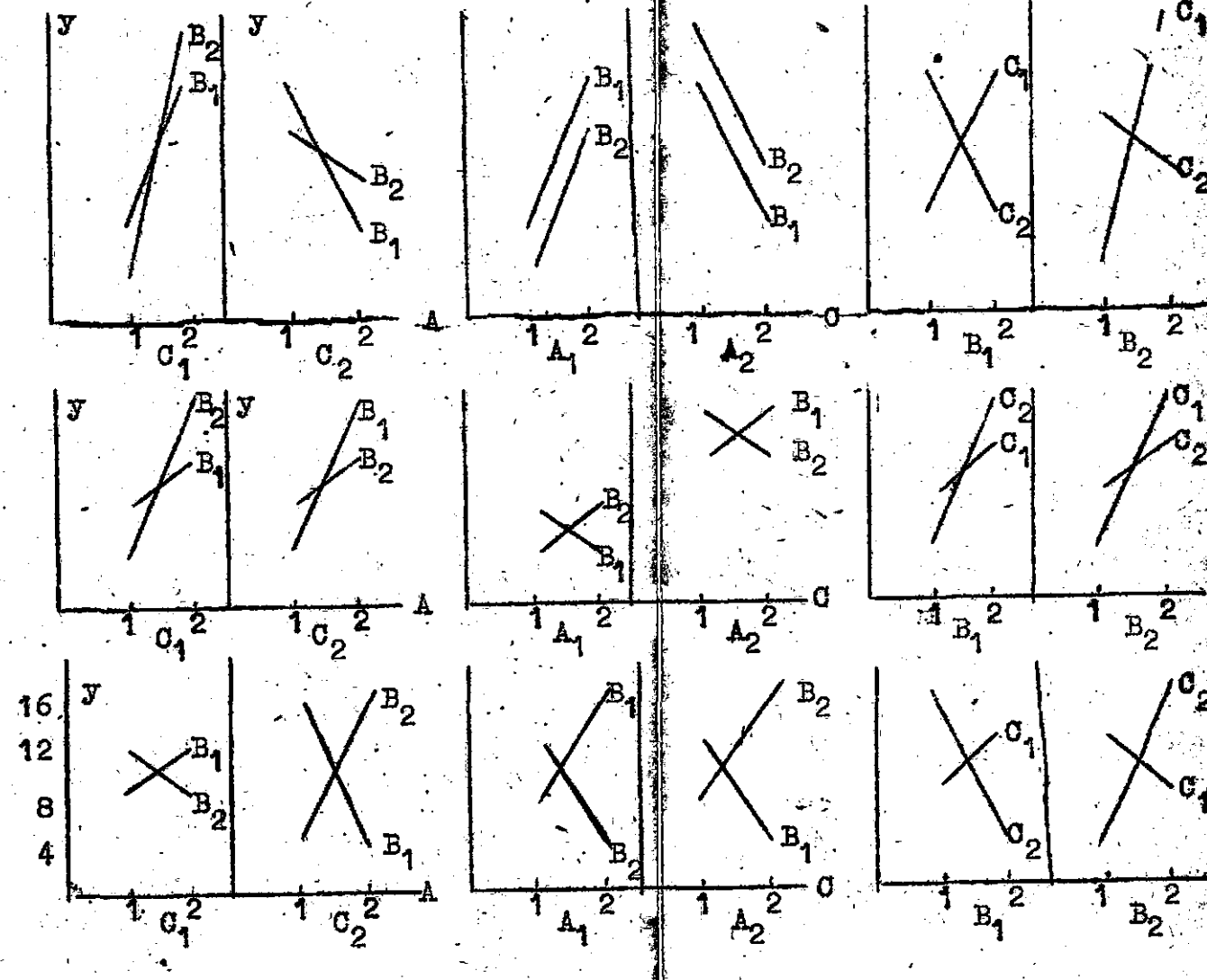
geval 9⁶	
V(A)	-
V(B)	-
V(C)	x
V(AB)	x
V(AC)	-
V(BC)	-
V(ABC)	-

Thans weer vervorming, maar nu géén analoge figuren met 9 of 9^3!



Dit wordt veroorzaakt, doordat in 9 en 9^3 met een $\neq 0$ variance van de interactie AB samengaat een $\neq 0$ variance van of van den factor A of van den factor B. In 9^6 echter geen van beide factoren, doch van den factor C. Hier gaat de analogie dus bijna mank; men zou dit geval als bijna een apart geval bij de reeds genoemde 32 gevallen kunnen rangschikken. Zoo zijn er trouwens nog meer. Inderdaad, het spel der factoren is niet eenvoudig!

Tot slot nog een drietal voorbeelden:



geval 12	
V(A)	x
V(B)	-
V(C)	-
V(AB)	x
V(AC)	x
V(BC)	-
V(ABC)	-

geval 18	
V(A)	x
V(B)	-
V(C)	-
V(AB)	-
V(AC)	-
V(BC)	-
V(ABC)	x

geval 30	
V(A)	-
V(B)	-
V(C)	-
V(AB)	x
V(AC)	-
V(BC)	-
V(ABC)	x

Thans moeten we terugkomen op het gezegde op pg. 55, nl. dat wisselwerking en interactie niet altijd dezelfde betcekenis hebben.

Zooals we hierboven in de verschillende schema's gezien hebben, kan een interactie bij de definities zooals die gegeven zijn, zeer vaak = 0 zijn, terwijl er toch zeker sprake is van een wisselwerking!

Het is duidelijk dat dit door het middelen veroorzaakt wordt! Evenzoo zijn sommige main-effects = 0 terwijl er toch zeer zeker een reactie op deze factoren is, alleen wordt deze verdoezeld door het middelen!

Nemen we b.v. het geval 5: hierin is het resultaat van de analysis of variance, dat alleen de variance van de interactie AB reëel is. Toch zien we duidelijk, b.v. in de rechter figuur, dat er een reactie op factor A is, zoowel bij B_1 als bij B_2. Alleen door het middelen komen we tot het resultaat, dat het main-effect van factor A = 0 is, c.q. de variance voor factor A = 0 is. Hetzelfde geldt voor factor B.

Nemen we vervolgens b.v. het geval No. 8. Hierin is in de analysis of variance alleen de interactie van de 2e orde ABC ≠ 0. Toch zien we duidelijk in b.v. de linker figuur dat de interactie AB reëel is, zoowel bij C_1 als bij C_2, alleen door het middelen over C wordt de interactie, zooals die gedefinieerd is = 0. Hetzelfde geldt voor de interacties AC en BC; ook zijn er hier zeer zeker reacties op de factoren A, B en C, alleen door het middelen worden de main-effects c.q. de variances van de factoren = 0.

§ 5. Twee nieuwe begrippen: main-action en interplay.

We zien dus, dat we de met onze definities verkregen uitdrukkingen, nog zeer voorzichtig moeten interpreteeren, om de juiste voorstelling van het factorenspel te verkrijgen.

Schrijver voert daarom nog de volgende uitdrukkingen hier nieuw in, n.l. main-action en interplay.

Zijn er in werkelijkheid wel reacties op de factoren, dan spreken we van main-action, zijn er geen reacties, dan is de main-action = 0, maar ook het main-effect = 0. Is de main-action ≠ 0, dan kan het main-effect al of niet = 0 zijn, is het = 0, dan wordt dat veroorzaakt door het middelen.

Zijn er in werkelijkheid wel reacties op de combinaties van verschillende factoren, is er dus een wisselwerking, dan spreken we van interplay. Is er geen wisselwerking dan is de interplay= 0 maar ook de interaction = 0. Is de interplay ≠ 0, dan kan de interaction al of niet = 0 zijn. Is de interactie in dat geval = 0, dan wordt dat veroorzaakt door het middelen.

Ook deze uitdrukkingen zullen nader moeten definieeren en wel met behulp van de absolute waarden van de diverse verschillen.

Bij 2 factoren A en B, elk in 2 trappen wordt de main-action A gedefinieerd door het gemiddelde van de absolute waarden van de reacties op factor A bij de trappen B_1 en B_2 dus:

$$M^x(A)=\tfrac{1}{2}\{|(A_2B_2-A_1B_2)|+|(A_2B_1-A_1B_1)|\}$$

evenzoo:

$$M^x(B)=\tfrac{1}{2}\{|(A_2B_2-A_2B_1)|+|(A_1B_2-A_1B_1)|\}$$

De interplay kan net zoo gedefinieerd worden als de interaction bij 2 factoren, daar de interactie bij 2 factoren nooit een gemiddelde is. Voor de uniformiteit en voor de analogie met en uitbreiding tot het geval met 3 factoren, zullen we echter de absolute waarde van de interactie nemen dus:

$$I^x(AB)=\tfrac{1}{2}\{|(A_2B_2-A_1B_2)|-|(A_2B_1-A_1B_1)|\}$$

Bij 3 factoren worden de mainactions van de factoren A B en C gedefinieerd als volgt:

$$M^x(A)=\tfrac{1}{2}\Big[\tfrac{1}{2}\{|(A_2B_1C_1-A_1B_1C_1)|+|(A_2B_2C_1-A_1B_2C_1)|\}+\tfrac{1}{2}\{|(A_2B_1C_2-A_1B_1C_2)|+|(A_2B_2C_2-A_1B_2C_2)|\}\Big]$$

$$M^x(B)=\tfrac{1}{2}\Big[\tfrac{1}{2}\{|(A_1B_2C_1-A_1B_1C_1)|+|(A_2B_2C_1-A_2B_1C_1)|\}+\tfrac{1}{2}\{|(A_1B_2C_2-A_1B_1C_2)|+|(A_2B_2C_2-A_2B_1C_2)|\}\Big]$$

$$M^x(C)=\tfrac{1}{2}\Big[\tfrac{1}{2}\{|(A_1B_1C_2-A_1B_1C_1)|+|(A_2B_1C_2-A_2B_1C_1)|\}+\tfrac{1}{2}\{|(A_1B_2C_2-A_1B_2C_1)|+|(A_2B_2C_2-A_2B_2C_1)|\}\Big]$$

De interplay AB wordt, daar de interactie AB een gemiddelde is, thans gedefinieerd als het gemiddelde van de absolutewaarden van de interactie AB bij C_1 en bij C_2, dus als het gemiddelde van de interplays AB bij C_1 en C_2. De interplays AC en BC worden op analoge wijze gedefinieerd.

$$I^x(AB)=\tfrac{1}{2}\Big[\tfrac{1}{2}\big\{(A_2B_1C_1-A_1B_1C_1)-(A_2B_2C_1-A_1B_2C_1)\big\}+\tfrac{1}{2}\big\{(A_2B_1C_2-A_1B_1C_2)-(A_2B_2C_2-A_1B_2C_2)\big\}\Big]$$

$$I^x(AC)=\tfrac{1}{2}\Big[\tfrac{1}{2}\big\{(A_1B_1C_2-A_1B_1C_1)-(A_2B_1C_2-A_2B_1C_1)\big\}+\tfrac{1}{2}\big\{(A_1B_2C_2-A_1B_2C_1)-(A_2B_2C_2-A_2B_2C_1)\big\}\Big]$$

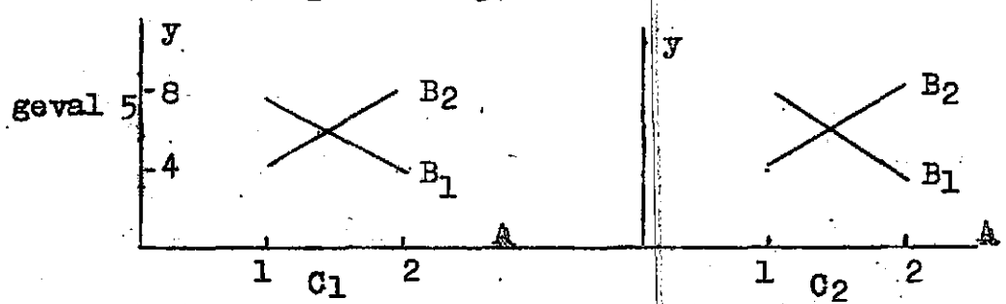
$$I^x(BC)=\tfrac{1}{2}\Big[\tfrac{1}{2}\big\{(A_1B_2C_1-A_1B_1C_1)-(A_1B_2C_2-A_1B_1C_2)\big\}+\tfrac{1}{2}\big\{(A_2B_2C_1-A_2B_1C_1)-(A_2B_2C_2-A_2B_1C_2)\big\}\Big]$$

De interactie van de 2e orde is bij 3 factoren geen gemiddelde en de interplay zou dus net zoo gedefinieerd kunnen worden. Voor de uniformiteit met en voor de uitbreiding tot het geval met 4 factoren zullen we de absolute waarde nemen. Bij 4 factoren kunnen dan de 4 interplays van de 2e orde gedefinieerd worden als een gemiddelde van interplays van de 2e orde, evenals hierboven de interplays van de 1e orde gedefinieerd konden worden, als een gemiddelde van interplays van de 1e orde.

We krijgen dan:

$$I^x(ABC)=\tfrac{1}{2}\Big[\tfrac{1}{2}\big\{(A_2B_1C_1-A_1B_1C_1)-(A_2B_2C_1-A_1B_2C_1)\big\}-\tfrac{1}{2}\big\{(A_2B_1C_2-A_1B_1C_2)-(A_2B_2C_2-A_1B_2C_2)\big\}\Big]$$

Met behulp van deze uitdrukkingen kunnen we de wisselwerkingen, de reacties in het algemeen nu volledig beschrijven. Nemen we weer als voorbeeld de genoemde gevallen 5 en 8.



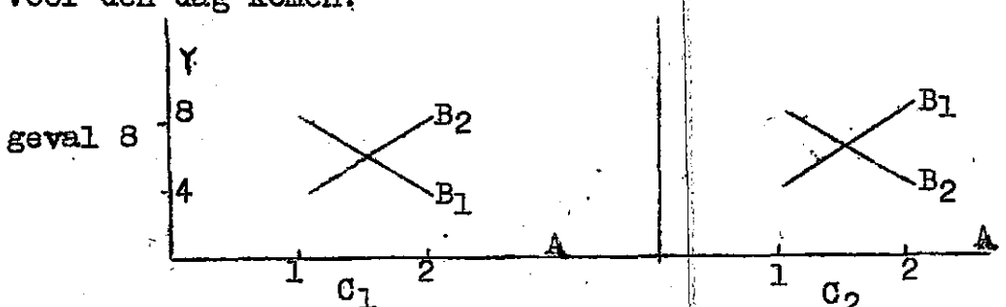
main-effects en interacties:

$M(A) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)+(8-4)\}+\tfrac{1}{2}\{(4-8)+(8-4)\}\big] = 0$
$M(B) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)+(8-4)\}+\tfrac{1}{2}\{(4-8)+(8-4)\}\big] = 0$
$M(C) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(8-8)+(4-4)\}+\tfrac{1}{2}\{(4-4)+(8-8)\}\big] = 0$
$I(AB) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}+\tfrac{1}{2}\{(4-8)-(8-4)\}\big] = -4$
$I(AC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(8-8)-(4-4)\}+\tfrac{1}{2}\{(4-4)-(8-8)\}\big] = 0$
$I(BC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(4-8)\}+\tfrac{1}{2}\{(8-4)-(8-4)\}\big] = 0$
$I(ABC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}-\tfrac{1}{2}\{(4-8)-(8-4)\}\big] = 0$

mainactions en interplays:

$M^x(A) = \tfrac{1}{2}\big[\tfrac{1}{2}\{|(4-8)|+|(8-4)|\}+\tfrac{1}{2}\{|(4-8)|+|(8-4)|\}\big] = 4$
$M^x(B) = \tfrac{1}{2}\big[\tfrac{1}{2}\{|(4-8)|+|(8-4)|\}+\tfrac{1}{2}\{|(4-8)|+|(8-4)|\}\big] = 4$
$M^x(C) = \tfrac{1}{2}\big[\tfrac{1}{2}\{|(8-8)|+|(4-4)|\}+\tfrac{1}{2}\{|(4-4)|+|(8-8)|\}\big] = 0$
$I^x(AB) = \tfrac{1}{2}\big[|\tfrac{1}{2}\{(4-8)-(8-4)\}|+|\tfrac{1}{2}\{(4-8)-(8-4)\}|\big] = 4$
$I^x(AC) = \tfrac{1}{2}\big[|\tfrac{1}{2}\{(8-8)-(4-4)\}|+|\tfrac{1}{2}\{(4-4)-(8-8)\}|\big] = 0$
$I^x(BC) = \tfrac{1}{2}\big[|\tfrac{1}{2}\{(4-8)-(4-8)\}|+|\tfrac{1}{2}\{(8-4)-(8-4)\}|\big] = 0$
$I^x(ABC) = \tfrac{1}{2}\big[|\tfrac{1}{2}\{(4-8)-(8-4)\}-\tfrac{1}{2}\{(4-8)-(8-4)\}|\big] = 0$

Inderdaad zien we thans het gestelde op pg. 36 over het optreden van de diverse reacties, die door het middelen vorborgen waren, thans voor den dag komen!



main-effects en interacties:

$M(A) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)+(8-4)\}+\tfrac{1}{2}\{(8-4)+(4-8)\}\big] = 0$
$M(B) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)+(8-4)\}+\tfrac{1}{2}\{(8-4)+(4-8)\}\big] = 0$
$M(C) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)+(8-4)\}+\tfrac{1}{2}\{(8-4)+(4-8)\}\big] = 0$
$I(AB) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}+\tfrac{1}{2}\{(8-4)-(4-8)\}\big] = 0$
$I(AC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}+\tfrac{1}{2}\{(8-4)-(4-8)\}\big] = 0$
$I(BC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}+\tfrac{1}{2}\{(8-4)-(4-8)\}\big] = 0$
$I(ABC) = \tfrac{1}{2}\big[\tfrac{1}{2}\{(4-8)-(8-4)\}-\tfrac{1}{2}\{(8-4)-(4-8)\}\big] = -4$

main actions en interplays:

$M^x(A) = \frac{1}{2}[\frac{1}{2}\{|(4-8)| + |(8-4)|\} + \frac{1}{2}\{|(8-4)| + |(4-8)|\}] = 4$

$M^x(B) = \frac{1}{2}[\frac{1}{2}\{|(4-8)| + |(8-4)|\} + \frac{1}{2}\{|(8-4)| + |(4-8)|\}] = 4$

$M^x(C) = \frac{1}{2}[\frac{1}{2}\{|(4-8)| + |(8-4)|\} + \frac{1}{2}\{|(8-4)| + |(4-8)|\}] = 4$

$I^x(AB) = \frac{1}{2}[|\frac{1}{2}\{(4-8) - (8-4)\}| + |\frac{1}{2}\{(8-4) - (4-8)\}|] = 4$

$I^x(AC) = \frac{1}{2}[|\frac{1}{2}\{(4-8) - (8-4)\}| + |\frac{1}{2}\{(8-4) - (4-8)\}|] = 4$

$I^x(BC) = \frac{1}{2}[|\frac{1}{2}\{(4-8) - (8-4)\}| + |\frac{1}{2}\{(8-4) - (4-8)\}|] = 4$

$I^x(ABC) = |\frac{1}{2}[\frac{1}{2}\{(4-8) - (8-4)\} - \frac{1}{2}\{(8-4) - (4-8)\}]| = 4$

Alle reacties en wisselwerkingen, die moeten optreden, als er een interactie van de 2e orde ABC is, maar die verborgen waren door het middelen, komen thans voor den dag!

Alle andere genoemde schema's kunnen evenzoo uitgewerkt worden; b.v. schema 6 waarin we de interacties AB en AC vinden, geeft dan tot resultaat de main actions A, B en C en de interplays AB en AC.

Schema 9^6 waarin we het maineffect C en de interactie AB vinden, geeft dan tot resultaat de main-actions A, B en C en de interplay AB.

Schema 30 waarin we de interacties AB en ABC vinden geeft dan tot resultaat de main actions A, B en C en de interplays AB, AC, BC en ABC.

(De directe bepaling der main-effects en interacties zouden we "analysis of effects" kunnen noemen, de bepaling van de main actions en interplays "analysis of actions").

We behoeven deze definities voor main action en interplay niet uit te breiden, zooals we de main effects en interacties uitbreidden met behulp van de variances als er vele factoren zijn, alle in veel trappen (vergl. pg.28).

Uit de analysis of variance kunnen we nl. zonder meer besluiten, welke main actions en interplays een rol spelen. Is gegeven een reëele variance voor de interactie AB dan bestaat er een main action A en B en interplay AB. Is er daarnaast een reëele interactie AC, dan bestaan de main actions A en C en de interplay AC. De main action A kan door het middelen zooals wij dat gedaan hebben in onze definities met behulp van de absolute waarde van de verschillen nooit wegvallen!

Bestaat er een variance voor de interactie van de 2e orde ABC, dan bestaan de 3 main-actions A, B en C, de 3 interplays van de 1e orde AB, AC en BC en de interplay ABC van de 2e orde.

Naar schrijver hoopt is door het invoeren van deze twee nieuwe begrippen main-action en interplay het spel der factoren, zooals dat geanalyseerd wordt door de analysis of variance, thans wat doorzichtiger geworden. Men zou kunnen zeggen, dat men de resultaten van een analysis of variance zou moeten interpreteeren met behulp van deze twee nieuwe begrippen.

Het kan nu dus duidelijk zijn, dat de woorden reactie en wisselwerking de algemeene termen zijn. De interactie is een zeer bepaald gedefinieerd begrip, dat het woord wisselwerking niet altijd dekt, alleen als de interactie, de interactie van de hoogste orde is, bij een gegeven aantal factoren, kan zij samenvallen met het begrip wisselwerking. Het woord interplay is altijd synoniem aan wisselwerking, maar bovendien is interplay een zeer speciale maat om de wisselwerking te meten, speciaal in het geval dat de factoren in 2 trappen onderzocht worden[1]).

De nomenclatuur ware duidelijker, indien het woord interactie, dat thans zoowel voor wisselwerking, als voor het resultaat van een zeer bepaald rekenproces gebruikt wordt, voor dit laatste geval vervangen werd door intereffect. Doch het woord interactie voor het re-

[1]) In het Engelsch zouden we het algemeene woord wisselwerking m.i. moeten vertalen door het woord interacting.

kenproces is nu zoo vast ingeburgerd, dat schrijver dit niet doenlijk acht.

Aan het einde gekomen van deze algemeene inleiding, speciaal bedoeld voor die lezers, die zich tot nog toe weinig in deze zaken verdiept hadden, hoopt schrijver dat men eenig inzicht in het spel der factoren gekregen heeft, en dat in het bijzonder men van de noodzakelijkheid overtuigd is, om zijn experimenten volgens het factorial design te ontwerpen om de werking der factoren in zijn macht te krijgen, dat men overtuigd is van de noodzakelijkheid om zijn waarnemingen a posteriori te overkappen met een factorial design, om orde te scheppen in de factoren, m.a.w. dat men zijn waarnemingen klassificeert volgens de verschillende factoren met behulp van het factorial design.

Het laatste heeft veel overeenkomst met een algemeene correlatieanalyse. Alleen is tot nog toe de meervoudige en kromlijnige correlatie waarbij de factoren bovendien niet onafhankelijk van elkaar zijn, niet erg populair, (onafhankelijk hier gebruikt in den zin van geen interactie tusschen de factoren). Men beperkt zich meestal zeer sterk, en wel tot de meervoudige lineaire correlatie, een geval, dat men in de natuur zelden verwezenlijkt vindt. Als men echter goed van de beteekenis van het factorial design doortrokken is, is het veel gemakkelijker, vertrouwd te worden met de algemeene correlatie analyse, waarbij men zich geen beperkingen behoeft op te leggen[1]).

---

[1]) Speciaal uitgewerkt, echter voor een klein aantal factoren, door Erekiel, zie: Methods of Correlation-analysis. New York 1930.

HOOFDSTUK IV

VOORTGEZETTE PROBLEEMSTELLING.

Na aldus het factorial design besproken te hebben, kunnen we thans het probleem, dat het spel der factoren ons stelt, nader definieeren.

Het is het factorial design, dat in het spel van tientallen of honderdtallen factoren, die elkaar wederzijds beinvloeden, en waardoor honderdtallen en duizendtallen interacties ontstaan, orde moet scheppen!

Het spel der factoren kan beheerscht worden door een factorial design met een oneindig aantal factoren, laten we voorloopig eenvoudigheidshalve zeggen: met zeer veel factoren. Het is hiermede, dat we weer een stap voorwaarts kunnen doen, naar het ideaal van de beheersching der natuur.

Dus het zij door een experiment a priori uit te voeren volgens zoodanig factorial design, dan wel de waarnemingen a posteriori te overkappen met zoodanig factorial design.

We moeten ons dus thans bezig houden met de problemen, die het factorial design met een oneindig aantal of liever gezegd met zeer vele factoren ons stelt.

Het moge schijnen, dat dit slechts een blooteverlegging van de moeilijkheden is, van het spel der factoren naar het factorial design met oneindig veel factoren. Dit is gedeeltelijk ook zoo, maar het factorial design is exact te definieeren en, zooals we zullen zien, mathematisch te bewerken.

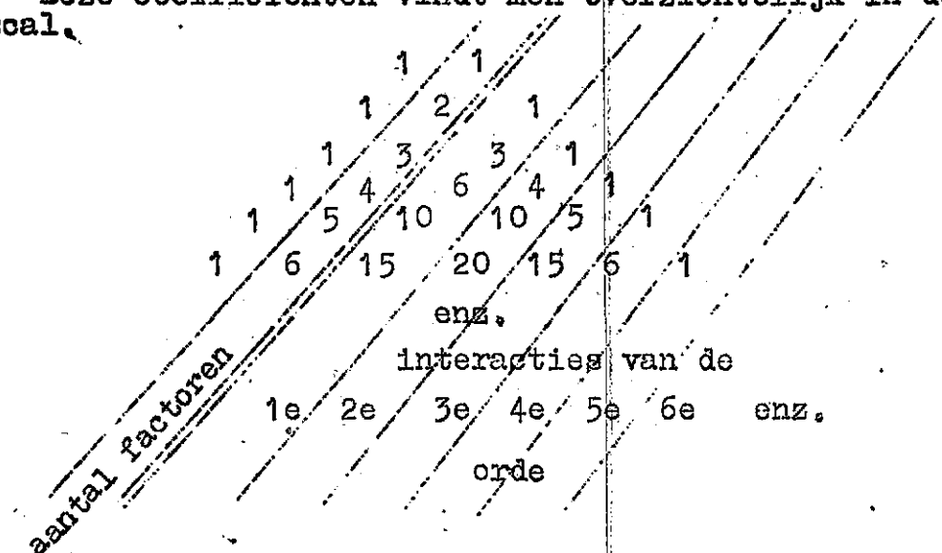
Het spel der factoren is daarmee evenzoo gedefiniecrd en wiskundig te bewerken.

De allergrootste moeilijkheid, die het factorial design ons stelt is ongetwijfeld het groote aantal objecten (combinaties van 2 of meer factoren), waartoe zij noodwendig voert, zoodra het aantal factoren maar even toeneemt.

Dat ook het aantal interacties zeer sterk toeneemt moge uit het volgende blijken:

Het aantal interacties wordt weergegeven door de coëfficienten van het binomium.

Deze coëfficienten vindt men overzichtelijk in de driehoek van Pascal.



waaruit het aantal interacties is af te lezen. Bij toenemend aantal factoren neemt het aantal interacties zeer sterk toe. Bij x factoren zijn er 2^x-x-1 interacties. Bij b.v. 10 factoren dus 1013 interacties!

Door de wiskundige formuleering der interacties kunnen wij ze echter alle ondubbelzinnig onderscheiden en bepalen.

De groote, vooral practische moeilijkheid is echter het aantal objecten (combinaties van trappen van 2 of meer factoren).

Zouden we b.v. de werking van 10 factoren - wat toch geen overmatige eisch is - willen onderzoeken, dan voert het factorial design, als we elken factor zeer poovertjes in b.v. 5 trappen zouden onderzoeken, ons reeds tot $5^{10} \backsimeq 9.765.000$ objecten[1]).

Of een ander voorbeeld: Vragen we hoeveel factoren we kunnen onderzoeken op een proefveld, zoo groot als het landoppervlak van de aarde d.i. 150.000.000 km^2, als we elken factor in slechts 5 trappen onderzoeken en de veldjesgrootte slechts 30 m^2 is. Het aantal factoren n vinden we dan uit: $5^n \times 30 = 150.000.000 \times 1.000.000$ bij uitdrukking in m^2. of $5^n = 5 \times 10^{12}$

$$n \log 5 = \log 5 + 12 \qquad n-1 = \frac{12}{\log 5} \backsimeq 17 \quad \text{of } n = 18$$

Inderdaad geen overdreven aantal, al mag dit aantal tegenwoordig niet gebruikelijk zijn, of liever, kan dit aantal met de huidige theorien nog niet bewerkt worden; geen wonder dat men het spel der factoren niet meester is!

Zoo kunnen we ons probleem thans weer opnieuw stellen:

Hoe moet ons ontwerp zijn, om een zelfde resultaat te benaderen met een zeer gereduceerd aantal objecten, of juister, welke beperkingen moeten we ons opleggen om toch een groot aantal factoren met succes te kunnen onderzoeken en beheerschen, met een voorafgestelde benadering van gegeven waarde.

Zonder meer is duidelijk, dat deze benadering gezocht moet worden in de verontachtzaming van de interacties van hooge orde. Uit het vorig hoofdstuk moge blijken, dat deze interacties niet alleen moeilijk voor te stellen zijn, doch ook dat hun betoekenis vermoedelijk steeds geringer wordt bij hoogere orde.

Stellen we zoodoende interacties van zekere orde èn alle hoogere a priori = 0, dan houdt dit in, dat het factorial design uiteenvalt in een aantal kleine sub-factorial designs. Stellen we in een factorial design van n factoren, de interacties van de pe orde en hooger a priori = 0, waarbij $p<n$, dan bestaan de sub-factorial designs uit p factoren. Aangezien tusschen elke mogelijke combinatie van p factoren, wel interacties van hoogstens $(p-1)^e$ orde kunnen bestaan is het aantal sub-factorial designs noodzakelijk $C_p(n) = \frac{n!}{p!(n-p)!}$

We moeten dus thans onderzoeken in hoeverre de interacties van hooge orde = 0 kunnen worden, hetgeen de inhoud van het volgende Hoofdstuk (V) vormt. Nadat we hierop antwoord hebben verkregen, kunnon we de mogelijkheid van een gereduceerd factorial design bespreken en de eventueele uitvoering daarvan, wat de inhoud van hoofdstuk VI en VII vormt.

---

[1]) volgens log-tafel met 5 decimalen.

HOOFDSTUK V

INTERPRETATIE VAN HET FACTORIAL DESIGN

§ 1. Fisher's analysis of variance.

De wiskundige uitwerking van experimenten genomen volgens het factorial design geschiedt haast zonder uitzondering met behulp van Fisher's analysis of variance. Voor zoover deze proeven met werkelijke herhalingen zijn genomen, levert deze methode geen moeilijkheden op, daar tot nog toe zelden meer dan 4 factoren tegelijkertijd onderzocht werden, en deze methode heeft bovendien het voordeel van zeer snel te zijn. Worden de proeven echter in enkelvoud genomen, wat bij toepassing van het factorial design met toenemend aantal factoren steeds noodzakelijker wordt, dan levert deze methode wel moeilijkheden op. Er is dan n.l. geen toevals variance, waaraan de overige variances getoetst kunnen worden. Fisher beveelt dan aan, de variances van de interacties van hoogere orde voor dit doel te gebruiken[1]). Hierin merkt Fisher echter ook reeds op, dat het soms kan voorkomen, dat één of meer van deze variances voor interacties van hoogere orde een groote waarde kunnen hebben, zoodat deze variances niet als toevalsvariance kunnen dienen.

In Hoofdstuk III § 4 zagen we reeds eenige geïdealiseerde gevallen, modellen, waar de variance van de interactie van hoogere orde grooter is ($\neq 0$), dan de variance van de interactie van lagere orde ($= 0$).

Bezien de we gevallen met 3 factoren, dan kunnen alle voorgesteld worden door de volgende z.g. regressievergelijking:
$y = aA + bB + cC + dAB + eAC + fBC + gABC + h.$
Zoo wordt geval 2 blz. 31 voorgesteld door:
$y = 8A - 4$ b t/m g = 0
Geval 5 wordt voorgesteld door:
$y = -12A - 12B + 8AB + 24$ c, e, f en g = 0
Geval 8 wordt voorgesteld door:
$y = -36A - 36B - 36C + 24AB + 24AC + 24BC - 16ABC + 64$
en geval 30 door:
$y = 60A + 60B + 72C - 40AB - 48AC - 48BC + 32ABC - 80.$

Bij eenzelfde type regressiefunctie wordt het gedrag der variances, dus beheerscht door de waarden, die de coëfficienten aannemen, als ze voor dezelfde trappen onderzocht worden.

Zouden we thans een experiment uitvoeren met 3 factoren, waarbij het mechanisme van het factorenspel geïdealiseerd wordt door b.v. het geval No. 30 (Hoofdstuk III § 4) dan treden door toevallige oorzaken eenige kleine schommelingen op in de waarnemingen.

Het geïdealiseerde geval zij weergegeven door

	C_1			C_2					B_1	B_2	
	B_1	B_2		B_1	B_2						
A_1	8	12	20	16	4	20	40	A_1	24	16	40
A_2	12	8	20	4	16	20	40	A_2	16	24	40
	20	20	40	20	20	20	80		40	40	80

Laten nu de werkelijke waarnemingen van een 1e proef zien:

	C_1			C_2					B_1	B_2	
	B_1	B_2		B_1	B_2						
A_1	7	13	20	18	4	22	42	A_1	25	17	42
A_2	14	5	19	4	17	21	40	A_2	18	22	40
	21	18	39	22	21	43	82		43	39	82

[1]) Fisher R.A. The design of experiments. Chapter VI. Experiments without replication

analysis of variance:

SS totaal $= 7^2 + \ldots\ldots\ldots 17^2 - \frac{82^2}{8} = 1084 - 840.5 = 243.5$

SS(A) $= \frac{42^2 + 40^2}{4} - cf = \frac{2^2}{8_2} = 0.5$

SS(B) $= \frac{43^2 + 39^2}{4} - cf = \frac{4^2}{8} = 2$

SS(C) $= \frac{39^2 + 43^2}{4} - cf = \frac{4^2}{8} = 2$

SS(AB) $= \frac{25^2 + 17^2 + 18^2 + 22^2}{2} - cf - A - B = 18$

SS(AC) $= \frac{20^2 + 19^2 + 22^2 + 21^2}{2} - cf - A - C = 0$

SS(BC) $= \frac{21^2 + 18^2 + 22^2 + 21^2}{2} - cf - B - C = 0.5$

SS(ABC) $= 243.5 - 23 = 220.5$

Daar het aantal vrijheidsgraden voor elke variance één is, geven bovenstaande cijfers tevens de variances weer.

Laten de werkelijke waarnemingen van de 2e proef vervolgens zijn:

	C_1			C_2							
	B_1	B_2		B_1	B_2				B_1	B_2	
A_1	8	14	22	18	4	22	44	A_1	26	18	44
A_2	15	7	22	6	18	24	46	A_2	21	25	46
	23	21	44	24	22	46	90		47	43	90

analysis of variance:

SS totaal $8^2 + \ldots\ldots\ldots 18^2 - \frac{90^2}{8} = 1234 - 1012.5 = 221.5$

SS(A) $= \frac{44^2 + 46^2}{4} - cf = \frac{2^2}{8_2} = 0.5$

SS(B) $= \frac{47^2 + 43^2}{4} - cf = \frac{4^2}{8_2} = 2$

SS(C) $= \frac{44^2 + 46^2}{2} - cf = \frac{2^2}{8} = 0.5$

SS(AB) $= \frac{26^2 + 18^2 + 21^2 + 25^2}{2} - cf - A - B = 18$

SS(AC) $= \frac{22^2 + 22^2 + 22^2 + 24^2}{2} - cf - A - C = 0.5$

SS(BC) $= \frac{23^2 + 21^2 + 24^2 + 22^2}{2} - cf - B - C = 0$

SS(ABC) $= 221.5 - 21.5 = 200$

welke waarden weer tevens de variances zijn.

Beschouwen we beide proeven thans gezamenlijk als herhalingen, dan kunnen we een toevals variance vinden.

		1e herhaling			2e herhaling			
		C_1	C_2		C_1	C_2		
B_1	A_1	7	18	25	8	18	26	51
	A_2	14	4	18	15	6	21	39
	A	13	4	17	14	4	18	35
B_2	A	5	17	22	7	18	25	47
		39	43	82	44	46	90	172

analysis of variance:

SS totaal $7^2 + \ldots\ldots\ldots 18^2 - \frac{172^2}{16} = 2318 - 1849 = 469$

SS herhalingen $\frac{82^2 + 90^2}{8} - cf = \frac{8^2}{16} = 4$

SS objecten $\frac{15^2 + \ldots\ldots\ldots 35^2}{2} - cf = 2312 - 1849 = 463$

SS toeval $469 - (4 + 463) = 2$

$$SS(A) = \frac{86^2 + 86^2}{8} - cf = 0$$

$$SS(B) = \frac{90^2 + 82^2}{8} - cf = \frac{8^2}{16} = 4$$

$$SS(C) = \frac{83^2 + 89^2}{8} - cf = \frac{6^2}{16} = 2.25$$

$$SS(AB) = \frac{51^2 + 39^2 + 35^2 + 47^2}{4} - cf - A - B = 36$$

$$SS(AC) = \frac{42^2 + 41^2 + 44^2 + 45^2}{4} - cf - A - C = 0.25$$

$$SS(BC) = \frac{44^2 + 39^2 + 46^2 + 43^2}{4} - cf - B - C = 0.25$$

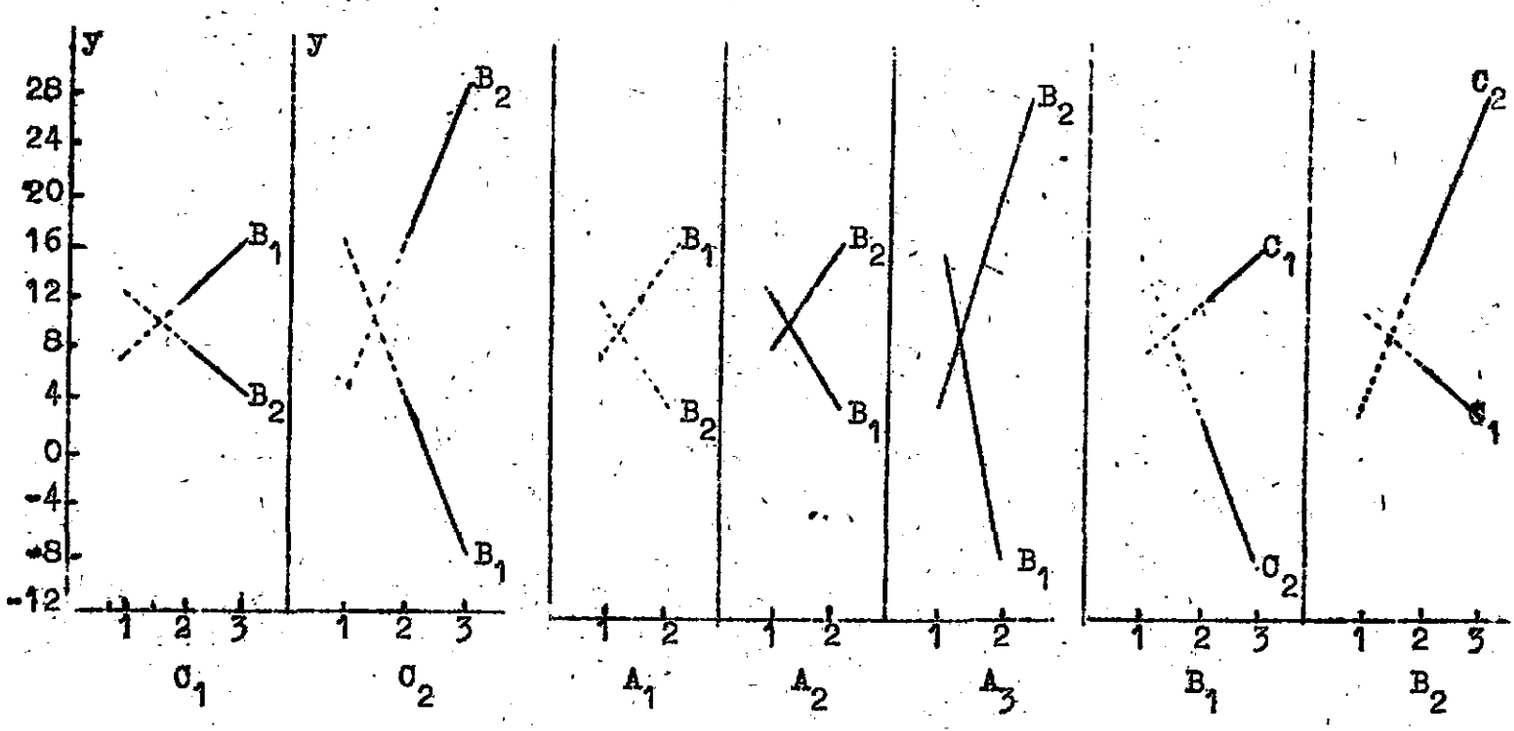
$SS(ABC) = 463 - 42.75 = 420.25$

Het totaal aantal vrijheidsgraden is 16 - 1 = 15, hiervan worden in de toevalsvariance 7 gebruikt, zoodat de toevalsvariance $\frac{7}{7} = 0.29$ wordt. De overige variances hebben alle 1 vrijheidsgraad, zoodat bovenstaande waarden tevens de variances voorstellen.

Vergelijken we thans de toevalsvariance met al de gevonden variances, dan blijkt de variance van de interactie van de hoogste orde een zeer slechte maat te zijn voor de toevalsvariance, zooals ook verwacht kan worden uit het geidealiseerde geval. Hetzelfde geldt ook voor de variance van de interactie van de 1e orde AB.

Hier ziet men dus een duidelijk voorbeeld, waar de door Fisher aangewezen weg, zonder meer, niet te bewandelen is. Veeleer zijn het hier de andere variances, die als schatting zouden kunnen dienen, nl. die variances die in het geidealiseerde geval = 0 zijn.

Voeren we thans nog een manipulatie uit met het geval No. 30, n.l. het toevoegen van een 3en trap aan den factor A.
We krijgen dan de volgende graphische voorstelling:



Beperken we onze beschouwingen, thans alleen tot de trappen A_2 en A_3 (naast B_1 en B_2 en C_1 en C_2, dus ook als een 2 x 2 x 2 experiment opgevat), dan zou thans een eventueele analysis of variance tot resultaat hebben:

		(No.30 pg	
V(A)	= 0		V(A) = 0
V(B)	≠ 0		V(B) = 0
V(C)	= 0		V(C) = 0
V(AB)	≠ 0		V(AB) ≠ 0
V(AC)	= 0		V(AC) = 0
V(BC)	≠ 0		V(BC) = 0
V(ABC)	≠ 0		V(ABC) ≠ 0)

Zooals we zien zijn thans, vergeleken met No. 30 pg. 35 ook de variances B en BC $\neq$ 0 geworden! Toch is het mechanisme van het factorenspel niet gewijzigd; in beide gevallen, dus het geval waarin we A_1 en A_2 beschouwen en het geval waarin we A_2 en A_3 beschouwen, kunnen de waarden, die Y voor de verschillende trappencombinaties aanneemt, zooals de berekening leert, gevonden worden uit de (regressie-) vergelijking $Y = 60A + 60B + 72X - 40AB - 48AC - 48BC - 32\,ABC - 80$. Het verschil in het gedrag der variances wordt geheel veroorzaakt doordat thans een ander interval van de gegeven functie $y = f(ABC)$ onderzocht wordt.

Zouden we als voren weer een experiment uitvoeren met dit factorenmechanisme, waarin we weer een kleine spreiding door toevallige oorzaken opnemen, dan kunnen thans ook de variances voor B en BC geen dienst meer doen voor een schatting van de toevalsvariance.

De $\neq$ 0 variances kunnen echter zeer klein zijn, wanneer ze als overgangsgevallen, die bijna de geïdealiseerde gevallen benaderen, beschouwd kunnen worden, zoodat ze onder zekere omstandigheden nog wel dienst zullen kunnen doen voor een schatting van de toevalsvariance.

Blijkbaar hangt een en ander af 1e van de vergelijking die het factorenspel weergeeft, 2e het interval waarin deze vergelijking onderzocht wordt.

§ 2. Ezekiel's "additive and joint functions".

Een andere methode van wiskundige verwerking is te ontleenen aan Ezekiel's methoden van "correlatie analyse"[1])(correlatie analyse hier in algemeener zin dan oorspronkelijk bedoeld werd). Ezekiel gebruikt hierbij de z.g. "additive and joint functions" om het statistisch materiaal te beschrijven en te rangschikken naar de verschillende factoren. Deze functies kan men opvatten als z.g. regressiefuncties. De meest bekende en veelvuldigst toegepaste correlatie theorie bij n veranderlijken is die van de rechte of normale correlatie, waarbij de factoren alle onafhankelijk van elkaar zijn, onafhankelijk hier weer gebruikt in den zin van geen interactie tusschen de factoren en de uitwerking van elken factor afzonderlijk, rechtevenredig is met den factor, kortom de regressiefunctie de volgende gedaante heeft:

$$y = a_1x_1 + a_2x_2 + a_3x_3 + \ldots\ldots + a_nx_n + c$$

Bij Ezekiel's additive functions vervalt de voorwaarde van lineariteit, zoodat de regressiefunctie als volgt geschreven kan worden.

$$y = a_{11}x_1 + a_{12}x^2 + a_{13}x^3 + \ldots\ldots$$
$$+ a_{21}x_2 + a_{22}x_2^2 + \ldots\ldots$$
$$+ a_{n1}x_n + a_{n2}x_n^2 + \ldots\ldots a_{nn}x_n^n + C$$

of afgekort $y = f_1(x_1) + f_2(x_2) + \ldots\ldots f_n(x_n)$

Hierbij kunnen de functies f_1, f_2 enz. alle een verschillende gedaante hebben. In den vervolge zal dit niet telkens aangeduid worden, zoodat we schrijven: $y = f(x_1) + f(x_2) + \ldots\ldots f(x_n)$ (1).

Bij Ezekiel's joint functions vervalt ook de voorwaarde van afhankelijkheid der factoren.

Bij 2 factoren kunnen we deze afhankelijkheid uitdrukken als $y = f(x_1, x_2)$: voor elke combinatie van waarden van x_1 en x_2 bestaat een bepaalde waarde voor y.

Een eenvoudig voorbeeld is het volgende:

$$y = a + bx_1 + (c + dx_1)\,x_2$$

d.w.z. voor een constante waarde van x_1 is de regressie van y t.o.v. x_2 lineair, maar als x_1 verandert, verandert deze regressie eveneens.

In sommige gevallen is het noodig de gecombineerde invloed van 2 factoren te onderzoeken, terwijl daarnaast één of meer onafhankelijke factoren onderzocht worden, we kunnen dit dan schrijven als $y = f(x_1x_2) + f(x_3)$. Bij een groot aantal factoren kunnen ook meerdere joint-functions tegelijkertijd onderzocht worden b.v. $y = f(x_1x_2) + f(x_3x_4) + f(x_5)$.

---

[1]) Ezekiel, M. Methods of correlation analysis. New York 1930; in het bijzonder hoofdstuk 20 en 21.

Maar er zullen gevallen zijn, waar een zoo flexibele vergelijking als bovenstaande nog niet voldoende is; de verandering in een factor kan een verandering in alle andere factoren teweeg brengen, dit wordt dan uitgedrukt door $y = f(x_1, x_2, x_3, \ldots\ldots, x_n)$ (2)

Vergelijkingen (1) en (2) stellen de meest algemeene vormen voor van Ezekiel's addditive respectievelijk joint functions. De vrije-hand curven geven dan de basis om een definitieven vorm van de functies te kiezen, die dan aangepast kunnen worden, volgens de methode der kleinste quadraten. Bij onafhankelijke factoren is dit niet zoo moeilijk, wel echter bij joint-functions. Het eenvoudigste geval is dat waar de regressie van y t.o.v. x_2 lineair is voor een gegeven waarde van x_1, maar waarbij de regressie lineair verandert als de waarde van x_1 verandert:

$$y = a + b(c + dx_1)\ x_2 \text{ of uitgeschreven}$$

$y = a + bcx_2 + bdx_1x_2$ of $y = a + ex_2 + gx_1x_2$

Nemen we nog aan dat y, behalve door x_1 in den term $g\,x_1\,x_2$, bovendien verandert op een andere wijze, dan kunnen we nog een term toevoegen b.v. $y = a + ex_2 + gx_1x_2 + hx_1$ een geval boven reeds vermeld.

Dezelfde methode kan uitgebreid worden tot n factoren met eveneens lineaire regressie van y t.o.v. elk der factoren x_1 t/m x_n, terwijl de overige constant zijn (partieele regressie), terwijl die regressie verandert als die factoren veranderen. Voor 3 factoren vinden we dan

$y = a + bx_1 + cx_2 + dx_3 + ex_1x_2 + gx_1x_3 + hx_2x_3 + ix_1x_2x_3$

Ingewikkelder worden de volgende gevallen: nemen we aan, dat de regressie y t.o.v. x_1 bij een bepaalde waarde van x_2 verloopt volgens een kromme lijn, en dat de vorm van deze kromme lijn zich wijzigt volgens een veranderlijke maat als x_2 verandert, dan kunnen we dit weergeven door $y = a + f(b + \varphi x_2)\ x_1$. Deze vergelijking kan benaderend worden voorgesteld door $y = a + f(x_1) + f(x_1x_2) + f(x_2)$. Gebruiken we vervolgens x_1, x_2 en het product x_1x_2 als onafhankelijke factoren, dan kunnen de vormen van de 3 functies vastgesteld worden. Werken we daarbij de waarden y uit bij de verschillende combinaties van x_1 en x_2, dan kunnen zeer grillig gekromde regressieoppervlakken ontstaan. Ook kan dit voor meerdere veranderlijken uitgebreid worden, voor 3 veranderlijken vinden we:

$y = a + f(x_1) + f(x_2) + f(x_3) + f(x_1x_2) + f(x_1x_3) + f(x_2x_3) + f(x_1x_2x_3)$.

Een andere methode, de z.g. Court methode berust op rotatie van het regressieoppervlak; hoewel in principe gelijk aan het voorgaande, is dit toepasbaar op een iets ander type van oppervlak. Het karakteristieke er van is het gebruik van een samengestelde functie, die 2 of meer onafhankelijke factoren vertegenwoordigt. Bij gebruik van deze combinatie functies, bij een verschillende rotatie, kunnen zeer grillige regressieoppervlakken vereffend worden. Er is echter zeer veel overleg noodig, om deze rotaties goed uit te kiezen voor de vereffening begint. Bij uitbreiding tot n factoren geldt dit in zeer ernstige mate, terwijl het aantal constanten zeer groot wordt. Overigens zij verwezen naar Ezekiel. Daar ons doel is, juist met zeer veel factoren te werken, moeten we een meer rationeele methode zeeken, dan in bovenstaande kort geciteerd is.

Nog zij vermeld wat Derksen[1]) hierover zegt, hetgeen tevens als een korte samenvatting van bovenstaande kan dienen: "Men zou verwachten, dat we als joint functions allereerst de vergelijking zouden gebruiken, die men krijgt, door de reeks van Taylor voor meerdere variabelen achter de termen van den n'graad af te breken. Dit zou dan overeenkomen met hetgeen we bij de enkelvoudige correlatie gedaan hebben, toen wij een parabool van den n'graad aanpasten. Ezekiel doet dit echter niet, maar beschouwt wel verschillende lineaire combinaties van machten en producten van de veranderlijken". (De Court-methode refereert D. blijkbaar niet).

§ 3. De φ-regressiefuncties.

Schrijver gaat nu als volgt te werk, waarbij ook alleen machten en producten der variabelen beschouwd worden. De eenvoudigste joint-function is dan $y = kAB + b$. (Hierbij worden de factoren in plaats van x_1 x_2..... x_n nu A, B, C enz. genoemd, wat bij veel factoren overzichtelijker is). De aanpassingsmogelijkheden van de gegevens aan dit regres-

1) J.B.M. Derksen. Inleiding tot de correlatierekening. pg. 45, 46. Delft 1935.

sievlak kan nu sterk uitgebreid worden, door den oorsprong van de veranderlijken willekeurig aan te nemen. Dit komt dus neer op een translatie: $A' = A+\alpha$, $B' = B+\beta$ zoodat
$y = kAB + l = k(A'-\alpha)(B'-\beta) = kA'B' - k\beta A' - k\alpha B' + k\alpha\beta$ of algemeen en met weglating der accenten, $y = a + bA + cB + dAB$. Omgekeerd kunnen we dus dadelijk, door deze laatste formule te gebruiken met willekeurige waarden voor de coëfficienten, een betere aanpassing verkrijgen $y = a + bA + cB + dAB$ stelt een algebraïsch oppervlak (hyperbolische paraboloïde) voor, verkregen door translatie van de hyperbolische paraboloïde $y = kAB + l$.

$$y = a + bA + cB + dAB = d\left(AB + \frac{c}{d}B + \frac{b}{d}A + \frac{a}{d}\right) =$$
$$d\left(A + \frac{c}{d}\right)\left(B + \frac{b}{d}\right) + \left(a - \frac{bc}{d}\right)$$

Evenzoo ontstaat uit $y = kABC + l$ de algemeene formule
$y = a + bA + cB + dC + eAB + fAC + gBC + hABC$
Uit $y = kA^2B^2 + l$ ontstaat:

$$y = k(A'-\alpha)^2(B'-\beta)^2 + l = k(A'^2 - 2\alpha A' + \alpha^2)(B'^2 - 2\beta B' + \beta^2) + l =$$
$$= kA'^2B'^2 - 2\beta kA'^2B' + \beta^2 kA'^2 - 2\alpha kA'B'^2 + 4\alpha\beta kA'B' - 2\alpha k\beta^2 A'$$
$$+ k\alpha^2 B'^2 - 2k\alpha^2\beta B' + k\alpha^2\beta^2 = 0 \text{ of algemeen:}$$

$$y = a + bA + cA^2 + dB + eB^2 + fAB + gA^2B + iAB^2 + jA^2B^2$$

Uit $y = kA^3B + l$ ontstaat:
$y = a + bA + cA^2 + dA^3 + eB + gA^3B + hA^2B + iAB;$

Waren we in deze laatste formule uitgegaan van een vorm waarin ook B tot de 3e macht voorkwam dus $y = kA^3B^3 + l$, dan was het aantal termen waar A en B gezamenlijk in voorkomen gestegen van 3 tot 9. Dit is eveneens het geval als we de indicatie van Derksen volgen door de reeks van Taylor voor meer veranderlijken achter de termen van den n'graad af te breken. Dit maakt het aantal termen onnoodig zeer hoog.

Dit is te illustreeren aan het volgende voorbeeld: twee factoren A en B geven grafisch voorgesteld de volgende y-waarden:

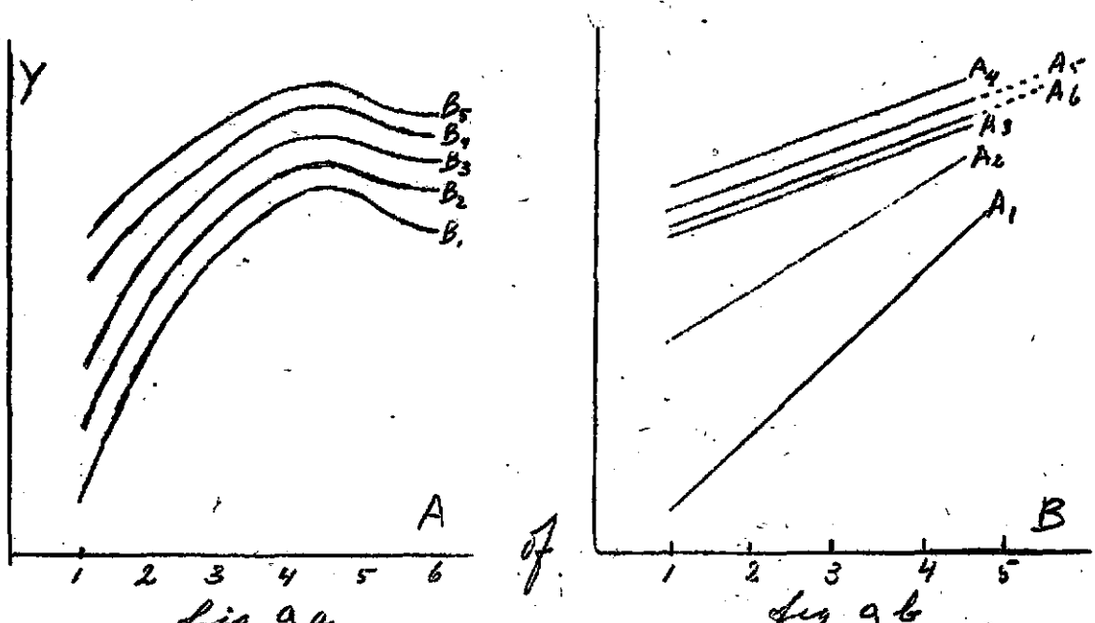


fig 9a fig 9b

Het resultaat van factor A kan bij constante waarde van B weergegeven worden door een 3'graadsfunctie, terwijl het resultaat van factor B bij constante waarde van factor A weergegeven kan worden door een lineaire functie. Was er geen interactie tusschen de factoren (b.v. kenbaar aan het feit dat alle rechten in fig. 9b evenwijdig zouden loopen) dan kon het regressievlak voorgesteld worden

door $y = a + bA + cA^2 + dA^3 + eB$. Is er wel interactie dan kan in dit geval, waar factor B eeb lineaire functie is, voldoende aanpassing verkregen worden door de 3 termen uit de laatst genoemde formule $y = a + bA + CA^2 + dA^3 + eB + gA^3B + hA^2B + iAB$. Het gebruik van 9 termen is hier overbodig. Dit laten voorkomen van de verschillende factoren tot een verschillenden graad kan een groote vereenvoudiging beteekenen.

We zullen nu de aldus verkregen regressie functies aldus afkorten:

bij 2 factoren $y = \varphi_1(A) + \varphi_2(B) + \varphi_3(A,B) +$ constante

De functies φ_1, φ_2 en φ_3 kunnen alle verschillend zijn. In den vervolge laten we echter deze indices eenvoudigheidshalve weg! In $\varphi_3(A,B)$ komen alleen producten van machten van A en B voor, in den vervolge laten we daarom de komma weg en schrijven eenvoudigheidshalve $\varphi(AB)$, zoo ook $\varphi(ABC)$ enz.

bij 3 factoren $y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB) + \varphi(AC) + \varphi(BC) + \varphi(ABC) +$ constante enz.; in $\varphi(A)$ komen dus alleen machten van A voor in $\varphi(AB)$ alleen producten van machten van A en B, in $\varphi(AC)$ alleen producten van machten van A en C enz., in $\varphi(ABC)$ alleen gedurige producten van machten van A, B en C.

Hierbij zij nog nadrukkelijk vermeld, dat als de graad van de functie $\varphi(A), \varphi(B), \varphi(C)$,..... enz gegeven is, de functies $\varphi(AB)$, $\varphi(AC)$ ..... $\varphi(ABC)$ ..... $\varphi(ABCD)$ ..... enz. dus volkomen bepaald zijn volgens het voorgaande, omgekeerd geldt hetzelfde. Na de vereffening is echter wel verschil tusschen de graad, waarin b.v. B in $\varphi(B)$ en $\varphi(AB)$ voorkomt mogelijk! B.v. $\varphi(A)$ is van den 1e graad, $\varphi(B)$ is van den 3' graad. Dan is $\varphi(AB)$ in de vereffeningsfunctie volkomen bepaald, de regressiefunctie luidt dan $y = a + bA + cB + dB^2 + eB^3 + gAB + hAB^2 + iAB^3$. Na de vereffening kan door een bijzonderen stand van het regressieoppervlak t.o.v. het coördinatenstelsel, blijken dat de vereffende formule b.v. moet zijn $y = a + bA + cB + gAB + hAB^2 + iAB^3$, waarin $\varphi(B)$ dus van lageren graad is dan wel de andere mogelijkheid b.v. $y = a + bA + cB + dB^2 + eB^3 + gAB$, waarin $\varphi(AB)$ van een lageren graad is.

Een dergelijke schrijfwijze leidt tot zeer groote vereenvoudigingen, zooals we later zullen zien, weshalve ook het symbool angstvallig voor deze beteekenis bewaard blijve. $\varphi(A)$, $\varphi(AB)$ enz. noemen we de termen van de φ-regressiefunctie.

§ 4. De betrekkingen tusschen de analysis of variance en de φ-regressiefuncties bij 2 factoren.

We zullen thans onderzoeken hoe Fisher's analysis of variance samenhangt met deze regressiefuncties. Bij 2 factoren luidde deze regressie vergelijking $y = \varphi(A) + \varphi(B) + \varphi(AB) +$ constante. Deze constante laten we in den vervolge weg, daar deze geen invloed op de variances heeft.

We kunnen nu afleiden welke variances hierbij behooren. De totale variance kan, zooals bekend, volgens Fisher gesplitst worden in 3 afzonderlijke variances, welke splitsing gebaseerd is op de volgende identiteit:

$$\sum_{k=1}^{n}\sum_{\rho=1}^{m}(y_{k\rho}-\bar{y})^2 = \sum_{k=1}^{n}\sum_{\rho=1}^{m}(\bar{y}_{k.}-\bar{y})^2 + \sum_{k=1}^{n}\sum_{\rho=1}^{m}(\bar{y}_{.\rho}-\bar{y})^2 + \sum_{k=1}^{n}\sum_{\rho=1}^{m}(y_{k\rho}-\bar{y}_{k.}-\bar{y}_{.\rho}+\bar{y})^2$$

Hierin is dus

$$y_{k\rho} = \varphi(A_k) + \varphi(B_\rho) + \varphi(A_kB_\rho)$$

$$\bar{y}_{k.} = \frac{\sum_{\rho=1}^{m}\{\varphi(A_k)+\varphi(B_\rho)+\varphi(A_kB_\rho)\}}{m}$$

$$\bar{y}_{.\rho} = \frac{\sum_{k=1}^{n}\{\varphi(A_k)+\varphi(B_\rho)+\varphi(A_kB_\rho)\}}{n}$$

$$\bar{y} = \frac{\sum_{k=1}^{n}\sum_{\rho=1}^{m}\{\varphi(A_k)+\varphi(B_\rho)+\varphi(A_kB_\rho)\}}{nm}$$

We laten in den vervolge de vele indices weg, indien hier geen verwarring door ontstaat.
Nu is dus:

$$\sum_{k=1}^{n}\sum_{l=1}^{m}(y_{kl}-\bar{\bar{y}})^2=\sum_{}^{n}\sum_{}^{m}\left\{\varphi(A)+\varphi(B)+\varphi(AB)-\frac{\sum\sum\{\varphi(A)+\varphi(B)+\varphi(AB)\}}{nm}\right\}^2=$$

$$\sum\sum\left\{\varphi(A)+\varphi(B)+\varphi(AB)-\frac{\sum\varphi(A)}{n}-\frac{\sum\varphi(B)}{m}-\frac{\sum\sum\varphi(AB)}{nm}\right\}^2=$$

$$\sum\sum\left\{\varphi^2(A)+\varphi^2(B)+\varphi^2(AB)+\frac{\left(\sum\varphi(A)\right)^2}{n^2}+\frac{\left(\sum\varphi(B)\right)^2}{m^2}+\frac{\left(\sum\sum\varphi(AB)\right)^2}{n^2m^2}+\right.$$

$$2\varphi(A)\varphi(B)+2\varphi(A)\varphi(AB)-\frac{2\varphi(A)\sum\varphi(A)}{n}-\frac{2\varphi(A)\sum\varphi(B)}{m}$$

$$-\frac{2\varphi(A)\sum\sum\varphi(AB)}{nm}+2\varphi(B)\varphi(AB)-\frac{2\varphi(B)\sum\varphi(A)}{n}-\frac{2\varphi(B)\sum\varphi(B)}{m}$$

$$-\frac{2\varphi(B)\sum\sum\varphi(AB)}{nm}-\frac{2\varphi(AB)\sum\varphi(A)}{n}-\frac{2\varphi(AB)\sum\varphi(B)}{m}-\frac{2\varphi(AB)\sum\sum\varphi(AB)}{nm}$$

$$\left.\frac{2\sum\varphi(A)\sum\varphi(B)}{nm}+\frac{2\sum\varphi(A)\sum\sum\varphi(AB)}{nm^2}+\frac{2\sum\varphi(B)\sum\sum\varphi(AB)}{n^2m}\right\}=$$

$$=m\sum\varphi^2(A)+n\sum\varphi^2(B)+\sum\sum\varphi^2(AB)+\frac{m}{n}\left(\sum\varphi(A)\right)^2+\frac{n}{m}\left(\sum\varphi(B)\right)^2+$$

$$+\frac{1}{nm}\left(\sum\sum\varphi(AB)\right)^2+2\sum\varphi(A)\sum\varphi(B)+2\sum\left(\varphi(A)\sum\varphi(AB)\right)$$

$$-\frac{2m}{n}\left(\sum\varphi(A)\right)^2+2\sum\varphi(A)\sum\varphi(B)-\frac{2}{n}\sum\varphi(A)\sum\sum\varphi(AB)$$

$$+2\sum\left(\varphi(B)\sum\varphi(AB)\right)-2\sum\varphi(A)\sum\varphi(B)-\frac{2n}{m}\left(\sum\varphi(B)\right)^2$$

$$-\frac{2}{m}\sum\varphi(B)\sum\sum\varphi(AB)-\frac{2}{m}\sum\sum\varphi(AB)\sum\varphi(A)-\frac{2}{m}\sum\sum\varphi(AB)\sum\varphi(B)$$

$$-\frac{2}{nm}\left(\sum\sum\varphi(AB)\right)^2+2\sum\varphi(A)\sum\varphi(B)+\frac{2}{n}\varphi(A)\sum\sum\varphi(AB)+$$

$$+\frac{2}{m}\sum\varphi(B)\sum\sum\varphi(AB)=$$

$$m\sum\varphi^2(A)-\frac{m}{n}\left(\sum\varphi(A)\right)^2+n\sum\varphi^2(B)-\frac{n}{m}\left(\sum\varphi(B)\right)^2+$$

$$+\sum\sum\varphi^2(AB)-\frac{1}{nm}\left(\sum\sum\varphi(AB)\right)^2+2\sum\left(\varphi(A)\sum\varphi(AB)\right)-$$

$$-\frac{2}{n}\sum\varphi(A)\sum\sum\varphi(AB)+2\sum\left(\varphi(B)\sum\varphi(AB)\right)-$$

$$-\frac{2}{m}\sum\varphi(B)\sum\sum\varphi(AB).$$

De totale variance is dus van $\varphi(A)$, $\varphi(B)$ en $\varphi(AB)$ afhankelijk.
Bepalen we thans:

$$\sum\sum\left(\bar{y}_{k.}-\bar{\bar{y}}\right)^2=\sum\sum\left\{\frac{\sum\{\varphi(A)+\varphi(B)+\varphi(AB)\}}{m}-\frac{\sum\sum\{\varphi(A)+\varphi(B)+\varphi(AB)\}}{nm}\right\}^2=$$

$$=\sum\sum\left\{\varphi(A)+\frac{1}{m}\sum\varphi(AB)-\frac{1}{n}\sum\varphi(A)-\frac{1}{nm}\sum\sum\varphi(AB)\right\}^2=$$

$$= \sum_n \sum_m \Big\{ \varphi^2(A) + \frac{1}{m^2}\Big(\sum_m \varphi(AB)\Big)^2 + \frac{1}{n^2}\Big(\sum_n \varphi(A)\Big)^2 + \frac{1}{n^2 m^2}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 +$$

$$+ \frac{2}{m}\varphi(A)\sum_m \varphi(AB) - \frac{2}{m}\varphi(A)\sum_n \varphi(A) - \frac{2}{nm}\varphi(A)\sum_n \sum_m \varphi(AB)$$

$$- \frac{2}{nm}\sum_m \varphi(AB)\sum_n \varphi(A) - \frac{2}{nm^2}\sum_m \varphi(AB)\sum_n \sum_m \varphi(AB) + \frac{2}{nm^2}\sum_n \varphi(A)\sum_n \sum_m \varphi(AB)\Big\} =$$

$$= m\sum_n \varphi^2(A) + \frac{1}{m}\sum_n\Big(\sum_m \varphi(AB)\Big)^2 + \frac{m}{n}\Big(\sum_n \varphi(A)\Big)^2 + \frac{1}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 +$$

$$+ 2\sum_n\Big(\varphi(A)\sum_m \varphi(AB)\Big) - \frac{2m}{n}\Big(\sum_n \varphi(A)\Big)^2 - \frac{2}{n}\sum_n \varphi(A)\sum_n \sum_m \varphi(AB)$$

$$- \frac{2}{nm}\sum_n \sum_m \varphi(AB)\sum_n \varphi(A) - \frac{2}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 + \frac{2}{n}\sum_n \varphi(A)\sum_n \sum_m \varphi(AB) =$$

$$m\sum_n \varphi^2(A) - \frac{m}{n}\Big(\sum_n \varphi(A)\Big)^2 + \frac{1}{m}\sum_n\Big(\sum_m \varphi(AB)\Big)^2 - \frac{1}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2$$

$$+ 2\sum_n\Big(\varphi(A)\sum_m \varphi(AB)\Big) - \frac{2}{n}\sum_n \varphi(A)\sum_n \sum_m \varphi(AB) \qquad (I)$$

De variance voor factor A is alleen afhankelijk van $\varphi(A)$ en $\varphi(AB)$. Uitgevallen is $\varphi(B)$. Evenzoo is de variance

$\sum_n \sum_m (\bar{y}_{\cdot i} - \bar{\bar{y}})^2 / m-1$ alleen afhankelijk van $\varphi(B)$ en $\varphi(AB)$

Beschouwen we thans nog:

$$\sum_n \sum_m (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{\bar{y}})^2 = \sum_n \sum_m \Big\{ \varphi(A) + \varphi(B) + \varphi(AB) -$$

$$- \frac{\sum_m \{\varphi(A) + \varphi(B) + \varphi(AB)\}}{m} - \frac{\sum_n \{\varphi(A) + \varphi(B) + \varphi(AB)\}}{n} + \frac{\sum_n \sum_m \{\varphi(A) + \varphi(B) + \varphi(AB)\}}{nm}\Big\}^2$$

$$= \sum_n \sum_m \Big\{ \varphi(A) + \varphi(B) + \varphi(AB) - \varphi(A) - \frac{1}{m}\sum_m \varphi(B) - \frac{1}{m}\sum_m \varphi(AB) - \frac{1}{n}\sum_n \varphi(A) - \varphi(B)$$

$$- \frac{1}{n}\sum_n \varphi(AB) + \frac{1}{n}\sum_n \varphi(A) + \frac{1}{m}\sum_m \varphi(B) + \frac{1}{nm}\varphi(AB)\Big\}^2 =$$

$$= \sum_n \sum_m \Big\{ \varphi(AB) - \frac{1}{m}\sum_m \varphi(AB) - \frac{1}{n}\sum_n \varphi(AB) + \frac{1}{nm}\sum_n \sum_m \varphi(AB)\Big\}^2 =$$

$$= \sum_n \sum_m \Big\{ \varphi^2(AB) + \frac{1}{m^2}\Big(\sum_m \varphi(AB)\Big)^2 + \frac{1}{n^2}\Big(\sum_n \varphi(AB)\Big)^2 + \frac{1}{n^2 m^2}\Big(\sum_n \sum_m \varphi(AB)\Big)^2$$

$$+ \frac{2}{m}\varphi(AB)\sum_m \varphi(AB) - \frac{2}{m}\varphi(AB)\sum_n \varphi(AB) + \frac{2}{nm}\varphi(AB)\sum_n \sum_m \varphi(AB)$$

$$+ \frac{2}{nm}\sum_m \varphi(AB)\sum_n \varphi(AB) - \frac{2}{nm^2}\sum_m \varphi(AB)\sum_n \sum_m \varphi(AB) -$$

$$- \frac{2}{n^2 m}\sum_n \varphi(AB)\sum_n \sum_m \varphi(AB)\Big\} =$$

$$\sum_n \sum_m \varphi^2(AB) + \frac{1}{m}\sum_n\Big(\sum_m \varphi(AB)\Big)^2 + \frac{1}{n}\sum_m\Big(\sum_n \varphi(AB)\Big)^2 + \frac{1}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2$$

$$- \frac{2}{m}\sum_n\Big(\sum_m \varphi(AB)\Big)^2 - \frac{2}{n}\sum_m\Big(\sum_n \varphi(AB)\Big)^2 + \frac{2}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2$$

$$+ \frac{2}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 - \frac{2}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 - \frac{2}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 =$$

$$\sum_n \sum_m \varphi^2(AB) - \frac{1}{m}\sum_n\Big(\sum_m \varphi(AB)\Big)^2 - \frac{1}{n}\sum_m\Big(\sum_n \varphi(AB)\Big)^2 + \frac{1}{nm}\Big(\sum_n \sum_m \varphi(AB)\Big)^2 \qquad (II)$$

In de variance AB komt dus slechts nog de term $\varphi(AB)$ van de regressievergelijking voor; zoowel $\varphi(A)$ als $\varphi(B)$ zijn er uitgevallen!

Uit de formules (I) en (II) volgt:
Is dus in de regressievergelijking $y = \varphi(A) + \varphi(B) + \varphi(AB)$ de term $\varphi(AB) = 0$, dan gaat dit samen met het feit dat de variance voor de interactie van de 1e orde = 0 is, onafhankelijk van de waarde van $\varphi(A)$ en $\varphi(B)$.

Is daarentegen $\varphi(AB) \neq 0$ dan zal de interactievariance nooit = 0 zijn, de variances van de factoren A en B in het algemeen ook niet, hoewel hierop uitzonderingen mogelijk zijn.

Een overzicht over de waarden die de variances voor den factor A en voor de interactie AB aannemen, bij gegeven waarden voor $\varphi(A)$ en $\varphi(AB)$ volgt hieronder:

	$\varphi(AB) = 0$		$\varphi(AB) \neq 0$	
Gegeven:	$\varphi(A) = 0$	$\varphi(A) \neq 0$	$\varphi(A) = 0$	$\varphi(A) \neq 0$
Var AB	$= 0$	$= 0$	$\neq 0$	$\neq 0$
Var A	$= 0$	$\neq 0$	$= 0$ of $\neq 0$	$= 0$ of $\neq 0$

Voor den factor B geldt een analoog overzicht. (=0 of ≠ 0 beteekent natuurlijk in het algemeen ≠ 0,= 0 is slechts een bijzonder geval, is een uitzondering, maar kan voorkomen).

Omgekeerd, vinden wij in een experiment met 2 factoren een betrouwbare variance voor de interactie van de 1e orde, getoetst aan de toevalsvariance, die we vinden op een nog later te beschrijven methode, dan moet in de regressiefunctie de term $\varphi(AB)$ voorkomen. Daarnaast zal in het algemeen $\varphi(A)$ voorkomen, doch deze term kan ook = 0 zijn. Vinden wij een betrouwbare variance voor den factor A, dan moet als de interactie variance = 0, cq. $\varphi(AB) = 0$, in de regressiefunctie de term $\varphi(A)$ voorkomen, in het andere geval kan $\varphi(A)$ ook = 0 zijn.

Een overzicht over het optreden van de termen uit de regressiefunctie, bij gegeven variances volgt hieronder:

	Var AB = 0		Var AB ≠ 0	
Gegeven	Var A = 0	Var A ≠ 0	Var A = 0	Var A ≠ 0
$\varphi(AB)$	$= 0$	$= 0$	$\neq 0$	$\neq 0$
$\varphi(A)$	$= 0$	$\neq 0$	$= 0$ of $\neq 0$	$= 0$ of $\neq 0$

Bij de voorbeelden zullen enkele gevallen besproken worden.

In onze terminologie van main-action en interplay kunnen we zeggen, dat de interplay uitgedrukt wordt door den term $\varphi(AB)$ de main-action door $\varphi(A)$ of door $\varphi(AB)$ of door beide tegelijk!

§ 5. Variances uitgedrukt in de contanten van de regressievergelijking.

Door voor de termen $\varphi(A)$, $\varphi(B)$ en $\varphi(AB)$ uit de regressievergelijking bepaalde vormen aan te nemen, kunnen we de variances in de contanten van deze vormen uitdrukken. We houden ons daar bij aan het gestelde in § 3. We zullen dit alleen uitwerken voor vormen van $\varphi(AB)$ waarin A en B tot den 1en of 2en graad voorkomen om ruimte te besparen. Komen A en B beide tot de 1e macht in $\varphi(AB)$ voor, dan wordt de regressievergelijking $y = aA + bB + cAB + d$.

We vinden dan voor $\sum\sum(\bar{y}_{i.} - \bar{y})^2$ door substitutie van $\varphi(A) = aA$ $\varphi(B) = bB$ en $\varphi(AB) = cAB$ in formule (I) (de constante d heeft geen invloed op de variances)

$$\sum_n\sum_m(\bar{y}_{i.} - \bar{y})^2 = m\sum_n a^2A^2 + \frac{1}{m}\sum_n\left(\sum_m cAB\right)^2 - \frac{m}{n}\left(\sum_n aA\right)^2 -$$

$$- \frac{1}{nm}\left(\sum_n\sum_m cAB\right)^2 + 2\sum_n\left(aA\sum_m cAB\right) - \frac{2}{n}\sum_n aA\sum_n\sum_m cAB =$$

$$= mna^2\Sigma A^2 + \frac{c^2}{m}\Sigma A^2(\Sigma B)^2 - \frac{a^2m}{n}(\Sigma A)^2 - \frac{c^2}{nm}(\Sigma A)^2(\Sigma B)^2 +$$

$$+2ac\,\Sigma A^2\Sigma B - \frac{2ac}{n}(\Sigma A)^2\Sigma B =$$

$$\left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{m a^2 + 2ac\,\Sigma B + \frac{c^2}{m}(\Sigma B)^2\right\} =$$

$$\left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{a\sqrt{m} + \frac{c\,\Sigma B}{\sqrt{m}}\right\}^2 \qquad (3)$$

terwijl we met behulp van (II) voor den teller van de interactie-variance vinden:

$$\sum_m\sum_m(y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{y})^2 = \sum_m\sum_m c^2A^2B^2 - \frac{1}{m}\sum_n\left(\sum_m cAB\right)^2 - \frac{1}{m}\sum_m\left(\sum_n cAB\right)^2 +$$

$$+ \frac{1}{nm}\left(\sum_m\sum_n cAB\right)^2 = c^2\Sigma A^2\Sigma B^2 - \frac{c^2}{m}\Sigma A^2(\Sigma B)^2 - \frac{c^2}{n}(\Sigma A)^2\Sigma B^2 + \frac{c^2}{nm}(\Sigma A)^2(\Sigma B)^2 =$$

$$= c^2\left\{\Sigma B^2 - \frac{(\Sigma B)^2}{m}\right\}\left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\} \qquad (4)$$

In een orthogonaal proefplan b.v. het factorial design, zullen n en m in den regel klein zijn, terwijl ΣA, ΣB, ΣA^2 en ΣB^2 eenvoudige waarden aannemen.

In de voorbeelden zal dit voor een paar gevallen uitgewerkt worden.

Deelen door (n-1) respectievelijk (n-1)(m-1) geeft de variances voor den factor A en de interactie AB.

We moeten nu de sommatievormen nog anders schrijven. Dit zijn n.l. de veranderlijken in de formules (3) en (4) als we n en m constant houden c.q. een bepaald proefplan voor oogen hebben. We nemen daartoe aan, dat de factoren onderzocht kunnen worden in de trappen ..... -k ..... -2, -1, 0, +1, +2, +3 .....k .....
Het is nu duidelijk, dat de waarde der sommatievormen afhankelijk is, van het interval der regressiefunctie, waarin we ons bevinden, daar we altijd slechts een beperkt aantal trappen in een proefplan onderzoeken.

Is b.v. $m^* = 4$, d.w.z. wordt factor B onderzocht in 4 trappen, dan is, als deze trappen zijn b.v. 3, 4, 5 en 6 de sommatievorm $\Sigma B = 18$, zijn de trappen echter 12, 13, 14 en 15 dan is $\Sigma B = 54$, waarmee zich de waarde voor de variance A wijzigt.

Nemen we den laagsten trap k^*, dan worden bij $n^* = 4$ de trappen k^*, k^*+1, k^*+2 en k^*+3 en wordt $\Sigma A = 4k^* + 6$. ΣA^2 wordt dan reeds ingewikkelder. We willen echter ook n^* veranderlijk laten, de trappen kunnen we dan schrijven als k^*, k^*+1, ... k^*+n^*-1 en we vinden dan $\Sigma A = n^*k^* + n^{*2} - \frac{n^*}{2}$. Om zoo ook ΣA^2 en later bij de 2e graads vormen ook ΣA^3 en ΣA^4 te bepalen moeten we onze toevlucht nemen tot de formules voor de reeksen van hoogere orde of tot de veeltermen van Bernoulli.

§ 6. Substitutie van de reeksen van hoogere orde.

Hierbij wordt de som van een aantal termen uitgedrukt in de 1e termen van de betrokken reeks en haar verschilreeksen met als coëfficienten de binomiaal coëfficienten.

Wij hebben dan in ons geval

de reeks: k k+1 k+2 ..... k+n-1
1e verschil reeks: 1 1 1 .....
Som van n termen dus b.v.

$$\sum_k^{k+n-1} A = C_1(n).k + C_2(n).1 = nk + \frac{n(n-1)}{1.2} =$$

$$= nk + \frac{n^2}{2} - \frac{n}{2} = n\left(k + \frac{n-1}{2}\right) \qquad (5)$$

de reeks: k^2 $(k+1)^2$ $(k+2)^2$ $(k+3)^2$ ..... $(k+n-1)^2$
1e verschil reeks: 2k+1 2k+3 2k+5 .....
2e verschil reeks: 2 2 2 .....

Som van n termen, dus b.v. $\sum_{k}^{k+n-1} A^2 =$

$$= C_1(n)k + C_2(n).(2k+1) + C_3(n).2 =$$
$$= nk^2 + n^2k - nk + \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} =$$
$$= n\left\{k^2 + (n-1)k + \frac{(n-1)(2n-1)}{6}\right\} =$$
$$= n\left\{\left(k + \frac{n-1}{2}\right)^2 + \frac{n^2-1}{12}\right\} \qquad (6)$$

de reeks: $k^3 \quad (k+1)^3 \quad (k+2)^3 \quad (k+3)^3 \quad (k+4)^3 \ldots\ldots\ldots (k+n-1)^3$

1e verschil reeks: $3k^2 + 3k + 1 \quad 3k^2 + 9k + 7 \quad 3k^2 + 15k + 19$
$3k^2 + 21k + 37 \ldots\ldots\ldots\ldots$

2e verschil reeks: $6k + 6 \quad 6k + 12 \quad 6k + 18 \ldots\ldots\ldots$

3e verschil reeks: $6 \quad 6 \quad 6 \ldots\ldots\ldots$

Som van n termen, dus b.v. van $\sum_{k}^{k+n-1} A^3 =$

$$= C_1(n)k^3 + C_2(n).(3k^2+3k+1) + C_3(n).(6k+6) + C_4(n).6 =$$
$$= nk^3 + \left(\frac{3n^2}{2} - \frac{3n}{2}\right)k^2 + \left(n^3 - \frac{3}{2}n^2 + \frac{n}{2}\right)k + \frac{n^4}{4} - \frac{n^3}{2} + \frac{n^2}{4} =$$
$$= n\left\{k^3 + \frac{3(n-1)}{2}k^2 + \frac{(n-1)(2n-1)}{2}k + \frac{n(n-1)^2}{4}\right\} \qquad (7)$$

de reeks: $k^4 \quad (k+1)^4 \quad (k+2)^4 \quad (k+3)^4 \quad (k+4)^4 \quad (k+5)^4 \ldots\ldots (k+n-1)^4$

1e verschil reeks: $4k^3 + 6k^2 + 4k + 1 \quad 4k^3 + 30k^2 + 76k + 65 \quad 4k^3 + 54k^2 + 244k + 369$
$4k^3 + 18k^2 + 28k + 15 \quad 4k^3 + 42k^2 + 148k + 175$

2e verschil reeks: $12k^2 + 24k + 14 \quad 12k^2 + 48k + 50 \quad 12k^2 + 72k + 110$
$12k^2 + 96k + 194 \ldots\ldots\ldots$

3e verschil reeks: $24k + 36 \quad 24k + 60 \quad 24k + 84 \ldots\ldots\ldots$

4e verschil reeks: $24 \quad 24 \ldots\ldots\ldots\ldots$

Som van n termen, dus b.v. van $\sum_{k}^{k+n-1} A =$

$$= C_1(n)k + C_2(n).(4k^3+6k^2+4k+1) + C_3(n).(12k^2+24k+14) + C_4(n).(24k+36) + C_5(n).(24)_5 =$$
$$nk^4 + 2(n^2-n)k^3 + (2n^3+3n^2+n)k^2 + (n^4-2n^3+n^2)k + \frac{n^5}{5} - \frac{n^4}{2} + \frac{n^3}{3} - \frac{n}{30} =$$
$$= n\left\{k^4 + 2(n-1)k^3 + (n-1)(2n-1)k^2 + n(n-1)^2k + \frac{(n-1)(2n-1)(3n^2-3n-1)}{30}\right\} \qquad (8)$$

Evenzoo beteekent ℓ^* de 1e trap waarin de factor B onderzocht wordt, waarbij dus n^* en m^* het aantal trappen is, waarin respectievelijk de factoren A en B onderzocht worden. We spreken af, daar dit hier geen verwarring kan wekken, in den vervolge deze indices* weg te laten, om de schrijfwijze zoo eenvoudig mogelijk te houden; k beteekent dan dus zoowel de 1e trap van de trappen waarin factor A onderzocht wordt, als de argumentwaarde voor de regressiefunctie! Hetzelfde geldt voor ℓ.

Substitueeren we nu deze formules (5) en (6) in de formules (3) en (4) die de tellers der variances weergeven:

$$\sum^{n}\sum^{m}(\bar{y}_{\ell} - \bar{\bar{y}})^2 = \left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{m a^2 + 2ac\,\Sigma B + \frac{c^2}{m}(\Sigma B)^2\right\} =$$
$$= \left\{nk^2 + n^2k - nk + \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} - \frac{\left(nk + \frac{n^2}{2} - \frac{n}{2}\right)^2}{n}\right\} \times$$
$$m\left\{a^2 + 2ac\left(\ell + \frac{m}{2} - \frac{1}{2}\right) + c^2\left(\ell + \frac{m}{2} - \frac{1}{2}\right)\right\} =$$
$$= m\left\{\frac{n^3}{12} - \frac{n}{12}\right\}\left\{c^2\ell^2 + (2ac + c^2m - c^2)\ell + a^2 + acm - ac + \frac{c^2m^2}{4} + \frac{c^2}{4} - \frac{c^2m}{2}\right\} =$$
$$= \frac{mn(n^2-1)}{12}\left\{c^2\ell^2 + c(2a+cm-c)\ell + \frac{(2a+cm-c)^2}{4}\right\} - \frac{mn(n^2-1)}{12}\left\{c\ell + \frac{(2a+cm-c)}{2}\right\}^2 \qquad (9)$$

$$\sum_{k=1}^{n}\sum_{l=1}^{m}(y_{kl}-\bar{y}_{k.}-\bar{y}_{.l}+\bar{\bar{y}})^2 = c^2\left\{\sum A^2-\frac{(\sum A)^2}{n}\right\}\left\{\sum B^2-\frac{(\sum B)^2}{m}\right\} =$$

$$c^2\left\{\frac{m^3}{12}-\frac{m}{12}\right\}\left\{\frac{n^3}{12}-\frac{n}{12}\right\} = \frac{c^2 mn(n^2-1)(m^2-1)}{144} \qquad (10)$$

Hiermede hebben we de tellers der variances uitgedrukt als een functie van k, ℓ, m en n en de constanten uit de regressiefunctie, dus als een functie van plaats en grootte van het interval, waarin de regressiefunctie beschouwd wordt, wat dus aangegeven is door de trappen van het proefplan, waarbij de factoren onderzocht worden.

§ 7. Het gedrag der variances.

Beschouwen we thans eerst de variances afzonderlijk. De interactie variance is dan:

$$V(AB) = \frac{\sum_{k=1}^{n}\sum_{l=1}^{m}\{y_{kl}-\bar{y}_{k.}-\bar{y}_{.l}+\bar{\bar{y}}\}^2}{(n-1)(m-1)} = \frac{c^2 mn(n+1)(m+1)}{144}$$

dus voor een gegeven proefplan alleen afhankelijk van de constante c uit de regressievergelijking, dus constant bij een bepaalde regressievergelijking!
De variance voor factor A is:

$$V(A) = \frac{\sum_{k=1}^{n}\sum_{l=1}^{m}(\bar{y}_{k.}-\bar{\bar{y}})^2}{n-1} = \frac{mn(n+1)}{12}\left\{c\ell+\frac{(2a+cm-c)}{2}\right\}^2$$

dus bij een gegeven aantal trappen nog afhankelijk van het interval waarin de regressiefunctie t.o.v. de factor B onderzocht wordt.

Gaan we het verloop van deze variance na:
Uit $V(A) = \frac{mn(n+1)}{12}\left\{c\ell+\frac{(2a+cm-c)}{2}\right\}^2$ volgt, dat de waarde voor variance voor den factor A een minimum vertoont en overigens parabolisch verloopt. De plaats van het minimum vinden we uit

$$c\ell+\frac{2a+cm-c}{2} = 0 \quad \text{of} \quad \ell_{mn} = \frac{-(2a+cm-c)}{2c} = \frac{a}{c}+\frac{1-m}{2}$$

Substitutie geeft de grootte van deze variance in dit minimum nl.

$$\frac{mn(n+1)}{12}\left\{\frac{-(2a+cm-c)}{2}+\frac{(2a+cm-e)}{2}\right\} = 0$$

Voor waarden van $\ell >$ of $<$ dan ℓ_{mn} neemt de variance toe in waarde. Alles conform het overzicht op pg.51 waar bij $\varphi(AB) \neq 0$ en $\varphi(A) \neq 0$ voor de Var. A opgegeven staat = 0 of $\neq$ 0. Hier vinden we dus de voorwaarden daartoe.

§ 8. De onderlinge verhouding der variances.

Door thans de verhouding $H = \frac{V(A)-V(AB)}{V(AB)}$ (11) na te gaan, is het mogelijk te onderzoeken welke waarden de variances t.o.v. elkaar kunnen aannemen. Bovenstaande verhouding is ook te schrijven:

$$H = \frac{V(A)}{V(AB)}-1 \Rightarrow \frac{\frac{\sum_{k=1}^{n}\sum_{l=1}^{m}(\bar{y}_{k.}-\bar{\bar{y}})^2}{(n-1)}}{\frac{\sum_{k=1}^{n}\sum_{l=1}^{m}(y_{kl}-\bar{y}_{k.}-\bar{y}_{.l}+\bar{\bar{y}})^2}{(n-1)(m-1)}}-1 = \frac{(m-1)\sum_{k=1}^{n}\sum_{l=1}^{m}(\bar{y}_{k.}-\bar{\bar{y}})^2}{\sum_{k=1}^{n}\sum_{l=1}^{m}(y_{kl}+\bar{y}_{k.}-\bar{y}_{.l}+\bar{\bar{y}})^2}-1$$

Krachtens formules (9) en (10) vinden we dan:

$$H = \frac{\left\{c\ell+\frac{(2a+cm-c)}{2}\right\}^2}{\frac{c^2(m+1)}{12}}-1 \qquad (12)$$

Daar ℓ niet in V(AB) voorkomt is de minimum voorwaarde dezelfde, die we hierboven reeds vonden:
nl. $\ell = -\frac{a}{c}+\frac{1-m}{2}$ (15)

Daar H van den 2en graad is in ℓ, kan het verloop der verhouding dus voor elke waarde van m als een parabool geschetst worden:

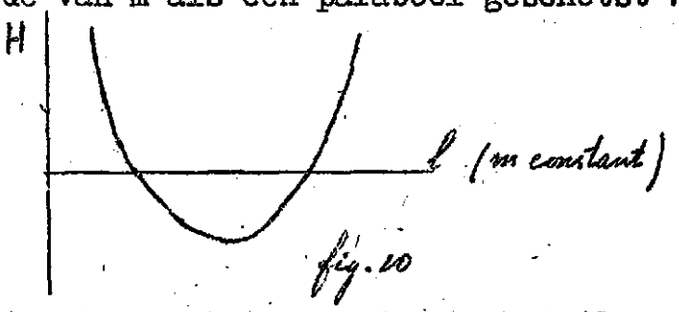


fig. 10

M.a.w. mogen we nu reeds concludeeren dat in het algemeen de variance van de interactie van de 1e orde kleiner zal zijn, dan de variance van factor A (natuurlijk bij deze bijzondere regressievergelijking).

Het aantal trappen waarin de factor B wordt onderzocht, is op zijn minst 2 ($m \geq 2$) en zal zelden grooter dan 10 zijn. Loopt m van 2 tot 10 en hooger dan neemt ℓ de volgende waarden aan:

$m = 2$	$\ell = -\frac{1}{2} - \frac{a}{c}$	$m = 5$	$\ell = -2 - \frac{a}{c}$
$m = 3$	$\ell = -1 - \frac{a}{c}$	$m = 6$	$\ell = -3\frac{1}{2} - \frac{a}{c}$
$m = 4$	$\ell = -1\frac{1}{2} - \frac{a}{c}$	$m = 7$	$\ell = -3 - \frac{a}{c}$ enz.

Bezien we dit graphisch, dan vinden we, dat het interval, waarvoor H minimum wordt, zich symmetrisch van uit een bepaald punt $-\frac{a}{c}$ uitbreidt. Het punt $-\frac{a}{c}$ wordt door de regressievergelijking bepaald.

Voor enkele waarden van m vinden we b.v. onderstaand schema voor de trappen waarin factor B onderzocht wordt, wil H minimaal zijn, als we daarbij nog bedenken, dat ℓ beteekent zoowel 1e trap van factor B, als argument waarde voor de regressiefunctie.

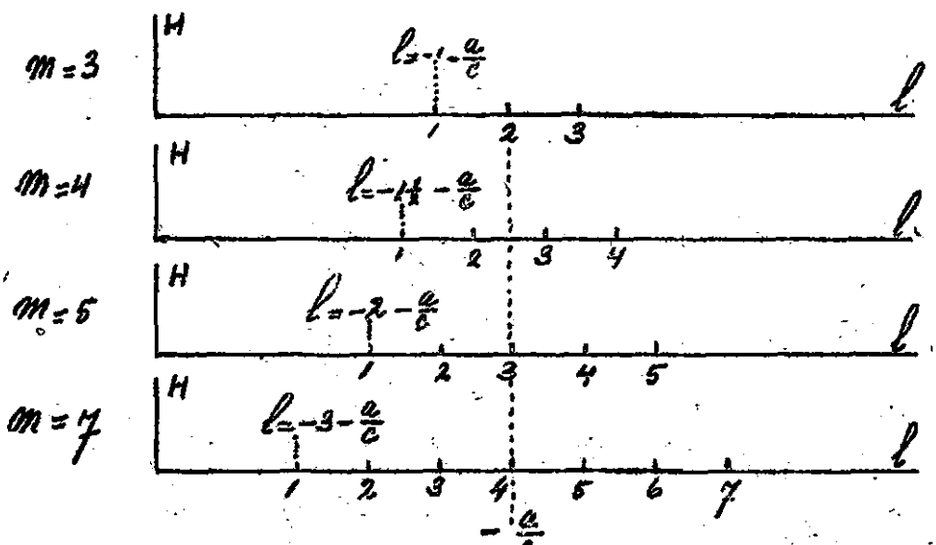


fig 11

De waarde van het minimum vinden we door substitutie van $\ell = -\frac{a}{c} + \frac{1-m}{2}$ in vergelijking (12)
We vinden dan H = -1 (16) d.w.z. als de variance voor factor A $\neq$ 0, hetgeen ook uit (11) kon volgen.

Vragen we thans nog naar de snijpunten met de ℓ-as van de parabool, die de verhouding H weergeeft d.w.z. we vragen wanneer V(A) = V(AB).

$$\frac{c^2 nm(n+1)(m+1)}{144} = \frac{nm(n+1)}{12}\left\{c\ell + \frac{(2a+cm-c)}{2}\right\}^2 \quad \text{of}$$

$$\left(c\sqrt{\frac{m+1}{12}}\right)^2 = \left\{c\ell + \frac{(2a+cm-c)}{2}\right\}^2 \quad \text{of}$$

$$\ell = -\frac{a}{c} + \frac{1-m}{2} \pm \sqrt{\frac{m+1}{12}} \qquad (17)$$

Bij een bepaalde m vinden we dus de volgende situatie:

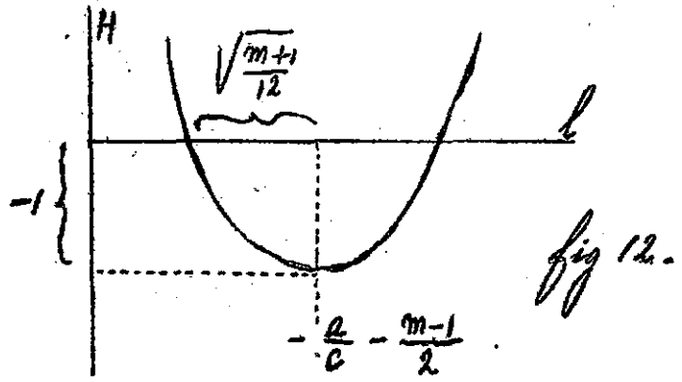


fig 12.

Heeft dus de 1e trap een kleinere waarde dan $-\frac{a}{c}+\frac{1-m}{2}-\sqrt{\frac{m+1}{12}}$ of een grootere waarde dan $-\frac{a}{c}+\frac{1-m}{2}+\sqrt{\frac{m+1}{12}}$ dan zal de variance van de interactie altijd kleiner zijn dan de variance voor den factor A.

Opmerking: Is a = 0 en b = 0, dus is de regressievergelijking $y=\varphi(AB)$ i.c. y = cAB + d dan vinden we voor

$$V(AB)=\frac{c^2}{(n-1)(m-1)}\left\{\Sigma A^2-\frac{(\Sigma A)^2}{n}\right\}\left\{\Sigma B^2-\frac{(\Sigma B)^2}{m}\right\}$$

$$V(A)=\frac{c^2}{(n-1)m}\left\{\Sigma A^2-\frac{(\Sigma A)^2}{n}\right\}(\Sigma B)^2$$

We vinden dan op dezelfde wijze V(A) minimaal en = 0 voor $l=\frac{1-m}{2}$ eveneens H minimaal en = -1 voor $l=\frac{1-m}{2}$

Dit is ook direct te vinden uit de formule $l=-\frac{a}{c}+\frac{1-m}{2}$ door hierin a = 0 te stellen.

Stellen we daarentegen c = 0 in deze formule dan vinden we $l=-\infty$ Deze uitkomst heeft geen zin, daar als c = 0 er geen interactie is. Hoe eenvoudig dit ook moge zijn, in het hierna te behandelen geval, waarin A en B in den vorm $\varphi(AB)$ beide tot de 2e macht voorkomen en 9 coëfficienten optreden, zijn soortgelijke gevallen minder doorzichtig.

Hiermede zijn dus de aanwijzingen van Fisher als vermeld in § 1 voor een eenvoudig geval reeds exact aangegeven. De verhouding der variances is dus afhankelijk van grootte en plaats van het interval waarin de regressiefunctie onderzocht wordt. Voor elk gegeven interval kunnen we de variance berekenen met behulp van de coëfficienten uit de regressievergelijking.

Evenzoo wordt het gedrag van de variance voor den factor B t.o.v. de variance van de interactie, bepaald door grootte en plaats van het interval waarin de regressiefunctie onderzocht wordt, in het bijzonder door het aantal en waarde der trappen waarbij factor A onderzocht wordt.

§ 9. Voorbeeld 1.

We beginnen met $y=\varphi(AB)$ b.v. y = 2AB + 11.5

A loopt van ..... -4, -3, -2, -1, 0 , +1, +2 ..... k .....

B loopt van ..... -4, -3, -2, -1, 0 , +1, +2 ..... l .....

Graphisch voorgesteld aldus

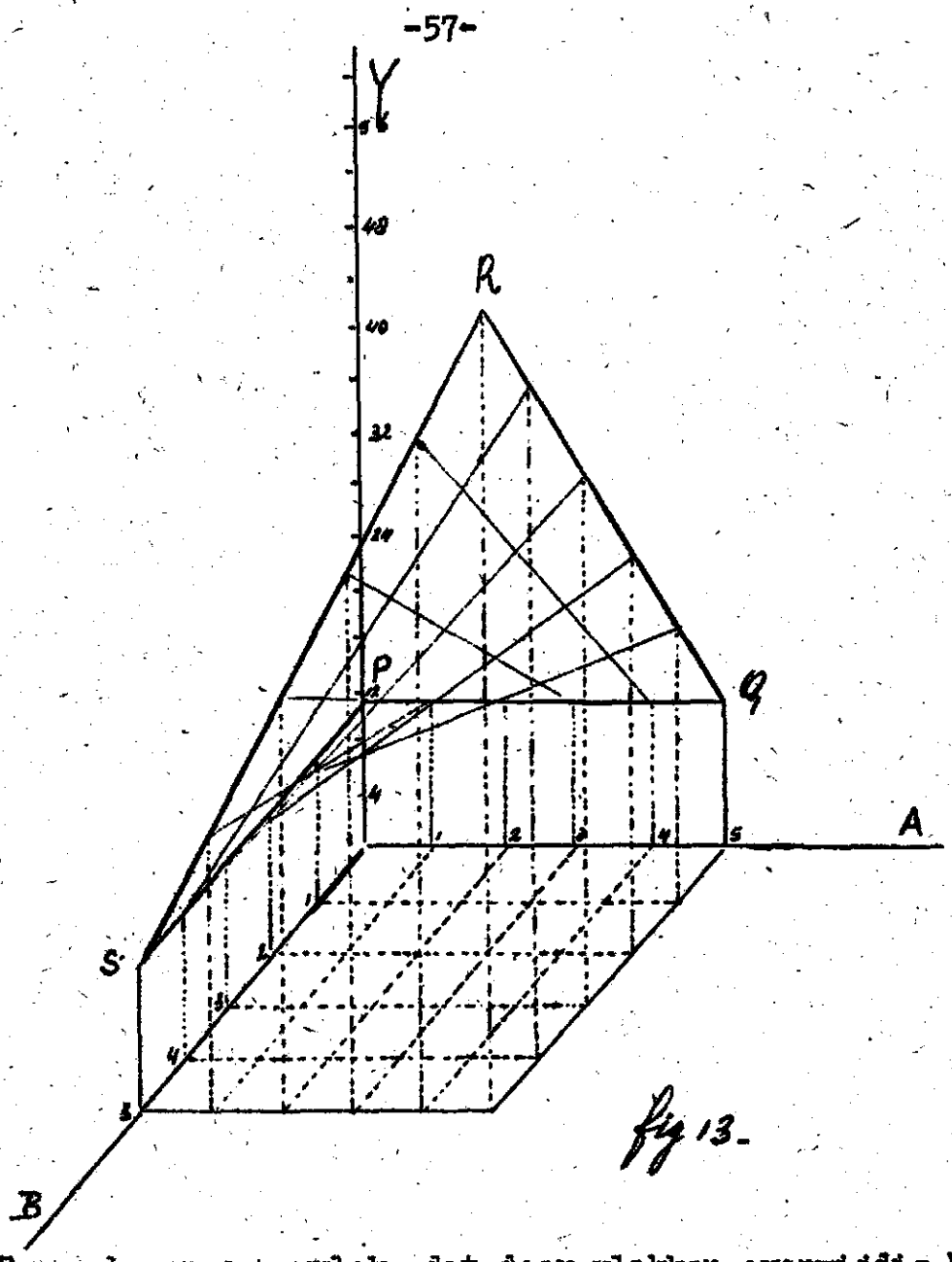


fig 13.

Een gebogen oppervlak, dat door vlakken evenwijdig het A-Y vlak en evenwijdig het B-Y vlak volgens rechte lijnen gesneden wordt; het A-Y vlak en het B-Y vlak zelve worden gesneden door de rechten PQ en PS evenwijdig het grondvlak op een hoogte y = 11.5. Bezien wij het vlak van uit een ander standpunt en iets meer perspectivisch, dan wordt de kromming duidelijker zichtbaar:

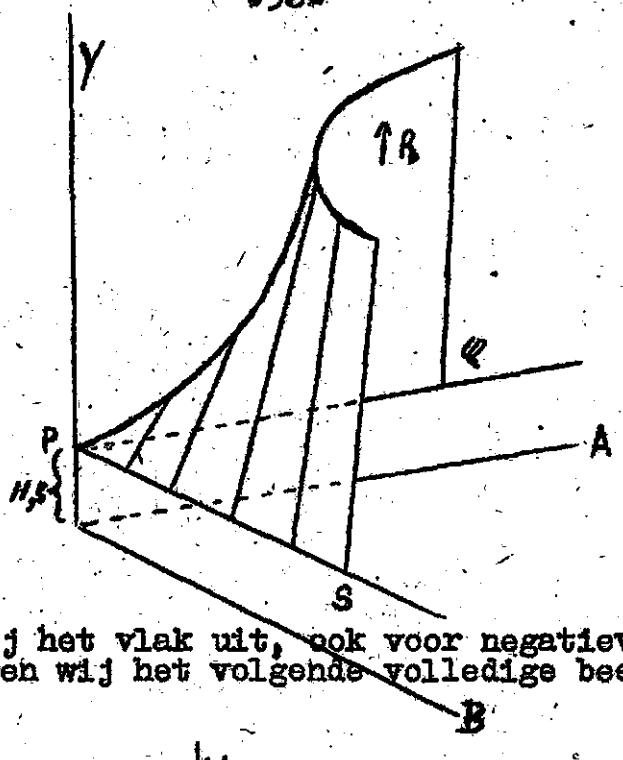


fig 14.

Breiden wij het vlak uit, ook voor negatieve waarden van A en B, dan verkrijgen wij het volgende volledige beeld.

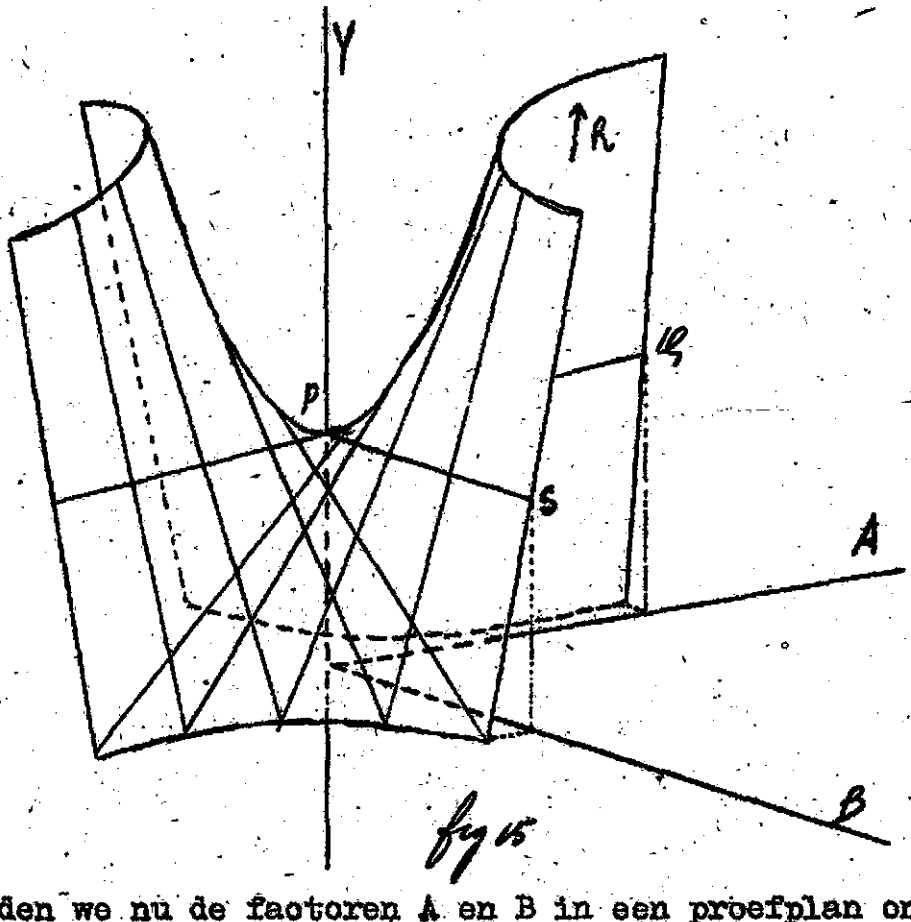


fig 15

Zouden we nu de factoren A en B in een proefplan onderzoeken, elke factor in b.v. 4 trappen A_2 A_3 A_4 en A_5 en B_1 B_2 B_3 en B_4 dan zouden we de volgende waarden voor y vinden:

	A_2	A_3	A_4	A_5	
B_1	15.5	17.5	19.5	21.5	74
B_2	19.5	23.5	27.5	31.5	102
B_3	23.5	29.5	35.5	41.5	130
B_4	27.5	35.5	43.5	51.5	158
	86	106	126	146	464

Een analysis of variance geeft dan het volgende resultaat

SS totaal $15.5^2 + \ldots\ldots 51.5^2 - \frac{464^2}{16} = 1.5036 - 1.3456 = 1580$

SS(A) $\frac{86^2 + \ldots 146^2}{4} - cf = 1.3956 - 1.3456 = 500$

SS(B) $\frac{74^2 + \ldots 158^2}{4} - cf = 1.4436 - 1.3456 = 980$

SS(AB) $1580 - (500 + 980) = 100$

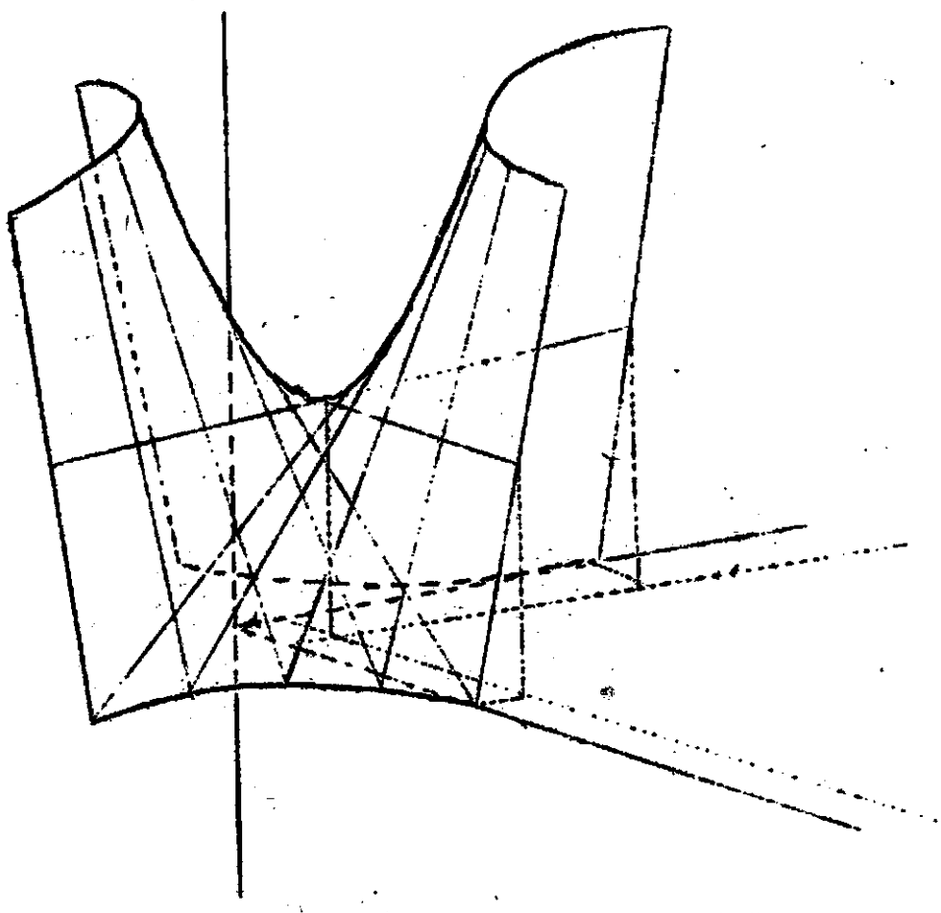
variatieoorzaak	sommen van quadraten	vrijheids-graden	variance
A	500	3	167
B	980	3	327
AB	100	9	11
	1580	15	

Leiden we uit de regressiefunctie thans de analysis of variance af. In de regressievergelijking was $\varphi(A) = 0, \varphi(B) = 0$ $\varphi(AB) \neq 0$. Volgens het 1e overzicht volgende uit de formules (I) en (II) moet dan een interactievariance gevonden worden, en in het algemeen ook een variance voor de factoren A en B, wat in dit voorbeeld dus het geval is.

Leiden we omgekeerd uit de resultaten van de analysis of variance de regressiefunctie af: vinden we een interactievariance, dan moet volgens het 2e overzicht volgende uit de formules (I) en (II) in de regressiefunctie de term $\varphi(AB)$ voorkomen en in het algemeen ook $\varphi(A)$ en $\varphi(B)$, hoewel uitzonderingen mogelijk zijn; hier hebben wij dus zoo'n uitzondering.

Deze uitzondering wordt als volgt begrijpelijk: passen we op de gegeven regressiefunctie een translatie toe en gaan we b.v. over op de coördinaten $A' = A + 1.5$ en $B' = B + 1.5$.

Wij vinden dan $y = 2AB + 11.5 = 2(A'-1.5)(B' - 1.5) + 11.5$
In het volgende laten we de accenten gemakshalve weg en bemoeien ons niet meer met het oude coördinatenstelsel. We vinden dan
$y = -3A-3B+2AB+16$ dus $y = \varphi(A) + \varphi(B) + \varphi(AB)$. Graphisch voorgesteld aldus:



Zouden we nu weer de factoren A en B in een proefplan onderzoeken elke factor weer in 4 trappen en hemen we daarvoor thans $A_{3.5}$ $A_{4.5}$ $A_{5.5}$ en $A_{6.5}$ en voor den factor B de trappen $B_{2.5}$ $B_{3.5}$ $B_{4.5}$ en $B_{5.5}$ dan zouden wc de volgende waarden vinden:

	$A_{3.5}$	$A_{4.5}$	$A_{5.5}$	$A_{6.5}$
$B_{2.5}$	15.5	17.5	19.5	21.5
$B_{3.5}$	19.5	23.5	27.5	31.5
$B_{4.5}$	23.5	29.5	35.5	41.5
$B_{5.5}$	27.5	35.5	43.5	51.5

Dus precies dezelfde waarden als bij het 1e proefplan. De analysis of variance zal natuurlijk ook precies hetzelfde resultaat geven. En de conclusies daaruit ten aanzien van de regressievergelijking zijn gelijkluidend. Alleen hebben we thans niet meer met een uitzondering te doen: $\varphi(A)$ en $\varphi(B)$ komen ook in de regressievergelijking voor zooals volgens § 4 in het algemeen ook voorkomt. De uitzondering werd dus veroorzaakt door een zeer bijzondere ligging van het regressievlak t.o.v. het coördinatenstelsel.

We bemoeien ons verder alleen maar weer met het meer algemeene geval i.c. $y = -3A - 3B + 2AB + 16$.

We hadden nu de uitkomsten van de analysis of variance ook volgens voorgaande theorie uit deze regressiefunctie kunnen bepalen. En wel met behulp van de formules (I) en (II) uit § 4. Voor deze regressiefunctie bij 2 factoren is dit reeds uitgewerkt in de theorie van § 5, in de formules (3) en (4):

$$\sum_k^n \sum_l^m (\bar{y}_{k.} - \bar{\bar{y}})^2 = \left\{ \sum A^2 - \frac{(\sum A)^2}{n} \right\} \left\{ a\sqrt{m} + \frac{c \sum B}{\sqrt{m}} \right\}^2$$

$$\sum_k^n \sum_l^m (y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{\bar{y}})^2 = c^2 \left\{ \sum A^2 - \frac{(\sum A)^2}{n} \right\} \left\{ \sum B^2 - \frac{(\sum B)^2}{m} \right\}$$

In ons laatste proefplan is dus $m = n = 4$

$\sum A = 3.5 + 4.5 + 5.5 + 6.5 = 20$ $\quad \sum A^2 = 105$

$\sum B = 2.5 + 3.5 + 4.5 + 5.5 = 16$ $\quad \sum B^2 = 69$

terwijl $a = -3$ $\quad b = -3$ $\quad c = +2$

Gesubstitueerd geeft dit

$$\sum_k^n \sum_l^m (\bar{y}_{k.} - \bar{\bar{y}})^2 = \left\{ 105 - \frac{20^2}{4} \right\} \left\{ -3 \times 2 + \frac{2 \times 16}{2} \right\}^2 = 500$$

$$\sum_k^n \sum_l^m (y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{\bar{y}}) = 4 \left\{ 105 - \frac{20^2}{4} \right\} \left\{ 69 - \frac{16^2}{4} \right\} = 100$$

gelijk wij in de analysis of variance ook vonden. De totale variance en de variance voor factor B kunnen desgewenscht evenzoo berekend worden.

Omgekeerd is het in dit geval mogelijk, uit de analysis of variance de coëfficienten van de regressievergelijking te bepalen, op de constante na, eveneens met dezelfde formules:

$$\sum_k^n \sum_l^m (\bar{y}_{k.} - \bar{\bar{y}})^2 = \left\{ 105 - \frac{20^2}{4} \right\} \left\{ a \times 2 + \frac{2 \times 16}{2} \right\}^2 = 500$$

$$\sum_k^n \sum_l^m (\bar{y}_{.l} - \bar{\bar{y}})^2 = \left\{ 69 - \frac{16^2}{4} \right\} \left\{ b \times 2 + \frac{2 \times 20}{2} \right\}^2 = 980$$

$$\sum_k^n \sum_l^m (y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{\bar{y}})^2 = c^2 \left\{ 105 - \frac{20^2}{4} \right\} \left\{ 69 - \frac{16^2}{4} \right\} = 100$$

welke vereenvoudigd geven:

$5\,(2a + 16)^2 = 500$ $\quad$ of $a = -3$

$5_2(2b + 20)^2 = 980$ $\quad$ of $b = -3$

$c^2 \times 5 \times 5 = 100$ $\quad$ of $c = 2$

Voor A en B loopende van 0 tot 5 neemt y de volgende waarden aan:

	A	A_1	A_2	A_3	A_4	A_5
B_0	16	13	10	7	4	1
B_1	13	12	11	10	9	8
B_2	10	11	12	13	14	15
B_3	7	10	13	16	19	22
B_4	4	9	14	19	24	29
B_5	1	8	15	22	29	36

We zullen hieruit thans op 5 manieren een 4 x 4 proefplan lichten en daarvoor een analysis of variance uitvoeren ons beperkende tot de splitsing der sommen van quadraten:

a)

	A_0	A_1	A_2	A_3	
B_0	16	13	10	7	46
B_1	13	12	11	10	46
B_2	10	11	12	13	46
B_3	7	10	13	16	46
	46	46	46	46	184

SS totaal: $16^2 + \ldots. 16^2 - \frac{184^2}{16} =$

$2216 - 2116 = 100$

$SS(A) = 0$

$SS(B) = 0$

$SS(AB) = 100 - (0+0) = 100$

b)

	A_1	A_2	A_3	A_4	
B_1	12	11	10	9	42
B_1	11	12	13	14	50
B_2	10	13	16	19	58
B_3	9	14	19	24	66
	42	50	58	66	216

SS totaal: $12^2 + \ldots. 24^2 - \frac{216^2}{16} =$

$3176 - 2916 = 260$

$SS(A) \frac{42^2 + \ldots.66^2}{4} - cf =$

$2996 - 2916 = 80$

$SS(B)$ idem $= 80$

$SS(AB) = 260 - (80+80) = 100$

c)

	A_2	A_3	A_4	A_5	
B_2	12	13	14	15	54
B_3	13	16	19	22	70
B_4	14	19	24	29	86
B_5	15	22	29	36	102
	54	70	86	102	312

SS totaal $12^2 + \ldots. 36^2 - \frac{312^2}{16} =$

$6824 - 6084 = 740$

$SS(A) = \frac{54^2 + \ldots. 102^2}{4} - cf =$

$6404 - 6084 = 320$

$SS(B)$ idem $= 320$

$SS(AB) = 740 - (320 + 320) = 100$

d)

	A_0	A_1	A_2	A_3	
B_1	13	12	11	10	46
B_1	10	11	12	13	46
B_2	7	10	13	16	46
B_3	4	9	14	19	46
	34	42	50	58	184

SS totaal $13^2 + \ldots. 19^2 - \frac{184^2}{16} =$

$2296 - 2116 = 180$

$SS(A) = \frac{34^2 + \ldots 58^2}{4} - cf =$

$2196 - 2116 = 80$

$SS(B) = 0$

$SS(AB) = 180 - (80+0) = 100$

e)

	A_0	A_1	A_2	A_3	
B_2	10	11	12	13	46
B_3	7	10	13	16	46
B_4	4	9	14	19	46
B_5	1	8	15	22	46
	22	38	54	70	184

SS totaal $10^2 + \ldots. 22^2 - \frac{184^2}{16} =$

$2536 - 2116 = 420$

$SS(A) = \frac{22^2 + \ldots 70^2}{4} - f =$

$2436 - 2116 = 320$

$SS(B) = 0$

$SS(AB) = 420 - (320 + 0) = 100$

Het eerste wat opvalt in deze 5 proefplannen is dat de variance AB constant is. Doch dit vonden we reeds in de theorie; in § 7 vonden we: $\sum_{n}\sum_{m}(y_{ik} - \bar{y}_{i.} - \bar{y}_{.l} + \bar{y})^2 = c^2\left\{\frac{n^3}{12} - \frac{n}{12}\right\}\left\{\frac{m^3}{12} - \frac{m}{12}\right\}$

Hierin is $c = 2$ de constante uit de regressievergelijking n en m zijn het aantal trappen waarin respectievelijk factor A en B onderzocht worden. Hier is $n = m = 4$, zoodat we vinden

$2^2\left\{\frac{4^3}{12} - \frac{4}{12}\right\}\left\{\frac{4^3}{12} - \frac{4}{12}\right\} = 4 \times 5 \times 5 = 100$ voor alle 4 x 4 proefplannen, waar dit interval van de regressiefunctie ook gelegen moge zijn.

Dan werd in § 4 opgemerkt: is $\varphi(AB) \neq 0$, dan zal de interactie variance nooit = 0 zijn en de varianees voor de factoren A en B in het algemeen ook niet, hoewel hierop uitzonderingen mogelijk zijn. Geval a) is nu zo'n uitzondering. Verder wordt daar vermeld: het is daarbij onverschillig of $\varphi(A)$ en $\varphi(B)$ al dan niet = 0 zijn. In geval a) zijn ze $\neq 0$. In de eerste regressiefunctie $y = 2AB + 11.5$ zijn ze wel = 0 en daar komt dit geval voor bij de trappen $A_{-1.5}$ $A_{-0.5}$ $A_{+0.5}$ en $A_{+1.5}$ en $B_{-1.5}$ $B_{-0.5}$ $B_{+0.5}$ en $B_{+1.5}$.
De regressiefunctie $y = 2AB + 11.5$ neemt daarbij de volgende waarde aan:

	$A_{-1.5}$	$A_{-0.5}$	$A_{+0.5}$	$A_{+1.5}$	
$B_{-1.5}$	16	13	10	7	46
$B_{-0.5}$	13	12	11	10	46
$B_{+0.5}$	10	11	12	13	46
$B_{+1.5}$	7	10	13	16	46
	46	46	46	46	184

De analysis of variance is precies gelijk aan die van geval a). Ook deze uitzonderlijke mogelijkheid is dus veroorzaakt door den bijzonderen stand van het regressievlak t.o.v. het coördinatenstelsel.

Uit de theorie van § 8 hadden we deze trappen waarvan de varianees van factor A en of B = 0 wordt en dus de interactie variance maximaal is vergeleken met die variance, kunnen vinden uit formule (15) de variance voor factor A wordt minimum voor $\ell = -\frac{a}{c} + \frac{1-m}{2}$ (in casu is dit minimum = 0), a en c ontleenen we aan de regressievergelijking
$y = -3A - 3B + 2AB + 16$
$a = -3 \qquad c = 2 \qquad m = 4$
dus $\ell = -\frac{-3}{2} + \frac{1-4}{2} = 0$ dus de trappen B_0 B_1 B_2 en B_3 zie geval a).
Evenzoo is de variance voor factor B = 0 als $k = -\frac{b}{c} + \frac{1-m}{2}$ dus als $k = -\frac{-3}{2} + \frac{1-4}{2} = 0$, dus bij de trappen A_0 A_1 A_2 en A_3 we vinden dit bij de gevallen a) d) en e).

Ook in het volgende 4 x 6 experiment:

	A_0	A_1	A_2	A_3	
B_0	16	13	10	7	46
B_1	13	12	11	10	46
B_2	10	11	12	13	46
B_3	7	10	13	16	46
B_4	4	9	14	19	46
B_5	1	8	15	22	46
	51	63	75	87	276

Duidelijk blijkt ook hieruit, dat het aantal trappen en de plaats van het interval, waarin factor B onderzocht wordt in dit geval geen invloed op de variance van factor B heeft, zooals ook uit de formules blijkt.

Nog valt na te gaan, bij welke trappen de interactie variance kleiner wordt, dan de variance voor een der factoren A of B.

Volgens de theorie uit § 8 moet, wil de interactie variance kleiner zijn dan de variance voor den factor A, de eerste trap van factor B een kleinere waarde hebben dan $-\frac{a}{c} + \frac{1-m}{2} - \sqrt{\frac{m+1}{12}}$ (17) of een grootere waarde dan $-\frac{a}{c} + \frac{1-m}{2} + \sqrt{\frac{m+1}{12}}$ (zie fig. 12).

In ons geval moet de eerste trap dan b.v. hooger zijn dan $-\frac{-3}{2} + \frac{1-4}{2} + \sqrt{\frac{4+1}{12}} = \sqrt{\frac{5}{12}} \simeq 0{,}65$. Dus b.v. de trappen B_1 B_2 B_3 en B_4.
Wil de interactie variance ook kleiner zijn dan de variance voor den factor B, dan moet de 1e trap van factor A b.v. een grootere waarde hebben dan $-\frac{-b}{c} + \frac{1-n}{2} + \sqrt{\frac{m+1}{12}}$ dus ook grooter dan 0.65 dus b.v. A_1 A_2 A_3 en A_4.

Dit is geval b) en c). In geval b) is de Var A = Var B = $\frac{80}{3} \approx 27$ en de interactie variance $\frac{100}{9} \approx 11$. In de gevallen d) en e) geldt het alleen t.o.v. de variance voor factor A.

§ 10. Herhaling bij hoogere graads regressie vergelijking

Bij 2 factoren, waarbij in $\varphi(AB)$ de termen A en B beide tot den 2en graad voorkomen, wordt de regressievergelijking
$y = aA + bA^2 + cB + dB^2 + eAB + fA^2B + gAB^2 + hA^2B^2 + i.$

We vinden dan weer door substitutie van
$\varphi(A) = aA + bA^2$ $\quad$ $\varphi(B) = cB + dB^2$
$\varphi(AB) = eAB + fA^2B + gAB^2 + hA^2B^2$ in formule (II):

$$\sum_n \sum_m \left(y_{kl} - \bar{y}_{k\cdot} - \bar{y}_{\cdot l} + \bar{\bar{y}}\right)^2 = \sum_n \sum_m \left(eAB + fA^2B + gAB^2 + hA^2B^2\right)^2 -$$
$$- \frac{1}{m} \sum_n \left(\sum_m \left(eAB + fA^2B + gAB^2 + hA^2B^2\right)\right)^2 - \frac{1}{n} \sum_m \left(\sum_n \left(eAB + fA^2B + gAB^2 + hA^2B^2\right)\right)^2 +$$
$$+ \frac{1}{nm} \left(\sum_n \sum_m \left(eAB + fA^2B + gAB^2 + hA^2B^2\right)\right)^2 =$$
$$= \sum_n \sum_m \left(e^2A^2B^2 + f^2A^4B^2 + g^2A^2B^4 + h^2A^4B^4 + 2efA^3B^2 + 2egA^2B^3 + 2ehA^3B^3 + 2fgA^3B^3 + 2fhA^4B^3 + 2ghA^3B^4\right) -$$
$$- \frac{1}{m} \sum_n \left(eA\Sigma B + fA^2\Sigma B + gA\Sigma B^2 + hA^2\Sigma B^2\right)^2 -$$
$$- \frac{1}{n} \sum_m \left(eB\Sigma A + fB\Sigma A^2 + gB^2\Sigma A + hB^2\Sigma A^2\right)^2 +$$
$$+ \frac{1}{nm} \left(e\Sigma A\Sigma B + f\Sigma A^2\Sigma B + g\Sigma A\Sigma B^2 + h\Sigma A^2\Sigma B^2\right)^2 =$$
$$= e^2\Sigma A^2\Sigma B^2 + f^2\Sigma A^4\Sigma B^2 + g^2\Sigma A^2\Sigma B^4 + h^2\Sigma A^4\Sigma B^4$$
$$+ 2ef\Sigma A^3\Sigma B^2 + 2eg\Sigma A^2\Sigma B^3 + 2eh\Sigma A^3\Sigma B^3$$
$$+ 2fg\Sigma A^3\Sigma B^3 + 2fh\Sigma A^4\Sigma B^3 + 2gh\Sigma A^3\Sigma B^4$$
$$- \frac{1}{m} \sum_n \left(e^2A^2(\Sigma B)^2 + f^2A^4(\Sigma B)^2 + g^2A^2(\Sigma B^2)^2 + h^2A^4(\Sigma B^2)^2\right.$$
$$+ 2efA^3(\Sigma B)^2 + 2egA^2\Sigma B\Sigma B^2 + 2ehA^3\Sigma B\Sigma B^2 +$$
$$\left. + 2fgA^3\Sigma B\Sigma B^2 + 2fhA^4\Sigma B\Sigma B^2 + 2ghA^3(\Sigma B^2)^2\right)$$
$$- \frac{1}{n} \sum_m \left(e^2B^2(\Sigma A)^2 + f^2B^2(\Sigma A^2)^2 + g^2B^4(\Sigma A)^2 + h^2B^4(\Sigma A^2)^2\right.$$
$$+ 2efB^2\Sigma A\Sigma A^2 + 2egB^3(\Sigma A)^2 + 2ehB^3\Sigma A\Sigma A^2$$
$$\left. + 2fgB^3\Sigma A\Sigma A^2 + 2fhB^3(\Sigma A^2)^2 + 2ghB^4\Sigma A\Sigma A^2\right) +$$
$$+ \frac{1}{nm} \left(e^2(\Sigma A)^2(\Sigma B)^2 + f^2(\Sigma A^2)^2(\Sigma B)^2 + g^2(\Sigma A)^2(\Sigma B^2)^2 + h^2(\Sigma A^2)^2(\Sigma B^2)^2\right.$$
$$+ 2ef\Sigma A\Sigma A^2(\Sigma B)^2 + 2eg(\Sigma A)^2\Sigma B\Sigma B^2 + 2eh\Sigma A\Sigma A^2\Sigma B\Sigma B^2$$
$$\left. + 2fg\Sigma A\Sigma A^2\Sigma B\Sigma B^2 + 2fh(\Sigma A^2)^2\Sigma B\Sigma B^2 + 2gh\Sigma A\Sigma A^2(\Sigma B^2)^2\right) =$$

$$e^2\Sigma A^2\Sigma B^2 + f^2\Sigma A^4\Sigma B^2 + g^2\Sigma A^2\Sigma B^4 + h^2\Sigma A^4\Sigma B^4 + 2ef\Sigma A^3\Sigma B^2$$
$$+2eg\,\Sigma A^2\Sigma B^3 + 2eh\,\Sigma A^3\Sigma B^3 + 2fg\,\Sigma A^3\Sigma B^3 + 2fh\,\Sigma A^4\Sigma B^3 +$$
$$+2gh\,\Sigma A^3\Sigma B^4 - \frac{e^2}{m}\Sigma A^2(\Sigma B)^2 - \frac{f^2}{m}\Sigma A^4(\Sigma B)^2 - \frac{g^2}{m}\Sigma A^2(\Sigma B^2)^2 - \frac{h^2}{m}\Sigma A^4(\Sigma B^2)^2$$
$$-\frac{2ef}{m}\Sigma A^3(\Sigma B)^2 - \frac{2eg}{m}\Sigma A^2\Sigma B\Sigma B^2 - \frac{2eh}{m}\Sigma A^3\Sigma B\Sigma B^2$$
$$-\frac{2fg}{m}\Sigma A^3\Sigma B\Sigma B^2 - \frac{2fh}{m}\Sigma A^4\Sigma B\Sigma B^2 - \frac{2gh}{m}\Sigma A^3(\Sigma B^2)^2$$
$$-\frac{e^2}{m}(\Sigma A)^2\Sigma B^2 - \frac{f^2}{m}(\Sigma A^2)^2\Sigma B^2 - \frac{g^2}{m}(\Sigma A)^2\Sigma B^4 - \frac{h^2}{m}(\Sigma A^2)^2\Sigma B^4$$
$$-\frac{2ef}{m}\Sigma A\Sigma A^2\Sigma B^2 - \frac{2eg}{m}(\Sigma A)^2\Sigma B^3 - \frac{2eh}{m}\Sigma A\Sigma A^2\Sigma B^3$$
$$-\frac{2fg}{m}\Sigma A\Sigma A^2\Sigma B^3 - \frac{2fh}{m}(\Sigma A^2)^2\Sigma B^3 - \frac{2gh}{m}\Sigma A\Sigma A^2\Sigma B^4$$
$$+\frac{e^2}{nm}(\Sigma A)^2(\Sigma B)^2 + \frac{f^2}{nm}(\Sigma A^2)^2(\Sigma B)^2 + \frac{g^2}{nm}(\Sigma A)^2(\Sigma B^2)^2 + \frac{h^2}{nm}(\Sigma A^2)^2(\Sigma B^2)^2$$
$$+\frac{2ef}{nm}\Sigma A\Sigma A^2(\Sigma B)^2 + \frac{2eg}{nm}(\Sigma A)^2\Sigma B\Sigma B^2 + \frac{2eh}{nm}\Sigma A\Sigma A^2\Sigma B\Sigma B^2$$
$$+\frac{2fg}{nm}\Sigma A\Sigma A^2\Sigma B\Sigma B^2 + \frac{2fh}{nm}(\Sigma A^2)^2\Sigma B\Sigma B^2 + \frac{2gh}{nm}\Sigma A\Sigma A^2(\Sigma B^2)^2 =$$

Een schikken naar ΣA^4, ΣA^3 $(\Sigma A^2)^2$ en $\Sigma A\Sigma A^2$:

$$= \Sigma A^4\left\{f^2\Sigma B^2 + h^2\Sigma B^4 + 2fh\Sigma B^3 - \frac{f^2}{m}(\Sigma B)^2 - \frac{h^2}{m}(\Sigma B^2)^2 - \frac{2fh}{m}\Sigma B\Sigma B^2\right\} +$$
$$+\Sigma A^3\left\{2ef\Sigma B^2 + 2gh\Sigma B^4 + (2eh+2fg)\Sigma B^3 - \frac{2ef}{m}(\Sigma B)^2 - \frac{2gh}{m}(\Sigma B^2)^2\right.$$
$$\left.-\left(\frac{2eh}{m}+\frac{2fg}{m}\right)\Sigma B\Sigma B^2\right\} +$$
$$+(\Sigma A^2)^2\left\{-\frac{f^2}{m}\Sigma B^2 - \frac{h^2}{m}\Sigma B^4 - \frac{2fh}{m}\Sigma B^3 + \frac{f^2}{nm}(\Sigma B)^2 + \frac{h^2}{nm}(\Sigma B^2)^2 + \frac{2fh}{nm}\Sigma B\Sigma B^2\right\} +$$
$$\Sigma A\Sigma A^2\left\{-\frac{2ef}{m}\Sigma B^2 - \left(\frac{2eh}{m}+\frac{2fg}{m}\right)\Sigma B^3 - \frac{2gh}{m}\Sigma B^4\right.$$
$$\left.+\frac{2ef}{nm}(\Sigma B)^2 + \left(\frac{2eh}{nm}+\frac{2fg}{nm}\right)\Sigma B\Sigma B^2 + \frac{2gh}{nm}(\Sigma B^2)^2\right\} +$$
$$+e^2\Sigma A^2\Sigma B^2 + g^2\Sigma A^2\Sigma B^4 + 2eg\,\Sigma A^2\Sigma B^3 - \frac{e^2}{m}\Sigma A^2(\Sigma B)^2 - \frac{g^2}{m}\Sigma A^2(\Sigma B^2)^2$$
$$-\frac{2eg}{m}\Sigma A^2\Sigma B\Sigma B^2 - \frac{e^2}{m}(\Sigma A)^2\Sigma B^2 - \frac{g^2}{m}(\Sigma A)^2\Sigma B^4 - \frac{2eg}{m}(\Sigma A)^2\Sigma B^3$$
$$+\frac{e^2}{nm}(\Sigma A)^2(\Sigma B)^2 + \frac{g^2}{nm}(\Sigma A)^2(\Sigma B^2)^2 + \frac{2eg}{nm}(\Sigma A)^2\Sigma B\Sigma B^2 =$$
$$= \left\{\Sigma A^4 - \frac{(\Sigma A^2)^2}{m}\right\}\left[h^2\left\{\Sigma B^4 - \frac{(\Sigma B^2)^2}{m}\right\} + 2fh\left\{\Sigma B^3 - \frac{\Sigma B\Sigma B^2}{m}\right\} + f^2\left\{\Sigma B^2 - \frac{(\Sigma B)^2}{m}\right\}\right] +$$
$$+\left\{\Sigma A^3 - \frac{\Sigma A\Sigma A^2}{m}\right\}\left[2gh\left\{\Sigma B^4 - \frac{(\Sigma B^2)^2}{m}\right\} + 2(eh+fg)\left\{\Sigma B^3 - \frac{\Sigma B\Sigma B^2}{m}\right\} + 2ef\left\{\Sigma B^2 - \frac{(\Sigma B)^2}{m}\right\}\right] +$$
$$\left\{\Sigma A^2 - \frac{(\Sigma A)^2}{m}\right\}\left[g^2\left\{\Sigma B^4 - \frac{(\Sigma B^2)^2}{m}\right\} + 2eg\left\{\Sigma B^3 - \frac{\Sigma B\Sigma B^2}{m}\right\} + e^2\left\{\Sigma B^2 - \frac{(\Sigma B)^2}{m}\right\}\right] \quad (18)$$

Van een constant zijn van de interactie variance is hier dus geen sprake meer!
Door substitutie in formule (I) vinden we:

$$\sum_{k}^{n}\sum_{l}^{m}\left(\bar{y}_{kl}-\bar{\bar{y}}\right)^2 = m\sum^{n}\left(aA+bA^2\right)^2 - \frac{m}{mn}\left(\sum^{n}\left(aA+bA^2\right)\right)^2$$
$$+\frac{1}{m}\sum^{n}\left(\sum^{m}\left(eAB+fA^2B+gAB^2+hA^2B^2\right)\right)^2$$
$$-\frac{1}{mn}\left(\sum^{n}\sum^{m}\left(eAB+fA^2B+gAB^2+hA^2B^2\right)\right)^2$$
$$+2\sum^{n}\left(aA+bA^2\right)\sum^{m}\left(eAB+fA^2B+gAB^2+hA^2B^2\right)$$
$$-\frac{2}{mn}\sum^{n}\left(aA+bA^2\right)\sum^{n}\sum^{m}\left(eAB+fA^2B+gAB^2+hA^2B^2\right)$$

Door weer alle quadraten uit te schrijven, daarna alle sommaties uit te voeren, vervolgens gelijke termen samen te vatten en te rangschikken volgens ΣA^4, ΣA^3, $(\Sigma A^2)^2$ en $\Sigma A \Sigma A^2$ en deze buiten haken te brengen, vinden we op eender wijze:

$$\left\{\Sigma A^4-\frac{(\Sigma A^2)^2}{n}\right\}\left[m\left(b^2+2bf\Sigma B+f^2(\Sigma B)^2+g^2\left\{\frac{4}{m}\Sigma B^2+\frac{4fh}{m}\Sigma B+4h^2\right\}\right]+\right.$$
$$+\left\{\Sigma A^3-\frac{\Sigma A\Sigma A^2}{n}\right\}\left[2mab+2(af+be)\Sigma B+\frac{2ef}{m}(\Sigma B)^2+\right.$$
$$\left.\left.+\Sigma B\left\{\frac{2gf}{m}\Sigma B^2+\frac{2}{m}(eh+fg)\Sigma B+2(bh+gf)\right\}\right]\right.$$
$$+\left\{\Sigma A^2-\frac{(\Sigma A)^2}{n}\right\}\left[ma^2+2ae\Sigma B+\frac{e^2}{m}(\Sigma B)^2+\Sigma B^2\left\{\frac{g^2}{m}\Sigma B^2+\frac{2eg}{m}\Sigma B+e^2g\right\}\right] \quad (19)$$

Substitutie van de formules der rekenkundige reeksen van hoogere orde, dus de formules (5), (6), (7) en (8) in de $\{\ \}$ vormen geeft:

$$\left\{\Sigma A^4-\frac{(\Sigma A^2)^2}{n}\right\}=\frac{n(n^2-1)}{3}\left\{k^2+(n-1)k+\frac{6n^2-3n+11}{60}\right\}=$$
$$=\frac{n(n^2-1)}{3}\left\{\left(k+\frac{n-1}{2}\right)^2+\frac{n^2-4}{60}\right\} \quad (20)$$
$$\left\{\Sigma A^3-\frac{\Sigma A\Sigma A^2}{n}\right\}=\frac{n(n^2-1)}{3}\cdot\frac{1}{2}\cdot\left(k+\frac{n-1}{2}\right) \quad (21)$$
$$\left\{\Sigma A^2-\frac{(\Sigma A)^2}{n}\right\}=\frac{n(n^2-1)}{3}\cdot\frac{1}{4} \quad (22)$$

Voor de sommatie vormen met B vinden we gelijksoortige vormen, maar daar is k door l en n door m vervangen.

Voeren we nu vervolgens als nieuwe veranderlijken in $k+\frac{n-1}{2}=k'$ en $l+\frac{m-1}{2}=l'$ bij oneven aantal trappen zijn k' en l' eigenlijk de middelste trappen, dan vinden we voor de interactievariance:

$$\frac{1}{(n-1)(m-1)} \cdot \frac{n(n^2-1)}{3} \cdot \frac{m(m^2-1)}{3} \times$$

$$\times \Big[\Big(k' + \frac{n^2-4}{60}\Big) \Big\{ h^2 \Big(l'^2 + \frac{m^2-4}{60}\Big) + fh\,l' + \frac{f^2}{4} \Big\} +$$

$$+ \frac{k'}{2} \Big\{ 2gh \Big(l'^2 + \frac{m^2-1}{60}\Big) + (eh+fg)\,l' + \frac{ef}{4} \Big\} +$$

$$+ \frac{1}{4} \Big\{ g^2 \Big(l'^2 + \frac{m^2-4}{60}\Big) + eg\,l' + \frac{e^2}{4} \Big\} \Big]$$

Gaan we thans het gedrag van deze interactie variance na, dan is het duidelijk dat, als er geen bijzondere betrekkingen tusschen de coefficienten a t/m h heerschen, we zeer ingewikkelde en lange vormen krijgen. Bovendien geldt hier eveneens dat het schrappen van coefficienten in de verkregen formules, in overeenstemming met een gereduceerde regressievergelijking, tot onjuiste conclusies kan voeren, evenals in ons eerste geval waar schrappen van de coefficient c een onjuiste voorstelling geeft (zie pg. 56).

We beperken ons daarom tot het eenvoudigste geval $y = h\,A^2B^2 + i$. We vinden dan:

$$V(AB) = \frac{h^2}{(n-1)(m-1)} \Big\{ \Sigma A^4 - \frac{(\Sigma A^2)^2}{n} \Big\} \Big\{ \Sigma B^4 - \frac{(\Sigma B^2)^2}{m} \Big\}$$

$$V(A) = \frac{h^2}{m(n-1)} \Big\{ \Sigma A^4 - \frac{(\Sigma A^2)^2}{n} \Big\} (\Sigma B^2)^2$$

Substitueeren we hierin thans de formules (20), (21) en (22) en gaan we het gedrag van de varianoes na:

Stellen we weer $k + \frac{n-1}{2} = k'$ en $l + \frac{m-1}{2} = l'$ dan vinden we

$$V(AB) = \frac{n(n+1)m(m+1)}{9} \Big\{ k'^2 + \frac{n^2-4}{60} \Big\} \Big\{ l'^2 + \frac{m^2-4}{60} \Big\}$$

$$\frac{\partial V(AB)}{\partial k'} = C \cdot 2k' \Big\{ l'^2 + \frac{m^2-4}{60} \Big\}$$

$$\frac{\partial^2 V(AB)}{\partial k'^2} = C \cdot 2 \Big\{ l'^2 + \frac{m^2-4}{60} \Big\}$$

$$\frac{\partial V(AB)}{\partial l'} = C \Big\{ k'^2 + \frac{n^2-4}{60} \Big\} 2l'$$

$$\frac{\partial^2 V(AB)}{\partial l'^2} = C \Big\{ k'^2 + \frac{n^2-4}{60} \Big\} 2$$

$$\frac{\partial^2 V(AB)}{\partial k' \partial l'} = C \cdot 2k' \cdot 2l'$$

$V(AB)$ wordt extreem voor $\frac{\partial V(AB)}{\partial k'} = 0$ als $C \cdot 2k' \Big\{ l'^2 + \frac{m^2-4}{60} \Big\} = 0$

a) $k' = 0$ of $k = \frac{1-n}{2}$

b) $l'^2 + \frac{m^2-4}{60} = 0$ of $l' = \pm\sqrt{\frac{4-m^2}{60}}$

Was in het 1e geval ($y=aA+bB+cAB+d$) n en m ≥ 2, hierbij 2e graadsvormen hebben we n en m ≥ 3. Geval b) vervalt dus en voor $\frac{\partial V(AB)}{\partial l'} = 0$ wat op dezelfde wijze voert tot $l = \frac{1-m}{2}$. Dus V(AB) extreem voor $k = \frac{1-n}{2}$ en $l = \frac{1-m}{2}$

voor deze waarden wordt

$$\frac{\partial^2 V(AB)}{\partial k'^2} = l \cdot 2 \cdot \frac{m^2-4}{60} \quad \text{dus} > 0 \quad (m \geq 3)$$

$$\frac{\partial^2 V(AB)}{\partial l'^2} = l \cdot \frac{n^2-4}{60} \cdot 2 \quad \text{dus} > 0$$

$$\frac{\partial^2 V(AB)}{\partial k' \partial l'} = 0$$

$$\left(\frac{\partial^2 V(AB)}{\partial k' \partial l'}\right)^2 - \frac{\partial^2 V(AB)}{\partial k'^2} \cdot \frac{\partial^2 V(AB)}{\partial l'^2} = < 0$$

Extremum is dus een minimum. De waarde van de variance in het minimum is:

$$V(AB) = \frac{n(n+1)\,m(m+1)}{9} h^2 \cdot \frac{n^2-4}{60} \cdot \frac{m^2-4}{60}$$

Voor V(A) minder we $\frac{h^2}{m(m-1)} \cdot \frac{n(n^2-1)}{3}\left\{l'^2 + \frac{n^2-4}{60}\right\} m\left\{l'^2 + \frac{m^2-1}{12}\right\}$,

$$= \frac{n(n+1)h^2}{3}\left\{l'^2 + \frac{n^2-4}{60}\right\}\left\{l'^2 + \frac{m^2-1}{4}\right\}$$

Op dezelfde wijze vinden we weer V(A) extreem voor $k = \frac{1-n}{2}$ en $l = \frac{1-m}{2}$. Dit extremum is eveneens een minimum en heeft de waarde

$$V(A) = \frac{n\,(n+1)}{3} \cdot h^2 \cdot \frac{n^2-4}{60} \cdot \frac{m^2-1}{12}$$

Beschouwen we thans de verhouding

$$H = \frac{V(A) - V(AB)}{V(AB)} = \frac{V(A)}{V(AB)} - 1$$

$$H = \frac{m-1}{m} \cdot \frac{(\Sigma B^2)^2}{\left\{\Sigma B^4 - \frac{(\Sigma B^2)^2}{m}\right\}} - 1 = \frac{m-1}{m} \cdot \frac{m\left\{l'^2 + \frac{m^2-1}{12}\right\}}{\frac{n(n^2-1)}{3}\left\{l'^2 + \frac{n^2-4}{60}\right\}} - 1 =$$

$$= \frac{-3}{m\,(m+1)} - 1 + \frac{\frac{4\,m^2-1}{20\,m\,(m+1)}}{l'^2 + \frac{m^2-4}{60}}$$

Dus H maximaal voor $l' = 0$ of $l = \frac{1-m}{2}$

De waarde van H in dit maximum is

$$\frac{-3}{m(m+1)} - 1 + \frac{\frac{4\,m^2-1}{10\,m(m+1)}}{\frac{m^2-4}{60}} = -1 + \frac{15\,m^2-15}{m^2+m^2-4\,m^2-4\,m}$$

Voor m = 3 wordt dit $-1+2 = +1$
Voor m = ∞ nadert dit tot -1
Voor $l = \infty$ nadert H tot $\frac{-3}{m(m+1)} - 1$
in dat geval is voor m = 3 H = $-\frac{3}{4}$
m = ∞ H = -1
We krijgen dus het volgende verloop voor H als m = 3

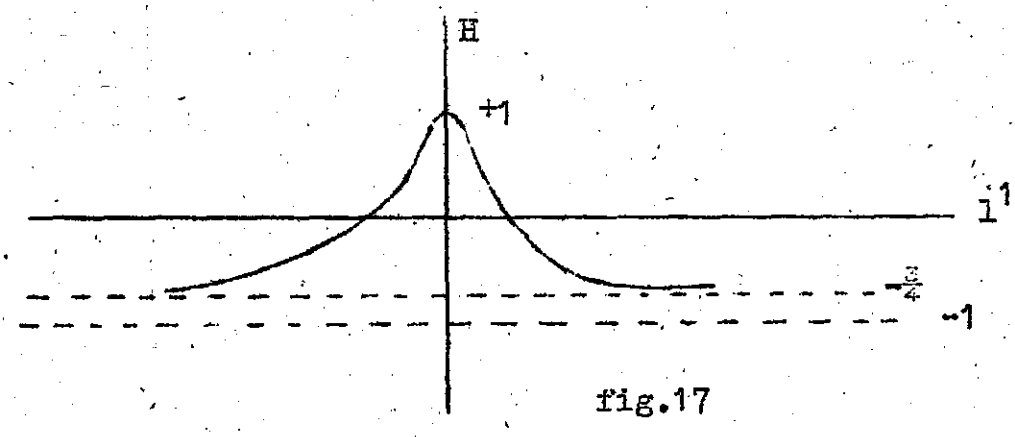


fig.17

Voor hooge waarden van m het volgende verloop:

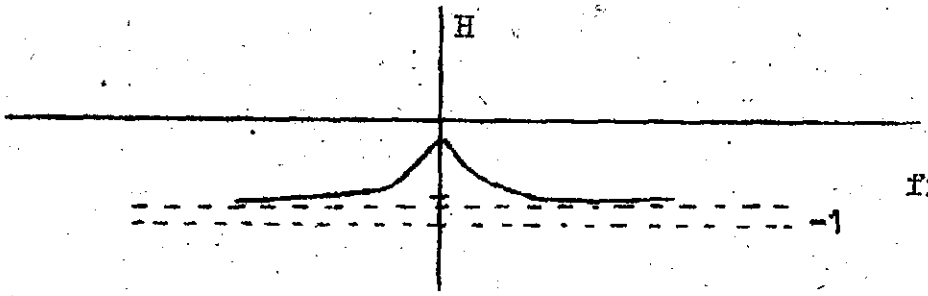


fig. 18

Hierbij is dus $H<0$ en de asymptoot tusschen $-\frac{3}{4}$ en -1. $V(A)=V(AB)$ als $H=0$. De waarde van ℓ waarvoor dit plaats heeft vinden we uit:

$$\frac{n(n+1)}{9}\frac{m(m+1)}{9}h^2\left\{k^{1^2}+\frac{n^2-4}{60}\right\}\left\{\ell^{1^2}+\frac{m^2-4}{60}\right\}=$$

$$=\frac{n(n+1)}{3}h^2\left\{k^{1^2}+\frac{n^2-4}{60}\right\}\left\{\ell^{1^2}+\frac{m^2-1}{12}\right\} \quad \text{of}$$

$$\frac{m(m+1)}{3}\left\{\ell^{1^2}+\frac{m^2-4}{60}\right\}=\left\{\ell^{1^2}+\frac{m^2-1}{12}\right\}$$

$$\therefore\ \ell^{1^2}=\frac{-m^4-m^3+19m^2+4m-15}{60(m^2+m-3)}=\frac{-(m+1)(m^3-19m+15)}{60(m^2+m-3)}$$

($m^3-19m+19=0$ voor $m=3.892$).

Voor hooge waarden van m wordt ℓ^1 imaginair hetgeen zonder meer duidelijk is uit fig. 18. Voor $m=3$ vinden we

$$\ell^{1^2}=\frac{-81-27+171+12-15}{60(9+3-3)}=\frac{60}{540}=\frac{1}{9}$$

$$\ell^1=\pm\frac{1}{3}\quad\left(\ell=\frac{1-m}{2}\pm\frac{1}{3}=-\frac{4}{3}\text{ of }-\frac{2}{3}\right)$$

Voor de volledigheid nog het volgende:
Voor m is klein b.v. $m=3$ is het verloop der varianbes dus als volgt voor te stellen:

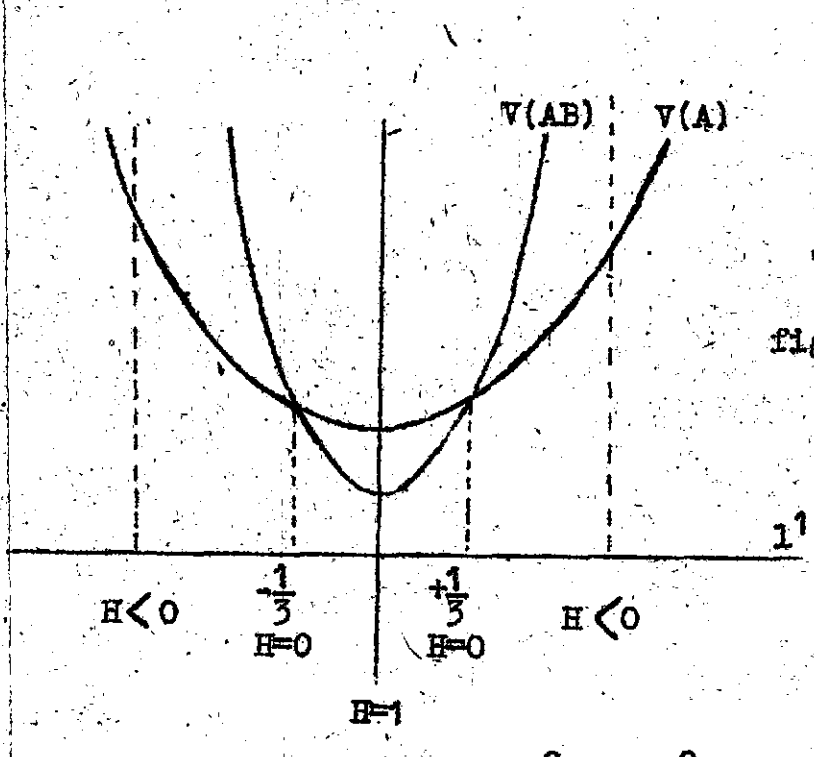


fig.19

Minimumwaarde $V(AB) = \frac{n(n+1)m(m+1)}{9} h^2 . \frac{m^2-4}{60} . \frac{m^2-4}{60}$

$= C \frac{m(m+1)}{3} . \frac{m^2-4}{60} = \frac{C}{3}$

de minimumwaarde van V(A) is $\frac{n(n+1)}{3} h^2 . \frac{n^2-4}{60} . \frac{m^2-1}{12} =$

$= C . \frac{m^2-1}{12} = \frac{2C}{3}$

Voor m = groot vinden we daarentegen:

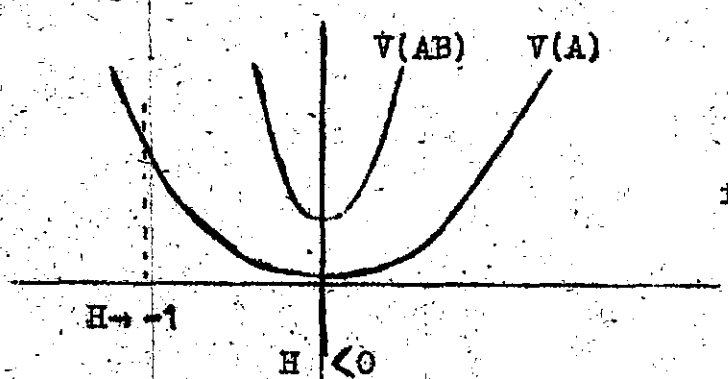


fig: 20

H altijd <0, dit is reeds het geval als

$\frac{15m^2-15}{m^4+m^3-4m^2-4m} < +1$, dat is voor $m \geq 4$

We kunnen dus beter zeggen, dat het geval bij m = 3 dus feitelijk een uitzondering is!

In het algemeen zal dus bij $y = hA^2B^2 + i$ de interactie variance steeds grooter zijn dan de variance voor den factor A (of factor B). Alleen bij m = 3 is er een klein interval waarin de interactie variance kleiner is. Dit is dus totaal tegengesteld aan ons eerste geval!

Illustreeren we dit nog aan een voorbeeld:

§ 11. Voorbeeld 2.

Twee factoren en kromlijnige regressie. We nemen een eenvoudig geval: $y = \varphi(A) + \varphi(AB)$ i.c. $y = -2A^2 - AB^2 + 50$.
Graphisch voorgesteld aldus:

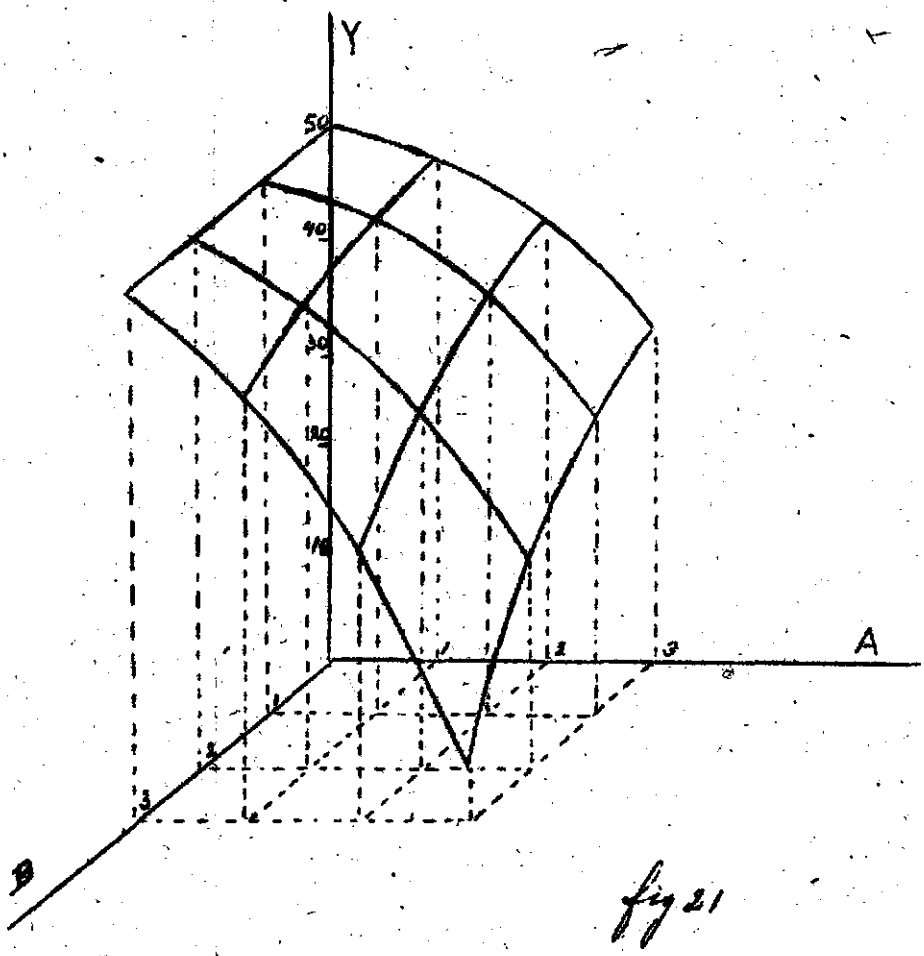


fig 21

Zouden we nu de factoren A en B in een experiment onderzoeken, elke factor in b.v. 4 trappen: A_0, A_1, A_2, A_3 en B_0, B_1, B_2 en B_3 dan zouden we de volgende waarden voor y vinden.

	A_0	A_1	A_2	A_3	
B_0	50	48	42	32	172
B_1	50	47	40	29	166
B_2	50	44	34	20	148
B_3	50	39	24	5	118
	200	178	140	86	604

Analysis of variance geeft het volgende resultaat:

SS totaal $50^2 + \ldots . 5^2 - \frac{604^2}{16} = 2.5536 - 2.2801 = 2555$

SS(A) $\frac{200^2 + \ldots . 86^2}{4} - cf = 2.4670 - 2.2801 = 1869$

SS(B) $\frac{172^2 + \ldots . 118^2}{4} - cf = 2.3242 - 2.2801 = 441$

SS(AB) $2555 - (1869 + 441) = 245$

Leiden we uit de regressiefunctie de analysis of variance af: In de regressievergelijking is $\varphi(A) \neq 0$ $\varphi(B) = 0$ $\varphi(AB) \neq 0$.

Volgens het 1e overzicht volgende uit de formules (I) en (II) moet dan een interactievariance gevonden worden en in het algemeen ook een variance voor de factoren A en B, wat in dit voorbeeld dus het geval is.

Omgekeerd kunnen we uit de analysis of variance het type van de regressiefunctie afleiden. Daar we een interactievariance vinden, moet volgens het 2e overzicht volgende uit de formules (I) en (II) in de regressiefunctie de term $\varphi(AB)$ voorkomen en in het algemeen ook $\varphi(A)$ en $\varphi(B)$ hoewel uitzonderingen mogelijk zijn. Hier hebben we dus in tegenstelling met het eerste voorbeeld een gedeeltelijke uitzondering, immers $\varphi(B) = 0$. Passen we weer evenals in het voorbeeld 1 een translatie toe op de gegeven regressievergelijking en gaan we over op b.v. de coördinaten $A' = A+2$, $B' = B+2$. We vinden dan $y = -2(A'-2)^2 - (A'-2)(B'-2)^2 + 50$. In het volgende laten we de accenten gemakshalve weg. We vinden dan:
$y = -2A^2+4A+2B^2-8B-AB^2+4AB+50$.
Graphisch voorgesteld aldus:

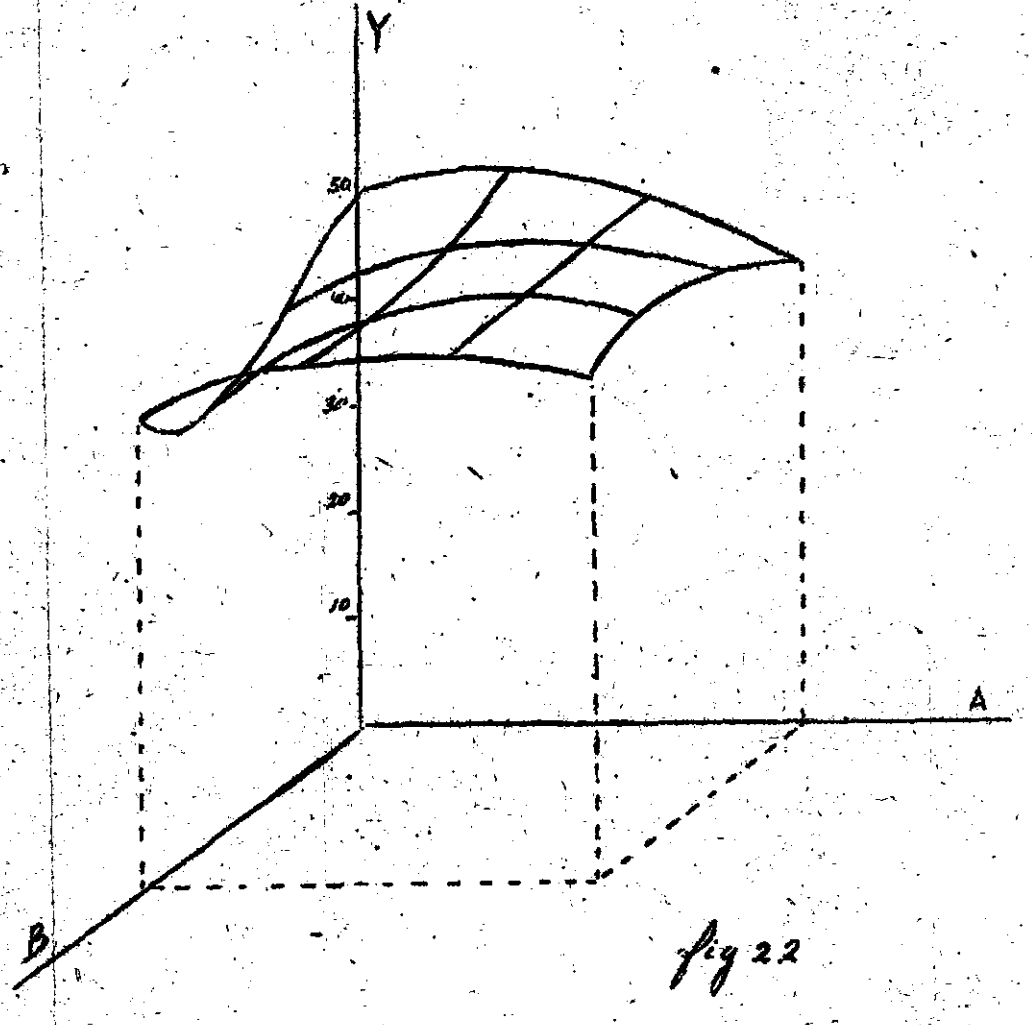


fig 22

Zouden we nu weer de factoren A en B in een proefplan onderzoeken, elke factor in 4 trappen n.l. A_2, A_3, A_4, A_5 en B_2, B_3, B_4 en B_5 dan zouden we weer dezelfde waarden vinden als in het 1e proefplan bij $y = -2A^2 - AB^2 + 50$. De analysis of variance had precies hetzelfde resultaat gegeven en de conclusie t.a.v. de regressievergelijking luiden hetzelfde, alleen hebben we thans niet meer met een uitzondering te doen. Ook hier dus een uitzondering tengevolge van een bijzondere ligging van het regressievlak t.o.v. het coördinatenstelsel.

We bemoeien ons verder alleen met het algemeenere geval $y = -2A^2 + 4A + 2B^2 - 8B - AB^2 + 4AB + 50$. We hadden de uitkomsten van de analysis of variance ook dadelijk uit deze regressievergelijking kunnen bepalen en wel met behulp van de formules (I) en (II) uit § 4. Voor het geval van 2 factoren bij kromlijnige regressie waarbij A en B tot in den 2en graad in $\varphi(AB)$ voorkomen, is dit reeds uitgewerkt in § 10 in de formules (18) en (19).

In ons voorbeeld is a = +4 b = -2 c = -8 d = +2 e = +4 f = 0 g = -1 h = 0

Gesubstitueerd geeft dit:

$$\sum_n^m \sum (\bar{y}_{k.} - \bar{\bar{y}})^2 = \left\{\Sigma A^4 - \frac{(\Sigma A^2)^2}{n}\right\}\left\{mx - 2^2\right\} +$$

$$+ \left\{\Sigma A^3 - \frac{\Sigma A \Sigma A^2}{n}\right\}\left\{2.m.4.-2 + 2.-2.4\,\Sigma B + \Sigma B^2(2.-2.-1)\right\} +$$

$$+ \left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{m4^2 + 2.4.4.\Sigma B + \frac{4^2}{m}(\Sigma B)^2 + \right.$$

$$\left. + \Sigma B^2\left(\frac{-1^2}{m}\Sigma B^2 + \frac{2.4.-1}{m}\Sigma B + 2.4.-1\right)\right\} =$$

$$= \left\{\Sigma A^4 - \frac{(\Sigma A^2)^2}{n}\right\} 4m +$$

$$\left\{\Sigma A^3 - \frac{\Sigma A \Sigma A^2}{n}\right\}\left\{-16m - 16\Sigma B + 4\Sigma B^2\right\}$$

$$+ \left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{16m + 32\Sigma B + \frac{16}{m}(\Sigma B)^2 + \Sigma B^2\left(\frac{\Sigma B^2}{m} - \frac{8}{m}\Sigma B - 8\right)\right\}$$

$$\sum_n^m \sum \left\{y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{\bar{y}}\right\}^2 =$$

$$\left\{\Sigma A^2 - \frac{(\Sigma A)^2}{n}\right\}\left\{-1^2\left(\Sigma B^4 - \frac{(\Sigma B^2)^2}{m}\right) + 2.4.-1\left(\Sigma B^3 - \frac{\Sigma B \Sigma B^2}{m}\right) + 4^2\left(\Sigma B^2 - \frac{(\Sigma B)^2}{m}\right)\right\}$$

In ons laatste proefplan hadden we de trappen A_2 t/m A_5 en B_2 t/m B_5, zoodat dus m = n = 4

$\Sigma A_2 = \Sigma B_2 = 2+3+4+5 = 14$

$\Sigma A^2 = \Sigma B^2 = 54$ $\Sigma A^3 = \Sigma B^3 = 224$ en $\Sigma A^4 = \Sigma B^4 = 978$

zoodat we vinden voor $\sum_4^4 \sum (\bar{y}_{k.} - \bar{\bar{y}})^2 =$

$$\left\{978 - \frac{54^2}{4_2}\right\}.4.4 + \left\{224 - \frac{14.54}{4}\right\}.\left\{-16.4 - 16.4 + 4.54\right\} +$$

$$\left\{54 - \frac{14^2}{4}\right\}\left\{16.4 + 32.14 + \frac{16}{4}:14^2 + 54\left(\frac{54}{4} - \frac{8}{4}.14 - 8\right)\right\} =$$

$$= 249 \times 16 + 35 \times -72 + 5 \times 81 = 1869$$

en voor $\Sigma\Sigma(y_{kl} - \bar{y}_{k.} - \bar{y}_{.l} + \bar{\bar{y}})^2$

$$\left\{54 - \frac{14^2}{4}\right\}\left\{\left(978 - \frac{54^2}{4}\right) - 8\left(224 - \frac{14.54}{4}\right) + 16\left(54 - \frac{14^2}{4}\right)\right\} =$$

$$5.249 - 8.35 + 16.5 = 245$$

gelijk wij in de analysis of variance ook vonden.

De totale variance en de variance voor factor B kan desgewenscht evenzoo berekend worden. Thans kan echter niet omgekeerd uit de analysis of variance de regressiefunctie bepaald worden. De analysis of variance geeft ons slechts 3 onafhankelijke waarden, terwijl hier in de regressiefunctie reeds 9 constanten voorkomen!

We kunnen ons thans afvragen wat het gedrag van de interactie variance zal zijn.

Substitueeren we in (21) de waarden van de coëfficienten uit de regressievergelijking dan vinden we:

$$V(AB) = \frac{1}{(n-1)(m-1)} \left\{ \Sigma A^2 - \frac{(\Sigma A)^2}{n} \right\} \times$$

$$\times \left[\left\{ \Sigma B^4 - \frac{(\Sigma B^2)^2}{m} \right\} - 8 \left\{ \Sigma B^3 - \frac{\Sigma B \Sigma B^2}{m} \right\} + 16 \left\{ \Sigma B^2 - \frac{(\Sigma B)^2}{m} \right\} \right] =$$

$$= \frac{1}{(n-1)(m-1)} \frac{n(n^2-1)}{12} \left[\frac{m(m^2-1)}{3} \left\{ (\ell + \frac{m-1}{2})^2 + \frac{m^2-4}{60} \right. \right.$$

$$\left. \left. -8.\tfrac{1}{2}(\ell + \frac{m-1}{2}) + 16.\tfrac{1}{4} \right\} \right] =$$

$$\frac{n(n+1)m(m+1)}{36} \left\{ \ell^{1^2} - 4\ell^1 + \frac{m^2+236}{60} \right\}$$

De waarde der variance verloopt parabolisch en wordt extreem voor $\frac{V(AB)}{\partial \ell} = 0 = \frac{n(n+1)m(m+1)}{36} (2\ell^1 - 4)$

of $\ell^1 = 2$ d.w.z. $\ell = \frac{5-m}{2}$

$$\frac{\partial^2 V(AB)}{\partial \ell^{1^2}} = \frac{n(n+1)m(m+1)}{36} .2 > 0$$

Bij b.v. 4 trappen wordt de variance dus minimaal bij de trappen $\ell = \frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, 3\frac{1}{2}$.

De waarde van de interactie variance in dat minimum is:

$$\frac{n(n+1)m(m+1)}{36} \left\{ 2^2 - 4.2 + \frac{m^2+236}{60} \right\}$$

wat voor n = m = 4 overgaat in:

$$\frac{400}{36} \left\{ 4 - 8 + \frac{252}{60} \right\} = \frac{400}{36} \cdot \frac{12}{60} = 2\frac{2}{9}.$$

Een analysis of variance doet hetzelfde vinden.

Nemen we factor A in de trappen 1, 2, 3 en 4, we vinden dan de volgende y-waarden:

	$B_{\frac{1}{2}}$	$B_{1\frac{1}{2}}$	$B_{2\frac{1}{2}}$	$B_{3\frac{1}{2}}$	
A_1	50.25	48.25	48.25	50.25	197
A_2	50	50	50	50	200
A_3	45.75	47.75	47.75	45.75	187
A_4	37.50	41.50	41.50	37.50	158
	183.50	187.50	187.50	183.50	748

SS totaal $50.25^2 + \ldots\ldots 37.50^2 - \frac{742^2}{16} = 299.25$

SS(A) $\frac{197^2 + \ldots\ldots 158^2}{4} - cf = 275.25$

SS(B) $\frac{183.50^2 + \ldots\ldots 183.50^2}{4} - cf = 4$

SS(AB) $299.25 - (275.25 + 4) = 20$

variatiecoorzaak	sommen van quadraten	vrijheids-graden	variance
A	275.25	3	91.75
B	4	3	1
AB	20	9	$2\frac{2}{9}$
	299.25	15	

We zouden nu ook nog het gedrag van de variance van factor A kunnen bepalen en de verhouding van deze variance met de interactie variance. Eventueel ook voor uitgebreidere vormen, doch we zullen het hierbij laten. Met deze beide voorbeelden zij voorgaande theorie trouwens voldoende toegelicht, geillustreerd.

§ 12. De betrekkingen bij 3 factoren.

Hierbij hebben we te doen met

$$y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB) + \varphi(AC) + \varphi(BC) + \varphi(ABC).$$

We kunnen nu weer evenals in § 4 afleiden welke variances hierbij behooren. De totale variance kan thans bij 3 factoren gesplitst worden in 7 variances, welke splitsing gebaseerd is op de volgende identiteit:

$$\sum_{k}^{n}\sum_{l}^{m}\sum_{i}^{j}(y_{kli}-\bar{\bar{\bar{y}}})^2 = \sum\sum\sum(\bar{y}_{k..}-\bar{\bar{\bar{y}}})^2 + \sum\sum\sum(\bar{y}_{.l.}-\bar{\bar{\bar{y}}})^2 + \sum\sum\sum(\bar{y}_{..i}-\bar{\bar{\bar{y}}})^2$$

$$+\sum\sum\sum(\bar{y}_{.li}-\bar{y}_{.l.}-\bar{y}_{..i}+\bar{\bar{\bar{y}}})^2 + \sum\sum\sum(\bar{y}_{k.i}-\bar{y}_{k..}-\bar{y}_{..i}+\bar{\bar{\bar{y}}})^2 +$$

$$+\sum\sum\sum(\bar{y}_{kl.}-\bar{y}_{k..}-\bar{y}_{.l.}+\bar{\bar{\bar{y}}})^2 +$$

$$+\sum\sum\sum(y_{kli}-\bar{y}_{.li}-\bar{y}_{k.i}-\bar{y}_{kl.}+\bar{y}_{i..}+\bar{y}_{.l.}+\bar{y}_{..i}-\bar{\bar{\bar{y}}})^2$$

Door deze 7 sommen van quadraten respectievelijk door (n-1), (m-1), (j-1), (m-1)(j-1), (n-1)(j-1), (n-1)(m-1) en (n-1)(m-1)(j-1) te deelen, vinden wij de variances voor de factoren A, B en C, de variances voor de interacties van de 1e orde BC, AC en AB en nog de variance van de interactie van de 2e orde ABC.

Hierin is dus

$$y_{kli} = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB) + \varphi(AC) + \varphi(BC) + \varphi(ABC)$$

$$\bar{y}_{.li} = \frac{\sum_{k}^{n}\{\varphi(A)+\cdots\varphi(ABC)\}}{n} \qquad \bar{y}_{k.i} = \frac{\sum_{l}^{m}\{\varphi(A)+\cdots\cdots\varphi(ABC)\}}{m}$$

$$\bar{y}_{kl.} = \frac{\sum_{i}^{j}\{\varphi(A)+\cdots\varphi(ABC)\}}{j} \qquad \bar{y}_{k..} = \frac{\sum_{l}^{m}\sum_{i}^{j}\{\varphi(A)+\cdots\cdots\varphi(ABC)\}}{mj}$$

$$\bar{y}_{.l.} = \frac{\sum_{k}^{n}\sum_{i}^{j}\{\varphi(A)+\cdots\cdots\varphi(ABC)\}}{nj} \qquad \bar{y}_{..i} = \frac{\sum_{k}^{n}\sum_{l}^{m}\{\varphi(A)+\cdots\varphi(ABC)\}}{nm}$$

$$\bar{\bar{\bar{y}}} = \frac{\sum_{k}^{n}\sum_{l}^{m}\sum_{i}^{j}\{\varphi(A)+\cdots\varphi(ABC)\}}{nmj}$$

Nu is dus $\sum_{k}^{n}\sum_{l}^{m}\sum_{i}^{j}(\bar{y}_{kl.}-\bar{\bar{\bar{y}}})^2 = \sum_{k}^{n}\sum_{l}^{m}\sum_{i}^{j}\left\{\frac{\sum_{i}^{j}\{\varphi(A)+\cdots\varphi(ABC)\}}{j} - \frac{\sum_{k}^{n}\sum_{i}^{j}\{\varphi(A)+\cdots\varphi(ABC)\}}{nj}\right\}^2 =$

$$\sum_{k}^{n}\sum_{l}^{m}\sum_{i}^{j}\left\{\frac{\sum_{i}^{j}\varphi(A)}{j} + \frac{\sum_{i}^{j}\varphi(B)}{j} + \frac{\sum_{i}^{j}\varphi(C)}{j} + \frac{\sum_{i}^{j}\varphi(AB)}{j} + \frac{\sum_{i}^{j}\varphi(AC)}{j} + \right.$$

$$\left. + \frac{\sum_{i}^{j}\varphi(BC)}{j} + \frac{\sum_{i}^{j}\varphi(ABC)}{j} - \frac{\sum_{k}^{n}\sum_{i}^{j}\varphi(A)}{nj} - \frac{\sum_{k}^{n}\sum_{i}^{j}\varphi(B)}{nj} - \right.$$

$$-\frac{\sum\sum\sum_{n\,m\,j}\psi(C)}{n\,m\,j}-\frac{\sum\sum\sum_{n\,m\,j}\psi(AB)}{n\,m\,j}-\frac{\sum\sum\sum_{n\,m\,j}\psi(AC)}{n\,m\,j}-\frac{\sum\sum\sum_{n\,m\,j}\psi(BC)}{n\,m\,j}$$

$$-\frac{\sum\sum\sum_{n\,m\,j}\psi(ABC)}{n\,m\,j}\Big\}^2=\sum\sum\sum\Big\{\psi(A)+\frac{\sum\psi(B)}{m}+\frac{\sum\psi(C)}{j}+\frac{\sum\psi(AB)}{m}$$

$$+\frac{\sum\psi(AC)}{j}+\frac{\sum\sum\psi(BC)}{m\,j}+\frac{\sum\sum\psi(ABC)}{m\,j}-\frac{\sum\psi(A)}{n}-\frac{\sum\psi(B)}{m}$$

$$-\frac{\sum\psi(C)}{j}-\frac{\sum\sum\psi(AB)}{n\,m}-\frac{\sum\sum\psi(AC)}{n\,j}-\frac{\sum\sum\psi(BC)}{m\,j}-\frac{\sum\sum\sum\psi(ABC)}{n\,m\,j}\Big\}^2=$$

$$\sum\sum\sum\Big\{\psi(A)-\frac{\sum\psi(A)}{n}+\frac{\sum\psi(AB)}{m}-\frac{\sum\sum\psi(AB)}{n\,m}+\frac{\sum\psi(AC)}{j}$$

$$-\frac{\sum\sum\psi(AC)}{n\,j}+\frac{\sum\sum\psi(ABC)}{m\,j}-\frac{\sum\sum\sum\psi(ABC)}{n\,m\,j}\Big\}^2$$

Zooals we zien komen $\psi(B)$, $\psi(C)$ en $\psi(BC)$ hierin niet meer voor, zoodat deze termen uit de regressiefunctie geen invloed op de variance voor factor A hebben.
Verder uitwerken geeft:

$$\sum\sum\sum\Big\{\psi^2(A)+\Big(\frac{\sum\psi(A)}{n}\Big)^2+\Big(\frac{\sum\psi(AB)}{m}\Big)^2+\Big(\frac{\sum\sum\psi(AB)}{n\,m}\Big)^2+\Big(\frac{\sum\psi(AC)}{j}\Big)^2+$$

$$\Big(\frac{\sum\sum\psi(AC)}{m\,j}\Big)^2+\Big(\frac{\sum\sum\psi(ABC)}{m\,j}\Big)^2+\Big(\frac{\sum\sum\sum\psi(ABC)}{n\,m\,j}\Big)^2-\frac{2\psi(A)\sum\psi(A)}{n}+$$

$$\frac{2\psi(A)\sum\psi(AB)}{m}-\frac{2\psi(A)\sum\sum\psi(AB)}{n\,m}+\frac{2\psi(A)\sum\psi(AC)}{j}$$

$$-\frac{2\psi(A)\sum\sum\psi(AC)}{n\,j}+\frac{2\psi(A)\sum\sum\psi(ABC)}{m\,j}-\frac{2\psi(A)\sum\sum\sum\psi(ABC)}{n\,m\,j}-$$

$$-\frac{2\sum\psi(A)\sum\psi(AB)}{n\,m}+\frac{2\sum\psi(A)\sum\sum\psi(AB)}{n\,m^2}-\frac{2\sum\psi(A)\sum\psi(AC)}{n\,j}$$

$$+\frac{2\sum\psi(A)\sum\sum\psi(AC)}{n^2\,j}-\frac{2\sum\psi(A)\sum\sum\psi(ABC)}{n\,m\,j}+\frac{2\sum\psi(A)\sum\sum\sum\psi(ABC)}{n^2\,m\,j}$$

$$-\frac{2\sum\psi(AB)\sum\sum\psi(AB)}{n\,m^2}+\frac{2\sum\psi(AB)\sum\psi(AC)}{m\,j}+\frac{2\sum\sum\psi(AB)\sum\psi(AC)}{n\,m\,j}$$

$$+\frac{2\sum\psi(AB)\sum\sum\psi(ABC)}{m^2\,j}-\frac{2\sum\psi(AB)\sum\sum\sum\psi(ABC)}{n\,m^2\,j}-\frac{2\sum\sum\psi(AB)\sum\psi(AC)}{n\,m\,j}$$

$$+\frac{2\sum\sum\psi(AB)\sum\sum\psi(AC)}{n^2\,m\,j}-\frac{2\sum\sum\psi(AB)\sum\sum\psi(ABC)}{n\,m^2\,j}+\frac{2\sum\sum\psi(AB)\sum\sum\sum\psi(ABC)}{n^2\,m^2\,j}$$

$$-\frac{2\sum\psi(AC)\sum\sum\psi(AC)}{n\,j^2}+\frac{2\sum\psi(AC)\sum\sum\psi(ABC)}{m\,j^2}-\frac{2\sum\psi(AC)\sum\sum\sum\psi(ABC)}{n\,m\,j^2}$$

$$-\frac{2\sum\sum\psi(AC)\sum\sum\psi(ABC)}{n\,m\,j^2}+\frac{2\sum\sum\psi(AC)\sum\sum\sum\psi(ABC)}{n^2\,m\,j^2}-\frac{2\sum\sum\psi(ABC)\sum\sum\sum\psi(ABC)}{n\,m^2\,j^2}\Big\}=$$

$$
\begin{aligned}
&mj\sum_{}^{n}(\psi(A))^2+\frac{mj}{n}\Big(\sum^{n}\psi(A)\Big)^2+\frac{j}{m}\sum^{n}\Big(\sum^{m}\psi(AB)\Big)^2+\frac{1}{nm}\Big(\sum^{n}\sum^{m}\psi(AB)\Big)^2\\
&+\frac{m}{l}\sum^{n}\Big(\sum^{l}\psi(AC)\Big)^2+\frac{m}{nl}\Big(\sum^{n}\sum^{l}\psi(AC)\Big)^2+\frac{1}{ml}\sum^{n}\Big(\sum^{m}\sum^{l}\psi(ABC)\Big)^2+\frac{1}{nml}\Big(\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\Big)^2\\
&-\frac{2mj}{n}\Big(\sum^{n}\psi(A)\Big)^2+2j\Big(\psi(A)\sum^{m}\psi(AB)\Big)-\frac{2j}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\psi(AB)+2m\sum^{n}\Big(\psi(A)\sum^{l}\psi(AC)\Big)\\
&-\frac{2m}{n}\sum^{n}\psi(A)\sum^{n}\sum^{l}\psi(AC)+2\sum^{n}\Big(\psi(A)\sum^{m}\sum^{l}\psi(ABC)\Big)-\frac{2}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&-\frac{2j}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\psi(AB)+\frac{2j}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\psi(AB)-\frac{2m}{n}\sum^{n}\psi(A)\sum^{n}\sum^{l}\psi(AC)\\
&+\frac{2m}{n}\sum^{n}\psi(A)\sum^{n}\sum^{l}\psi(AC)-\frac{2}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)+\frac{2}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&-\frac{2j}{nm}\Big(\sum^{n}\sum^{m}\psi(AB)\Big)^2+2\sum^{n}\Big(\sum^{m}\psi(AB)\sum^{l}\psi(AC)\Big)-\frac{2}{n}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{l}\psi(AC)\\
&+\frac{2}{m}\sum^{n}\Big(\sum^{m}\psi(AB)\sum^{m}\sum^{l}\psi(ABC)\Big)-\frac{2}{nm}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&-\frac{2}{n}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{l}\psi(AC)+\frac{2}{n}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{l}\psi(AC)-\frac{2}{nm}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&-\frac{2}{nm}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)-\frac{2m}{nl}\Big(\sum^{n}\sum^{l}\psi(AC)\Big)^2+\frac{2}{l}\sum^{n}\Big(\sum^{l}\psi(AC)\sum^{m}\sum^{l}\psi(ABC)\Big)\\
&-\frac{2n}{l}\sum^{n}\sum^{l}\psi(AC)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)-\frac{2n}{l}\sum^{n}\sum^{l}\psi(AC)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&+\frac{2}{nl}\sum^{n}\sum^{l}\psi(AC)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)-\frac{2}{nml}\Big(\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\Big)^2=\\
&=mj\sum^{n}(\psi(A))^2-\frac{mj}{n}\Big(\sum^{n}\psi(A)\Big)^2+\frac{j}{m}\sum^{n}\Big(\sum^{m}\psi(AB)\Big)^2-\frac{j}{nm}\Big(\sum^{n}\sum^{m}\psi(AB)\Big)^2\\
&+\frac{m}{l}\sum^{n}\Big(\sum^{l}\psi(AC)\Big)^2-\frac{m}{nl}\Big(\sum^{n}\sum^{l}\psi(AC)\Big)^2+\frac{1}{ml}\sum^{n}\Big(\sum^{m}\sum^{l}\psi(ABC)\Big)^2-\frac{1}{nml}\Big(\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\Big)^2\\
&+2j\sum^{n}\Big(\psi(A)\sum^{m}\psi(AB)\Big)-\frac{2j}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\psi(AB)+2m\sum^{n}\Big(\psi(A)\sum^{l}\psi(AC)\Big)\\
&-\frac{2m}{n}\sum^{n}\psi(A)\sum^{n}\sum^{l}\psi(AC)+2\sum^{n}\Big(\psi(A)\sum^{m}\sum^{l}\psi(ABC)\Big)-\frac{2}{n}\sum^{n}\psi(A)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)\\
&+2\sum^{n}\Big(\sum^{m}\psi(AB)\sum^{l}\psi(AC)\Big)-\frac{2}{n}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{l}\psi(AC)+\frac{2}{m}\sum^{n}\Big(\sum^{m}\psi(AB)\sum^{m}\sum^{l}\psi(ABC)\Big)\\
&-\frac{2}{nm}\sum^{n}\sum^{m}\psi(AB)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)+\frac{2}{l}\sum^{n}\Big(\sum^{l}\psi(AC)\sum^{m}\sum^{l}\psi(ABC)\Big)-\frac{2}{nl}\sum^{n}\sum^{l}\psi(AC)\sum^{n}\sum^{m}\sum^{l}\psi(ABC)
\end{aligned}
$$

Beziehen we thans de interactie variance AB

$$
\sum^{n}\sum^{m}\sum^{l}\Big(\bar{y}_{kl.}-\bar{y}_{k..}-\bar{y}_{.l.}+\bar{y}\Big)^2=\sum^{n}\sum^{m}\sum^{l}\Bigg\{\frac{\sum^{l}\{\psi(A)+\cdots\psi(ABC)\}}{l}-\frac{\sum^{m}\sum^{l}\{\psi(A)+\cdots\psi(ABC)\}}{mj}-\frac{\sum^{n}\sum^{l}\{\psi(A)+\cdots\psi(ABC)\}}{nj}+\frac{\sum^{n}\sum^{m}\sum^{l}\{\psi(A)+\cdots\psi(ABC)\}}{nmj}\Bigg\}^2=\tag{II}
$$

$$
\sum^{n}\sum^{m}\sum^{l}\Bigg\{\psi(AB)-\frac{\sum^{m}\psi(AB)}{m}-\frac{\sum^{n}\psi(AB)}{n}+\frac{\sum^{n}\sum^{m}\psi(AB)}{nm}+\frac{\sum^{l}\psi(ABC)}{l}-\frac{\sum^{m}\sum^{l}\psi(ABC)}{ml}-\frac{\sum^{n}\sum^{l}\psi(ABC)}{nl}+\frac{\sum^{n}\sum^{m}\sum^{l}\psi(ABC)}{nml}\Bigg\}^2
$$

Hierin komen alleen nog maar de termen $\varphi(AB)$ en $\varphi(ABC)$ voor!
Uitschrijven voert tenslotte tot:

$$\frac{1}{nm}\sum\sum\left(\varphi(AB)\right)^2+\frac{1}{nm}\left(\sum\sum\varphi(AB)\right)^2-\frac{1}{m}\sum\left(\sum\varphi(AB)\right)^2-\frac{1}{m}\sum\left(\sum\varphi(AB)\right)^2$$
$$+\frac{1}{j}\sum\sum\left(\sum\varphi(ABC)\right)^2-\frac{1}{nj}\sum\left(\sum\sum\varphi(ABC)\right)^2-\frac{1}{mj}\sum\left(\sum\sum\varphi(ABC)\right)^2+$$
$$+\frac{1}{nmj}\left(\sum\sum\sum\varphi(ABC)\right)^2+\frac{2}{nm}\sum\sum\varphi(AB)\sum\sum\varphi(ABC)-\frac{2}{n}\sum\left(\sum\varphi(AB)\sum\sum\varphi(ABC)\right)$$
$$-\frac{2}{m}\sum\left(\sum\varphi(AB)\sum\varphi(ABC)\right)+2\sum\sum\left(\varphi(AB)\sum\varphi(ABC)\right) \qquad (IV)$$

Bezien we tot slot de interactievariance ABC

$$\sum\sum\sum\left(y_{ijk}-\bar{y}_{ij.}-\bar{y}_{i.k}-\bar{y}_{.jk}+\bar{y}_{i..}+\bar{y}_{.j.}+\bar{y}_{..k}-\bar{y}\right)^2=$$
$$\sum\sum\sum\Big\{\{\varphi(A)+\cdots\varphi(ABC)\}-\frac{\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{n}-\frac{\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{m}$$
$$-\frac{\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{j}+\frac{\sum\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{mj}+\frac{\sum\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{nj}$$
$$+\frac{\sum\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{nm}-\frac{\sum\sum\sum\{\varphi(A)+\cdots\varphi(ABC)\}}{nmj}\Big\}^2=$$
$$\sum\sum\sum\Big\{\varphi(ABC)-\frac{\sum\varphi(ABC)}{n}-\frac{\sum\varphi(ABC)}{m}-\frac{\sum\varphi(ABC)}{j}+$$
$$+\frac{\sum\sum\varphi(ABC)}{mj}+\frac{\sum\sum\varphi(ABC)}{nj}+\frac{\sum\sum\varphi(ABC)}{nm}-\frac{\sum\sum\sum\varphi(ABC)}{nmj}\Big\}^2$$

Hierin komt alleen nog maar de term (ABC) voor!
Verder uitwerken geeft ten lange leste:

$$\sum\sum\sum\left(\varphi(ABC)\right)^2-\frac{1}{m}\sum\sum\left(\sum\varphi(ABC)\right)^2-\frac{1}{m}\sum\sum\left(\sum\varphi(ABC)\right)^2-\frac{1}{j}\sum\sum\left(\sum\varphi(ABC)\right)^2$$
$$+\frac{1}{mj}\sum\left(\sum\sum\varphi(ABC)\right)^2+\frac{1}{nj}\sum\left(\sum\sum\varphi(ABC)\right)^2+\frac{1}{nm}\sum\left(\sum\sum\varphi(ABC)\right)^2-$$
$$-\frac{1}{nmj}\left(\sum\sum\sum\varphi(ABC)\right)^2 \qquad (V)$$

Met de vormen (III), (IV) en (V) kan thans elke regressievergelijking met 3 factoren verder uitgewerkt worden.

Bij 3 factoren kunnen we nu dus zeggen, als in de regressie vergelijking $y=\varphi(A)+\varphi(B)+\varphi(C)+\varphi(AB)+\varphi(AC)+\varphi(BC)+\varphi(ABC)$ de term $\varphi(ABC)=0$, de variance van de interactie van de 2e orde ABC altijd = 0 is.

Is daarentegen $(ABC)\neq 0$ dan zal deze variance nooit = 0 zijn en de variances van de interactie AB en van den factor A in 't algemeen ook niet, hoewel hierop uitzonderingen mogelijk zijn, waarbij het onverschillig is of $\varphi(A)$, $\varphi(AB)$, $\varphi(AC)$ al dan niet = 0 zijn; meestal zullen ze $\neq 0$ zijn.

Vormen we weer een overzicht:

	$\varphi(ABC)=0$								$\varphi(ABC)\neq 0$							
	$\varphi(AC)=0$				$\varphi(AC)\neq 0$				$\varphi(AC)=0$				$\varphi(AC)\neq 0$			
	$\varphi(AB)=0$		$\varphi(AB)\neq 0$		$\varphi(AB)=0$		$\varphi(AB)\neq 0$		$\varphi(AB)=0$		$\varphi(AB)\neq 0$		$\varphi(AB)=0$		$\varphi(AB)\neq 0$	
$\varphi(A)$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$
Var ABC	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$
Var AC	$=0$	$=0$	$=0$	$=0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$								
Var AB	$=0$	$=0$	$\neq 0$	$\neq 0$	$=0$	$=0$	$\neq 0$	$\neq 0$			$=0$ of $\neq 0$					
Var A	$=0$	$\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$								

Is $\varphi(ABC)=0$, dan gaat de variance voor de interactie van de 1e orde AB over in denzelfden vorm op den factor j na, die we reeds vonden bij beschouwing van 2 factoren n.l. in:

$$j \sum_{n} \sum_{m} \left\{ \varphi(AB) - \frac{\sum_{n} \varphi(AB)}{n} - \frac{\sum_{m} \varphi(AB)}{m} + \frac{\sum_{n} \sum_{m} \varphi(AB)}{n\,m} \right\}^2$$

De variance voor den factor A gaat over in:

$$j \sum_{n} \sum_{m} \left\{ \varphi(A) - \frac{\sum_{n} \varphi(A)}{n} + \frac{\sum_{m} \varphi(AB)}{m} + \frac{\sum_{n} \sum_{m} \varphi(AB)}{n\,m} + \right.$$
$$\left. \frac{\sum_{j} \varphi(AC)}{j} - \frac{\sum_{n} \sum_{j} \varphi(A)}{n\,j} \right\}^2$$

Nog moet nu hierin $\varphi(AC)=0$ zijn, opdat de variance voor den factor A dezelfde vorm aanneemt als bij 2 factoren. Is dan naast $\varphi(AC)=0$ ook nog $\varphi(AB)=0$ dan vinden we $j\,m \sum_{n} \left\{ \varphi(A) - \frac{\sum_{n} \varphi(A)}{n} \right\}^2$ weer op denzelfden factor j na hetzelfde als bij 2 factoren, als daarbij $\varphi(AB)=0$ is. Hetzelfde geldt voor de factoren B en C.

Omgekeerd vinden we bij een experiment met 3 factoren een betrouwbare interactie-variance van de 2e orde, dan moet in de regressiefunctie de term $\varphi(ABC)$ voorkomen (waarnaast in het algemeen ook $\varphi(A)$, $\varphi(B)$, $\varphi(C)$, $\varphi(AB)$, $\varphi(AC)$ en $\varphi(BC)$ zullen voorkomen, hoewel uitzonderingen mogelijk zijn.

Vinden we een betrouwbare interactie variance van de 1e orde AB, dan moet als $\varphi(ABC)=0$ in de regressiefunctie de term $\varphi(AB)$ voorkomen, is daarentegen $\varphi(ABC)\neq 0$, dan geldt de 1e regel, zoodat we niet zeker zijn over het optreden van $\varphi(AB)$.

Vinden we een betrouwbare variance voor factor A, dan moet als $\varphi(ABC)$ en $\varphi(AB)$ en $\varphi(AC)=0$ de term $\varphi(A)$ in de regressiefunctie optreden. Is één of meer der termen $\varphi(ABC)$, $\varphi(AB)$ of $\varphi(AC) \neq 0$ dan zijn we niet geheel zeker over het optreden van $\varphi(A)$.

Maken we weer een overzicht.

	$V(ABC)=0$								$V(ABC)\neq 0$							
	$V(AC)=0$				$V(AC)\neq 0$				$V(AC)=0$				$V(AC)\neq 0$			
	$V(AB)=0$		$V(AB)\neq 0$		$V(AB)=0$		$V(AB)\neq 0$		$V(AB)=0$		$V(AB)\neq 0$		VA(AB)=0		V(AB)≠0	
V(A)	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$	$=0$	$\neq 0$
$\varphi(ABC)$	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$=0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$
$\varphi(AC)$	$=0$	$=0$	$=0$	$=0$	$\neq 0$	$\neq 0$	$\neq 0$	$\neq 0$								
$\varphi(AB)$	$=0$	$=0$	$\neq 0$	$\neq 0$	$=0$	$=0$	$\neq 0$	$\neq 0$			$=0$ of $\neq 0$					
$\varphi(A)$	$=0$	$\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$	$=0$ of $\neq 0$								

We zullen thans het geval van 3 factoren uitwerken, voor het geval, dat in $\varphi(ABC)$ de termen A, B en C tot de 1e macht voorkomen. De regressievergelijking wordt dan:

$y = aA + bB + cC + dAB + eAC + fBC + gABC + h$

dus $\varphi(A)=aA \quad \varphi(B)=bB \quad \varphi(C)=cC \quad \varphi(AB)=dAB$

$\varphi(AC)=eAC \quad \varphi(BC)=fBC \quad \varphi(ABC)=gABC$

Voor den teller van de variance van factor A vinden we dan:

$$\sum_k^n \sum_l^m \sum_{kl}^{ij} \left(\bar{y}_{k..} - \bar{\bar{y}}\right)^2 = mj \sum^n a^2A^2 - \frac{mj}{n}\left(\sum^n aA\right)^2 + \frac{j}{m}\sum^n\left(\sum^m dAB\right)^2 - \frac{j}{nm}\left(\sum^n\sum^m dAB\right)^2$$
$$+ \frac{m}{j}\sum^n\left(\sum^{ij} eAC\right)^2 - \frac{mj}{n}\left(\sum^n\sum^{ij} cAC\right)^2 + \frac{1}{mj}\sum^n\left(\sum^m\sum^{ij} gABC\right)^2 - \frac{1}{nmj}\left(\sum^n\sum^m\sum^{ij} gABC\right)^2$$
$$+ 2j\sum^n\left(aA\sum^m dAB\right) - \frac{2j}{n}\sum^n aA\sum^n\sum^m dAB + 2m\sum^n\left(aA\sum^{ij} eAC\right)$$
$$- \frac{2m}{n}\sum^n aA\sum^n\sum^{ij} eAC + 2\sum^n\left(aA\sum^m\sum^{ij} gABC\right) - \frac{2}{n}\sum^n aA\sum^n\sum^m\sum^{ij} gABC$$
$$+ 2\sum^n\left(\sum^m dAB\sum^{ij} eAC\right) - \frac{2}{n}\sum^n\sum^m dAB\sum^{ij} eAC + \frac{2}{m}\sum^n\left(\sum^m dAB\sum^{ij} gABC\right)$$
$$- \frac{2}{nm}\sum^n\sum^m dAB\sum^n\sum^{ij} gABC + \frac{2}{j}\sum^n\left(\sum^{ij} eAC\sum^m\sum^{ij} gABC\right)$$
$$- \frac{2}{nj}\sum^n\sum^{ij} eAC\sum^n\sum^m\sum^{ij} gABC =$$
$$= a^2mj\sum A^2 - \frac{a^2mj}{n}\left(\sum A\right)^2 + \frac{d^2j}{m}\sum A^2\left(\sum B\right)^2 - \frac{d^2j}{nm}\left(\sum A\right)^2\left(\sum B\right)^2$$
$$+ \frac{e^2m}{j}\sum A^2\left(\sum C\right)^2 - \frac{e^2m}{j}\left(\sum A\right)^2\left(\sum C\right)^2 + \frac{g^2}{mj}\sum A^2\left(\sum B\right)^2\left(\sum C\right)^2$$
$$- \frac{g^2}{nmj}\left(\sum A\right)^2\left(\sum B\right)^2\left(\sum C\right)^2 + 2jad\sum A^2\sum B - \frac{2jad}{n}\left(\sum A\right)^2\sum B$$
$$+ 2aem\sum A^2\sum C - \frac{2aem}{n}\left(\sum A\right)^2\sum C + 2agj\sum A^2\sum B\sum C$$
$$- \frac{2agj}{n}\left(\sum A\right)^2\sum B\sum C + 2deg\sum A^2\sum B\sum C - \frac{2deg}{n}\left(\sum A\right)^2\sum B\sum C$$
$$+ \frac{2dgj}{m}\sum A^2\left(\sum B\right)^2\sum C - \frac{2dgj}{nm}\left(\sum A\right)^2\left(\sum B\right)^2\sum C + \frac{2egj}{j}\sum A^2\sum B\left(\sum C\right)^2$$
$$- \frac{2egj}{nj}\left(\sum A\right)^2\sum B\left(\sum C\right)^2 =$$
$$= \left\{\sum A^2 - \frac{\left(\sum A\right)^2}{n}\right\}\left\{a^2mj + 2jad\sum B + \frac{d^2j}{m}\left(\sum B\right)^2 + 2aem\sum C + \frac{e^2m}{j}\left(\sum C\right)^2\right.$$
$$\left. + 2(ag+de)\sum B\sum C + \frac{g^2}{mj}\left(\sum B\right)^2\left(\sum C\right)^2 + \frac{2dgj}{m}\left(\sum B\right)^2\sum C + \frac{2egj}{j}\sum B\left(\sum C\right)^2\right\}$$
$$= \left\{\sum A^2 - \frac{\left(\sum A\right)^2}{n}\right\}\left\{j\left(a\sqrt{m} + \frac{d\sum B}{\sqrt{m}}\right)\left(a\sqrt{m} + \frac{d\sum B}{\sqrt{m}}\right) + \right.$$
$$2\sum C\left(a\sqrt{m} + \frac{d\sum B}{\sqrt{m}}\right)\left(e\sqrt{m} + \frac{g\sum B}{\sqrt{m}}\right) +$$
$$\left.\frac{\left(\sum C\right)^2}{j}\left(e\sqrt{m} + \frac{g\sum B}{\sqrt{m}}\right)\left(e\sqrt{m} + \frac{g\sum B}{\sqrt{m}}\right)\right\} =$$
$$= \left\{\sum A^2 - \frac{\left(\sum A\right)^2}{n}\right\}\left\{\sqrt{j}\left(a\sqrt{m} + \frac{d\sum B}{\sqrt{m}}\right) + \frac{\sum C}{\sqrt{j}}\left(e\sqrt{m} + \frac{g\sum B}{\sqrt{m}}\right)\right\}^2 =$$
$$= \left\{\sum A^2 - \frac{\left(\sum A\right)^2}{n}\right\} mj\left(a + \frac{d}{m}\sum B + \frac{e}{j}\sum C + \frac{g}{mj}\sum B\sum C\right)^2. \qquad (23)$$

Voor den teller van de interactie van de 1e orde AB vinden we

$$\sum\sum(\bar{y}_{kl.} - \bar{y}_{k..} - \bar{y}_{.l.} + \bar{y}_{...})^2 = j\sum\sum d^2A^2B^2 + \frac{1}{nm}\left(\sum\sum dAB\right)^2$$
$$- \frac{1}{n}\sum\left(\sum dAB\right)^2 - \frac{1}{m}\sum\left(\sum dAB\right)^2 + \frac{1}{j}\sum\sum\left(\sum gABC\right)^2 - \frac{1}{nj}\sum\left(\sum gABC\right)^2$$
$$- \frac{1}{mj}\sum\left(\sum gABC\right)^2 + \frac{1}{nmj}\left(\sum\sum gABC\right)^2 + \frac{2}{nm}\sum\sum dAB\sum\sum gABC$$
$$- \frac{2}{n}\sum\left(\sum dAB\sum gABC\right) - \frac{2}{m}\sum\left(\sum dAB\sum gABC\right)$$
$$+ 2\sum\sum\left(dAB\sum gABC\right)$$

hetgeen na uitschrijven en samenvoegen voert tot:

$$\left\{\sum A^2 - \frac{(\sum A)^2}{n}\right\}\left\{\sum B^2 - \frac{(\sum B)^2}{m}\right\}\left\{d\sqrt{j} + \frac{g\sum C}{\sqrt{j}}\right\}^2 \qquad (24)$$

Vervolgens nog de teller van de interactie variance ABC:

$$\sum\sum\sum(y_{klr} - \bar{y}_{.l.} - \bar{y}_{k..} - \bar{y}_{..r} + \bar{y}_{k..} + \bar{y}_{.l.} + \bar{y}_{..r} - \bar{y}_{...})^2 =$$
$$\sum\sum\sum g^2A^2B^2C^2 - \frac{1}{n}\sum\sum\left(\sum gABC\right)^2 - \frac{1}{m}\sum\sum\left(\sum gABC\right)^2 + \frac{1}{j}\sum\sum\left(\sum gABC\right)^2$$
$$+ \frac{1}{mj}\sum\left(\sum gABC\right)^2 + \frac{1}{nj}\sum\left(\sum gABC\right)^2 + \frac{1}{nm}\sum\left(\sum gABC\right)^2$$
$$- \frac{1}{nmj}\left(\sum\sum gABC\right)^2$$

hetgeen na uitschrijven en samenvoegen voert tot:

$$g^2\left\{\sum A^2 - \frac{(\sum A)^2}{n}\right\}\left\{\sum B^2 - \frac{(\sum B)^2}{m}\right\}\left\{\sum C^2 - \frac{(\sum C)^2}{j}\right\} \qquad (25)$$

Hiermede zijn thans de variances te berekenen van elke regressievergelijking met 3 factoren van den gegeven vorm. We zien een groote analogie met het geval van 2 factoren, waar A en B eveneens tot den 1en graad in $\varphi(AB)$ voorkomen, als we Var(ABC) en Var(AB) bij 3 factoren vergelijken met de Var(AB) en Var(A) bij 2 factoren.

We passen thans weer toe de formules (5) en (6) uit §6 voor de sommatietermen.
Voor de som van quadraten van de interactie variance van de 2e orde ABC vinden we dan:
$g^2 \cdot \frac{n(n^2-1)}{12} \cdot \frac{m(m^2-1)}{12} \cdot \frac{j(j^2-1)}{12}$ en voor de variance zelve na deelen door (n-1)(m-1)(j-1) vinden we $\frac{g^2 n(n+1)m(m+1)j(j+1)}{12.12.12}$ (26)
dus constant voor een gegeven proefschema.

Voor de som van quadraten van de interactie variance van de 1e orde AB vinden we:
$j\frac{n(n^2-1)}{12} \cdot \frac{m(m^2-1)}{12} \cdot \left\{g^2i^2 + g(2d + gj - g)i + \frac{(2d+gj-g)^2}{4}\right\}$
en voor de variance zelve na deelen door (n-1)(m-1)
$j\frac{n(n+1)}{12} \cdot \frac{m(m+1)}{12}\left\{g^2i^2 + g(2d+gj-g)i + \frac{(2d+gj-g)^2}{4}\right\}$ op $\frac{m(m+1)}{12}$ na analoog met wat we vonden voor de variance voor den factor A bij 2 factoren.
Deze variance wordt naar analogie minimaal voor $i = -\frac{d}{g} + \frac{1-j}{2}$ (27)
waarvoor de variance AB = 0.

Voor waarden van i $<$ of $>$ dan $i_{min.}$ neemt de variance toe in waarde.

Voor de som van quadraten van de variance voor den factor A vinden we dan:

$$\frac{n(n^2-1)}{12}.mj.\left\{a+\frac{d}{m}m(1+\frac{m-1}{2}) + \frac{e}{j}j(i+\frac{j-1}{2}) + \frac{g}{mj}mj(1+\frac{m-1}{2})(i+\frac{j-1}{2})\right\}^2$$

en voor de variance zelve na deelen door n-1.

terwijl we $1+\frac{m-1}{2} = 1^1$ en $i+\frac{j-1}{2} = i^1$ stellen

$$\frac{n(n+1)}{12}\,mj\left\{a + dl^1 + ci^1 + gl^1i^1\right\}^2$$

Deze variance wordt minimaal als de vorm tusschen $\{\}$ = 0
dus als $a + ei^1 + (d+gi^1)l^1 = 0$

of als $l^1 = -\frac{a+ei}{d+gi^1} = -\frac{a}{d} - (e-\frac{ag}{d})i^1$

of $l = -\frac{a}{d} - (e-\frac{ag}{d})(i+\frac{j-1}{2}) - \frac{m-1}{2}$ (28)

De minimum waarde van V(A) = 0

Onderzoeken we thans de verhouding

$H = \frac{V(AB) - V(ABC)}{V(ABC)}$ der variances

$$\frac{j.\frac{n(n+1)}{12}.\frac{m(m+1)}{12}\left\{g^2i^2+g(2d+gj-g)i + \frac{(2d+gj-g)^2}{4}\right\}}{g^2\,\frac{n(n+1)}{12}\,\frac{m(m+1)}{12}\,\frac{j(j+1)}{12}} - 1 =$$

$$= \frac{\left\{g^2i^2+g(2d+gj-g)i + \frac{(2d+gj-g)^2}{4}\right\}}{g^2.\frac{j+1}{12}} - 1 = \frac{\left\{gi + (d + \frac{j-1}{2}g)\right\}^2}{g^2.\frac{j+1}{2}}$$

Een vorm op enkele termen na, analoog met wat we vonden bij 2 factoren voor de verhouding $\frac{V(A) - V(AB)}{V(AB)}$

Deze verhouding wordt minimaal voor dezelfde voorwaarden waarvoor V(AB) minimaal wordt.
Zooals we gezien hebben is dit het geval bij $i = \frac{1-j}{2} - \frac{d}{g}$

De waarde van deze verhouding is dan = -1.
Voor waarden $<$ of $>$ $i_{min.}$ neemt deze verhouding toe.
V(AB) = V(ABC) en dus H = 0 naar analogie van voorgaande voor

$$i = -\frac{d}{g} + \frac{1-j}{2} \pm \sqrt{\frac{j+1}{12}} \qquad (29)$$

Voor waarden $<$ of $>$ dan deze waarde van i, is de variance ABC altijd kleiner dan de variance AB, welk verschil steeds sneller toeneemt.

Bezien we hiernaast nog de verhouding $\frac{V(A) - V(ABC)}{V(ABC)}$

$$H = \frac{\frac{n(n+1)}{12}\,mj\left\{a+d(1+\frac{m-1}{2}) + e(i+\frac{j-1}{2}) + g(1+\frac{m-1}{2})(i+\frac{j-1}{2})\right\}^2}{g^2.\frac{n(n+1)}{12}.\frac{m(m+1)}{12}.\frac{j(j+1)}{12}} - 1 =$$

$$= \frac{(a + dl^1 + ei^1 + gl^1i^1)^2}{g^2.\frac{m+1}{12}.\frac{j+1}{12}} - 1.$$

Deze verhouding wordt minimaal als de vorm tusschen $\{\}$ minimaal wordt, dus voor dezelfde voorwaarden, waarvoor V(A) minimaal wordt.

$$l = -\frac{a}{d} - (e-\frac{ag}{d})(i+\frac{j-1}{2}) - \frac{m-1}{2}$$

De waarde van H in dit minimum = -1

$V(A)=V(ABC)$ dus $H = 0$ als

$$\frac{n(n+1)}{12} mj\,(a + dl^1 + ei^1 + gl^1i^1)^2 = g^2 \frac{n(n+1)}{12} \frac{m(m+1)}{12} \frac{j(j+1)}{12}$$

of $(a + dl^1 + ei^1 + gl^1i^1)^2 = g^2 \cdot \frac{m+1}{12} \cdot \frac{j+1}{12}$

$$(a + ei^1) + (d + gi^1)l^1 = \pm \frac{g}{12}\sqrt{(m+1)(j+1)}$$

dus $V(A) = V(ABC)$ voor

$$l = -\frac{a}{d} - (e\frac{ag}{d})(i+\frac{j-1}{2}) - \frac{m-1}{2} \pm \frac{\frac{g}{12}\sqrt{(m+1)(j+1)}}{d+g(i+\frac{j-1}{2})} \qquad (30)$$

Voor l < of > dan deze 2 waarden zal $V(ABC)$ altijd < $V(A)$ zijn!
De grootte van dit interval is echter thans niet constant meer voor een gegeven proefschema, maar afhankelijk van de coëfficienten d en g en de waarde i, de eerste trap waarin factor C onderzocht wordt, in tegenstelling met het 1e geval ($y=aA+bB+cAB+d$) waar we $\pm \frac{m+1}{12}$ vonden; zie pg.56.

§ 13. Voorbeeld 3

Nemen we de regressievergelijking
$Y=-9A-9B-9C+6AB+6AC+6BC-4ABC+85$ tot voorbeeld. Zouden we thans een 4 x 4 x 4 experiment genomen hebben, met als trappen van factor A, A_1, A_2, A_3 en A_4 voor factor B de trappen B_1, B_2, B_3 en B_4 en voor factor C de trappen C_1, C_2, C_3 en C_4, dan hadden we de volgende y-waarden verkregen:

	C_1					C_2					C_3					C_4					
	B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4		B_1	B_2	B_3	B_4		
A_1	72	71	70	69	282	71	72	73	74	290	70	73	76	79	298	69	74	79	84	306	1176
A_2	71	72	73	74	290	72	71	70	69	282	73	70	67	64	274	74	69	64	59	266	1112
A_3	70	73	76	79	298	73	70	67	64	274	76	67	58	49	250	79	64	49	34	226	1048
A_4	69	74	79	84	306	74	69	64	59	266	79	64	49	34	226	84	59	34	9	186	984
	282	290	298	306	1176	290	282	294	266	1112	298	274	250	226	1048	306	266	226	186	984	4320

Sommeeren over C geeft nog de volgende tabel

	B_1	B_2	B_3	B_4	
A_1	282	290	298	306	1176
A_2	290	282	274	266	1112
A_3	298	274	250	226	1048
A_4	306	266	226	186	984
	1176	1112	1048	984	4320

analysis of variance heeft tot resültaat:

SS totaal $72^2 + \ldots 9^2 - \frac{4320^2}{64} = 302.240 - 291.600 = 10.640$

SS(A) $\frac{1176^2 + \ldots 984^2}{16} - cf = 292.880 - 291.600 = 1.280$

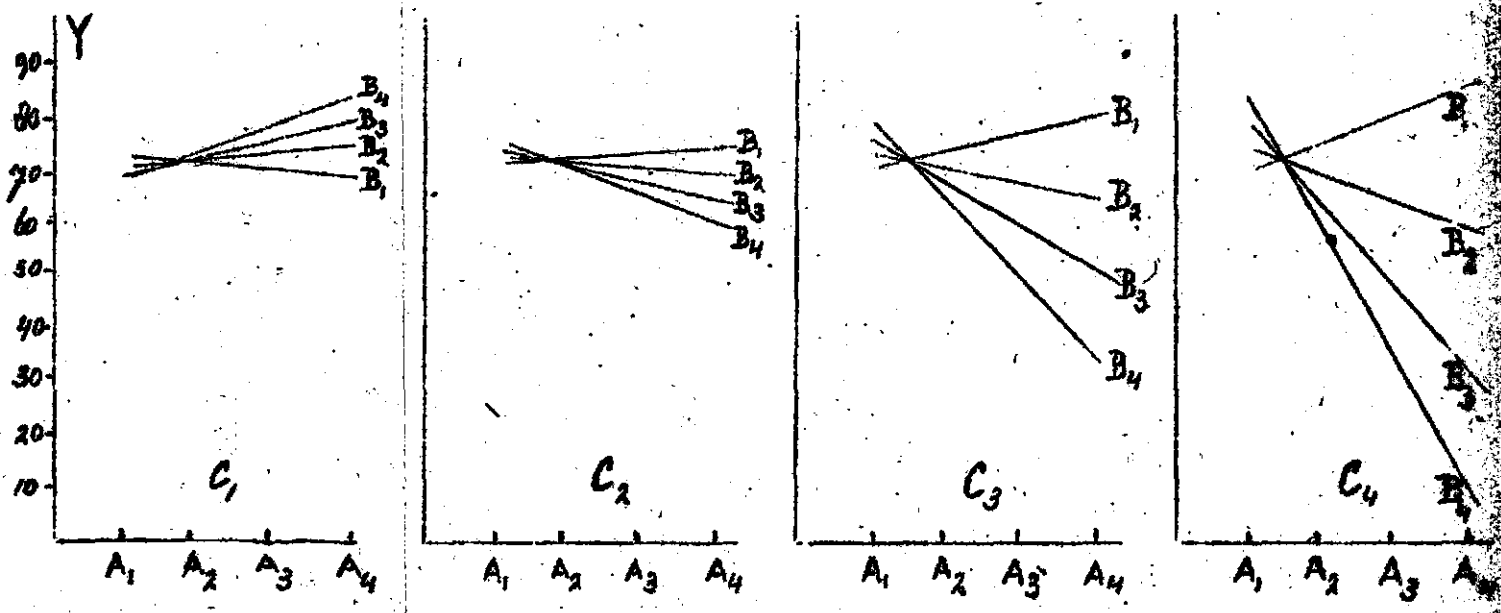
SS(B) en SS(C) idem

SS(AB) $\frac{282^2 + \ldots 186^2}{4} - cf - A - B = 295.760 - 291.600 - 2560 = 1600$

SS(AC) en SS(BC) iden

SS(ABC) = 10.640 - (3 x 1280 + 3 x 1600)= 2000

De graphische voorstelling is als volgt:



Berekenen we thans de analysis of variance weer uit de regressie-vergelijking, dan is a=b=c=-9 d=e=f=6 en g=-4
verder is n=m=j=4 en k=l=i=1
$\Sigma A = \Sigma B = \Sigma C = 10$ $\Sigma A^2 = \Sigma B^2 = \Sigma C^2 = 30$
Voor den teller van de variance voor den factor A vinden we dan:
formule(23): $\left\{30-\frac{10^2}{4}\right\} 4.4.\left\{-9+\frac{6}{4}.10+\frac{-4}{4.4}\ 10.10\right\}^2 = 1280$
Voor den teller van de variance voor de interacte AB
formule(24): $\left\{30-\frac{10^2}{4}\right\}\left\{30-\frac{10^2}{4}\right\}\left\{4.6^2 +2.6.-4.10+\frac{-4}{4}\ 10^2\right\} = 1600$
Voor den teller van de variance voor de interactie ABC
formule(25): $\left\{30-\frac{10^2}{4}\right\}\left\{30-\frac{10^2}{4}\right\}\left\{30-\frac{10^2}{4}\right\} = 2000$
gelijk wij ook vonden in de analysis of variance.

We zagen in de theorie dat de interactie van de 1e orde AB=0 wordt voor $i = -\frac{d}{g} + \frac{1-1}{2}$ dus voor $-\frac{6}{4} + \frac{1-4}{2} = 0$ dus factor C moet dan in de trappen C_0, C_1, C_2 en C_3 onderzocht worden, in welke trappen factor A en factor B onderzocht worden heeft geen invloed.
Nemen we 2 voorbeelden: 1ê de trappen A_1, A_2, A_3 en A_4, B_1, B_2, B_3, B_4 en C_0, C_1, C_2 en C_3 we vinden dan:

	C_0				C_1				C_2				C_3			
	B_1	B_2	B_3	B_4	B_1	B_2	B_3	B_4	B_1	B_2	B_3	B_4	B_1	B_2	B_3	B_4
A_1	73	70	67	64	72	71	70	69	71	72	73	74	70	73	76	79
A_2	70	73	76	79	71	72	73	74	72	71	70	69	73	70	67	64
A_3	67	76	85	94	70	73	76	79	73	70	67	64	76	67	58	49
A_4	64	79	94	109	69	74	79	84	74	69	64	59	79	64	49	34

Sommeeren over C geeft de volgende tabel

	B_1	B_2	B_3	B_4
A_1	286	286	286	286
A_2	286	286	286	286
A_3	286	286	286	286
A_4	286	286	286	286

waaruit zonder meer volgt dat de interactie AB=0 in dit geval. Als 2e voorbeeld de trappen A_3, A_4, B_6, B_7 en B_8 en weer C_0, C_1, C_2 en C_3. We vinden dan:

	C_0			C_1			C_2			C_3		
	B_6	B_7	B_8	B_6	B_7	B_8	B_6	B_7	B_8	B_6	B_7	B_8
A_3	112	121	130	94	97	100	76	73	70	31	22	13
A_4	139	154	169	103	108	113	67	62	57	4	-11	-26

Sommeeren over C geeft de volgende tabel:

	B_6	B_7	B_8
A_3	313	313	313
A_4	313	313	313.

waaruit weer zonder meer te zien is dat de interactie van de 1e orde AB in dit geval = 0 wordt, het is onverschillig in welke trappen B en C onderzocht worden.

De variance van factor A wordt = 0 als

$$1 = -\frac{a}{d} - (e - \frac{ag}{d})(i + \frac{j-1}{2}) - \frac{m-1}{2}$$

Gesteld dat we factor C in de trappen C_1 en C_2 onderzoeken in welke trappen moeten we dan factor B onderzoeken, opdat V(A)=0 zij (bij gegeven aantal trappen voor factor B, b.v. ook 2 om de berekeningen klein te houden).

$$1 = -\frac{-9}{6} - (6 - \frac{-9-4}{6})(1 + \frac{2-1}{2}) - \frac{2-1}{2} = \frac{3}{2} - 0 \times \frac{3}{2} - \frac{1}{2} = 1$$

Nemen we dus de trappen B_1 en B_2 naast de gegeven trappen C_1 en C_2 dan wordt V(A)=0.

Nemen we verder voor A de trappen 1e A_1, A_2 en A_3 en 2e A_4, A_5, A_6 en A_7 dan vinden we de volgende y-waarden:

	C_1			C_2			
	B_1	B_2		B_1	B_2		
A_1	72	71	143	71	72	143	286
A_2	71	72	143	72	71	143	286
A_3	70	73	143	73	70	144	286

	C_1			C_2			
	B_1	B_2		B_1	B_2		
A_4	69	74	143	74	69	143	286
A_5	68	75	143	75	68	143	286
A_6	67	76	143	76	67	143	286
A_7	66	77	143	77	66	143	286

Zonder verdere berekeningen kunnen we hieruit zien, dat de variance voor factor A=0 onafhankelijk van de trappen waarin factor A onderzocht wordt.

Zal V(ABC) kleiner zijn dan V(AB) dan moet

$$i < -\frac{d}{g} + \frac{1-j}{2} - \sqrt{\frac{j^2-1}{12}} \quad \text{of} \quad i > -\frac{d}{g} + \frac{1-j}{2} + \sqrt{\frac{j^2-1}{12}}$$

Voor 2 trappen van factor C vinden we dan:

$$i < +\frac{3}{2} - \frac{1}{2} - \frac{1}{2} \quad \text{of} \quad i < +\frac{1}{2}$$

$$i > +\frac{3}{2} - \frac{1}{2} + \frac{1}{2} \quad \text{of} \quad i > +1\frac{1}{2}$$

Dus is er slechts een zeer klein gebied waarin $V(ABC) > V(AB)$.
Nemen we als trappen A_3 en A_4, B_3 en B_4 en C_2 en C_3 dus de 1e trap van C grooter dan $1\frac{1}{2}$ en de trappen van A en B willekeurig, dan vinden we de volgende y-waarden:

	C_2			C_3						
	B_3	B_4		B_3	B_4				B_3	B_4
A_3	67	64	131	58	49	107	238	125	113	238
A_4	64	59	123	49	34	83	206	113	93	206
	131	123	254	107	83	190	444	238	206	444

analysis of variance geeft:

SS totaal $67^2 + \ldots\ldots - 34^2 - \frac{444^2}{8} = 842$

SS(A) $\frac{238^2 + 206^2}{4} - cf = \frac{32^2}{8} = 128$

SS(B) idem

SS(C) $\frac{254^2 + 190^2}{4} - cf = \frac{64^2}{8} = 512$

SS(AB) $\frac{125^2 + 113^2 + 113^2 + 93^2}{2} - cf - A - B = 8$

SS(AC) $\frac{131^2 + 123^2 + 107^2 + 83^2}{2} - cf - A - C = 32$

SS(BC) idem

SS(ABC) 842-840= 2

$V(AB) = \frac{8}{(2-1)(2-1)} = 8$ $\qquad V(ABC) = \frac{2}{(2-1)(2-1)(2-1)} = 2$

dus inderdaad $V(ABC) < V(AB)$.

Nemen we voor i een tussenliggende waarde dan moet $V(ABC) > V(AB)$ b.v. $i = 1$ dus de trappen C_1 en C_2.
We vinden dan de volgende y-waarden

	C_1			C_2						
	B_3	B_4		B_3	B_4				B_3	B_4
A_3	76	79	155	67	64	131	286	143	143	286
A_4	79	84	163	64	59	123	286	143	143	286
	155	163	318	131	123	254	572	286	286	572

Zonder verdere berekening zien we uit de tabel dat $V(AB)=0$ en $V(ABC)\neq 0$ dus inderdaad $V(ABC) > V(AB)$.

Zal V(ABC) kleiner zijn dan V(A) dan moet

$$1 \gtrless -\frac{a}{d} - (c-\frac{ag}{d})(i+\frac{j-1}{2})\frac{m-1}{2} \pm \frac{\frac{e}{12}\sqrt{(m+1)(j+1)}}{d+g(i+\frac{j-1}{2})}.$$

Gesteld we onderzoeken C in 8 trappen C_2 t/m C_9 en B in 3 trappen dan vinden we

$$1 \gtrless -\frac{-9}{6} - (6-\frac{-9-4}{6})(2+\frac{8-1}{2})\frac{3-1}{2} \pm \frac{\frac{-4}{12}\sqrt{(3+1)(8+1)}}{6-4(2+\frac{8-1}{2})}$$

of $1 \gtrless \frac{1}{2} \pm \frac{1}{8}$ $\qquad$ b.v. $1 = 0$

(tevens kunnen we hieruit nog vooruit zeggen hoe V(AB) zich zal gedragen:
Zal V(ABC) ook kleiner dan V(AB) zijn dan moet

$$i \gtrless -\frac{d}{g} + \frac{1-j}{2} \pm \sqrt{\frac{1+j}{12}} \quad \text{of} \quad i \gtrless -2 \pm \sqrt{\tfrac{3}{4}}$$

$i = +2$ dus $V(ABC) < V(AB)$.
We nemen voor C dus de trappen C_2 t/m C_9.
Voor B de trappen B_0, B_1 en B_2 ($f=0$, $m=3$), terwijl de trappen van A geen invloed hebben, kiezen we A_1 en A_2.

	A_1				A_2				
	B_0	B_1	B_2		B_0	B_1	B_2		
C_0	70	71	72	213	73	72	71	216	429
C_1^2	67	70	73	210	76	73	70	219	429
C_2^3	64	69	74	207	79	74	69	222	429
C_3^4	61	68	75	204	82	75	68	225	429
C_4^5	58	67	76	201	85	76	67	228	429
C_5^6	55	66	77	198	88	77	66	231	429
C_6^7	52	65	78	195	91	78	65	234	429
C_7^8	49	64	79	192	94	79	64	237	429
C_8^9	476	540	604	1620	668	604	540	1812	3432

analysis of variance geeft:

SS totaal $70^2 + \ldots. 64^2 - \frac{3432^2}{48} = 3740$

$SS(A)$ $\frac{1620^2 + 1812^2}{24} - cf = \frac{192^2}{48} = 768$

$SS(B)$ $\frac{1144^2 + 1144^2 + 1144^2}{16} - cf = 0$

$SS(C) = 0$

$SS(AB) = \frac{476^2 + \ldots. 540^2}{8} - cf - A - B = 798$

$SS(AC) = \frac{213^2 + \ldots. 237^2}{3} - cf - A - C = 252$

$SS(BC) = 0$

$SS(ABC) = 3740 - (768 + \ldots 0) = 1922$
derhalve $V(A) = \frac{768}{24} = 768$

$$V(AB) = \frac{798}{(2-1)(3-1)} = 399$$

$$V(ABC) = \frac{1922}{(2-1)(3-1)(8-1)} = 137.3$$

dus inderdaad $V(ABC) < V(A)$ en ook $V(ABC) < V(AB)$

§ 14. De toevalsvariance als interactievariance

In werkelijkheid hebben we in de experimenten natuurlijk altijd te doen met een stochastisch verband, dat we kenschetsen door een regressievergelijking. In onze voorbeelden werkten we altijd met functioneel verband, hoewel we steeds de naam regressiefunctie gebruikten. In werkelijkheid komt natuurlijk nooit (hoogstens zeer toevallig) een variance voor, die = 0 is: alle variances zijn in principe gelijk of grooter dan de toevalsvariance (alleen door het toeval soms kleiner). De gegeven voorbeelden dienen alleen als model. Om het spel der factoren en haar analyse te begrijpen is het gemakkelijk geïdealiseerde modellen voor oogen te hebben.

We hebben thans dus ons eerste gestelde doel bereikt nl. onderzoek wanneer de variances van hoogere orde = 0 worden, of feitelijk

en wat belangrijker is, hun onderlinge verhouding. Het blijkt dat in deze experimenten in enkelvoud de variances van interacties van hoogere orde lang niet altijd de kleinste zijn en bijgevolg lang niet altijd dienst kunnen doen als toevalsvariance. Een en ander is afhankelijk van plaats en grootte van het interval waarin de regressiefunctie onderzocht wordt, of wat hetzelfde is, het aantal en de waarde der trappen waarin de factoren onderzocht worden.

We kunnen hier volstaan met het vergelijken van de variances van interacties om na te gaan of de variances van interacties van hoogere orde de toevalsvariance zullen benaderen, aangezien de toevalsvariance bij het vergelijken van meerdere objecten altijd opgevat kan worden als een variance van een interactie of de som van variances van een aantal interacties.

(Ter vergelijking diene het volgende b.v. een blokkenproeven in den landbouw: onderzoeken we een aantal objecten b.v. varieteiten van een gewas in een aantal herhalingen c.q. blokken, dan vindt men uit de analysis of variance de variances voor rassen, blokken en toeval. Deze toevalsvariance kunnen we opvatten als de variance van de interactie tusschen rassen en blokken. Evenzoo als we 2 factoren A en B b.v. 2 meststoffen onderzoeken; dan vinden we in de analysis of variance een variance voor den factor A, voor den factor B, voor de interactie AB, een voor de blokken en een voor toeval; deze toevalsvariance is op te bouwen uit de som van de variances van 3 interacties nl. blokken xA, blokken xB en blokken xAxB).

§ 15. Bewijsgang voor niet-aequistante ordinaten.

Tot nog toe namen we trappen met gelijke intervallen aan, waarin de factoren onderzocht worden. De trappen kunnen echter ook vrij willekeurig zijn, mits men maar de orthogonaliteit (d.i. het voorkomen van alle combinaties) van het proefplan behoudt. Dit komt er op neer, dat de regressiefuncties dan bij niet-aequidistante ordinaten onderzocht worden en hiervoor gelden de formules van de reeksen van hoogere orde die wij gebruikt hebben, niet meer.
De bewijsgang moeten we daarom als volgt wijzigen.

De niet-aequidistante abciswaarden x zijn uit te drukken als een functie $F(x')$ van de aequidistante abciswaarden x'.
We kunnen nu vragen: wat is de regressiefunctie $y=(x')$ die voor de aequidistante abciswaarden dezelfde ordinaatwaarden geeft als de niet-aequidistante abciswaarden voor de functie $f(x)$.
v.b. Niet-aequidistante ordinaten

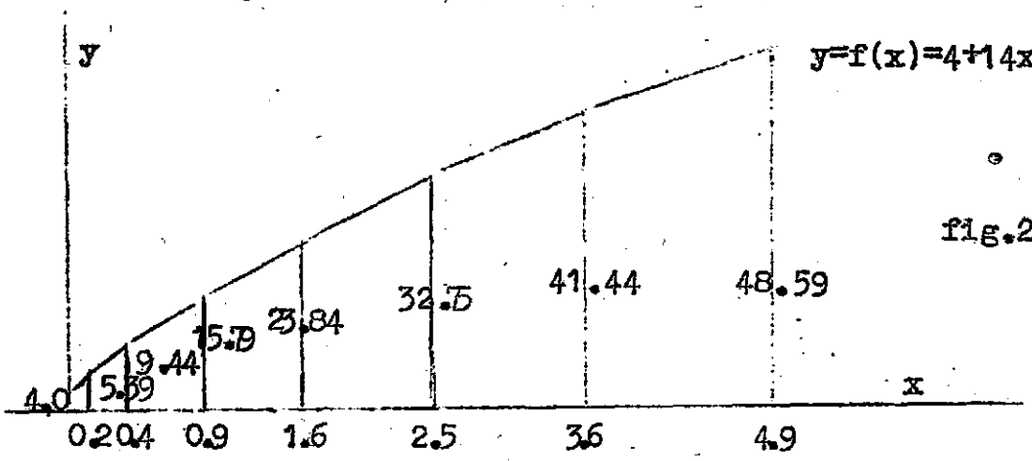


fig.23

De niet-aequidistante abciswaarden x zijn uit te drukken als een functie $F(x')$ van de aequidistante abciswaarden x'.
In osn voorbeeld is $x = F(x') = 0.1\ x'^2$

De gevraagde regressiefunctie $_2(x^1)$ is dan $y=_2(x^1) = f(F(x^1)$
of $y = 4 + 14(0.1\ x^{1^2}) - (0.1\ x^{1^2}) = 4 + 1.4\ x^{1^2} - 0.01\ x^{1^4}$

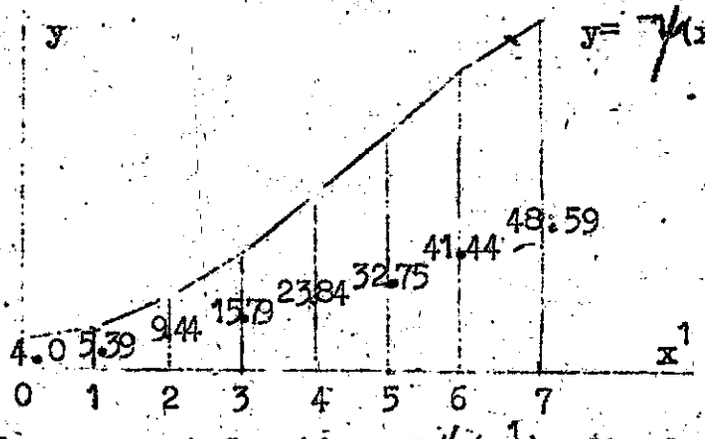


fig. 24

Voor de regressiefunctie $y = \psi(x^1)$, die dan uit te breiden is tot $y = \varphi(x^1) + \varphi(z^1) + \ldots\ldots \varphi(x^1 z^1) + \ldots\ldots$
kan dan weer dezelfde bewijsvoering als in voorgaande toegepast worden.

Bij willekeurige aboiswaarden wordt $F(x^1)$ natuurlijk zeer inGewikkeld, hetgeen echter aan het principe niets afdoet
b.v. x=10, 14, 28, 29, 31, 41, 50, 51 en 52.
dan krijgen we voor $x = F(x^1)$ de volgende graphische voorstelling als we de x-waarden volgens hun grootte rangschikken.
$F(x^1)$ is hoogstens van den 8en graad.

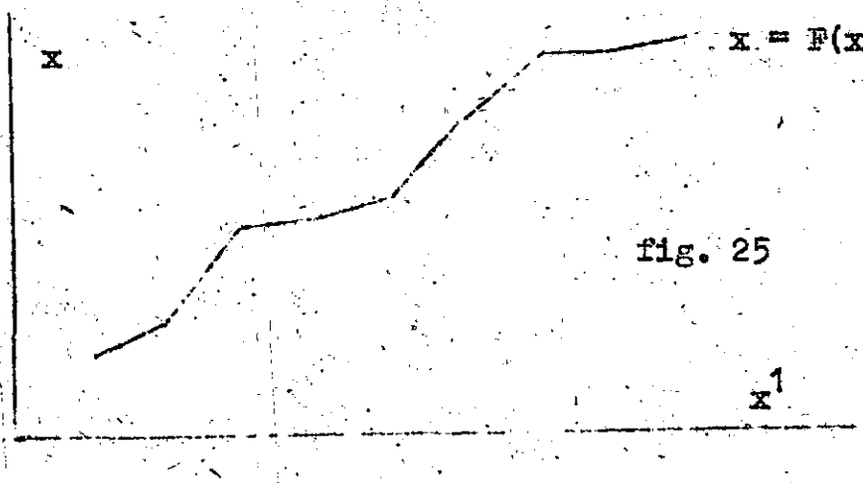


fig. 25

§ 16. Drie korte voorbeelden over het afleiden van de regressievergelijking uit de analysis of variance en omgekeerd.

We gaan hierbij uit van de gegeven formules (I) t/m (V) met de bijbehoorende overzichten, zonder verdere berekeningen uit te voeren. Het doel is de weg aan te geven hoe men de vereffeningsfunctie voor een reeks gegeven waarnemingen vindt! We geven daarbij als voorbeelden ook hoogere graadsvormen in de regressievergelijking bij 3 factoren, die niet in de theorie uitgewerkt zijn. Voor de toepassing van de genoemde overzichten is dit echter geen bezwaar, daar die onafhankelijk van den graad der regressievergelijking zijn.

1e voorbeeld: $y = 7A - 2A^2 + 8B - B^2 + 6C - C^2 + AB + 2AC + BC - 20$ dus
$y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB) + \varphi(AC) + \varphi(BC)$
Graphisch voorgesteld aldus:

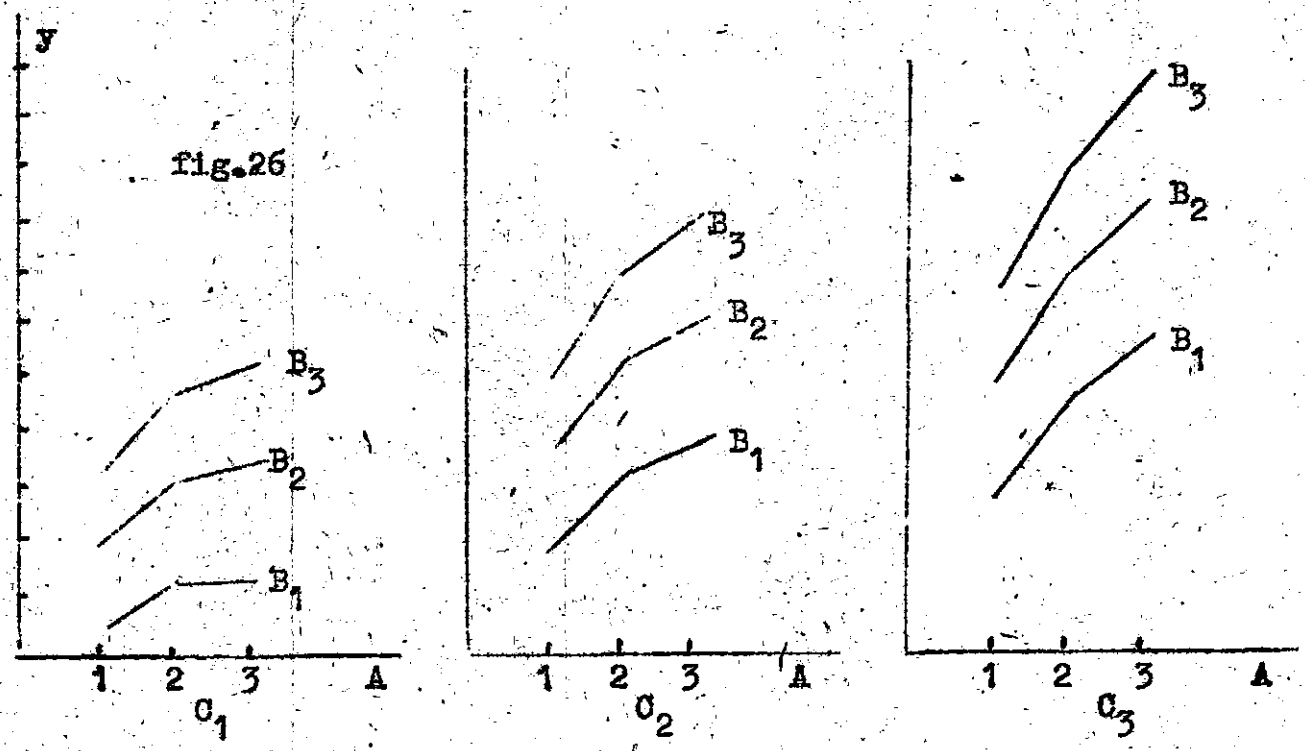


fig.26

We leiden de analysis of variance af uit de regressievergelijking $\varphi(ABC) = 0$, dus de variance van de interactie van de 2e orde zal eveneens = 0 zijn. In de variances van de interacties van de 1e orde komen alleen voor $\varphi(AB)$ en $\varphi(ABC)$ respectievelijk $\varphi(AC)$ en $\varphi(ABC)$ en nog $\varphi(BC)$ en $\varphi(ABC)$ m.a.w. zullen in dit geval alle 3 interacties van de 1e orde in de analysis of variance moeten voorkomen.

In de variance van den factor A komt voor $\varphi(A)$, $\varphi(AB)$, $\varphi(AC)$ en $\varphi(ABC)$; alleen $\varphi(ABC) = 0$ het is dus niet a priori te zeggen of de variance van factor A aanwezig zal zijn over 't algemeen echter wel, hetzelfde geldt voor de factoren B en C.

Voeren we de analysis of variance uit met als trappen A_1 A_2 A_3, B_1 B_2 B_3 en C_1 C_2 en C_3 dus als 3 x 3 x 3 experiment.

	B_1	B_2	B_3		B_1	B_2	B_3		B_1	B_2	B_3			B_1	B_2	B_3
A_1	1	8	13	22	7	15	21	48	11	20	27	58	123	19	43	61
A_2	5	13	19	37	13	22	29	64	19	29	37	85	186	37	64	85
A_3	5	14	21	40	15	25	33	73	23	34	43	100	213	43	73	97
	11	35	53	99	35	62	83	180	53	83	107	243	522	99	180	243

SS totaal $1^2 + \ldots + 43^2 - \frac{522^2}{27} = 2862$

SS(A) $\frac{123^2+186^2+213^2}{9} - cf = 474$

SS(B) $\frac{99^2+180^2+243^2}{9} - cf = 1158$

SS(C) idem $= 1158$

SS(AB) $\frac{19^2+\ldots 97^2}{3} - cf - A - B = 12$

SS(AC) $\frac{22^2+\ldots 100^2}{3} - cf - A - C = 48$

SS(BC) $\frac{10^2+\ldots 107^2}{3} - cf - B - C = 12$

SS(ABC) $= 2862 - 2862 = 0$

We leiden thans uit de resultaten van de analysis of variance de regressievergelijking af. Er zijn 3 factoren. De regressievergelijking kan dus hoogstens van den volgenden vorm zijn
$y = \varphi(A)+\varphi(B)+\varphi(C)+\varphi(AB)+\varphi(AC)+\varphi(BC)+\varphi(ABC)$. De variance van de interactie van de 2e orde is nul (niet reëel), dus de term $\varphi(ABC)$ is nul en moet dus verdwijnen uit de regressievergelijking.

Alle 3 interacties van de 1e orde zijn $\neq 0$ (reëel) dus moeten in de regressievergelijking $\varphi(AB)$ $\varphi(AC)$ en $\varphi(BC)$ voorkomen.

Ook de 3 variances voor de 3 factoren zijn $\neq 0$ (reëel) doch dit kan veroorzaakt zijn door $\varphi(AB)$ $\varphi(AC)$ en $\varphi(BC)$. De regressievergelijking moet dus in ieder geval van den vorm $y = \varphi(AB)+\varphi(AC)+\varphi(BC)$ zijn en in het algemeen zal daar ook nog $\varphi(A)$, $\varphi(B)$ en $\varphi(C)$ in voorkomen; we kiezen dus als vereffeningsfunctie $y = \varphi(A)+\varphi(B)+\varphi(C)+\varphi(AB)+\varphi(AC)+\varphi(BC)$.

2e voorbeeld: $y=7A-2A^2+8B-B^2+6C-C^2+AB+AC-15$
dus $y=\varphi(A)+\varphi(B)+\varphi(C)+\varphi(AB)+\varphi(ABC)$.
Graphisch voorgesteld aldus:

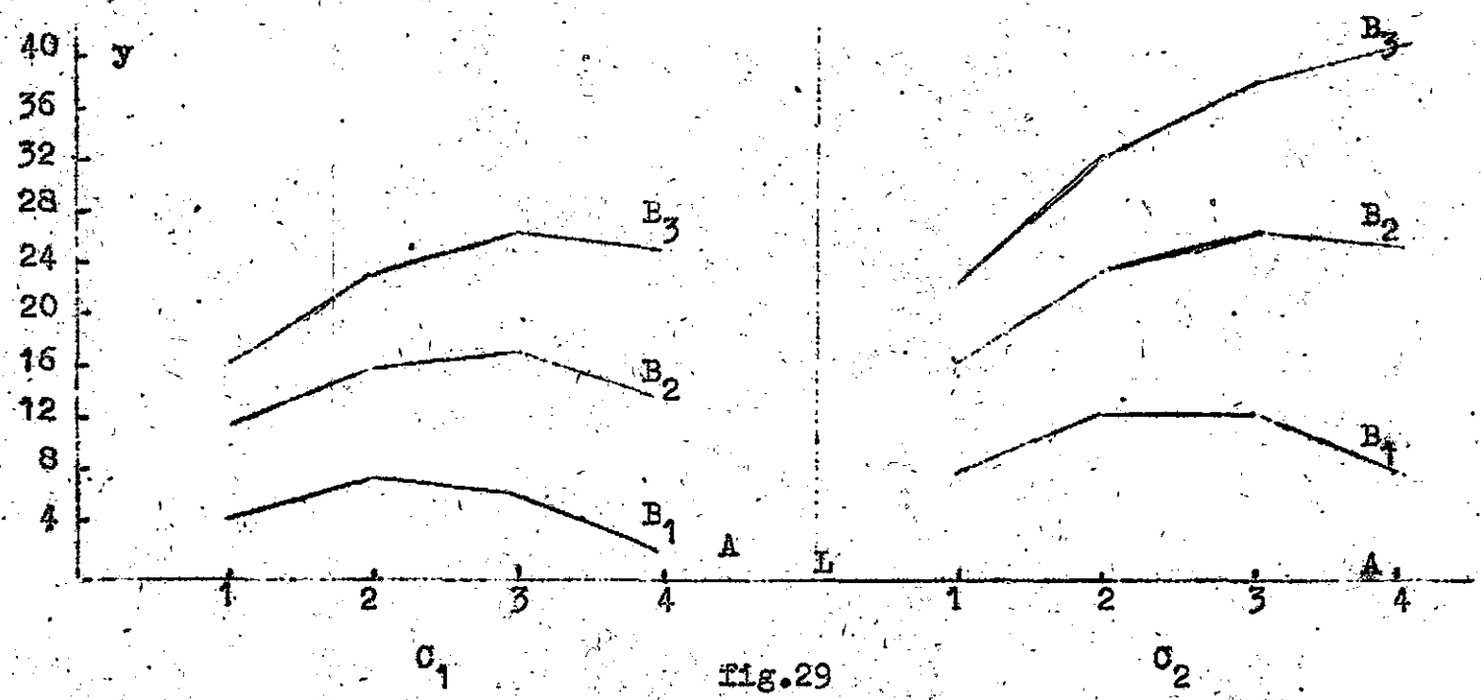


fig.29

We leiden thans de analysis of variance af uit de regressievergelijking: $\varphi(ABC) \neq 0$, dus moet de variance van de interactie van de 2e orde reëel zijn. Van de overige variances valt nu niets te zeggen, over 't algemeen zullen ze aanwezig zijn.
Voeren we een analysis of variance uit
(4 x 3 x 2 experiment):

	C_1				C_2							
	B_1	B_2	B_3		B_1	B_2	B_3			B_1	B_2	B_3
A_1	4	11	16	31	8	16	22	46	77	12	27	38
A_2	7	16	23	46	12	23	32	67	113	19	39	55
A_3	6	17	26	49	12	26	38	76	125	18	43	64
A_4	1	14	25	40	8	25	40	73	113	9	39	65
	18	58	90	166	40	90	132	262	428	58	148	222

SS totaal $4^2 + \ldots. 40^2 - \frac{428^2}{24} = 2472.3$

SS(A) $\frac{77^2+ \ldots. 113^2}{6} - cf = 216$

SS(B) $\frac{58^2+148^2+222^2}{8} - cf = 1686.3$

SS(C) $\frac{166^2 + 262^2}{12} - cf = 384$

SS(AB) $\frac{12^2 + \ldots. 65^2}{2} - cf - A - B = 125$

SS(AC) $\frac{31^2 + \ldots. 73^2}{3} - cf - A - C = 30$

SS(BC) $\frac{18^2 + \ldots. 132^2}{4} - cf - B - C = 25$

SS(ABC) $2471.\beta - 2466.\beta = 5$

We leiden thans de regressievergelijking af uit de analysis of variance. De variance voor de interactie van de 2e orde is reëel ($\neq 0$), dus moet φ(ABC) in de regressievergelijking voorkomen. φ(ABC) komt echter ook in de variances van de interacties AB, AC en BC voor en ook in de variances voor de factoren A, B en C. Over het optreden van φ(A), φ(B), φ(C), φ(AB), φ(AC) en φ(BC) laat zich dus a priori niets zeggen, in het algemeen zullen ze wel voorkomen, en we kiezen dus als vereffeningsfunctie $y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB) + \varphi(AC) + \varphi(BC) + \varphi(ABC)$.

3e voorbeeld: $y = 7A - 2A^2 + 6C - C^2 + AB$
dus $y = \varphi(A) + \varphi(C) + \varphi(AB)$
Graphisch voorgesteld aldus:

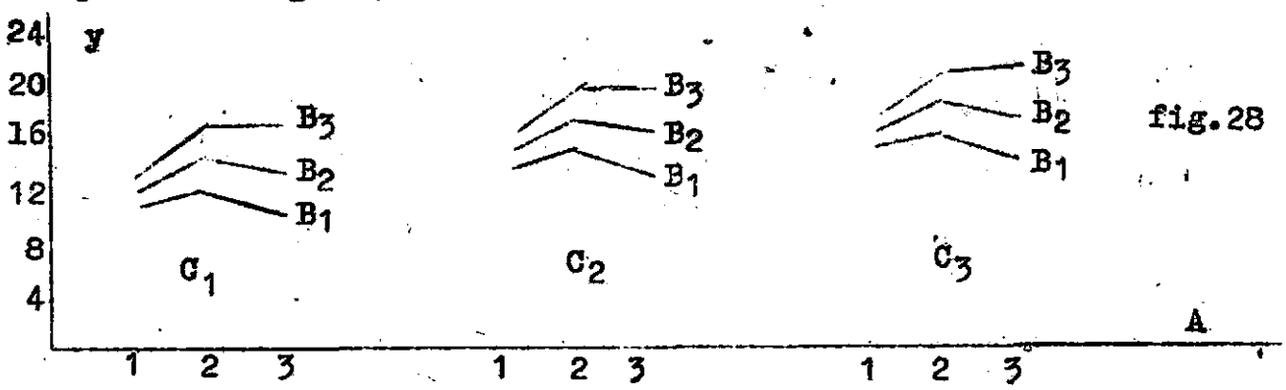


fig.28

Leiden we de analysis of variance uit de regressievergelijking af: φ(ABC) = 0 dus de variance voor de interactie van de 2e orde ABC moet = 0 zijn.
φ(AC) en φ(BC) beide = 0. Daar in de variances voor de interacties van de 1e orde AC en BC respectievelijk alleen voorkomen φ(AC), φ(ABC) en φ(BC), φ(ABC) zullen ook deze beide variances = 0 zijn. De variance voor de interactie AB echter niet, daar φ(AB) aanwezig is.

Of de variances voor de factoren A en B aanwezig zullen zijn, valt a priori niet te zeggen, in het algemeen zullen ze wel aanwezig zijn.

In de variance voor factor C komt voor: φ(C), φ(AC), φ(BC) en φ(ABC), deze zijn alle = 0 op φ(C) na; de variance voor factor C moet dus aanwezig zijn ! Voeren we een analysis of variance uit: (3x3x3 experiment)

	C_1				C_2				C_3							
	B_1	B_2	B_3		B_1	B_2	B_3		B_1	B_2	B_3			B_1	B_2	B_3
A_1	11	12	13	36	14	15	16	45	15	16	17	48	129	40	43	46
A_2	13	15	17	45	16	18	20	54	17	19	21	57	156	46	52	58
A_3	11	14	17	42	14	17	20	51	15	18	21	54	147	40	49	58
	35	41	47	123	44	50	56	150	47	53	59	159	432	126	144	162

SS totaal $11^2 + \ldots. 21^2 - \frac{432^2}{27} = 204$

SS(A) $\frac{129^2 + \ldots. 147^2}{9}$ - cf = 42

SS(B) $\frac{126^2 + \ldots. 162^2}{9}$ - cf = 72

SS(C) $\frac{123^2 + \ldots. 159^2}{9}$ - cf = 78

SS(AB) $\frac{40^2 + \ldots. 58^2}{3}$ - cf - A - B = 12

SS(BC) $\frac{35^2 + \ldots. 59^2}{3}$ - cf - B - C = 0

SS(AC) $\frac{36^2 + \ldots. 54^2}{3}$ - cf - A - C = 0

SS(ABC) = 204 - 204 = 0

Leiden we thans de regressievergelijking af uit deze analysis of variance.

De variance voor de interactie van de 2e orde ABC en van de 1e orde AC en BC zijn = 0; φ(ABC), φ(AC) en φ(BC) komen derhalve niet voor in de regressievergelijking. De variance voor de interactie AB is $\neq 0$ (reeel), φ(AB) moet derhalve voorkomen!

Of φ(A) en φ(B) zullen voorkomen, valt a priori niet te zeggen, over 't algemeen zullen ze wel voorkomen. De variance voor den factor C is $\neq 0$, daar de variance voor AC eh ABC wel = 0 zijn, moet φ(C) voorkomen in de regressiefunctie.

We kiezen dus als vereffeningsfunctie

$$y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB).$$

een belangrijke reductie dus, vergeleken met denmeest algemeenen vorm voor 3 factoren!

HOOFDSTUK VI

DE REGRESSIEFUNCTIES IN DE NATUUR.

§ 1. Algemeene types.

De tot nu toe behandelde regressiefuncties zijn zuiver algebraisch en wel in het byzonder rationeele geheele functies , waarbij het argument loopt van $-\infty$ tot $+\infty$.

Richten we ons thans tot de natuurwetenschappelijke proeven, dan beperken we ons tot reactiekrommen met een kleiner getallen gebied, in den landbouw in het byzonder opbrengstcurven, optimumkrommen of hoe men ze ook noemen wil.

De gebruikelijke opvatting van het begrip factor brengt met zich mee, dat we ons beperken tot positieve waarden van het argument.

Over de mathematische weergave van opbrengstcurven bestaat weinig eenstemmigheid. Bekend zijn de curven van Mitcherlich, de logistische curven en de curven van Pearl.

Een overzichtelijke rangschikking verkrijgt men door den vorm $\frac{dy}{dx}$ na te gaan.

a) $\frac{dy}{dx} = c.y$ of $y = e^{cx+p}$. Veel gebruikt in de biologie, het beschrijft ook de samengestelde interest: het resultaat van de aangroeiing, bepaalt tegelijkertijd ook weer het bedrag van de volgende aangroeiing.

b) $\frac{dy}{dx} = c(A-y)$ of $y = A(1-e^{-cx})$ (met de conditie $x = 0 \rightarrow y = 0$). Het is de bekende Mitcherlich curve in den landbouw; tevens beschrijft het de monomoleculaire reactie: de aangroeiing is evenredig met het bedrag dat nog ontbreekt aan het maximaal bereikbare resultaat.

c) $\frac{dy}{dx} = c.y.(A-y)$ of $y = \frac{A}{1+pe^{-Acx}}$, een soort combinatie van beide voorgaande vormen, veel gebruikt voor populatiecurven (demographen: Verhulst, Pearl). In de botanie beschrijft het de groote periode van Sachs e.d.: beschrijft tevens sommige autokatalytische reacties. De curve staat bekend onder den naam logistische curve (een enkele maal Robertsonsche curve).

Daarnaast zijn nog andere vergelijkingen in gebruik bv. Czuber gebruikt $y = Ax^m e^{nx^p}$. Een zeer "buigzame" kromme !

Mitcherlich gebruikte ook, om tegemoet te komen aan een dalende tak van de optimumcurve de vergelijking $y = A(1-10^{-cx})10^{-kx^2}$.
(Hierin in c evenals in den 1e vorm van Mitcherlich de bekende werkingsfactor, terwijl k de schadelijkheids constante genoemd wordt).

Feitelijk is het aantal opgestelde vergelijkingen voor opbrengstwetten zeer groot. Meestal heeft men daarbij slechts één factor beschouwd, er komt slechts één variabele in de vergelijking voor.

Waar ons doel is, met een groot aantal factoren tegelijk te werken is het noodzakelijk deze beperktheid nader onder oogen te zien, door de historische ontwikkeling van de theorie der opbrengstcurven zeer in het kort na te gaan.

§ 2. Historische ontwikkeling van de theorie der opbrengstcurven.

Het werken met wiskundige relaties in overeenstemming met de natuurwetenschappelijke opvatting van het begrip factor, veronderstelt, dat het te bestudeeren verschijnsel meetbaar is. Vooral in de economie heeft men zich de vraag gesteld of alle invloeden, waaronder vele gevoelsimpulsen wel voor meting vatbaar zijn. In de natuurwetenschappen stelde men zich deze vraag reeds eerder. Bekend is de uitspraak van Galileo Galilei: "Meet wat te meten is en maakt meetbaar, wat het nog niet is", iets wat de hedendaagsche psychologie maar al te graag doet en waarop terecht thans vele reacties komen. Nochtans duurde het lang, voordat men het quantitatieve aspect der relaties opzocht [1]). Bezien we eerst den Landbouw.

Anderson en Young, tijdgenooten van Adam Smith, huldigden de opvatting, dat in den landbouw van toenemende meeropbrengsten sprake is, terwijl naderhand Malthus, Ricardo en van Thünen de leer van de wet der afnemende meeropbrengsten in den landbouw verkondigd hebben.

---

[1]) Dr.O.Bakker. De Wetten der toe- en afnemende meeropbrengsten. Amsterdam 1932.

Wanneer in een nog betrekkelijk schaars bevolkt land, de grond bij toeneming der bevolking met behulp van nieuwe en betere methoden, intensiever bebouwd wordt, is de kans zeer groot, dat de landbouwproductie sneller toeneemt dan de bevolking. Deze mogelijkheid bestond ten tijde van, en werd ingezien door Anderson en Young, vandaar hun bewering van toenemende meeropbrengsten. Zij zagen vooral het dynamische element; in hun tijd veranderden de omstandigheden zeer sterk, waardoor toenemende meeropbrengsten konden ontstaan. Als later de bebouwing van den grond is toegenomen, de dynamische mogelijkheden tamelijk uitgeput zijn, groote niet agrarische bevolkingen ontstaan, de graanprijzen en de korenwetten tot veel discussie en onderzoek aanleiding geven, komen de economen tot de conclusie, dat in den landbouw een wet van afnemende meeropbrengsten heerscht, hetgeen zeer lang de klassieke opvatting blijft. Hierin treft men weer het statische element.

Turgot is de eenige, die het verband juist ziet en het theoretisch formuleert als een successie van toenemende meeropbrengsten, punt van relatief meest gunstige productiviteit, gevolgd door afnemende meeropbrengsten (dus een verloop, wat thans door een logistische curve aangeduid zou worden) (hij maakte van zijn theorie gebruik in zijn belastingwetten).

Steunende op de genoemde klassieke opvatting in den landbouw (wet der afnemende meeropbrengsten) ging men zich daarna meer richten op de opbrengstwetten in de industrie en in het bedrijf. Ook daar zag men afnemende meeropbrengsten, doch zij, Amerikaansche auteurs in het byzonder, gingen de productiewetten in de industrie als een afzonderlijke groep beschouwen, zij onderscheidden meer dan één opbrengstwet en voerden een nieuwe naamgeving in, waardoor groote verwarring ontstond. De wet der afnemende meeropbrengsten werd herdoopt in "Law of Proportion of Factors", de wet der toenemende meeropbrengsten in "Law of Advantage and Size". Carver en Davonport zeggen in hun discussies dat de wet der afnemende meeropbrengsten onderscheiden moet worden van de wet van de productie op groote en kleine schaal. De wet der afnemende meeropbrengsten heeft allereerst betrekking op de verhouding der productiefactoren. Deze wet zegt hoe de productiviteit van een productieeenheid varieert als de productieelementen in verschillende verhouding gecombineerd worden. Deze wet veronderstelt dat één productiefactor constant (gegeven) is, terwijl de andere variabel zijn. Is er nu een bepaalde verhouding der productiefactoren , die het meest gunstige resultaat opleveren, dan zal iedere andere verhouding een minder gunstig resultaat geven, maar uitgaande van deze gunstigste verhouding, constateert men steeds afnemende meeropbrengsten.

De wet van den omvang zou de wet der toenemende meeropbrengsten moeten vervangen en onderscheiden worden van de vorige. Hoe grooter de onderneming hoe economischer, hier zouden zich toenemende meeropbrengsten vertoonen. Op het feit dat bij toenemenden omvang van het bedrijf ook de verhouding der factoren zich zal wijzigen, werd geen acht geslagen. Trouwens het heele nieuwe ingevoerde onderscheid is onjuist en men laat het later weer vervallen. Men legde te zeer den nadruk op het onderscheid tusschen toe- en afnemende meeropbrengsten. Men realiseerde zich niet, dat het 2 phasen van eenzelfde proces zijn. In een statischen toestand kunnen beide voorkomen. Veranderen de omstandigheden, ziet men het proces dynamisch, dan ontstaan in den regel eerst toenemende, later weer afnemende meeropbrengsten. Men krijgt a.h.w. cycli van logistische relaties. De bovengenoemde theorieen waren meestal aangepast aan de omstandigheden van den tijd, waarin de onderzoeker leefde, zoo kwam men in den loop der tijden tot geheel verschillende inzichten, die sterk van elkaar afweken.

Zoo ziet men, dat het beperkte standpunt, van slechts één factor te beschouwen, zich ernstig wreekt in tal van elkaar bestrijdende theorieen. Hiernaast bestond nog de groote tekortkoming, dat men een veel te klein interval van den onderzochten factor beschouwde.

Nadat men deze beperkingen besefte,was het niet moeilijk meer, om door toepassing van deze nieuwe inzichten, i.c. dat door de werking van andere factoren de afnemende meeropbrengsten weer in toenemende meeropbrengsten kunnen overgaan, bv. in de bevolkingsleer, de pessimistische theorie van Malthus te weerleggen.

In den experimenteelen landbouw begon met Liebig de gedachte van toenemende meeropbrengsten, doch nadehand, mede tengevolge van uitgebreide proefnemingen, ging men gelooven aan het bestaan van een wet der afnemende meeropbrengsten in den landbouw, een gedachte, die een voorloopige afsluiting vond in de wetten van Mitcherlich.

In den experimenteelen landbouw kwam men ook gemakkelijker tot uitbreiding van het gebied, waarover de factor beschouwd wordt, en zoo ontdekte men ook een dalenden tak van de opbrengst krommen, hoewel men zich hiermee weinig bezig gehouden heeft. Ook hier was het, zij het veel te bescheiden, dat men meer factoren tegelijk ging beschouwen. In het experiment 2 zelden 3 factoren; In de theorie meer: de formule van Baule-Mitcherlich voor meerdere factoren zullen we nog nader noemen.

In overeenstemming met dit hierboven geschilderd, steeds wisselend aspect, is ook de wiskundige formuleering een strijdvraag gebleven.

Van uit het breedere biologisce-technische gezichtspunt werden de opbrengsttheorieen opnieuw in de economie ingevoerd. In de critiek op de econometrie, die in ruime mate gebruik maakt van vraag-en aanbodcurven, waarin de opbrengstcurven verdisconteerd zijn, wreekt zich nochtans andermaal het te enge standpunt van het beschouwen van slechts één factor. Op die manier kan men slechts zelden een overtuigend resultaat verkrijgen.

Den enkelen keer dat men meerdere factoren tegelijkertijd beschouwt, doet men dat met de methode der lineaire correlatie, een slechte benadering, daar de relaties niet lineair zijn en nog veen minder onafhankelijk (onafhankelijk hier weer in den zin van geen interactie tusschen de factoren) van elkaar. Ook hier moet men zijn abstracties laten leiden door het factorial design, of wel zijn waarnemingen overkoepelen met een factorial design.

De wiskundige formuleering, die een strijdvraag bleef, wordt uitvoerig besproken voor één factor door Pearl [1]), hetwelk hier eerst moge volgen, voor we de formuleering voor meerdere factoren onder oogen zien.

§ 3. Formuleering voor één factor volgens Pearl.

Het is mogelijk een gevonden empirische curve te vereffenen volgens een of andere wiskundige vergelijking, maar een dergelijke wiskundige relatie zegt ons niets over de wetten, die een rol spelen bij het tot stand komen van de empirische curve. De wiskundige vergelijking heeft alleen waarde als een beschrijving van het empirische materiaal. Hetzelfde type dient dan ook voor de beschrijving van zeer verschillende verschijnselen, sommige oogstopbrengstcurven en de monomoleculaire reactie kunnen door dezelfde wiskundige vergelijking weergegeven worden, doch beschrijven zeer verschillende zaken. Evenwel voert het aanpassen van een zoo eenvoudig mogelijke vergelijking aan alle beschikbare gegevens, lang niet altijd tot het beste resultaat.

Dit wordt duidelijk als men bv. met behulp van de gevonden vergelijking zou willen gaan extrapoleeren. De wiskundige relatie blijkt dan vaak ingewikkelder te moeten zijn om hieraan eenigszins tegemoet te kunnen komen. Men controleere dit door, bij tijdreeksen bv., een aantal van de laatste gegevens niet in de vereffeningscalculatie op te nemen en overtuige zich vervolgens van de overeenstemming , tusschen de, met behulp van de gevonden formule geëxtrapoleerde gegevens en de werkelijke (niet gebruikte) gegevens. Voor groeicurven van populaties (Drosophilacultures, bevolking van verschillende landen e.a.) gebruikte Pearl eerst $y = a_0 + a_1 t + a_2 t^2 + a_3 t^3$ (31) (y = aantal individuen, t = de tijd). De entrapolatie hiermede gaf natuurlijk slechte resultaten; toen gebruikte men $y = a_0 + a_1 t + a_2 t^2 + a_3 \log t$ (32), voldeed alleen voor extrapolatie over korte afstanden, doch voor hooge t geeft deze formule ook een zeer hooge waarde voor y, wat in strijd met de werkelijkheid is. Zoo zijn er aantal eischen op te stellen, waaraan de vereffeningsfunctie moet voldoen. Voor populatiegroeicurven noemt Pearl de volgende vijf:
1) de waarde van y is begrensd 2) de onderste grens = 0, negatieve populaties hebben geen practischen zin; 3) daar de groeiruimte begrensd is, moet ook de totale populatie begrensd zijn; elke qualitatieve verbetering staat een steeds kleinere vermeerdering toe, de curve moet dus de bovenste grens asymptetisch naderen;

[1]) R.Pearl. Studies in human biology.
Baltimore 1924. Chapter XXIV.

4) staat de ontwikkeling gedurende eenigen tijd bijna stil, dan nadert de populatie spoedig de bovenste grens, ontstaan plotseling nieuwe mogelijkheden, dan kan de populatie zoch plotseling zeer sterk uitbreiden, waarna het weer langzaam een 2e maximum nadert; bv. de opkomst van de industrie teekent zich in vele bevolkingscurven scherp af. De formule moet dus in staat zijn, dit cyclische karakter te kunnen weergeven, waarbij de opvolgende cycli additief zijn; 5) binnen zoo'n periode of cyclus heeft de populatiecurve een karakteristieken vorm, een langzaam begin, vervolgens neemt de mate van aangroei toe tot een maximum en vermindert vervolgens weer.

Voor andere relaties kan men soortgelijke eischen opstellen, waaraan de vergelijking moet voldoen. Bezien we deze 5 eischen nader, dan is het duidelijk dat geen der beide bovengenoemde vereffeningsfuncties aan deze eischen voldoen.

Vervolgens werd de logistische curve geprobeerd:

$$y = \frac{be^{at}}{1 + ce^{at}} = \frac{b}{c + e^{-at}} \quad ') \quad (33)$$

Deze voldoet aan de eischen 1); 2) en 3) en ook min of meer aan 5), genoemde curve is echter punt symmetrisch, het buigpunt ligt in het middeh; we hebben echter meestal met asymetrische vormen te doen. In geenen deele wordt echter aan punt 4) voldaan.

We kunnen (33) ook schrijven als volgt:

$$y = \frac{k}{1 + me^{-at}} \quad \text{als } k = b/c \text{ en } m = 1/c$$

Nu is $\frac{dy}{dt} = ay(k - y)$ $\quad \frac{\frac{dy}{dt}}{y(k-y)} = a \quad (34)$

dus $\frac{dy}{dt}$ varieert met het varieeren van y en (k - y).

Daar deze veranderingen afhankelijk zijn van factoren, die met den tijd veranderen, kunnen we (34) generaliseeren door a te vervangen door f(t), dan krijgen we:

$$\frac{\frac{dy}{dt}}{y(k - y)} = f(t) \quad \text{of} \quad y = \frac{k}{1 + me^{F(t)}}$$

waarin F(t) nog eens vervangen wordt door de reeks van Taylor:

$$y = \frac{k}{1 + me^{a_1 t + a_2 t^2 + \ldots\ldots a_n t^n}} \quad (35)$$

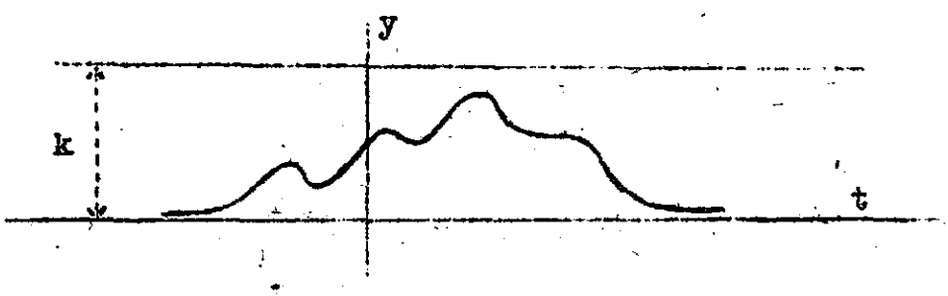
als m negatief is, is de curve discontinu bij eindige waarde van t

($y = \infty$ voor $a_1 t + a_2 t^2 + \ldots a_n t^n = -\log(-m)$)

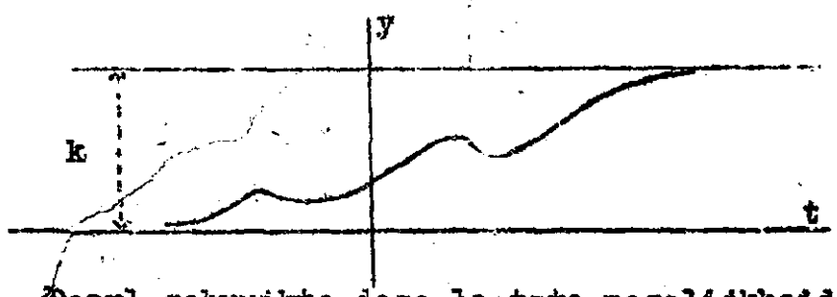
Dit is niet mogelijk bij populatiecurven, dus beperken we ons tot positieve waarden voor y, we beperken ons hier dus eveneens tot positieve waarden van k.

$\frac{dy}{dt} = y(k - y)F^1(t)$ De curve ligt dus tusschen de asymptoten y = 0 en y = k.

Als nu n = even en a_n positief dan heeft de curve den volgenden vorm:



Is n = oneven en a_n negatief, dan is de vorm als volgt:



Pearl gebruikte deze laatste mogelijkheid voor populatiecurven. In hoofdzaak beperkt Pearl zich tot

$$y = d + \frac{k}{1 + me^{a_1 t + a_2 t^2 + a_3 t^3}}$$

waarin d het resultaat is van alle vorige cycli en die zich additief gedragen met betrekking tot de volgende cycli, hetgeen in bovenstaande formule tot uiting komt.

Zoo kan men voor hetzelfde waarnemingsmateriaal verschillende vereffeningsfuncties opstellen, naar gelang van de eischen die men stelt, waarbij men zich dus niet in de eerste plaats laat leiden door de beste aanpassing aan het gegeven aanwezige materiaal.

Voor landbouwkundige opbrengstcurven, die een duidelijk optimum vertoonen kan dus zeer goed de eerste mogelijkheid n = even, a_n = positief dienen.

De eenvoudigste vorm moet dan zijn:

$$y = \frac{k}{1 + me^{a_1 x + a_2 x^2}} \qquad k > 0 \quad m > 0 \quad a_2 > 0$$

Voor zoover schrijver bekend, is hier tot op heden geen gebruik van gemaakt, bovendien is deze formuleering moeilijk uit te breiden voor meer variabelen. (Pearl geeft vervolgens nog een kort rekenschema voor de vereffening met deze formules).

Zoo zien we dus reeds eenige keus voor opbrengstcurven, doch voor ons doel is het noodzakelijk, dat deze opbrengstcurven voor meerdere variabelen ontwikkeld worden.

§ 4. Formuleering voor twee en meer factoren.

De curven van Mitcherlich zijn uitgebreid voor meer dan een factor.

$$y = E(1 - e^{-c_1 x_1})(1 - e^{-c_2 x_2})(1 - e^{-c_3 x_3}).....(1 - e^{-c_n x_n})$$

doch hieraan zijn natuurlijk dezelfde bezwaren verbonden als aan de curve van Mitcherlich voor een enkele factor, terwijl bovendien geen rekening gehouden is met het optreden van interacties.

Wordt x_k groot dan nadert $(1 - e^{-c_k x_k})$ tot 1, m.a.w. intensiveering van een groeifactor, die in overmaat aanwezig is, heeft weinig effect. Is x_k klein dan nadert $(1 - e^{-c_k x_k})$ tot 0, m.a.w. afwezigheid van een factor kan den heelen oogst doen mislukken. Wat deze twee punten betreft is de formule niet onverdienstelijk, bezwaar is echter dat er geen optimum door vertegenwoordigd wordt en verder dat ze slechts enkele mogelijkheden van interactie toestaat, optimum verschuivingen kunnen er niet mee beschreven worden.

Voor de andere types zijn schrijver geen ontwikkelingen voor meerdere variabelen bekend.

Noodgedwongen moeten we ons zoodoende thans trachten te redden met de regressieoppervlakken uit de correlatietheorie.

Het regressieoppervlak uit de theorie der normale correlatie is alweer ongeschikt, door zijn symmetrie en door het onafhankelijk zijn der factoren (onafhankelijk hier weer in den zin van geen interactie).

Volgens Derksen [1]) heeft men bij de niet normale correlatie

---

1.c. pg. 46 § 28.

gewoonlijk de overwegingen, die aan de verschillende types van één dimensionale verdeelingen ten grondslag liggen, tot meerdere veranderlijke trachten uit te breiden.

Men zou de verdeelingskromme type Pearson I zeer goed kunnen gebruiken als opbrengstcurve:

$$y = y(1 + \frac{x_1}{a_1})^{m_1}(1 - \frac{x_2}{a_2})^{m_2} \qquad (36)$$

begrensd gebied aan beide zijden, asymetrisch, terwijl de vereffening volgens de methode der momenten vlot kan verloopen !

Zoo heeft men getracht, als we Derksen even verder vervolgen, in analogie met de differentiaal vergelijking , waaruit de frequente curven van Pearson worden afgeleid, een differentiaal vergelijking op te stellen, die op een dergelijke wijze de basis zou moeten zijn van een stelsel meer dimensionale verdeelingen. Deze pogingen hebben echter volgens D. geen succes gehad (d.w.z. de verdeelingen van bv. Filon en Isserlis komen wel uit de differentiaal vergelijking van Pearson voort, doch de regressie is lineair).

Hetzelfde heeft men getracht met de frequentiecurven van het type van Charlier. Derksen verwijst daartoe naar Pretorius [1]). Bevredigende vormen, die tot meer dan 2 factoren uitgebreid kunnen worden, vinden we daar echter niet.

Onze eenige conclusie moet derhalve voorloopig luiden: dat we voor de opbrengstvergelijking voor meerdere variabelen nog geen vorm hebben, die voldoet, weshalve we in het volgende ons houden aan de regressiefuncties.

Het heeft dan ook geen zin een eventueel geschikte vorm uit voorgaande theorie met behulp van de formules I en II respectievelijk III, IV en V enz. te ontwikkelen.

We zouden dit immers toch doen in de valsche hoop het steeds onbelangrijker worden der interacties van hoogere orde tot resultaat te verkrijgen !

Moch zich dit geval met een of anderen vorm voordoen, welke zekerheid hebben we dan, dat deze vorm ook in niet onderzochte combinaties geldigheid heeft. Het zou slechts onverantwoorde inductie zijn, waarbij de kans bestaat dat elk experiment met nieuwe combinaties de onjuistheid ervan kan aantoonen.

In de extreme, althans niet gewone combinaties ligt juist de mogelijkheid van nieuwe, tot nog toe onbekende typen met nieuwe ontwikkelings kansen.

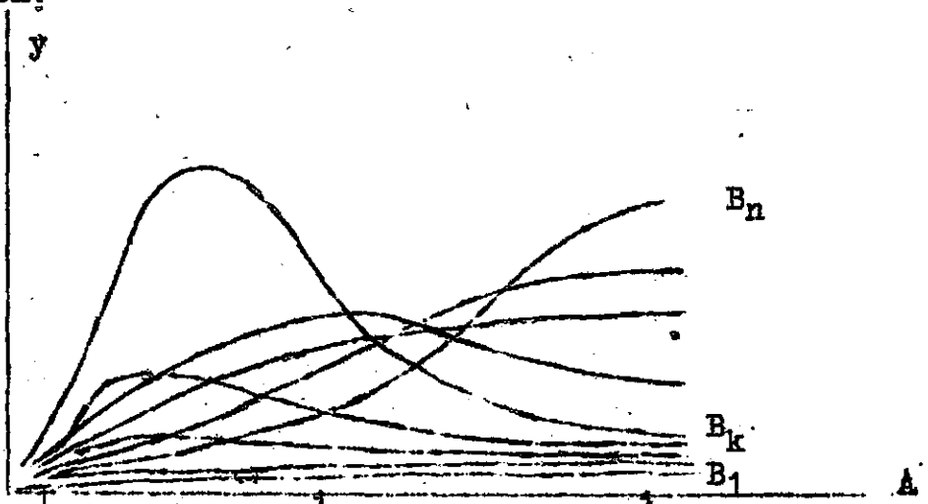
§ 5. Over het optreden van onbekende typen van opbrengstcurven.

Wij zijn nl. nog maar zeer slecht ingelicht over de mogelijke vormen, waarin de opbrengstcurven zich kunnen voordoen !

De reden hiervan ligt in het feit, dat in de natuur niet alle trappen c.q. alle waarden die een factor zou kunnen aannemen, voorkomen ! Sommige wearden zullen vaak, andere zelden of niet voorkomen, anders gezegd het optreden van de verschillende waarden kan worden weergegeven door een frequentieverdeeling.

Onderzoeken we bv. een factor A en een gegeven aantal trappen, loopende over i van o tot j en een 2den factor B, loopende over k van o tot n (dus volgens een factorial design met j n objecten), dan kunnen bv. de verkregen reactiecurven het volgende beeld geven:

fig. 26



[1])S.J.Pretorius. Skew bivariate frequency surfaces. Biometrika, Vol.22, pg.109. 1930.

In de natuur zullen nu sommige van deze trappen van factor B zeer vaak, andere zelden of niet optreden. Is nu bv. de frequentie van B_k zeer groot, terwijl deze naar hoogere en lagere waarden afneemt en zoowel de laagste waarde B_1, B_2, B_3 .... als de hoogste waarden .....B_{n-1}, B_n niet meer voorkomen (zie frequentieverdeeling fig.27), dan zal men alleen reactiecurven van het type van fig.28 waarnemen.

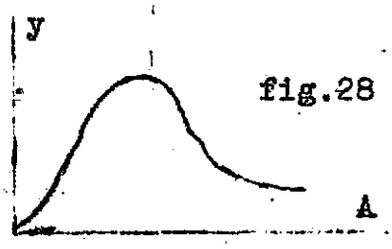


fig.28

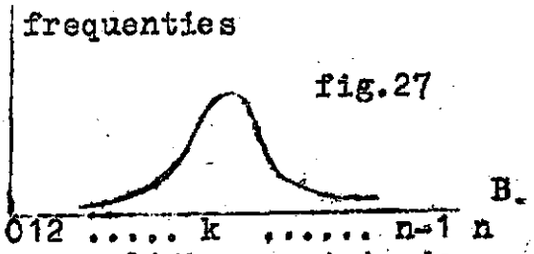


fig.27

Zou men echter in een experiment verwezenlijken, wat in de natuur niet voorkomt nl. de hooge en lage trappen van factor B, dan zou men ook reactiecurven van een ander type ontdekken fig. 29 en 30.

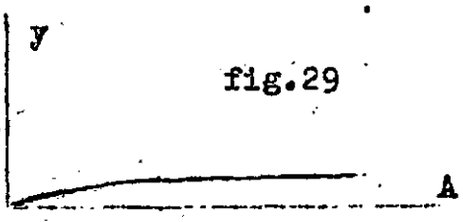


fig.29

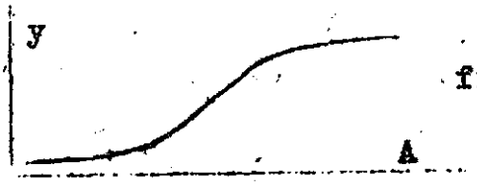


fig.30

Hetzelfde geldt bij 2 factoren B en C. In de natuur zullen niet alle combinaties van B en C, waarbij we den factor A willen onderzoeken, voorkomen, doch hun optreden, hun frequenties worden thans bepaald door een correlatieschema. Alleen in een experiment kunnen we de ontbrekende combinaties verwezenlijken en datgene tot stand brengen, wat in de natuur niet voorkomt, in het byzonder kunnen we dit bereiken in een experiment volgens het factorial design, waarin alle combinaties van B en C voorkomen (en alle even vaak, dus efficient, vergeleken met de waarnemingen uit de natuur, waarin de combinaties in het centrum van de correlatiefiguur zeer frequent voorkomen).

We krijgen dan bv. de volgende situaties:

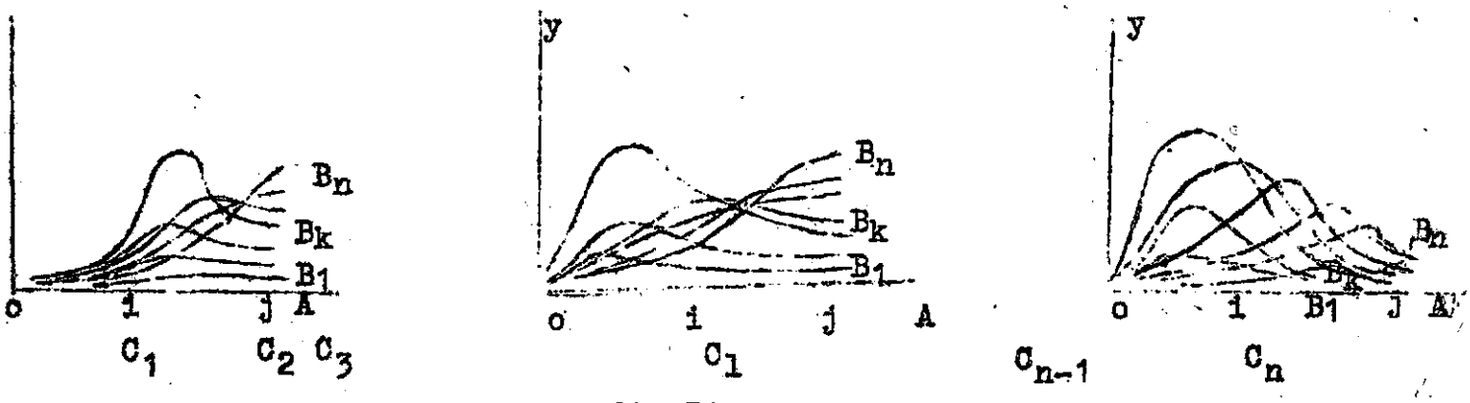


fig.31

(De middelste figuur is in dit voorbeeld dezelfde als fig.26). Ook hier zien we bij extreme waarden van factor C nieuwe types van reactiecurven optreden:

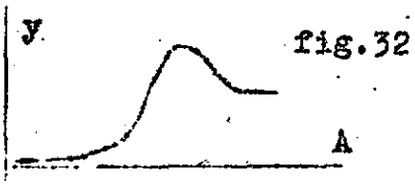


fig.32

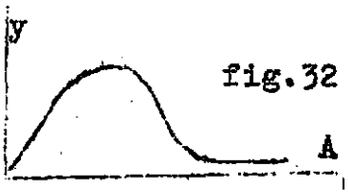


fig.32

waarvan de 1e zich scherp onderscheidt van fig. 28 en 30. de 2e

In feite lijkt deze redeneering dus op ontkenning van het goed recht van de inductieve methode of op scepticisme ten aanzien van de waarden van het experiment.

Dat dit niet zoo is, moge uit de volgende § blijken.

§ 6. De tragiek van het spel der factoren.

De maximale kans, dat bij onderzoek van een factor A, geclassificeerd volgens een 2den factor B, een onbekende reactiecurve, die tot nog toe niet gecontroleerd werd in de natuur, optreedt, wordt gegeven door de kansverdeeling van de waarden, die de 2e factor in de natuur aanneemt.

Het is noodig hierbij van maximale kans te spreken, omdat het zeer goed mogelijk is, dat de aard der interactie zoodanig is, dat geen, wat wij in de voorgaande voorbeelden een ander type noemden, kan optreden.

Voor deze frequentieverdeelingen kunnen we, in aansluiting met genoemde rationeele functies als regressiefunctie, § 1, pg. 93 aannemen de normale verdeeling.

Beperken we ons tot de natuurwetenschappelijke opvatting van het factorbegrip, dan kunnen we bv. de verschillende types van Pearson, die aan een zijde begrensd zijn, gebruiken.

Gemakshalve beperken we ons tot het eerste, de normale verdeeling. De maximale kans wordt dan:

$$p_1 = \frac{h}{\sqrt{\pi}} \int_{-\infty}^{-k} e^{-h^2u^2} du \qquad \frac{h}{\sqrt{\pi}} \int_{+k}^{+\infty} e^{-h^2u^2} du = 1 - \frac{h}{\sqrt{\pi}} \int_{-k}^{+k} e^{-h^2u^2} du \qquad (37)$$

waarin k en de hiermee correspondeerende waarde van u $(u_k = x_k - z)$ een zoodanige waarde heeft, dat de absolute waarde van u zeer zelden grooter zal uitvallen dan u_k;
dus k grooter dan 3σ of 4σ of zelfs 5σ misschien 6σ.

Bij beschouwing van nog een 3e factor C wordt deze maximale kans:

$$p_2 = 1 - \frac{h_1 h_2 \sqrt{1-r^2}}{\pi} \int_{-k_1}^{+k_1} \int_{-k_2}^{+k_2} e^{-(h_1^2 u_1^2 - 2r h_1 h_2 u_1 u_2 + h_2^2 u_2^2)} du_1 du_2$$

De maximale kans wordt dus grooter ! (en des te meer hoe nauwer de correlatie is).

Bij steeds toenemend aantal beschouwde factoren zou dus de maximale kans een nieuw type te vinden volgens het voorgaande steeds grooter worden en wel

$$p_n = 1 - \frac{\sqrt{P}}{(2\pi)^{n/2}} \int_{-k_1}^{+k_1} \int_{-k_2}^{+k_2} \cdots\cdots \int_{-k_n}^{+k_n} e^{-f_n(u_1 \ldots u_n)} du_1 \ldots du_n \qquad (38)$$

waarin $f_n(u_1 \ldots u_n)$ de homogene functie van den 2den graad is

$$\tfrac{1}{2}(p_{11}u_1^2 + 2p_{12}u_1u_2 + \ldots + p_{nn}u_n^2)$$

en $P = \begin{vmatrix} p_{11} \cdots p_{1n} \\ \vdots \\ p_n \cdots p_{nn} \end{vmatrix}$ is.

Echter is dit nog niet juist: slechts een gedeelte der factoren moet die extreme waarden hebben, een ander gedeelte juist de niet-extreme.

De kans op een niet-extreme waarde is:

$$q_m = \frac{\sqrt{Q}}{(2\pi)^{m/2}} \int_{-k_1}^{+k_1} \int_{-k_2}^{+k_2} \cdots\cdots \int_{-k_m}^{+k_m} e^{-f_m(u_1 \ldots u_m)} du_1 \ldots du_m \qquad (39)$$

waarin $f_m = \frac{1}{2}(q_{11}u_1^2 + 2q_{12}u_1u_2 + \ldots q_{mm}u_m^2)$ en $Q = \begin{vmatrix} q_{11} \cdots q_{1m} \\ \vdots \\ q_{m1} \cdots q_{mm} \end{vmatrix}$ (39)

De maximale kans wordt zoodoende bij nxm factoren een functie van p_n en q_m bv. $F(p_n, q_m)$ (40)

Over de waarde hiervan tasten we totaal in het duister, ze is onbepaalbaar, daar ons de verhouding tusschen n en m ten eenenmale onbekend is, doch ze wordt, indien m grooter wordt, in ieder geval zeer klein !
(Indien er geen correlatie bestaat tusschen een of meer der factoren uit de groep n en uit de groep m mogen we voor de maximale kans schrijven $p_n \times q_m$, waarmee het bovenstrande gemakkelijk is voor te stellen, daar p_n tot 0 nadert en q_m tot 1)

Van de kans $F(p_n , q_m)$ kan men slechts een schatting maken, nadat de proef genomen is, dan is pas bekend welke factoren tot de n groep en welke factoren tot de m groep behooren. A priori kunnen we deze kans dus niet schatten, noch benaderen langs dezen weg: ook zoo moet dus het bestaansrecht van een "omgekeerde waarschijnlijkheid" bestreden worden.

<u>In dit zeer klein worden van de maximale kans ligt juist de reden, dat we nooit een afwijkend type vinden, door telkens één factor te onderzoeken volgens ceteris paribus principe ! Algemeen gesproken is deze kleine kans de reden van het bestaansrecht van de inductieve methode in de natuurwetenschappen ! De mogelijkheid van het bestaan van de z.g. natuurwetten is eveneens hierin gelegen !</u>

Omgekeerd ! we moeten juist deze, in het voorgaande genoemde correlaties de natuur doorbreken en wel door extreme waarden te combineeren, of liever gezegd, door extreme combinaties te vormen, daar hier ook niet extreme waarden in voor moeten komen, en wel met behulp van het factorial design, waarbij de kans om iets byzonders te vinden in zekeren zin juist in tegengestelde richting verandert met de hierboven aangegeven kansen, m.a.w. hoe meer factoren we beschouwen, we combineeren, hoe grooter de kans iets nieuws te vinden. Hier is dus iets nader aangegeven, welke weg tot nieuwe mogelijkheden en vooruitgang van sommige (productie)processen kan leiden.

Hoe meer factoren we in het factorial design opnemen, hoe grooter de kans, zagen we ! het tragische, als we het zoo mogen zeggen, ligt nu in het feit, dat deze steeds grooter wordende kans, steeds moeilijker te verwezenlijken valt, immers bij toenemend aantal factoren wordt het aantal objecten van het factorial design zoo groot, dat onze hulpmiddelen ontoereikend worden !

Bij landbouwkundige proeven zagen we reeds in hoofdstuk IV, dat bij slechts ± 20 factoren het aardoppervlak te klein wordt. En we kunnen niet volstaan met slechts een deel der objecten te kiezen, omdat we niet weten, welke factoren extreem en welke niet extreem gekozen moeten worden.

Willen we slechts dé natuur leeren kennen zooals zij zich aan ons voordoet: het resultaat van vele factoren gegroepeerd volgens zekere frequentie verdeelingen, dan is inderdaad de inductieve methode de weg tot een algemeene kennis van deze stoffelijke natuur of, zooals Fisher ') zegt: "inductie is het eenig bekende proces ons bekend, waarmee wij nieuwe kennis kunnen verkrijgen". '').

Willen wij daarentegen verschijnselen tot stand brengen en leeren kennen, die niet in de natuur voorkomen, zoo moeten wij, zooals reeds eerder gezegd, deze correlaties doorbreken met behulp van het factorial design.

In Hoofdstuk I, § 1 werd reeds een onderscheid gemaakt tusschen een oudere Europeesche en een nieuwere Amerikaansche philosophie. Zooals we zagen knoopt Krzymowski bij het Europeesche standpunt aan, terwijl ook in hoofdstuk II, § 2, waar eenige kentheoretische vraagstukken ter sprake kwamen, van dit standpunt uitgegaan wordt. Het andere standpunt treedt juist in voorgaande speculatieve beschouwingen op den voorgrond en is ontsproten uit het radicaal empirisme van James, die in tegenstelling met de

') R.A.Fisher. The logic of the Laboratory.
'') Zoo moet ook Fisher's Null-hypothese" opgevat worden als een omschrijving van de inductieve methode: De Null-hypothese kan nooit worden bewezen, alleen misschien weerlegd; elk experiment wordt slechts genomen om als het ware de feiten een kans te geven de Null-hypothese te vernietigen

klassieke opvattingen den nadruk legt op de byzondere ervaringen, of deze nu den regel bevestigen of niet, verificatie van min of meer waarschijnlijk geachte buitenwetmatige vermoedens zoekt, het wil gids zijn voor nieuwe waarheden en daarbij pluralistisch zijn, waarbij aansluit de instrumenteele logica van Dewey.

Heeft aldus het factorial design, een conceptie uit het landbouwkundig onderzoek voortgesproten, dienst gedaan in bovenstaande speculatieve beschouwingen, het kan daarentegen ook met vrucht aangewend worden in betrekking tot het andere standpunt en wel in een poging om het natuurwetenschappelijk kenvermogen te definieeren, zooals in de volgende § behandeld zal worden.

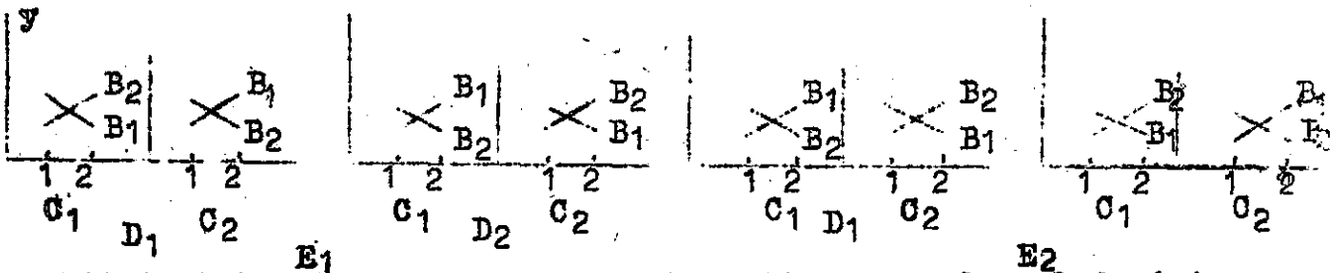
HOOFDSTUK VII.

HET PANPARALLELLE EXPERIMENT.

§ 1. Over de grenzen van het natuurwetenschappelijk kenvermogen.

Zooals we reeds zagen stelt de uitvoering van het factorial design met zeer veel factoren ons voor onoverkomenlijke moeilijkheden. De eenige mogelijkheid om met zeer veel factoren te werken, hetgeen nog op heel andere wijze noodzakelijk bleek in het voorgaande hoofdstuk, ligt in het verontachtzamen van interacties van hoogere orde, en, zooals we in Hoofdstuk IV zagen, in het vervangen van het factorial design door een combinatie van sub-factorial designs.

Daartoe moest echter het gedrag van deze interacties van hoogere orde onderzocht worden, hetgeen inmiddels in hoofdstuk V geschied is. De conclusie was, dat dit gedrag afhankelijk is van den aard der regressiefunctie die het factorenspel beschrijft en van het interval waarin die functie onderzocht wordt. Het vermoeden van het onbelangrijk worden der interacties van hoogere orde kon niet bevestigd worden. In het voorbeeld met $y = hA^2B^2 + i$ bleek de interactievariance bijna altijd grooter te zijn dan de variance voor een der factoren. Verder zijn er gevallen te construeeren waarbij de variance van de interactie van de hoogste orde gelijk is aan de totale variance; dit vindt plaats, als wij het geïdealiseerde geval no.8 uitbreiden voor meerdere factoren. Voegen we nog 2 factoren D en E toe, dan vinden we graphisch voorgesteld het volgende beeld:



Dit is tot het oneindige voort te zetten, steeds zal de interactie van de hoogste orde aanwezig zijn! Afbreken van het onderzoek naar een dergelijk factorenspel, door af te breken bij een interactie van bepaalde orde, geeft een onvolledige beschrijving der werkelijkheid.

Is de laatste reële interactie van een hoogere orde dan de hoogst bereikbare orde, die ons de uitvoering van een factorial design toestaat, dan zijn we aan een van de grenzen gekomen van ons natuurwetenschappelijk kenvermogen.

Wanneer we die grens bereiken kunnen we echter nooit weten: Nemen we bovenstaand voorbeeld: beschouwen we een factor minder, dan wordt eenvoudig de interactievariance van de orde (n-1) gelijk aan de totale variance. Omgekeerd, elk experiment met een factor meer, waarbij de orde der interactie vam de hoogste orde met een vermeerderd wordt, kan toch het verkregen beeld wijzigen; de grens van ons natuurwetenschappelijk kenvermogen is aldus gegeven door de interactie van de hoogste orde die we controleeren kunnen, wat weer door de uitvoerbaarheid van het factorial design beheerscht wordt.

De verantwoording van een aldus wiskundig gedefinieerde grens van het natuurwetenschappelijk kenvermogen zoeke men in Hoofdstuk II.

Maar we kunnen ons afvragen: komen dergelijke gevallen als het genoemde voorbeeld voor in de natuur? Een vraag die we, eenigszins anders geformuleerd, ook reeds in het vorige hoofdstuk stelden. De conclusie was: gaan we uit van een speciale regressiefunctie, dan is het gedrag der variances na te gaan (volgens de methode beschreven in Hoofdstuk V), De veralgemeening van deze regressiefuncties, door inductie, heeft in dit geval alleen dan zin, als we niet buiten het door de natuur gegevene gaan. Het is dan duidelijk in verband met het voorgaande hoofdstuk, dat we dan nooit de natuurwetenschappelijke kennis hooger kunnen opvoeren. We moeten het factorial design nl. van twee kanten bezien: in de diepte en in de breedte. In de diepte wil zeggen dat de interacties van steeds hoogere orde ons in het bij-

zonder interesseeren; in de breedte wil zeggen dat, zijn de interacties van zekere orde en hooger alle = 0, we door steeds meer factoren op te nemen, het factorial design in de breedte uitbouwen met behulp van sub-factorial designs. In ons geval nu zijn de interacties van hoogere orde niet a priori = 0, maar we vinden ze = 0 door de correlatie pg. 99 e.v. in de natuur, zooals uit voorgaand hoofdstuk bleek. Juist daardoor is onze kennis door van de, door de natuur gegeven waarnemingen uit te gaan, begrensd in de diepte, we kunnen het alleen in de breedte uitbreiden.

In landbouwkundige proeven waar 3 factoren herhaaldelijk, 4 factoren [1]) zelden zijn onderzocht, heeft men tot nog toe geen reeele interactie hooger dan van de 2e orde gevonden!

Verontachtzaming van interacties van de 3e orde en hooger stelt ons dus in staat, toch met zeer vele factoren te werken, immers behoeven niet meer dan 3 factoren gelijktijdig onderzocht te worden zoodat bij n factoren het totaal aantal objecten in de combinatie van sub-factorial designs C (n) x 5 wordt, als we elke factor in 5 trappen onderzoeken. Dit wordt voor 10 factoren $\frac{10!}{3!7!} \times 5 = 120 \times 125 = 15.000$ objecten, terwijl het volledig factorial design $5^{10} = 9.765.600$ objecten zou eischen. Deze methode komt dus ten volle tegemoet aan de eischen, om het spel der factoren zooals zich dat in de natuur aan ons voordoet, te beschrijven en te beheerschen.

Zouden we in dit geval ook nog de interacties van de 2e orde verwaarloozen dan wordt het aantal objecten slechts 120 x 25 = 3000. En reeds dan zijn de resultaten veel bruikbaarder, dan bij de huidige methode van elken factor op zich zelf te beschouwen volgens ceteris paribus.

Willen we we wel "buiten de natuur" gaan, in den zin als in het vorige hoofdstuk besproken, dan komt deze combinatie van sub-factorial designs bij zeer veel factoren, daaraan slechts gedeeltelijk tegemoet. Volledig aan tegemoet komen doet alleen het volledige factorial design bij zeer veel factoren; zie verder § 5 in hoeverre de combinatie van sub-factorial designs er wel toe meewerken kan. Uit deze § moge dus nogmaals volgen, dat er alle aanleiding bestaat deze combinatie voor sub-factorial designs uitvoerig te bespreken.

§ 2. Het incomplete factorial design en het incomplete complex experiment.

Het resultaat van een experiment met bv. 5 factoren en elken factor in 5 trappen volgens het factorial design (3125 objecten) kan beschreven worden met de volgende φ-regressiefunctie:

$$
\begin{aligned}
y = {} & \varphi(A) + \varphi(B) + \varphi(C) + \varphi(D) + \varphi(E) + \\
& + \varphi(AB) + \varphi(AC) + \varphi(AD) + \varphi(AE) + \varphi(BC) + \\
& + \varphi(BD) + \varphi(BE) + \varphi(CD) + \varphi(CE) + \varphi(DE) + \\
& + \varphi(ABC) + \varphi(ABD) + \varphi(ABE) + \varphi(BCD) + \varphi(BCE) \\
& + \varphi(CDE) + \varphi(ACD) + \varphi(ADE) + \varphi(BDE) + \varphi(ACE) \\
& + \varphi(ABCD) + \varphi(ABCE) + \varphi(ABDE) + \varphi(ACDE) + \varphi(BCDE) + \\
& + \varphi(ABCDE).
\end{aligned}
$$

Veronderstellen we thans, dat de interacties van de 2e orde en hooger alle = 0 zijn, dan kunnen de resultaten beschreven worden in:

$$
\begin{aligned}
y = {} & \varphi(A) + \varphi(B) + \varphi(C) + \varphi(D) + \varphi(E) + \\
& + \varphi(AB) + \varphi(AC) + \varphi(AD) + \varphi(AE) + \varphi(BC) + \\
& + \varphi(BD) + \varphi(BE) + \varphi(CD) + \varphi(CE) + \varphi(DE) .
\end{aligned}
$$

terwijl de uitvoering van het experiment geschieden kan, met behulp van 10 sub-factorial designs, elk bestaande uit een 5 x 5 design, dus met 10 x 25 = 250 objecten (in plaats van 3125).

Een en ander berust toch geheel op de formules (I) en (II), respectievelijk (III), (IV) en (V) enz:
alleen als de interactie van de hoogste orde (hier 4e orde) = 0 is, kan $\varphi(ABCDE) = 0$ zijn; en eveneens $\varphi(ABCD)$, $\varphi(ABCE)$ enz. = 0, als de interacties van de 3e orde = 0 zijn, naast het = 0 van de interactie van de 4e orde enz. enz.

---

[1]) bv. de veldproeven met katoen in de Soedan: 3 varieteiten, 3 plantafstanden, 2 irrigaties, 3 bemestingen, dus 4 factoren en wel in 4 herhalingen, zoodat de interacties nauwkeurig getest konden worden.

Nu behoeven slechts alle combinaties van 2 factoren onderzocht te worden A en B, A en C, A en D enz.; de interactie AB wordt verder niet beinvloed door de andere factoren, immers de interacties van de 2e orde en hooger zijn = 0. Factor A kan alleen beinvloed worden door de factoren B, C, D en E, welnu dit wordt ook in de 4 sub-factorial designs onderozcht. Evenzoo de andere factoren.

Over de naamgeving nog het volgende:

Fisher e.s. noemen alle proeven genomen volgens het factorial design als proefschema "complex experiments". Splitsen we een factorial design in kleinere factorial designs, dan noemen we deze laatste sub-factorial designs. bv. Een A x B x C x D experiment kan, als we de interactie van de hoogste orde, dus van de 3e orde = 0 stellen, gesplitst worden in 4 sub-factorial designs nl. A x B x C, A x B x D, A x C x D en B x C x D of in 6 sub-factorial designs als we naast de interactie van de 3e orde, ook de interacties van de 2e orde = 0 stellen nl. A x B, A x C, A x D, B x C, B x D en C x D. Het complex experiment wordt zoodoende gesplitst in 4 of 6 sub-complex experiments. De verzameling van alle sub-complex experiments noemen we het "incomplete complex experiment" of ook wel pan-parallelexperiment. Evenzoo spreken we van het incomplete factorial design.

Feitelijk kan zoowel elk factorial design opgevat worden als een sub-factorial design van een hypothetisch incomplete factorial design, dan wel als een gedeelte van een hypothetisch grooter factorial design waarop we echter hier niet in zullen gaan.

In het algemeen zullen we hier alleen gebruik maken van de naam incomplete complex experiment, vooral in tegenstelling . . de complex-experiments, die hier meestal complete complex experiments genoemd zullen worden, om in de redeneering de tegenstelling goed te doen uitkomen. In werkelijkheid is het toevoegsel complete overbodig.

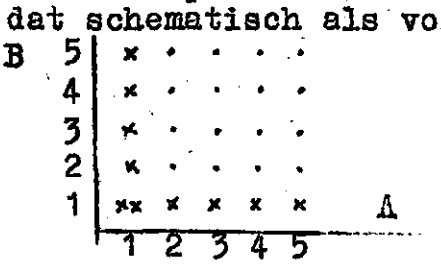
Evenzoo zal vaak gesproken worden van complete factorial design in tegenstelling tot het incomplete factorial design.

Werden de complete complex experiments onderscheiden door het aantal factoren en het aantal trappen van elken factor (men spreekt van 3x3x4 of van 3x5x5x6 complex experiments), zoo kan men de incomplete complex experiments evenzoo onderscheiden met toevoeging van een combinatiefactor (binomiaal coefficient), die het aantal sub-complex experiments geeft.

§ 3. Uitvoering van een incomplete complex experiment met 3 factoren.

Ter nadere uiteenzetting beginnen we met het eenvoudigste geval, waarbij

a) 2 factoren A en B, waarbij we a priori weten dat er geen interactie is. We kunnen dan volstaan met het onderzoeken van factor A bij slechts één trap van B en het onderzoeken van factor B bij slechts één trap van A. Geen interactie beteekent immers: de reactie op factor A is onafhankelijk van factor B, of de reactie is bij elken trap van factor B dezelfde. Onderzoeken we elken factor in bv. 5 trappen, factor A bij den 1sten trap van B en factor B bij den 1sten trap van A, dan kunnen we dat schematisch als volgt voorstellen:



schema 1

fig.34

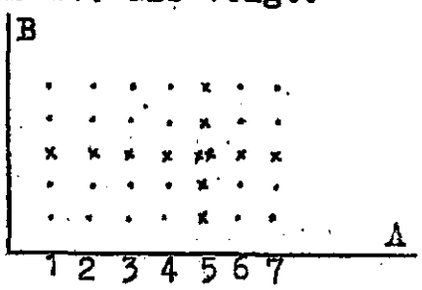
In plaats van de factoren in een volledig factorial design te onderzoeken bestaande uit 5x5 = 25 objecten, volstaan we hier met slechts een deel i.c. 10 objecten te onderzoeken. Een object n.l. A_1B_1, is hierbij zooals we zien in 2-voud. Door deze in enkelvoud te nemen, zouden we het aantal objecten nog tot 9 kunnen reduceeren, doch dit verdient geen aanbeveling, le kan dit dienen voor een fouten bepaling en 2e voor vereffening. Zouden we n.l. beide series van vijf objecten in afzonderlijke proeven onderzoeken, dan is het mogelijk, dat de omstandigheden, onder welke beide series onderzocht worden niet geheel gelijk zijn. In het laboratorium kan dit in hooge mate voorkomen worden, in den landbouw (veldproeven)

in den regel niet, daar heeft men altijd te doen met een niet homogenen achtergrond van het experiment. Het object in duplo geeft nu de gelegenheid om hier eenige correctie aan te brengen. Tot hun volle recht komen beide genoemde punten pas in experimenten met meer factoren, waarin ook interacties van lagere orde onderzocht worden; daar treden dan veel meer objecten in duplo op, naast objecten in 3-voud, 4-voud enz., terwijl daar toch ook hun totale aantal gering blijft, in vergelijking met het gereduceerde aantal. Hierover later meer.

Nu is het een bekend feit (vooral bij landbouwkundige proeven) dat de bepaling van de lagere trappen onzekerder resultaten geven dan de andere, het verdient dus geen aanbeveling factor A bij den laagsten trap van B en factor B bij den laagsten trap van A te onderzoeken. Het beste schijnt te zijn in de buurt van het optimum, daar we ons daar het meest voor interesseeren. Deze optima willen we echter juist vinden in het experiment. Daar we echter meestal het werkingstraject van de factoren min of meer kennen uit vorige of voorloopige proeven, nemen we den middelsten of nog iets hoogeren trap, ons de algemeene gedaante van een optimum-curve voor de oogen houdend, waarbij het optimum meestal iets rechts in het beschouwde gebied ligt. Schematisch dus bv. als volgt:

schema 2

fig.35



Om nu de waarde van de niet onderzochte objecten te vinden, kunnen we de gevonden waarden van de wel onderzochte objecten vereffenen met de regressievergelijking $y = \varphi(A) + \varphi(B)$. Den graad van $\varphi(A)$ en $\varphi(B)$ stellen we vast, na beschouwing van de graphische voorstelling der verkregen resultaten.

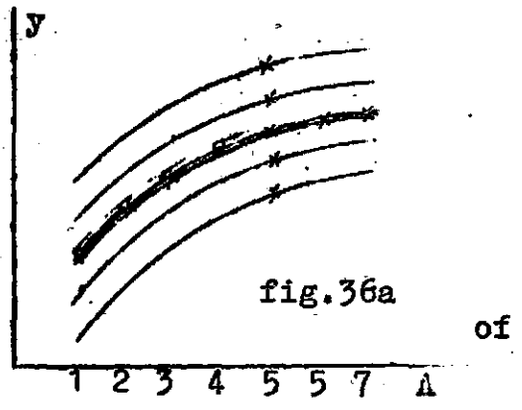


fig.36a

of

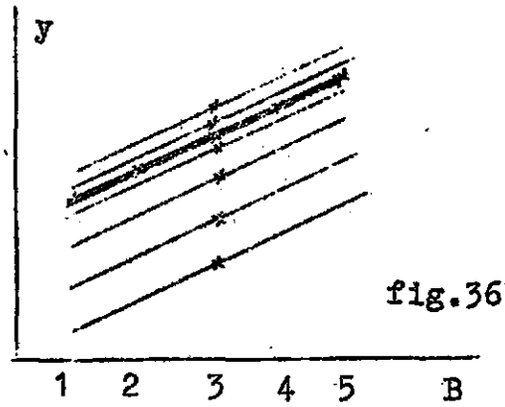


fig.36b

Beide figuren stellen weer hetzelfde geval voor. Evenals in den aanvang van hoofdstuk III duidelijkheidshalve telkens meerdere figuren voor hetzelfde geval werden gegeven, zal dit ook hier worden gedaan. Uit fig. 36a is dadelijk af te lezen dat de reactie op factor B lineair is, daar de kromme lijnen daar alle op even grooten afstand van elkaar liggen, doch bij ingewikkelder reactie krommen en vooral bij meerdere factoren wordt dit zeer moeilijk. De dun getrokken lijnen stellen de reactiekrommen voor, die we gevonden zouden hebben, als we alle objecten van het factorial design onderzocht hadden.
De dikgetrokken lijnen, die reactiekrommen, die we met ons gereduceerd proefschema gevonden hebben. Op grond hiervan kunnen we als vereffeningsfunctie kiezen $y = a_0 + a_1 A + a_2 A^2 + b_0 + b_1 B = a_1 A + a_2 A^2 + b_1 B + c$ (in dit eenvoudige geval van 2 factoren zonder interactie kan ook zeer goed graphisch vereffend worden).

Hiermede kunnen dan de waarden, die bij andere trappencombinaties behooren, gevonden worden, of eigenlijk geinterpoleerd worden, mits deze trappen niet liggen buiten het gebied, dat door het volledig factorial design aangegeven wordt. (wil men dit toch doen, dus extrapoleeren, dan gelden dezelfde restricties, die ook gesteld worden, als men het volledig factorial design uitgevoerd had.).

De nauwkeurigheid kunnen we beschrijven met de error of estimation (Ezekiel). Het quadraat hiervan is de som der quadraten van de residuen (fouten), gedeeld door het aantal vrijheidsgraden, in dit geval dus het aantal waarnemingen minus het aantal constantem van de regressiefunctie, hier dus 7 + 5 - 4. Zouden we in het geval, dat we met een niet-homogenen achtergrond van het experiment te doen hebben, het object in duplo ook nog gebruiken om beide serie met elkaar in overeenstemming te brengen, dan zouden we het aantal vrijheidsgraden ook hierop nog moeten corrigeeren.

b) We weten a priori dat er een geringe interactie aanwezig is. Passen we dan toch voorgaande schema's toe, m.a.w. nemen we aan, dat er geen interactie is, dan maken we dus een benadering, hetgeen in een grootere error of estimation tot uiting komt.

Om deze benadering zoo goed mogelijk te doen zijn, is het niet onverschillig welk schema we gebruiken.

Nemen we het volgende voorbeeld, graphisch aldus voorgesteld:

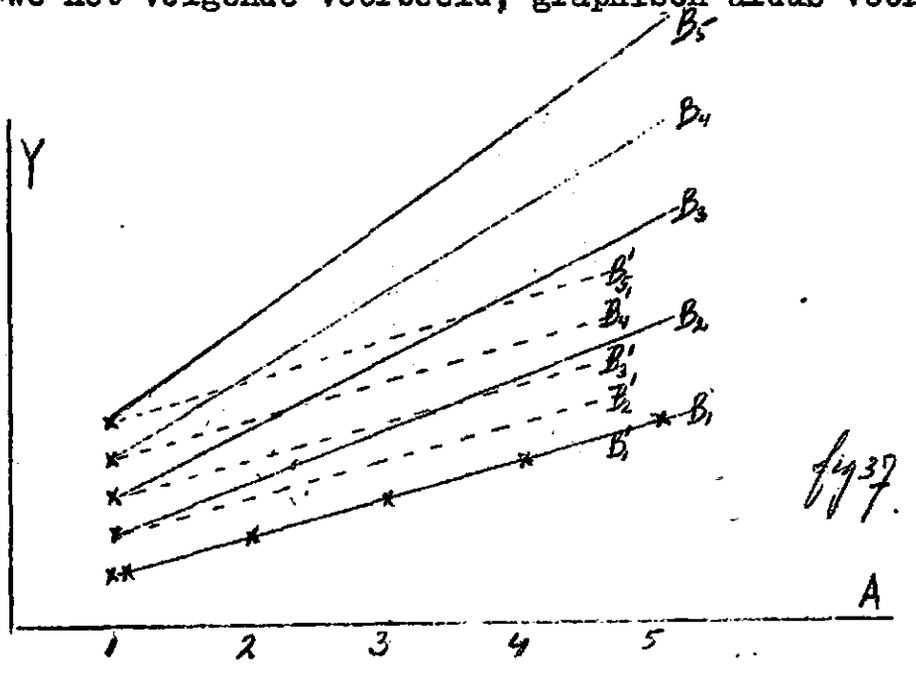


fig 37.

Passen we het eerst besproken schema toe: dus onderzoek van factor A bij den laagsten trap van B, onderzoek van factor B bij den laagsten trap van A; in de figuur is dit weer door kruisjes aangegeven. Vereffening als besproken in het voorgaande, doet, zooals direct kan blijken, door graphische vereffening de gestippelde lijnen (reactie - curven) ontstaan. Dus zeer aanzienlijke afwijkingen, de grootste voor het object A_5B_5. Hadden we volgens het 2e schema gewerkt, dan komen we tot de volgende figuur:

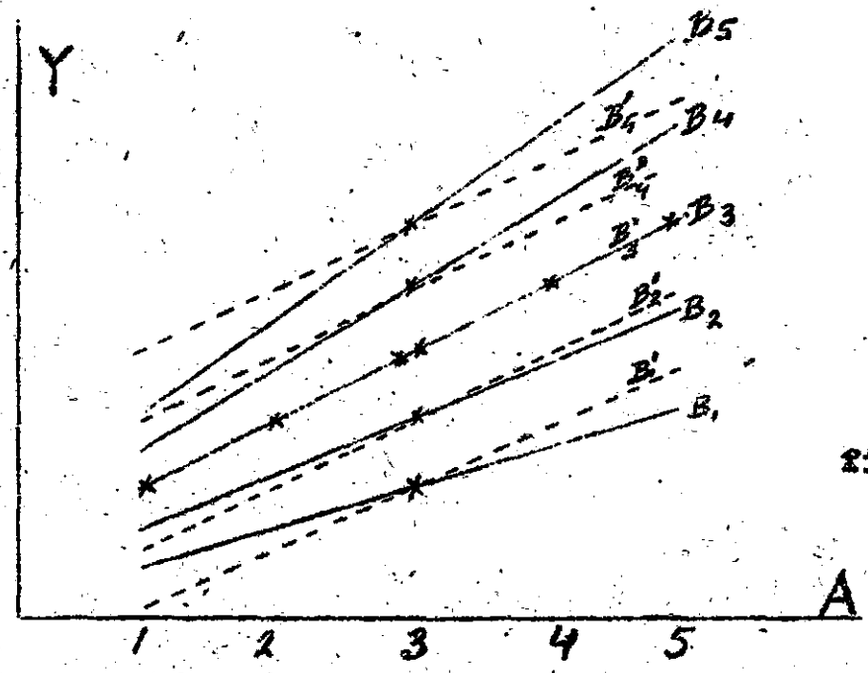


fig. 38

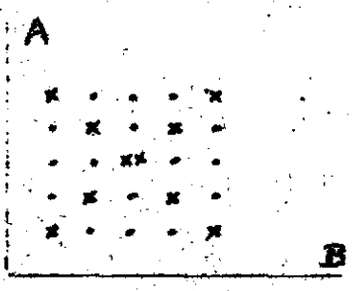
Thans zijn de afwijkingen veel geringer en bovendien gelijkmatiger verdeeld.

De gelijkmatigheid zou nog verder op te voeren zijn, door nog een ander, een 3e schema te kiezen.

Bovendien bestaat er tegen beide genoemde schema's nog een klein bezwaar, dat tot nog toe niet besproken is.

In het volledig factorial design van bv. 5 x 5 komt elke trap van de eene factor even vaak (n.l. 5 x) voor bij elken trap van den anderen factor z.g. verborgen herhalingen (hidden replication). In ons schema komt een trap van den eenen factor 5 x voor, de andere trappen slechts 1 x en omgekeerd, dus zeer onregelmatig. We kunnen dit opheffen door de objecten diagonaalsgewijze uit te zoeken voor ons gereduceerd schema, schematisch als volgt voorgesteld:

3e schema
fig.39



Thans hebben we voor elken trap van beide factoren 2 verborgen herhalingen.

Het is duidelijk dat bij een zoo regelmatige verdeeling van de onderzochte objecten over het geheele te onderzoeken gebied ook de afwijkingen en dus de error of estimation zoo laag mogelijk wordt.

Het is mogelijk deze regelmatige verdeeling bij 2 factoren ook uit te breiden voor meer factoren.

Bij 2 factoren zijn de indices der objecten A B gegeven door het volgende schema:

A	B	A	B	
1	1	1	5	
2	2	2	4	totaal 10 objecten
3	3	3	3	1 object in 2-voud
4	4	4	2	
5	5	5	1	

Bij 3 factoren wordt dit:

AB	C	AB	C	AB	C	AB	C	
11	1	11	5	15	1	15	5	
22	2	22	4	24	2	24	4	totaal 20 objecten
33	3	33	3	33	3	33	3	1 object in 4-voud
44	4	44	2	42	4	42	2	
55	5	55	1	51	5	51	1	

Bij 4 factoren zijn de indices:

ABC	D	ABC	D	ABC	D	ABC	D	ABC	D	
111	1	111	5	115	1	115	5	151	1	
222	2	222	4	224	2	224	4	242	2	
333	3	333	3	333	3	333	3	333	3	enz.
444	4	444	2	442	4	442	2	424	4	totaal 40 objecten
555	5	555	1	551	5	551	1	515	5	1 object in 8-voud

Bij 3 factoren kunnen we dit schematisch in een ruimtefiguur voorstellen.

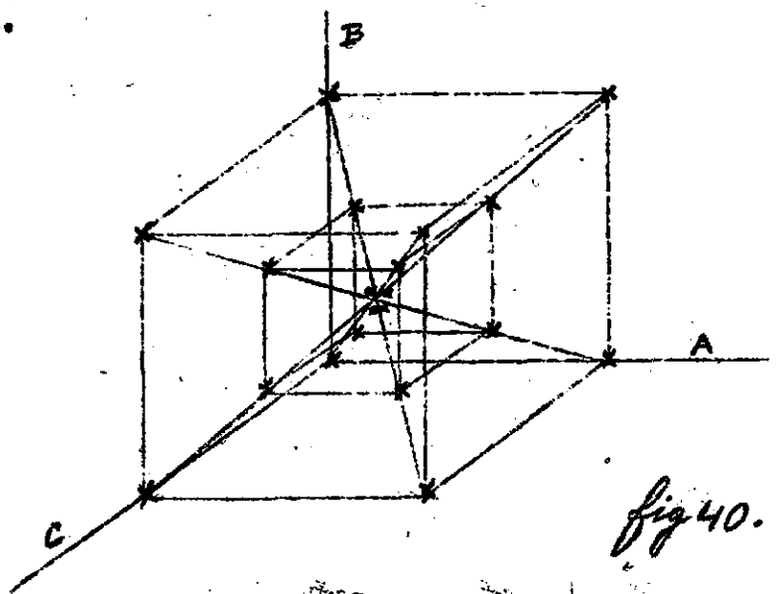


fig 40.

totaal, volgens het volledig factorial design 5 = 125 objecten; we zullen niet alle 125 door een punt voorstellen, doch de te onderzoeken objecten zijn weer door een x aangegeven. We kunnen het ook in een platte figuur weergeven:

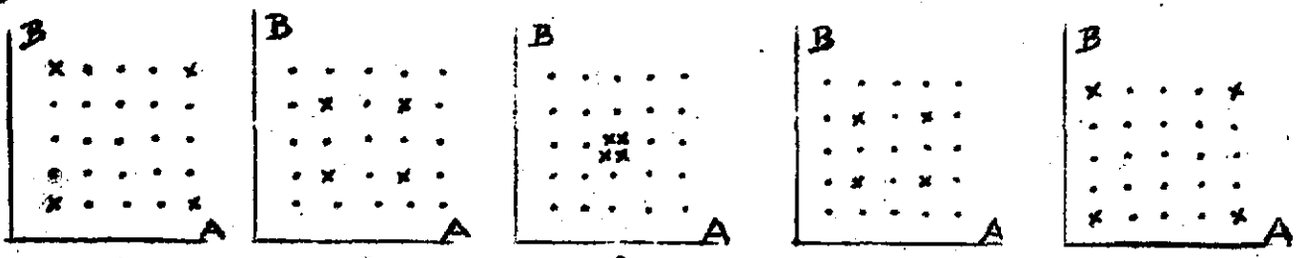


fig.41

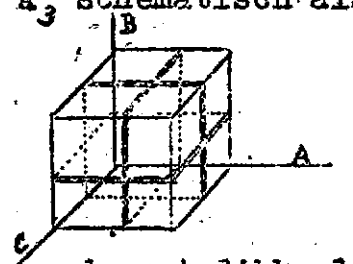
tot welke methode van weergave we ons ook moeten wenden bij 4 factoren, of door 5 cubussen of door 25 platte figuren.

Hoe verleidelijk het 3e homogene schema ook moge lijken, het heeft een aantal bezwaren:
1e. de factoren moeten hierbij alle in een gelijk aantal trappen onderzocht worden; 2e. onderzoekt men ze alle bij een even aantal trappen, dan vallen de herhalingen weg, bepaling van een middelbare fout en correctie op een heterogenen achtergrond van het experiment kunnen niet uitgevoerd worden; 3e. bij meerdere factoren worden ook interacties onderzocht, het volledig factorial design valt daarbij uiteen in een groot aantal kleinere factorial designs, waarop, zoals we nog zullen zien, een analysis of variance toegepast wordt, hetwelk een orthogonaal proefplan vereischt.
c) Is de interactie niet = 0 of bijna = 0, dus groot, dan is de gevolgde methode niet toe te passen, dan moeten de factoren, volgens een volledig factorial design onderzocht worden.

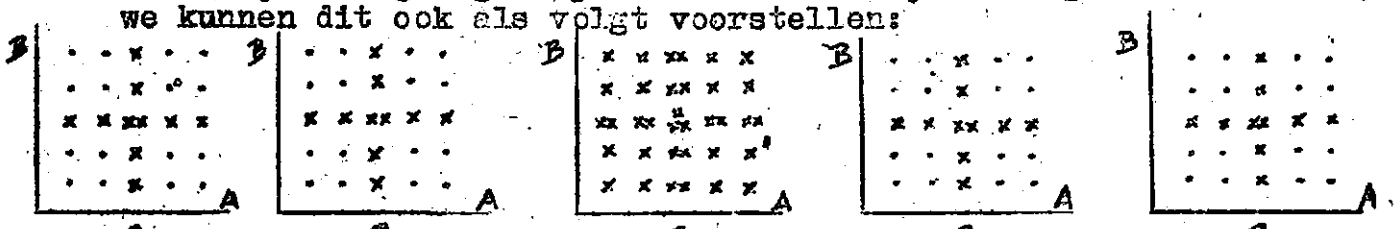
Willen we het voorgaande toepassen bij meer dan 2 factoren, dan moeten ook alle interacties van hoogere orde = 0 zijn. We krijgen daarbij één object in evenveel-voud als het aantal onderzochte factoren bedraagt.

In den regel zullen echter de interacties van lagere orde niet = 0 of bijna = 0 zijn. Het hierboven staande zullen we dan ook nooit toepassen, het dient alleen als inleiding voor de thans volgende bespreking met veel factoren en met te verwaarloozen interacties van hoogere orde. Laten we weer met het eenvoudigste voorbeeld beginnen n.l. 3 factoren A, B en C, hierbij treden 3 interacties van de 1e orde en 1 interactie van de 2e orde op. We willen de interacties van de 1e orde in ons onderzoek opnemen, doch nemen thans aan, dat de interactie van de 2e orde = 0 of bijna = 0. We moeten dan de factoren A en B volgens een volledig factorial design onderzoeken; factor C heeft op de interactie AB geen invloed, het doet er dus niet toe, bij welken trap van C we deze interactie onderzoeken; we nemen weer den middelsten of een iets hoogeren trap om dezelfde redenen als vermeld bij schema 2 hierboven. We nemen dus de volgende objecten in ons gereduceerd proefplan op: bv. bij een 5 x 5 x 5 experiment; alle 25 combinaties van de factoren A en B bij trap C_3, alle 25 combinaties van de factoren A en C bij trap B_3 en alle 25 combinaties van de factoren B en C bij trap A_3 schematisch als volgt voorgesteld:

fig.42



alle objecten gelegen op de met dikke lijnen aangeduide vlakken, we kunnen dit ook als volgt voorstellen:



C_1 C_2 fig.43 C_3 C_4 C_5

totaal dus 3 x 25 = 75 objecten in plaats van 5 x 5 x 5 = 125 bij uitvoering van het volledig factorial design, gespecificeerd als volgt:

48	objecten in enkelvoud	= 48
12	objecten in twee-voud	= 24
1	object in drievoud	= 3
61		75

totaal dus 75 objecten, waarbij 61 verschillende.

Eventueel zou dus met 61 objecten volstaan kunnen worden. Bij de uitvoering van de proef moeten, vooral bij een niet homogenen achtergrond van het experiment, deze 61 objecten volgens het toeval gerangschikt worden in de laboratoriumruimte of bv. bij landbouwkundige veldproeven op het proefveld. Er is dan echter geen gelegenheid meer tot correctie op de inhomogeniteit van den achtergrond van het experiment.

Verder kunnen we de trappen ook zoo kiezen dat geen herhalingen voorkomen bv. bij een 5 x 5 experiment onderzoeken we factor A en 5 trappen A_1, A_2, A_2,A_4 en A_5 bij trap 3½ van B en factor B eveneens in 5 trappen B_1 t/m B_5 bij trap 3½ van A. We krijgen dan 25 objecten zonder herhalingen. Evenzoo bij een 5 x 5 x 5 experiment met de factoren A, B en C; onderzoeken we de factoren A en B bij bv. den trap 3½ van C, de factoren A en C bij den trap 3½ van B en de factoren B en C bij den trap A 3½, dan krijgen 75 objecten zonder herhalingen. Correctie is dan eveneens niet mogelijk; de bepaling van een middelbare fout kan soms nog wel langs een omweg plaats vinden, nl. bij veel factoren is het mogelijk dat enkele variances onbeteekenend blijken te zijn, waaruit dan een schatting van de toevalsvariance gemaakt kan worden.

De mogelijkheid voor correctie is echter zoo noodzakelijk, vooral bij veel factoren, dat we ons verder alleen bezig zullen houden met het schema, waarbij herhalingen optreden.

We kunnen deze 75 objecten schematisch nog als volgt weergeven:

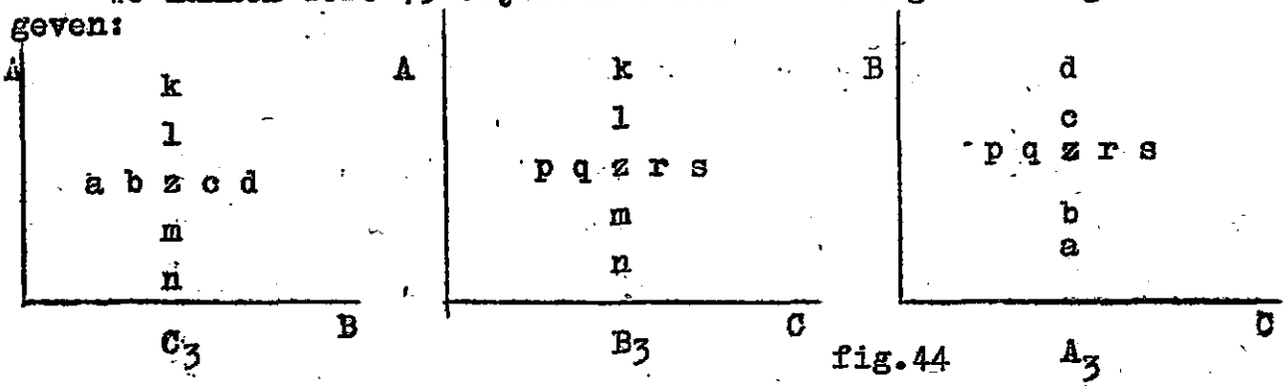


fig.44

De 48 objecten in enkelvoud zijn door een x aangegeven; 3 stel objecten komen in 2-voud voor nl: a, b, c en d; k, l, m en n en nog p, q, r en s, dit zijn de objecten, die in fig.42 op de 3 snijlijnen van de 3 met dikke lijnen aangegeven vlakken liggen; terwijl object z in 3-voud voorkomt, het snijpunt van alle 3 vlakken in fig.42.

Bij de uitvoering van het incomplete complex experiment houden we de 3 sub-complex experiments gescheiden, dus de objecten van elk sub-complex experiment bij elkaar; de interacties worden op die manier het nauwkeurigst vastgesteld. De 25 objecten van elk sub-complex experiment in ons voorbeeld, rangschikken we op het proefveld volgens het toeval. De dishomogeniteit van het proefveld heeft zoodoende op een zoo klein aantal objecten weinig invloed. Wel kan zij de drie sub-complex experiments in hun geheel beinvloeden, maar dit kan nu, dank zij de herhalingen, gecorrigeerd worden.

Het AB sub-complex experiment en het AC sub-complex experiment hebben de objecten k, l, z, m en n gemeen; het AB sub-complex experiment en het BC sub-complex experiment de objecten a, b, z, c en d en het AC sub-complex experiment en het BC sub-complex experiment hebben de objecten p, g, z, r en s gemeen.

Zouden we eerst de sub-complex experiments AB en AC corrigeeren dan moeten de 2 reeksen objecten k, l, z, m en n aan elkaar gelijk gemaakt worden. Dit is voor elk object afzonderlijk niet mogelijk, we zullen daarom hun sommen aan elkaar gelijk maken, zoodat elk object van eenzelfde sub-complex experiment dezelfde correctie krijgt.

Noemen we in sub-complex experiment AB de som der objecten k l z m en n = K_{AB} en in sub-complex experiment AC K_{AC}, dan is de correctie voor elk object voor het sub-complex experiment AB

$= - \frac{K_{AB}-K_{AC}}{2 \times 5}$ als $K_{AB} > K_{AC}$ en van sub-complex experiment AC

$= + \frac{K_{AB}-K_{AC}}{2 \times 5}$. Evenzoo noemen we in sub-complex experiment AB de som der objecten a b z c en d A_{AB} en in sub-complex experiment BC : A_{BC}. Op dezelfde wijze P_{AC} en P_{BC}.

We kunnen de sub-complex experiments echter niet 2 aan 2 corrigeeren op deze manier; zouden we eerst zoodoende sub-compl.exp.AB en AC gecorrigeerd hebben, daarna sub-compl.exp. AC en BC, dan zou de eerste correctie weer teniet gedaan worden. De correcties moeten zoodanig zijn, dat aan de drie voorwaarden gelijktijdig wordt voldaan. Deze 3 voorwaarden zijn dus $A_{AB}-A_{BC}$, $K_{AB}-K_{AC}$ en $P_{AC}-P_{BC}$ minimaal.

We moeten voor elk sub-complex experiment een correctie bepalen, die voor alle objecten van elk sub-complex experiment apart, gelijk is. Daar we met de sommen K_{AB}, P_{BC} enz. werken, is het het eenvoudigst de correcties voor deze sommen te bepalen. De correctie voor A_{AB} en K_{AB} noemen we C_{AB}; de correctie voor K_{AC} en P_{AC} noemen we C_{AC} en die voor A_{BC} en P_{BC} noemen we C_{BC}, zoodat we onze voorwaarden als volgt kunnen formuleeren:

$$\phi = \left\{A_{BC}+C_{BC}-(A_{AB}+C_{AB})\right\}^2 + \left\{K_{AB}+C_{AB}-(K_{AC}+C_{AC})\right\}^2 + \left\{P_{AC}+C_{AC}-(P_{BC}+C_{BC})\right\}^2 \quad \text{minimaal}$$

$$\frac{\partial \phi}{\partial C_{AB}} = -2(A_{BC}+C_{BC}-A_{AB}-C_{AB}) + 2(K_{AB}+C_{AB}-K_{AC}-C_{AC}) = 0$$

$$\frac{\partial \phi}{\partial C_{AC}} = -2(K_{AB}+C_{AB}-K_{AC}-C_{AC}) + 2(P_{AC}+C_{AC}-P_{BC}-C_{BC}) = 0$$

$$\frac{\partial \phi}{\partial C_{BC}} = 2(A_{BC}+C_{BC}-A_{AB}-C_{AB}) - 2(P_{AC}+C_{AC}-P_{BC}-C_{BC}) = 0$$

of $2C_{AB}-C_{AC}-C_{BC}+A_{AB}-A_{BC}+K_{AB}-K_{AC}=0 \quad (41)$

$2C_{AC}-C_{AB}-C_{BC}+K_{AC}-K_{AB}+P_{AC}-P_{BC}=0 \quad (42)$

$2C_{BC}-C_{AB}-C_{AC}+A_{BC}-A_{AB}+P_{BC}-P_{AC}=0 \quad (43)$

welk stel door eliminatie te reduceeren is tot:

$3C_{AB}-3C_{BC}+P_{AC}-P_{BC}+2A_{AB}-2A_{BC}+K_{AB}-K_{AC}=0 \quad (44)$

$3C_{AC}-3C_{AB}+2K_{AC}-2K_{AB}+A_{BC}-A_{AB}+P_{AC}-P_{BC}=0 \quad (45)$

$3C_{BC}-3C_{AC}+A_{BC}-A_{AB}+2P_{BC}-2P_{AC}+K_{AB}-K_{AC}=0 \quad (46)$

welk stel afhankelijk blijkt te zijn, maar oplosbaar is als we nog de volgende voorwaarden stellen:

$$S = C_{AB}^2 + C_{AC}^2 + C_{BC}^2 \text{ minimaal}$$

Uit (44) volgt:

$$C_{BC} = C_{AB} + \frac{P_{AC}-P_{BC}+2A_{AB}-2A_{BC}+K_{AB}-K_{AC}}{3} \quad (47)$$

Uit (45) volgt:

$$C_{AC} = C_{AB} + \frac{2K_{AB}-2K_{AC}+A_{AB}-A_{BC}+P_{BC}-P_{AC}}{3} \quad (48)$$

Uit (47) en (48) volgt:

$$C_{BC}^2 = C_{AB}^2 + 2C_{AB}\,\frac{P_{AB}-\cdots\cdots}{3} + \frac{(P_{AB}-\cdots\cdots)^2}{3}$$

$$C_{AC}^2 = C_{AB}^2 + 2C_{AB}\,\frac{2K_{AB}-\cdots\cdots}{3} + \frac{(2K_{AB}-\cdots\cdots)^2}{3}$$

$$C_{AB}^2 = C_{AB}^2$$

$$S = 3C_{AB}^2 + 2C_{AB}\left(\frac{P_{AB}\cdots\cdots}{3} + \frac{2K_{AB}-\cdots\cdots}{3}\right) + \cdots\cdots$$

$$\frac{\partial S}{\partial C_{AB}} = 6C_{AB} + 2\left(\frac{P_{AB}\cdots\cdots}{3} + \frac{2K_{AB}-\cdots\cdots}{3}\right) = 0$$

of $C_{AB} = \frac{K_{AC}-K_{AB}+A_{BC}-A_{AB}}{3} \quad (49)$

Substitutue van (49) in (47) en (48) geeft:

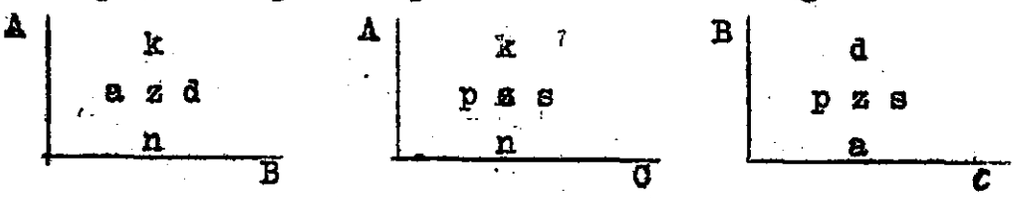
$$C_{BC} = \frac{P_{AC}-P_{BC}+A_{AB}-A_{BC}}{3} \quad (50)$$

$$C_{AC} = \frac{K_{AB}-K_{AC}+P_{BC}-P_{AC}}{3} \quad (51)$$

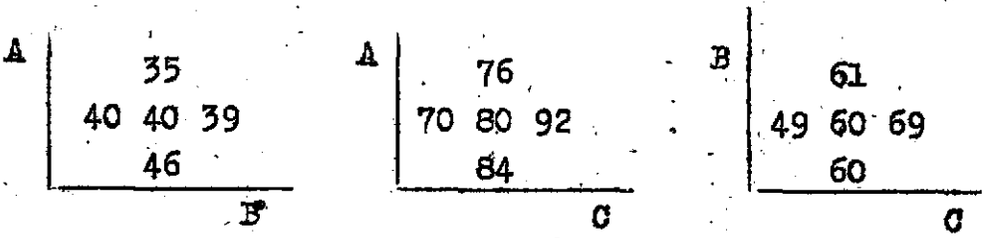
De correctie van elk object van de 3 sub-complex experiments is het 1/5 deel van deze correcties.

Dus alle 3 sub-complex experiments doen mee in een correctie waarin één sub-complex experiment zwaarder weegt dan de andere. We kunnen beide minimumvoorwaarden ook samenvatten in één minimumvoorwaarde, door een meetkundige voorstelling aan deze vereffening te verbinden. Aangezien bij de hierna te bespreken gevallen we dan te doen krijgen met meer dan 3 coördinaten, is de voorkeur gegeven aan de hier gevolgde methode, die voor meer lezers toegankelijk is.

Ter illustratie het volgende korte voorbeeld:
In het incomplete complex experiment van het volgende schema:



vinden we de volgende waarden:

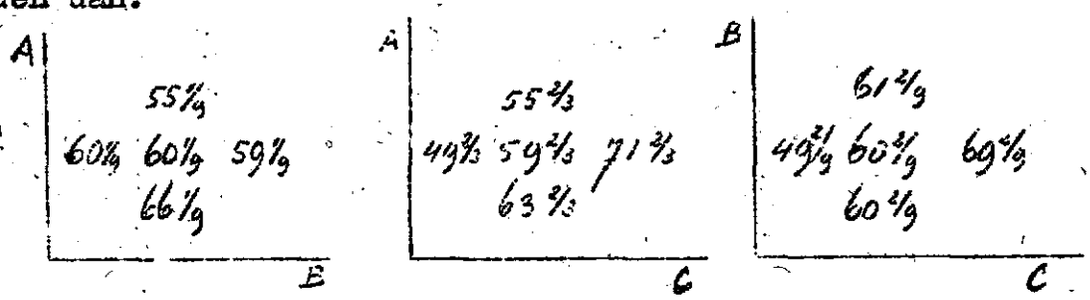


$A_{AB} = 40 + 40 + 39 = 119$
$A_{BC} = 60 + 60 + 61 = 181$
$K_{AB} = 35 + 40 + 46 = 121$
$K_{AC} = 76 + 80 + 84 = 240$
$P_{AC} = 70 + 80 + 92 = 242$
$P_{BC} = 49 + 60 + 69 = 178$

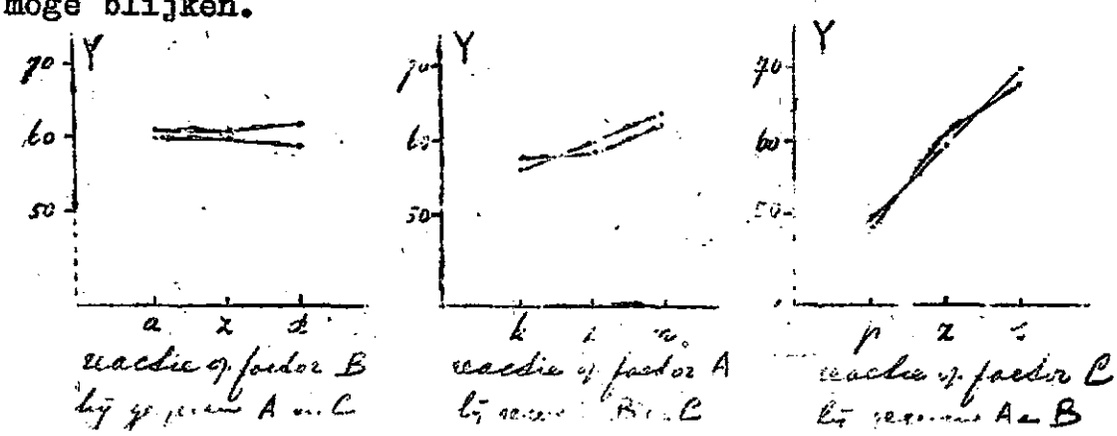
$$c_{AB} = \frac{240-121+181-119}{3} = \frac{181}{3} = 60\tfrac{1}{3}$$
$$c_{BC} = \frac{242-178+119-181}{3} = \frac{2}{3}$$
$$c_{AC} = \frac{121-240+178-242}{3} = -\frac{183}{3} = -61$$

rekencontole:
$60\tfrac{1}{3} + 2/3 - 61 = 0$

Het object wordt met 1/3 deel van deze correcties gecorrigeerd. We vinden dan:



Zooals we zien zijn door de correctie de 3 sub-complex experiments, die eerst zeer sterk verschilden, thans alle op hetzelfde niveau gekomen. Tevens zijn de objecten in duplo (en triplo) in goede overeenstemming gekomen, zooals uit de volgende graphische voorstelling moge blijken.



Een schatting van de toevalsvariance kunnen we uit de objecten in tweevoud gemakkelijk maken met de formule:

$$\sigma_1^2 = \frac{(l_{AB}-l_{AC})^2+(l_{AB}-l_{AC})^2+\ldots\ldots(r_{BC}-r_{BC})^2+\ldots\ldots(a_{AB}-a_{BC})^2+\ldots}{2+2+\ldots\ldots+2+\ldots\ldots+2\ldots} = \frac{\Sigma d^2}{2n_1}$$

waarin dus n_1 het aantal objecten in tweevoud is. (In het 5 x 5 x 5 experiment dus $n_1 = 12$).

Uit het object in drievoud vinden we als schatting

$$\sigma_2^2 = \frac{\Sigma (z-\bar{z})^2}{3-1}$$

zoodat we uiteindelijk vinden voor de toevalsvariance de schatting:

$$\sigma_3^2 = \frac{\Sigma d^2 + \Sigma (z-\bar{z})^2}{2n_1+2} \qquad (52)$$

Zooals uit bovenstaande blijkt, hebben we in de vereffening van de 3 sub-complex experiments onderling geen speciaal gebruik gemaakt van het object in drievoud, doch dit feitelijk opgenomen in de vereffening met behulp van de objecten in tweevoud, we hebben het als het ware gesplitst in 3 objecten in tweevoud, elk object van het object in drievoud behoort immers tot 2 reeksen van de objecten in tweevoud.

In verband met bovenstaande kunnen we nu nog een eenvoudiger formule voor σ^2 gebruiken:

$\frac{\Sigma^n (x_i-\bar{x})^2}{n-1}$ is ook te schrijven als $\frac{\Sigma^m d^2}{2m}$

als we uit $z_1\ z_2\ \ldots\ldots z_k\ \ldots\ldots z_n$ alle mogelijke verschillen d vormen. Dit aantal mogelijke verschillen is $C_2(n) = m$.

Bij het bepalen van de verschillen tusschen de objecten in tweevoud gaan we nu dus op dezelfde wijze te werk als bij de vereffening we slaan geen acht op het object in drievoud, deze wordt opgenomen in de verschillen d. In de reeksen A_{AB} en A_{BC} bepalen we dus de 5 verschillen $a_{AB} - a_{BC}$, $b_{AB} - b_{BC}$, $z_{AB} - z_{BC}$, $c_{AB} - c_{BC}$ en $d_{AB} - d_{BC}$ en de formule voor σ^2 wordt dus in zijn algemeene gedaante

$\sigma^2 = \frac{\Sigma d^2}{2n}$ waarin n dus het aantal verschillen d is (dus 3 x 5 = 15 of $n_1 + C_2 m = 12 + 3 = 15$). (53)

Thans kunnen we voor elk sub-complex experiment een analysis of variance uitvoeren. De plaats gevonden correctie heeft hierop geen invloed gehad, daar alle objecten met eenzelfde bedrag verhoogd of verlaagd zijn!!

De variances kunnen nu getest worden met de toevalsvariance, zooals we die gevonden hebben uit de herhalingsobjecten. Daarna kunnen we de vereffeningsformule opstellen.

Vinden we bv. in het sub-complex experiment AB een betrouwbare variance voor de factor A en voor de interactievariance AB; in sub-complex experiment AC een betrouwbare variance voor de factoren A en C; en tenslotte in sub-complex experiment BC een betrouwbare variance voor de factor B dan is de vereffeningsfunctie

$$y = \varphi(A) + \varphi(B) + \varphi(C) + \varphi(AB)$$

Tot den vorm van elke φ-functie kunnen we besluiten uit beschouwing van de graphische voorstellingen.

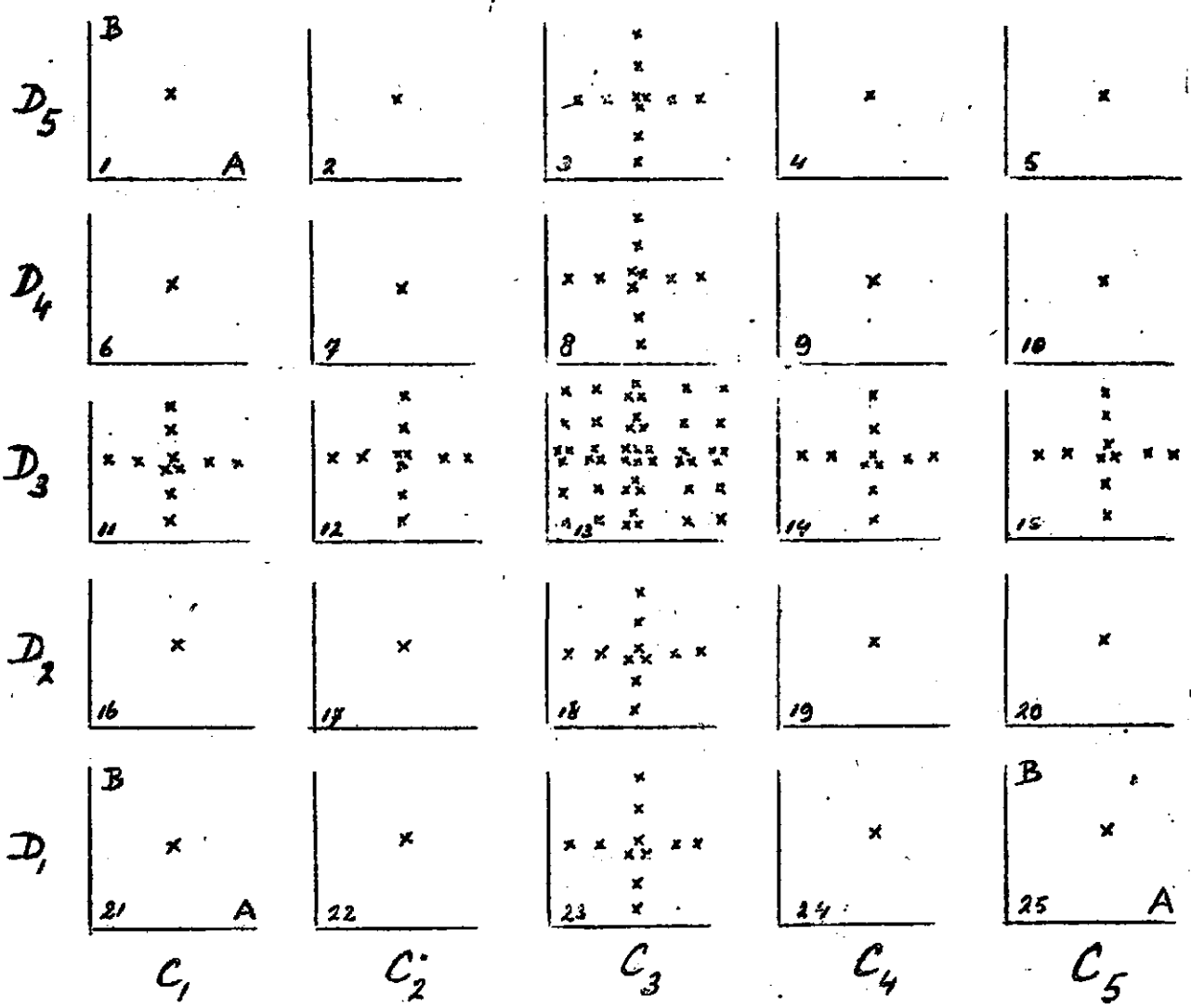
Na vereffening met de methode der kleinste quadraten is het nu mogelijk met behulp van de gevonden regressiefunctie de waarde van elk object uit het volledige 5 x 5 x 5 factorial design te bepalen, waarmede ons vraagstuk opgelost is!

Zoo hebben we het 5 x 5 x 5 complex experiment met 125 objecten dus gereduceerd tot een incompleet complex experiment met 3 x 25 = 75 objecten, die ons dezelfde resultaten, c.q. inlichtingen geeft, als het volledige complex-experiment mits de interactie van de 2e orde = 0 is!!

De zeer groote besparingen aan objecten, waarmee dus de mogelijkheid van uitvoering van experimenten met zeer veel factoren volgens het factorial design staat en valt, treden juist op bij veel factoren.

§ 4. Bespreking van het incomplete complex experiment bij 4 en 5 factoren.

Bezien we een complex experiment met 4 factoren bv. een 5 x 5 x 5 x 5 experiment. Uitvoering van het complete factorial design eischt 625 objecten. Uitvoering volgens een incomplete factorial design waarbij de interacties van de 2e en 3e orde = 0 zijn, eischt slechts 6 x 25 = 150 objecten, dus hier wordt de reductie reeds zeer aanzienlijk. We hebben de volgende 6 sub-designs: AB, AC, AD, BC, BD en CD. AB wordt uitgevoerd bij C_3D_3, AC bij B_3D_3, enz. enz. Door deze telkens "af te kruisen" in het overzicht van de 625 mogelijke objecten van het volledig factorial design, vinden we onderstaand schema waaruit na te gaan is, hoeveel objecten in meervoud voorkomen. Bij grootere experimenten, waarbij het volledig factorial design 10.000en en meer objecten bedraagt, zoekt men in het incomplete factorial design direct de meervoudige objecten op. Bij deze kleine experimenten is bovenstaand procédé gevolgd, omdat deze schema's mooi illustreeren, wat uit het volledig factorial design wegvalt en wat meervoudig onderzocht wordt:



gespecificeerd:	96	objecten in enkelvoud	= 96
	16	" " drievoud	= 48
	1	" " zesvoud	= 6
	113		150

Voor de uitvoering moeten we weer weten, welke objecten in meervoud aanwezig zijn om een correctie tusschen de 6 sub-complex experiments te kunnen aanbrengen:

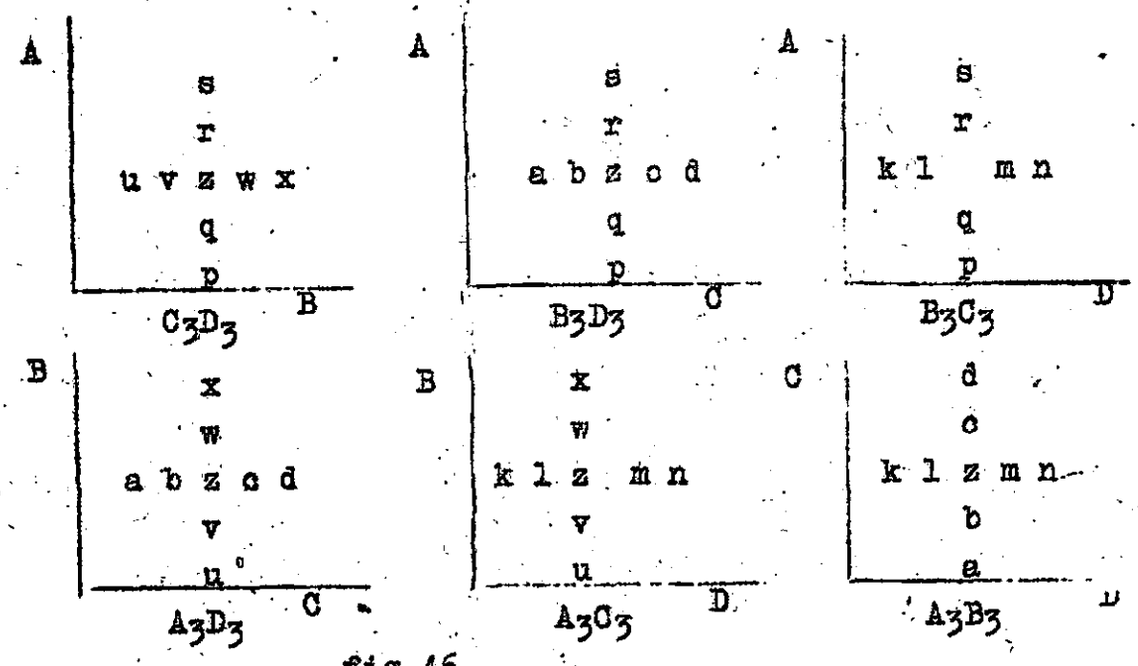


fig.46

We construeeren bovenstaand schema het gemaakelijkst door uit de voorgaande figuur de trappen van de meervoudige objecten af te lezen en deze in de bovenstaande figuur op te zoeken, bv in de voorgaande figuur vinden we in No.11 object $A_3B_3C_1D_3$ in 3-voud, noemen we dit object a en vullen dit in bovenstaand schema in in 2) 4) en 6), vervolgens vinden we in No.12 object $A_3B_3C_2D_3$ in 3-voud, noemen we deze b en vullen dit weer in enz. enz.

Voor de correctie van de sub-compl.exp. gebruiken we hier weer dezelfde notatie. In sub-compl.exp. AB is de som der objecten p q z r s weer P_{AB} enz.

Thans hebben we echter 3 P-waarden, 3 U-waarden enz. de minimum voorwaarde kan hier dus niet zoo eenvoudig zijn.
We vormen daartoe:

$$\frac{P_{AB}+C_{AB}+P_{AC}+C_{AC}+P_{AD}+C_{AD}}{3} = M_P$$

vervolgens $\pi_{AB} = P_{AB}+C_{AB}-M_P = \frac{2P_{AB}+2C_{AB}-P_{AC}-C_{AC}-P_{AD}-C_{AD}}{3}$

$$\pi_{AC} = P_{AC}+C_{AC}-M_P = \frac{2P_{AC}+2C_{AC}-P_{AB}-C_{AB}-P_{AD}-C_{AD}}{3}$$

$$\pi_{AD} = P_{AD}+C_{AD}-M_P = \frac{2P_{AD}+2C_{AD}-P_{AB}-C_{AB}-P_{AC}-C_{AC}}{3}$$

en eischen vervolgens $\Sigma \pi^2$ minimaal
naast het minimaal zijn van $\Sigma \alpha^2$, $\Sigma \varkappa^2$, $\Sigma \upsilon^2$ die op dezelfde wijze gevormd worden, dus uiteindelijk

$$\Phi' = \Sigma \pi^2 + \Sigma \alpha^2 + \Sigma \varkappa^2 + \Sigma \upsilon^2 \quad \text{minimaal}$$

$$\frac{U_{AB}+C_{AB}+U_{BC}+C_{BC}+U_{BD}+C_{BD}}{3} = M_U$$

$$\upsilon_{AB} = U_{AB}+C_{AB}-M_U = \frac{2U_{AB}+2C_{AB}-U_{BC}-C_{BC}-U_{BD}-C_{BD}}{3}$$

$$\upsilon_{BC} = U_{BC}+C_{BC}-M_U = \frac{2U_{BC}+2C_{BC}-U_{AB}-C_{AB}-U_{BD}-C_{BD}}{3}$$

$$\upsilon_{BD} = U_{BD}+C_{BD}-M_U = \frac{2U_{BD}+2C_{BD}-U_{AB}-C_{AB}-U_{BC}-C_{BC}}{3}$$

evenzoo $\varkappa_{AD} = \frac{2K_{AD}+2C_{AD}-K_{BD}-C_{BD}-K_{CD}-C_{CD}}{3}$

$$\varkappa_{BD} = \frac{2K_{BD}+2C_{BD}-K_{AD}-C_{AD}-K_{CD}-C_{CD}}{3}$$

$$\alpha_{CD} = \frac{2K_{CD} + 2C_{CD} - K_{AD} - C_{AD} - K_{BD} - C_{BD}}{3}$$

en $$\alpha_{AC} = \frac{2A_{AC} + 2C_{AC} - A_{BC} - C_{BC} - A_{CD} - C_{CD}}{3}$$

$$\alpha_{BC} = \frac{2A_{BC} + 2C_{BC} - A_{AC} - C_{AC} - A_{CD} - C_{CD}}{3}$$

$$\alpha_{CD} = \frac{2A_{CD} + 2C_{CD} - A_{AC} - C_{AC} - A_{BC} - C_{BC}}{3}$$

$$\phi = 9\phi' = (2P_{AB}+2C_{AB}-P_{AC}-C_{AC}-P_{AD}-C_{AD})^2 + (2P_{AC}+2C_{AC}-P_{AB}-C_{AB}-P_{AD}-C_{AD})^2$$
$$+ (2P_{AD}+2C_{AD}-P_{AB}-C_{AB}-P_{AC}-C_{AC})^2 + (2U_{AB}+2C_{AB}-U_{BC}-C_{BC}-U_{BD}-C_{BD})^2$$
$$+ (2U_{BC}+2C_{BC}-U_{AB}-C_{AB}-U_{BD}-C_{BD})^2 + (2U_{BD}+2C_{BD}-U_{AB}-C_{AB}-U_{BC}-C_{BC})^2 +$$
$$+ (2K_{AD}+2C_{AD}-K_{BD}-C_{BD}-K_{CD}-C_{CD})^2 + (2K_{BD}+2C_{BD}-K_{AD}-C_{AD}-K_{CD}-C_{CD})^2 +$$
$$+ (2K_{CD}+2C_{CD}-K_{AD}-C_{AD}-K_{BD}-C_{BD})^2 + (2A_{AC}+2C_{AC}-A_{BC}-C_{BC}-A_{CD}-C_{CD})^2 +$$
$$+ (2A_{BC}+2C_{BC}-A_{AC}-C_{AC}-A_{CD}-C_{CD})^2 + (2A_{CD}+2C_{CD}-A_{AC}-C_{AC}-A_{BC}-C_{BC})^2$$

$$\frac{\partial\phi}{\partial C_{AB}} = 4(2P_{AB}+2C_{AB}-P_{AC}-C_{AC}-P_{AD}-C_{AD})$$
$$-4(2P_{AC}+2C_{AC}-P_{AB}-C_{AB}-P_{AD}-C_{AD})$$
$$-4(2P_{AD}+2C_{AD}-P_{AB}-C_{AB}-P_{AC}-C_{AC})$$
$$+4(2U_{AB}+2C_{AB}-U_{BC}-C_{BC}-U_{BD}-C_{BD})$$
$$-4(2U_{BC}+2C_{BC}-U_{AB}-C_{AB}-U_{BD}-C_{BD})$$
$$-4(2U_{BD}+2C_{BD}-U_{AB}-C_{AB}-U_{BC}-C_{BC}) =$$

$$16P_{AB} - 8P_{AC} - 8P_{AD} + 16U_{AB} - 8U_{BC} - 8U_{BD}$$
$$+ 16C_{AB} - 8C_{AC} - 8C_{AD} + 16C_{AB} - 8C_{BC} - 8C_{BD} = 0$$

of $$4C_{AB} - C_{AC} - C_{AD} - C_{BC} - C_{BD} + R_{AB} = 0 \qquad (54)$$

op analoge wijze vinden we:

$$4C_{AC} - C_{AB} - C_{AD} - C_{BC} - C_{CD} + R_{AC} = 0 \qquad (55)$$

$$4C_{AD} - C_{AB} - C_{AC} - C_{BD} - C_{CD} + R_{AD} = 0 \qquad (56)$$

$$4C_{BC} - C_{AB} - C_{AC} - C_{BD} - C_{CD} + R_{BC} = 0 \qquad (57)$$

$$4C_{BD} - C_{AB} - C_{AD} - C_{BC} - C_{CD} + R_{BD} = 0 \qquad (58)$$

$$4C_{CD} - C_{AC} - C_{AD} - C_{BC} - C_{BD} + R_{CD} = 0 \qquad (59)$$

waarin $R_{AB} = 2P_{AB} - P_{AC} - P_{AD} + 2U_{AB} - U_{BC} - U_{BD}$

$R_{AC} = 2P_{AC} - P_{AB} - P_{AD} + 2A_{AC} - A_{BC} - A_{CD}$

$R_{AD} = 2K_{AD} - K_{BD} - K_{CD} + 2P_{AD} - P_{AC} - P_{AB}$

$R_{BC} = 2A_{BC} - A_{AC} - A_{CD} + 2U_{BC} - U_{BD} - U_{AB}$

$R_{BD} = 2K_{BD} - K_{AD} - K_{CD} + 2U_{BD} - U_{AB} - U_{BC}$

$R_{CD} = 2K_{CD} - K_{BD} - K_{AD} + 2A_{CD} - A_{BC} - A_{AC}$

Uit (54) t/m (59) volgt: $4C_{AB} + R_{AB} - 4C_{CD} - R_{CD} = 0$

$4C_{AC} + R_{AC} - 4C_{BD} - R_{BD} = 0$

$4C_{AD} + R_{AD} - 4C_{BC} - R_{BC} = 0$

of $C_{AB} - C_{CD} = \frac{R_{CD} - R_{AB}}{4}$ (60)

$C_{AC} - C_{BD} = \frac{R_{BD} - R_{AC}}{4}$ (61)

$C_{AD} - C_{BC} = \frac{R_{BC} - R_{AD}}{4}$ (62)

Substitutie van (62) en (61) in (54) geeft:

$4C_{AB} - C_{AC} - C_{AD} - C_{AD} + \frac{R_{BC} - R_{AD}}{4} - C_{AC} + \frac{R_{BD} - R_{AC}}{4} + R_{AB} = 0$

of $4C_{AB} - 2C_{AC} - 2C_{AD} = \frac{R_{AD} - R_{BC} + R_{AC} - R_{BD} - 4R_{AB}}{4}$ (63)

Substitutue van (62) en (60) in (55) geeft:

$4C_{AC} - C_{AB} - C_{AD} - C_{AD} + \frac{R_{BC} - R_{AD}}{4} - C_{AB} + \frac{R_{CD} - R_{AB}}{4} + R_{AC} = 0$

of $4C_{AC} - 2C_{AB} - 2C_{AD} = \frac{R_{AD} - R_{BC} + R_{AB} - R_{CD} - 4R_{AC}}{4}$ (64)

$(63) - (64) = C_{AB} - C_{AC} = \frac{5R_{AC} - R_{BD} - 5R_{AB} + R_{CD}}{24}$ (65)

$(63) + 2x(64) = C_{AC} - C_{AD} = \frac{3R_{AD} - 3R_{BC} - 2R_{AB} - 7R_{AC} - R_{BD} - 2R_{CD}}{24}$ (66)

$2x(63) + (64) = (65) + (66) = C_{AB} - C_{AD} = \frac{3R_{AD} - 3R_{BC} - 7R_{AB} - 2R_{AC} - 2R_{BD} - R_{CD}}{24}$ (67)

Thans zijn de correcties op te lossen, als we nu alle correcties in C_{AB} uitdrukken en vervolgens nog als voorwaarde stellen

$S = C_{AB}^2 + C_{AC}^2 + C_{AD}^2 + C_{BC}^2 + C_{BD}^2 + C_{CD}^2$ =minimaal, daar we uit ϕ niet meer dan 5 onafhankelijke vergelijkingen kunnen vormen om de C's op te lossen.

$C_{AB} = C_{AB}$

$C_{AC} = C_{AB} + \frac{R_{BD} - 5R_{AC} + 5R_{AB} - R_{CD}}{24}$

$C_{AD} = C_{AB} + \frac{3R_{BC} - 3R_{AD} + 7R_{AB} + 2R_{AC} + 2R_{BD} + R_{CD}}{24}$

$C_{BC} = C_{AB} + \frac{-3R_{BC} + 3R_{AD} + 7R_{AB} + 2R_{AC} + 2R_{BD} + R_{CD}}{24}$

$C_{BD} = C_{AB} + \frac{-5R_{BD} + R_{AC} + 5R_{AB} - R_{CD}}{24}$

$C_{CD} = C_{AB} + \frac{6R_{AB} - 6R_{CD}}{24}$

Hieruit volgt:

$C_{AB}^2 = C_{AB}^2$

$C_{AC}^2 = C_{AB}^2 + 2C_{AB}(\frac{R_{BD} \cdots\cdots}{24}) + (\cdots\cdots)^2$

$C_{AD}^2 = C_{AB}^2 + 2C_{AB}(\frac{3R_{BC} \cdots\cdots}{24}) + (\cdots\cdots)^2$

$C_{BC}^2 = C_{AB}^2 + 2C_{AB}(\frac{-3R_{BC} \cdots\cdots}{24}) + (\cdots\cdots)^2$

$$C_{BD}^2 = C_{AB}^2 + 2C_{AB}\left(\frac{-5R_{BD}\cdots}{24}\right) + (\ldots\ldots)^2$$

$$C_{CD}^2 = C_{AB}^2 + 2C_{AB}\left(\frac{6R_{AB}\cdots}{24}\right) + (\ldots\ldots)^2$$

$$S = 6C_{AB}^2 + 2C_{AB}\left(\frac{30R_{AB} - 6R_{CD}}{24}\right) + \ldots\ldots$$

$$\frac{\partial S}{\partial C_{AB}} = 12C_{AB} + 2\left(\frac{30R_{AB} - 6R_{CD}}{24}\right) = 0$$

$$C_{AB} = \frac{R_{CD} - 5R_{AB}}{24} \qquad (68)$$

(Uit bovenstaande berekening is te zien dat hetzelfde resultaat berecht wordt als we als voorwaarde stellen:
$S = C_{AB} + C_{AC} + C_{AD} + C_{BC} + C_{BD} + C_{CD} = 0$ hetgeen eenvoudiger is.
We vinden dan:

$$S = 6C_{AB} + \frac{30R_{AB} - 6R_{CD}}{24} = 0\ .).$$

$$C_{AC} = \frac{R_{CD}-5R_{AB}}{24} + \frac{R_{BD}-5R_{AC}+5R_{AB}-R_{CD}}{24} = \frac{R_{BD}-5R_{AC}}{24} \qquad (69)$$

$$C_{AD} = \frac{R_{CD}-5R_{AB}}{24} + \frac{3R_{BC}-3R_{AD}+7R_{AB}+2R_{AC}+2R_{BD}+R_{CD}}{24} =$$

$$= \frac{3R_{BC}-3R_{AD}+2R_{AB}+2R_{AC}+2R_{BD}+2R_{CD}}{24} \qquad (70)$$

$$C_{BC} = \frac{R_{CD}-5R_{AB}}{24} + \frac{-3R_{BC}+3R_{AD}+7R_{AB}+2R_{AC}+2R_{BD}+R_{CD}}{24} =$$

$$= \frac{-3R_{BC}+3R_{AD}+2R_{AB}+2R_{AC}+2R_{BD}+2R_{CD}}{24} \qquad (71)$$

$$C_{BD} = \frac{R_{CD}-5R_{AB}}{24} + \frac{-5R_{BD}+R_{AC}+5R_{AB}-R_{CD}}{24} =$$

$$= \frac{R_{AC}-5R_{BD}}{24} \qquad (71a)$$

$$C_{CD} = \frac{R_{CD}-5R_{AB}}{24} + \frac{6R_{AB}-6R_{CD}}{24} = \frac{R_{AB}-5R_{CD}}{24} \qquad (72)$$

Thans dienen alle R's nog weer vervangen te worden, we vinden dan:

$$C_{AD} = \frac{2K_{CD}-K_{BD}-K_{AD}+2A_{CD}-A_{BC}-A_{AC}-10P_{AB}+5P_{AC}+5P_{AD}-10U_{AB}+5U_{BC}+5U_{BD}}{24} \qquad (73)$$

$$C_{AC} = \frac{2K_{BD}-K_{AD}-K_{CD}+2U_{BD}-U_{AB}-U_{BC}-10P_{AC}+5P_{AB}+5P_{AD}-10A_{AC}+5A_{BC}+5A_{CD}}{24} \qquad (74)$$

$$C_{AD} = \frac{2A_{BC}-A_{AC}-A_{CD}+2U_{BC}-U_{BD}-U_{AB}-10K_{AD}+5K_{BD}+5K_{CD}-10P_{AD}+5P_{AC}+5P_{AB}}{24} \qquad (75)$$

$$C_{BC} = \frac{2K_{AD}-K_{BD}-K_{CD}+2P_{AD}-P_{AC}-P_{AB}-10A_{BC}+5A_{AC}+5A_{CD}-10U_{BC}+5U_{BD}+5U_{AB}}{24} \qquad (76)$$

$$C_{BD} = \frac{2P_{AC}-P_{AB}-P_{AD}+2A_{AC}-A_{BC}-A_{CD}-10K_{BD}+5K_{AD}+5K_{CD}-10U_{BD}+5U_{AB}+5U_{BC}}{24} \qquad (77)$$

$$C_{CD} = \frac{2P_{AB}-P_{AC}-P_{AD}+2U_{AB}-U_{BC}-U_{BD}-10K_{CD}+5K_{BD}+5K_{AD}-10A_{CD}+5A_{BC}+5A_{AC}}{24} \qquad (78)$$

Ter illustratie het volgende voorbeeld:

A | 6 8 10 8 16 B	A | 12 15 15 14 21 C	A | 25 14 20 26 27 D
B | 21 19 20 20 19 C	B | 25 21 25 29 26 D	C | 29 25 30 37 29 D

$A_{AC} = 15 + 15 + 14 = 44$ $\quad K_{AD} = 14 + 20 + 26 = 60$

$A_{BC} = 19 + 20 + 20 = 59$ $\quad K_{BD} = 21 + 25 + 29 = 75$

$A_{CD} = 29 + 30 + 29 = 88$ $\quad K_{CD} = 25 + 30 + 37 = 92$

$P_{AB} = 16 + 10 + 6 = 32$ $\quad U_{AB} = 8 + 10 + 18 = 26$

$P_{AC} = 21 + 15 + 12 = 48$ $\quad U_{BC} = 19 + 20 + 21 = 60$

$P_{AD} = 27 + 20 + 15 = 62$ $\quad U_{BD} = 26 + 25 + 25 = 76$

$$C_{AB} = \frac{184-75-60+176-59-44-320+240+310-260+300+380}{24} = +32\frac{4}{24}$$

$$C_{AC} = \frac{150-60-92+152-26-60-480+160+310-440+295+440}{24} = +14\frac{13}{24}$$

Evenzoo $C_{AD} = +\frac{19}{24}$ $\quad C_{BC} = -\frac{23}{24}$ $\quad C_{BD} = -15\frac{17}{24}$ $\quad C_{CD} = -30\frac{20}{24}$

rekencontrole $C_{AB} + C_{AC} + C_{AD} + C_{BC} + C_{BD} + C_{CD} = 0$

De afzonderlijke objecten krijgen alle 1/3 van deze correctie. We zullen ze niet alle geven; alleen de gecorrigeerde z-waarden, waaraan we mooi kunnen zien dat de 6 sub-compl.exp. op hetzelfde niveau gekomen zijn.

$Z_{AB} = 20\frac{13}{18}$ $\quad Z_{AC} = 19\frac{61}{72}$ $\quad Z_{AD} = 20\frac{19}{72}$

$Z_{BC} = 19\frac{49}{72}$ $\quad Z_{BD} = 19\frac{55}{72}$ $\quad Z_{CD} = 19\frac{52}{72}$

Het object in 6-voud is hierin weer verwerkt, analoog aan het vorig geval, waarin het object in 3-voud verwerkt was in de vereffening van de objecten in 2-voud.

Overigens kan hier weer hetzelfde gezegd worden van de error of estimation als op pg. 114. De eenvoudige formule met de verschillen d is hier echter niet van toepassing.

Ook de 6 analysis of variances en de opstelling van de regressiefunctie levert hier geen byzondere moeilijkheden op.

Op dezelfde wijze kunnen we verder incomplete complex experiments ontwerpen. Bij 4 factoren kan men ook alleen de interactie van de 3e orde = 0 stellen. Onderzoeken we weer elken factor in 5 trappen, dan bestaat het volledig compl.experiment uit $5^4 = 625$ objecten; het incomplete complex experiment bestaat uit 4 sub-complex experiments elk van $5^3 = 125$ objecten of totaal 500 objecten en wel nader gespecificeerd:

256	objecten in enkelvoud	= 256
96	objecten in 2-voud	= 192
16	objecten in 3-voud	= 48
1	object in 4-voud	= 4
369		500

Vergeleken met het volledig complex experiment is dit een geringe reductie van het aantal objecten, zoodat dit ontwerp natuurlijk nooit voor uitvoering in aanmerking komt.

Is het aantal factoren echter 5 dan bestaat het volledig complex experiment uit $5^5 = 3125$ objecten. Het incomplete complex experiment, als we de interacties van de 3e orde en van de 4e orde = 0 stellen uit $C_3(5) \times 5^3 = 1250$ objecten. Hier dus wel een belangrijke reductie.

§ 5. Tabellen

Het aantal objecten van een incomplete complex experiment kan ook grooter worden dan het complete complex experiment !

Een volledig 3x3x3x3 complex experiment bestaat uit 81 objecten. Bij een incomplete complex experiment, waarin we de interactie van de 3e orde = 0 stellen, is het aantal objecten $C_3(4) \times 3^3 = 108$.

We moeten nu dus een overzicht maken, waaruit te zien is, hoe groot de reductie is bij

1e een gegeven aantal factoren

2e bij een gegeven aantal trappen, waarin de factoren onderzocht worden,

als gegeven is welke de interactie van hoogste orde is, die men nog onderzocht wil hebben.

Is in een of ander experiment met n factoren de interactie van de 3e orde, de interactie van hoogste orde, die men onderzoeken wil, zoo bestaat het incomplete complex experiment uit $C_4(n)$ sub-complex experiments en elk sub-complex experiment uit $m_i \times m_j \times m_k \times m_l$ objecten, als m het aantal trappen is, waarin de factoren onderzocht worden.

Wordt elke factor in een verschillend aantal trappen onderzocht, dan bestaan ook de sub-complex experiments uit een verschillend aantal objecten.

Om os overzicht te vergemakkolijken, zullen we aannemen, dat alle factoren in een gelijk aantal trappen m onderzocht zullen worden. Het aantal objecten van een incomplete complex experiment is dan eenvoudig

$$C_4(n) \times m^4$$

We hebben dus noodig een tabel voor de combinatiecoefficienten $C_2(n), C_3(n)$ enz. waarin n het argument is. Dit is tabel I. We kunnen ook zeggen, dat deze tabel het aantal sub-complex experiments weergeeft. Het is duidelijk dat deze tabel niets anders is dan de driehoek van Pascal.

Daarna volgt voor m = 2, 3, 4, 5 en 6 de tabellen II t/m VI, waarin het totaal aantal objecten is aangegeven.

Ter vergelijking is in de 2e kolom telkens het aantal objecten van het complete complex experiment vermeld = m^n.

Wat de onderzochte interactie van de hoogste orde is kan eveneens worden aangegeven met de combinatie coefficienten. In de bovenste horizontale kolom wil bv. $C_4(x)$ zeggen dat de subcompl.exp. alle m x m x m x m experimenten zijn, dus dat de interactie van de 3e orde, de hoogste is, die nog onderzocht wordt.

Verder is met een vette lijn aangegeven onder welke omstandigheden een reductie in het aantal objecten optreedt !

Aant. fact. x	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$	$C_{11}(x)$	$C_{12}(x)$	$C_{13}(x)$	$C_{14}(x)$	$C_{15}(x)$	$C_{16}(x)$	$C_{17}(x)$	$C_{18}(x)$	$C_{19}(x)$	$C_{20}(x)$
1																			
2	1																		
3	3	1																	
4	6	4	1																
5	10	10	5	1															
6	15	20	15	6	1														
7	21	35	35	21	7	1													
8	28	56	70	56	28	8	1												
9	36	84	126	126	84	36	9	1											
10	45	120	210	252	210	120	45	10	1										
11	55	165	330	462	462	330	165	55	11	1									
12	66	220	495	792	924	792	495	220	66	12	1								
13	78	286	715	1287	1716	1716	1287	715	286	78	13	1							
14	91	364	1001	2002	3003	3432	3003	2002	1001	364	91	14	1						
15	105	455	1365	3003	5005	6435	6435	5005	3003	1365	455	105	15	1					
16	120	560	1820	4368	8008	11440	12870	11440	8008	4368	1820	560	120	16	1				
17	136	680	2380	6188	12376	19448	24310	24310	19448	12376	6188	2380	680	136	17	1			
18	153	816	3060	8568	18564	31824	43758	48620	43758	31824	18564	8568	3060	816	153	18	1		
19	171	969	3876	11628	27132	50388	75582	92378	92378	75582	50388	27132	11628	3876	969	171	19	1	
20	190	1140	4845	15504	38760	77520	125970	167960	184756	167960	125970	77520	38760	15504	4845	1140	190	20	1

Tabel II. Aantal objecten van het 2x2x2x.... pan parallelle experiment .

Aant. fact. x	compl. complex exp.	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$	$C_{11}(x)$	$C_{12}(x)$	$C_{13}(x)$	$C_{14}(x)$
1	2													
2	4	4												
3	8	12	8											
4	16	24	32	16										
5	32	40	80	80	32									
6	64	60	160	240	192	64								
7	128	84	280	560	672	448	128							
8	256	112	448	1120	1792	1792	1024	256						
9	512	144	672	2016	4032	5376	4608	2304	512					
10	1024	180	960	3360	8064	13440	15360	11520	5120	1024				
11	2048	220	1320	5280	14784	29568	42240	42240	28160	11264	2048			
12	4096	264	1760	7920	25344	59136	101376	126720	112640	67584	24576	4096		
13	8192	312	2288	11440	41184	109824	219648	329472	366080	292864	159744	53248	8192	
14	16384	364	2912	16016	64064	192192	439296	768768	1025024	1025024	745472	372736	144688	16384
15	32768	420	3640	21840	96096	320320								
16	65536	480	4480	29120	139776	512512								
17	131072	544	5440	38080	198016	792064								
18	262144	612	6528	48960	274176	1118096								
19	524288	684	7752	62016	372096	1736448								
20	1048576	760	9120	77520	496128	2480640								

Tabel III. Aantal objecten van het 3x3x3x.... pan parallelle experiment.

Aant. fact. x	compl. complex exp.	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$	$C_{11}(x)$	$C_{12}(x)$
1	3											
2	9	9										
3	27	27	27									
4	81	54	108	81								
5	243	90	270	405	243							
6	729	135	540	1215	1458	729						
7	2187	189	945	2835	5103	5103	2187					
8	6561	252	1512	5670	13608	20412	17496	6561				
9	19683	324	2268	10206	30608	61236	78732	59049	19683			
10	59049	405	3240	17010	61236	153090	262440	295245	196830	59049		
11	177147	495	4455	26730	112266	336798	721710	1082565	1082565	649539	177147	
12	531441	594	5940	40095	192456	673596	1732104					
13	1594323	702	7722	57915	312741	1250964	3752892					
14	4782969	819	9828	81081	486486	2189187	7505784					

Tabel IV. Aantal objecten van het 4x4x4x.... pan parallelle experiment.

Aant. fact. x	complete complex exp.	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$
1	4									
2	16	16								
3	64	48	64							
4	256	96	256	256						
5	1024	160	640	1280	1024					
6	4096	240	1280	3840	6144	4096				
7	16384	336	2240	8960	21504	28672	16384			
8	65536	448	3584	17920	57344	114688	131072	65536		
9	262144	576	5376	32256	129024	344064	589824	589824	262144	
10	1048576	720	7680	53760	258048	860160	1966080	2949120	2621440	1048576

Tabel V. Aantal objecten van het 5x5x5x.... pan parallelle experiment.

Aant. fact. x	complete complex exp.	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$
1	5									
2	25	25								
3	125	75	125							
4	625	150	500	625						
5	3125	250	1250	3125	3125					
6	15625	375	2500	9375	18750	15625				
7	78125	525	4375	21875	65625	109375	78125			
8	390625	700	7000	43750	175000	437500	625000	390625		
9	1953125	900	10500	78750	393750	1312500	2812500	3515625	1953125	
10	9765625	1125	15000	131250	787500	3281250	9375000	17578125	19531250	9765625

Tabel VI. Aantal objecten van het 6x6x6x.... pan parallelle experiment.

Aant. fact. x	complete complex exp.	$C_2(x)$	$C_3(x)$	$C_4(x)$	$C_5(x)$	$C_6(x)$	$C_7(x)$	$C_8(x)$	$C_9(x)$	$C_{10}(x)$
1	6									
2	36	36								
3	216	108	216							
4	1296	216	864	1296						
5	7776	360	2160	6480	7776					
6	46656	540	4320	19440	46656	46656				
7	279936	756	7560	45360	163296	326592	279936			
8	1679616	1008	12096	90720	435456	1306368	2239488	1679616		
9	10077696	1296	18144	163296	979776	3919104	33592320	15116544	10077696	
10	60466176	1620	25920	272160	1959552	9797760	10077696	75582720	100776960	60466176

Duidelijk zien we hieruit, dat als de factoren in een gering aantal trappen onderzocht worden, de pan-parallelle experimenten pas beteekenis krijgen bij veel factoren, dat als de factoren in een grooter aantal trappen onderzocht worden de pan-parallelle experimenten ook reeds bij een gering aantal factoren van groote practische beteekenis zijn.

Verder zien we, dat als we interacties van hooge orde onderzocht willen hebben, het aantal objecten ook in de pan-parallelle experimenten zeer hoog wordt !

Verder geldt, dat als we interacties van zeer hooge orde onderzocht willen hebben, het incomplete complex experiment geen zin meer heeft, daarbij wordt het aantal objecten veel grooter dan in het complete complex experiment, terwijl het ons toch in sommige gevallen minder inlichtingen verschaffen kan, doordat niet alle objecten van het complete complex experiment in het incomplete complex experiment voorkomen, hoewel het toch meer objecten telt.

Na deze uiteenzettingen in voorgaande paragraphen valt thans beter aan te geven in hoeverre het incomplete complex experiment tegemoet komt aan het 2e punt genoemd in § 1 n.l. in hoeverre met het incomplete complex experiment aan "de natuurwetten te ontkomen is". Een en ander hield verband met de 2 groepen, waarin we in het vorig hoofdstuk de factoren onderverdeelden n.l. in een groep, waarvan de factoren in extreme trappen voorkomen en in een groep, waar de factoren in niet-extreme trappen voorkomen.

Beschouwen we thans het incomplete complex experiment met 4 factoren, waarbij de interacties van de 2e en 3e orde = 0 gesteld zijn, beschouwen we in het bijzonder de daarbij behoorende figuur no.45. Hierbij worden telkens 2 factoren, zoowel bij extreme als bij niet extreme-trappen onderzocht en wel altijd bij niet-extreme waarden der 2 andere factoren. Zoo worden de factoren A en B beide bij de trappen 1 t/m 5 onderzocht, dus bij extreme en niet-extreme trappen, terwijl de beide andere factoren C en D bij niet-extreme trap C_3 en D_3 onderzocht worden.

De factor C wordt daarentegen met bv. B samen onderzocht in extreme en niet-extreme trappen, bij niet-extreme waarden van A en D enz.

Door de trappen van de factoren C en D, waarbij A en B in een sub-factorial design onderzocht worden, anders te kiezen, kunnen we deze groepeering beinvloeden. In de fig.45 uit zich dit door verlegging van het "kruisjescentrum". Inmiddels, wilden we weer alle mogelijkheden verwezenlijken, dan zou dit neerkomen op het nemen van het volledige complex experiment.

Algemeen gezegd: storen we ons niet aan de interacties hooger dan de (m-1) orde, dan kan de verhouding tusschen beide groepen de verhouding m : n nooit te boven gaan (hierbij is m steeds klein, anders vervalt de waarde van het incomplete complex experiment). Door echter voor de factoren uit de n groep niet steeds de midden waarde te kiezen, zooals we dat tot nog toe fixeerden bij het incomplete complex experiment, is hierop echter nog invloed uit te oefenen.

Algemeen gezegd komt het pan-parallelle experiment weinig aan het 2e punt tegemoet!

§ 6. Splitsen van de factoren in groepen onafhankelijke factoren met behulp van het 2^2 experiment.

Heeft men zeer veel factoren, dan bestaat de mogelijkheid dat een aantal factoren onafhankelijk van elkaar zijn (onafhankelijk weer in den zin van geen interactie tusschen de factoren), hetgeen nieuwe mogelijkheden kan openen voor ons probleem.

Beschouwen we ter orientatie eerst eens de 6 factoren A, B, C, D, E en F dan kunnen zich de volgende gevallen voordoen, als we willen trachten de factoren in 2 groepen te splitsen, zoodanig, dat tusschen géén enkele factor van de eene groep en een factor van de andere groep interactie bestaat.

a) tusschen A en B t/m F geen interactie; de factoren zijn te splitsen in 2 groepen, groep A en groep B t/m F
b) tusschen A en B, C en D geen interactie de factoren zijn niet te splitsen, wel als E en F constant gehouden worden.

c) tusschen A en B geen interactie, als voren; wel te splitsen als C t/m F constant gehouden worden.
d) tusschen A en B eenerzijds en C, D, E en F anderzijds geen interacties de factoren zijn te splitsen in 2 groepen, een groep A en B en een groep C t/m F.
e) tusschen A en B eenerzijds en E en F anderzijds geen interactie; de factoren zijn niet te splitsen; wel als C en D constant gehouden worden.
f) tusschen A en B eenerzijds en C t/m F anderzijds geen interactie, eveneens tusschen C en D eenerzijds en A, B, E en F anderzijds geen interactie; de factoren zijn te splitsen in 3 groepen.

Heeft men nu zeer veel factoren, waarvoor het complete factorial design een onbereikbaar ideaal is, dan is een eventueele splitsing zeer belangrijk, want voor die gesplitste groepen zal vermoedelijk wel het complete factorial design uitvoerbaar zijn. Hoe echter deze splitsing te onderzoeken ?

Zekerheid over de splitsing is slechts te krijgen door een volledig factorial design uit te voeren.

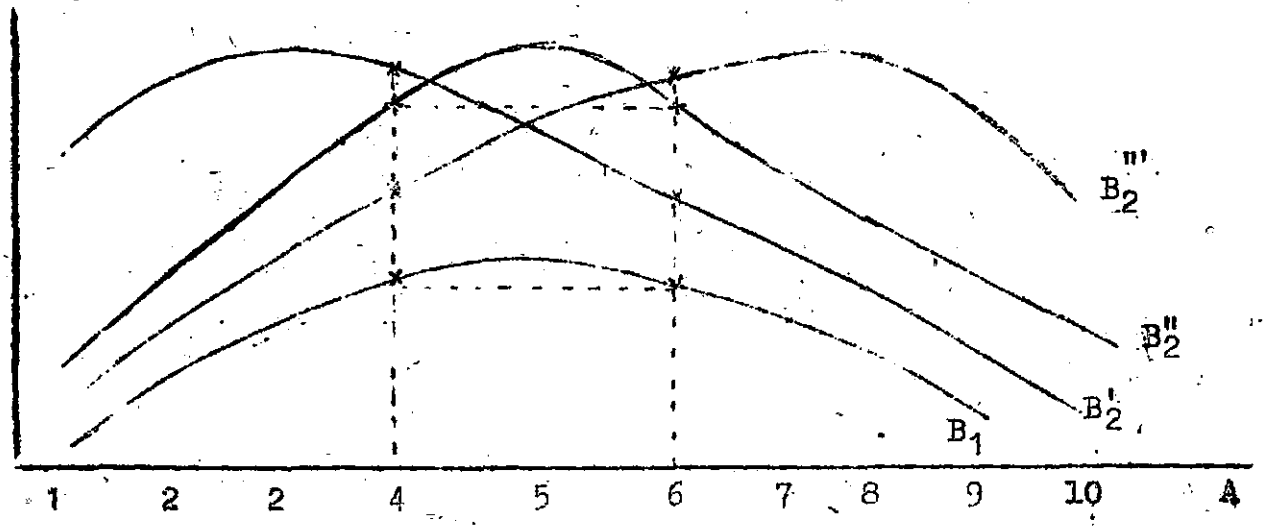
Bij gegeven aantal factoren wordt het aantal objecten zoo klein mogelijk als we het minimum aantal trappen kiezen, dus een 2x2x2x.... design.

Bij 10 factoren omvat dit volledig factorial design 1024 objecten (tabel II). Blijkt hieruit (of dit inderdaad kan, zullen we dadelijk bespreken) dat de factoren in 2 onafhankelijke groepen te splitsen zijn bv. 2 groepen van 5 factoren, dan kunnen we deze 2 groepen in een complete cq incomplete factorial design verder onderzoeken. Nemen we bv. 5 trappen per factor, wat wel het minimum is, om een optimum te benaderen, dan is aantal objecten van elk der beide complete factorial designs 3125 (tabel V), terwijl het complete factorial design van 10 factoren, elken factor ook in 5 trappen 9.765.625 objecten vereischt had. Een ontzaggelijke reductie dus van het aantal objecten. Bij veel grooter aantal factoren moet ook dit 2x2x2x.... experiment uitgevoerd worden als een incomplete complex experiment, immers een 2x2x2x.... experiment met 20 factoren eischt 1.048.576 objecten (tabel II) en de splitsing in eventueel te vinden groepen eischt dit dus ook. De zekerheid, dat deze groepen inderdaad bestaan, is daarmede ten deele vervallen, we moeten het dan ook zien als een hulpmiddel - dat echter zeer dienstig kan zijn - evenzeer als het incomplete factorial design al een hulpmiddel is in zekeren zin.

We moeten nu nog de vraag onder oogen zien of met het 2x2x2x.... experiment de interacties voldoende vast te stellen zijn.

De minder belangrijke interacties zijn die, waar het optimum op zijn plaats blijft, de belangrijkste, waarbij het optimum verschuift, hetzij naar links, hetzij naar rechts.

Beschouwen we daartoe de wisselwerking tusschen 2 factoren A en B, factor A wordt in een 10-tal trappen onderzocht en B in 2 trappen. We kunnen dan de volgende gevallen krijgen:



B_1 geeft den 1en trap van den factor B aan, terwijl we voor den 2en trap de gevallen B_2 B_2' en B_2'' onderscheiden.

a) Onderzoeken we in 2 lage trappen A_1 en A_2 dan kunnen voor alle 3 gevallen de interacties moeilijk vastgesteld worden.
b) Onderzoeken we in 1 lagen en 1 gemiddelden trap bv. trap A_1 en A_4, dan komen we tot dezelfde conclusie. Analoog zijn 2 hooge of 1 hooge en 1 gemiddelde trap.
c) Onderzoeken we in 1 minder lagen en 1 gemiddelden trap dan kunnen we voor het 1e geval een interactie duidelijk aantoonen, voor het 2e en 3e geval moeilijk.
d) Nemen we 2 gemiddelde trappen links en rechts van het gemiddelde bv A_4 en A_6, dan kunnen we voor het 1e en 3e geval duidelijk een interactie aantoonen, voor het 2e geval moeilijk of niet.
e) Onderzoeken we in 1 lagen en 1 hoogen trap, dan is het in het 1e en 3e geval zeer duidelijk een interactie aan te toonen, in het 2e geval zeer moeilijk.

Het beter of slechter aantoonbaar zijn hangt af van het verloop der reactiecurven; de hierboven gegeven situatie dient slechts als voorbeeld, ter bepaling van de gedachten.

Hierbij is onze conclusie, dat de belangrijkste interacties het duidelijkst bij een lagen trap en bij een hoogen trap van factor A gevonden worden.

In het algemeen bij een 2x2x2x.... experiment, hoe ook het preciese verloop der reactiecurven moge zijn, heeft men vermoedelijk de beste kansen bij 2 min of meer extreme en tegengestelde trappen voor alle factoren. Zekerheid verkrijgt men natuurlijk niet, het is slechts een voorgesteld hulpmiddel. Aanvaardt men dit, dan is het noodig 2x2x2x.... experiment nog eensstatistisch te beschouwen.

Wordt het 2x2x2x.... experiment uitgevoerd als complete complex experiment in enkelvoud, dan is bij n factoren het aantal objecten 2^n, het vertegenwoordigt 2^n vrijheidsgraden. Het aantal main effecten en interacties wordt, zooals we zagen in Hoofdstuk IV, gegeven door de som van binomiaal coefficienten, is dus 2^n-1.

Het aantal coefficienten van de regressievergelijking, die dit materiaal beschrijft:
$y = aA + bB + ...kAB + lAC + ...pABC + ...c = 2^n$. Directe beschrijving van het materiaal met behulp van de regressievergelijking is zeer inefficient, het vereischt het oplossen van 2^n vergelijkingen met 2^n onbekenden. Tevens vinden we dan geen middelbare fout voor deze coefficienten, aangezien het aantal onbekenden even groot is als het aantal waarnemingsvergelijkingen.

We laten daarom een analysis of variance voorafgaan, deze is veel minder bewerkelijk, terwijl een schatting van de middelbare fout gemaakt kan worden, Dit laatste biedt nog eenige moeilijkheden, daar in de analysis of variance alle varianties slechts één vrijheidsgraad hebben. Het vergelijk van dergelijke varianties levert groote moeilijkheden op, daar dan voor een verschil met zekere betrouwbaarheid een zeer groote Z of F waarde vereischt is [1]).

Bij 95% en 99% kansgrens is F respectievelijk 161 en 4052. Bij 2 vrijheidsgraden voor de toevalsvariance respectievelijk 18,5 en 98,5 enz. We moeten daarom een aantal varianties gaan samenvatten, om een toevalsvariance te krijgen. Daaraan vooraf moet gaan een test op de homogeniteit van de samen te vatten varianties. Wishart [2]) bespreekt er twee, waarbij hij de voorkeur geeft aan de test van Brandt, die een soort ruwe χ^2 waarde bepaalt.

Bij k gegeven varianties s_r^2 met n_r vrijheidsgraden, berekenen we deze "crude value" van χ^2 als volgt:

---

[1]) F is de directe verhouding der varianties, hetgeen gemakkelijker is. F wordt gedefinieerd door $F = e^{2z}$

[2]) J.Wishart. Field Experiments of factorial design. Journ.Agr.Science 1938, pg.299-306.

als S^2 is de samengevatte variance (verkregen door optellen van de sommen van quadraten, resp. vrijheidsgraden) met $n = \Sigma n_r$ vrijheidsgraden. Als deze waarde van χ^2 die k-1 vrijheidsgraden heeft, betrouwbaar is, berekene men een gecorrigeerde kleinere χ^2 waarde, door te delen door

$$C = 1 + \frac{1}{3(k-1)}\left(\Sigma \frac{1}{n_r} - \frac{1}{n}\right)$$

en teste opnieuw.

(Bartlett veronderstelt echter, dat deze nieuwe waarde de neiging heeft de anders overschatte beteekenis van de ruwe χ^2 waarde te flatteeren, d.w.z. men vindt dan een te kleine waarde, waardoor men een groep variances als homogeen zou beschouwen, zonder dat ze het zijn.). Hiermede valt dus min of meer te beoordeelen of de genoemde variances al of niet samengevat kunnen worden tot een nieuwe, de toevalsvariance.

Na dit onderzoek met de analysis of variance kan eventueel de gereduceerde Ω-regressievergelijking opgesteld worden.

De uitvoering van het 2x2x2x.... experiment stelt ons echter bij veel factoren nog voor eenige moeilijkheden bij een niet homogenen achtergrond van het experiment. Aanleg in herhalingen kan deze moeilijkheden niet ten volle overwinnen. Een mogelijkheid is de aanleg als "split-plot design" (blokken, sub-blokken, sub-sub-blokken enz.). Confouding (voor interacties) kan soms ook bruikbaar zijn, het is bedoeld als vereffening van een heterogenen achtergrond van het experiment op grond van het a priori = 0 stellen van interacties van hoogere orde, alle combinaties worden echter uitgevoerd !

Zijn deze confounded interacties niet de hoogste, dan kan de methode bruikbaar voor ons doel zijn, als de hoogste interacties juist reëel blijken te zijn, in een zoodanige combinatie, dat op grond daarvan een splitsing in groepen mogelijk is. Zijn de interacties die van een hoogere orde zijn, dan de confounded interacties, niet reeel, dan is het mogelijk, dat de interacties die juist "unconfounded" zijn, wel reeel waren, de mogelijkheid van splitsing in groepen hangt dan nog af van de eventueele interacties van lagere orde. Het is dus van te voren niet te zeggen of het toepasbaar is of niet, zoodat we het niet gebruiken kunnen.

Bij partial confounding zouden we deze moeilijkheden niet hebben, doch dit is niet mogelijk, als de factoren alle in 2 trappen onderzocht worden.

Inmiddels kan nog opgemerkt worden, dat het pan-parallelle experiment ook in zekeren zin confounding is: sommige interacties worden = 0 gesteld. In beide gevallen is daardoor vereffening van den heterogenen achtergrond van het experiment mogelijk. Het pan-parallelle experiment heeft daarbij het groote voordeel, dat het aantal objecten veel kleiner is. Bij partial-confounding daarentegen zijn alle interacties echter weer in zekere mate te testen en zoodoende in het voordeel tegenover het pan-parallelle experiment, hoewel dit natuurlijk bij veel factoren spaak loopt, om dezelfde redenen als het complete complex experiment spaak loopt bij veel factoren.

Voeren we het 2x2x2x.... experiment uit als incomplete complex experiment, dan is hierbij correctie op de niet-homogeniteit van den achtergrond zonder meer mogelijk. Hierbij zouden we dadelijk kunnen vereffenen met de gereduceerde regressievergelijking in overeenstemming met de = 0 gestelde interacties, doch beter is het eerst nog de analysis of variance uit te voeren om de regressievergelijking verder te reduceeren.

SAMENVATTING.

In het methodologische gedeelte van zijn philosophie der landbouwwetenschap betoogt Krzymowski de onbereikbaarheid van de huidige experimenteele methode. Hij verklaart deze omstandigheid uit het feit, dat de landbouwwetenschap nog steeds op een eng rationalistisch standpunt staat en uit het feit, dat in de historische ontwikkeling van de landbouwwetenschap de opkomst van deze wetenschap samenviel met een groote bloeiperiode van de natuurwetenschappen. Vooral het overgenomen ceteris paribus principe, in het bijzonder de methode der natuurkunde, is ongeschikt, te eenvoudig, voor de landbouwwetenschap; het is onvoldoende, noch om de vele factoren, die de verschijnselen in den landbouw beheerschen, te analyseeren, noch laat zich omgekeerd, uit de aldus verkregen gegevens, de werkelijkheid reconstrueeren met al zijn wisselwerkingen tusschen de verschillende factoren. Het ideaal van de beheersching der natuur blijft zoo onbereikbaar. Krzymowski zoekt een uitweg in de niet-experimenteele methode i.c. de empirische waarnemingen en de practijk, de overlevering enz.

Na op en dergelijken ontwikkelingsgang in de biologie gewezen te hebben, die tot een zgn. synthetische methode voerde, waarbij de zgn. causale structuren in hun geheel beschouwd worden, in plaats van de factoren en hun wisselwerkingen afzonderlijk, wat onmogelijk geacht werd, wordt door schrijver getracht in dit geschrift het probleem direct op te lossen, door de analytische methode als het ware te verfijnen. En wel door het ceteris paribus principe te verlaten en te vervangen door het zgn. factorial design, waarmede Fisher voor enkele factoren het spel der wisselwerkingen wel kon analyseeren.

Dit wordt nu uitgewerkt voor een onbepaald aantal factoren, waarbij blijkt, dat voor zeer veel factoren dit factorial design een onbereikbaar, een onuitvoerbaar ideaal is, hoewel theoretisch het probleem er geheel door bepaald is. De wiskundige formuleering, die hierdoor echter verkregen is, maakt een benaderende oplossing met gegeven graad van inrichtingen mogelijk, waarbij de onderste grens die inrichtingsgraad is, die met het verlaten ceteris paribus principe bereikt wordt. De hierbij behoorende practische oplossing wordt uitvoerig aangegeven, met inbegrip van het geval van een niet-homogenen achtergrond van het experiment, waarvoor een op een minimum voorwaarde berustend correctiesysteem is opgesteld. Naar schrijvers meening worden hiermede de door Krz. opgesomde bezwaren tegen de experimenteele methode grootendeels geëlimineerd, terwijl veel grootere resultaten bereikt kunnen worden.

Het besproken factorial design met zeer vele factoren geeft verder aanleiding tot beschouwingen over een grens van ons natuurwetenschappelijk kenvermogen, en over het optreden van tot nu toe onbekende types van reactiecurven, in verband waarmede de inductieve methode in de natuurwetenschappen beschouwd worden. Met het thans volgende inhoudsoverzicht, kan men zich een weg zoeken in het geschrift.

OVERZICHT.

Hoofdstuk I (Krzymowski philosophie der landbouwwetenschap). De philosophie der landbouwwetenschap moet opgevat worden als een zgn. vakphilosophie. Wat deze vakphilosophie in het bijzonder inhoudt, wordt uiteengezet in § 2. Hierin wordt in het kort gewezen op de veel verder ontwikkelde vakphilosophieën van andere wetenschappen; verder wordt weergegeven hoe Krz. enkele tot nog toe als vrijwel vaststaande beschouwde principes aan twijfel onderwerpt, om zoo als het ware tot een aanloop ter rechtvaardiging van een vakphilosophie, waarin dergelijke principes critisch behandeld worden, te komen. Ten slotte wordt ter orientatie de inhoud van Krz. werk weergegeven. Het is hierbij niet onverschillig hoe men zich tegenover de philosophie in het algemeen stelt. In § 1 wordt daarom onderscheid gemaakt in een oudere, Europeesche philosophie en een nieuwere, Amerikaansche philosophie (James, Dewey), waarbij opgemerkt wordt dat Krz. philosophie der landbouwwetenschap van eerstgenoemd standpunt uitgaat, wat bv. al een van de redenen is, waar-

waarom genoemde philosophie van Krz. onvolledig moet zijn, waarop in § 3 nader gewezen wordt. Ook liggen aan dit onderscheid en het daarmee samenhangende hiaat, de speculatieve beschouwingen van Hoofdstuk VI (§ 5, over het optreden van onbekende typen van opbrengstcurven en § 6, de tragiek van het spel der factoren) ten grondslag (waarbij men een parallel moet zoeken met het radicaalempirisme van James). Wanneer men nu geneigd is, het bestaansrecht van een philosophie der landbouwwetenschap te erkennen, zoo zal men, zooals in § 3 vermeld wordt, tot de ontdekking komen, dat over dit onderwerp nog slechts weinig gegeven is. Er wordt om daar vermelde redenen aanknooping gezocht met het geschrift van Krz. Dit blijkt zeer onvolledig te zijn, terwijl een aantal problemen, die daar in drie opeenvolgende hoofdstukken behandeld worden, onopgelost gebleven zijn. Hier een oplossing voor te zoeken en zoodoende een bijdrage te leveren tot de philosophie der landbouwwetenschap, is het hoofddoel van dit geschrift. In § 4 worden deze hoofdstukken kort weergegeven om een overzicht van de problemen te verkrijgen. In het eerste geciteerde hoofdstuk (Het rationalisme in de landbouwwetenschap en het factorengewirwar) wordt gewezen op de zelden naar voren gebrachte verwantschap tusschen het philosophisch rationalisme en de rationalistische richting der landbouwwetenschap. De oorzaak van deze over het hoofd geziene betrekking zoekt Krz. in het feit, dat de huidige landbouwwetenschap nog steeds rationalistisch is. Vervolgens wordt door Krz. critiek op deze eenzijdige instelling uitgeoefend. We zullen de bezwaren hier niet alle noemen, gedeeltelijk zijn ze van denzelfden aard, als de bezwaren die het rationalisme in de philosophie aangemerkt worden, doch slechts vermelden, dat van uit practisch standpunt gezien, de bezwaren culmineeren in de onoplosbaarheid van het door Krz. zoo genoemde factorengewirwar. Kort gezegd verstaat hij hieronder, dat het zeer groote aantal verschillende factoren, dat op een landbouwkundig verschijnsel invloed heeft, onmogelijk alle te analyseeren zijn, en zoo deze factoren en hun werking toch bekend waren, het onmogelijk is, deze factoren alle tot op vrij groote nauwkeurigheid te meten, noch hun wederkeerigen invloed vast te stellen, laat staan het resultaat van dit spel der factoren, zooals schrijver dit liever noemt, vooruit berekenen. Aan de hand van eenige concrete voorbeelden wordt dit nader besproken. In verband hiermede bespreekt Krz. de waarde der overlevering, de plaatselijke gebruiken enz. In het tweede geciteerde hoofdstuk (Het experimenteele onderzoek in den landbouw) wordt het feit becritiseerd, dat de groote beteekenis van het experiment in de natuurwetenschappen er toe geleid heeft, de landbouwwetenschap als bijna uitsluitend een experimenteele wetenschap te zien. Krz. wijst op de grenzen, waaraan de experimenteele methode, waarbij men tal van factoren afzonderlijk onder overigens gelijke omstandigheden (ceteris paribus) onderzoekt, zonder op hun wederkeerige beinvloeding te letten - welk laatste trouwens door Krz. onmogelijk geacht wordt - gebonden is. Deze beschouwingen voeren weer tot het probleem van het spel der factoren. De overschatting van de experimenteele methode wijt Krz. aan de rationalistische instelling van de landbouwwetenschap. Verder wordt het feit besproken, dat de vele zgn. exacte experimenten geen ander resultaat hebben, dan dat ze de regels van de practijk bevestigen. In het derde geciteerde hoofdstuk (Het niet-experimenteele onderzoek in den landbouw) wordt, gezien het feit dat de experimenteele methode niet bij machte is, het spel der factoren te beheerschen, nader ingegaan op de empirische waarnemingen in de practijk.

Hoofdstuk II (Inleidende probleemstelling). In § 1 wordt er op gewezen, dat het onopgelost probbleem van het spel der factoren een overheerschende rol speelt in genoemde hoofdstukken, terwijl vele moeilijkheden, die hiermee samenhangen door Krz. teruggevoerd worden op een door hem bestreden rationalisme in den landbouw. De hoofdopgave van dit geschrift is daarom, dit probleem op te lossen, en indien dit slechts gedeeltelijk mogelijk blijkt, de grenzen der te geven methode nauwkeurig aan te geven, om de waarde van het rationalisme in de landbouwwetenschap scherper te kunnen beoordeelen (Vergelijk in verband hiermede Hoofdstuk VII § 1 Over de grenzen van het

natuurwetenschappelijk kenvermogen). In § 2 wordt van uit kentheoretisch standpunt (neo-kantianisme) de wetenschappen gekarakteriseerd, waarbij de natuurwetenschappen met als centrale categorieën causaliteit en meetbaarheid in het bijzonder behandeld worden. Daarna wordt de gevraagde methode gequalificeerd als een natuurwetenschappelijke. Op de groote beperking hieraan verbonden wordt gewezen, waarna besproken wordt, hoe bij deze omschrijving van het begrip natuurwetenschap toch het overgroote deel der huidige landbouwwetenschap hier onder valt. Vervolgens wordt op de verschillende interpretaties van het causaliteitsbeginsel gewezen. Om de methode nog verder af te bakenen, wordt ook de meetbaarheid besproken, aangezien zonder dit begrip ook een natuurwetenschappelijke oplossing van het probleem van het spel der factoren mogelijk is, volgens een zgn. synthetische methode (in het bijzonder in de biologie ontwikkelt met behulp van de begrippen amboceptor-, polyceptor- en pantoceptorverband, daar men de analyse der wisselwerkingen onmogelijk achtte). Het natuurwetenschappelijke van de methode staat een formeel wiskundige behandeling van het spel der factoren toe en wel met behulp van het "factorial design", dat het volle gewicht legt juist op het onderzoek der wisselwerkingen, waarmee dus a.h.w. getracht wordt de analytische methode op een hooger plan te brengen.

Hoofdstuk III (Ceteris paribus en factorial design). Reeds werd op de ontoereikendheid van de proefplannen volgens het ceteris paribus principe gewezen. Prof.Fisher stelt hier het factorial design voor in de plaats: men moet een factor onderzoeken bij een aantal trappen van een andere factor, waardoor de wisselwerking tusschen beide factoren bekend wordt. Bij meer factoren worden alle combinaties onderzocht. De voordeelen van het factorial design zijn zoodoende:

1e. haar efficiency (in verband met de verborgen herhalingen);
2e. haar volledigheid (in verband met het feit dat alle interacties onderzocht worden);
3e. haar geschiktheid als inductieve basis voor generalisatie, hetgeen uitvoerig besproken wordt in § 1.

In § 2 wordt de afbeelding van deze interacties besproken, waartegen bij veel factoren bezwaren ontstaan. In § 3 wordt de wiskundige formuleering besproken; bij directe formuleering (main effects en interacties) ontstaan ook hier onoverkomenlijke bezwaren bij veel factoren in meer dan twee trappen, waarna overgegaan wordt op een indirecte formuleering met behulp van hun variances. In § 4 zijn door schrijver modellen (d.w.z. de toevallige afwijkingen zijn uitgeschakeld) van het spel der factoren ontworpen, waarbij voor 2 en 3 factoren een overzicht gegeven wordt van de respectievelijk 8 en 128 verschillende resultaten dieeen analysis of variance kan hebben, als elke factor in slechts 2 trappen onderzocht wordt. Daarbij blijkt tevens, dat het spel der factoren onvoldoende gekarakteriseerd wordt door de main-effects en interacties, die slechts het uiteindelijk effect vertolken. Daarom werden door schrijver twee nieuwe begrippen: main-action en interplay ingevoerd, die vooral de reacties karakteriseeren, die tot de genoemde effecten leiden. In § 5 worden zij nader geformuleerd. (Verder onderscheidt schrijver nog: analysis of effects, analysis of actions en interacting). Hiermede is het spel der factoren volledig en overzichtelijk te beschrijven, hetgeen weer aan eenige modellen geillustreerd wordt.

Hoofdstuk IV (Voortgezette probleemstelling). Hierin wordt het spel der factoren door schrijver teruggebracht op een factorial design met een oneindig aantal (practischer: met zeer veel) factoren, waardoor tevens min of meer reeds een wiskundige formuleering van het probleem gegeven is. De moeilijkheid wordt dan verplaatst naar het groote aantal interacties en het groote aantal objecten (combinaties van trappen van 2 of meer factoren) dat in de millioenen kan loopen. De nieuwe opgave is dus het aantal interacties en objecten te verminderen

bij gelijkblijvend aantal factoren bij een minimum van nieuwe voorwaarden. Het is daartoe noodig na te gaan, welke waarden de variances van hoogere orde zullen aannemen.

Hoofdstuk V (Interpretatie van het factorial design). Aan de reeds genoemde modellen wordt in § 1 geillustreerd dat de waarden der variances afhankelijk zijn:

1e. van de coefficienten van de zgn. regressievergelijking, die het spel der factoren kan beschrijven;

2e. van het interval waarin deze vergelijking onderzocht wordt.

In § 2 worden daarom een aantal gebruikelijke regressiefuncties besproken. Deze blijken echter alle ongeschikt te zijn, indien men met veel factoren te doen heeft (resp. variabelen in de regressievergelijking). Daarom wordt in § 3 door schrijver de zgn. φ-regressiefuncties opgesteld, die zich onderscheiden van de gebruikelijke, door eenige beperkende voorwaarden en door hun symbolische schrijfwijze. In § 4 worden, eerst nog voor 2 factoren, met behulp van de door Fisher opgestelde identiteit voor de sommen van quadraten, de variances uitgedrukt in de symbolisch geschreven termen van deze φ-regressiefuncties. Hierdoor is het mogelijk, uit de resultaten van elke analysis of variance, een dergelijke φ-regressiefunctie op te stellen voor evenveel variabelen, als factoren in de analysis of variance beschouwd worden, terwijl omgekeerd uit elke zoodanige regressiefunctie, de resultaten van een eventueele analysis of variance voorspeld kunnen worden. Voor beide afleidingen is een overzicht opgesteld, die alle mogelijke gevallen weergeven. Tevens wordt dit nog nader toegelicht met de nieuwe begrippen main-action en interplay. In § 5 wordt de symbolische schrijfwijze van de φ-regressiefuncties gewijzigd, door concrete vormen in te voeren, die voldoen aan de gestelde beperkingen als genoemd in §,3, zoodat de variances uitgedrukt worden in de constanten van deze nieuwe regressievergelijking en in de sommatievormen van de variabelen, m.a.w. we zijn thans langs een algemeene bewijstrant teruggekomen op de twee voorwaarden, die reeds in § 1 genoemd werden, maar thans met dit groote voordeel, dat we met deze nieuw geformuleerde voorwaarden, de waarden die de variances t.o.v. elkaar kunnen aannemen, kunnen bepalen. Hiervoor is nog noodig de sommatievormen anders te schrijven. We nemen daarbij voorloopig nog aan, dat het factorenspel c.q. de regressiefunctie onderzocht wordt in trappen gelijk aan een geheele waarde van het argument. De waarden der sommatievormen zijn dan afhankelijk van de 1e trap en 2e, van het aantal trappen, welkerafhankelijkheid in §6 uitgewerkt wordt, door de sommatievormen met behulp van de formules voor de reeksen van hoogere orde (of met de veeltermen van Bernoulli) in deze twee grootheden uit te drukken, waarmede tevens het interval waarin de regressiefunctie beschouwd wordt, omschreven is. Vervolgens wordt in § 7 eerst nog het gedrag van elke variance afzonderlijk gegeven, nu dus uitgedrukt in de constanten van de regressie functie en in de 1e trap en het aantal trappen waarin deze regressiefunctie onderzocht wordt, en in § 8 wordt hun onderlinge verhouding bepaald. We kunnen dan nauwkeurig voor elke bepaalde regressiefunctie aangeven, voor welk interval de interactievariance kleiner, gelijk of grooter is dan de variance voor één der factoren. Bij een bilineaire regressievergelijking blijkt de interactievariance bijna steeds het kleinst te zijn, hetgeen in § 9 aan een getallenvoorbeeld geillustreerd wordt. In § 10 wordt het onderzoek herhaald bij een regressievergelijking van hoogeren graad. In eenbepaald voorbeeld blijkt thans, dat de interactievariance daar juist bijna altijd grooter is, dan de variance voor één der factoren, wat in § 11 aan een getallenvoorbeeld geillustreerd wordt. Hierna wordt in § 12 en § 13 het geheel nog eens herhaald voor 3 factoren, om zoo tot uitbreiding tot n factoren te kunnen komen, waarbij blijken zal (Hoofdstuk VII, § 1) dat er gevallen zijn te construeeren, waarbij de totale variance steeds

gelijk is aan de variance van de interactie van de hoogste orde Dit compliceert het probleem ten zeerste, daar thans niet onder alle omstandigheden op het onbelangrijk worden der interacties van hoogere orde gerekend mag worden. Tevens blijkt hieruit, dat deze niet altijd kunnen dienen als toevalsvariance, zie § 14. Tot slot wordt de beperking in § 5 betreffende geheele argumentwaarden opgeheven, door den gang van het bewijs in het kort na te gaan voor willekeurige waarden van het argument van de regressiefunctie c.q. voor niet-aequidistante ordinaten. Waarna in § 16 nogmaals drie voorbeelden gegeven worden, waarin alleen het behandelde in § 4 uitgewerkt wordt, daar dit in het bijzonder van belang is voor hoofdstuk VII.

Hoofdstuk VI (De regressiefuncties in de natuur). Om thans het verband met landbouwkundige experimenten weer op te nemen, gaan we over van de algemeene regressiefunctie, waarbij het argument loopt van $-\infty$ tot $+\infty$, tot reactiecurven c.q. opbrengstcurven (Mitcherlichcurve, logistische curve e.a.) welke in § 1 kort worden weergegeven. Ze blijken echter niet geschikt te zijn om de in Hoofdstuk V door schrijver ontwikkelde theorie op toe te passen, daar ze alle slechts voor een factor, een variabele opgesteld zijn. In § 2 wordt deze beperktheid nader onder oogen gezien, door de historische ontwikkeling van deze opbrengstcurven na te gaan; ook de opbrengstwetten in bv. de economie blijken aan deze beperktheid te lijden. Om een formuleering voor meerdere factoren te vinden, wordt in § 3 eerst nagegaan, hoe men tot een formuleering voor een factor komt, aan de hand van een schema van Pearl, die eerst, al naar gelang waarvoor de curve dienen zal, een aantal eischen opstelt en deze vervolgens in de formuleering verwerkt. In § 4 wordt dan de formuleering van opbrengstcurven voor meer veranderlijken onder oogen gezien, zonder tot een belangrijk resultaat te komen, weshalve we in het volgende ons houden aan de φ-regressiefuncties. Echter wordt er op gewezen, dat al hadden we een geschikte formuleering gevonden, en hierop de theorie uit voorgaand hoofdstuk ontwikkeld, ons dit nog geenszins zekerheid verschaft over een eventueel algemeen wetmatig gedrag van de variances van interacties van hoogere orde, omdat we nog slechts zeer slecht ingelicht zijn over de mogelijke vormen, waarin opbrengstcurven c.q. reactiecurven zich kunnen voordoen, waarmede de inleiding tot eenige speculatieve beschouwingen van schrijver in § 5 en § 6 gegeven is. In § 5 wordt verband gelegd tusschen de interacties der factoren en de frequenties waarmee in de natuur de verschillende objecten (combinaties van trappen van 2 of meer factoren) optreden; doordat deze frequenties zich wetmatig uiten in een frequentieverdeeling (correlatieschema) wordt het bestaan van tot nog toe onbekende typen van reactiecurven waarschijnlijk gemaakt. In § 6 wordt getracht de kans hierop nader te schatten; deze blijkt zeer klein te worden, waarom zooals daar nader aangegeven wordt, we nooit een afwijkend type, dat theoretisch bestaanbaar is vinden, door telkens slechts een factor te onderzoeken volgens het ceteris paribus principe, terwijl in deze kleine kans tevens het bestaansrecht van de inductieve methode in de natuurwetenschappen ligt, en de mogelijkheid van het bestaan van zgn. natuurwetten. Omgekeerd, willen we iets nieuws vinden, dan moeten we deze correlaties "doorbreken" door de experimenten volgens het factorial design uit te voeren, in plaats van volgens het oude ceteris paribus principe. Hoe meer factoren we opnemen in het factorial design, hoe grooter de kans iets nieuws te vinden, het tragische ligt nu juist in het feit, dat het aantal objecten dan zoo groot wordt, dat het onmogelijk wordt het experiment uit te voeren, (voor landbouwkundige proeven wordt bij 20 factoren het aardoppervlak reeds te klein, zooals in Hoofdstuk IV becijferd werd). Tot slot wordt er op gewezen, dat hier weer de tegenstelling waarop in Hoofdstuk I § 1 reeds gewezen werd tot uiting komt. Deze beschouwingen, die parallel loopen

aan het radicaal-empirisme zijn gegrond op overwegingen die verband houden met het factorial design, hetwelk echter ook dienen kan voor beschouwingen over het standpunt dat in Hoofdstuk II, § 2 genoemd is, waarover in de volgende §.

Hoofdstuk VII (Het pan-parallelle experiment). Waar in hoofdstuk V reeds een voorbeeld gegeven werd, waarbij de interactievariance bij 2 factoren bijna altijd grooter is, dan de variance voor één der factoren, wordt hier een model gegeven, waarbij de variance van de interactie van de hoogste orde altijd gelijk is aan de totale variance, hoe groot ook het aantal factoren wordt. In verband hiermede is een grens aan het natuurwetenschappelijk kenvermogen op te stellen. Met de vraag of een dergelijk model in de natuur voor kan komen, zijn we weer bij de voorgaande §§ terug. In verband hiermede wordt nog het factorial design van twee kanten beschouwd, zgn. in de diepte met interacties van steeds hoogere orde en in de breedte met negeeren met steeds hooger aantal factoren met behulp van zgn. sub-factorial designs (waarover in de volgende §). De eerste zienswijze, speciaal van belang voor het nieuwere bovengenoemde standpunt, is aan gegeven grenzen gebonden en eischt uitvoering van het volledige factorial design; de tweede zienswijze, van belang voor het oudere standpunt, eischt een nieuwen proefopzet door schrijver ontworpen, nl. het zgn. pan-parallelle experiment, hetwelk in de volgende §§ wordt besproken. § 2 Gezien de onmogelijkheid om een volledig factorial design met veel factoren uit te voeren, worden op bepaalde wijze interacties van hooge orde verwaarloosd, waardoor het aantal objecten sterk daalt, doordat men zekere groepen van combinaties van factoren volgens een volledig factorial design (complex experiment) onderzoekt in zgn. sub-factorial designs (sub-complex exp.), die samengevat in een volledig stelsel, dat alle factoren gelijk recht doet wedervaren, het incomplete factorial design (incomplete complex experiment, pan-parallelle experiment) genoemd wordt. Bij de verwaarloozing van deze interacties moet men zich echter aan zeer bepaalde regels houden, die op de resultaten van Hoofdstuk V gegrond zijn. In § 3 en 4 worden eerst uitvoerig de waarden besproken, die men de rest der factoren, die niet in een bepaald sub-factorial design onderzocht worden, geven moet; daarna worden besproken de herhalingen van enkele objecten, die optreden tengevolge van deze groepeering; deze kunnen weer gebruikt worden voor correctie op een eventueele dishomogeniteit van den achtergrond van het experiment, waarvoor een minimum voorwaarde opgesteld en uitgewerkt wordt, zoodanig dat alle objecten van een bepaald sub-factorial design (sub-complex experiment) dezelfde correctie krijgen, waardoor de waarden der interacties zich niet wijzigen. Daarna worden deze interacties aan een toevalsvariance getoetst, die gebaseerd is op de objecten in veelvoud na correctie, met behulp waarvan, volgens de gegeven schema's uit Hoofdstuk V (en de uitbreidingen daarvan) de φ-regressiefunctie opgesteld kan worden. Na vereffening hiervan kunnen alle niet onderzochte objecten van het volledige factorial design berekend worden. Vervolgens wordt in § 5 nagegaan wat de practische waarde van deze incomplete factorial designs is, door de reductie van het aantal objecten bij gegeven aantal factoren en verwaarloosde interacties na te gaan in uitvoerige tabellen, die opgesteld zijn voor experimenten, die volgens het volledig factorial design 60 millioen objecten eischen, waarna nog een beschouwing gegeven wordt over het tegemoet komen van dit incomplete factorial design aan het optreden van de buitenwetmatige verschijnselen c.q. reactiecurven. Tot slot wordt in § 6 getracht de benadering tot het onbereikbare ideaal van het volledige factorial design door het incomplete factorial design nog hooger op te voeren, door het splitsen van de factoren in onafhankelijke groepen met behulp van het 2^n experiment, waarbij besproken wordt in hoeverre hierbij confounding is toegestaan. Tevens wordt nagegaan in hoeverre confounding en partial confounding een rol spelen in het incomplete factorial design.