BBM
B. 69
2)

VERHANDELINGEN VAN HET
INSTITUUT VOOR PRAEVENTIEVE GENEESKUNDE

III

DE GEDACHTENGANG VAN DE STATISTICA

DOOR

Prof. Dr S. T. BOK

HOOFD VAN DE AFDELING STATISTIEK VAN HET
INSTITUUT VOOR PRAEVENTIEVE GENEESKUNDE
TE LEIDEN

TWEEDE DRUK

1948
H. E. STENFERT KROESE'S UITGEVERS-MAATSCHAPPIJ N.V.
LEIDEN

Dit boek is een uitgave van het Instituut voor Praeventieve Geneeskunde, Wassenaarsche weg 56, Leiden.

Volgens zijn statuten stelt dit Instituut zich ten doel het bevorderen van de praeventieve geneeskunde in het algemeen en van de arbeidshygiëne in het bijzonder. Dit doel vereischt de samenwerking met zeer veel instanties en personen. Vooral in ons land speelt het particulier initiatief op dit gebied een voorname rol: mannen en vrouwen uit alle rangen en standen, uit kringen van de geneeskunde zoowel als van de industrie, de handel, de administratie en zoo veel meer, van werkgevers zoowel als werknemers, besteden een groot deel van hun krachten en van hun vrije tijd ten bate van het lichamelijk en geestelijk welzijn hunner medemensen.

Deze werkers wensen de gegevens, die zij hebben vergaard, gaarne te algemenen nutte dienstbaar te maken en daartoe is in den regel een statistische bewerking nodig. Het Instituut voor Praeventieve Geneeskunde is steeds bereid, daarbij behulpzaam te zijn en dit boekje vormt een eerste stap in deze richting. Velen toch, die zich op praeventief geneeskundig terrein zo nuttig maken, weten niet wat de wetenschappelijk gevoerde statistiek vermag en welke eisen aan de zo te bewerken gegevens moeten worden gesteld. De uiteenzetting van Prof. Bok kan worden beschouwd als een leiddraad voor hen, die, om hieraan tegemoet te komen, zich een inzicht willen verwerven in de gedachtengang, die aan de statistica ten grondslag ligt. Zij kan zonder bijzondere wiskundige ontwikkeling worden begrepen.

Moge dit boekje echter ook velen, die op andere terreinen van wetenschap en praktijk werkzaam zijn, zulk een inzicht bijbrengen!

Wie het boek ter hand neemt bedenke evenwel, dat het niet als een romannetje bedoeld is maar als een werk, dat moet worden bestudeerd. En studeren vereist geestelijke inspanning.

De Directeur
van het Instituut voor Praeventieve Geneeskunde,

J. P. Bijl

INHOUD

blz.

Voorwoord
Inleiding 1
Statistiek en statistica 5
Telling van het gehalte in een steekproef 8
Intermezzo: telling in een kleine steekproef 10
Het gehalte in een grote steekproef 12
Meetfouten 23
De standaardfout van de enkele meting. 26
De standaardfout van een gemiddelde als beste schatting. . . 30
Standaardfout en betrouwbaarheidsgrenzen 32
Simultane kansdistributies 34
Steekproef uit meer dan twee soorten individuen 42
Het frequentie-vierkant 45
Zijn twee afmetingen, die elk enige malen werden gemeten, in werkelijkheid verschillend van grootte? 48
De verhouding tussen twee strooiingen 51
Zijn enige afmetingen, die elk enkele malen werden gemeten, in werkelijkheid uiteenlopend van grootte? De variantie-analyse 55
Meervoudige variantie-analyse 62
Correlatie. 71
Regressie 75
Rechtlijnige regressie 80
De statistische bewerking van historische gegevens 93
De aard van de statistische conclusie 101
De draagwijdte van een statistische conclusie 109
De meest economische opzet van een onderzoek, waarvan de uitkomsten later statistisch moeten worden onderzocht 120
Het nut van een globale kennis van de statistica 125
Appendix (De becijfering van enkele der besproken vraagstukken). 127
De betekenis van de in dit boek gebruikte **symbolen** 146
Overzicht van de in dit boek gebruikte **formules** 148
De in dit boek gebruikte **nomogrammen**. 150

VOORWOORD BIJ DE EERSTE DRUK

Deze korte schets van de gedachtengang, die ten grondslag ligt aan de werkmethoden van de statistiek, is de inhoud van enkele voordrachten, door den schrijver gehouden voor de wetenschappelijke en praktische werkers van het Instituut voor Praeventieve Geneeskunde te Leiden.

Nadat in dat Instituut een afdeling voor de statistiek was opgericht bleek n.l. al spoedig, dat de werkers der andere afdelingen daar meer profijt van zouden kunnen trekken als zij enig inzicht in de hier bedoelde gedachtengang zouden hebben: zij zouden dan beter in staat zijn op de juiste momenten de hulp van de statistiek voor hun eigen onderzoekingen in te roepen.

Doordat de kennis omtrent het exact statistisch bewerken van getallen, die door meten of tellen zijn verkregen, bij zeer veel onderzoekers op de meest uiteenlopende terreinen van wetenschap en praktijk uiterst gering is, blijven deze zo nuttige bewerkingen in heel veel gevallen achterwege, tot schade zowel van die onderzoekers zelf als van hun resultaten. Op grond daarvan werd gemeend, dat publicatie dezer voordrachten ook voor velen buiten genoemd Instituut van nut zou kunnen zijn.

Dit boek is dus bedoeld voor werkers op de meest uiteenlopende terreinen van wetenschap en praktijk. Het richt zich niet alleen tot artsen, *hoewel de meeste der voorbeelden, die er in verwerkt zijn, aan de geneeskunde ontleend werden.*

De schrijver spreekt hier graag zijn dank uit voor de vele hulp, die hij van zijn assistent, den Heer Ch. A. G. Nass, heeft ontvangen en voor de goede raadgevingen van Dr. J. B. D. Derksen, hoofdcommies aan het Centraal Bureau voor de Statistiek te 's-Gravenhage en privaat-docent aan de Rijks Universiteit te Leiden (na het ter perse gaan van dit boekje benoemd tot hoogleraar te Rotterdam), die zo vriendelijk was het manuscript door te lezen.

Een deel van dit boekje is geschreven in een gevangeniscel van het Politiebureau aan het Haagse Veer te Rotterdam. De schrijver is vervuld van bijzonder hartelijke dankbaarheid ten opzichte van den Directeur dier gevangenis, den Heer J. J. Swart, die hem daartoe in staat stelde lijnrecht tegen de orders van de bezetters in, en ten aanzien van den chef van het politielaboratorium aldaar, den helaas later waarschijnlijk in gevangenschap gevallen Hoofdinspecteur J. V. Tas, die hem toestond af en toe in dat laboratorium te werken. Zij beiden hebben niet alleen hem maar zeer velen in een moeilijke periode op buitengewone wijze krachtdadig bijgestaan.

Leiden, November 1945. *S. T. Bok*

BIBLIOTHEEK - NEDERLANDS INSTITUUT
VOOR PRAEVENTIEVE GENEESKUNDE
WASSENAARSEWEG 56 - LEIDEN

INLEIDING

Statistisch onderzoek kan bij zeer veel onderzoekingen op de meest uiteenlopende terreinen van wetenschap en praktijk belangrijk nut afwerpen. Bijzonder talrijk toch zijn de onderzoekingen waarbij het bereikte eindoordeel onder meer steunt op getallen, die door meting of steekproeftelling waren verkregen. Dit geldt niet alleen voor de natuurwetenschappen en hun praktische toepassingen, maar ook voor de geesteswetenschappen, waar immers ook steeds meer onderzoekers hun inzicht mede door tellen of meten trachten te verdiepen en te versterken. Voor al deze onderzoekingen is het in beginsel onontbeerlijk en vaak van het grootste nut, de beschouwde getallen op correcte wijze volgens exacte statistische methoden te bewerken, zowel om tot een zo betrouwbaar mogelijke conclusie te komen als ook om een inzicht te krijgen in de *mate van betrouwbaarheid* dier conclusie.

Het is echter opvallend, dat de meeste onderzoekers hun conclusie uit meet- of telgetallen trekken zonder er zulk een exacte statistische analyse op toe te passen. En dat moet in hoge mate worden betreurd omdat zij daarmee telkens de gelegenheid tot het verkrijgen van — vaak niet eens vermoede — voordelen ongebruikt laten voorbijgaan.

Voor een deel vindt dit verschijnsel zijn verklaring in het feit, dat men in vele gevallen wel kan vermoeden, welke uitkomst zulk een exacte analyse zou geven. Daarmee laat men die analyse echter niet alleen in deze speciale gevallen achterwege, maar helaas ook al gauw in omstandigheden, waarin de uitkomst niet zo gemakkelijk kan worden voorspeld. En verder werkt er toe mee, dat de statistica lang niet alles vermag en zij dus vaak in de steek laat en teleurstelt waar te hoge verwachtingen waren gesteld.

Maar in hoofdzaak is het toch wel te wijten aan *onvoldoende kennis omtrent de exacte statistische methoden:* de meeste dezer onderzoekers weten niet hoe die methoden werken, noch wat er mee kan worden bereikt of in welke omstandigheden zij voordeel kunnen opleveren.

Het gaat met deze statistische methoden als met elk ander instrument: om er profijt van te kunnen trekken dient men er althans enig inzicht in te hebben, ook al zou men het aan een ander willen overlaten

er mee te werken. Hij, die de hulp van een laboratorium voor microscopie inriep om kleine kleurverschillen te bestuderen, gaf daarmee blijk niet te weten, dat een microscoop slechts afstanden vergroot weergeeft en de kleuren onveranderd laat; het gevolg was, dat hij nul op het request kreeg, maar bovendien kan worden verwacht, dat hij in die gevallen, waarin microscopisch onderzoek hem wèl zou kunnen helpen, dit niet zal inzien en dat hij dus de gunstige gelegenheden zal laten voorbijgaan. En dat is meer te betreuren dan de zoëven bedoelde teleurstelling.

Dit boekje is bedoeld om deze onderzoekers dat inzicht in de gedachtengang der statistica te geven, dat nodig is om aan te voelen, in welke gevallen het eigen onderzoek zou kunnen worden gebaat door een exacte analyse van het verkregen getallenmateriaal, m.a.w. om hen in staat te stellen op het juiste moment de hulp van de statistica in te roepen.

En daarbij is er in de eerste plaats rekening mee gehouden, dat het zich in hoofdzaak moet richten tot lezers, die geen speciale wiskundige opleiding achter de rug hebben. Want zij, die meer wiskundig geschoold zijn, passen de statistische methoden al vaak toe, in hoofdzaak doordat zij zich die methoden gemakkelijker eigen konden maken.

In dit verband is het niet ondienstig op te merken, dat het waarschijnlijk aan deze laatstgenoemde bijzonderheid, dus aan de kennis der onderzoekers omtrent de wiskunde, te danken is, dat in de z.g. exacte vakken — waarbij aan de sterrekunde als prototype kan worden gedacht — de verkregen meet- en telgetallen vrijwel steeds exact statistisch worden bewerkt. De andere verklaringen daarvoor gaan alle mank.

Zo wordt wel als verklaring gewezen op het feit, dat de fouten juist bij sterrekundige metingen en tellingen beter de z.g. „wetten der waarschijnlijkheid" zouden volgen dan bij andere vakken zoals b.v. de geneeskunde. Deze verklaring gaat niet geheel op doordat bij een exacte statistische analyse onder meer wordt nagegaan, in hoeverre de toevallige fouten aan die bedoelde wetten beantwoorden: doen zij dat wel — en dat blijkt heel vaak het geval te zijn — dan zijn de statistische methoden er even goed op van toepassing of de getallen een sterrekundig dan wel b.v. een medisch onderwerp uitdrukken.

Ook wordt wel als verklaring vernomen, dat de meetfouten in de sterrekunde zo veel kleiner zouden zijn dan in de andere vakken, zodat het in die andere vakken „gevaarlijker" zou zijn, op een statistische analyse te vertrouwen of toch niet gewenst een zo nauwkeurige conclusie uit de meetgetallen te trekken als dat met behulp van de statistische methoden mogelijk zou zijn. De sterrekundige weet echter,

dat, al zijn zijn meetmethoden uiterst verfijnd, datgene wat hij tegenwoordig wenst te weten te komen in zo uiterst kleine grootten optreedt, dat dáárvoor zijn metingen eigenlijk als bijzonder grof moeten worden beschouwd.

En tenslotte de tegenovergestelde verklaring: de zaken, die de sterrekundige tegenwoordig met zijn metingen wenst vast te stellen, zouden zonder eliminering van de meetfouten — dat is zonder exacte statistische analyse — niet uit de meetgetallen kunnen worden afgeleid juist doordat die fouten zo groot zijn. Hier wringt de schoen bij een gebrek aan zelfkennis in die „andere vakken". De medicus heeft er wel enig besef van, dat het gevaarlijk is om een oordeel uit te spreken op grond van een ervaring bij enkele ziektegevallen, maar hij heeft er in de regel geen kijk op, hòe gevaarlijk dat is. Dat wil zeggen: hij heeft geen gemotiveerde indruk noch omtrent de onnauwkeurigheid van zijn gegevens noch omtrent de graad van onnauwkeurigheid, die voor zijn oordeel nog gevaarlijk zou zijn. En enige ervaring met het toepassen van de exacte methoden der statistica op medische onderzoekingen is al voldoende om te weten, dat die gevaren groter zijn dan men in het algemeen vermoedt. Ook in die „andere vakken" is het heel vaak, net als in de sterrekunde, niet mogelijk om zonder exacte eliminatie der foutenbronnen een behoorlijk gefundeerd oordeel uit de verkregen meet- of telgetallen af te leiden, juist doordat ook daar die fouten zo groot zijn.

De schrijver van deze regelen is er dan ook van overtuigd, dat het ruimere gebruik van de statistische werkmethoden in de z.g. exacte vakken niet in de eerste plaats te danken is aan de aard van de studieobjecten dier vakken maar aan de aard van de onderzoekers in die vakken, die door hun ruimere wiskundige opleiding gemakkelijker in staat zijn zich van die methoden op hoogte te stellen. Dit houdt voor de onderzoekers in de andere vakken een duidelijke aansporing in om ook te trachten enig inzicht in deze methoden te verkrijgen, al valt dat hen ook moeilijker.

Het wordt de niet-wiskundig-geschoolde werker inderdaad niet gemakkelijk gemaakt, zich dat inzicht te verschaffen. En dat is de belangrijkste oorzaak van de veel te geringe toepassing der exacte statistische analyse. Bijna alle boeken over dit onderwerp zijn door wiskundigen geschreven en al richten dezen zich in vele hunner boeken speciaal tot niet-wiskundigen, deze schrijvers blijven toch de wiskundige, dat wil in dit verband niet zeggen de man, die met formules werkt, maar de man die nu eenmaal anders denkt, van andere uitgangspunten uitgaat en zijn conclusies anders formuleert dan de jurist, de medicus, de socioloog.

Moge dit boekje er in geslaagd zijn, deze anders-dan-wiskundig-denkenden een inzicht te geven in de gedachtengang der statistica, voldoende reëel om er hen toe te brengen vaker en op de juiste momenten de hulp van die statistica in te roepen, hetzij om te weten te komen of hun conclusie voldoende betrouwbaar is, of dat er een gelukkiger conclusie kan worden getrokken, of hoe hun onderzoek op bepaalde punten kan worden verscherpt, of wel om een aanwijzing te krijgen hoe het onderzoek, dat zij van plan zijn te gaan ondernemen, zo gunstig mogelijk kan worden opgezet om met de geringste moeite (en met de geringste kosten!) het grootste effect te sorteren.

Het ligt niet in de bedoeling, de lezer door middel van dit boekje in staat te stellen, zelf exacte statistische onderzoekingen te gaan uitvoeren (behalve mogelijk enkele uiterst eenvoudige in situaties, waarvan men op goede gronden overtuigd zou zijn dat zij met die van een der hier behandelde voorbeelden overeenkomen). Van dezelfde schrijver is een uitvoeriger boek in bewerking, waarin hij wel bij dezelfde soort lezers het inzicht in de statistica zo ver hoopt te verdiepen, dat zij er door in staat zullen worden gesteld de meeste van hun onderzoekingen door eigen statistische analyse te verrijken.

STATISTIEK EN STATISTICA

De meest in het oog springende trek van de statistiek is wel, dat deze de werkelijkheid door middel van getallen tracht te karakteriseren. In den regel zijn dit reeksen van getallen en daaraan is ook de naam ontleend: in het hoogtij der Latijnsche beschaving werden zulke getallenreeksen voornamelijk in het staatsbestel gebruikt (b.v. voor het regelen der belastingen) waardoor zulk een getallenrij de nu nog gebruikelijke naam van een „staat" kreeg en daarnaar is deze branche statistiek genoemd. De theoretische wetenschap, die ten dienste van de statistiek is ontwikkeld, kan daarvan worden onderscheiden als de „statistica".

Een deel van de taak der statistiek is dan ook het verkrijgen van de nodige getallen en een ander deel het overzichtelijk weergeven van die getallen. In veler ogen vormen deze twee werkzaamheden zelfs de belangrijkste taak van de statistiek. Veel werk wordt er zeker voor vereist. De getallen moeten worden verkregen door tellingen of door metingen en vaak dienen daar eerst uitgebreide organisaties voor in het leven te worden geroepen: wij denken aan een volkstelling of aan de meting van het resultaat in een moeilijk natuurkundig experiment, dat uitgebreide voorzorgen vereist. Soms moeten veel gegevens uit diverse archieven of publicaties bij elkaar worden gegaard, waarbij dan vaak nog extra moeilijkheden zijn te overwinnen doordat de getallen op die verschillende plaatsen op uiteenlopende manieren waren verkregen. Of wel een uitvoerige proef, b.v. over het nut van een praeventief geneeskundige maatregel, moet op vele plaatsen tegelijk in gang worden gezet, waarbij zorg dient te worden gedragen dat de mogelijk zeer talrijke personen, die bij de uitvoering een rol spelen, hun taak allen op de gewenste manier vervullen.

Het is echter de vraag of dit alles, wat nodig is voor het verkrijgen van de getallen, het werk van de statisticus moet zijn. Wel is een organisatie denkbaar, waarbij ook dit werk aan de statisticus zou worden opgedragen, maar in den regel zullen de gewenste getallen beter en gemakkelijker kunnen worden verkregen door b.v. de ambtenaren van de bevolkingsregisters, de natuurkundige, de medicus. Het meest spreekt dit wel bij het zoëven genoemde natuurkundige experi-

ment, dat zonder twijfel in handen van de deskundige physicus beter wordt uitgevoerd dan door de statisticus, die tevens op zoveel andere gebieden statistische vraagstukken zal hebben te bewerken en als zodanig allicht niet speciaal in de experimentele natuurkunde deskundig zal zijn.

Zelfs het overzichtelijk weergeven van de verkregen getallen zal vaak even goed gelukken aan de speciale deskundige in die branche van de wetenschap of van de praktijk, in welker belang die getallen waren verzameld. Die overzichtelijkheid kan meestal in hoge mate worden bevorderd door grafische voorstellingen en om de suggestieve werking daarvan te verhogen en dus de bedoeling er van gemakkelijker te doen begrijpen, worden daar tegenwoordig graag de bekende moderne figuurtjesgrafieken voor gebruikt. Maar ook die kunnen door de bedoelde deskundige in den regel minstens even gelukkig ontworpen worden als door de statisticus.

Wat is dan het „eigenlijke" werk van de statisticus?

Dit komt pas aan bod als uit de verkregen getallen een *conclusie* moet worden getrokken omtrent de werkelijkheid, waar die getallen aan ontleend zijn. En ook dan is de hulp van de statisticus nog niet in alle gevallen nodig. Als de tuberculose-arts de aantallen sterfgevallen aan tuberculose kent, die in bepaalde jaren in een bepaalde streek zijn voorgekomen, en in elk van die jaren blijkt dit aantal kleiner te zijn dan in het onmiddellijk daaraan voorafgegane jaar, dan kan hij ook zonder de hulp van een statisticus concluderen, dat in die jaren en in die streek de sterfte aan tuberculose afnam.

Het eigenlijke werk van de specialist op het gebied der statistiek komt pas aan de orde als de getallen, waaruit een conclusie over de werkelijkheid moet worden afgeleid, de bedoeling, die met die getallen werd nagestreefd, *slechts onnauwkeurig weergeven*, m.a.w. **als een conclusie moet worden getrokken uit onnauwkeurige getallen.**

En de meeste getallen zijn in deze zin onnauwkeurig.

Een getal, dat door meten is verkregen, moet steeds worden geacht, de bedoeling onnauwkeurig weer te geven, want elke meting kan behept zijn met een meetfout en bovendien wordt bij het aflezen van een meting steeds op een bepaalde eenheid afgerond.

Alleen een getal, dat door telling is verkregen, kan in bepaalde omstandigheden worden geacht, de werkelijkheid nauwkeurig weer te geven. Als b.v. wordt gevraagd hoeveel jongen door een bepaalde hond ter wereld zijn gebracht, dan kunnen de omstandigheden zodanig zijn, dat het getelde aantal zeker met het werkelijke aantal overeenkomt. En ook kan, als voldoende voorzorgen in acht zijn genomen, door tellingen nauwkeurig worden vastgesteld, hoeveel jongen ge-

middeld per worp door een bepaalde hond of door enige bepaalde honden ter wereld werden gebracht. Maar anders wordt het, als gevraagd wordt naar het gemiddeld aantal jongen per worp bij het genus hond: het is ondoenlijk de jongen te tellen in alle worpen van alle honden, die nu, in het verleden en in de toekomst over de gehele aarde bestaan, hebben bestaan en zullen bestaan. Voor de beantwoording van die vraag dient wel genoegen te worden genomen met het tellen in een steekproef van een matig aantal honden en het is duidelijk, dat de verhoudingen in zulk een steekproef niet nauwkeurig behoeven overeen te komen met die over de hele wereld en in alle tijden. Een getal, dat verkregen is door het tellen in een steekproef met de bedoeling, een eigenschap van een ruimere werkelijkheid weer te geven, moet dus eveneens worden beschouwd als een onnauwkeurig getal. En in de praktijk worden de meeste tellingen aan steekproeven uitgevoerd.

Elk meetgetal en de meeste telgetallen moeten dus worden opgevat als onnauwkeurige getallen, n.l. als *schattingen* van de werkelijkheid, zij het dan ook dat zij door het benutten van een meetinstrument of van een steekproef allicht betere schattingen zijn dan die, die zonder zulk een hulpmiddel zijn verkregen.

De vraag kan worden gesteld of het mogelijk is, juiste conclusies uit deze in wezen onnauwkeurige gegevens te trekken. En het antwoord moet luiden, dat deze conclusies in beginsel steeds de sporen van deze onnauwkeurigheid dragen: een statistische conclusie is n.l. nooit volledig betrouwbaar. Wel kan de ene statistische conclusie een grotere mate van betrouwbaarheid hebben dan de andere en daarom dient de eis te worden gesteld, dat bij een statistische conclusie steeds *de mate van betrouwbaarheid* op de een of andere manier behoort te worden vermeld.

Om een conclusie uit onnauwkeurige getallen af te leiden en er de mate van betrouwbaarheid van te bepalen is een afzonderlijke tak van wetenschap ontstaan, de statistica. En het is van deze statistica dat hier de theoretische grondslagen zullen worden geschetst. Op de statistica toch berusten de praktische werkmethoden der statistiek en pas een zekere mate van kennis omtrent deze grondslagen maakt het voor de onderzoeker, die op zijn eigen speciale terrein met meet- of telgetallen werkt, mogelijk, telkens de gelegenheden te onderkennen, waar een statistische bewerking zijner getallen hem verder kan brengen.

Bij deze schets zal worden uitgegaan van enkele eenvoudige „vraagstukken", die elk de gelegenheid zullen geven om telkens een volgende sport in de ladder dezer theoretische grondslagen te belichten. Pas daarna zal de positie van deze ladder in haar geheel worden besproken,

dat wil zeggen de grond, waarop zij staat en de muur, die er mee kan worden beklommen.

Wiskundige bewijzen zullen hier niet worden weergegeven, wel zal telkens uitgangspunt en resultaat van zulk een wiskundige afleiding op zodanige wijze worden aangeduid, dat geen enkele formule „uit de lucht komt vallen": de grond dient te worden ontnomen aan de foutieve indruk, die bij velen wordt aangetroffen, dat in de waarschijnlijkheidsleer een mysterieuze, ja zelfs mystieke inslag zou bestaan.

TELLING VAN HET GEHALTE IN EEN STEEKPROEF

Gaan wij uit van een concreet onderzoek op het gebied van de erfelijkheidsleer.

Een tweelingpaar kan uit één of uit twee eieren zijn ontstaan. In het eerste geval is het ei na de bevruchting gespleten in twee volledig gescheiden delen, die dan elk tot een levensvatbaar individu zijn uitgegroeid. Vandaar dat éénei̇ge tweelingen eenzelfde kapitaal aan erfelijke eigenschappen op hun levenspad hebben meegekregen, waardoor zij opvallend sterk op elkaar gelijken en o.a. hetzelfde geslacht hebben. Om na te gaan, of bij de cavia cobaya (het Guinese biggetje) zulke éénei̇ge tweelingen voorkomen, werd door Dr. A. L. en A. J. W. HAGEDOORN gedurende een zekere tijd in hun cavia-stal van elke worp het geslacht van elk jong genoteerd. Als van 400 getelde worpen er 103 bestonden uit twee mannetjes, 104 uit twee vrouwtjes en 193 uit een mannetje en een vrouwtje, pleiten dan deze getallen voor het voorkomen van éénei̇ge tweelingen bij de cavia? [1])

Daar bij bedoelde eisplijting één eicel uitgroeit tot twee jongen van hetzelfde geslacht, is bij dieren, die o.a. éénei̇ge tweelingen voortbrengen, de kans op een gelijkslachtige tweeling groter dan zij zou geweest zijn als er geen éénei̇ge tweelingen zouden worden geboren. Hierop richt zich ons statistisch onderzoek.

Indien de cavia geen éénei̇ge tweelingen voortbrengt, kunnen wij de kans op gelijkslachtige tweelingen even groot achten als die op ongelijkslachtige.

Dit blijkt uit de volgende overweging. Er zijn geen aanwijzingen, dat de kansen op het ontstaan van manlijke en van vrouwelijke vruchten bij de cavia ongelijk zouden zijn [2]). Bij een tweevoudige zwangerschap kunnen de

---

[1]) Om didactische redenen worden de worpen met andere aantallen jongen hier niet in de beschouwingen opgenomen en zijn de getallen gefingeerd. De juiste volledige gegevens en hun statistische bewerking zijn gepubliceerd in Genetica 1943, blz. 315.

[2]) Het statistisch onderzoek, waar deze uitspraak op berust, is te vinden in het zoëven aangehaalde artikel van Dr. A. L. en A. J. W. HAGEDOORN.

twee kort na elkaar ontstane vruchten achtereenvolgens de geslachten hebben: mm, mv, vm of vv en de kansen van deze vier mogelijke combinaties zijn dus ook als gelijk te beschouwen. De kans op een gelijkslachtige combinatie (mm of vv) is dus gelijk te stellen aan die op een ongelijkslachtige (mv of vm).

Als in een steekproef van 400 worpen 200 gelijkslachtige en 200 ongelijkslachtige zouden voorkomen, zou deze steekproef dus overeenkomen met de kansen bij afwezigheid van eisplijting. Zulk een steekproef wordt enigszins oneigenlijk de „theoretisch verwachte" genoemd. In onze steekproef zijn 207 gelijkslachtige worpen geteld (n.l. 103 mm en 104 vv) zodat de „afwijking" tussen het „gevonden aantal" en het „theoretisch verwachte aantal" 7 bedraagt.

Deze overwegingen zijn in tabel I nog eens overzichtelijk opgeschreven. Daarbij is er van uitgegaan, dat de som van de kansen van alle mogelijkheden ener gegeven situatie altijd 100 % of 1 bedraagt, zodat de combinaties mm, mv, vm en vv, die bij de cavia gelijke kansen zouden hebben, elk een kans van $\frac{1}{4}$ hebben en de kans van een gelijkslachtige tweeling (mm of vv) dus $\frac{1}{2}$ zou zijn.

TABEL I

Zonder eisplijting

Soort	Kans	Soort	Kans	Theoretische verwachting in steekproef van 400 worpen	Gevonden aantal worpen	Afwijking
mm	$\frac{1}{4}$	gelijkslachtig	$\frac{1}{2}$	200	207	+ 7
vv	$\frac{1}{4}$					
mv	$\frac{1}{4}$	ongelijkslachtig	$\frac{1}{2}$	200	193	(—7)
vm	$\frac{1}{4}$					

De afwijking van 7 worpen, die onze steekproef te zien geeft ten opzichte van de onderstelling dat er geen eisplijting zou zijn voorgekomen, gaat in de richting die theoretisch bij eisplijting te verwachten is: er zijn „te veel" gelijkslachtige worpen aangetroffen. Mogen wij daarom uit deze steekproef zonder meer concluderen, dat bij de cavia ééneiige tweelingen voorkomen?

Neen: de gevonden afwijking behoeft immers niet op eisplijting te berusten, zij kan ook het gevolg zijn geweest van het bekende feit, dat een steekproef niet nauwkeurig met de verhouding bij „alle" cavia's

overeen behoeft te komen (beter gezegd: niet met de werkelijke kansen behoeft overeen te komen). De afwijking kan m.a.w. ook het gevolg zijn van de toevallige materiaalselectie, die steeds bij het nemen van een steekproef optreedt.

Wat de oorzaak der gevonden afwijking betreft, daarbij kunnen wij dus twee mogelijkheden onderscheiden: de afwijking berust òf uitsluitend op de toevallige steekproefselectie òf mede op een fout in de getoetste onderstelling, (n.l. de onderstelling, dat alle tweelingen tweeeiïg zouden zijn). Alleen als de eerste mogelijkheid zeer onwaarschijnlijk is, wordt de tweede zeer waarschijnlijk, d.w.z.: als de kans klein is, dat de toevallige steekproefselectie in een steekproef als deze een afwijking van + 7 veroorzaakt, dan zijn er waarschijnlijk wel ééneiïge tweelingen bij.

Onze conclusie hangt dus af van de kans, dat de toevallige steekproefselectie in een steekproef van 400 individuen, waarbij de kans op een „individu" van de bedoelde soort (i.c. een gelijkslachtige worp) $\frac{1}{2}$ bedraagt, een afwijking van + 7 veroorzaakt. Het is dus onze opgaaf, die kans te berekenen. Om technische redenen wordt in de statistica niet deze kans zelf bepaald, maar de kans op een afwijking, even ongunstig voor de te toetsen mening als de gevondene of nog ongunstiger, in ons geval dus de kans op 207, 208, 209. . . . 399 of 400 gelijkslachtige worpen in een steekproef van 400 worpen. Deze kans heet de *overschrijdingskans* van de gevonden afwijking. Hoe kleiner deze is, des te kleiner is ook de kans, dat de toevallige steekproefselectie een afwijking als de gevondene veroorzaakt en des te kleiner is de kans, dat de getoetste hypothese juist zou zijn.

In ons vraagstuk zijn wij bij het berekenen van de kansen in tabel I uitgegaan van de onderstelling, dat er geen eisplijting zou zijn geweest. Indien onze overschrijdingskans heel klein zou blijken te zijn, zou er waarschijnlijk wel eisplijting zijn voorgevallen.

Hoe groot is in ons vraagstuk de overschrijdingskans? Onze steekproef had een afwijking van + 7. De gevraagde overschrijdingskans is dus de kans dat, zonder eisplijting, een steekproef van 400 worpen een afwijking heeft van + 7 of groter.

Om deze vraag te beantwoorden bespreken wij eerst een nog eenvoudiger geval, n.l. een veel kleinere steekproef.

INTERMEZZO: TELLING IN EEN KLEINE STEEKPROEF

Wij stellen ons een loterij voor, bestaande uit „prijzen" en „nieten" en omtrent die loterij heeft iemand de mening verkondigd, dat de kans op het trekken van een prijs $\frac{1}{4}$ zou bedragen.

Om die hypothese te toetsen wordt uit die loterij een steekproef van 10 loten genomen en daarin blijken 7 prijzen voor te komen. Is de verkondigde hypothese in strijd met deze steekproef te achten?

In deze steekproef is het theoretisch verwachte aantal prijzen $\frac{1}{4} \times 10 = 2\frac{1}{2}$ (in het algemeen PN als P de kans op een „prijs" en N het aantal individuen van de steekproef voorstelt). De afwijking bedraagt dus $7 - 2\frac{1}{2}$ of $4\frac{1}{2}$. Voor de te toetsen hypothese zou het nog ongunstiger zijn geweest, indien de steekproef 8, 9 of 10 prijzen zou hebben bevat, zodat de overschrijdingskans in ons geval de kans is op een steekproef met 7, 8, 9 of 10 prijzen.

Een steekproef van 10 loten kan 0, 1, 2, 3 .... of 10 prijzen bevatten. De kans op elk dezer mogelijkheden is gemakkelijk te berekenen.

Stellen wij ons daartoe voor, dat de steekproef is genomen doordat eerst één lot is getrokken, dat dan afgelezen en genoteerd en daarna weer in de loterij teruggeschud wordt, welk spelletje 10 maal wordt uitgevoerd.

Als de te toetsen hypothese juist is bedraagt de kans, dat het eerstgetrokken lot een „niet" is, $\frac{3}{4}$ (want de kans op een prijs is dan $\frac{1}{4}$, de kans op een „prijs" en die op een „niet" samen, dat is de som van de kansen op alle mogelijkheden, is 1, dus is de kans op een „niet" $1 - \frac{1}{4} = \frac{3}{4}$).

Bij een steekproef van 2 loten bedraagt de kans, dat beide een niet zijn, dus $\frac{3}{4} \times \frac{3}{4}$ (volgens het z.g. theorema van de gecombineerde kans, dat als een axioma een van de theoretische grondslagen der statistica vormt). De kans, dat beide loten een prijs zijn, is $\frac{1}{4} \times \frac{1}{4}$. Verder is de kans, dat het eerste een prijs en het tweede een niet is, $\frac{1}{4} \times \frac{3}{4}$, en die, dat omgekeerd het eerste een niet en het tweede een prijs zou zijn, ook $\frac{1}{4} \times \frac{3}{4}$, zodat de kans op één prijs en één niet $2 \times \frac{3}{4} \times \frac{1}{4}$ bedraagt. Overzien wij deze kansen:

Steekproef van 2 loten (kans op een prijs $\frac{1}{4}$):

aantal prijzen:	0	1	2
kans:	$(\frac{3}{4})^2$	$2 \times \frac{3}{4} \times \frac{1}{4}$	$(\frac{1}{4})^2$

dan herkennen wij hierin de termen van $(\frac{3}{4} + \frac{1}{4})^2$. En zo vinden wij op dezelfde manier voor een

steekproef van 3 loten (kans op een prijs $\frac{1}{4}$):

aantal prijzen:	0	1	2	3
kans:	$(\frac{3}{4})^3$	$3 \times (\frac{3}{4})^2 \times \frac{1}{4}$	$3 \times \frac{3}{4} \times (\frac{1}{4})^2$	$(\frac{1}{4})^3$

of de termen van $(\frac{3}{4} + \frac{1}{4})^3$. Hierna zal het niet verwonderen dat bewezen kan worden, dat de diverse kansen voor een steekproef van N loten, als de kans op een prijs P is en die op een niet Q, gelijk zijn aan de termen van

$$(Q + P)^N$$

Naar aanleiding van deze overeenkomst van de kansen, die in deze omstandigheden optreden, met de termen van een macht van een binomium wordt deze soort kansverdeling de *binomiale kansdistributie* genoemd.

Met behulp van de geschetste eigenschap zijn in tabel II de kansen berekend voor de steekproef van 10 loten uit een loterij, waarin de kans op een prijs $\frac{1}{4}$ bedraagt. Wij zagen reeds dat de overschrijdingskans de kans is op een steekproef met 7, 8, 9 of 10 prijzen en uit tabel II lezen wij af, dat de kans op 7 prijzen, afgerond 0,0031 bedraagt, op 8 prijzen 0.0004, op 9 prijzen 0.00003 en op 10 prijzen 0.000001, zodat de overschrijdingskans 0.0031 + 0.0004 + 0.00003 + 0.000001 of 0.0035 of, nog meer afgerond, $^1/_{300}$ bedraagt.

Dit betekent dat, indien de kans op een prijs $\frac{1}{4}$ zou bedragen, de toevallige steekproefselectie slechts in 1 op de ongeveer 300 steekproeven een afwijking veroorzaakt even ongunstig als of nog ongunstiger dan die, die wij in onze steekproef hebben gevonden. Wij zouden het dus al heel ongelukkig hebben getroffen als juist in onze steekproef de toevallige materiaalselectie een voor onze gedachte zó ongunstig aantal prijzen had doen optreden en dus achten wij, op grond van de kleine overschrijdingskans, de getoetste hypothese waarschijnlijk onjuist: waarschijnlijk is het prijsgehalte in deze loterij niet $\frac{1}{4}$. En van deze conclusie geven wij de betrouwbaarheid aan door mee te delen, dat de overschrijdingskans $^1/_{300}$ bedraagt.

HET GEHALTE IN EEN GROTE STEEKPROEF

Op dezelfde manier kan ook voor ons tweelingen vraagstuk de overschrijdingskans worden berekend, maar doordat daar de steekproef zo veel groter is (400 in plaats van 10) zou dat zeer veel cijferwerk kosten en daarom is een andere methode bedacht, die het mogelijk maakt de overschrijdingskans in zulke omstandigheden, zij het ook slechts bij benadering, met weinig moeite te bepalen.

GAUSS stelde n.l. een schematisch distributietype op, dat twee voor ons doel gelukkige eigenschappen bezit: het lijkt veel op het binomiale distributietype en men kan er de grootte van een overschrijdingskans met weinig moeite uit afleiden.

TABEL II

De binomiale kansdistributie van het aantal prijzen in een steekproef van 10 loten uit een loterij, waarin de kans op het trekken van een prijs $\frac{1}{4}$ bedraagt.

Mogelijke aantallen prijzen in de steekproef	Kans op het links vermelde aantal prijzen
0	$\frac{10!}{0!\,10!} \times (\frac{3}{4})^{10} = 0.056\ldots$
1	$\frac{10!}{1!\,9!} \times (\frac{3}{4})^{9} \times \frac{1}{4} = 0.188\ldots$
2	$\frac{10!}{2!\,8!} \times (\frac{3}{4})^{8} \times (\frac{1}{4})^{2} = 0.282\ldots$
3	$\frac{10!}{3!\,7!} \times (\frac{3}{4})^{7} \times (\frac{1}{4})^{3} = 0.250\ldots$
4	$\frac{10!}{4!\,6!} \times (\frac{3}{4})^{6} \times (\frac{1}{4})^{4} = 0.146\ldots$
5	$\frac{10!}{5!\,5!} \times (\frac{3}{4})^{5} \times (\frac{1}{4})^{5} = 0.058\ldots$
6	$\frac{10!}{6!\,4!} \times (\frac{3}{4})^{4} \times (\frac{1}{4})^{6} = 0.016$
7	$\frac{10!}{7!\,3!} \times (\frac{3}{4})^{3} \times (\frac{1}{4})^{7} = 0.0031$
8	$\frac{10!}{8!\,2!} \times (\frac{3}{4})^{2} \times (\frac{1}{4})^{8} = 0.00039\ldots$
9	$\frac{10!}{9!\,1!} \times (\frac{3}{4}) \times (\frac{1}{4})^{9} = 0.000029\ldots.$
10	$\frac{10!}{10!\,0!} \times (\frac{1}{4})^{10} = 0.000001\ldots$

De coëfficienten $\frac{10!}{0! \times 10!}$ enz. kunnen ook worden afgelezen uit de bekende driehoek van PASCAL. Hier werden zij afgeleid met behulp van de formule, die NEWTON gaf voor de termen van $(a + b)^n$:

$$\frac{n!}{m!\,(n-m)!}\,a^{n-m}\,b^{m},$$

waarin m alle gehele getallen van 0 tot n voorstelt.

10! (te lezen: 10 faculteit) betekent $1 \times 2 \times 3 \times \ldots \times 10$, zodat b.v.

$$\frac{10!}{7!\,3!} = \frac{1\times2\times3\times4\times5\times6\times7\times8\times9\times10}{1\times2\times3\times4\times5\times6\times7\times1\times2\times3} = \frac{8\times9\times10}{1\times2\times3} =$$

$$= 4\times3\times10 = 120. \qquad (0! = 1! = 1)$$

Hij noemde dit type de *normale kansdistributie.* De grafische voorstelling er van is algemeen bekend: de z.g. normale kanslijn (ook wel genoemd de kromme van GAUSS of van QUETELET). Dat deze normale kansdistributie veel op de binomiale lijkt wordt gedemonstreerd door fig. 1, waar een normale kanskromme is getekend door de binomiale

Fig. 1.

Grafische voorstelling van de binomiale kansdistributie van het aantal prijzen in een steekproef van 20 loten uit een loterij, waarin de kans op het trekken van een prijs $\frac{1}{4}$ bedraagt (de tredenlijn) en van die normale kansdistributie, die er het best bij past (de normale kanskromme) [1]).

kansdistributie van het aantal prijzen in een steekproef van 20 loten uit een loterij, waarin de kans op een prijs $\frac{1}{4}$ bedraagt. De overeenstemming tussen beide typen van kansdistributie wordt te groter naarmate de steekproef groter is, op eenzelfde manier als een regelmatige veelhoek meer op een cirkel gaat lijken naarmate het aantal zijden groter is. De normale kansdistributie wordt dan ook wel beschouwd als de limiet, waartoe de binomiale

---

[1]) De kans op meer dan 12 prijzen is te klein om de kanslijn bij deze schaal van tekenen van de horizontale as te kunnen onderscheiden, reden waarom het rechter deel van de grafiek is weggelaten.

nadert als het aantal individuen van de steekproef steeds groter wordt gekozen.

Zoals er verschillende binomiale kansdistributies bestaan, afhankelijk van de grootten van P, Q en N in het binomium $(Q + P)^N$, zo zijn er ook verschillende normale kansdistributies, juist zoals er ook verschillende cirkels (n.l. met verschillend middelpunt en verschillende

Fig. 2.

Twee normale kanskrommen met verschillende centra (M) en verschillende standaard-deviaties (S).

straal) bestaan als de limiet van veelhoeken met verschillend middelpunt en verschillende afstand van de hoekpunten tot dat middelpunt. In fig. 2 zijn twee uiteenlopende normale kanslijnen getekend. Verschillende normale kansdistributies verschillen in twee opzichten, n.l. in de grootte van het z.g. *centrum* en in de grootte van de z.g. *standaard-deviatie*. Wat deze twee termen betekenen wordt het gemakkelijkst begrepen uit fig. 2. Van elk der twee normale kanslijnen is het centrum door de letter M aangegeven en de standaard-deviatie door de letter S. De variabele grootheid (in ons geval het aantal gelijkslachtige

worpen in een steekproef van 400 cavia-worpen) is in deze figuur langs de horizontale as (de X-as) uitgezet, zodat verschillende punten van de horizontale as verschillende waarden dier variabele grootheid aanduiden (i.c. verschillende aantallen gelijkslachtige worpen) waarbij dan de hoogte van de kanslijn boven dat punt een inzicht geeft in de grootte van de kans op dat aantal. Het centrum (M) van een normale kanslijn is die waarde, waar de symmetrie-as van de kanskromme de X-as snijdt, dat is tevens de waarde met de grootste kans (omdat de symmetrie-as door het hoogste punt van de normale kanskromme loopt). De symmetrie-as van de linker kanskromme snijdt de X-as dichter bij het nulpunt (maar nog rechts er van) dan die van de rechter kromme; de links getekende normale kansdistributie heeft dus een kleiner centrum dan de rechter. De standaard-deviatie (S) is de afstand van het buigpunt tot de symmetrie-as (het buigpunt is het punt, waar de convexe kromming van de top omslaat in de concave kromming van de voet). De links getekende kansdistributie heeft dus een grotere standaard-deviatie dan de rechter (vandaar dat de linker een bredere indruk maakt en lager blijft dan de rechter).

Hoe wordt nu de normale kansdistributie gebruikt als een hulpmiddel voor het bepalen van de overschrijdingskans in een binomiale kansdistributie? Daartoe wordt die binomiale kansdistributie vervangen gedacht door de normale, die er het best bij past [1]) en dan wordt de overschrijdingskans bepaald alsof die normale gold. Wel wordt dan een fout gemaakt doordat die twee kansdistributies enigszins verschillend zijn, maar die fout is slechts klein en het voordeel van het geringere werk weegt daar in den regel ruimschoots tegen op.

Wiskundig kan worden afgeleid, dat de normale kansdistributie, die het best bij de binomiale $(Q + P)^N$ past, dat centrum en die standaarddeviatie heeft welke volgen uit

$$\boxed{M = PN}$$

$$\boxed{S^2 = PQN}$$

In ons tweelingenvraagstuk gingen wij uit van de onderstelling, dat er geen eisplijting zou zijn voorgekomen. De kans op een gelijkslachtige tweeling bleek dan $\frac{1}{2}$ te zijn en ook die op een ongelijkslachtige was dan $\frac{1}{2}$, zodat in ons vraagstuk $P = \frac{1}{2}$ en $Q = \frac{1}{2}$. De steekproef

---

[1]) Voor het begrip „best passend" zie blz. 30 e.v.

bestond uit 400 worpen, dus $N = 400$, zodat de best passende normale kansdistributie bepaald is door

$$M = \tfrac{1}{2} \times 400 = 200$$
$$S^2 = \tfrac{1}{2} \times \tfrac{1}{2} \times 400 = 100 \text{ of } S = 10$$

Fig. 3 stelt deze best passende normale kansdistributie voor. Hierin is dus het centrum $M = 200$ en wij merken op, dat dit eenvoudig het theoretisch verwachte aantal gelijkslachtige worpen is: wij zagen n.l. op blz. 8 dat dit theoretisch verwachte aantal PN bedraagt en volgens het bovenstaande is ook het centrum $M = PN$. Als het punt

Fig. 3.

De eenzijdige overschrijdingskans (het gearceerde oppervlak P_1) van A in een normale kansdistributie (D = de deviatie van A).

A in deze figuur het aantal gelijkslachtige worpen voorstelt, dat in onze steekproef gevonden was (dus $A = 212$) dan is MA de gevonden afwijking. Het deel van de X-as, dat rechts van A ligt, bevat alle aantallen, die nog ongunstiger zouden zijn voor de onderstelling, dat er geen eisplijting zou voorkomen. Het oppervlak tussen dit gedeelte van de X-as en de kanskromme, dat in deze figuur gearceerd werd, is dus gelijk aan de gezochte overschrijdingskans. [1]).

---

[1]) Ter wille van de eenvoud der voorstelling werd hier niet ingegaan op het feit, dat bij zulk een steekproeftelling het aantal, dat langs de X-as is uitgezet, slechts in discrete waarden kan optreden, n.l. als hele getallen. Omdat in deze omstandig-

Het hier behandelde vraagstuk heeft de bijzonderheid, dat alleen die aantallen, die groter zijn dan dat in de genomen steekproef, ongunstiger voor de getoetste hypothese zijn: een aantal gelijkslachtige worpen met een negatieve afwijking zou, zelfs al was deze afwijking in absolute waarde groter dan de gevondene, immers niet wijzen op het optreden van eisplijting omdat het aantal gelijkslachtige worpen door eisplijting steeds zou toenemen. In vele gevallen, waar een geopperde hypothese wordt getoetst door het gehalte in een steekproef te tellen,

Fig. 4.

De dubbelzijdige overschrijdingskans (P_2) van A in een normale kansdistributie (D = de deviatie van A).

zijn ook die aantallen ongunstiger, waarvan de afwijking in absolute zin groter is en het tegengestelde teken heeft als de gevondene. Voor zulke gevallen is de z.g. dubbelzijdige overschrijdingskans gearceerd in fig. 4.

Hoe wordt nu de overschrijdingskans in een normale kansdistributie bepaald als wij het centrum (M), de standaard-deviatie (S) en de afwijking of deviatie (D) kennen?

Het is weer een wiskundige afleiding die aantoont, dat deze kans uitsluitend afhangt van de verhouding tussen de gevonden deviatie en

---

heden de kans op een afwijking van + 12 of meer hetzelfde is als de kans op een waarde, groter dan $11\frac{1}{2}$, zou het punt A in figuur 3 niet de waarde 12 maar de waarde $11\frac{1}{2}$ zou moeten voorstellen, wilde het gestippelde gebied inderdaad „gelijk" zijn aan de overschrijdingskans in de best passende normale kansdistributie.

de standaard-deviatie, dus van $D : S$. Deze verhouding heet de excentriciteit (T) van het gevonden getal, omdat deze verhouding aangeeft, hoe excentrisch het gevonden getal (A) ligt ten opzichte van de kanskromme. In fig. 5 zijn weer twee verschillende normale kanskrommen getekend en daarin stellen de punten A_1 en A_2 getallen voor, die beide op een afstand $1\frac{1}{2}$ S van M liggen, zodat de excentriciteit van beide $1\frac{1}{2}$ bedraagt; de overschrijdingskansen (gearceerd) zijn gelijk.

Fig. 5.

Twee waarden (A) met gelijke excentriciteit ($T = D : S = 1\frac{1}{2}$) in twee verschillende normale kansdistributies hebben gelijke overschrijdingskansen (gearceerd).

In ons vraagstuk vinden wij voor de excentriciteit:

$$T = D : S = 7 : 10 = 0.7\ ^{1)}$$

De formule, die het verband aangeeft tussen de grootte van de excentriciteit en de grootte van de overschrijdingskans, is tamelijk ingewikkeld en het kost veel cijferwerk om met behulp van deze formule de overschrijdingskans bij een gegeven excentriciteit uit te rekenen. Daarom werd eens voor altijd voor een groot aantal waarden van de excentriciteit telkens de daarbij behorende waarde van de overschrijdingskans

---

[1]) Ook hier zou men volgens de opmerking in de voetnoot op blz. 17 $D = 6\frac{1}{2}$ moeten nemen, waardoor $T = 6\frac{1}{2} : 10 = 0.65$ zou worden.

berekend en werden de resultaten in een tabel gepubliceerd. Nadat voor een gegeven geval de excentriciteit is gevonden, kan de overschrijdingskans zonder enige becijfering uit deze tabel worden afgelezen. Fig. 6 geeft die tabel weer in de vorm van een nomogram [1]), bestaande uit twee schaalverdelingen langs één lijn, de bovenste voor de excentriciteit (T), de onderste voor de dubbelzijdige overschrijdingskans (P_2). Voor ons vraagstuk lezen wij bij $T = 0.7$ daaruit de overschrijdingskans $P_2 = \frac{1}{2}$ af. Maar dit is, zoals gezegd, de dubbelzijdige overschrijdingskans (P_2), n.l. de kans op positieve en negatieve afwijkingen, even groot of groter dan de gevondene. Voor ons probleem zijn alleen de positieve afwijkingen ongunstig voor de getoetste onderstelling, zodat voor ons vraagstuk de éénzijdige overschrijdingskans (P_1) geldt en deze bedraagt, vanwege

Fig. 6.

Het T-P-nomogram, bedoeld om de dubbelzijdige overschrijdingskans (P_2) in een normale kansdistributie af te lezen bij een gegeven waarde der excentriciteit (T).

het symmetrische karakter van de normale kansdistributie, de helft van de dubbelzijdige ($P_1 = \frac{1}{2} P_2$). Door de werkelijke binomiale kansdistributie te vervangen door de best passende normale, vinden wij dus bij benadering als overschrijdingskans van de gegeven steekproef de helft van $\frac{1}{2}$ of $\frac{1}{4}$.

Dit betekent het volgende: zouden telkens weer in overeenkomstige omstandigheden steekproeven worden genomen, elk weer bestaande uit 400 cavia-worpen, dan is te verwachten dat de getallen in ten naastenbij 1 op de 4 zulke steekproeven even ongunstig of nog ongunstiger uitvallen ten opzichte van ons uitgangspunt, dat wil zeggen ten opzichte van de hypothese, dat de cavia geen eeneiïge tweelingen ter wereld brengt.

Dat is het nieuws, dat de statistische analyse van ons vraagstuk ons leert, niets meer en niets minder.

De statistica geeft dus geen afdoend antwoord op de door ons ge-

---

[1]) Een diagram, bedoeld om er de waarde van een bepaalde grootheid, behorende bij de waarde(n) van een of meer andere grootheden, uit af te lezen, wordt gewoonlijk een nomogram genoemd.

stelde vraag, of er bij de cavia ééneiïge tweelingen voorkomen: uit de gegeven telgetallen leidt de statistica noch het antwoord „ja" noch het antwoord „neen" af. De gegeven getallen waren niet nauwkeurig in overeenstemming met de gedachte, dat de cavia geen ééneiïge tweelingen voortbrengt. Verkregen tel- of meetgetallen zijn vrijwel nooit nauwkeurig met een te toetsen gedachte in overeenstemming; dat behoeft ook niet om die gedachte toch als juist te kunnen aanvaarden omdat wij weten dat bij praktisch elke meting of steekproeftelling een meet- of toevalsfout optreedt. *De statistica beantwoordt alleen de vraag hoe groot de kans is, dat de toevallige fouten even ongunstige of nog ongunstiger getallen doen optreden.*

In ons vraagstuk blijkt die kans, afgerond, $\frac{1}{4}$ te bedragen. Wie zou, waar het toch een enigszins belangrijke beslissing betreft, op grond van een afwijking, die in 1 op de 4 keer door een toevalsfout zou worden veroorzaakt, willen concluderen dat deze afwijking toe te schrijven is niet aan een toevalsfout maar aan het voorkomen van ééneiïge tweelingen? Indien iemand op grond van deze afwijking, waarvan de overschrijdingskans $\frac{1}{4}$ bedraagt, toch tot het voorkomen van ééneiïge tweelingen zou besluiten, dan zou dat toch een heel onbetrouwbare conclusie zijn.

Stel daartegenover eens het geval, dat er 231 gelijkslachtige paren in de steekproef van 400 waren geweest. De afwijking tussen het gevonden aantal (231) en het theoretisch verwachte aantal (200) zou dan 31 hebben bedragen en de excentriciteit $T = D : S = 31 : 10 = 3.1$. Volgens het T-P-diagram is de dubbelzijdige overschrijdingskans P_2 daarbij 0.002, zodat de gezochte eenzijdige overschrijdingskans $P_1 = 0.001$ zou zijn geweest. Dat wil zeggen: de toevalsfout zou in slechts 1 op de 1000 steekproeven van deze grootte zulk een afwijking van 31 gelijkslachtige paren of een nog ongunstigere veroorzaakt hebben. Wij zouden het dus al heel ongelukkig hebben getroffen als juist onze steekproef die éne ongunstige van de 1000 zou zijn geweest. Het lijkt dus heel waarschijnlijk, dat deze afwijking niet op een toevalsfout berust maar op het voorkomen van ééneiïge tweelingen. Bij die steekproef met 231 gelijkslachtige worpen zouden wij dus op grond van de kleine waarde der overschrijdingskans — 0.001 — concluderen, dat er waarschijnlijk wèl ééneiïge tweelingen bij de cavia voorkomen.

Zo blijkt dus de grootte van de overschrijdingskans een maat te vormen voor de graad van betrouwbaarheid, waarmee de getoetste hypothese — i.c. de hypothese dat er geen ééneiïge tweelingen zouden zijn — wordt verworpen. Hoe kleiner de overschrijdingskans is, des te groter is de betrouwbaarheid, waarmee de getoetste hypothese wordt verworpen. Een steekproef met een grote overschrijdingskans

zou dit verwerpen zeer onbetrouwbaar maken: een steekproef met een grote overschrijdingskans vormt geen betrouwbaar argument tegen de getoetste hypothese.

Hoe klein moet de overschrijdingskans dan zijn, willen wij concluderen, dat een getoetste hypothese onjuist is? Of, anders gezegd, hoe klein moet de overschrijdingskans zijn, willen we concluderen dat de gevonden afwijking „betekenis" heeft, of, in de internationale term, dat die afwijking *significant* is?

De statistica zelf geeft op deze vraag geen antwoord. Wetenschappelijk kan geen enkel criterium worden gevonden dat hier een grenswaarde aan zou wijzen: elke statistische conclusie is in beginsel onzeker, al kan zij in bepaalde gevallen onbeperkt tot zekerheid naderen doordat de overschrijdingskans onbeperkt klein kan uitvallen. Als in onze steekproef 250 van de 400 worpen gelijkslachtig waren geweest, dan zou de afwijking 50, de excentriciteit $T = D : S = 50 : 10 = 5$ en de overschrijdingskans kleiner dan een millioenste zijn geweest. In slechts één op de millioen steekproeven van deze grootte (n.l. van 400 worpen) uit een materiaal zonder ééneiige tweelingen veroorzaken de toevalsfouten een afwijking van 50 of meer, zodat het haast wel zeker zou zijn geweest dat er ééneiige tweelingen bij waren. Maar strikt genomen blijft die conclusie toch altijd enigszins onzeker en er kan dus geen grenswaarde voor de overschrijdingskans bestaan waarbij de onzekerheid in zekerheid zou verkeren.

Strikt genomen zou de statisticus in het rapport, dat hij over zulk een door hem verricht statistisch onderzoek uitbrengt, dan ook niet verder moeten gaan dan het meedelen van de door hem gevonden betrouwbaarheidsgraad, die dan zou moeten worden aangeduid door de grootte van de overschrijdingskans te vermelden. De exact-wetenschappelijke basis der statistica brengt hem immers niet verder. Maar het is de gewoonte, dat hij, om misverstanden te voorkomen, nog wèl een stap verder gaat, n.l. door mee te delen, welke eindconclusie hij geneigd is op grond van de gevonden betrouwbaarheid praktisch te aanvaarden. Daartoe neemt hij een practisch gekozen grenswaarde der overschrijdingskans als richtsnoer aan en in de praktijk voldoet daarvoor het best de grenswaarde van $^1/_{20}$ (in een decimale breuk: 0.05 of 5 %). Blijkt de overschrijdingskans in een gegeven onderzoek kleiner te zijn dan deze grenswaarde van 0.05, dan wordt de getoetste onderstelling practisch verworpen als waarschijnlijk in strijd met de gegeven getallen. Blijkt de overschrijdingskans veel groter te zijn dan 0.05, dan luidt de practische conclusie dat de gegeven getallen geen aanwijzing tegen de getoetste onderstelling inhouden. En als zij slechts weinig groter dan de grenswaarde blijkt te zijn, dan wordt gezegd dat

de getallen enige suggestie vormen om de getoetste mening te verwerpen maar dat meer gegevens (b.v. een grotere steekproef) nodig zijn om tot een duidelijker conclusie te kunnen komen.

Vereist een bepaald doel een grotere betrouwbaarheid, dan wordt een kleinere overschrijdingskans als grens aangenomen. Maar steeds is het nuttig, de bereikte betrouwbaarheid op de een of andere wijze bij de uitkomst te vermelden.

De behandeling van ons tweelingenvraagstuk heeft door het bespreken van allerlei theoretische begrippen vele bladzijden in beslag genomen; de oplossing zelf kostte echter weinig moeite en om dat te demonstreren wordt deze oplossing hier nog eens volledig, maar nu zonder uitweidingen, weergegeven.

Als bij de cavia geen eisplijting plaats vindt en er dus geen ééneiige tweelingen ter wereld worden gebracht, is de kans op gelijkslachtige tweelingworpen $\frac{1}{2}$. Het aantal gelijkslachtige worpen, aan te treffen in een steekproef van 400 tweelingworpen, heeft een binomiale kansdistributie, die weinig verschilt van die normale, waarvoor geldt:

$$M = PN = \tfrac{1}{2} \times 400 = 200$$
$$S^2 = PQN = \tfrac{1}{2} \times \tfrac{1}{2} \times 400 = 100$$
$$S = 10$$

zodat

$$D = 207 - 200 = 7$$
$$T = D : S = 7 : 10 = 0.7$$
$$P_2 = \tfrac{1}{2} \text{ (volgens het T-P-nomogram)}$$
$$P_1 = \tfrac{1}{4}.$$

Op grond van deze grote waarde van de overschrijdingskans (duidelijk groter dan $^1/_{20}$) wordt de onderstelling van afwezige eisplijting geacht, niet in strijd te zijn met de genomen steekproef: in 1 op de 4 steekproeven van deze grootte veroorzaakt de toevalsfout op zichzelf reeds zulk een afwijking of een nog ongunstigere. De steekproef behelst dus geen voldoende aanwijzing voor het bestaan van ééneiige cavia-tweelingen.

MEETFOUTEN

Als één afmeting vele malen gemeten wordt, dan zullen de afgelezen meetgetallen allicht niet alle gelijk aan elkaar zijn (tenzij bij het aflezen op een te grove eenheid was afgerond): bij elke meting kan een meetfout optreden en deze meetfouten, die door allerlei factoren

worden veroorzaakt, variëren in grootte. Bij vele soorten van metingen is gebleken, dat de kansdistributie van de daarbij optredende meetfout veel lijkt op de normale kansdistributie. Een voorbeeld daarvan brengt fig. 7, waarvan de tredenlijn de uitkomsten weergeeft van 2414 metingen van de geografische breedte van een bepaald punt in Washington; de best er bij passende normale kanskromme, die er doorheen is getekend, sluit er vrij goed bij aan (de stippen geven de theoretisch verwachte hoogten der treden aan).

Fig. 7.

De uitkomsten van 2414 metingen van één geografische breedte. (Klassebreedte 0.06 sec.)

Dat de kansdistributie van een meetfout in den regel veel op een normale lijkt behoeft ons niet te verwonderen. Wij zagen al dat de normale kansdistributie de limiet is van de binomiale en deze binomiale kansdistributie treedt op indien een getal een zeker aantal malen over een bepaald bedrag verandert, waarbij telkens aan het toeval wordt overgelaten of die verandering een toeneming of een afneming is (dit is n.l. een andere formulering voor hetgeen vroeger in termen van prijzen en nieten uit een loterij werd bedoeld). Wiskundig kan

bovendien worden bewezen dat de normale kansdistributie ook de limiet is voor die kansdistributie, die optreedt als een getal door vele oorzaken over verschillend grote bedragen wordt veranderd, mits die oorzaken onafhankelijk van elkaar al dan niet optreden en elk slechts een kleine wijziging meebrengen. En vermoedelijk is zulks het geval bij een meetfout, die immers in den regel het gezamenlijke gevolg is van vele ,,toevallig'' optredende ,,kleine'' oorzaken.

Men is daarom gewoon aan te nemen, dat zulk een meetfout (b.v. bij het meten van een lengte, een gewicht, een tijd, een hoek enz.) een normale kansdistributie heeft. Dat dit uitgangspunt niet geheel juist behoeft te zijn en de daarop gebaseerde uitkomsten dus slechts benaderende waarde kunnen hebben, is duidelijk. Zo is het b.v. bekend, dat bij het landmeten naast die meetfouten, die wel ongeveer aan bovenstaande karakteristiek beantwoorden, af en toe bij het neerleggen van het meetlint een opvallend grote fout wordt gemaakt, zodat het meetgetal, dat uiteindelijk voor de gevonden afstand tussen twee plaatsen wordt opgeschreven, o.a. bepaald is door enkele oorzaken, die elk een vrij grote invloed hadden. Dit is blijkbaar ook het geval geweest bij de metingen van de geographische breedte, die in fig. 7 zijn voorgesteld. Men stelt zich echter in den regel toch tevreden met de benaderende uitkomsten, die verkregen worden door de berekeningen te baseren op de voorstelling, dat de meetfouten een normale kansdistributie zouden hebben.

Dit uitgangspunt opent de mogelijkheid, overschrijdingskansen bij meetgetallen te bepalen op dezelfde grondslag als bij tellingen in grote steekproeven.

Daartoe dient bij de metingen in kwestie eerst te worden vastgesteld, welke normale kansdistributie het best past bij de werkelijke kansdistributie van de meetfout der toegepaste meetmethode, dat wil zeggen, wij dienen het best passende centrum en de best passende standaarddeviatie te bepalen. De formules $M = PN$ en $S^2 = PQN$, die bij het steekproefgehalte worden gebruikt, hebben hier geen zin, doordat de kansen op prijs en niet (P en Q) en het aantal loten (N) hier niet van toepassing zijn. Het is een zuiver wiskundig vraagstuk om, gegeven een aantal ,,individuele waarden'' van een ,,variabele grootheid'' (hier: een aantal verkregen meetgetallen uit ,,alle mogelijke'' meetgetallen, die met dezelfde meetmethode voor dezelfde afmeting zouden kunnen worden verkregen) de best er bij passende normale kansdistributie af te leiden. Het resultaat van deze wiskundige overwegingen is, dat de best passende middenwaarde (m) gelijk is aan het rekenkundig gemiddelde van de gegeven individuele waarden en dat de best passende standaarddeviatie gelijk is aan hun ,,strooiïng''.

Noemen wij de gegeven getallen x (welke x dus evenveel waarden heeft als er metingen zijn uitgevoerd: het zijn de meetgetallen), hun aantal n, hun som $[x]$ en hun rekenkundig gemiddelde $\bar{x}$, dan is

$$\bar{x} = [x] : n$$

(het rekenkundig gemiddelde van een aantal getallen is immers hun som gedeeld door hun aantal). De strooiïng (s_x) wordt berekend door eerst van elk getal het verschil te bepalen met dat rekenkundig gemiddelde. Stellen wij dat verschil voor door $x_{,}$ (uit te spreken als x-accent), dan geldt dus voor elk meetgetal

$$x_{,} = x - \bar{x}$$

Daarna wordt elk verschil in het kwadraat gebracht en wordt de som van al die kwadraten gedeeld door de „valentie", dat is één minder dan het aantal. Het verkregen getal, dat de „variantie" van deze groep getallen heet, is het kwadraat van de strooiïng, zodat voor de strooiïng van onze n gegeven x-waarden geldt:

$$s_x^2 = [x_{,}^2] : N \qquad \text{als } N = n - 1$$

Het best passende centrum en de dito standaard-deviatie volgen dus uit

$$\boxed{m = [x] : n}$$

$$\boxed{s^2 = [x_{,}^2] : N} \qquad \text{waarin } N = n - 1$$

Bij herhaling van de reeks metingen zouden allicht andere meetgetallen worden gevonden en daaruit zouden allicht iets andere waarden voor het best passende centrum en de best passende standaarddeviatie worden becijferd. In tegenstelling tot de werkwijze bij een steekproefgehalte zijn de verkregen getallen hier dus slechts schattingen voor de bedoelde maten, vandaar dat zij met kleine letters (m en s) werden aangeduid. De wiskundige analyse toont echter aan, dat het de beste schattingen zijn, die uit de verkregen meetgetallen kunnen worden afgeleid (zie blz. 31).

DE STANDAARDFOUT VAN DE ENKELE METING

Hoe kunnen deze beste schattingen m en s in een gegeven geval worden benut om er een overschrijdingskans mee te bepalen?

Nemen wij daartoe een reeks eenvoudige vraagstukjes, waarvan de eerste weer om didactische redenen gefingeerd zijn.

Een onderzoeker werkt over de phosphaatstofwisseling bij de mens en als maat daarvoor gebruikt hij de hoeveelheid

anorganische phosphaten, die per dag met de urine wordt afgescheiden. Wij noemen deze hoeveelheid hier kortheidshalve de phosphaatuitscheiding. Hij meent nu op gronden, die hier niet ter zake doen, dat deze bij een neger 1.5 gr. zou bedragen. Hij kan maar één keer een neger onderzoeken en vindt daarbij 1.6 gr. Is dit in strijd met zijn theorie?

In dit vraagstuk komt duidelijk uit, dat een meetgetal slechts een onnauwkeurig getal is, dat het n.l. de bedoeling slechts onnauwkeurig weergeeft: hier zijn niet alleen de meetfouten in het spel, die optreden bij het chemisch bepalen van het phosphaatgehalte in urine, noch ook van het meten van de totale hoeveelheid per 24 uur geproduceerde urine, maar hier wordt de phosphaatuitscheiding langs de nieren gemeten met de bedoeling, de phosphaatstofwisseling van de persoon in kwestie te karakteriseren, hoewel deze misschien slechts heel gebrekkig in de urine-phosphaatuitscheiding tot uitdrukking komt.

De vraag, of de geopperde hypothese in strijd is met de meting, kan op grond van de meegedeelde gegevens niet worden beantwoord: die gegevens zijn daartoe onvoldoende, doordat zij geen enkele aanwijzing bevatten omtrent de nauwkeurigheid, waarmee de phosphaatstofwisseling was gemeten.

De onderzoeker had dit al begrepen en daarom deelde hij bovendien mee, dat hij vroeger bij een blanke volgens dezelfde methode een 20-tal metingen had uitgevoerd, waarbij hij de volgende waarden voor de phosphaatuitscheiding had gevonden: 0.92, 1.04, 0.87, 0.84, 0.92, 0.87, 1.07, 0.79, 0.92, 1.05, 0.77, 1.11, 0.73, 1.03, 0.94, 1.05, 0.91, 0.84, 1.12, 1.01 gr[1]). Vormt het uiteenlopen van deze getallen, zo vraagt hij, voldoende aanwijzing voor de nauwkeurigheid van deze soort metingen?

Inderdaad, nemen wij aan dat de meetfouten bij de blanke van dezelfde aard zijn — met name dezelfde kansdistributie hebben — als die bij de neger, dan kunnen wij het door hem gegeven vraagstuk beantwoorden.

Wij onderstellen daartoe verder, dat de meetfouten een normale kansdistributie hebben (deze onderstelling zal wel onjuist zijn, maar wij nemen genoegen met het daardoor slechts benaderende karakter van de daaruit voortvloeiende uitkomsten). De beste schattingen voor centrum en standaard-deviatie zijn in achterstaande tabel berekend.

---

[1]) Deze getallen zijn niet gefingeerd; zij zijn afkomstig van Dr. A. W. J. H. HOITINK, Instituut voor Praeventieve Geneeskunde.

Berekening van gemiddelde en strooiing van de 20 meetgetallen op blz. 27, verkregen voor de dagelijkse phosphaatuitscheiding [1]).

x (in c.g.)	$x_,$ $(= x - \bar{x})$	$x_,^2$
92	— 2	4
104	+10	100
87	— 7	49
84	—10	100
92	— 2	4
87	— 7	49
107	+13	169
79	—15	225
92	— 2	4
105	+11	121
77	—17	289
111	+17	289
73	—21	441
103	+ 9	81
94	0	0
105	+11	121
91	— 3	9
84	—10	100
112	+18	324
101	+ 7	49
$[x] = 1880$		$[x_,^2] = 2528$
$n = 20$		$N = 19$
$\bar{x} = 94$ cg		$s_x^2 = 133.05$
		$s_x = 11.5$ cg

De eerste kolom bevat de gegeven meetgetallen (x), die uitgedrukt staan in cg om met hele getallen te kunnen werken. Hun som bedraagt 1880, hun aantal 20, hun gemiddelde dus 1880 : 20 of 94. Voor de tweede kolom ($x_,$) werd dit gemiddelde van elk meetgetal afgetrokken, waarna het kwadraat van elk verschil in de derde kolom ($x_,^2$) werd genoteerd. De som van deze kwadraatjes bedraagt 2528, de valentie

[1]) Een handiger methode om gemiddelde en strooiïng te becijferen, vooral als dit, zoals gewoonlijk het geval is, uit een groter aantal getallen dient te gebeuren, wordt op blz. 128 besproken.

20 — 1 = 19, de variantie (s_x^2) is dus 2528 : 19 = 133.05, zodat de strooiïng $s_x = \sqrt{133.05} = 11.5$ cg.

Indien de meetfout bij de negermeting dezelfde kansdistributie heeft als bij de blanke, dan geldt als standaard-deviatie van de best passende normale kansdistributie bij de neger dus ook $s = 11.5$ cg. En indien de te toetsen gegeven onderstelling juist zou zijn, dan zou het centrum 1.5 g of 150 cg zijn, zodat de afwijking tussen de gevonden waarde (160 cg, zie het opgegeven vraagstuk) en de theoretisch verwachte waarde 10 cg zou bedragen. Voor de excentriciteit hiervan vinden wij

$$t = d : s = 10 : 11.5 = 0.87.$$

Uit het T-P-nomogram lezen wij bij $T = 0.87$ de dubbelzijdige overschrijdingskans P_2 van bijna 0.4 af.

De overschrijdingskans blijkt dus veel groter te zijn dan de gebruikelijke grenswaarde 0.05, zodat de getoetste hypothese (dat de uitscheiding bij de neger 1.5 g zou bedragen) niet in strijd mag worden geacht met de verrichte meting (waarbij deze uitscheiding was gemeten als 1.6 g), aangenomen dat de meetfout dezelfde normale kansdistributie zou hebben gehad als in de vermelde reeks metingen bij de blanke: bij herhaling van de negermeting zou, indien zijn uitscheiding inderdaad 1.5 g zou hebben bedragen, in ongeveer de helft van alle metingen (n.l. bijna 4 op de 10) een meetfout van 0.1 g of groter te verwachten zijn, zodat het vinden van de afwijking 0.1 g geen aanleiding vormt om te denken, dat de uitscheiding bij de neger niet 1.5 g zou bedragen.

Bij de behandeling van dit vraagstuk werd de strooiïng van de meetgetallen gebruikt als een maat voor de nauwkeurigheid van een enkele meting. Deze strooiïng heet daarom de *standaardfout van de enkele meting* (vroeger de middelbare fout van de enkele meting).

Met behulp van de standaardfout (s) berekenen wij, zoals dat hierboven werd gedaan, de excentriciteit van de gevonden afwijking. Volgens de definitie is de excentriciteit de deviatie, gedeeld door de standaard-deviatie: $T = D : S$. In plaats van de standaard-deviatie (S), die ons niet bekend is, gebruiken wij echter de beste schatting daarvoor, die wij weer met kleine letters aanduiden: $t = d : s$. Is die schatting uit vele meetgetallen berekend, dan zal zij wel zo dicht bij de werkelijke waarde van T liggen, dat de overschrijdingskans, die daarbij in het T-P-diagram wordt afgelezen, een voor de praktijk voldoende benadering van de werkelijke overschrijdingskans vormt. Is het aantal meetgetallen echter klein, dan is dit niet meer het geval. Om dan toch zonder verder cijferwerk de beste schatting voor de overschrijdingskans te weten te komen wordt deze niet uit het T-P-nomogram afgelezen, maar uit het t-N-P-nomogram, dat op blz. 150

is afgebeeld: hierin is, behalve de excentriciteit, die uit $t = d : s$ becijferd werd, en de te bepalen overschrijdingskans P_2 ook nog de valentie N uitgezet (van dit diagram is de bovenste horizontale lijn, gemerkt $N = \infty$, identiek met het T-P-nomogram).

De hierboven gegeven oplossing van ons vraagstuk is dus niet helemaal juist: het aantal meetgetallen is slechts 20, zodat de overschrijdingskans niet behoorde te worden afgelezen uit het T-P-nomogram (fig. 6) maar uit het t-N-P-nomogram (blz. 150). Veel verschil maakt dat in ons geval intussen niet, want uit dat nomogram lezen wij bij $t = 0.87$ en $N = 20 - 1 = 19$ de dubbelzijdige overschrijdingskans $P_2 = 0.4$ af, hetgeen nauwelijks groter is dan de waarde, die bij $N = \infty$ en dus uit het T-P-nomogram werd afgelezen.

DE STANDAARDFOUT VAN EEN GEMIDDELDE ALS BESTE SCHATTING

Bekijken wij nu een volgend vraagstuk.

> Dezelfde onderzoeker meent (op grond van theoretische overwegingen, die hier niet ter zake doen) dat de phosphaatstofwisseling bij een blanke proefpersoon zou zijn gekarakteriseerd door een uitscheiding van 1.0 g en hij vindt bij 20 metingen, die hij op verschillende dagen bij die persoon uitvoert, de getallen, die in het vorige vraagstuk werden vermeld. Is de door hem geopperde theorie met deze meetgetallen in strijd?

Blijkbaar is hier uitgegaan van de mening, dat de phosphaatstofwisseling constant is en het uiteenlopen der meetgetallen berust op de onnauwkeurigheid, waarmee die getallen de mate dier stofwisseling aanduiden [1]). Onder deze gezichtshoek bekeken, heeft hij eenzelfde grootheid 20 maal gemeten, zodat hij over 20 schattingen voor die grootheid beschikt. Om ons vraagstuk te beantwoorden, dienen eerst deze 20 „elementaire schattingen" te worden gecombineerd tot één gecombineerde schatting dier grootheid, om dan van deze ene „beste schatting" de overschrijdingskans te bepalen.

Als „beste schatting" zouden wij natuurlijk graag die gebruiken, die het dichtst bij de werkelijkheid komt, maar wij kennen die werkelijkheid niet, zodat dit kenmerk geen bruikbare definitie levert van het begrip: beste schatting. Een bruikbare definitie vloeit voort uit de volgende overweging.

---

[1]) Ter wille van het betoog wordt hier niet gesproken over het andere gezichtspunt, dat de bedoelde stofwisseling zelf van dag tot dag zou wisselen. De conclusies zouden dan echter dezelfde zijn.

Nemen wij ter vereenvoudiging de vier eerste meetgetallen: 92, 104, 87 en 84. Het zou kunnen zijn dat de werkelijkheid ligt bij 90. In dat geval zouden de meetfouten zijn geweest +2, +14, —3 en —6. Maar het kan ook zijn dat de werkelijkheid b.v. 80 is, in welk geval de meetfouten zouden zijn +12, +24, +7 en +4. Het is duidelijk dat de kans op de eerstgenoemde meetfouten groter is dan die op de laatstgenoemde (de laatstgenoemde zijn alle positief en bovendien zijn ze groter dan de eerstgenoemde). Wij zijn dus geneigd, 90 een betere schatting te noemen dan 80. *En zo wordt de beste schatting genoemd die eventuele werkelijkheid, die meer kans biedt op het optreden van de gevonden meetgetallen dan enige andere werkelijkheid.*

Het woord is weer aan de wiskunde om ons een formule te leveren voor de beste schatting, die uit deze definitie volgt voor een aantal meetgetallen, gesteld dat deze een normale kansdistributie hebben. Die formule blijkt gelukkig heel eenvoudig te zijn: het is het rekenkundig gemiddelde dier meetgetallen (hetgeen trouwens hetzelfde betekent als het reeds meegedeelde feit, dat dit gemiddelde gelijk is aan het centrum van de best passende normale kansdistributie).

Uit de 20 metingen volgt dus als beste schatting van „de" phosphaatuitscheiding 94 cg (zie blz. 28).

Bij herhaling van deze metingen zou allicht weer een iets andere waarde van het gemiddelde worden gevonden, zodat ook het gemiddelde een variabele grootheid is. Voor het geval dat de meetgetallen zelf een normale kansdistributie hebben, kan wiskundig worden aangetoond, dat ook hun gemiddelde een normale kansdistributie heeft; de variantie daarvan is n-maal kleiner dan die van de meetgetallen zelf, zodat

$$\boxed{s_{\bar{x}}^2 = s_x^2 : n}$$

Reeds vonden wij op blz. 28 voor onze 20 getallen $s_x^2 = 133.05$, zodat $s_{\bar{x}}^2 = 133.05 : 20 = 6.6525$ of $s_{\bar{x}} = \sqrt{6.6525} = 2.58$ cg. Deze „strooiing" van het gemiddelde draagt de naam: *standaardfout van het gemiddelde.* Het is de beste schatting voor de standaard-deviatie van de normale kansdistributie voor het gemiddelde.

Wij weten nu alles wat nodig is voor de beantwoording van ons vraagstuk. De theoretisch verwachte waarde van de phosphaatuitscheiding is 100 cg, uit de metingen volgt als beste schatting 94 cg, de afwijking tussen de theoretisch verwachte en de gevonden waarde bedraagt dientengevolge 100 — 94 = 6 cg, zodat de excentriciteit volgt uit $t = d : s = 6 : 2.58 = 2.32$. Het t-N-P-nomogram levert bij $t = 2{,}32$ en $N = 19$ de dubbelzijdige overschrijdingskans $P_2 = 0.03$, dat is kleiner dan de traditionele grens van 0.05. Bij de gebruikelijke

betrouwbaarheidsdrempel dient aan dit verschil dus wel degelijk betekenis te worden toegekend: in minder dan 1 op de 20 keren (n.l. in ongeveer 1 op de 33 keren) kan bij herhaling van de meetreeks een dergelijke of nog grotere afwijking worden verwacht als gevolg van de meetfouten. Bij deze proefpersoon bedraagt de uitscheiding waarschijnlijk niet 1.00 gram.

STANDAARDFOUT EN BETROUWBAARHEIDSGRENZEN

Niet steeds is een metingenreeks bedoeld om er onmiddellijk een hypothese aan te toetsen. Dan wordt als uitkomst zowel de beste schatting van het gemetene vermeld als ook de standaardfout. Als resultaat van de eerstgenoemde metingenreeks van de phosphaat-uitscheiding (zonder vitaminetoediening) zou b.v. worden gepubliceerd, dat deze uitscheiding bij 20 metingen werd gevonden als 94 $\pm$ 2.6 cg, waarbij het getal vóór het $\pm$ teken de beste schatting en dat er na de standaardfout voorstelt.

Het vermelden van de standaardfout heeft een groot nut. Vagelijk zal degene, die leek op statistisch terrein is, er reeds enige kijk door krijgen op de verkregen nauwkeurigheid, waarmee het gevonden getal de grootte der gemeten afmeting aanduidt. Voor dengene, die thuis is in het statistisch rekenen, valt deze vaagheid weg. Hij kan n.l., als later een hypothese wordt opgesteld volgens welke deze gemeten grootheid een bepaalde waarde zou moeten hebben, deze hypothese onmiddellijk toetsen aan de destijds gepubliceerde meetresultaten: hij kan dan immers de afwijking vinden tussen de door hem theoretisch verwachte en de vroeger gepubliceerde gevonden waarde en hij kan tevens met behulp van de eveneens gepubliceerde standaardfout de excentriciteit en dus ook de overschrijdingskans bepalen, zodat hij op grond daarvan een oordeel kan uitspreken over het al dan niet in strijd zijn van de nieuwe hypothese met de oude metingen. De eerste publicatie kan dientengevolge worden beschouwd als *een nog niet volledig statistisch onderzoek*, n.l. als een onderzoek, dat pas volledig zal worden als er later een hypothese aan zal worden getoetst.

Alle hypothesen, die later aan de gepubliceerde uitkomst zouden kunnen worden getoetst, kunnen reeds tevoren worden verdeeld in zulke, die dan zullen worden geacht in strijd met deze uitkomst te zijn en de overige, die er niet door zullen worden weersproken: elke afwijking, die groter is dan 2 maal de standaardfout, heeft volgens fig. 6 een kleinere overschrijdingskans dan bij de betrouwbaarheidsdrempel behoort ($^1/_{20}$) en zijn dus significant, alle kleinere zijn insignificant. Alle hypothesen, volgens welke de gemeten

grootheid een waarde zou hebben; groter dan de beste schatting minus 2 maal de standaardfout en kleiner dan de beste schatting plus 2 maal de standaardfout, zullen dus als betrouwbaar worden betiteld, terwijl die, volgens welke de waarde buiten deze grenzen zou vallen, als onbetrouwbaar zullen worden afgewezen (in het zoëven aangehaalde geval zou een hypothese betrouwbaar worden geacht als deze voor de uitscheiding een waarde tussen $94 - 2 \times 2.6$ cg en $94 + 2 \times 2.6$ cg — dus tussen 88.6 en 99.2 cg — voorspelt en onbetrouwbaar als zij een waarde buiten die grenzen aangeeft). Vandaar dat deze grenzen (de beste schatting + en — 2 maal de standaardfout) als de *betrouwbaarheidsgrenzen* bekend staan. Dit zijn dan die bij een betrouwbaarheidsdrempel, overeenkomende met een overschrijdingskans van 5 %. Het zal duidelijk zijn, hoe de betrouwbaarheidsgrenzen voor een willekeurig te kiezen andere drempel worden bepaald. Zo wordt een overschrijdingskans (P_2) van 0.001 volgens het T-P-nomogram gevonden bij een excentriciteit $T = 3^1/_3$, zodat de betrouwbaarheidsgrenzen daarbij zijn de beste schatting + en — $3^1/_3$ maal de standaardfout.

De betekenis van de standaardfout kan nog iets eenvoudiger en dus sprekender, zij het ook iets minder exact, onder woorden worden gebracht. Een betrouwbaarheid, overeenkomend met een overschrijdingskans van 5 % of $^1/_{20}$ houdt in zekere zin de theoretische verwachting in, dat van 20 getoetste hypothesen met deze betrouwbaarheid er 19 juist en 1 fout zijn. (Geheel exact is deze uitspraak niet: zij eist de toepassing van het beginsel der „inverse waarschijnlijkheid" en dit beginsel wordt door velen op grond van algemene overwegingen van de hand gewezen.) Volgens deze theorieën ligt de werkelijke waarde tussen de betrouwbaarheidsgrenzen, gelijk aan de beste schatting — en + 2 maal de standaardfout. Dat wil dus zeggen, dat (bij normale kansdistributie) theoretisch kan worden verwacht, dat de werkelijke waarde in 19 op de 20 gevallen tussen deze twee grenzen ligt, dat m.a.w. de werkelijke waarde in 19 op de 20 gevallen minder dan 2 maal de standaardfout van de beste schatting verschilt en in 1 op de 20 gevallen er meer van verschilt. Berekenen wij hetzelfde voor een aantal betrouwbaarheidsdrempels, dan krijgen wij het volgende lijstje, dat de eenvoudigste manier is om de practische betekenis van de standaardfout als aanduiding van de verkregen nauwkeurigheid onder woorden te brengen, zij het dan ook dat deze wijze van formuleren misschien niet geheel exact is.

Bij ten naastenbij normale kansdistributie
is te verwachten dat
het verschil tussen werkelijkheid en beste schatting

in 1	op de	2	gevallen	meer	dan	$^2/_3$
„ 1	„ „	3	„	„	„	1
„ 1	„ „	20	„	„	„	2
„ 1	„ „	1000	„	„	„	$3^1/_3$

maal zo groot is als de standaardfout.

SIMULTANE KANSDISTRIBUTIES

In elk vraagstuk, dat wij tot nu toe ten tonele voerden, was telkens slechts sprake van één enkele kansdistributie. Bij de meeste vraagstukken dienen echter tegelijkertijd twee of meer kansdistributies onder ogen te worden gezien. Dit maakt het nodig, naast de theorie van de „solitaire" kansdistributie, die het onderwerp van de voorafgaande paragrafen uitmaakte, die van de „simultane" kansdistributies te behandelen.

Een zeer schematisch gesteld (en daarom weer gefingeerd) vraagstuk moge dit toelichten.

Een chemicus opperde het vermoeden, dat de verschillende atomen van een zekere organische verbinding op een bepaalde regelmatige wijze als in een soort kristalrooster gerangschikt zouden liggen. Zijn vermoeden zou slechts juist kunnen zijn als twee bepaalde afstanden in dat rooster, die hij de lengte en de breedte van de roostercel noemde, nauwkeurig gelijk zouden zijn aan 12 en 6 Å [1]). Om dat vermoeden te toetsen werden deze lengte en breedte elk 40 maal gemeten met als resultaat: lengte 12.1 $\pm$ 0.6 Å en breedte 6.7 $\pm$ 0.3 Å. Is het hier te toetsen vermoeden in strijd met deze metingen?

Bij oppervlakkige beschouwing lijkt het, dat dit vraagstuk zou kunnen worden gesplitst in twee afzonderlijke vragen:

1. is de onderstelling, dat de lengte 12 Å zou bedragen, in strijd met het meetresultaat 12.1 $\pm$ 0.6 Å en
2. is de onderstelling, dat de breedte 6 Å zou bedragen, in strijd met het meetresultaat 6.7 $\pm$ 0.3 Å?

---

[1]) 1 Å = 1 Angstrom = 10^{-10} meter of 0.1 millimicron.

Elk van deze twee vragen zou gemakkelijk kunnen worden beantwoord op de manier, die in de vorige paragrafen uiteengezet is, want in elk er van treedt slechts één solitaire kansdistributie op:

1. Voor de lengte bedraagt de beste schatting, die uit de metingen volgt, volgens de opgaaf 12.1 Å en de theoretisch verwachte waarde 12.0 Å, de afwijking tussen beide is dus 0.1 Å. Daar de standaardfout van de beste schatting volgens de opgaaf 0.6 Å bedraagt, is de excentriciteit $t = d : s = 0.1 : 0.6 = 0.17$ en de bovenste lijn van het t-N-P-nomogram (het aantal metingen is groter dan 30) geeft bij $T = 0.17$ een overschrijdingskans $P_2 = 0.85$. De lengtemetingen blijken dus niet in strijd te zijn met het vermoeden, dat de lengte 12 Å zou bedragen.
2. Voor de breedte is de afwijking $6.7 - 6.0 = 0.7$ Å en de excentriciteit $0.7 : 0.3 = 2.33$, waarbij de overschrijdingskans $P_2 = 0.02$ of iets kleiner dan de traditionele betrouwbaarheidsgrens van 0.05, zodat de breedtemetingen wel in strijd moeten worden geacht met het vermoeden, dat de kern 6.0 Å breed zou zijn.

Door ons oorspronkelijke vraagstuk in tweeën te splitsen, krijgen wij dus twee tegengestelde antwoorden: wat de lengte betreft „niet in strijd", wat de breedte betreft „wel in strijd". Hoe moet nu onze eindconclusie luiden?

De lezer zou misschien geneigd zijn de eindconclusie te trekken, dat het te toetsen aangeboden vermoeden in strijd moet worden geacht met de verrichte metingen, omdat de vermoede breedte in strijd zou zijn met de breedte-metingen, zodat althans een deel van het vermoeden in strijd zou zijn met een deel van de metingen. Maar tegen die opvatting zijn twee ernstige bezwaren in te brengen. Tot nu toe werd een conclusie steeds op *alle* desbetreffende metingen gebaseerd en werden er niet bepaalde metingen uitgeschakeld. In het algemeen moet het zelfs niet toelaatbaar worden geacht, bepaalde metingen uit te schakelen: door van een reeks meetgetallen achteraf b.v. de ongunstigste te schrappen zou het resultaat willekeurig gunstiger worden gemaakt en door de gunstigste te schrappen zou het eveneens zonder behoorlijk motief ongunstiger worden. Is het dan in het onderhavige voorbeeld vergund, achteraf de lengte-metingen buiten beoordeling te laten, omdat ze gunstig zijn, en de conclusie uitsluitend te baseren op de ongunstig gebleken breedte-metingen? En in de tweede plaats zouden wij, door de conclusie uitsluitend te baseren op de breedte-metingen, geen rekening houden met de merkwaardige tegenstelling, dat de breedte-metingen de betrouwbaarheidsgrens slechts weinig overschrijden, maar de lengte-metingen ver binnen die grenzen zijn uitgevallen. Om het plastisch — zij het misschien weinig exact — uit te

drukken: de breedte-metingen pleiten maar zwak tegen het getoetste vermoeden, de lengte-metingen pleiten er sterk voor. In deze lijn doordenkend zouden wij willen zoeken naar een soort gemiddelde overschrijdingskans, gelegen ergens tussen de twee afzonderlijke overschrijdingskansen, gevonden uit de twee solitaire kansdistributies, in. Maar dan zou de splitsing van ons vraagstuk niet meer gehandhaafd zijn en bovendien zouden wij dan toch nog niet weten, welk soort gemiddelde hier logischerwijze zou moeten worden gekozen.

Al deze moeilijkheden zijn het gevolg van het splitsen van ons vraagstuk in tweeën en daaruit blijkt wel, dat de twee kansdistributies, die hier in het spel zijn, niet mogen worden beschouwd als twee solitaire kansdistributies, maar dat zij als een simultane kansdistributie dienen te worden opgevat.

De theorie van de simultane kansdistributies geeft een logische weg aan om ons antwoord te vinden zonder het vraagstuk te splitsen en deze weg berust op dezelfde algemene gedachtengang als gevolgd werd bij het bepalen van de overschrijdingskans in een solitaire kansdistributie. De (dubbelzijdige) overschrijdingskans in een solitaire kansdistributie is (zie blz. 10 en 16) de kans op een afwijking met een even grote kans als de gevonden afwijking of met een kleinere kans. Bij ons tweetal kansdistributies (die van de lengte en die van de breedte) hebben wij niet met één afwijking te maken maar met *een combinatie van twee afwijkingen* (n.l. die in de lengtematen en die in de breedtematen) en hier is de overschrijdingskans dus de kans op een *combinatie* (van een lengte- en een breedteafwijking) met dezelfde kans als de gevonden combinatie of met een kleinere kans.

Uit de (solitaire) kansdistributie van de lengtematen zou gemakkelijk de kans kunnen worden berekend, dat juist de gevonden afwijking in lengte zou optreden en zo zou uit de (solitaire) kansdistributie van de breedtematen de kans op de gevonden afwijking in breedte volgen (tenminste als beide kansdistributies als normaal kunnen worden opgevat, zoals bij deze metingen aangenomen wordt). Volgens het op blz. 11 ingevoerde theorema van de gecombineerde kans zou de kans van de gevonden combinatie (de kans op het samen of simultaan optreden van de gevonden lengte-afwijking en de gevonden breedte-afwijking) gelijk zijn aan het product van de twee solitaire kansen. De overschrijdingskans in onze simultane kansdistributies is dan gelijk aan deze kans op de gevonden combinatie en op al die combinaties, die volgens overeenkomstige berekeningen een nog kleinere kans hebben.

Een grafische voorstelling van deze kansen kan ons helpen om in te zien, van welke factor de grootte van deze simultane overschrij-

dingskans afhangt. Dit kan echter geen vlakke — twee-dimensionale — grafische voorstelling zijn, het is in ons geval van twee kansdistributies een drie-dimensionale, dus ruimtelijke figuur.

De weg daartoe wordt gewezen door in een twee-dimensionaal schema (tabel V) voor twee variabele grootheden X en IJ bij een aantal individuele waarden de (hier min of meer willekeurig gekozen) kansen op te schrijven, voor de grootheid X horizontaal, voor IJ loodrecht daarop. In het daardoor bepaalde vierkant kan dan in elk hokje de kans worden genoteerd voor de combinatie van een bepaalde waarde van X en een bepaalde waarde van IJ, eenvoudig door de kansen op die twee bepaalde waarden met elkaar te vermenigvuldigen. B.v. als de kans, dat X de waarde 16 heeft, 0.2 bedraagt en de kans, dat IJ de waarde 103 heeft, 0.4 is, dan is de kans op de combinatie $X = 16$ en $IJ = 103$ volgens het axioma van de gecombineerde kans $0.2 \times 0.4 = 0.08$. (De som van al deze kansen is natuurlijk weer 1.)

TABEL V

De kans op elke mogelijke combinatie van een X- en een IJ-waarde, van welke waarden de solitaire kansen buiten het vierkant vermeld staan

Y	KANS					
101	0.1	0.01	0.02	0.04	0.02	0.01
102	0.2	0.02	0.04	0.08	0.04	0.02
103	0.4	0.04	0.08	0.16	0.08	0.04
104	0.2	0.02	0.04	0.08	0.04	0.02
105	0.1	0.01	0.02	0.04	0.02	0.01
		0.1	0.2	0.4	0.2	0.1 KANS
		15	16	17	18	19 X

Door nu op elk veldje een blok te zetten, waarvan het volume (in een willekeurig groot te kiezen eenheid) gelijk is aan de daarin vermelde kans, ontstaat de ruimtelijke grafische voorstelling dezer gecombineerde kansen die in fig. 8 is weergegeven.

Hier moge de opmerking worden tussengevoegd, dat de overschrijdingskans van een willekeurig te kiezen combinatie hier reeds

gemakkelijk uit tabel V kan worden gevonden. Volgens die tabel heeft de combinatie $X = 18$ en $IJ = 104$ een kans van 0.04 en alle veldjes, waarin een even grote of kleinere kans vermeld staat, werden grijs gekleurd: hun som, 0.52, is de overschrijdingskans dezer combinatie. In de grafische voorstelling van fig. 8 is de overschrijdingskans gelijk aan het gezamenlijk volume van alle blokken, die even hoog of lager zijn dan dat van de bedoelde combinatie.

In het voorbeeld van fig 8 zijn de kansen, zoals gezegd, vrij willekeurig gekozen (n.l. in de buurt van wat de normale kansdistributie zou voorschrijven). Bovendien zijn de grootheden X en IJ hier gekozen als discontinu (sprongsgewijs) variabel. In de statistiek worden de

Fig. 8.
Grafische voorstelling van de kansen, bedoeld in Tabel V.

meeste berekeningen, zoals wij zagen, uitgevoerd op grond van de benaderende onderstelling, dat de solitaire kansdistributies der gegeven getallen normaal zouden zijn. Nemen wij twee normale kansdistributies als uitgangspunt, dan ontstaat op soortgelijke wijze de drie-dimensionale grafische voorstelling van fig. 9, de „GAUSS-berg".

In deze figuur zijn langs de X- en de IJ-as niet de meetwaarden zelf afgezet, maar hun excentriciteiten, dus hun deviatie gedeeld door hun standaard-deviatie volgens $T = D : S$. Daardoor wordt elke horizontale doorsnee door de GAUSSberg een cirkel. (Anders zouden het ellipsen zijn, waarvan de lange en de korte as zich verhouden als de twee standaard-deviaties.)

Voor een twee-dimensionale normale kanskromme kan de overschrijdingskans van een bepaalde X-waarde worden aangegeven, n.l.

op de manier zoals dat in fig. 4 gedaan is. In een drie-dimensionale GAUSS-berg kan dat worden gedaan voor elke willekeurig te kiezen combinatie van een X- en een IJ-waarde. Die twee waarden bepalen een punt van het grondvlak, b.v. punt A in fig. 10. Een loodlijn, in dat punt op het grondvlak opgericht, snijdt het oppervlak van de berg in een bepaald punt (B) en de hoogte, AB, van dat punt B boven het

Fig. 9.
De (standaard-) GAUSS-berg.

grondvlak is een aanduiding van de kans dier combinatie. De horizontale cirkel op het bergoppervlak, die door dat punt gaat, duidt op die wijze dus alle combinaties van X- en IJ-waarden met gelijke kans aan. En alle lager op het bergoppervlak gelegen punten — die tevens verder van de as liggen — duiden combinaties met kleinere kansen aan, zodat de overschrijdingskans gelijk is aan dat deel van de GAUSS-berg, dat gelegen is buiten de getekende cylindermantel, dat is de cylinder, waarvan de as samenvalt met de as van de GAUSS-berg en die verder door het punt A gaat, dat de bedoelde combinatie van een X- en een IJ-waarde representeert.

Om het korter te zeggen: door fig. 4 om zijn as te wentelen wordt een GAUSS-berg beschreven, waarbij het gestippelde deel van de wente-

lende figuur de ruimte beschrijft, die gelijk is aan de bedoelde overschrijdingskans.

Hieruit blijkt, dat de overschrijdingskans uitsluitend afhangt van de afstand tussen de as en het punt, dat de gegeven combinatie van een X- en een IJ-waarde representeert, in fig. 10 de afstand OA. Die afstand is immers de straal van de bedoelde cylindermantel. En de grootte van die straal, die wij daarom de *simultane excentriciteit* noemen, is gemakkelijk te vinden. In het grondvlak is die straal de

Fig. 10.

$G^2 = T^2_x + T^2_{ij}$.

diagonaal van een rechthoek, waarvan de zijden gelijk zijn aan de twee gegeven excentriciteiten. Duiden wij die straal aan door de letter G (oorspronkelijk gebruikte PEARSON hiervoor de Griekse letter χ) dan is dus volgens de stelling van PYTHAGORAS $G^2 = T^2_x + T_y^2$. De overschrijdingskans van onze combinatie hangt dus uitsluitend af van de som van de kwadraten der excentriciteiten, die onze twee X- en IJ-waarden in hun eigen kansdistributie hebben, en de wortel uit deze som noemen we de simultane excentriciteit (G) onzer combinatie.

De formule, volgens welke de overschrijdingskans uit een gegeven

simultane excentriciteit G kan worden berekend, eist weer veel cijferwerk en daarom is ook hiervoor een tabel opgesteld, die bij een aantal waarden van G^2 de bijbehorende overschrijdingskans aangeeft.

Het kost dus maar weinig cijferwerk om in een geval van zulk een simultane kansdistributie de overschrijdingskans te bepalen: van elk der twee beste schattingen ($\bar{x}$ en $\bar{y}$) wordt de excentriciteit berekend (uit $t_{\bar{x}} = d_{\bar{x}} : s_{\bar{x}}$ en $t_{\bar{y}} = d_{\bar{y}} : s_{\bar{y}}$) waarna hun kwadraten bij elkaar worden opgeteld (volgens $g^2 = t^2_{\bar{x}} + t^2_{\bar{y}}$). De G^2-N-P-tabel levert dan bij deze waarde van G^2 de gevraagde overschrijdingskans.

Dezelfde regel als hierboven voor 2 simultane kansdistributies werd afgeleid, geldt voor elk aantal (N) van zulke kansdistributies, mits zij, zoals in ons voorbeeld, normaal zijn en onafhankelijk van elkaar (waarover later). Er zijn dan meer solitaire excentriciteiten (n.l. N), het kwadraat van de simultane excentriciteit is dan de som van de kwadraten van al die excentriciteiten. De overschrijdingskans blijkt dan echter niet alleen van de grootte dier simultane excentriciteit af te hangen, maar ook van haar valentie, dat is het aantal (N) van die kansdistributies [1]). Deze valentie kan duidelijkheidshalve als index aan de letter G worden toegevoegd, zodat voor een simultane kansdistributie, die uit N normale kansdistributies bestaat, geldt:

$$G^2_N = T^2_1 + T^2_2 + \ldots\ldots + T^2_N.$$

Op blz. 151 staat het z.g. G^2-N-P-nomogram, waarin het kwadraat van de simultane excentriciteit (G^2) horizontaal is uitgezet en de valentie (N) verticaal. Een uit de gegeven getallen becijferde waarde van G^2

---

[1]) Om dit af te leiden kan moeilijk de grafische voorstelling te hulp geroepen worden, omdat deze bij meer dan 2 kansdistributies meer dan 3-dimensionaal zou moeten zijn. Het volgt echter eenvoudig uit de tot nu toe niet genoemde formule, waarmee GAUSS zijn normale kansdistributie definieerde:

$$I^2 = 1 : 2\,\pi\, e^{T^2}$$

waarin I de kansdichtheid bij de gekozen waarde van T voorstelt (de kansdichtheid is de limiet, waartoe de kans op een interval van de X-as, gedeeld door de grootte van dat interval, onbeperkt nadert als dat interval steeds kleiner wordt gekozen). De kansdichtheid van een combinatie van de T-waarden in N onafhankelijke normale kansdistributies is volgens het axioma der gecombineerde kans gelijk aan het product van de solitaire kansdichtheden, dus gelijk aan

$$1 : (2\,\pi)^N\, e^{T^2_1 + T^2_2 + \ldots. + T^2_N},$$

zodat de overschrijdingskans uitsluitend afhangt van N en van G^2 (zijnde $T^2_1 + T^2_2 + \ldots\ldots + T^2_N$). De overschrijdingskans voor $G = g$ bedraagt n.l.

$$\int_g^{+\infty} d_x\, d_y : (2\,\pi)^N\, e^{G^2}$$

en een dito van N wijzen samen in dit rechthoekige coördinatensysteem op de gebruikelijke manier een bepaald punt aan en uit de ligging van dat punt ten opzichte van de schuine lijnen lezen wij de éénzijdige overschrijdingskans P_1 af, die aan het eind van die schuine lijnen is genoteerd.

Het beantwoorden van ons vraagstuk omtrent de lengte en de breedte van de roostercel eist dus slechts weinig werk. Voor de lengtemetingen vonden wij de excentriciteit $t = 0.17$ en voor de breedtemetingen $t = 2.33$, zodat de simultane excentriciteit volgt uit

$$g^2_2 = t^2_x + t^2_y = 0.17^2 + 2.33^2 = 5.5$$

en bij $G^2 = 5.5$ en $N = 2$ lezen wij uit het G^2-N-P-nomogram een éénzijdige overschrijdingskans van ongeveer 0.07 af, zodat de dubbelzijdige 0.14 bedraagt of duidelijk groter dan de traditionele betrouwbaarheidsdrempel van 0.05.

Het vermoeden, dat de cel 12 Å lang en 6 Å breed zou zijn, is dus niet in strijd met de verrichte metingen (in ongeveer 1 op de 7 maal kan worden verwacht dat de meetfouten even ongunstige of nog ongunstigere afwijkingen veroorzaken).

De simultane excentriciteit (die door PEARSON is opgesteld) wordt zeer veel toegepast, niet alleen doordat vaak een aantal onafhankelijke kansdistributies tegelijk in het spel zijn maar ook doordat wiskundige analyse heeft aangetoond, dat vele andere situaties eveneens tot een simultane excentriciteit kunnen worden teruggebracht. De volgende paragrafen geven daar enkele voorbeelden van, waarbij eindelijk vraagstukken uit de praktijk kunnen worden gekozen.

STEEKPROEF UIT MEER DAN TWEE SOORTEN INDIVIDUEN

Bij het onderzoek naar ééneiïge tweelingen (zie blz. 8) werden ook drielingen geteld; van de 320 drielingworpen bestonden 46 uit 3 mannetjes, 106 uit 2 mannetjes en 1 vrouwtje, 133 uit 1 mannetje en 2 vrouwtjes en 35 uit 3 vrouwtjes. Pleit dit tegen eisplijting?

Ter wille van ons verhaal stellen wij de vraag iets anders: zijn deze aantallen in strijd met de mening, dat de kansen op deze 4 soorten worpen overeenkomen met de bevruchtingskansen, dus met de kansen $^1/_8$, $^3/_8$, $^3/_8$ en $^1/_8$? [1])

---

[1]) Deze bevruchtingskansen volgen n.l. op eenvoudige wijze uit de regels van MENDEL: bij drie onmiddellijk op elkaar volgende bevruchtingen zijn er 8 geslachtsverdelingen mogelijk, n.l. *mmm, vmm, mvm, mmv, mvv, vmv, vvm* en *vvv* en

Gesteld dat de 4 soorten worpen inderdaad deze kansen hebben, dan zouden in een steekproef van 320 worpen theoretisch verwacht worden $^1/_8 \times 320 = 40$ worpen „3*m*", $^3/_8 \times 320 = 120$ worpen „2*m*, 1*v*", en dus ook 120 worpen „1*m* 2*v*" en 40 worpen „3*v*". De getelde aantallen geven dus afwijkingen te zien van +6, —14, +13, en —5 worpen, zoals overzichtelijk nog eens in de eerste 5 kolommen van tabel VI te zien is.

Beschouwen we eerst de worpen, bestaande uit louter mannetjes, dan kunnen we deze vergelijken met de „prijzen" en dan vormen de overige drie soorten worpen samen de „nieten", zoals dat op blz. 10 in ons eerste vraagstuk werd genoemd. In steekproeven van deze grootte uit dit materiaal heeft het gehalte aan 3*m*-worpen dus een binomiale en ten naastenbij een normale kansdistributie. Op dezelfde wijze kunnen we achtereenvolgens elk van de andere soorten als „prijs" beschouwen, daarbij telkens de overige drie samen als „nieten" opvattend; daaruit blijkt dat het gehalte van elke soort worpen een ten naastenbij normale kansdistributie moet hebben, zodat wij hier een geval hebben van vier normale kansdistributies.

Intussen is deze situatie niet volledig vergelijkbaar met een 4-waardige simultane kansdistributie zoals die in de vorige paragraaf werd behandeld, want hier geldt het één steekproef uit een materiaal van 4 soorten of „klassen" van individuen en daar gold het 4 steekproeven elk uit een ander materiaal. De 4 kansdistributies zijn hier niet onafhankelijk van elkaar, zoals b.v. blijkt uit het feit, dat als van de eerste drie soorten in de steekproef van 320 individuen (hier worpen) een bepaald totaal aanwezig is, daaruit reeds vaststaat, hoeveel er van de vierde soort in voorkomen.

Wiskundig kan deze situatie toch tot een simultane excentriciteit worden herleid, dat blijkt dan echter een te zijn, waarvan de valentie kleiner is dan het aantal klassen: valentie en grootte volgen uit de formule:

$$\boxed{G^2_N = [d^2 : f_v]} \quad \text{waarin } N = k - c - 1.$$

Deze simultane excentriciteit kan uit de gegeven getallen worden becijferd. Voor elke klasse afzonderlijk wordt eerst uitgerekend, hoeveel individuen er volgens de te toetsen hypothese theoretisch in te verwachten zijn: dit theoretisch verwachte aantal, f_v, is gelijk

---

deze hebben alle 8 dezelfde kans, dus elk $^1/_8$. En van deze verdelingen zijn er: 1 van 3 mannetjes, 3 van 2 mannetjes en 1 vrouwtje, 3 van 1 mannetje en 2 vrouwtjes en 1 van 3 vrouwtjes, zodat de kansen van deze drielingcombinaties achtereenvolgens $^1/_8$, $^3/_8$, $^3/_8$ en $^1/_8$ bedragen.

aan het product van de kans op deze klasse, P_{kl}, en het totale aantal individuen van de steekproef, n, volgens

$$f_v = P_{kl}\, n.$$

Dan wordt voor die klasse de afwijking, d, bepaald, dat is het verschil tussen het aantal individuen dat in die klasse van de steekproef was gevonden, f, en het theoretisch daarin verwachte aantal, f_v, volgens

$$d = f - f_v.$$

Het kwadraat van deze afwijking wordt gedeeld door het theoretisch verwachte aantal, waarmee voor deze klasse het getal $d^2 : f_v$ is verkregen. Nadat dit voor elke klasse is becijferd, worden deze getallen bij elkaar opgeteld: de verkregen som is de gezochte waarde van G^2_N. Volgens $N = k - c - 1$ is de valentie gelijk aan het aantal klassen, k, minus het aantal te schatten constanten, c, minus 1. De bedoelde constanten komen in ons vraagstuk niet voor en worden daarom pas later besproken.

TABEL VI

Geslachtsverdeling bij drielingen in een steekproef van 320 worpen

Geslacht	Kans P_{kl}	Theoretisch verwachte frequentie f_v $(= P_{kl}n)$	In steekproef gevonden frequentie f	Afwijking d $(= f - f_v)$	d^2	$d^2 : f_v$
3 *m* 0 *v*	$^1/_8$	40	46	+ 6	36	0.900
2 *m* 1 *v*	$^3/_8$	120	106	— 14	196	1.633
1 *m* 2 *v*	$^3/_8$	120	133	+ 13	169	1.408
0 *m* 3 *v*	$^1/_8$	40	35	— 5	25	0.625
	1	320	320	0	$G^2_3 = 4.566$	
					$P_1 = 0.3$	

Voor ons vraagstuk werd deze simultane excentriciteit in tabel VI berekend met als resultaat $G^2 = 4.566$. De valentie volgde uit $N = k - c - 1 = 4 - 0 - 1 = 3$. Het G^2-N-P-nomogram (blz. 151) geeft bij $G^2 = 3.6$ en $N = 3$ een overschrijdingskans $P_1 = 0.3$. Daar deze belangrijk groter is dan 0.05 wordt de onderstelling, dat de geboortekansen gelijk zouden zijn aan de bevruchtingskansen, niet in strijd met de genomen steekproef geacht: als gevolg van de toevallige

steekproefselectie zijn in 3 op de 10 even grote steekproeven even waarschijnlijke of minder waarschijnlijke afwijkingen te verwachten.

Op eenvoudige wijze kan aan dit voorbeeld worden gedemonstreerd, hoe deze methode bij vermindering van het aantal klassen overgaat in de eerder besproken methode van de solitaire kansdistributie van de steekproef uit 2 soorten individuen (b.v. prijs en niet). Beschouwen wij hier alleen de worpen van louter mannetjes (3 *m*) tegenover alle andere, dan is het aantal klassen slechts 2 en wordt de valentie van de simultane excentriciteit dus 2 — 1 of 1, d.w.z. de methode voert dan automatisch tot een solitaire excentriciteit, zoals dat ook het geval is indien wij, zoals dat op blz. 8 e.v. werd gedaan, van een solitaire kansdistributie uitgingen.

Deze methode, die door PEARSON ontworpen werd, is toepasselijk op al die vraagstukken, waarbij een onderstelling moet worden getoetst omtrent een bepaalde verdeling van de kansen, dat een individu uit een reeks verschillende kenmerken er telkens één zou vertonen.

In vele statistisch te onderzoeken gevallen, waarin sprake is van een simultane excentriciteit, daalt de valentie door het aftrekken van het aantal te schatten constanten tot 1, zodat wij uiteindelijk niet met een simultane maar met een solitaire excentriciteit — *T* — te maken krijgen. Wij bespreken twee van zulke situaties.

HET FREQUENTIE-VIERKANT

Hierbij is het de vraag of de kans, dat een individu een bepaalde eigenschap vertoont, verschilt naarmate dat individu wel of niet een bepaalde andere eigenschap bezit. In de taal der statistica heet dit de vraag, of er correlatie tussen deze twee eigenschappen bestaat.

Dr. VAN KREVELEN stelde de vraag, of bij kinderen met een hypomaan karakter de fantasie meer dan gewoonlijk tot uiting komt. Hij beschikte daartoe over vragenlijsten, die door onderwijzers waren ingevuld en waaruit o. a. werd geconcludeerd of een kind volgens de onderwijzer tekenen van actieve fantasie vertoonde en ook of het een hypomaan karakter heeft (over de betrouwbaarheid van het vaststellen dezer eigenschappen spreken wij hier niet).

Beperken wij ons tot de daarbij beschreven „enige" kinderen (d.w.z. kinderen zonder broertje of zusje) dan leveren die lijsten de aantallen enige kinderen, die binnen het vierkant in tabel VII vermeld staan. Pleiten die getallen voor de opvatting, dat het hypomane karakter de kans verhoogt op de aantekening „behept met actieve fantasie" ?

TABEL VII

		hypomaan karakter: wel	hypomaan karakter: niet	totalen
actieve fantasie	wel	32	15	47
	niet	37	86	123
totalen		69	101	170

Wij toetsen weer een nul-hypothese, n.l. de onderstelling dat het hypomane karakter geen invloed op de fantasie zou hebben. In dat geval zou een hypomaan kind een even grote kans op de qualificatie „aktieve fantasie" hebben als een niet-hypomaan kind. Bovendien zou de kans op een hypomaan karakter bij een kind met aktieve fantasie even groot zijn als bij een kind zonder aktieve fantasie. De te toetsen onderstelling vermeldt niet, hoe groot die twee kansen (te weten 1e de kans op de qualificatie: aktieve fantasie en 2e de kans op de qualificatie hypomaan karakter) zouden zijn, maar wel kan uit de gegeven aantallen voor elk van die twee kansen de beste schatting worden berekend. Ziehier de twee constanten, die uit de gegevens moeten worden geschat. Zijn eenmaal die twee kansen geschat, dan volgen daaruit ook de beste schattingen voor elk van de vier hier onderscheiden klassen (wel-wel, wel-niet, niet-wel en niet-niet) en kan op die klassen dus de formule van PEARSON worden toegepast.

Van de 170 kinderen onzer steekproef hebben er 47 de qualificatie „aktieve fantasie" gekregen (zie de totalen langs het vierkant) zodat de beste schatting, die uit onze steekproef volgt op „fantasie", $47 : 170 = 0.277$ bedraagt. En zo bedraagt de beste schatting voor de kans op een hypomaan karakter $69 : 170 = 0.406$.

Uit deze twee beste schattingen volgt ook voor elk onzer vier klassen de beste schatting voor de kans en dus ook het theoretisch in die klasse verwachte aantal. Voor de klasse „wel hypomaan en wel fantasie" geeft het theorema van de gecombineerde kans de kans als $0.277 \times 0.406 = 0.1125$ zodat het theoretisch in die klasse van onze steekproef te verwachten aantal kinderen $0.1125 \times 170 = 19.125$ bedraagt. Daar in die klasse 32 kinderen aangetroffen waren, is de afwijking in die klasse $d = 32 - 19.125 = 12.875$ en voor de be-

rekening van g^2 levert deze klasse dus een bijdrage van $d^2 : f_v =$ $= 12.875^2 : 19.125 = 8.667$. Berekenen wij op dezelfde manier de bijdragen van de overige 3 klassen, dan blijkt hun som 20.37 te bedragen en dat is dus de grootte van g^2 (kleine letter g omdat het hier, als consequentie van het schatten der twee kansen, niet G^2 zelf maar slechts de beste schatting voor G^2 betreft).

Doordat het aantal klassen hier 4 is en er voor 2 constanten de beste schattingen moesten worden bepaald, bedraagt de valentie 1:

$$N = k - c - 1 = 4 - 2 - 1 = 1$$

Wij zouden de overschrijdingskans dus uit het G^2-N-P-nomogram kunnen aflezen bij $G^2 = 20.37$ en $N = 1$, maar het kan ook anders: een „simultane excentriciteit" met valentie 1 is een solitaire excentriciteit, T, zodat in ons geval $t^2 = 20.37$ of $t = 4.5$ (met als valentie $N = n - c - 1 = 170 - 2 - 1 = 167$). Uit het t-N-P-nomogram lezen wij af, dat de overschrijdingskans, P_2, daarbij zo veel kleiner dan 0.001 is, dat ook de eenzijdige, die wij hier wensen te kennen, veel kleiner dan 0.001 blijkt te zijn. De nul-hypothese wordt dus zeer betrouwbaar afgewezen, waarmee met even grote betrouwbaarheid de eindconclusie wordt getrokken, dat het hypomane karakter de uitingen der fantasie, zoals hier bedoeld, bevordert: als een even grote steekproef uit hetzelfde kindermateriaal vele malen zou worden genomen, dan is te verwachten dat de toevallige steekproefselectie slechts in 1 op de veel meer dan 1000 zulke steekproeven even grote of nog grotere tegenstellingen zou veroorzaken.

	wel	niet	totalen
wel	a	b	p
niet	c	d	q
totalen	r	s	n

Bij elk onderzoek, waarbij uit een steekproef „met een frequentievierkant" de vraag dient beantwoord, of er correlatie tussen twee eigenschappen bestaat, zou de berekening op geheel analoge wijze geschieden, alleen met andere getallen. De berekening kan daardoor worden bekort door de vier gegeven aantallen volgens bovenstaand schema met de letters a, b, c en d aan te geven, de regel- en kolom-

sommen door p, q, r en s, de totale som, dat is het totale aantal individuen van de steekproef door n, en dan algebraïsch de grootte van de excentriciteitsschatting t uit te rekenen. Het blijkt dan dat

$$\boxed{t = (ad - bc) \sqrt{n : pqrs}}$$

met $N = n - 3$ [1])

Het gebruik van deze formule wordt slechts beperkt door de voorwaarde, dat geen van de vier gegeven aantallen kleiner mag zijn dan 6 (anders levert n.l. het vervangen van de binomiale — hier polynomiale — kansdistributies door normale immers geen voldoende benadering op).

De grootte van deze excentriciteit (t) geeft uitsluitend de betrouwbaarheid aan, waarmee de gegeven aantallen pleiten voor de conclusie, dat er al dan niet correlatie tusschen de twee onderzochte eigenschappen bestaat; de grootte van deze correlatie zelf (de sterkte van het verband, i.c. tussen het hypomane karakter en de fantasie) moet op een andere manier worden afgeleid. Daarover wordt op blz. 74 e.v. gesproken.

ZIJN TWEE AFMETINGEN, DIE ELK ENIGE MALEN WERDEN GEMETEN, IN WERKELIJKHEID VERSCHILLEND VAN GROOTTE?

De behandeling van bovenstaande vraag vormt ons laatste voorbeeld van een gereduceerde simultane kansdistributie.

De op blz. 27 bedoelde onderzoeker had de phosphaatuitscheiding bij zijn blanke proefpersoon niet alleen in een normale periode herhaaldelijk gemeten, maar ook in een tweede periode, waarin deze persoon regelmatig een bepaald vitaminepreparaat gebruikte. Deze metingen (in werkelijkheid deel van nog meer reeksen, waarover later) waren bedoeld om na te gaan, of dit vitaminegebruik invloed op de phosphaatstofwisseling heeft.

In de periode zonder het extra vitaminegebruik was het resultaat van zijn 20 metingen, zoals op blz. 27—32 werd uiteengezet,

---

[1]) Als een (of meer) van de 4 gegeven aantallen (a, b, c en d) klein is, dient eerst, in verband met de opmerking in de voetnoten van blz. 17 en 19, bij twee diagonaal tegenover elkaar staande aantallen (dus bij a en b of bij c en d) $\frac{1}{2}$ te worden opgeteld en van de andere twee $\frac{1}{2}$ te worden afgetrokken, waarbij de twee te verhogen en de twee te verlagen aantallen zodanig dienen te worden gekozen, dat t kleiner uitvalt dan bij de onveranderde aantallen het geval zou zijn, d.w.z. zodanig, dat ad en bc minder van elkaar gaan verschillen.

94 $\pm$ 2.6 cg (waarbij $s_{\bar{x}} = 2.6$ was berekend uit $s_{\bar{x}}^2 = 6.65$). Voor de periode met vitaminenpreparaat [1]) luidde het resultaat van de verkregen 22 meetgetallen (die wij y noemen) 95,8 $\pm$ 1.9 cg (waarbij $s_{\bar{y}} = 1.9$ volgde uit $s_{\bar{y}}^2 = 3.48$). Gevraagd werd, of op grond van deze metingen geconcludeerd mag worden, dat de phosphaatafscheiding in deze twee perioden verschilde.

Om deze vraag te beantwoorden stellen wij als nul-hypothese, dat deze twee phosphaatuitscheidingen niet verschilden — m.a.w. dat hun verschil nul zou bedragen — en wij toetsen deze hypothese aan de verkregen meetgetallen.

Een wiskundige afleiding leert dat de beste schatting, die uit de meetgetallen volgt voor het verschil tussen de twee uitscheidingen, het verschil is tussen de beste schattingen voor de twee uitscheidingen zelf, hier dus het verschil tussen de twee gemiddelden of $\bar{x} - \bar{y} =$ 94 — 95.8 = —1.8 cg. Daar de theoretisch (op grond van de te toetsen nul-hypothese) verwachte waarde 0 bedraagt, is dus ook de afwijking tussen de gevonden en verwachte waarde —1.8, dat wil zeggen als de nul-hypothese juist is, is het verschil van —1.8 cg uitsluitend door de meetfouten veroorzaakt.

Als deze metingreeksen nog eens zouden worden uitgevoerd, dan zou voor het verschil tussen de twee gemiddelden allicht een ander getal worden gevonden, doordat de meetfouten dan allicht anders zouden uitvallen. Ook het „gevonden verschil" is dus weer een individuele waarde van een variabele grootheid en wij vragen ons daarom af hoe de kansdistributie is van het verschil tussen twee gemiddelden, in de hoop daaruit de overschrijdingskans van de gevonden afwijking van —1.8 cg te kunnen afleiden.

Ook bij deze vraagstelling hebben wij in beginsel te maken met een simultane kansdistributie, n.l. een, die uit 2 normale kansdistributies bestaat: die van de metingen in de eerste proefperiode en die in de tweede. Wiskundige analyse toont aan, dat ook hier één van de twee valenties wegvalt (in wezen door het schatten van één constante: het gezamenlijk gemiddelde) zodat de valentie 1 wordt en wij hier dus uiteindelijk met een solitaire kansdistributie te maken krijgen. Dit is het eenvoudigst onder woorden te brengen in de vorm van de stelling, dat, als twee getallen elk een normale kansdistributie hebben, ook hun verschil een normale kansdistributie heeft, waarbij de variantie

---

[1]) In deze tweede periode vond hij op 22 dagen een phosphaatuitscheiding van 103, 91, 99, 97, 89, 111, 102, 97, 97, 70, 94, 94, 93, 103, 102, 102, 83, 83, 100, 98, 98 en 102 cg.

van het verschil de som is van de varianties der getallen zelf. Noemen wij die getallen a en b dan is dus volgens deze stelling

$$\boxed{S^2_{a-b} = S^2_a + S^2_b}$$

Door deze stelling op onze twee gemiddelden — $\bar{x}$ en $\bar{y}$ — toe te passen (hetgeen veroorloofd is omdat deze gemiddelden elk een normale kansdistributie hebben) vinden wij

$$s^2_{\bar{x}-\bar{y}} = s^2_x + s^2_{\bar{y}}$$

of in ons vraagstuk

$$s^2_{\bar{x}-\bar{y}} = 6.65 + 3.48 = 10.13$$

of

$$s_{\bar{x}-\bar{y}} = 3.2 \text{ cg}$$

Voor de gevonden afwijking van —1.8 cg vinden wij dus de (solitaire) excentriciteit

$$t = d : s = 1.8 : 3.2 = 0.56$$

(het minteken is weggelaten omdat de kansdistributie toch symmetrisch is).

Voor de valentie geldt dezelfde regel als voor de variantie: de valentie van de strooiïng van het verschil tussen twee gemiddelden is de som van de valenties van de strooiïngen dier gemiddelden zelf:

$$\boxed{N_{a-b} = N_a + N_b}$$

In ons geval bedraagt de valentie dus $N = (20-1) + (22-1) = 40$ of meer dan 30. Wij lezen van de bovenste lijn van het t-N-P-nomogram ($N = \infty$) bij $t = 0.56$ een dubbelzijdige overschrijdingskans van 0.6 af, dat is veel groter dan 0.05. De metingen vormen dus geen aanwijzing om aan onze nul-hypothese te twijfelen, behelzen dus ook geen aanwijzing voor een invloed van dit vitamine-gebruik op de phosphaatstofwisseling (in ongeveer 6 op de 10 keren zouden de meetfouten bij herhaling van evenveel metingen eenzelfde of nog groter verschil veroorzaken).

Voor die conclusie hadden wij de excentriciteit niet eens behoeven te berekenen. Het t-N-P-diagram geeft aan, dat bij een excentriciteit, die kleiner is dan 2, de dubbelzijdige overschrijdingskans steeds groter is dan 0.05: *bij de gebruikelijke betrouwbaarheidsdrempel wordt geen betekenis gehecht aan een verschil, dat kleiner is dan 2 maal de standaardfout.* In ons vraagstuk was het verschil (1.8 cg) kleiner dan de standaard-

fout zelf (3.2 cg) en dus veel kleiner dan het dubbele, zodat reeds deze eenvoudige vaststelling zonder meer leert, dat aan het gevonden verschil geen betekenis mag worden toegekend en de metingen dus geen invloed van het vitamine hebben aangetoond.

Nog sneller hadden wij dezelfde conclusie kunnen trekken op grond van de volgende overweging.

Als een afmeting in twee verschillende omstandigheden telkens ongeveer even veel malen gemeten is en die omstandigheden ten opzichte van dat meten als praktisch gelijkwaardig te beschouwen zijn, dan kan worden verwacht, dat de twee gemiddelden ten naastenbij gelijke standaardfouten zullen vertonen. Als die twee standaardfouten gelijk zijn is de standaardfout van het verschil bijna $1\frac{1}{2}$ maal zo groot: noemen we de standaardfout van elk gemiddelde a, dan volgt die van het verschil uit $s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}} = a^2 + a^2 = 2a^2$ of $s = a\sqrt{2} = 1.41a$. Pas als een vele malen gemeten maat meer dan ruim 2 maal zijn standaardfout bedraagt, mag er, zoals wij zagen, bij de gebruikelijke betrouwbaarheidsdrempel betekenis aan worden toegekend. Hieruit volgt de praktische regel: *als twee gemiddelden zijn berekend uit ongeveer even veel en samen meer dan 30 getallen, die verkregen zijn door praktisch even nauwkeurige metingen, mag aan hun verschil slechts betekenis worden toegekend indien het groter is dan 3 maal de standaardfout van die gemiddelden zelf* ($2 \times 1.41 = 2.82$ of bijna 3).

In ons vraagstuk zijn die voorwaarden vervuld. Wel komen de standaardfouten der twee gemiddelden niet opvallend mooi overeen (2.6 en 1.9) maar het gevonden verschil (1.8 cg) is kleiner dan elke en dus zo veel kleiner dan 3 maal „de fout der gemiddelden", dat er wel zeker geen betekenis aan mag worden toegekend.

DE VERHOUDING TUSSEN TWEE STROOIINGEN

Werd in de vorige paragraaf het onderzoek besproken omtrent de vraag, of twee gemiddelden significant verschillen, in vele vraagstukken komt het probleem aan de orde, of twee gevonden strooiïngen significant uiteenlopen, m.a.w. of de twee standaard-deviaties, waarvan deze strooiïngen de beste schattingen vormen, moeten worden opgevat als verschillend in grootte.

Een eenvoudig voorbeeld daarvan is de vraag, of twee meetmethoden verschillend nauwkeurig zijn. De nauwkeurigheid van een meetmethode kan n.l. worden gekenmerkt door eenzelfde afmeting een groot aantal malen volgens die methode te meten en dan de strooiïng van de verkregen meetgetallen te berekenen — uit $s^2_x = [x_,^2] : N$ met $N = n-1$ — want deze strooiïng of de „standaardfout van de

enkele meting" is immers de beste schatting, die uit de verkregen meetgetallen voortvloeit voor de standaard-deviatie van die normale kansdistributie, die geacht wordt aan de meetfouten eigen te zijn; en hoe kleiner deze is des te nauwkeuriger is de meetmethode. Levert de ene meetmethode meetgetallen met een kleinere strooiïng dan de andere, dan kan dat berusten op een grotere nauwkeurigheid van de eerstgenoemde. Doordat de strooiïng echter slechts de beste schatting is voor de bedoelde standaard-deviatie en de strooiïng bij een volgende meetreeks dan ook allicht een iets andere waarde zou hebben, kan het ook zijn, dat het gevonden uiteenlopen van de twee strooiïngen uitsluitend op de toevallig deze malen gemaakte meetfouten berust en de twee standaard-deviaties zelf gelijk, de meetmethoden dus even nauwkeurig zijn.

Wanneer mogen wij aannemen, dat twee gevonden strooiïngen significant verschillen, d.w.z. dat de twee standaard-deviaties, waarvan zij de beste schattingen zijn, ongelijke waarde zouden hebben?

Terwijl het gemiddelde van een aantal getallen, die zelf een normale kansdistributie hebben, weer een normale kansdistributie heeft (zie blz. 50) heeft de strooiing van die getallen een andere soort kansdistributie. Is het aantal getallen, waar die strooiïng uit berekend werd, groot, dan nadert deze kansdistributie echter onbeperkt tot de normale.

De kansdistributie van de strooiïng van n getallen, die zelf een normale kansdistributie hebben, is n.l. te herleiden tot die kansdistributie, die eigen is aan de simultane excentriciteit met dezelfde valentie als de strooiïng, dus van G_N (waarin $N = n - 1$). Niet de strooiïng zelf heeft deze kansdistributie, maar de grootheid $(s : S) \sqrt{N}$, dat is dus $\sqrt{N}$ maal de verhouding tussen de strooiïng en de standaard-deviatie, waar zij de beste schatting van is (n.l. de standaard-deviatie van de normale kansdistributie der getallen zelf). En bij toenemende valentie (N) nadert de kansdistributie van de simultane excentriciteit (G_N) tot de normale, doordat dan elke van de samenstellende T-waarden slechts een kleine en onafhankelijke invloed heeft op het tot stand komen van de simultane excentriciteit.

Dat $(s : S) \sqrt{N}$ dezelfde kansdistributie heeft als G_N moge door de volgende overweging plausibel worden. De variantie s^2 is gelijk aan $[x_,^2] : N$ en daarin hebben alle $x_,$-waarden dezelfde kansdistributie met standaard-deviatie S en een middenwaarde, die niet veel van $\bar{x}$ verschilt. Verwaarlozen wij even dit kleine verschil, dan beschouwen wij elke $x_,$-waarde gelijk aan haar deviatie en is haar excentriciteit $T = D : S = x_, : S$ en is $T^2 = x_,^2 : S^2$. De som van de n waarden van T^2 heeft volgens de definitie dezelfde kansdistributie als G^2_n (zie blz. 41) en deze som bedraagt $[x_,^2] : S^2$ of $Ns^2 : S^2$ of $(s^2 : S^2)\, N$ zodat de wortel eruit — $(s : S) \sqrt{N}$ — dezelfde kansdistributie zou hebben als G_n. Verdere analyse toont aan dat de valentie n door onze verwaarlozing 1 te groot is geworden, dus dat zij moet worden gerekend als $n - 1$ of N, zodat $(s : S) \sqrt{N}$ dezelfde kansdistributie heeft als G_N.

Dit vervangen van n door $n - 1$ of N heeft dezelfde grond als waarom

de beste schatting van een standaard-deviatie niet is $[x_,^2) : n$ maar $[x_,^2] : N$. De strooiïng is n.l. bedoeld als de beste schatting voor $[D^2] : n$.

Bij een strooiïng, berekend uit een groot aantal getallen, mogen wij de ware kansdistributie ter benadering vervangen denken door een normale en daarvan blijkt de standaard-deviatie $\sqrt{2}$ maal zo klein te zijn als die van het gemiddelde, zodat de „standaardfout van de strooiïng" volgt uit

$$\boxed{s^2_s = s^2_{\bar{x}} : 2}$$

of

$$\boxed{s^2_s = s^2_x : 2n}$$

of

$$\boxed{s^2_s = [x_,^2] : 2nN} \qquad \text{waarin } N = n-1$$

Met behulp van een van deze formules kunnen wij, op dezelfde manier als dat in de vorige paragraaf voor twee gemiddelden is uiteengezet, de vraag beantwoorden of twee strooiïngen significant verschillen, tenminste als die strooiïngen elk uit meer dan 30 getallen berekend waren.

Als b.v. een afmeting 40 maal gemeten werd volgens de ene meetmethode en 50 maal volgens de andere, en de strooiïng van de 40 getallen der eerste methode bedroeg 0.100 mm, die van de 50 getallen der tweede 0.110 mm, dan toetsen wij de nulhypothese, dus de opvatting dat de twee methoden niet in nauwkeurigheid zouden verschillen. In dat geval zouden de twee strooiïngen volgens de theoretische verwachting niet verschillen, zij verschillen echter 0.010 mm, zodat de afwijking tussen de gevonden en de theoretisch verwachte waarde 0.010 mm bedraagt. Om daarvan de overschrijdingskans te bepalen stellen wij een normale kansdistributie in de plaats van de slechts bij benadering normaal zijnde kansdistributie van elk dier strooiïngen. De standaardfout van de eerste strooiïng volgt uit $s^2_s = s^2_x : 2n = 0.100^2 : 80 = 0.000125$ en de tweede uit $s^2_s = 0.110^2 : 100 = 0.000121$. Volgens de stelling $s^2_{a-b} = s^2_a + s^2_b$ geldt voor de standaardfout van het verschil der twee strooiïngen $s^2_v = 0.000125 + 0.000121 = 0.000246$ of $s_v = 0.016$ mm. Gevonden was een verschil van 0.010 mm, zodat het gevonden verschil kleiner is dan zijn standaardfout en dus niet significant. (Het was trouwens al kleiner dan de kleinste standaardfout onzer strooiïngen, welke immers $\sqrt{0.000121}$ of 0.011 mm bedroeg.)

Alleen als een strooiïng uit een groot aantal getallen berekend is, lijkt haar kansdistributie voldoende op een normale om de zo juist geschetste benaderingsmethode te mogen toepassen. Daar bij vele statistische vraagstukken ook strooiïngen van weinig getallen met elkaar dienen te worden vergeleken, is een methode uitgewerkt, die

in alle omstandigheden kan worden toegepast, dus ook bij een strooiïng van enkele getallen (= een strooiïng van geringe valentie).

Deze methode beoordeelt niet het *verschil* tussen de twee strooiïngen in kwestie maar hun *verhouding*, dus niet $s_1 - s_2$ maar $s_1 : s_2$. Is de nul-hypothese gesteld, dan bedraagt de theoretisch verwachte waarde voor deze strooiïngsverhouding 1 (want met die nul-hypothese zou het overeenkomen als de twee strooiïngen aan elkaar gelijk zouden zijn — $s_1 = s_2$ — en in dat geval is hun verhouding $s_1 : s_2 = 1$). Is uit de twee gegeven reeksen van getallen een andere verhouding tussen de twee strooiïngen gevonden, b.v. 1.5, dan moet de overschrijdingskans van deze uitkomst worden bepaald om een oordeel omtrent de juistheid of onjuistheid van die nul-hypothese te kunnen vaststellen. De theorie van de simultane kansdistributies leert, dat die overschrijdingskans van 3 grootheden afhangt, n.l. van de verhouding der twee strooiïngen en van hun valenties, dus van $s_1 : s_2$, van N_1 en van N_2.

In de kleine letters van blz. 52 en 53 werd besproken, dat bij één strooiïng de grootheid $(s : S) \sqrt{N}$ dezelfde kansdistributie heeft als G_N. Bij twee strooiïngen s_1 en s_2, die volgens een nul-hypothese de beste schattingen zijn van eenzelfde standaard-deviatie S, heeft dus het quotient

$$\{(s_1 : S) \sqrt{N_1}\} : \{(s_2 : S) \sqrt{N_2}\}$$

dezelfde kansdistributie als het quotient $G_{N_1} : G_{N_2}$ zodat de strooiïngsverhouding $s_1 : s_2$ dezelfde kansdistributie heeft als $(G_{N_1} : G_{N_2}) \sqrt{N_2 : N_1}$. Deze kansdistributie hangt dus uitsluitend van G_{N_1} en G_{N_2} af, m.a.w. van N_1 en N_2.

Een volledig nomogram, dat voor alle waarden van de strooiïngsverhouding $s_1 : s_2$ en bij alle valenties N_1 en N_2 dier strooiïngen de grootte van de overschrijdingskans zou aangeven, zou 3-dimensionaal moeten zijn (omdat deze kans van 3 grootheden afhangt, n.l. van $s_1 : s_2$, van N_1 en van N_2). Om aan die praktische moeilijkheid tegemoet te komen, behelpen wij ons met enkele 2-dimensionale dus vlakke ,,doorsneden'' door dit 3-dimensionale nomogram. Zo geeft het ,,eerste strooiïngs-nomogram'' op blz. 152 aan, hoe groot bij diverse valenties N_1 en N_2 die strooiïngsverhouding — $s_1 : s_2$ — is, waarbij de overschrijdingskans 0.2 bedraagt en de volgende nomogrammen geven hetzelfde voor de overschrijdingskansen 0.05, 0.01 en 0.001. Met behulp van deze nomogrammen kan dus steeds een oordeel worden gevormd of de betrouwbaarheidsdrempel, overeenkomende met een overschrijdingskans van 0.05 of 0.01 of 0.001, overschreden is.

Bij deze strooiïngs-nomogrammen dient nog te worden opgemerkt, dat N_1 de valentie van de grootste strooiïng voorstelt en de (eenzijdige!) overschrijdingskans P_1 de kans is op een even grote of grotere strooiïngsverhouding $s_1 : s_2$. Deze overschrijdingskans is dus het richtsnoer

voor het beantwoorden van de vraag, of de strooiïng s_1 significant *groter* is dan de strooiïng s_2.

Deze nomogrammen zullen in de volgende vraagstukken worden toegepast.

Twee opmerkingen mogen er hier eerst nog over gemaakt worden.

Als een van de twee strooiïngen uit weinig en de andere uit zeer veel getallen is berekend, dan kan de overschrijdingskans beter uit de G^2-N-P-tabel worden afgelezen. De strooiïng uit de vele getallen heeft dan n.l. een zo kleine onnauwkeurigheid, dat deze in het niet valt tegenover die van de andere strooiïng, zodat het probleem dan neerkomt op de vraag, of de andere strooiïng significant verschilt van een bepaalde getalswaarde (n.l. de waarde van de nauwkeurige strooiïng). Daartoe wordt de simultane excentriciteit berekend uit

(als $N_2 > 30$) $\boxed{G^2_N = N_1 (s^2_1 : s^2_2)}$ waarin $N = N_1$

waarbij s_1 de strooiïng is uit de weinige getallen en s_2 die uit de vele getallen. Het G^2-N-P-diagram levert dan de gezochte overschrijdingskans.

Is een van de twee strooiïngen uit slechts 2 getallen berekend en de andere uit zeer veel, dan heeft de simultane excentriciteit, die op dezelfde wijze berekend wordt, een valentie $N_1 = 1$ zodat het een solitaire excentriciteit is:

$$T^2 = G^2_1 = 1 \times (s^2_1 : s^2_2)$$

zodat, als

$N_1 = 1$ en $N_2 > 30$: $\boxed{T = s_1 : s_2}$

waarbij de dubbelzijdige overschrijdingskans — P_2 — dient te worden afgelezen omdat een strooiïng niet negatief kan uitvallen.

Is een van de twee strooiïngen (n.l. s_1) uit slechts 2 en de andere (s_2) uit een matig aantal getallen berekend, dan geldt

(als $N_1 = 1$) $\boxed{t = s_1 : s_2}$ met $N = N_2$

FISHER, die de methode van de strooiïngsverhouding heeft aangegeven, gebruikt niet deze verhouding zelf, maar de natuurlijke logarithme er van, omdat de kansdistributie daarvan bij grote valenties onbeperkt tot een normale kansdistributie nadert. Hij gebruikt voor deze logarithme het symbool z, waarvoor dus geldt $z = \text{nat log}\,(s_1 : s_2)$. Een principieel verschil met de hier gevolgde lijn vormt dit niet.

ZIJN ENIGE AFMETINGEN, DIE ELK ENIGE MALEN GEMETEN ZIJN, IN WERKELIJKHEID UITEENLOPEND VAN GROOTTE ? DE VARIANTIE-ANALYSE

Bij 5 proefpersonen werd de dagelijkse phosphaatuitscheiding gedurende 9 achtereenvolgende dagen gemeten. Pleiten de daarbij verkregen getallen, die hieronder binnen de rechthoek van Tabel VIII vermeld staan, voor de opvatting, dat deze uitscheiding bij verschillende personen uiteenloopt?

TABEL VIII

De phosphaatuitscheiding in cg bij de proefpersonen A tot E, voor elk hunner op 9 achtereenvolgende dagen gemeten

	A	B	C	D	E		
	84	139	122	98	115		
	92	149	140	88	127		
	87	147	141	115	111		
	107	146	131	92	102		
	79	158	150	90	96		
	92	150	160	92	111		
	105	174	143	74	136		
	77	183	152	94	121		
	111	141	121	89	133		
$[x]$	834	1387	1260	832	1052		
n	9	9	9	9	9		
$[x] : n = \bar{x}$	92.7	154.1	140.0	92.4	116.9		
$[x_{,}^2]$	1208	1816	1416	960	1464	samen	6864
$n - 1 = N$	8	8	8	8	8	samen	40
$[x_{,}^2] : N = s^2_x$	151	227	177	120	183	gemiddeld	171.6 [1])
s_x	12.3	15.1	13.3	11.0	13.5	gemiddeld	13.1 [2])

In elke kolom van deze tabel staan de 9 meetgetallen, die bij één proefpersoon voor de phosphaatuitscheiding op 9 achtereenvolgende dagen waren verkregen. Daaronder werd hun som — $[x]$ — gedeeld door hun aantal — $n = 9$ — waarmee voor elke proefpersoon het gemiddelde — $\bar{x}$ — was gevonden. Deze 5 gemiddelden lopen inderdaad uiteen: 92.7 voor persoon A en voor de personen B, C, D en E achtereenvolgens 154.1, 140.0, 92.4 en 116.9 cg.

Als de phosphaatuitscheiding bij één persoon elke dag dezelfde grootte heeft [3]) dan kan het feit, dat deze 5 gemiddelden uiteenlopen, in het algemeen gesproken op tweeërlei oorzaken berusten:

1. als de 5 personen inderdaad verschillend sterk phosphaat uitscheidden, zou dit één oorzaak voor het uiteenlopen der gemiddelden zijn geweest,
2. doordat meetfouten telkens anders uitvallen, vallen de meetgetallen en hun gemiddelde bij de verschillende personen allicht

---

[1]) 6864 : 40.

[2]) $\sqrt{171.6}$.

[3]) De mogelijkheid, dat de phosphaatuitscheiding van dag op dag varieert, wordt op blz. 59 besproken.

verschillend uit, zelfs al zouden de 5 personen in werkelijkheid evenveel phosphaat uitscheiden.

Hiermee is de weg voor ons onderzoek aangewezen. Als wij konden vaststellen dat de 5 gemiddelden sterker uiteenlopen dan alleen door de meetfouten zou kunnen worden veroorzaakt, dan zouden wij weten dat de werkelijke phosphaatuitscheidingen uiteenlopen. Het eerste kunnen wij in beginsel nooit met zekerheid vaststellen, doordat de meetfouten wel eens bijzonder ongelijk zouden kunnen uitvallen. Maar als wij konden vaststellen, dat de 5 gemiddelden zó sterk uiteenlopen als slechts uiterst zelden door de meetfouten zou worden veroorzaakt, dan zouden wij de conclusie mogen trekken dat de 5 personen hoogst waarschijnlijk verschillend sterk phosphaat uitscheiden.

Hoe sterk varieert een gemiddelde tengevolge van de meetfouten? Wij zagen op blz. 31 dat, als een groep van n getallen op een normale kansdistributie berust, hun gemiddelde een normale kansdistributie met een $\sqrt{n}$ maal kleinere standaard-deviatie heeft. In ons vraagstuk is elk gemiddelde uit 9 getallen berekend, zodat, als er alleen meetfouten in het spel zijn geweest, de standaard-deviatie van de gemiddelden $\sqrt{9}$ of 3 maal zo klein zou zijn als die van de meetfouten. Mocht uit een nadere analyse van de getallen blijken, dat de standaard-deviatie van de gemiddelden waarschijnlijk groter is dan $^1/_3$ van de standaard-deviatie der meetfouten, dan zouden de 5 personen waarschijnlijk uiteenlopende phosphaatuitscheidingen hebben gehad.

In symbolen weergegeven is, als er geen verschil tussen de personen is,

$$S^2_{\bar{x}} = S^2_x : n \quad \text{of} \quad nS^2_{\bar{x}} = S^2_x$$

terwijl, als er wel verschil tussen de personen is, geldt dat

$$nS^2_{\bar{x}} \text{ groter is dan } S^2_x.$$

Hierin heet de grootheid $nS^2_{\bar{x}}$ de variantie tussen de personen en de grootheid S^2_x de variantie binnen de personen.

De meetfouten komen tot uiting in het uiteenlopen van de meetgetallen binnen één proefpersoon: de strooiïng van de 9 meetgetallen van persoon A is de beste schatting, die persoon A levert voor de standaard-deviatie der meetfouten. En zo levert elke persoon een schatting voor die ene standaard-deviatie. Deze strooiïngen werden onder elke kolom berekend (alleen werd de becijfering van de kwadraatjessom $[x_,^2]$ van elke persoon hier niet weergegeven; zij geschiedde volgens een op blz. 127 behandelde formule). Het gemiddelde van die 5 strooiïngen is dus de beste schatting, waarover wij uiteindelijk kunnen beschikken. Wij vinden daarvoor $\bar{s}_x = 13.1$.

De beste schatting voor de standaard-deviatie van het gemiddelde is de strooiïng van de 5 berekende gemiddelden (de strooiïng van de

5 $\bar{x}$-waarden). Deze strooiïng — $s_{\bar{x}}$ — werd in tabel IX berekend: boven de streep zijn de zoëven gevonden 5 gemiddelden $\bar{x}$ uit tabel VIII overgeschreven en onder de streep werd daar dezelfde berekening op toegepast als in de vorige tabel op de meetgetallen van telkens één proefpersoon. Voor de strooiïng der gemiddelden werd gevonden $s_{\bar{x}} = 27.8$ cg en dat is veel groter dan $^1/_3$ van de „strooiïng binnen de personen": $3 \times$ de strooiïng der gemiddelden is $3 \times 27.8 = 83.4$ en de „strooiïng binnen de personen" is slechts 13.1. De strooiïng der gemiddelden is dus 83.4 : 13.1 = 6.4 maal „te groot".

TABEL IX

Berekening van de strooiing der gemiddelden

gemiddelde	van	A	= 92.7
„	„	B	= 154.1
„	„	C	= 140.0
„	„	D	= 92.4
„	„	E	= 116.9
		$[\bar{x}]$	= 596.1
		$n_{\bar{x}}$	= 5
		$\bar{\bar{x}}$	= 119.2
		$[\bar{x}_{,}^2]$	= 3088.4
		$n_{\bar{x}} - 1 = N_{\bar{x}}$	= 4
		$[\bar{x}_{,}^2 : N_{\bar{x}} = s^2_{\bar{x}}$	= 772
		$s_{\bar{x}}$	= 27.8

Nog weten wij hieruit niet, dat 3 maal de standaard-deviatie der gemiddelden werkelijk groter is dan de standaard-deviatie der meetfouten: wij weten alleen, dat de beste schatting voor de eerste (de z.g. strooiïng tussen de personen) 6.4 maal de beste schatting voor de tweede (de z.g. strooiïng binnen de personen) is. Maar in blz. 54 e.v. leerden wij de manier om nu een oordeel over deze vraag uit te spreken. Wij kennen n.l. zowel de verhouding van de twee strooiïngen — $s_1 : s_2 = 6.4$ — als hun valenties: de valentie van de strooiïng der 5 gemiddelden is 5 — 1 of 4, de valentie van de strooiing der 9 meetgetallen van één proefpersoon is 9 — 1 of 8, de valentie van het gemiddelde dezer 5 strooiïngen is dus 5×8 of 40. En uit het strooiïngsnomogram op blz. 155 lezen wij af, dat bij $N_1 = 4$ en $N_2 = 40$ de strooiïngsverhouding van ong. 2.4 een overschrijdingskans van 0.001 heeft. De gevonden strooiïngsverhouding van 6.4 heeft dus een nog veel kleinere overschrijdingskans, zodat wij met zeer grote betrouwbaarheid kunnen concluderen,

dat 3 maal de standaard-deviatie van het gemiddelde groter is dan de standaard-deviatie der meetgetallen binnen de personen.

De 5 gemiddelden lopen dus hoogstwaarschijnlijk niet alleen tengevolge van de meetfouten uiteen, maar ook door iets anders, d.w.z. hoogstwaarschijnlijk liep ook de werkelijke phosphaatuitscheiding bij onze 5 proefpersonen uiteen (in slechts 1 op veel meer dan 1000 zulke onderzoekingen zouden de meetfouten zulke grote afwijkingen veroorzaken).

Bij de bovenstaande behandeling van dit vraagstuk was uitgegaan van de onderstelling, dat de phosphaatuitscheiding van één persoon constant zou zijn. De gevonden meetgetallen laten echter ook de duiding toe, dat die uitscheiding van dag tot dag wisselde. In dat geval zou bij die éne persoon de phosphaatuitscheiding een bepaalde kansdistributie hebben waarvan de middenwaarde het bedrag voorstelt, waarom de dagelijkse phosphaatuitscheiding van die persoon schommelt, en waarvan de standaard-deviatie de maat is voor de grootte dier schommelingen. Deze kansdistributie van zijn werkelijke phosphaatuitscheiding is nog niet de kansdistributie van zijn meetgetallen: de laatste heeft wel dezelfde middenwaarde maar een grotere standaard-deviatie, doordat de variantie van de meetgetallen in dit geval de som is van de variantie der werkelijke phosphaatuitscheiding en de variantie der meetfouten.

Bij het beantwoorden van ons vraagstuk kunnen wij in het midden laten welke duiding de juiste is: wij vatten dan de factoren, die de phosphaatuitscheiding bij één persoon van dag tot dag doen variëren èn de factoren, die de meetfouten veroorzaken, samen als „de hier niet nader te analyseren factoren". Wat wij wèl wensen te onderzoeken is, of er factoren waren, die de middenwaarden der diverse personen verschillend deden uitvallen: als er zulke factoren zijn dan „scheiden de diverse personen verschillend sterk phosphaat uit", in welke formulering in het midden wordt gelaten of de phosphaatuitscheiding van telkens één dier personen constant is dan wel van dag tot dag verschilt.

De oplossing is dan geheel dezelfde als hierboven werd beschreven voor het geval de phosphaatuitscheiding van telkens één persoon constant is. Alleen denken we ons het woord „meetfouten" telkens vervangen door „de invloed van de hier niet nader te analyseren factoren"[1]). De variantie binnen de personen wordt veroorzaakt

---

[1]) Dezelfde opmerking geldt voor het vraagstuk van blz. 48, waarin gevraagd werd of twee personen in hun phosphaat-uitscheiding verschillen.

uitsluitend door die niet nader te analyseren factoren, de variantie tussen de klassen zowel door deze als ook, indien zij er zijn, door de factoren, die eventueel verschillen tussen de personen hebben veroorzaakt. Zijn de laatstgenoemde verschillen er niet, dan zijn beide varianties aan elkaar gelijk (d.w.z. dan is $\sqrt{n}$ maal de standaard-deviatie der gemiddelden gelijk aan de standaard-deviatie van de meetwaarden telkens binnen één persoon), verschillen de personen wèl, dan is de variantie tussen de personen groter dan die er binnen (d.w.z. dan is $\sqrt{n}$ maal de standaard-deviatie der gemiddelden groter dan die binnen de personen).

Een tweede opmerking betreft de beste schattingen voor deze twee varianties.

Om de strooiïng binnen de personen te berekenen werd in tabel VIII eerst voor elke persoon afzonderlijk zijn kwadraatjessom becijferd, dan werden die 5 sommen bij elkaar geteld, waarna deze eindsom (6864) werd gedeeld door de som van de valenties binnen elke persoon (40). Tabel X geeft dat nog eens in een andere vorm weer. Als wij nu, om misverstanden te voorkomen, de meetgetallen van één persoon een klasse noemen en het gemiddelde van één persoon een klassegemiddelde (symbool: $\hat{x}$), dan zal het duidelijk zijn, dat de „kwadraatjessom binnen de klassen" — dat is bedoelde eindsom 6864 — de som is van de kwadraten van alle afwijkingen van een meetgetal tot zijn klassegemiddelde.

Om de strooiïng tussen de klassen te berekenen, werd de kwadraatjessom van de 5 gemiddelden (die 3088.4 bedraagt) gedeeld door de valen-

TABEL X

Overzicht van de variantie-analyse der 5 × 9 phosphaat uitscheidingen

		kwadr. som		valentie		variantie-schatting		strooiïng
strooiïng binnen A:	$s^2_A = [x_{,}^2]_A$: $(f-1)$ =	1208:		8	=	151, dus	s_A =	12.3
„ „ B:	$s^2_B = [x_{,}^2]_B$: $(f-1)$ =	1816:		8	=	227, dus	s_B =	15.1
„ „ C:	$s^2_C = [x_{,}^2]_C$: $(f-1)$ =	1416:		8	=	177, dus	s_C =	13.3
„ „ D:	$s^2_D = [x_{,}^2]_D$: $(f-1)$ =	960:		8	=	120, dus	s_D =	11.0
„ „ E:	$s^2_E = [x_{,}^2]_E$: $(f-1)$ =	1464:		8	=	183, dus	s_E =	13.5
		—+		—+				
binnen de 5 pers.:	$s^2_2 = [x_{.}^2]$: $k(f-1)$ =	**6864**:		**40**	=	**172**, dus	s_2 =	**13.1**
								($= s_{x\cdot}$)
tussen de 5 pers.:	$s_1^2 = [f\hat{x}_{,}^2]$: $(k-1)$ =	**27796**:		**4**	=	**6949**, dus	s_1 =	**83.4**
		—+		—+				
								($= fs_{\hat{x}}$)
totale strooiïng:	$s^2_x = [x_{,}^2]$: $(n-1)$ =	**34660**:		**44**	=	788, dus	s_x =	28.1

tie tussen de 5 klassen (dus door 4) en daarna vermenigvuldigd met het aantal meetgetallen per persoon (dat 9 bedraagt, welk aantal daar door n werd weergegeven en nu beter de klassefrequentie f heet). Wij hadden ook de kwadraatjessom eerst met de klassefrequentie kunnen vermenigvuldigen ($9 \times 3088.4 = 27796$) en dat getal, dat de kwadraatjessom tussen de klassen heet, kunnen delen door 4, de valentie tussen de klassen (zie tabel X). Daar deze kwadraatjessom van de 5 gemiddelden de som is van de kwadraten van de afwijkingen der 5 gemiddelden tot het totale gemiddelde, is de kwadraatjessom tussen de klassen nog anders te interpreteren: als elk meetgetal vervangen wordt door zijn klassegemiddelde en dan de afwijking van dat klassegemiddelde tot het totale gemiddelde wordt opgeschreven ($\hat{x}_{,}$), dan is de som van deze n afwijkingen gelijk aan de kwadraatjessom tussen de klassen.

Dit voert tot de volgende beschouwing, die de methode doorzichtiger maakt. Elk meetgetal (x) heeft een afwijking tot het totale gemiddelde ($\bar{x}$). Denken wij voor elk meetgetal deze totale afwijking ($x_{,} = x - \bar{x}$) gesplitst in twee porties, n.l.

1. de afwijking van het meetgetal tot zijn klassegemiddelde (voor welke afwijking wij het symbool $x_{.}$ invoeren) en
2. de afwijking van dat klassegemiddelde tot het totale gemiddelde (symbool: $\hat{x}_{,}$)

dan zijn de onder 1 bedoelde afwijkingen veroorzaakt uitsluitend door de niet nader te analyseren factoren en dan zijn de onder 2 bedoelde afwijkingen het gevolg zowel van de niet als van de wel te analyseren factoren. Dan is verder

1. de kwadraatjessom binnen de klassen de som van de kwadraten der eerstgenoemde afwijkingen, dus $[x_{.}^2]$ en
2. de kwadraatjessom tussen de klassen de som van de kwadraten der laatstgenoemde afwijkingen.

Nu leert een rekenkundige stelling, dat als alle totale afwijkingen op deze manier elk in twee gedeelten gesplitst worden, de som van de kwadraten van al de eerste gedeelten en de som van de kwadraten van alle tweede gedeelten samen gelijk zijn aan de som van de kwadraten der totale afwijkingen [1]), m.a.w.

---

[1]) $x_{,} = x - \bar{x} = x - \hat{x} + \hat{x} - \bar{x} = x_{.} + \hat{x}_{,}$

$[x_{,}^2] = [(x_{.} + x_{,})^2] = [x_{.}^2] + 2\,[x_{.}x_{,}] + [\hat{x}_{,}^2]$

en hierin is voor elke klasse afzonderlijk $[x_{.}\hat{x}_{,}] = [x_{.}]\,\hat{x}_{,} = 0$

omdat in die klasse $[x_{.}] = [x - \hat{x}] = [x] - f\hat{x} = [x] - [x] = 0$

zodat ook voor alle klassen samen $[x_{.}\hat{x}_{,}] = 0$ en dus

$[x_{,}^2] = [x_{.}^2] + [\hat{x}_{,}^2]$

de totale kwadraatjessom =
= de kwadraatjessom tussen de klassen +
+ de kwadraatjessom binnen de klassen.

Hetzelfde geldt voor de valenties der strooiïngen:

de totale valentie =
= de valentie tussen de klassen +
+ de valentie binnen de klassen.

Dit is in tabel X gedemonstreerd door als laatste regel de berekening van de totale strooiïng op te schrijven: de totale kwadraatjessom (34660) blijkt inderdaad de som te zijn van de kwadraatjessom binnen de klassen (6864) en die tussen de klassen (27796) en de totale valentie (44) is inderdaad de som van die binnen de klassen (40) en die tussen de klassen (4).

De bovenbeschreven methode komt daardoor neer op een splitsing van de totale kwadraatjessom in één tussen de klassen en één binnen de klassen. De eerste is veroorzaakt alleen door de niet nader te analyseren factoren, de tweede zowel door deze als ook door de te onderzoeken factor. En de totale valentie is in twee overeenkomstige porties gesplitst.

Dit opent een manier om in bepaalde omstandigheden de kwadraatjessom tussen de klassen gemakkelijker te berekenen dan door eerst in elke klasse afzonderlijk een kwadraatjessom te becijferen: als de totale kwadraatjessom van alle gegeven meetwaarden wordt becijferd en verder die van de (meest klein in aantal zijnde) klassegemiddelden, dan is hun verschil de kwadraatjessom binnen de klassen. Hetzelfde geldt voor de valenties. Deling geeft dan de kwadraten der twee gezochte strooiïngen, waarna worteltrekking de strooiïngen zelf en deling hun verhouding leert kennen. Dit komt er op neer, dat van tabel X alleen de dik gedrukte getallen behoeven te worden berekend.

Bovendien heeft deze manier het grote voordeel, dat zij ook kan worden toegepast als in de diverse klassen verschillende aantallen meetgetallen zijn gegeven (zie voorbeeld blz. 131).

Op grond van de beschreven splitsingen werd deze methode door haar ontwerper, FISHER, variantie-analyse genoemd.

DE MEERVOUDIGE VARIANTIE-ANALYSE

Bij het laatstbehandelde vraagstuk werd, om vast te stellen of een bepaalde factor aanwezig is, de invloed van een andere factor geëlimineerd: vastgesteld moest worden of er een factor is, die verschil veroorzaakte tussen de phosphaatstofwisseling van de ene en die van

de andere persoon; geëlimineerd werd, voor zover mogelijk, de invloed van de factor, die bij eenzelfde persoon de phosphaatstofwisseling van dag tot dag verschillend deed uitvallen (c.q. alleen het meetgetal daarvoor verschillend deed uitvallen). Beide factoren hadden invloed op de meetgetallen uitgeoefend en die meetgetallen waren op een zodanige manier gegeven, dat de invloeden van die twee factoren er op verschillende manier in tot uiting kwamen.

In vele gevallen zijn de meetgetallen op een zodanige manier gegeven, dat zij nog méér inlichting over storende factoren geven en dan kan het onderzoek vaak door een meervoudige variantie-analyse gevoeliger worden gemaakt.

Als voorbeeld wordt een ander onderzoek van Dr HOITINK besproken.

Bij 20 personen (A, B, C . . . . T) werd telkens 3 maal het gehalte aan vitamine C in het bloed gemeten:

I. even vóór het verrichten van zware arbeid,
II. 8 minuten er na,
III. 50 minuten er na.

Mag uit de meetgetallen, die in tabel XI vermeld staan, worden opgemaakt, dat zware arbeid invloed op het vitamine-C-gehalte van het bloed heeft?

De eventuele invloed van de arbeid zou tot uiting komen als wij van de groepen meetgetallen I, II en III telkens het gemiddelde bepalen en deze 3 gemiddelden (dat zijn de gemiddelde vitamine-gehalten bij deze proefpersonen voor, kort na en lang na de arbeid) met elkaar vergelijken. Deze 3 gemiddelden staan onder aan tabel XI vermeld.

Elk van deze gemiddelden volgt uit 20 meetgetallen en deze lopen uiteen doordat het gehalte telkens bij 20 verschillende personen gemeten was. Tot zo ver is dit, statistisch gezien, dezelfde situatie als bij het vorige vraagstuk, waar elk gemiddelde volgde uit (9) getallen, die uiteenliepen doordat de phosphaatuitscheiding van dag tot dag uiteenliep. Maar in dit nieuwe vraagstuk komt er een nieuwe inlichting bij: één persoon heeft telkens 3 vitamine-gehalten geleverd en wel uit elke reeks één: een vóór de arbeid, een er kort na en een er lang na. Als het gemiddelde vitamine-gehalte in het bloed van een speciale persoon b.v. ongewoon hoog was, dan heeft deze persoon in elke reeks een meetgetal extra hoog doen uitvallen. De getallen geven m.a.w. nog een extra inlichting omtrent een van de storende factoren: de invloed van de uiteenlopende personen kan er uit worden afgelezen.

Als de metingen in het vorige vraagstuk telkens op voor alle personen dezelfde dagen waren uitgevoerd, dan zou op overeenkomstige manier de invloed van de datum — dus b.v. de invloed van het weer —

TABEL XI

Vitamine-C-gehalte in het bloed van 20 proefpersonen, A tot T, gemeten kort voor, 8 min. na en 50 min. na zware arbeid (uitgedrukt in mgr per liter)

	kort voor,	8 min. na,	50 min. na zware arbeid	
A	24	25	31	
B	23	23	22	
C	18	20	21	
D	18	19	18	
E	20	23	29	
F	13	16	17	
G	18	21	22	
H	17	17	21	
I	22	30	19	
J	17	17	15	
K	13	9	9	
L	22	26	26	
M	14	20	22	
N	16	17	24	
O	19	21	21	
P	11	13	15	
Q	14	16	18	
R	17	15	16	
S	15	14	15	
T	11	14	12	
	342	376	393	$[x] = 1111$
Gemiddeld:	17.1	18.8	19.65	

afzonderlijk tot uiting zijn gekomen. Ze waren echter op verschillende data uitgevoerd en daardoor bevatte dat vorige vraagstuk geen extra inlichting omtrent de storende invloeden.

In ons nieuwe vraagstuk kunnen de 60 gegeven meetgetallen dientengevolge op 2 zinrijke manieren worden gegroepeerd, n.l. 1e in drie *kolommen*, die elk de meetgetallen bij één toestand ten opzichte van de arbeid weergeven (voor, kort na of lang na de arbeid) en 2e in 20 *regels*, die elk de meetgetallen bij één persoon vermelden. De meetgetallen vormen hier, zoals dat heet, een rechthoek, waarvan de twee richtingen elk een betekenis hebben.

De onzekerheid van onze blik wordt door deze tweede inlichting

kleiner. Wij bepalen weer de totale kwadraatjessom (de berekening, die op blz. 134 e.v. uitgevoerd en toegelicht is, geeft daarvoor 1385) en dan zowel de totale kwadraatjessom tussen de 20 personenklassen (1093) als die tussen de 3 arbeidsklassen (67.4) en dan blijft voor de kwadraatjessom van het „residu" de totale over minus de *beide* andere kwadraatjessommen:

$$\begin{aligned}[\text{residu}] &= [\text{totaal}] - [\text{personen}] - [\text{arbeid}] \\ &= 1385 - 1093 - 67.4 \\ &= 224.6\end{aligned}$$

Voor de valenties geldt hetzelfde. De totale valentie is het totale aantal meetgetallen minus 1, dus $60 - 1 = 59$, de valentie tussen de 20 personen is 19, die tussen de 3 arbeidsklassen 2, zodat de valentie van het residu is

$$\begin{aligned}N_{\text{res.}} &= N_{\text{tot.}} - N_{\text{pers.}} - N_{\text{arb.}} = \\ &= 59 - 19 - 2 = \\ &= 38.\end{aligned}$$

Wij vragen nu of de strooiïng tussen de arbeidsklassen (s_1) significant groter is dan die van het residu (s_2):

$$s^2_1 = [\text{arbeid}] : N_{\text{arb.}} = 67.4 : 2 = 33.7$$
$$s^2_2 = [\text{residu}] : N_{\text{res.}} = 224.6 : 38 = 5.91$$

en vinden voor de strooiïngsverhouding uit

$$s^2_1 : s^2_2 = 33.7 : 5.91 = 5.70$$
$$s_1 : s_2 = \sqrt{5.70} = 2.4 \qquad \text{met } N_1 = 2 \text{ en } N_2 = 38$$

De nomogrammen van de strooiïngsverhouding leren, dat bij $N_1 = 2$ en $N_2 = 38$ de strooiïngsverhouding 2.4 een overschrijdingskans P_1 van ongeveer 0.01 heeft, de G^2-methode, die een nauwkeuriger aflezing toelaat, levert $P_1 = 0.004$. De overschrijdingskans is dus veel kleiner dan de gebruikelijke grenswaarde 0.05, de strooiïng tussen de arbeidsklassen is dus significant groter dan die van het residu, m.a.w. de arbeid heeft waarschijnlijk wèl invloed op het vitamine-C-gehalte in het bloed uitgeoefend (in slechts 1 op de 250 gevallen zouden de toevalsfouten zodanige afwijkingen tussen de gemiddelden als hier gevonden waren of nog grotere veroorzaken).

In dit vraagstuk kan duidelijk worden gedemonstreerd, dat de gevoeligheid van het onderzoek door het toepassen van de dubbele variantie-analyse is toegenomen. Dat is een gevolg van het feit, dat de

verschillende personen een sterk verschillende invloed op de meetgetallen hebben uitgeoefend [1]).

Indien wij niet deze dubbele variantie-analyse hadden toegepast maar de enkelvoudige — dat wil zeggen als wij de invloed van de personen niet nader hadden geanalyseerd en dus de kwadraatjessom en de valentie van de personen niet mede van de totalen hadden afgetrokken — dan was de strooiïng binnen de klassen veel groter uitgevallen dan nu die van het residu. Wij hadden dan van de totale kwadraatjessom immers alleen de kwadraatjessom tussen de arbeidsklassen afgetrokken (voor de overzichtelijkheid noemen wij hier de kwadraatjessom binnen de klassen ten onrechte „residu"):

$$[\text{residu}] = [\text{totaal}] - [\text{arbeid}] = \\ = 1385 - 67.4 = 1317.6$$

$$N_{\text{res.}} = N_{\text{tot.}} - N_{\text{arb.}} = \\ = 59 - 2 = 57$$

$$s^2_2 = [\text{residu}] : N_{\text{res.}} = \\ = 1317.6 : 57 = 23.1$$

De strooiïng tussen de arbeidsklassen zou dezelfde waarde hebben gehad als bij de dubbele analyse, zodat voor de invloed van de arbeid de strooiïngsverhouding zou gevolgd zijn uit

$$s^2_1 : s^2_2 = 33.7 : 23.1 = 1.46$$

$$s_1 : s_2 = \sqrt{1.46} = 1.2 \quad \text{met } N_1 = 2 \text{ en } N_2 = 57$$

In plaats van de zeer significante strooiïngsverhouding 2.4 zouden wij dus de niet significante strooiïngsverhouding 1.2 hebben gevonden (overschrijdingskans $P_1 = 0.2$).

---

[1]) Wij stellen die invloed van de personen vast door de strooiïng tussen de personen-klassen met die van het residu te vergelijken. Voor de personenklassen vinden wij

$$s_1^2 = [\text{pers.}] : N_{\text{pers.}} = 1093 : 19 = 57.5$$

$$s_2^2 = [\text{residu}] : N_{\text{res.}} = 224.6 : 38 = 5.91 \text{ (als boven)}$$

$$s_1^2 : s_2^2 = 57.5 : 5.91 = 97$$

$$s_1 : s_2 = \sqrt{9.7} = 3.1 \text{ met } N_1 = 19 \text{ en } N_2 = 38$$

$$\text{bij } s_1 : s_2 = 1.8 \text{ en } N_1 = 19,\ N_2 = 38 \text{ is } P_1 = 0.001$$

$$P_1 \ll 0.001$$

of, omdat N_2 vrij groot is:

$$G^2_{19} = 38 \times 1093 : 224.6 = 185$$

$$P_1 \ll 0.0001$$

De oorzaak van dit verschil tussen beide methoden blijkt het duidelijkst als wij de kwadraatjessommen, de valenties en de daaruit volgende strooiingskwadraten en strooiingen telkens in een tabel overzichtelijk weergeven, waarbij dan op te merken is, dat de (vetgedrukte) kwadraatjessom van het residu telkens zó groot is, dat de optelling de (constante) totale kwadraatjessom levert en dat hetzelfde voor de valentie van het residu geldt:

TABEL XII

Overzicht enkelvoudige en dubbele variantie-analyse van het vitamine-C-gehalte in het bloed

	kwadr. som	valentie	s^2	s
Enkelvoudige variantie-analyse				
tussen de arbeidsklassen	67.4	2	33.7	5.81
residu (binnen de klassen)	**1317.6**	**57**	23.1	**4.81**
	+	+		
totaal	1385.0	59	23.5	4.85
Dubbele variantie-analyse				
tussen de arbeidsklassen	67.4	2	33.7	5.81
tussen de personenklassen	1093.0	19	57.5	7.68
residu	**224.6**	**38**	5.9	**2.44**
	+	+		
totaal	1385.0	59	23.5	4.85

De dubbele variantie-analyse, waarbij de invloed van de personen wordt geëlimineerd (althans zover zulks mogelijk is) brengt dus uit de gegeven meetgetallen aan het licht, dat de arbeid invloed op het vitamine-C-gehalte heeft; de enkelvoudige variantie-analyse, waarbij de invloed van de personen niet in rekening wordt gebracht, is daar niet toe in staat, ontleent aan dezelfde meetgetallen zelfs geen suggestie daarvoor.

Meetgetallen zijn, zoals reeds in de inleiding werd gezegd, steeds onnauwkeurig, d.w.z. zij geven datgene wat wij willen uitdrukken slechts bij benadering weer. Wat wij willen te weten komen zien wij als het ware door een mist van onnauwkeurigheid heen. Indien de meetgetallen echter ook over deze mist enige nadere inlichting geven, dan kan dat gedeelte van die mist worden opgeklaard, zodat wij meer kans krijgen door de dunner geworden rest van de mist heen datgene te zien, wat wij wensen waar te nemen.

In algemene zin vormt dit laatste vraagstuk daardoor een aanwijzing, hoe nuttig en dus hoe nodig het is om bij elk statistisch onderzoek eerst helder na te gaan, welke soorten van inlichtingen door de gegevens worden geboden: pas als al deze inlichtingen worden benut krijgt het statistisch onderzoek de grootst mogelijke gevoeligheid en biedt het de grootst mogelijke kans op het bevestigen van de ter toetsing aangeboden opvatting.

Op dezelfde manier als hier werd aangegeven voor het geval dat de meetgetallen op twee manieren kunnen worden gegroepeerd — als zij in een rechthoek met twee betekenisvolle richtingen kunnen worden gerangschikt — kan ook tewerk worden gegaan als ze op meer dan twee manieren kunnen worden gegroepeerd, b.v. als zij kunnen worden gerangschikt in een ruimtelijk (rechthoekig) blok, waarvan de drie richtingen elk een betekenis hebben. In dat laatste geval worden drie kwadraatjessommen (tussen de drie soorten van klassen, die elk uit zulk een groepering, dus uit zulk een richting, volgen) van de totale kwadraatjessom afgetrokken en ook drie valenties van de totale valentie, waardoor meer invloeden worden geëlimineerd en de gevoeligheid nog weer groter wordt.

Meervoudige variantie-analyse kan ook met grote winst voor de gevoeligheid worden toegepast indien de meetgetallen over even grote groepen zijn verdeeld en die groepen op meer dan één manier in klassen kunnen worden samengevat, d.w.z. als die groepen in een rechthoek met twee betekenissen kunnen worden geplaatst (of in een recht blok, waarvan de drie richtingen elk een betekenis hebben, enz.), zoals in het vorige vraagstuk de meetgetallen zelf.

Als voorbeeld daarvan kan nu het volledige onderzoek naar de invloed van het vitamine-preparaat op de phosphaatuitscheiding aan de orde worden gesteld.

> Bij 4 proefpersonen werd gedurende twee perioden van 10 dagen elke dag de phosphaatuitscheiding in de urine gemeten. Alleen in de tweede periode nuttigden zij een bepaald vitamine-C-preparaat. Tonen de verkregen meetgetallen, die in tabel XIII weergegeven zijn, aan, dat de phosphaatuitscheiding door dit vitamine-gebruik veranderd is? [1]).

De situatie in deze proefopstelling is, statistisch gezien, een zodanige als bij zeer vele proefnemingen voorkomt.

---

[1]) Het werkelijke onderzoek omvatte belangrijk meer; voor dit boekje werd het beschreven gedeelte er uit gelicht, zoals ook eerder behandelde voorbeelden slechts kunstmatige brokstukjes van dit grotere onderzoek zijn.

Als in elke periode bij elke proefpersoon maar één meting was verricht, dan zouden wij de 8 verkregen meetgetallen in een rechthoek van 4 regels en 2 kolommen hebben kunnen rangschikken: telkens een regel voor een proefpersoon, telkens een kolom voor de perioden zonder of voor die met het vitaminepreparaat. Statistisch zou de probleemstelling dan dezelfde zijn geweest als in het vorige vraagstuk. Nu in elke periode 10 metingen zijn verricht, en wel op 10 achtereenvolgende dagen die voor de diverse preparaten niet samenvielen, nu is als het ware elke „plaats" in deze rechthoek door 10 meetgetallen bezet. Voor het statistisch onderzoek brengt dit geen principiële verandering met zich mee, doordat dit aantal voor alle „plaatsen" even groot is.

TABEL XIII

De dagelijkse phosphaatuitscheiding in de urine, gemeten in cg bij de 4 proefpersonen A—D in een periode zonder (—) en in een periode met (+) regelmatig gebruik van een vitamine-C-preparaat

A		B		C		D		
—	+	—	+	—	+	—	+	
107	97	139	159	141	137	88	76	
79	89	149	126	131	115	115	77	
92	111	147	147	150	124	92	80	
105	102	146	111	160	128	90	118	
77	97	158	128	143	111	92	76	
111	97	150	137	152	134	74	78	
73	70	174	164	121	132	94	103	
103	94	183	137	115	137	89	89	
94	94	141	142	133	134	100	72	
105	93	134	125	111	131	110	89	
946	944	1521	1376	1357	1283	944	858	$[x] = 9229$

Weer kunnen de meetgetallen op twee manieren in klassen worden verdeeld. Als alle getallen in twee kolommen zouden worden opgeschreven, in de eerste kolom die uit de 4 perioden zonder het vitaminepreparaat en in de tweede kolom die mèt het preparaat, en verder zo, dat de eerste 10 getallen van elke kolom bij de proefpersoon A waren verkregen en telkens elk volgend 10-tal bij een volgende proefpersoon, dan zouden zij gemakkelijk op twee manieren te groeperen zijn, n.l.

1. naast elkaar in de twee kolommen: resp. zonder en met vitamine,

2. van boven naar onder in vier groepen, elk bestaande uit 2 × 10 getallen: resp. proefpersoon A, B, C en D, en daaruit zouden de gemiddelden volgen 1e van de 2 kolommen, zijnde de 2 gemiddelden van de 2 „vitamine-klassen" en 2e van de 4 groepen, zijnde de 4 gemiddelden van de 4 „personenklassen".

Gemakkelijker vinden wij die gemiddelden door de 8 kolomtotalen van tabel XIII op deze wijze in een rechthoek te plaatsen (tabel XIV). De 4 regelsommen van die rechthoek zijn dan de personen-totalen, $[x]_{\text{pers.}}$, de 2 kolomsommen de totalen der „vitamine-klassen", $[x]_{\text{vit.}}$. Deling door de betrokken aantallen getallen, waar deze sommen uit bestaan, levert dan de bedoelde gemiddelden $\hat{x}_{\text{pers.}}$ en $\hat{x}_{\text{vit.}}$ (waarbij elke frequentie in een „personen-klasse" 20 is en in een „vitamine-klasse" 40).

TABEL XIV

Rechthoek, bestaande uit de kolomsommen van tabel XIII

	zonder vit.	met vit.	$[x]_{\text{pers.}}$	$f_{\text{pers.}}$	$\hat{x}_{\text{pers.}}$
A	946	944	1890	: 20 =	94.5
B	1521	1376	2897	: 20 =	144.85
C	1357	1283	2640	: 20 =	132.0
D	944	858	1802	: 20 =	90.1
$[x]_{\text{vit.}}$	4768	4461	$[x] = 9229$		
$f_{\text{vit.}}$	40	40	$n = 80$		
$\hat{x}_{\text{vit.}}$	119.2	111.525		$\bar{x} = 115.3625$	

Voor de kwadraatjessom tussen de vitamine-klassen volgt uit de 2 vitamine-gemiddelden 119.2 en 111.525 — voor de becijfering zie blz. 139 — $[f\hat{x}_{,}^{2}]_{\text{vit.}} = 1178$ en voor die tussen de 4 personenklassen volgt uit de personengemiddelden 94.5 enz. $[fx_{,}^{2}]_{\text{pers.}} = 44395$. Voor de totale kwadraatjessom van alle meetgetallen werd gevonden $[x_{,}^{2}] = 59615$, zodat voor de kwadraatjessom van het residu overblijft $59615 - 1178 - 44395 = 14042$ [1]). De totale valentie bedraagt $N_{\text{tot.}} = n - 1 = 80 - 1 = 79$, de valentie van de 2 vitamine-klassen is 1, die van de 4 personen-klassen 3 en die van het residu dus

[1]) Hetzelfde getal zou ook kunnen worden gevonden door in elke kolom van de oorspronkelijke meetgetallen (tabel XIII) van elk meetgetal het verschil met het kolomgemiddelde in het kwadraat te verheffen en deze 80 kwadraten bij elkaar op te tellen.

79 — 1 — 3 = 75, zodat voor de vitamine-klassen geldt (zie blz. 55) $t^2 = 79 \times 1178 : 14042 = 6.63$ of $t = 2.6$. Bij $N = 79$ wijst het t-N-P-nomogram hierbij de overschrijdingskans 0.01 aan, zodat de strooiïng van de vitamine-gemiddelden significant groter is dan die van het residu: de conclusie luidt, dat het gebruik van het vitamine-preparaat waarschijnlijk invloed op de phosphaat-uitscheiding heeft uitgeoefend: de meet- en toevalsfouten kunnen althans geacht worden, in slechts 1 op ongeveer 100 gevallen even grote of ernstiger afwijkingen te veroorzaken (zie echter nog blz. 117).

Ook in dit vraagstuk zou geen significant verschil zijn gevonden als de invloed van de diverse proefpersonen niet, door middel van een meervoudige variantie-analyse, in rekening was gebracht.

CORRELATIE

In de laatstbehandelde vraagstukken werd onderzocht, of een bepaalde factor (het vitamine-preparaat) invloed had op een bepaalde afmeting (de phosphaatstofwisseling). Wij kunnen daar nog aan toevoegen, dat dit ook mogelijk is als de eerstbedoelde factor continu-variabel is, b.v. als gevraagd wordt of de lichaamslengte invloed heeft op de polsfrequentie: is daartoe van vele personen de lichaamslengte en de polsfrequentie gemeten, dan kan de schaal, waarin de lichaamslengten zijn genoteerd, in een aantal stukken worden verdeeld — b.v. de stukken van 0—140 cm, van 140—150 cm, van 150—160 cm enz. — en dan worden alle personen, waarvan de lichaamslengte in eenzelfde stuk van de schaal valt, opgevat als individuen van eenzelfde klasse, waarna de methode der strooiïngsverhouding op die — kunstmatige — klassen kan worden toegepast.

Blijkt zulk een invloed te bestaan, dan noemt men de bedoelde factor en de bedoelde afmeting of de twee variabele afmetingen „gecorreleerd". Waarop berust zulk een gecorreleerd zijn van twee afmetingen?

Het feit, dat een variabele grootheid bij een bepaald individu een zekere waarde heeft aangenomen, is in den regel het gevolg van het samenwerken van een aantal factoren: dat een bepaalde persoon 176 cm lang is, is het gevolg van de lengte van verscheidene zijner beenderen, van de kromming zijner wervelkolom, van de hoeveelheid vet onder de huid van zijn voetzool en van dergelijke factoren meer. Dat twee variabele grootheden — b.v. de lichaamslengte en de polsfrequentie — bij één individu elk een bepaalde waarde hebben, is het gevolg van twee reeksen factoren; hebben die twee reeksen een of meer factoren gemeen, dan zijn de waarden, die die twee afmetingen bij één individu hebben,

ten dele door dezelfde factoren bepaald. In dat geval variëren die twee afmetingen niet onafhankelijk van elkaar maar zijn zij gecorreleerd.

Het feit van de correlatie wijst dan ook aan, dat de individuele waarden van twee variabele grootheden geheel of ten dele door een of meer gemeenschappelijke oorzaken zijn bepaald.

In het eene geval van correlatie hebben de gemeenschappelijke oorzaken slechts een geringe rol gespeeld bij het tot stand komen van de individuele waarden der twee gecorreleerde afmetingen, die dan dus in hoofdzaak tot stand zijn gebracht door de niet-gemeenschappelijke oorzaken. In dat geval noemt men de correlatie gering. De correlatie is daarentegen groot als de gemeenschappelijke factoren een grote rol spelen ten opzichte van de niet-gemeenschappelijke. In al deze gevallen noemt men het verband tussen de twee afmetingen „stochastisch": bij één waarde van de ene afmeting komen — als gevolg van de niet-gemeenschappelijke oorzaken — uiteenlopende waarden van de andere afmeting voor, maar bij verschillende waarden van de ene loopt het gemiddelde van de andere uiteen. Zijn er echter geen gemeenschappelijke oorzaken, dan is de correlatie afwezig (nul) en variëren de twee afmetingen „onafhankelijk van elkaar". Zijn er uitsluitend gemeenschappelijke oorzaken, dan is de correlatie maximaal en heet het verband tussen de twee afmetingen„functioneel": bij één waarde van de ene afmeting behoort dan één bepaalde afmeting van de andere, zoals in cirkels bij één lengte van de straal slechts één lengte van de omtrek voorkomt; de ene afmeting kan dan in beginsel door middel van een wiskundige vergelijking worden voorgesteld als een „functie" van de andere. Vandaar de naam: functioneel verband.

Zoals wij reeds meermalen zagen kan de mate, waarin een variabele afmeting varieert, worden gekenschetst door haar standaarddeviatie, S. Deze deviatie is m.a.w. een maat voor de invloed, die die afmeting ondervindt van alle factoren samen, die haar doen variëren. Is zij gecorreleerd aan een andere variabele afmeting, dan wordt een deel van haar standaard-deviatie veroorzaakt door de zoëven bedoelde gemeenschappelijke factoren, dat zijn die, die op beide afmetingen invloed uitoefenen. Bedoeld deel van de totale standaard-deviatie is dus een maat voor de invloed van de gemeenschappelijke factoren. En de verhouding tussen dit deel van de standaard-variatie tot de totale is dientengevolge een maat voor de correlatie. Deze verhouding heet de *correlatie-coëfficient*, dat is dus *het getal dat aangeeft, welk deel van de totale standaard-deviatie door de gemeenschappelijke factoren veroorzaakt is.* Stellen wij de correlatie-coëfficient voor door het symbool R en dat deel van de standaard-deviatie, dat door de gemeen-

schappelijke factoren veroorzaakt wordt, door S_g, dan is dus

$$R = S_g : S$$

Daar elke standaard-deviatie positief is heeft de correlatie-coëfficient, R, steeds een positieve waarde. De grootste waarde, die zij kan bereiken, is 1, want S_g is een deel van de totale deviatie S en kan dus hoogstens gelijk zijn aan die totale deviatie. Dit is het geval als er uitsluitend gemeenschappelijke factoren zijn, dus bij een functioneel verband. En de kleinste waarde, die R kan hebben, is 0: er zijn dan geen gemeenschappelijke factoren, zodat $S_g = 0$; dit is het geval bij twee grootheden, die onafhankelijk van elkaar variëren.

Uit een complex van gegeven getallen kan de beste schatting voor de correlatie-coëfficient, r, op verschillende manieren worden berekend.

Als een reeks getallenparen is gegeven, verkregen door bij elkaar behorende metingen van twee variabele afmetingen, dan wordt het best de methode van de variantie-analyse gevolgd op de manier zoals in een vroeger hoofdstuk werd aangegeven, waarna de gevraagde schatting wordt becijferd uit de daarbij gevonden strooiingsverhouding $s_1 : s_2$ volgens de formule

$$\boxed{r^2 = \frac{s_1^2 : s_2^2}{N + s_1^2 : s_2^2}} \qquad \text{waarbij } N = n-2$$

Van de overige situaties, waarin de correlatie-coëfficient kan worden bepaald, wordt hier slechts de vier-velden-methode vermeld: de beste schatting voor de correlatie-coëfficient tussen de twee bestudeerde eigenschappen volgt uit

$$\boxed{r = (ad - bc) : \sqrt{pqrs}}$$

waarin a, b, c en d, zoals op blz. 48, de gegeven aantallen voorstellen en p, q, r en s hun dubbelsommen.

De correlatie-schatting, r, die met behulp van deze formule becijferd is, heeft ten naaste bij een normale kansdistributie, zodat de schatting voor de standaard-deviatie daarvan kan worden gebruikt als de standaardfout voor de berekende correlatie-coëfficient. Daarvoor geldt

$$\boxed{s_r^2 = n^2 : pqrs\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}$$

waarin n het totale aantal individuen van de steekproef voorstelt (dus $n = a + b + c + d$). Als voorwaarde voor het gebruik van deze

twee formules geldt, bij correct genomen steekproef, dezelfde eis als reeds op blz. 48 voor deze methode werd gesteld, n.l. dat geen van de vier gegeven aantallen (*a*, *b*, *c* en *d*) kleiner mag zijn dan 6.

Als voorbeeld wordt het volgende vraagstuk uit de praktijk besproken.

Bij het onderzoek over de correlatie van het hypomane karakter en uitingen van aktieve fantasie, waarover op blz. 45 werd gesproken, verkreeg Dr. VAN KREVELEN bij niet-enige kinderen de vier aantallen, die hieronder binnen het vierkant van tabel XV weergegeven zijn. Hij stelde de vraag, of de correlatie tussen het hypomane karakter en de fantasie volgens deze gegevens bij de enige kinderen, waarvoor het vierkant van blz. 46 was verkregen, een andere grootte heeft dan bij niet-enige kinderen [1]).

TABEL XV (niet-enige kinderen)

		hypomaan karakter		
		wel	niet	totalen
actieve fantasie	wel	12	5	17
	niet	17	16	33
totalen		29	21	50

Op blz. 145 wordt met behulp van de eerste der zo juist vermelde formules uit deze gegevens voor de niet-enige kinderen als beste schatting voor de correlatie-coëfficient uit het rechter vierkant becijferd $r = 0.183$, terwijl uit het vierkant van blz. 46 op dezelfde manier voor de enige kinderen was berekend $r = 0.346$, zodat het verschil tussen

---

[1]) Uit de gegevens omtrent de niet-enige kinderen blijkt niet betrouwbaar, dat er correlatie tussen beide eigenschappen bestaat: de berekening, uitgevoerd volgens blz. 48, levert $t = 1.29$, zodat de overschrijdingskans van de nul-hypothese groter is dan 0.05. Daaruit mag dus niet worden geconcludeerd, dat de correlatie bij de niet-enige kinderen kleiner is dan die bij enige kinderen (waarvan de gegevens de correlatie zeer betrouwbaar aantonen (met $t = 4.5$). Uit het verschil der gevonden excentriciteiten t volgt slechts, dat de correlatie door de steekproef uit de enige kinderen betrouwbaarder werd aangetoond dan door de steekproef uit de niet-enige kinderen en dat kàn het gevolg zijn van verschillend sterke correlatie, maar ook van het verschil in grootte der steekproeven.

deze twee beste schattingen 0.346 — 0.183 = 0.163 bedraagt. Om na te gaan of betekenis aan dit gevonden verschil mag worden gehecht stellen wij als nul-hypothese, dat bij de twee soorten kinderen de correlatie tussen hypomanie en fantasie dezelfde zou zijn. In dat geval zou het gevonden verschil uitsluitend op de toevallige steekproefselectie berusten en daar elk van de twee schattingen geacht wordt bij benadering een normale kansdistributie te bezitten, achten wij ook de kansdistributie van hun verschil normaal. De variantie van dit verschil is de som van de varianties der schattingen zelf. Voor de niet-enige kinderen vinden wij met behulp van de tweede der zo juist gegeven formules $s_r^2 = 0.00525$ en voor de enige kinderen $s_r^2 = 0.01806$, zodat voor de standaardfout van het verschil geldt $s_d^2 = 0.00525 + 0.01806 = 0.0233$ of $s_d = 0.153$. Het verschil der correlaties (0.163) is nauwelijks groter dan de standaardfout (veel kleiner dan 2 maal de standaardfout) en dus niet significant, zelfs niet suggestief (nauwkeuriger: $t = d : s = 0.163 : 0.153 = 1.07$, $P_2 = {}^1/_3$ of veel groter dan ${}^1/_{20}$). De gegevens vormen dus geen aanwijzing, dat de correlatie tussen hypomanie en fantasie bij enige kinderen anders zou zijn dan bij niet-enige kinderen.

Wel werd in de bovenstaande alinea's aangewezen hoe een correlatie wordt veroorzaakt, maar een definitie moet er nog van worden gegeven. Als twee variabele afmetingen samen gemeten zijn, dan vormen de verkregen twee meetgetallen een „meetpaar"; zijn zij een aantal malen telkens samen gemeten, dan is een even groot aantal meetparen verkregen. Noemen wij de meetgetallen van de ene grootheid x en van de andere y, dan bestaat elk paar uit een x-waarde en *de daarbij behorende y-waarde.* Als bij deze metingen (of bij herhaling ervan) eenzelfde x-waarde meer dan eens wordt aangetroffen, dan zijn de y-waarden, die bij diezelfde x-waarde behoren, allicht niet alle aan elkaar gelijk. Wij kunnen dus spreken van de kansdistributie van die y-waarden, die bij één bepaalde x-waarde behoren. En zo kan ook worden gesproken over de kansdistributie der y-waarde(n) bij elke willekeurig gekozen andere x-waarde. En daarmee kan het begrip correlatie worden gedefinieerd: *de getallen x en y heten gecorreleerd indien bij verschillende waarden van x de kansdistributie van y verschillend is.* In de praktijk wordt hier gewoonlijk van gemaakt: indien bij verschillende waarden van x de kansdistributies van y *verschillende centra* hebben.

REGRESSIE

Terwijl een onderzoek naar de correlatie tussen twee variabele afmetingen nagaat òf bij verschillende waarden van de eerste afmeting het centrum van de tweede afmeting verschillend is, wordt bij een

onderzoek naar hun regressie nagegaan, *hoe* dat centrum van de tweede verandert bij veranderingen van de eerste. Anders gezegd: terwijl de correlatie aanwijst *hoe sterk* het verband tussen twee grootheden is, dus o.a. òf er verband tussen is, wijst de regressie aan *welke vorm* dat verband heeft.

Lichten wij deze abstracte omschrijving door een concreet voorbeeld toe.

Het onderzoek naar de z.g. sterrennevels heeft het heel waarschijnlijk gemaakt, dat de sterren in het voor ons zichtbare heelal niet gelijkmatig over de ruimte verspreid zijn maar gegroepeerd in een aantal z.g. sterrenstelsels: binnen elk sterrenstelsel liggen de sterren op betrekkelijk kleine onderlinge afstanden, terwijl de verschillende sterrenstelsels zich op veel grotere afstanden van elkaar bevinden. Alle met ons ongewapende oog of zelfs door een matige kijker afzonderlijk zichtbare sterren behoren tot „ons" sterrenstelsel, dat is dat waarvan de zon een der sterren is (doordat het de vorm van een zeer platte schijf heeft zien wij de meeste sterren in de richtingen, die met de vlakteuitbreiding van die schijf overeenkomen, dus langs een grote band, de melkweg). De andere sterrenstelsels zijn de z.g. sterrenwolken, de extragalactische nevels, waarvan er maar heel enkele zonder kijker afzonderlijk zichtbaar zijn. Volgens de bekende opvatting van het „uitdijend heelal" zouden al deze sterrenstelsels eens bij of misschien zelfs door elkaar heen hebben gelegen; een kosmische catastrophe zou dit oorspronkelijke, grote complex uit elkaar hebben doen spatten in een aantal brokstukken, de sterrenstelsels, welke daarbij verschillende snelheden zouden hebben gekregen die echter sindsdien elk constant zouden zijn gebleven.

Deze hypothese van het uitdijend heelal is slechts houdbaar indien de snelheden, waarmee de diverse sterrenstelsels zich van ons sterrenstelsel verwijderen, evenredig zijn aan hun afstand tot ons sterrenstelsel. Tussen de bedoelde snelheden, die wij X zullen noemen, en de bedoelde afstanden, die wij met Y betitelen, zou dus het functionele verband $Y = BX$ bestaan, waarin B een constant getal voorstelt, de evenredigheidsfactor. Drukken wij de snelheden uit in kilometers per seconde en de afstanden in millioenen lichtjaren, dan zou B ongeveer $^1/_{150}$ bedragen, dus $Y = {}^1/_{150}\, X$.

Bij een aantal sterrenstelsels konden de snelheden — in hoofdzaak door het principe van DOPPLER toe te passen op de spectraallijnen — betrekkelijk nauwkeurig worden gemeten, de bedoelde afstand daarentegen slechts zeer onnauwkeurig. Tussen de verkregen meetgetallen bestaat dus geen functioneel, ten hoogste een stochastisch verband. Op beide variabele grootheden — x = de meetgetallen voor de snelheid

en y = de meetgetallen voor de afstand — zou immers in de eerste plaats een gemeenschappelijke factor hebben ingewerkt, n.l. de stoot, die het betreffende sterrenstelsel bij de catastrophe had ontvangen: deze stoot veroorzaakte de snelheid en verder bepaalde deze stoot, samen met de sinds de catastrophe verlopen tijd, de hedendaagse plaats en daarmee de onderlinge afstanden van de sterrenstelsels. In de tweede plaats werd één van de twee variabele grootheden — n.l. y, het meetgetal voor de afstand — nog door niet-gemeenschappelijke factoren beïnvloed, n.l. door die factoren, die de opgetreden meetfouten hebben veroorzaakt (wij verwaarlozen hier de meetfouten, die bij het meten van de bedoelde snelheden waren gemaakt, omdat deze zo heel veel kleiner zijn dan de gemeten grootheden en ook veel kleiner dan de fouten, gemaakt bij het meten van de afstanden). Volgens de opvatting van het uitdijend heelal zou dus een sterrenhoop, die zich met een snelheid van b.v. 900 km/sec van ons verwijdert, nu op een afstand van $^1/_{150} \times 900$ of 6 millioen lichtjaren van ons af zijn. Bij het meten van die afstand werd allicht een flinke meetfout gemaakt, zodat het daarbij gevonden meetgetal allicht niet nauwkeurig gelijk zal zijn aan 6 maar een kansdistributie zal hebben waarvan het centrum gelijk is aan 6 en waarvan de standaarddeviatie die is van de meetfouten. En bij een sterrenhoop, die zich met een snelheid van b.v. 1500 km/sec verwijdert, zal het meetgetal voor de afstand volgens de opvatting van het uitdijend heelal een centrum van $^1/_{150} \times 1500$ of 10 millioen lichtjaren hebben, zodat in het algemeen de vorm van het verband tussen deze twee variabele grootheden wordt aangegeven door de vergelijking $Y = {}^1/_{150}\, x$, waarbij Y telkens het centrum voorstelt van die y-waarden, die bij een willekeurig gekozen x-waarde behoren.

De vergelijking, die op deze wijze de vorm van het verband tussen twee variabelen aangeeft, heet een *regressievergelijking*, in het bovenbedoelde geval *de regressievergelijking van y op x*. In het besproken voorbeeld geeft deze vergelijking tevens het functionele verband tussen de gemetene eigenschappen zelf aan, wat echter niet steeds het geval behoeft te zijn.

Er zijn natuurlijk ontelbaar vele vormen van regressie-vergelijkingen denkbaar. Luidt die vergelijking $Y = 3 + 4x$ (of een waarin de getallen 3 en 4 door andere getallen vervangen zijn) dan heet de regressie rechtlijnig of lineair: bij graphische voorstelling van de x- en de y-waarden in een rechthoekig diagram liggen de centra, die door deze vergelijking worden aangegeven, op een rechte lijn. De vergelijking $Y = 3 + 4x - 2x^2 - 5x^3$ geeft een 3e-graadsregressie aan, $Y = 3 + 2 \log x$ een logarithmische enz. enz. Keert

telkens in een volgende periode van x-waarden eenzelfde reeks waarden voor de centra Y op, dan wordt van een periodische regressie gesproken, b.v. bij de regressie-vergelijking $Y = 23 + 15 \sin 3x$ (telkens als x met 120 is toegenomen en $3x$ dus met 360 heeft $\sin 3x$, dus ook Y, dezelfde waarde).

Uit gegeven meetgetallen kan de regressie-vergelijking in beginsel nooit worden uitgerekend als ware het een sommetje, waarop maar één antwoord past. Dit zou wel kunnen bij een functioneel verband, want daarbij komt naast elke waarde van de afmeting x slechts één waarde van de afmeting ij voor (doordat er geen niet-gemeenschappelijke oorzaken op ij inwerken). Hierboven werd reeds als voorbeeld de verhouding tussen de straal en de omtrek van verschillend grote cirkels genoemd; stellen wij de lengte van de straal voor door x en die van de omtrek door ij, dan geldt voor elke cirkel het functionele verband $ij = 2\pi x$. Bij het meten van deze stralen en cirkels zouden echter allicht uiteenlopende meetfouten optreden, waardoor het verband tussen de meetgetallen niet meer zuiver functioneel is maar stochastisch: bij één meetwaarde voor de stralen zullen allicht uiteenlopende meetwaarden voor de omtrekken worden gevonden. En daardoor wijzen meetgetallen nooit nauwkeurig aan, welke de juiste regressie-formule is. Zo zouden zij in het voorbeeld van de cirkel wel de suggestie inhouden, dat de regressievergelijking een evenredigheid tussen straal en cirkel zou moeten aanduiden ($Y = Ax$) en ook zou volgens die meetgetallen de evenredigheidsfactor (A) dan in de buurt van 2×3.14 moeten liggen, maar zij zouden toch ook de mogelijkheid openlaten, dat deze evenredigheidsfactor een iets andere waarde zou hebben en zelfs dat het verband iets anders dan een evenredigheid zou wezen.

Bij het opstellen van een regressievergelijking speelt dan ook steeds de fantasie van den onderzoeker een zekere rol, zij het natuurlijk geenszins de enige. De rol, die de meetgetallen bij dit ontwerpen van een regressie-vergelijking spelen, is drieërlei: zij vormen een suggestie in de richting van een mogelijke gedaante der regressie-vergelijking, zij kunnen de vermoedelijkerwijs ontworpen regressie-vergelijking nader quantitatief preciseren en zij bieden gelegenheid die vermoede regressie-vergelijking te toetsen.

Onder de *gedaante* van een regressie-vergelijking wordt verstaan de algebraïsche vergelijking, waarin, in plaats van de nodige getalswaarden, letters voorkomen. Een bepaalde gedaante betekent dus b.v. een evenredigheid ($Y = Ax$) of een logarithmisch verband ($Y = A + B \log x$) e.d. De suggestie, die de meetgetallen inhouden in de richting van zulk een gedaante, kan belangrijk worden versterkt door eerst

grafische voorstellingen te maken van de meetgetallen of van er van af te leiden grootheden (bepaalde machten ervan, hun logarithme, bepaalde gemiddelden, o.d.) en door dan in de op die manier verkregen puntenvelden rechte of eenvoudig gekromde lijnen zo goed mogelijk tussen de desbetreffende punten te schetsen en een algebraïsche formule voor zulk een lijn uit die figuur af te leiden.

Om een in gedaante ontworpen regressie-vergelijking nader quantitatief te preciseren, dienen de „constanten", die daarin als letters voorkomen (de letters A, B, C en D in de als voorbeelden gegeven regressie-vergelijkingen) te worden vervangen door getallen. Ook hierbij spelen de gegeven meetgetallen een zeer belangrijke rol: uit deze meetgetallen toch dienen de beste schattingen voor deze constanten te worden afgeleid. De manier, waarop dat gebeurt, loopt voor de diverse gedaanten der regressie-vergelijkingen sterk uiteen en in het kader van dit overzicht der gedachtengang van de statistiek is het niet nodig, deze manieren te behandelen. Er kan worden volstaan met de opmerking, dat het berekenen er van steeds weer is gebaseerd op de definitie van de beste schatting als eigen aan die denkbare werkelijkheid, die een grotere kans op het optreden der gevonden meetgetallen biedt dan elke andere werkelijkheid.

Het toetsen van de hypothetisch geopperde en in getallen gepreciseerde regressie-vergelijking aan de gegeven meetgetallen is als gewoonlijk het beantwoorden van de vraag, of de afwijkingen tussen de meetgetallen en deze hypothese een voldoend grote kans hebben om uitsluitend door de meetfouten te zijn ontstaan, dat wil zeggen ook de regressie-vergelijking wordt getoetst door de overschrijdingskans te bepalen van de afwijkingen tussen de verkregen meetgetallen en de waarden, die theoretisch zouden kunnen worden verwacht ingeval die regressie-vergelijking juist zou zijn. En daartoe kan, in het algemeen gesproken, weer het best de methode van de strooiïngsverhouding worden toegepast. Weer wordt dan de schaal der x-waarden in een aantal stukken verdeeld en wordt in elke daardoor ontstane klasse het gemiddelde der y-waarden bepaald. De theoretische verwachting luidt nu niet, zoals in de vorige paragrafen, dat deze gemiddelden aan elkaar gelijk zouden zijn, maar dat elk hunner gelijk zou zijn aan die waarde van y, die verkregen wordt door in de te toetsen regressie-vergelijking de letter x te vervangen door het gemiddelde van de x-waarden in de betrokken klasse. De afwijkingen der y-waarden ten opzichte van hun klassegemiddelden, de afwijkingen dier klassegemiddelden ten opzichte van hun theoretisch verwachte waarden en de afwijkingen dezer theoretisch verwachte waarden ten opzichte van het totale gemiddelde worden dan in beginsel op eenzelfde wijze onderzocht als dat bij de

behandeling der variantie-analyse werd besproken [1]). Op de details van dit onderzoek behoeft hier weer niet nader te worden ingegaan, noch ook op de wijze, waarop dit onderzoek kan worden vereenvoudigd als de te toetsen regressie-vergelijking bepaalde eenvoudige gedaanten heeft.

Uit dit summiere overzicht van het begrip der regressie moge echter wel blijken, dat van de meest eenvoudige tot in zeer samengestelde problemen der statistiek éénzelfde gedachtengang de onderzoekingsmethoden heeft gericht, een gedachtengang, die uit drie begrippen voortvloeit: overschrijdingskans, beste schatting en binomiale, resp. normale kansdistributie.

RECHTLIJNIGE REGRESSIE

Om een concreter beeld van het begrip der regressie te geven wordt van één regressie-vorm een voorbeeld behandeld, n.l. van de rechtlijnige regressie.

Een regressie heet rechtlijnig of lineair als zij voldoet aan de vergelijking van de eerste graad $Y = A + Bx$. De naam lineaire of rechtlijnige regressie is te danken aan de bijzonderheid, dat, als elke combinatie van een x-waarde en de volgens die vergelijking daarbij behorende y-waarde als een punt wordt weergegeven in een rechthoekig X-Y-diagram, de relatiepunten in deze grafische voorstelling op een rechte lijn liggen. Dit blijkt uit fig. 11, waarin voor een willekeurig punt P van de rechte lijn AD geldt.

$$\begin{aligned} y &= CP = \\ &= CE + EP = \\ &= CE + AE\, tg\, \alpha = \\ &= OA + OC\, tg\, \alpha = \\ &= OA + x\, tg\, \alpha \end{aligned}$$

Vervangen wij hierin OA door de letter A en $tg\, \alpha$ door de letter B, dan krijgen wij

$$y = A + B\, x$$

welke vergelijking blijkbaar voor alle punten van de rechte lijn AD geldt.

---

[1]) Indien een reeks meetgetallen de suggestie heeft gevormd voor het opstellen van een regressie-vergelijking, dan zou het toetsen van die regressie-vergelijking aan dezelfde meetgetallen geen correct resultaat opleveren. Voor het toetsen dient men dan gebruik te maken van een nieuwe reeks meetgetallen, die onafhankelijk van de eerste tot stand gekomen was. Het quantitatief preciseren en het toetsen van een regressievergelijking behoort daarentegen met behulp van dezelfde meetgetallen te worden uitgevoerd.

Indien van twee variabele grootheden, waarvan gevraagd wordt of hun verband rechtlijnig is, ettelijke meetparen zijn gegeven, dan zal de grafische voorstelling dier meetparen allicht niet zo zijn, dat de relatie-punten nauwkeurig op een rechte lijn liggen. Dat zou n.l. alleen het geval zijn als het verband behalve rechtlijnig ook functioneel is. Zodra ook niet-gemeenschappelijke factoren de individuele waarden hebben beïnvloed, liggen de punten in de graphische voorstelling om een rechte lijn „gestrooid", zoals in fig. 13.

Fig. 11.
De vergelijking $ij = A + Bx$ drukt een rechtlijnig verband uit.

Een hypothese, volgens welke de rechtlijnige regressie-vergelijking $Y = A + Bx$ op zulke gegeven meetgetallen zou passen, kan in deze vorm niet worden getoetst: een algemene gedaante der regressie-vergelijking geeft bij de diverse x-waarden geen theoretisch verwachte y-waarden aan. Daarom wordt eerst voor de constanten A en B de beste schatting uit de gegeven meetparen berekend, waarmee de „best passende regressie-vergelijking" is gevonden, dat is die waarin de letters A en B door getallen, n.l. door die beste schattingen, zijn vervangen. Deze, „in getallen gepreciseerde hypothese" kan dan wèl aan de gegeven meetgetallen worden getoetst, doordat zij wèl theoretisch te verwachten y-waarden aangeeft. Als deze regressie-vergelijking dan blijkt, niet in strijd met de meetgetallen te zijn, dan is ook de alge-

mene gedaante $Y = A + Bx$ en dus de rechtlijnige regressie er niet mee in strijd.

De beste schattingen voor de constanten A en B zijn verschillend naar gelang de x- en de y-waarden beide door „niet-gemeenschappelijke factoren" (zie blz. 72) waren beïnvloed dan wel alleen de x-waarden of alleen de y-waarden. Als b.v. twee variabele grootheden X en Y nauwkeurig evenredig aan elkaar variëren (zoals de omtrek en de middellijn van een cirkel) dan heerst tussen elke twee bij elkaar behorende individuele waarden telkens het lineaire verband $Y = AX$ (bij de genoemde afmetingen van cirkels zou $A = \pi$ zijn). Dit rechtlijnige verband is functioneel doordat de individuele waarden van beide grootheden uitsluitend door gemeenschappelijke oorzaken tot stand gekomen zijn. Wordt nu een reeks van deze individuele afmetingen gemeten, dan is elk meetgetal $x_,$ verkregen door het meten van een individuele waarde van de afmeting X, opgebouwd uit 1e een X-waarde en 2e een meetfout en dan is elk meetgetal y opgebouwd uit 1e een Y-waarde en 2e een meetfout. De x- en de y-waarden zijn dan niet uitsluitend door gemeenschappelijke factoren bepaald maar ook — n.l. voor zover zij uit een meetfout bestaan — door niet-gemeenschappelijke factoren. Is daarbij de meting van de grootheid X zoveel nauwkeuriger geweest dan die van Y, dat de meetfout in de x-waarden praktisch kan worden verwaarloosd, dan geldt bij benadering, dat de niet-gemeenschappelijke factoren uitsluitend op de y-waarden hebben ingewerkt.

Beperken wij ons hier tot het geval, dat de hypothese een functioneel rechtlijnig verband tussen de gemeten grootheden onderstelt — dus het verband $Y = A + Bx$ — en dat de grootten van X veel nauwkeuriger werden gemeten dan die van Y. Uitgaande van de onderstelling, dat de meetfouten van Y een normale kansdistributie hebben, leert een exacte wiskundige afleiding, dat in dat geval de beste schattingen voor de constanten A en B in de regressie-vergelijking $Y = A + Bx$ zijn:

$$\boxed{b = [x_, y_,] : [x^2_,]}$$

$$\boxed{a = \bar{y} - b\bar{x}}$$

Om hen te berekenen kan dus het gemiddelde van alle x-waarden worden becijferd en het gemiddelde van alle y-waarden en dan van elke x- en van elke y-waarde het verschil met hun gemiddelde, dus $x_,$ en $y_,$. Dan zou elke $x_,$-waarde 1e in het kwadraat worden verheven — $x_,^2$ — en 2e met de bijbehorende $y_,$-waarde vermenigvuldigd —

$x_{,}$ $y_{,}$ — waarna twee optellingen de waarden zouden leren kennen van $[x_{,}^2]$ en van $[x_{,}\ y_{,}]$. Deling geeft dan de waarde van $b_{,}$ zijnde de beste schatting voor de constante B, waarna $a_{,}$ de beste schatting voor de constante A, gemakkelijk uit de tweede formule volgt. (In de praktijk worden de becijferingen van de kwadraatjessom $[x_{,}^2]$ en de productensom $[x_{,}\ y_{,}]$ bekort door de weg te volgen, die op blz. 139 e.v. is aangegeven.)

Het volgende onderzoek is een instructief voorbeeld hiervan.

In grotere zenuwcellen neemt de kern een naar verhouding kleiner deel van de ruimte in dan in kleinere zenuwcellen van eenzelfde gebied van het zenuwstelsel: bij zenuwcellen is het celvolume blijkbaar niet, zoals bij vele andere soorten van cellen in het dierlijk of menselijk lichaam, evenredig aan het kernvolume.

Om te onderzoeken, of bij zenuwcellen het celoppervlak evenredig aan het kernvolume is (bij zenuwcellen oefent juist het celoppervlak een belangrijke specifieke functie uit, n.l. de receptie van prikkels) werd van 247 dicht bij elkaar gelegen zenuwcellen van de menselijke hersenschors het volume van de kern en het volume van de cel (kern en lichaam samen) gemeten.

Uit de aard der zaak waren dit vrij onnauwkeurige metingen, zodat tamelijk grote meetfouten werden verwacht, en wel vooral bij het meten van de celvolumina, daar de cel onregelmatiger van vorm is dan de kern en het microscoop slechts een vlakke (n.l. de „optische") doorsnee goed te zien geeft.

Indien de genoemde onderstelling juist is, moeten de celvolumina evenredig zijn aan de $^3/_2$ macht van de kernvolumina, afgezien van de benadering die voortvloeit uit de onderstelling dat de cellen onderling gelijkvormige figuren zouden zijn en ook de kernen. Stellen wij het celvolume voor door C en het kernvolume door N (van nucleus), dan moet dus de onderstelling worden getoetst dat voor de diverse cellen het functionele verband $C = kN^{3/2}$ geldt (waarin k een evenredigheidsfactor voorstelt). Voor de meetgetallen zou dan de regressievergelijking zijn:

$$c = kn^{3/2}$$

Deze regressie-vergelijking is niet rechtlijnig, want de variabele n staat niet in de eerste macht. Er kan echter een rechtlijnige regressievergelijking van worden gemaakt door de logarithmen op te schrijven:

$$\log c = \log k + {}^3/_2 \log n$$

Noemen wij hierin $\log c = y$, $\log n = x$, $\log k = A$ en de te onderzoeken exponent $^3/_2 = B$, dan wordt dit

$$Y = A + Bx$$

dat wil zeggen, dan hebben wij, door andere variabelen te kiezen, onze regressie-vergelijking in een rechtlijnige veranderd en dan is het de opgaaf om na te gaan, of de „regressie-coëfficient" B inderdaad gelijk kan zijn aan $^3/_2$.

Deze „transformatie" of „herleiding" onzer regressie-vergelijking heeft in de eerste plaats het voordeel, dat zij tot een rechtlijnige leidde en van een rechtlijnige regressie-vergelijking kunnen de beste schattingen

Fig. 12.

Het kern- en celvolume van 247 zenuwcellen in de menselijke hersenschors, uitgedrukt in de getallen die bij de metingen waren verkregen.

der constanten volgens eenvoudige formules uit de gegeven meetgetallen worden becijferd. Bovendien heeft de grafische voorstelling bij rechtlijnige regressie het voordeel, dat het puntenveld „recht" loopt, dat wil zeggen een zodanige vorm heeft, dat het het trekken van een rechte regressielijn wettigt. Op het oog kan zulk een rechte regressielijn veel juister worden getrokken dan een kromme, zo juist zelfs, dat daarmee de constanten „graphisch" kunnen worden geschat. Fig. 12 is een grafische voorstelling van de gemeten cel- en kernvolumina (c en n) zelf: het puntenveld loopt krom. Het is tamelijk moeilijk te beoordelen, of de kromme regressielijn $C = kN^{3/2}$, die er in getekend is, inderdaad

bij deze punten „past". Fig. 13 is daarentegen de graphische voorstelling van hun logarithmen (log c of y en log n of x) en daarin heeft het puntenveld duidelijk een rechte vorm. Wij zouden er de op het oog best

Fig. 13.

Grafische voorstelling van de logarithmen der getallen, weergegeven in fig. 12.

passende rechte regressielijn in kunnen tekenen en dan de tangens van de stijghoek kunnen aflezen, waarmee een schatting zou zijn verkregen voor de constante B, dat wil zeggen voor de exponent onzer

oorspronkelijke regressie-vergelijking. In werkelijkheid is er de rechte lijn in getekend, die voldoet aan $y = a + 1.5\,x$ [1]) en deze blijkt op het oog reeds goed te passen: de graphische voorstelling suggereert dus reeds de juistheid van onze onderstelling, dat het celoppervlak evenredig zou zijn aan het kernvolume.

Een tweede voordeel van deze transformatie hangt samen met het feit, dat de meetfouten bij dit onderzoek des te groter zijn naarmate de gemeten volumina groter zijn. De meetfouten bestaan hier n.l. uit twee componenten: 1. de fouten, die gemaakt werden bij het meten van de optische doorsnee en het daaruit berekenen van de inhoud van het theoretische omwentelingslichaam, dat zou ontstaan als deze optische doorsnee om de langste as zou wentelen (deze inhoud vormde het getal, dat als meetgetal fungeerde) en 2. de afwijkingen tussen dit theoretische omwentelingslichaam en het werkelijke cel- of kernlichaam. De eerste zijn veel kleiner dan de tweede, zodat zij praktisch kunnen worden verwaarloosd. En de afwijkingen van de tweede soort hangen alleen van de vormen af en nemen dus ongeveer evenredig aan het gemeten volume zelf toe. De statistische formules gaan uit van de onderstelling, dat de meetfouten onafhankelijk van de gemeten afmetingen variëren, dus dat zij bij grotere afmetingen ongeveer even groot uitvallen als bij kleine en deze onderstelling geldt voor onze meetgetallen dus niet. Voor de logarithmen geldt zij echter wel, want als bij het meten van twee cellen, waarvan wij het volume C_1 en C_2 noemen, beide keren een meetfout van b.v. $+10\%$ wordt gemaakt, zodat de meetfouten in absolute waarde verschillen ($0.1\,C_1$ en $0.1\,C_2$) en de twee volumina werden gemeten als $1.1\,C_1$ en $1.1\,C_2$, dan vinden wij na logarithmiseren

$$y_1 = \log 1.1\,C_1 = \log C_1 + \log 1.1$$
$$y_2 = \log 1.1\,C_2 = \log C_2 + \log 1.1$$

zodat de meetfout bij beide cellen in de logarithmen dezelfde waarde heeft, n.l. log 1.1. In fig. 12 neemt de verticale afmeting van het puntenveld dan ook van links naar rechts sterk toe, in fig. 13 niet.

Daar de meetfouten, die bij het meten van de kernvolumina werden gemaakt, veel kleiner zijn dan die bij de celvolumina, worden de eerstgenoemde verwaarloosd, zodat de bovenvermelde formules voor de beste schattingen der constanten A en B kunnen worden toegepast. Op blz. 143 wordt voor ons vraagstuk berekend:

$$a = -0.488$$
$$b = 1.425$$

---

[1]) De hoogte van het veld is hier immers 2.25 en de breedte 1.50, zodat de tangens van de stijghoek der diagonaal is 2.25 : 1.50 = 1.5.

zodat, als de regressie rechtlijnig is, de best passende regressie-vergelijking zou luiden:

$$y_r = 1.425x - 0.488.$$

De oorspronkelijke regressie-vergelijking zou dan zijn

$$C = 0.325\ N^{1.425}$$

(want log 0.325 = —0.488)

Twee vragen dienen nu nog beantwoord te worden:

1. is de regressie van y op x werkelijk rechtlijnig?
2. kan het verschil tussen de gevonden exponent 1.425 en de theoretisch verwachte waarde $^3/_2$ of 1.5 worden toegeschreven aan de gemaakte meetfouten?

Het feit, dat de best passende waarden voor de constanten A en B in de regressie-vergelijking $Y = A + Bx$ becijferd konden worden, bewijst nog niet, dat de daaruit voortvloeiende vergelijking $y = 1.425x - 0.488$ de werkelijke regressie zou weergeven en deze dus rechtlijnig zou zijn. Op alle mogelijke meetparen kan, ook al hebben zij een regressie van een heel andere gedaante, hetzelfde rekenrecept worden toegepast, zodat dus, ook als de regressie-vergelijking $Y = A + Bx$ in het geheel niet past, toch op deze manier getallen worden gevonden voor de constanten A en B. Het zal duidelijk zijn, dat deze getallen dan geen enkele betekenis hebben. Nagegaan dient dus te worden of de gevonden vergelijking $y = 1.425x - 0.488$ in strijd moet worden geacht met de gegeven getallenparen.

Dit probleem kan tot een ons reeds bekende situatie worden herleid door het puntenveld van fig. 13 zo te vertekenen, dat de bestpassende regressielijn zuiver horizontaal komt te liggen. Dat is in fig. 14 bereikt door boven elke gegeven x-waarde niet, zoals in fig. 13, de daarbij behorende y-waarde zelf uit te zetten, maar haar „gevonden afwijking", dat is haar verschil met de beste schatting voor het centrum van haar kansdistributie. Daar de bestpassende regressielijn juist deze beste schattingen weergeeft is van een willekeurig gekozen relatiepunt in fig. 13 de hoogte boven de regressielijn gelijk aan de bedoelde „gevonden afwijking", terwijl de hoogte van de regressielijn daar ter plaatse boven de X-as gelijk is aan de beste schatting voor haar centrum. Door nu van elke y-waarde de laatstbedoelde schatting af te trekken blijven de gevonden afwijkingen over en komt de bestpassende regressielijn overal op de hoogte 0 te liggen, dat is zuiver horizontaal.

Onze vraag is dus: laten de punten in fig. 14 het trekken van een horizontale, rechte regressielijn toe? Verdelen wij de X-as nu nog in

een aantal stukken en daarmee de gegeven meetparen in klassen, dan is ons probleem omtrent het passen van de gevonden regressievergelijking veranderd in de vraag: zijn de gevonden afwijkingen in strijd met de hypothese, dat de verschillen tussen de klassen uitsluitend op het toeval berusten?

En dat is eenzelfde vraagstelling als die op blz. 55: scheiden de diverse proefpersonen evenveel phosphaat uit, anders gezegd: berusten de verschillen tussen de meetgetallen, die bij de verschillende personen (klassen) werden gevonden, uitsluitend op het toeval?

Fig. 14.

De afwijkingen van de logarithmen der celvolumina tot hun best passende rechte regressielijn ten opzichte van de logarithmen der kernvolumina.

In fig. 14 is elk klassegemiddelde door een cirkeltje weergegeven. Wel liggen zij dicht bij de X-as, maar zij verschillen toch in hoogte. *Zijn deze verschillen significant*? Dit is de vraag waar het op aankomt en daar moet de variantie-analyse antwoord op geven.

Op blz. 143 wordt de strooiïng tussen deze klassen berekend op 0.150 en die binnen de klassen op 0.127, zodat de eerste 1.18 maal zo groot is als de tweede. Bij $N_1 = 11$ en $N_2 = 235$ volgt daaruit een overschrijdingskans $P_1 = 0.2$ of veel groter dan 0.05. De meetgetallen vormen dus geen aanwijzing tegen de rechtlijnigheid der regressie.

Het zou zeer tijdrovend zijn, indien de variantie-analyse op de hier gesuggereerde wijze zou moeten worden uitgevoerd. Reeds het becijferen der 247 gevonden afwijkingen zou een omvangrijke opgaaf vormen: voor

elk der gegeven getallen-paren eerst de gegeven x-waarde invoeren in de gevonden, best passende regressievergelijking

$$y_r = a + bx$$

en dan de verkregen beste schatting y_r aftrekken van de gegeven y-waarde van datzelfde paar:

$$d = y - y_r$$

$$\text{of} \quad d = y - a - bx$$

in ons vraagstuk:

$$d = y + 0.488 - 1.425x.$$

Op blz. 143 werd daarom een weg gevolgd, waarbij dat niet nodig is. Wiskundig is n.l. te bewijzen, dat voor het onderzoek naar de toepasselijkheid van een rechtlijnige regressievergelijking de hierbedoelde strooiingen tussen de klassen (k in aantal) en binnen de klassen voldoen aan

$$\boxed{s_1^2 = ([f\bar{y}_,^2] - [x_, y_,]^2 : [x_,^2]) : N_1} \quad \text{met } N_1 = k-2$$

$$\boxed{s_2^2 = ([y_,^2] - [f\bar{y}_,^2]) : N_2} \quad \text{met } N_2 = n-k$$

In deze formules komen uitsluitend afwijkingen voor ten opzichte van de totaal-gemiddelden $\bar{x}$ en $\bar{y}$, n.l. van de x- en y-waarden zelf en van de klassegemiddelden der y-waarden ($\bar{y}$) in de klassen, die verkregen werden door de meetparen zelf in te delen volgens dezelfde grenswaarden langs de X-as, als in fig. 17 werden gebruikt om de d-waarden in klassen te verdelen.

Pas nu wij weten, dat er geen bezwaar tegen bestaat, de regressie als rechtlijnig te beschouwen, heeft het zin om nader te spreken over de grootte van de regressie-coëfficient B in de lineaire regressie-vergelijking $Y = A + Bx$. De te toetsen hypothese, volgens welke het celoppervlak dezer cellen evenredig zou zijn aan het kernvolume, voorspelt hiervoor de waarde $^3/_2$ of 1.5, uit de metingen volgt als beste schatting 1.425, zodat een afwijking is gevonden van 0.075. Pleit deze afwijking tegen de te toetsen gegeven hypothese of mag zij aan de meetfouten worden toegeschreven?

Indien weer 247 zenuwcellen uit eenzelfde gebied zouden worden gemeten, zou daar allicht, doordat de meetfouten telkens anders uitvallen, een iets andere waarde van de beste schatting voor de regressie-coëfficient uit worden becijferd: de beste schatting b is weer een variabele grootheid. Wiskundig is aangetoond, dat, indien de meetgetallen zelf een normale kansdistributie hebben, ook b een normale kansdistributie heeft. De beste schatting voor de standaarddeviatie daarvan kan dus de standaardfout van de rechtlijnige regressiecoëfficient worden genoemd. Volgens de bedoelde afleiding geldt daarvoor

$$\boxed{s_b^2 = [d^2] : [x_,^2]\, N} \quad \text{waarin } N = n-2$$

Bij een rechtlijnige regressie kost het minder rekenwerk, deze standaardfout te berekenen uit de dan geldende formule

$$\boxed{s_b^2 = ([y_,^2] : [x_,^2] - b^2) : N} \quad \text{waarin } N = n-2$$

Op blz. 143 wordt voor ons vraagstuk hiermee berekend

$$s_b = 0.041$$

zodat de afwijking 1.425 — 1.5 = —0.075 een excentriciteit heeft van

$$t = d_b : s = 0.075 : 0.041 = 1.83$$

Het t-N-P-nomogram wijst bij $t = 1.83$ en $N = 245$ (praktisch ∞) een dubbelzijdige overschrijdingskans

$$P_2 = 0.07$$

aan, dus iets groter dan de gebruikelijke grenswaarde van 0.05. De afwijking tussen de gevonden waarde van 1.425 en de theoretisch verwachte van 1.5 is dus niet significant, wel suggestief. De metingen zijn dus niet in strijd met de opvatting, dat het oppervlak van de zenuwcellen evenredig is met het volume hunner kernen, maar wel houden zij de suggestie in, dat de getoetste opvatting niet geheel juist zou zijn.

Het is instructief, even op deze suggestie in te gaan. Uit de beschreven gedachtengang blijkt, dat de uitgevoerde toets niet uitsluitend de hypothese betrof, volgens welke het celoppervlak evenredig aan het kernvolume zou zijn. Om de berekening in eenvoudige banen te leiden, werden allerlei verdere onderstellingen gemaakt en de eventuele onjuistheid, waartoe de meetgetallen een suggestie vormen, kan dus òf in de te toetsen gegeven hypothese schuilen òf in de nadere onderstellingen.

Bij deze onderstellingen is er één, die wel zeker onjuist zal zijn en daarvan kan bovendien worden vastgesteld, dat zij de schatting voor de regressie-coëfficient te laag heeft doen uitvallen: het is de benaderende onderstelling, dat bij het meten van de kernen geen meetfouten zouden zijn gemaakt. Natuurlijk zijn daarbij wèl meetfouten gemaakt, alleen is het aannemelijk, dat zij kleiner waren dan die bij de cellen. Waren beide categorieën van meetfouten even groot geweest, dan zou voor het berekenen van de beste schatting voor de regressie-coëfficient een formule moeten zijn gebruikt, die geldig is voor situaties in welke de individuele waarden van *beide* variabele grootheden mee door niet-gemeenschappelijke oorzaken beïnvloed waren en zulke formules leveren hogere waarden voor de regressie-coëfficient. Wij

bespreken die formules hier niet nader, noch de keus, die eruit moet worden gemaakt. Eén er van[1]) geeft als beste schatting $b = 1.62 \pm 0.04$, dus groter dan 1.5.

De waarde 1.5, die voor deze coëfficient te verwachten is als het celoppervlak evenredig is aan het kernvolume, ligt tussen de twee berekende schattingen 1.425 en 1.62 in. De schatting 1.425 zou gelden als alleen bij het meten der celvolumina meetfouten waren opgetreden en de schatting 1.62 als bij de cellen en de kernen even grote meetfouten waren gemaakt. De meetmethode was zodanig, dat zowel bij de cellen als bij de kernen meetfouten moesten optreden, bij de kernen alleen veel kleinere dan bij de cellen. Het is in fraaie overeenstemming daarmee, dat de theoretisch verwachte waarde tussen de twee genoemde schattingen in ligt en het dichtst bij die schatting, die zou gelden als bij de kernen geen meetfouten waren gemaakt.

De suggestie, die de meetgetallen inhouden in de zin van een kleine fout in de getoetste situatie, kan dus met goede grond worden opgevat als een suggestie tegen de volledige geldigheid van een der bijkomstige onderstellingen, waarvan de onjuistheid trouwens praktisch vast staat. Alles samen kan dus worden geconcludeerd, dat de metingen in overeenstemming zijn gebleken met de opvatting, dat het celoppervlak evenredig is aan het kernvolume.

Alleen: de laatste beschouwingen, die de suggestie tegen deze opvatting te niet deden, waren niet meer exact. Wij raken hier aan de betrekkelijkheid van het statistisch onderzoek: lang niet alle theorieën kunnen tot in alle finesses volgens de exacte methoden der statistica aan gegeven meet- of telgetallen worden getoetst. Ook deze facet der statistica is het zeker waard, goed bewust onder ogen te worden gezien.

Tenslotte dient een andere onderstelling, die in de bovenstaande beschouwingen was ingevoerd, nog met de gegevens te worden geconfronteerd, n.l. de onderstelling dat de afwijkingen ten opzichte van de gevonden regressie een normale kansdistributie hebben. Daartoe zijn de frequenties van de gevonden d-waarden in fig. 15 grafisch weergegeven en is door hun tredenlijn de best passende normale kanslijn getekend. Deze figuur brengt geen grove verschillen met de normale kansdistributie aan het licht. Doordat de gebruikte methoden alleen voor grove afwijkingen ten opzichte van de normale kansdistributie gevoelig zijn (zodat zij b.v. niet kunnen worden toegepast bij een opvallend scheve of bij een typisch tweetoppige

---

[1]) n.l. $b = tg\, \alpha$ als $tg\, 2\, \alpha = 2\, [x_, y_,] : ([x_,^2] - [y_,^2])$, waarbij α de stijghoek is van de (langste) zwaartelijn van het puntenveld.

kansdistributie der niet-nader-geanalyseerde factoren) maar bij matige afwijkingen der normale kansdistributie nog zeer behoorlijke benaderingen leveren, daarom is in dit geval onze onderstelling van een normale kansdistributie niet zo ver naast de waarheid, dat de conclusie er door in gevaar zou zijn gebracht.

De behandeling van dit voorbeeld van rechtlijnige regressie worde niet afgesloten zonder dat er nog eens nadrukkelijk op gewezen wordt, dat een beste schatting van een constante in een regressie-vergelijking alleen dàn zin heeft, als deze regressie-vergelijking zelf in de gegeven situatie zin heeft, dus pas nadat gebleken is dat de gevonden best

Fig. 15.
De frequenties der afwijkingen, voorgesteld in fig. 14, met de best passende normale kanskromme.

passende regressie-vergelijking niet in strijd behoeft te worden geacht met de meetgetallen, waar zij uit werden afgeleid. Vaak wordt volstaan met het berekenen van de beste schatting voor een constante en het toetsen van de geopperde hypothese aan deze beste schatting. Verzuimd wordt dan na te gaan of de regressie-vergelijking, tot welke deze constante behoort, in strijd met de gegeven meetwaarden is. Daar de gevonden beste schatting slechts zin heeft als de regressie-vergelijking, waarin zij voorkomt, zin heeft, is dit een niet te vergeven nalatigheid. Dat deze nalatigheid toch zo vaak begaan wordt, is verklaarbaar uit het feit, dat er tot voor kort geen methode bestond om over de toelaatbaarheid van een regressie-vergelijking te oordelen. Nu zulk een methode gevonden is — en de aangegeven methode geldt

voor alle vormen van regressie — nu mag worden geëist dat geen oordeel over een uit de meetgetallen af te leiden constante wordt meegedeeld zonder dat er uitdrukkelijk bij wordt vermeld, dat de gebruikte regressie-vergelijking niet in strijd met die meetgetallen is gebleken en de bedoelde constante dus zin heeft.

DE STATISTISCHE BEWERKING VAN HISTORISCHE GEGEVENS

Als de lezer voor de taak zou worden gesteld, een conclusie te trekken uit z.g. historische gegevens, dan zou hij allicht een merkwaardige moeilijkheid gaan gevoelen. Doordat hij al het voorafgaande zorgvuldig heeft gelezen en doordacht, heeft hij zich de gedachtengang van de statistica enigszins eigen gemaakt en is hij er goed van doordrongen dat de statistica alleen dan kan helpen als een conclusie moet worden getrokken uit onnauwkeurige getallen. Getallen omtrent historische gebeurtenissen zoals de fluctuatie in de prijs van een bepaald artikel, in de grootte van de oogst, in het aantal gevallen per jaar van een bepaalde ziekte of dergelijke, zijn echter veelal met zodanige zorg vastgesteld, dat de daarbij gemaakte fouten te verwaarlozen klein zijn. Voor praktische doeleinden kunnen zij dus als *nauwkeurige getallen* worden opgevat. Meetfouten spelen dan geen rol meer en ook een steekproef is daarbij niet genomen, want alle desbetreffende gevallen kunnen praktisch worden geacht in het onderzoek te zijn opgenomen. Waar is dan plaats voor de statistica, die immers alleen kan helpen als de gegeven getallen onnauwkeurig zijn? Dat is de vraag, die bij bedoelde lezer kan opkomen en die hem dan voor een onoplosbaar lijkende moeilijkheid schijnt te plaatsen.

Zo gauw wij de situatie aan de hand van een voorbeeld wat nauwkeuriger gaan bekijken, zal het blijken, dat de nauwkeurigheid van deze getallen gelukkig slechts in schijn en geenszins in werkelijkheid een moeilijkheid vormt.

Tijdens de tweede wereldoorlog nam de sterfte aan tuberculose in Nederland sterk toe. Bij een daaromtrent ingesteld onderzoek werd onder meer de vraag gesteld, of de leeftijd van invloed op die stijging is geweest en, zo ja, welke invloed de leeftijd daarbij had. De gegevens, waaruit een antwoord moest worden afgeleid, waren de getallen, verkregen uit de officiële aangiften der artsen omtrent de sterfgevallen. Deze aangiften vermeldden o.a. de gemeente, de datum, de leeftijd en de diagnose. Wel zal bij deze opgaven een enkele keer een vergissing zijn gemaakt, b.v. ten opzichte van de diagnose, maar dat zal op de zeer talrijke opgaven toch slechts bij hoge uitzondering het geval zijn

geweest en daarom beschouwen wij deze getallen praktisch als volkomen juist. Dus: geen meetfouten, geen steekproef, maar nauwkeurige getallen.

Beperken wij ons hier voor de overzichtelijkheid tot de 8 grote steden (n.l. die met meer dan 100.000 inwoners). In elk dier steden werd de bevolking volgens de leeftijd in 11 groepen gesplitst, n.l. in de leeftijdsklassen 0—5, 5—10, 10—15, 15—20, 20—25, 25—30, 30—40, 40—50, 50—60, 60—70 en 70—80 jaar (de enkele nog oudere personen blijven hier, ter vereenvoudiging van de beschrijving, buiten beschouwing), zoodat in totaal 88 groepen van inwoners ontstonden. In elke groep afzonderlijk werd het aantal sterfgevallen aan tuberculose in de periode van de eerste drie oorlogsjaren gedeeld door dat aantal in de periode van drie jaar die aan de oorlog voorafging. Het zo verkregen verhoudingsgetal gaf de stijging van de tuberculosesterfte in die bevolkingsgroep weer. Zo werd b.v. voor de 30—40-jarigen in Amsterdam het verhoudingsgetal 1.36 gevonden: in deze groep stierven in de oorlogsjaren 1.36 maal zo veel personen aan tuberculose als voor de oorlog. De 88 groepen leverden op die manier 88 zulke verhoudingsgetallen en deze vormden het uitgangspunt voor de eigenlijke statistische analyse.

Fig. 16 brengt de 11 getallen van Amsterdam in beeld. De lijn stijgt naar rechts, dat is naar de hogere leeftijden, en vertoont bovendien een extra verheffing bij de leeftijden tussen 10 en 20 jaar.

De Amsterdamse gegevens wekken dus de indruk

1. dat de stijging bij toenemende leeftijd sterker wordt en
2. dat de leeftijden tussen 10 en 20 jaar de stijging extra bevorderen.

Gevraagd wordt, deze indruk volgens een exacte statistische methode aan de getallen van alle 8 steden te toetsen.

De 8 lijnen van deze steden zijn in fig 17 samen weergegeven. Wij zien, dat ze geenszins eenzelfde beeld vertonen. Als de verschillen tussen de 11 getallen van één stad uitsluitend door de verschillen in leeftijd waren veroorzaakt, dan zouden deze 8 lijnen eenzelfde vorm hebben gehad, want de leeftijd zal toch wel in die verschillende steden eenzelfde invloed op de stijging van de tuberculose hebben uitgeoefend. Er moeten dus naast de leeftijd nog andere factoren invloed op die stijging hebben gehad. Dat kunnen er vele zijn geweest: verschillen in welstand, voeding, woningtoestand, infectiekansen, de erfelijkheid, de medische voorgeschiedenis, de psychische factoren als zorg en angst en nog vele andere bekende en onbekende factoren meer. Alleen als al die factoren over alle 88 bevolkingsgroepen zodanig waren verdeeld, dat zij in elke groep nauwkeurig dezelfde invloed op de tuberculose zouden hebben gehad, dan zouden de verschillen tussen de 88 getallen

uitsluitend door de verschillen in leeftijd veroorzaakt zijn geweest. Maar het zou al heel toevallig en zelfs vrijwel ondenkbaar zijn, als al die factoren zoals de welstand, de voeding, de zorg enz. b.v. bij de 10—15-jarigen te Amsterdam precies dezelfde scala zouden hebben vertoond als b.v. bij de 30—40-jarigen te Groningen. Allicht zullen zij in de 8 verschillende steden enigszins verschillend verdeeld zijn geweest en het is daaraan, dat wij de verschillen kunnen toeschrijven tussen

Fig. 16.

De relatieve toeneming van de sterfte aan tuberculose bij personen van uiteenlopende leeftijden te Amsterdam.

de 8 getallen, die in die 8 steden voor de stijging der tuberculosesterfte in éénzelfde leeftijdsgroep — b.v. in de leeftijdsklasse van 15—20 jaar — waren gevonden.

De verschillen tussen de 88 getallen zijn dus niet alleen door de leeftijdsverschillen veroorzaakt, maar ook door allerlei andere factoren. Dientengevolge geven zij de invloed van de leeftijd op de stijging van de tuberculosesterfte niet zuiver weer en het was toch alleen de invloed van de leeftijd, die wij wensten te bestuderen en waarvoor wij deze 88 getallen als inlichting bedoelden. Nu worden wij ons

bewust, dat onze getallen *datgene, wat wij er mee bedoelen,* slechts onnauwkeurig weergeven; daardoor zijn het voor ons speciale doel, volgens de definitie die op blz. 6 werd gegeven, in de gewone zin van de statistica onnauwkeurige getallen.

En daarmee is de moeilijkheid, die wij meenden te bespeuren, vervallen:

in zo verre de 88 getallen de stijging van de tuberculosesterfte in

Fig. 17.
De relatieve toeneming van de sterfte aan tuberculose in de 8 grootste steden van Nederland.

elk van de 88 bevolkingsgroepen weergeven, mogen ze worden beschouwd als nauwkeurige getallen,

zodra wij hen beschouwen als een inlichting over de invloed van de leeftijd op die sterftestijging worden diezelfde 88 getallen onnauwkeurige getallen.

Het gaat er mee als met een getal, dat verkregen wordt door het tellen van de individuen met een bepaald kenmerk in een steekproef: dat getal is nauwkeurig in zoverre het die steekproef weergeeft — ook hier met de beperkende voorwaarde dat er geen vergissingen bij het

tellen zijn gemaakt — maar hetzelfde getal is een onnauwkeurig getal zo gauw het wordt beschouwd als een inlichting omtrent de gehele verzameling, waar die steekproef uit genomen was.

Ons vraagstuk wordt dan ook op dezelfde manier als andere statistische vraagstukken behandeld.

Daar wij de verschillen tussen de getallen, die telkens voor éénzelfde leeftijdsklasse in de 8 verschillende steden werden gevonden, toe-

Fig. 18.
De relatiepunten van fig. 17 met hun best passende rechte regressielijn ten opzichte van de leeftijd.

schrijven aan de niet nader te analyseren factoren zoals voeding, woningtoestand, erfelijkheid, angst, zorg e.d., interesseert het ons niet meer welke getallen van de diverse leeftijdsklassen in éénzelfde stad waren verkregen. Uit fig. 17 kunnen we dus de 8 lijnen, die elk voor één stad de 11 punten dier stad met elkaar verbinden, gevoegelijk weglaten en zo ontstaat fig. 18, waarin alleen de 88 „meetpunten" zijn opgenomen. En omdat wij, althans om ons te oriënteren, graag volgens een bekend rekenschema willen werken, gaan wij na of er een lineaire regressie bestaat tussen de stijging der tuberculosesterfte enerzijds en

de leeftijd anderzijds. Dit schema geldt, zoals op blz. 82 e.v. uiteengezet werd, alleen als

1. de gevolgen van de niet-nader-te-analyseren factoren een normale kansdistributie hebben,
2. de diverse getallen met eenzelfde nauwkeurigheid zijn vastgesteld,
3. de leeftijd in alle 8 steden eenzelfde invloed op de stijging der tuberculose heeft gehad.

Wij passen bedoeld rekenschema daarom toe in de voorlopige onderstelling, dat deze 3 voorwaarden vervuld zijn. Daarna kan dan nader worden bekeken, in hoeverre die voorwaarden misschien niet vervuld zijn en welke invloed dat op onze conclusies kan hebben uitgeoefend.

Volgens het schema, dat op blz. 98 e.v. aan een ander voorbeeld is behandeld, werd uitgerekend

a. dat de 88 gegeven getallen geen bezwaar opleveren tegen de gedachte van een lineaire regressie (de desbetreffende overschrijdingskans bleek n.l. 0.15 te bedragen, dus ruim groter dan de gebruikelijke grenswaarde van 0.05)

b. dat de beste schatting voor de regressiecoëfficient 0.0076 bedraagt (voor de regressievergelijking $IJ = A + Bx$, waarbij ij een gegeven sterfteverhoudingsgetal representeert en x de gemiddelde leeftijd van de leeftijdsklasse in welke dat getal was gevonden, werd n.l. becijferd $b = 0.0076 \pm 0.0022$)

c. dat deze regressiecoëfficiënt significant van 0 verschilt (de excentriciteit bedraagt $t = 0.0076 : 0.0022 = 3.47$, waarvoor de overschrijdingskans kleiner is dan 0.001)

d. dat de beste schatting voor de theoretische sterftestijging bij 0 jaar een te verwaarlozen klein bedrag is (n.l. 7 ± 5 %, dus niet significant van 0 verschillend : de berekening leerde $a = 1.07 \pm 0.05$).

Deze vier conclusies samenvattend kunnen wij zeggen, dat de 88 gegeven verhoudingsgetallen in strijd zijn met de onderstelling dat de leeftijd geen invloed op de stijging der tuberculosesterfte zou hebben gehad, en dat zij als beste schatting aangeven dat die stijging voor elke leeftijd zou bedragen $^3/_4$ % van die leeftijd in jaren (voor een 40-jarige dus $^3/_4 \times 40$ of 30%). Om de nauwkeurigheid van die beste schatting er bij aan te geven kan worden geschreven $^3/_4 \pm {}^1/_4$ % : voor een 40-jarige dus 30 ± 10 %, of, anders uitgedrukt, voor de 40-jarige Nederlandse ingezetene zou de verhouding tussen de sterftekans in de drie eerste oorlogsjaren tot die in de daaraan voorafgaande periode van drie jaren gemiddeld 1.3 ± 0.1 zijn geweest. Daar er geen significante afwijkingen ten opzichte van de rechtlijnige regressie

bleken te bestaan, vormen de gegeven getallen geen voldoende aanwijzing om aan bepaalde leeftijden een extra versterkende invloed op de tuberculosesterfte toe te schrijven.

De eerste hypothese (n.l. dat de sterftestijging met de leeftijd toeneemt) wordt dus door de gegevens zeer betrouwbaar bevestigd, voor de tweede hypothese (n.l. dat de leeftijd van 10 tot 20 jaar de tuberculosestijging extra zou hebben versterkt) bleken de gegevens geen voldoende steun te bieden.

Fig. 19.

De frequenties van de afwijkingen der relatiepunten in fig. 18 tot de daar getekende regressielijn met de best passende normale kanslijn.

Rest nog te bespreken de invloed van de drie nadere onderstellingen, die nodig waren om het bekende rekenschema der lineaire regressie te kunnen toepassen.

De onderstelling, dat de gevolgen van de niet-nader-te-analyseren factoren een normale kansdistributie hadden, kan exact worden getoetst, en wel door de distributie van de *d*-waarden te bestuderen. Dit zijn de verschillen tussen de 88 gegeven getallen en de waarde, die de gevonden best passende regressie-vergelijking voor elk van die getallen aangeeft, anders gezegd de hoogte van elk punt van fig. 18 boven of onder de daarin getekende best passende rechte regressielijn. Die distributie is in fig. 19 grafisch weergegeven door een tredenlijn;

deze verschilt zo weinig van de eveneens getekende best passende normale kanskromme, dat wij de getrokken conclusies niet op deze grond in twijfel behoeven te trekken.

Intussen kunnen wij wel opmerken, dat de tredenlijn een, zij het klein, toch systematisch verschil van de normale kanskromme vertoont: de tredenlijn is duidelijk scheef met de top links van het midden en met een te lange rechter voet. Nadere analyse leert, dat de tredenlijn niet significant verschilt van een proportionele kanslijn. Hadden wij, in plaats van de verhoudingsgetallen zelf, hun logarithmen voor dit onderzoek gebruikt, dan hadden wij dus een goede overeenstemming gevonden met de normale kansdistributie der *d*-waarden. Theoretisch kan trouwens worden aangetoond, dat het voor het statistisch onderzoeken van verhoudingsgetallen correcter is, dat onderzoek niet aan die getallen zelf maar aan hun logarithmen uit te voeren. Ter wille van de lezer werd hier echter de meest voor de hand liggende manier gevolgd en werden de verhoudingsgetallen zelf voor het onderzoek gebruikt.

De twee andere onderstellingen zijn vrijwel zeker enigszins buiten de waarheid. Het onderzoek is daardoor minder nauwkeurig geweest, maar daar het in deze minder nauwkeurige vorm reeds een positief resultaat heeft opgeleverd is het niet nodig, een iets nauwkeuriger beeld van de leeftijdsinvloed te ontwerpen via andere onderstellingen, die zeer ingewikkelde, misschien zelfs in de praktijk onmogelijke berekeningen zouden vergen.

De drie gemaakte onderstellingen hebben dus geen afbreuk gedaan aan ons resultaat, dat de leeftijd wel degelijk invloed op de stijging der tuberculosesterfte heeft gehad en wel een zodanige, dat deze stijging ten naaste bij $^3/_4$ % van de leeftijd in jaren bedraagt[1]).

De statistische bewerking van dit vraagstuk werd hier intussen niet besproken ter wille van deze uitkomst, maar alleen om aan een voorbeeld te demonstreren hoe getallen, die historische gegevens praktisch nauwkeurig weergeven, voor een bepaalde vraagstelling toch als onnauwkeurige getallen moeten gelden en hoe zij volgens normale methoden exact statistisch kunnen worden geanalyseerd.

---

[1]) De praktische betekenis van dit resultaat wordt verzwakt door het feit, dat tijdens dit oriënterende onderzoek nog geen gegevens beschikbaar waren omtrent de veranderingen in het aantal inwoners van de beschouwde steden. Het is echter onwaarschijnlijk, dat de daaromtrent nog te verwachten gegevens het resultaat teniet zouden doen: indien in de diverse leeftijdsgroepen van de verschillende steden het aantal inwoners zou gestegen zijn met $^3/_4$ % van de gemiddelde leeftijd in die groep (uitgedrukt in jaren), dan zouden de gevonden verhoudingsgetallen geen aanwijzing voor een invloed van de leeftijd bevatten en de inwoners-aantallen hebben zich waarschijnlijk juist in een andere richting bewogen.

DE AARD VAN DE STATISTISCHE CONCLUSIE

Om een algemene schouw te krijgen omtrent de statistica en de door die wetenschap mogelijk gemaakte practische werkmethoden der statistiek komen wij terug tot ons uitgangspunt: de statistica wordt daar te hulp geroepen waar het geldt, een mening omtrent een werkelijkheid te vormen op grond van onnauwkeurige getallen, die omtrent die werkelijkheid waren verkregen. Met onnauwkeurige getallen werden daarbij bedoeld getallen, die door meting waren verkregen of door telling in een steekproef; zulke getallen geven immers datgene, wat er mee bedoeld werd, slechts onnauwkeurig, n.l. slechts bij benadering, weer.

En verder sluiten wij onze beschouwingen aan aan de opmerking, die in de vorige paragraaf werd gemaakt over de rol, die deze meet- en telgetallen bij het onderzoek naar een regressie spelen. Deze opmerking is n.l. niet slechts op de regressie van toepassing, zij geldt voor het gehele veld van de statistica:

1. meet- en telgetallen kunnen een *suggestie* vormen voor het ontwerpen van een hypothese over de gemeten of getelde werkelijkheid,
2. zij kunnen helpen zulk een geopperde hypothese quantitatief te *preciseren*,
3. zij kunnen helpen die hypothese te *kritiseren*.

Voor het verkrijgen van een goed inzicht in de aard van de statistische conclusie is het nuttig, deze drie punten nader te bekijken.

Doordat meet- en telgetallen het bedoelde slechts bij benadering weergeven, wijzen zij niet met volledige scherpte een bepaalde opvatting over de bedoelde facet van de werkelijkheid als de enig juiste interpretatie aan, maar laten zij ruimte voor allerlei hypothesen over het gemeten of getelde object. Uit het bekende gezegde, dat maat en getal betrouwbare gidsen voor ons denken vormen, spreekt wel een hoopvol enthousiasme, maar bij nadere kennismaking blijkt de taal, die de getallen spreken, toch steeds weer aan duidelijkheid te wensen over te laten, zodat hun als gids steeds een zekere mate van vaagheid aankleeft.

Tengevolge van deze vaagheid berusten onze opvattingen omtrent de werkelijkheid nooit uitsluitend op meet- of telgetallen, maar heeft bij het concipiëren dier opvattingen steeds een andere factor krachtdadig meegeholpen, n.l. de scheppende verbeeldingskracht van den onderzoeker. Zonder deze fantasie zou geen enkele theorie omtrent de ons omringende werkelijkheid bestaan, hoeveel meetgetallen ons ook uit indrukwekkende tabellen zouden aanstaren. Terwijl deze

fantasie dus zelfs onmisbaar is voor het ontstaan van een theorie over de waargenomen werkelijkheid, kan dat zelfde niet eens van de meet- en telgetallen worden gezegd: de fantasie is in staat, ook zonder die getallen zulke theorieën op te bouwen. Talrijk toch zijn de opvattingen, die niet op grond van metingen of tellingen zijn opgekomen maar die op andere overwegingen berusten, b.v. op een vermoed parallelisme met de gang van zaken in een verwant gebied.

De grote rol, die de verbeeldingskracht van den onderzoeker vervult bij het opstellen van een theorie over de werkelijkheid, maakt, dat zulk een theorie, althans in het begin, uitsluitend de betekenis heeft van een hypothetisch vermoeden, dus eigenlijk de vorm heeft van de vraag: „zou de werkelijkheid misschien zus of zo zijn?"

Deze rol van de fantasie maakt verder, dat geen vaste recepten kunnen worden opgesteld om uit meet- of telgetallen theorieën omtrent het gemetene of getelde af te leiden: theorieën worden niet uit zulke getallen afgeleid alsof het een rekensom is, waarop maar één antwoord past. Bij het ontwerpen van zulk een theorie spelen de meet- en telgetallen dan ook slechts wat wij zouden kunnen noemen een suggestieve rol.

Al zijn er voor deze suggestieve werking der meet- of telgetallen uit de aard der zaak geen recepten te geven, wel kan hun suggestieve kracht vaak sterk worden verhoogd door hen op een handige manier graphisch voor te stellen. Een statistisch onderzoek wordt dan ook veelal begonnen met het oriënterend tekenen van een graphische voorstelling waarin òf wel de gegeven getallen zelf òf bepaalde daarvan afgeleide grootheden — machten, logarithmen o.d. — zijn uitgezet. Dat ook hiervoor weer geen vaste regelen zijn aan te geven zal na het voorafgegane wel geen nader betoog behoeven.

Tot zover het eerste punt: de meet- en telgetallen kunnen een suggestie vormen voor het ontwerpen van een hypothese omtrent de gemeten of getelde werkelijkheid. Deze suggestieve werking — die op zichzelf niet onmisbaar is — en die door graphische voorstellingen sprekender kan worden gemaakt, behoeft steeds als onmisbare aanvulling de hulp van de fantasie van den onderzoeker. De geopperde hypothese heeft dientengevolge in beginsel de betekenis van de onderstellenderwijs uitgesproken vraag: „zou de werkelijkheid misschien . . . .", waarbij in te vullen de door die theorie geformuleerde opvatting.

Als tweede punt, waar meet- en telgetallen een rol kunnen spelen, werd genoemd het quantitatief preciseren van de geopperde hypothese. Wat is dat en waartoe dient dat?

Stel de mening is geopperd, dat de kans op het krijgen van malaria evenredig zou zijn aan het aantal muggen in de huiskamer. En stel

dat in een bepaalde periode daarom in verscheidene huizen de malariagevallen en de muggen waren geteld.

Bij het toetsen van deze hypothese staan wij voor een moeilijkheid. Als er zulk een evenredigheid zou bestaan dan was er ook een bepaalde evenredigheids- of verhoudings*factor*: de kans op malaria was dan in alle huizen b.v. $^1/_{1000}$ van het (gemiddeld) aantal muggen in de huiskamer. De hypothese, zoals deze hierboven geformuleerd was, noemt de getallenwaarde van die evenredigheidsfactor echter niet, zij stelt alleen, dat er „een" evenredigheid — tussen de malariakans en bedoeld aantal muggen — bestaat. En toch zouden wij, om die hypothese te toetsen, de grootte van die factor graag kennen, want pas dan zouden wij bij elk gevonden aantal muggen kunnen uitrekenen, hoe groot volgens die hypothese het aantal malariagevallen in dat huis zou moeten zijn, waarna wij de overschrijdingskans zouden hebben willen berekenen van de afwijking tussen deze „theoretisch verwachte" waarden en de door de tellingen „gevonden" waarden. Zonder de evenredigheidsfactor kunnen wij de theoretisch verwachte waarden niet bepalen en dus ook niet de overschrijdingskans vinden, m.a.w. de geopperde hypothese is nog te vaag geformuleerd om aan de gegeven getallen te kunnen worden getoetst.

In zo'n geval vragen wij nu van de gegeven meet- of telgetallen een nadere precisering: onderstel voorlopig, dat de bedoelde malariatheorie juist zou zijn en er dus een evenredigheid en dus ook een evenredigheidsfactor bestaat, welke is dan, gezien de informatie, die onze telgetallen omtrent de werkelijkheid geven, de beste schatting van die evenredigheidsfactor? Laat die beste schatting blijken $^1/_{1100}$ te zijn, dan toetsen wij niet de oorspronkelijke hypothese aan de gegeven telgetallen, maar een nieuwe, die wij zelf maken, n.l. de hypothese, dat de kans op malaria $^1/_{1100}$ zou bedragen van het gemiddeld aantal muggen in de huiskamers. Blijkt die speciale hypothese — die maar één speciaal geval is van onze algemene malaria-hypothese — niet met die getallen in strijd te zijn, dan is ook de oorspronkelijke algemene hypothese er niet mee in strijd: als de kans op malaria $^1/_{1100}$ van het aantal muggen bedraagt, dan bestaat er „een" evenredigheid tussen beide.

Dit malariavoorbeeld betreft een hypothese, volgens welke een evenredigheid (tussen bepaalde grootheden) zou bestaan, maar waarbij de getalswaarde van de evenredigheidsfactor in het midden werd gelaten. Zo worden vele hypothesen geopperd en ter toetsing voorgelegd, die de getalswaarde van allerlei constanten in het midden laten. Het probleem van het kernvolume en de celoppervlakte, dat op blz. 83 e.v. behandeld werd, is er een van. Om deze te vage theorieën toch te

toetsen, worden zij eerst nader quantitatief gepreciseerd door hun nog onbepaald gelaten constanten door de beste schattingen te vervangen.

Het kan in twijfel worden getrokken of het verantwoord is, de algemeen geformuleerde hypothese voor deze toets te vervangen speciaal door die met de beste schatting voor de constante. Als de te toetsen hypothese de waarde van een bepaalde constante in het midden laat, dan zouden wij eerst een willekeurig getal kunnen kiezen en dan die hypothese kunnen toetsen, waarbij de constante die willekeurig gekozen waarde heeft. Zulks zou met vele willekeurig gekozen waarden kunnen worden gedaan en zodra ons zou blijken, dat bij een of meer van die waarden der constante onze hypothese een voldoend grote overschrijdingskans oplevert, dan zouden wij weten, dat ook de algemener gestelde oorspronkelijke hypothese niet in strijd met de feiten moet worden geacht. En dat onderzoek zou in beginsel kunnen worden bekort door eerst af te leiden bij welke waarde van de constante de overschrijdingskans van onze hypothese het grootst is. Wij zouden dan alleen de hypothese met deze waarde van de constante behoeven te toetsen.

Wij laten het mogelijke verschil tussen deze waarde van de constante en de beste schatting — welk verschil meestal niet groot en vaak 0 is — hier echter onbesproken.

Ook bij het bepalen van de best passende waarden van constanten, die door de te toetsen hypothese zelf nog niet waren gepreciseerd, kan de grafische voorstelling vaak diensten bewijzen, vooral indien niet de maximale nauwkeurigheid wordt verlangd. In ons malariavoorbeeld ging het er om, uit de gegeven telgetallen de best passende waarde af te leiden voor de z.g. regressie-coëfficient B in de lineaire regressieformule $Y = Bx$. Daarvoor bestaat een bepaald rekenrecept, (zie blz. 140) dat vrij tijdrovend is, maar deze waarde kan ook worden afgelezen uit een grafische voorstelling, waar alle x-waarden (aantallen muggen) horizontaal en alle y-waarden (aantallen malariagevallen) verticaal zijn uitgezet, zodat in dit diagram elk huis door een punt wordt weergegeven. Waren alle x- en y-waarden nauwkeurig evenredig aan elkaar (m.a.w. was voor elk huis de verhouding $y : x$ dezelfde) dan lagen al deze relatiepunten op een rechte lijn, die tevens door het nulpunt ging (het snijpunt van de X- en de Y-as) n.l. op de rechte lijn, waarvan de stijghoek een tangens heeft gelijk aan die algemeen geldende verhouding $y : x$. Trekken wij dus in ons diagram die rechte lijn, die door het nulpunt gaat en „op het oog" het best bij de gevonden relatiepunten past, dan zal de tangens van zijn stijghoek ten naastenbij overeenkomen met de beste schatting voor de bedoelde regressiecoëfficient. En zo bestaat voor vrijwel elke soort van constante wel een middel, om de best passende waarde ervan langs grafische weg te bepalen. Voornamelijk bij oriënterend onderzoek is deze weg aan te bevelen, omdat er veel tijd mee bespaard wordt en uiteindelijk kan

toch, als dat voor het eindresultaat nuttig is, deze grafische manier weer door de cijfermethode worden vervangen.

Bleek bij het eerste punt — de betekenis van de meetgetallen als een suggestie voor het concipiëren van een hypothese omtrent de gemeten werkelijkheid — de fantasie van den onderzoeker het onontbeerlijke instrument te zijn, bij het tweede punt — de nadere quantitatieve precisering van de geopperde hypothese — vormt de beste schatting het aangewezen werktuig. En bij beide punten kan de grafische voorstelling nuttige diensten bewijzen.

De derde en laatste manier, waarop de meet- en telgetallen kunnen helpen bij het verbeteren van ons inzicht in de ons omringende werkelijkheid, is, dat zij kritiek kunnen leveren op een geopperd vermoeden: een hypothese, onderstellenderwijs omtrent de gemeten of getelde werkelijkheid opgeworpen, kan aan de meet- of telgetallen worden getoetst. Het is ter bevordering van dit derde punt, dat de theoretische wetenschap der statistiek, de statistica, is opgebloeid, dat complex van gedachten en van daarop steunende rekenvoorschriften, waarvan dit boekske de gedachtengang in korte trekken aanduidde. En uit het tot nu toe behandelde is het reeds duidelijk, dat voor dit derde punt de overschrijdingskans het wapen vormt, en wel in beginsel uiteindelijk het enige wapen. [1]). Hier is geen plaats voor de fantasie van den onderzoeker, voor vrije keus uit vele antwoorden of voor suggestieve hulp van de kant der grafische voorstelling: hier past op de vraag maar één antwoord, eventuele andere antwoorden zouden fout zijn. Welk karakter heeft dat enig-juiste antwoord?

In eerste instantie zouden wij geneigd zijn te zeggen dat het een antwoord zou moeten zijn op de vraag, of de te toetsen gegeven hypothese juist is. De statistica wijst echter geen weg aan om die vraag met een simpel en exclusief „ja" of „neen" te beantwoorden. Dat is een gevolg van de onnauwkeurigheid onzer meet- of telgetallen. Indien het de schijn zou hebben, dat het antwoord zou moeten luiden: „ja, deze hypothese is juist", dan zou, door de onnauwkeurigheid van de

---

[1]) Vele statistische werkers, die in hun praktijk steeds met grote aantallen meetgetallen werken, zijn zich daar praktisch niet van bewust. Als zij een grote correlatie of een grote regressie-coëfficient uit hun meetgetallen becijferd hebben, dan nemen zij in den regel zonder meer aan, dat er in werkelijkheid een correlatie of zulk een regressie bestaat. Zij kunnen dat doen doordat, als het aantal meetgetallen heel groot is, de standaardfouten al gauw zo klein worden dat praktisch geen bezwaar tegen die hypothese bestaat. Maar in beginsel berust zulks ook dan, evenals bij het vaststellen van elke correlatie of regressie of welke andere statistische conclusie dan ook, op het feit, dat de overschrijdingskans klein genoeg is, al wordt dat dan in de praktijk vaak slechts stilzwijgend en zonder daarop gerichte becijfering aangenomen.

meetgetallen, mogelijk toch nog een klein verschil kunnen bestaan tussen de hypothese en de werkelijkheid, n.l. een zo klein verschil, dat de gebruikte meet- of telmethode niet nauwkeurig genoeg was om het aan te tonen. En als het de schijn zou hebben, dat het antwoord zou moeten luiden: ,,neen, die hypothese is fout", dan zou theoretisch nog altijd gewezen kunnen worden op de mogelijkheid, dat de hypothese de werkelijkheid zelf wèl goed zou hebben weergegeven, maar dat bij de meting of telling een fout zou zijn opgetreden, waardoor alleen het daarbij gevonden meet- of telgetal niet met de hypothese klopt. Vandaar dat het antwoord, waartoe de statistica ons in staat stelt, nooit een zekerheid kan uitdrukken.

Wij zagen reeds herhaaldelijk, welke weg de statistica hier inslaat: door de onnauwkeurigheid van de meet- of telgetallen zal er allicht een verschil bestaan tussen de uit die getallen ,,gevonden waarde" en de ,,theoretisch verwachte waarde", dat is de waarde, die dezelfde werkelijkheid zou hebben als de te toetsen hypothese volledig juist zou zijn. Indien die hypothese juist is, berust deze ,,gevonden afwijking" uitsluitend op de meet- of telfouten. Indien die hypothese niet juist is, dan berust zij niet uitsluitend op die meet- of telfout, maar ook op de fout van die hypothese. Daarom vraagt de statistica of de kans klein is, dat de meet- of telfouten in de gegeven omstandigheden een afwijking veroorzaken, even groot of nog ernstiger dan de gevonden afwijking. Is deze kans, de ,,overschrijdingskans" der gevonden afwijking, heel klein, dan zouden wij het al heel ongelukkig hebben getroffen, dat zulk een grote fout nu juist bij de door ons verrichte meting of telling zou zijn opgetreden en dus achten wij het waarschijnlijk, dat de getoetste hypothese niet juist is. En zo wordt dus, als in een bepaald statistisch onderzoek de overschrijdingskans heel klein blijkt te zijn, de conclusie getrokken, dat de getoetste hypothese onbetrouwbaar wordt geacht als zijnde waarschijnlijk in strijd met de verkregen meet- of telgetallen. En de mate van onbetrouwbaarheid, dat wil zeggen de mate van betrouwbaarheid dezer conclusie, volgt uit de kleinte van de gevonden overschrijdingskans. Indien geen bijzondere eisen worden gesteld is men tevreden met een betrouwbaarheidsdrempel, die overeenkomt met een overschrijdingskans van 0.05.

In beginsel is dit de enige vorm van conclusie, waartoe de statistica ons in staat stelt. En dat is een negatieve conclusie: de getoetste hypothese wordt erdoor afgewezen (zij het ook slechts met een beperkte mate van betrouwbaarheid). *De statistica zoekt dus alleen naar argumenten, om de te toetsen hypothese af te wijzen.*

Welke conclusie kan dan worden getrokken als zulke argumenten niet worden gevonden, d.w.z. als de overschrijdingskans groot is be-

vonden? Feitelijk kan dan geen conclusie worden getrokken: een grote overschrijdingskans betekent eenvoudig, dat er in de gegeven omstandigheden een grote kans bestaat, dat de meet- of telfouten een afwijking als de gevondene — of zelfs een nog ernstiger — veroorzaken. In een geval, waarbij de overschrijdingskans groot blijkt te zijn, is het dus heel wel denkbaar, dat de gevonden afwijking uitsluitend op de toevallige fout van de meting of telling berust en de hypothese dus juist is, maar het kan ook best zijn, dat deze meet- of telfout slechts een deel van de gevonden afwijking op zijn rekening heeft en een ander deel het gevolg zou zijn van een fout in de te toetsen hypothese. En zo is dus, als de overschrijdingskans groot is, geen aanwijzing beschikbaar om te kiezen tussen pro en contra.

Bij een grote overschrijdingskans luidt de juiste conclusie dan ook: „de gegeven meet- of telgetallen pleiten niet tegen de getoetste hypothese" of „zij bevatten geen argument om de te toetsen hypothese in twijfel te trekken". Wel is het in zulk een geval opmerkelijk, dat de meting of telling geen verschil tussen hypothese en werkelijkheid aan het licht heeft gebracht: mocht de hypothese het al mis hebben, dan heeft zij het blijkbaar toch niet ver mis. Het is deze facet, die naar voren komt in die andere manier, waarop de conclusie ook wel wordt geformuleerd bij het aantreffen van een grote overschrijdingskans: „binnen de onnauwkeurigheid van de gevolgde meet- of telmethode is de te toetsen hypothese juist gebleken". Intussen blijven de eerst gegeven zegswijzen helderder.

Het negatieve karakter van een statistische conclusie kan soms in een positief verkeren: als van twee mogelijkheden één met grote stelligheid van de hand wordt gewezen, en een derde mogelijkheid bestaat er niet — ziedaar echter het twijfelachtige beginsel tertium non datur! — dan is de eerste met even grote stelligheid door de meting bevestigd. Deze figuur treedt vaak op als het de vraag geldt, of de ene gebeurtenis een invloed op de andere uitoefent. Wij stellen dan de nulhypothese, volgens welke deze invloed niet zou bestaan (de waarde 0 zou hebben) en toetsen deze aan de gegeven meet- of telgetallen. Als dan blijkt, dat de overschrijdingskans heel klein is, b.v. 0.000001, en de nulhypothese dus met een bijzonder grote betrouwbaarheid moet worden afgewezen, dan is daarmee tevens praktisch met een even grote betrouwbaarheid bevestigd, dat bedoelde invloed wèl bestaat.

Dit voorbeeld vormt al een aanwijzing, dat wij er niet ernstig over behoeven te treuren, dat de statistica in beginsel slechts negatieve conclusies trekt, dat zij de te toetsen gegeven hypothesen alleen maar kan afwijzen (en dat 1. nog niet altijd als 't zou behoren en 2. nooit

met 100 % betrouwbaarheid!). Het werpt alleen een duidelijk licht op 's mensen pad naar de onbekende sferen, dat is: naar de ons omringende werkelijkheid. Zuiver aanschouwen kunnen wij die niet. Maar wat wij ervan zien — door de mist onzer zintuigen heen — geeft onze fantasie gerede aanleiding tot tal van vermoedens. Dan kijken wij gespannen verder of wij op het pad, dat wij vermoeden het juiste te zijn, geen gevaren zien opdoemen voor onze helaas zo besluierde blik. Zo lang wij daarvan geen tekenen ontdekken, zo lang gaan wij door op 't ingeslagen pad. En dan, eerst vaag, daarna helderder, duikt er een baken op, dat ons een „neen" toeroept. Dankbaar erkennen wij de eventuele duidelijkheid van dit „neen" en wij stoppen om, proberenderwijs, een nieuwe richting in te slaan. Tot ook daar weer een afwijzend baken ons waarschuwt. En zo sturen wij de wagen onzes geestes voort, steeds zonder positieve aanwijzing te ontvangen dat de weg, die wij volgen, de juiste is, maar wel telkens door duidelijke bakens gewaarschuwd als wij het juiste pad te ver verlaten.

De statistica is een hulpmiddel, om het „neen" dezer bakens zo helder mogelijk te onderscheiden. En daarvoor komt haar lof toe:

Gelukkig hij en vroom en goed,
Die neen durft zeggen, neen,
Dat schone woord vol mannenmoed,
Tot iedereen.

Het derde punt samenvattende kunnen wij dus zeggen, dat het toetsen van de geopperde hypothese aan de gegeven meet- of telgetallen neerkomt op de vraag, of die getallen een bezwaar tegen die hypothese inhouden. Het middel, om die vraag te beantwoorden, is de overschrijdingskans: blijkt deze klein te zijn, dan vormen de gegevens een bezwaar tegen die hypothese in zo verre dat dan eveneens de waarschijnlijkheid klein is dat het een juiste hypothese zou zijn. Wat de statistica tracht op te sporen is dus steeds iets afbrekends, iets negatiefs: een argument om de geopperde mening af te wijzen. En zelfs in dit negatieve werk kan de statistica nooit met zekerheid slagen, zodat zij uiteindelijk slechts in staat is de mate van betrouwbaarheid te berekenen van een tegen de getoetste hypothese in te brengen argument. En dit berekenen heeft in de meeste gevallen nog weer slechts het karakter van een schatting.

Datgene, dat de betekenis van de statistische conclusie beperkt, werd hiermee zo sterk op de voorgrond gebracht om te trachten de gevaren van een verkeerd begrip omtrent de statistica te keren. Het vrijwel onschatbaar grote nut, dat de statistiek kan brengen ondanks

de in beginsel zo merkwaardig beperkte betekenis harer conclusies, behoeft er niet minder door te worden.

De rol van de meet- en telgetallen overziende kunnen wij dus in 't kort vaststellen, dat deze getallen drieërlei rol spelen bij het ontwikkelen van ons inzicht omtrent de gemeten of getelde werkelijkheid:

eerste rol: een *suggestie* bij het opperen van een hypothese omtrent de gemeten of getelde werkelijkheid
werktuig daarbij: de fantasie van den onderzoeker
hulpwerktuig: de grafische voorstelling
gevolg van deze rol: de geopperde hypothesen dragen een hypothetisch karakter (zou de werkelijkheid misschien zus of zo zijn?)

tweede rol: het quantitatief *preciseren* van de geopperde hypothese
werktuig: de beste schatting
hulpwerktuig: de grafische voorstelling
gevolg van deze rol: een te vage hypothese kan geschikt worden gemaakt om aan de meet- of telgetallen te worden getoetst

derde rol: het uitoefenen van *kritiek* op de geopperde hypothese
werktuig: de overschrijdingskans
gevolg van deze rol: het afwijzen van sommige der geopperde hypothese als zijnde waarschijnlijk in strijd met de verrichte meting of telling.

Het karakter van het statistisch onderzoek is daardoor het beantwoorden van de vraag, of de gegeven meet- of telgetallen een voldoend betrouwbaar *bezwaar* tegen een te toetsen hypothese behelzen.

DE DRAAGWIJDTE VAN EEN STATISTISCHE CONCLUSIE

De betekenis van een statistische conclusie wordt niet alleen beperkt door de mate harer betrouwbaarheid, maar ook doordat haar draagwijdte begrensd is. Haar betrouwbaarheid peilt de waarschijnlijkheid, dat zij juist zou zijn. Dat was het onderwerp, waarover wij tot nu toe spraken. De draagwijdte van een statistische conclusie is heel iets anders: stel dat die conclusie waar is, hoe ver strekt zich dan het terrein harer geldigheid uit? De betrouwbaarheid betreft de vraag: *geldt deze conclusie*? De draagwijdte betreft de vraag: *als zij geldt, waar geldt zij dan*?

Indien de Heren HAGEDOORN hun onderzoek naar het voorkomen van ééneiige tweelingen niet bij cavia's hadden ondernomen maar bij katten en indien dan hun conclusie had geluid: „het getelde materiaal

pleit met grote betrouwbaarheid voor het voorkomen van eéneiïge tweelingen", voor welke groep van levende wezens had dan deze conclusie gegolden? Gesteld dat zij juist was en er dus inderdaad zulke tweelingen voorkomen, bij welke dieren zouden die tweelingen dan volgens deze conclusie voorkomen? Bedoelde (slechts gefingeerde) conclusie berustte op een bepaalde steekproef, bestaande uit alle huiskatten, die tussen twee bepaalde tijdstippen in een bepaalde stal geboren waren. Geldt die conclusie nu uitsluitend voor deze ene steekproef of geldt zij ook voor andere huiskatten? Of zelfs ook voor andere soorten van katten, voor andere soorten van roofdieren, of zelfs ook voor planten? Waar eindigt, op deze lange scala, de draagwijdte van de bedoelde statistische conclusie?

De berekeningen, die tot een bepaalde statistische conclusie leiden, berusten uiteindelijk, zoals wij hebben gezien, op kansen bij het nemen van een steekproef. Onze eerste beschouwingen gingen uit van een willekeurig begrensde steekproef uit een loterij, bestaande uit prijzen en nieten, en uit die steekproef werd een conclusie afgeleid omtrent die loterij in haar geheel, n.l. omtrent de kans om een prijs aan te treffen als uit die loterij een lot wordt getrokken. De draagwijdte van deze conclusie strekte zich dus wel degelijk verder uit dan over de genomen steekproef alleen. Maar niet verder dan de betrokken loterij: de grens van die loterij vormde tevens de grens van haar draagwijdte.

Voordat de steekproef genomen was, was reeds aan al haar mogelijke begrenzingen een uiterste grens voorgeschreven: haar grenzen moesten blijven binnen de grenzen van onze ene loterij. En deze laatste grenzen waren niet meer door het toeval bepaald. De draagwijdte strekte zich dus uit zover als de mogelijkheid was opengelaten dat de grenzen van de steekproef door het toeval zouden worden aangegeven.

Om de grens aan te geven van de draagwijdte van een bepaalde statistische conclusie dienen wij de situatie, waar deze op berust, dan ook te vergelijken met de situatie van een willekeurig begrensde steekproef uit een loterij: zover als de twee situaties op het cardinale punt overeenkomen, zover reikt ook de bedoelde draagwijdte.

Welk is hier het cardinale punt van overeenstemming?

De steekproef, die wij telkens weer als uitgangspunt voor onze theoretische beschouwingen hebben gebruikt, behoefde niet in alle opzichten willekeurig begrensd te zijn: bedoelde beschouwingen zouden even goed hebben gegolden als deze begrenzing maar had berust op een eigenschap, die onafhankelijk van de te onderzoeken eigenschap varieert; zij zou n.l. alleen dan niet hebben gegolden als de eigenschap, op grond waarvan de steekproef begrensd was, gecorreleerd zou zijn

geweest aan de te bestuderen eigenschap. Als de loten gemaakt waren van papiertjes van uiteenlopende kleur en wel zo, dat alle prijzen uit groen papier bestonden, alle nieten uit andere kleuren — er was dan een maximale correlatie tussen de eigenschap „groen" en de eigenschap „prijs" — en als wij dan de steekproef zo hadden begrensd, dat we als steekproef alleen groene loten hadden uitgezocht, dan was het prijsgehalte in onze steekproef 100 % geweest en dan had de besproken binomiale kansdistributie er dus niet voor gegolden. Waren daarentegen de kleuren zodanig over de loten verdeeld, dat van de prijzen en de nieten eenzelfde percentage van groen papier was gemaakt — in welk geval er geen correlatie tussen de eigenschap „groen" en de eigenschap „prijs" had bestaan — en waren weer de groene loten als steekproef uitgezocht, dan was de kansdistributie van zijn prijsgehalte dezelfde geweest als wanneer deze steekproef geheel willekeurig begrensd zou zijn. Voor de geldigheid van onze beschouwingen is het slechts nodig dat de materiaalselectie *ten opzichte van de te onderzoeken eigenschap* zuiver toevallig tot stand is gekomen.

De cardinale eis, die aan een statistisch onderzoek dient te worden gesteld, is dus, dat er geen correlatie mag bestaan tussen de te onderzoeken eigenschap en de eigenschap, die heeft beslist over de begrenzing van het materiaal, waar de gegeven meet- of telgetallen aan ontleend zijn. En de draagwijdte van een statistische conclusie reikt zo ver als aan deze cardinale eis is voldaan. Als de conclusie juist is, geldt zij n.l. tot aan die grens, die niet meer onafhankelijk van de te onderzoeken eigenschap tot stand zou zijn gekomen.

De genoemde cardinale eis is vaak gemakkelijker te stellen dan dat er aan kan worden voldaan. En heel vaak kost het de grootste moeite om uit te maken, in hoe verre er aan voldaan is en hoe ver de draagwijdte van de verkregen statistische conclusie zich dus uitstrekt.

Beschouwen wij de (gefingeerde) conclusie omtrent het voorkomen van ééneiïge tweelingen bij katten. Het materiaal was hier door twee eigenschappen begrensd: de katten moesten 1. tussen twee bepaalde tijdstippen en 2. in een bepaalde stal geboren zijn. Volgens het bovenstaande zou de draagwijdte niet tot deze ene steekproef, dus niet tot deze speciale katten, beperkt zijn. Maar uit welke grotere verzameling was dat dan een steekproef, die ten opzichte van de kans op eisplijting willekeurig begrensd was? Indien de eisplijting een eigenschap is, die bij de katten aan het toenemen is, dan zou een soortgelijke steekproef, een eeuw vroeger genomen, op een andere kans hebben berust, zodat de begrenzing in tijd van de door ons genomen steekproef geenszins als zeker onafhankelijk van de te onderzoeken kans op eisplijting zou mogen worden beschouwd. En als de eisplijting invloed zou onder-

vinden van de grondsoort, de voeding of van het leven in een stal tegenover dat in de vrije natuur, dan was de te onderzoeken eigenschap niet onafhankelijk geweest van de ruimtelijke begrenzing onzer steekproef. Is dan bij dit onderzoek wel aan de cardinale eis voldaan?

Deze situatie laat tweeërlei interpretatie toe, waarbij toch een draagwijdte kan worden aangegeven.

In de eerste plaats: aangenomen, dat de kans op eisplijting bij de kat noch met de tijd noch met de plaats verandert, dan geldt onze conclusie voor ,,de'' huiskat (althans als zij juist is: zie de kwestie van de betrouwbaarheid) want de steekproef was, afgezien van de grenzen in tijd en plaats, uiteindelijk slechts beperkt door het voornemen, dat zij uit huiskatten zou bestaan. Op deze wijze bereiken wij dus een *voorwaardelijke draagwijdte* (i.c. een draagwijdte, die gebonden is aan de genoemde voorwaarden, dat de kans op eisplijting noch met de tijd noch met de plaats zou veranderen).

De andere interpretatie lijkt misschien wat meer gedrongen. Zij gaat uit van de overweging, dat het reeds van ontelbaar veel factoren — dus van het toeval — heeft afgehangen, dat in dat tijdsbestek en in die stal *juist deze* katten werden geboren, zodat zij dus wel degelijk een zuiver toevallig begrensde steekproef vormen uit al die huiskatten, die *mogelijk* in dit tijdsbestek en in deze stal zouden kunnen zijn geboren (hier wordt de werkelijkheid dus opgevat als de ene, *toevallig* gerealiseerde uit de vele mogelijke toestanden). Op grond van dit gezichtspunt geldt onze conclusie, zonder dat wij enige voorwaarde behoeven te stellen, voor dit tijdsbestek en voor deze stal. En degene, die vol schrik meent te moeten opmerken, dat de statistische conclusie dus ,,eigenlijk'' toch alleen maar voor de genomen steekproef geldt, krijgt ten antwoord, dat zelfs de geringe, hier bereikte draagwijdte, op de keper beschouwd, toch buiten de ene genomen steekproef uitgaat: de conclusie geldt ,,voor de omstandigheden'' van die steekproef.

Vele niet-statistisch onderlegde onderzoekers, die een statistische uitspraak over hun gegevens vragen, bezien zulk een uitspraak in het ,,licht'' van twee vage indrukken, die beide fout zijn, n.l. 1. de statistica werkt wiskundig en de statistische conclusie biedt dus wiskundige zekerheid; 2. het kansbegrip bezit een zodanig mystiek element, dat de statistische conclusie daardoor ongelimiteerd verder reikt dan de gegevens, waarop zij berust. Zoals gezegd, beide indrukken zijn fout. Wel gebruikt de statistica, evenals de andere natuurwetenschappen, wiskundige afleidingen ter vereenvoudiging van het nodige cijferwerk, maar zowel haar uitgangspunt, de onnauwkeurige gegevens, haar gereedschap, de kansdistributie, als haar doel, het schatten van de mate van betrouwbaarheid, gaan buiten de wiskunde uit en in

plaats van wiskundige zekerheid te bieden, beoogt zij slechts de betrouwbaarheid van bepaalde vermoedens zo goed mogelijk te schatten. Dat bespraken wij in het vorige hoofdstuk. En de geldigheid van haar conclusie strekt zich — juist ten koste van de zekerheid — veel verder uit dan de gegeven steekproef, waarop zij berust, maar niet ongelimiteerd verder, integendeel slechts tot een scherp gedefinieerde grens, n.l. tot daar, waar de cardinale eis niet meer is vervuld. En dat betekent: tot daar waar de omstandigheden zodanig afwijken van die der gegevens, dat mogelijk ook de te onderzoeken eigenschap anders is geworden. Zo geformuleerd is deze eis even eenvoudig en logisch als ook beperkend. Dat laatste is pijnlijk, maar waar, zodat het van groot belang is, dit punt steeds goed en open onder ogen te zien.

De cardinale eis, dat de materiaalbegrenzing onafhankelijk van de te onderzoeken eigenschap dient te zijn, geldt niet alleen voor de steekproef als geheel, maar ook voor de grenzen tussen de daarin te onderscheiden klassen.

Zulke klassen kunnen op twee manieren worden onderzocht: het kan de bedoeling zijn, dat geteld wordt hoeveel individuen in elke klasse terecht zijn gekomen — b.v. het aantal „prijzen" en het aantal „nieten" — of wel dat een bepaalde eigenschap aan alle individuen van elke klasse wordt gemeten. In het eerste geval is het van belang, dat het herkennen van een bepaald individu als behorend tot een bepaalde klasse — b.v. het herkennen als een „prijs" — niet wordt beïnvloed door de wetenschap, of dat individu de te onderzoeken eigenschap al dan niet kan bezitten. Zo wensten de Heren HAGEDOORN hun onderzoek over het voorkomen van ééneiïge tweelingen bij de cavia uit te breiden door ook gebruik te maken van de opvallende onderlinge gelijkenis van een ééneiïge tweeling. Die opvallende gelijkenis is echter geen scherp en objectief werkend kenmerk, zodat bijzondere maatregelen moesten worden genomen om te verhinderen, dat de kennis omtrent het geslacht van invloed zou kunnen zijn op de beslissing van den onderzoeker, een bepaald paar te rekenen tot de klasse met of tot die zonder opvallende onderlinge gelijkenis: alleen paren met gelijk geslacht kunnen immers ééneiïg zijn en de neiging zou dus kunnen bestaan bij gelijkslachtige paren eerder tot opvallende gelijkenis te besluiten en bij ongelijkslachtige paren eerder tot niet-opvallende gelijkenis. En als deze neiging, al was het nog zo onbewust, invloed zou hebben gehad, dan zou deze klasse-indeling niet aan de cardinale eis hebben voldaan. De oplossing, die in dit speciale geval voor deze moeilijkheid is aangegeven, is deze, dat uit de lijst, waarin van alle cavia's o.a. het geslacht, de kleur en de tekening waren genoteerd, een tweede lijst werd gemaakt, waarin, behalve het nummer van het

betrokken tweelingpaar, alleen de notities over de kleur en tekening werden overgenomen. Aan de hand van deze lijst en daardoor zonder iets te weten over de geslachten werd van elk paar beslist of het al dan niet tot de opvallend gelijkenden zou worden gerekend.

Indien metingen moeten worden uitgevoerd aan de individuen van de diverse klassen in een steekproef, dan grijpt de cardinale eis weer op een andere manier in.

Zou de groei kunnen worden bevorderd door een bepaald vitamine-D-preparaat aan de te drinken melk toe te voegen? Zo luidde de inzet van een onderzoek, dat mede onder auspicieën van het Instituut voor Praeventieve Geneeskunde bij enkele duizenden schoolkinderen is ondernomen en waarbij elk kind gedurende een bepaalde winter òf melk ontving, waaraan dat vitamine-preparaat wel was toegevoegd òf melk zonder die toevoeging en waarbij dat kind voor en na die winter o.a. werd gewogen. Bij de statistische bewerking van die gewichtsgetallen moest rekening worden gehouden met de manier, waarop voor elk kind de beslissing tot stand was gekomen of het de wel- of de niet-gevitamineerde melk zou ontvangen, d.w.z. met de manier, waarop de grens was vastgesteld tussen de klasse met en de klasse zonder vitamine-toevoeging.

De betrokken kinderen waren leerlingen van een 24-tal openbare lagere scholen te Leiden en ter vereenvoudiging van de melkuitdeling was besloten aan elke school slechts één soort melk te verstrekken, zodat kinderen, die op eenzelfde school „schoolmelk" ontvingen, allemaal melk met of allemaal melk zonder het vitamine-preparaat kregen. Is voor de grens, die op deze manier tussen de klasse met en de klasse zonder vitamine-preparaat tot stand is gebracht, de cardinale eis vervuld, dat de eigenschappen, die over deze grens hebben beslist, onafhankelijk zijn van de te onderzoeken eigenschap? Dat zou slechts het geval zijn geweest, indien de groei geheel onafhankelijk zou zijn geweest van de school. Het staat echter geenszins vast, dat de groeikansen op de verschillende scholen dezelfde zijn geweest, integendeel het is juist te verwachten, dat de verschillen in welstand, in buurtgewoonten, in leeftijdssamenstelling en bezigheden der bevolking en vele factoren meer evenzoveel oorzaken zullen zijn geweest voor het optreden van verschil in groeikans op de verschillende scholen, zodat tussen school en groei waarschijnlijk — afgezien van de te verstrekken soorten melk — wel een correlatie bestaat. Wij mogen de twee steekproefklassen dus niet opvatten als twee willekeurig begrensde steekproeven uit twee verzamelingen kinderen, waarin de groeikans uitsluitend door het verschil in wel- en niet-gevitamineerde melk zou verschillen. Ons oordeel over de invloed van het vitamine-preparaat

op de groei mogen wij dus ook niet baseren op de eenvoudige vraag, of de gemiddelde groei van de kinderen, die het vitamine-preparaat wel hebben ontvangen, significant groter is dan die van de andere kinderen. Hadden wij dit wel gedaan, dan hadden wij met behulp van de formule $s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$ in dit geval een zeer significante groeivermeerdering gevonden (n.l. met een overschrijdingskans, veel kleiner dan 0.001, zie blz. 131, maar dan zou de draagwijdte van die conclusie praktisch nul zijn geweest.

Om de realiteit hiervan goed in te zien, stellen wij ons even voor, dat dit onderzoek niet aan 24 maar aan slechts 2 scholen was gedaan, ook weer zo, dat aan de ene school de gevitamineerde melk verstrekt zou zijn en aan de andere school de niet-gevitamineerde. En wij stellen ons verder voor, dat ook tussen de gemiddelde groei op de ene school en die op de andere school een verschil zou zijn gevonden, dat in bovenbedoelde zin zeer significant zou zijn, zodat wij maar zullen aannemen, dat de groeikans op de ene school inderdaad groter is dan die op de andere school. Elke lezer zal echter onmiddellijk voelen, dat uit deze gegevens nog in het geheel niet kan worden afgeleid, dat juist het vitamine-preparaat de oorzaak moet zijn geweest van dit verschil in groeikansen. Het feit, dat de melk op de ene school dit vitamine-preparaat wel bevatte en die op de andere school niet, is n.l. geenszins het enige verschil tussen die twee scholen, dat oorzaak kon zijn van verschil in groeikansen: wij behoeven maar te wijzen op de boven reeds genoemde verschillen in welstand, buurtgewoonten, leeftijdsverdeling en broodwinning van de ouders. Ook als het vitamine-preparaat geen invloed op de groei had gehad, zouden deze andere factoren alleen de groeikansen best verschillend hebben kunnen doen uitvallen. Indien de proef op deze manier zou zijn ingericht, zou het zelfs onmogelijk zijn geweest, er enige statistische conclusie uit af te leiden omtrent een invloed van dit vitamine-preparaat op de groei.

Doordat het onderzoek aan meer dan 2 scholen was uitgevoerd, bood de methode van de variantie-analyse gelegenheid, om toch een statistische conclusie te trekken omtrent de invloed van de vitamine-toediening op de groei. De details daarvan behoeven hier niet te worden behandeld, alleen kan worden meegedeeld, dat, na splitsing van de totale variantie in die tussen de vitamine-klassen, die tussen de scholen [1]) en die binnen de klassen, geen significante ongelijkheid van de variantie tussen de vitamine-klassen en die binnen de klassen bleek te bestaan, zodat de verkregen meetgetallen niet pleiten voor een

---

[1]) en die tussen de leeftijdsgroepen, hetgeen in het verband van dit hoofdstuk echter niet ter zake doet.

invloed van dit vitamine-preparaat op de gewichtsgroei. Hetgeen nog eens bevestigt, dat het verwaarlozen van de invloed dezer bijzondere melkverdeling bij het beoordelen van de draagwijdte tot een foute conclusie zou hebben gevoerd.

Tenslotte over dit melkonderzoek nog een tegenwerping, gemaakt door hen, die het hadden opgezet. „Wij hebben juist", zo merkten zij op, „met grote zorg scholen uitgezocht, waar de algemene omstandigheden zo weinig mogelijk verschillen; naar onze overtuiging zó weinig, dat met de invloed dezer verschillen in maatschappelijke verhoudingen enz. geen rekening meer gehouden hoeft te worden bij het statistisch verwerken der verkregen meetgetallen". Door op deze suggestie in te gaan zou de statisticus zijn conclusie echter op een zeer subjectieve basis gaan bouwen, n.l. op een schatting van de invloed dezer maatschappelijke en andere verschillen tussen de gebruikte scholen en dan nog wel een schatting, die niet eens in maat en getal is uitgedrukt, maar waarvan alleen maar geponeerd wordt, dat zij heel klein zou zijn in vergelijking met het te vinden groeiverschil. Hier laat de geachte opponens zich in de kaart zien: toen hij die opzet maakte, wist hij nog niet in welke orde van grootte het verschil in groei zou worden gevonden! Hoe kon hij dan weten, dat de invloed van de maatschappelijke en andere ongemeten factoren „nog wel veel kleiner" zou zijn? Hij kòn dat niet weten en zijn overtuiging blijft een zuiver subjectieve mening, waarop geen statistische conclusie kan steunen.

Neen, ten opzichte van het streven, voor deze proef scholen te gebruiken, waar de algemene omstandigheden zo weinig mogelijk verschillen, staat de statisticus heel anders dan de bovenbedoelde opponens. De statisticus gaat niet uit van de subjectieve en ongewettigde onderstelling, dat de invloed van deze „overige" verschillen nul zou bedragen, hij meet deze invloed aan de hand van de verkregen meetgetallen (en wel volgens de methode van de variantie-analyse) om dan te kunnen uitmaken, of zij het gevonden groeiverschil hebben kunnen veroorzaken of niet. En daarmee maakt hij zich vrij van subjectieve argumenten. Anderzijds waardeert hij bedoeld streven om de „overige" verschillen tussen de scholen klein te houden, want hoe kleiner deze verschillen zijn, des te gevoeliger is het onderzoek: des te kleiner hoeft n.l. de invloed van het vitamine-preparaat te zijn om nog een significant effect te veroorzaken, n.l. een effect, dat o.a. boven de invloed van deze overige factoren uitkomt.

Soms komt de storende samenhang tussen de eigenschappen, op grond waarvan een steekproef of een klasse van een steekproef werd begrensd, en de eigenschap, die door middel van die steekproef moet worden bestudeerd, pas aan het licht als het onderzoek achteraf nog

eens zeer kritisch wordt bekeken. Als voorbeeld daarvan moge het onderzoek van HOITINK over de phosphaatuitscheiding dienen (blz. 68 e.v.).

Wij zagen al uitvoerig, dat pas de variantie-analyse in staat was, om uit de meetgetallen van een aantal proefpersonen een significant verschil aan te tonen tussen de phosphaatuitscheiding in de perioden met vitamine-toediening en die in de perioden zonder zulk een toediening. De draagwijdte gaat niet verder dan uit deze formulering blijkt; er kunnen verschillende factoren zijn, die in de ene groep van perioden anders waren dan in de andere perioden en in beginsel laat de statisticus het aan de medicus of aan de chemicus over om te beoordelen, welke van deze factoren hij verantwoordelijk zou durven stellen voor het gevonden verschil in phosphaatuitscheiding. Intussen wees het nadere statistische onderzoek van de verkregen meetgetallen zelf al op een bepaald gevaar. Bij het bekijken van de gemiddelde uitscheidingen in de diverse perioden bleek n.l. duidelijk een verband te bestaan tussen het jaargetijde en de phosphaatuitscheiding: in de tijd van het jaar, waarin de meeste dezer onderzoekingen waren verricht, nam de phosphaatuitscheiding gemiddeld af. Nu viel het op dat — om technische redenen — bij elke proefpersoon de phosphaatuitscheiding eerst een tijdlang was bepaald zonder dat vitamine was toegediend en pas daarna een tijdlang onder toediening van die vitamine. De periode met vitamine volgde dus steeds op die zonder vitamine en reeds door deze systematische opeenvolging zou, gezien de neiging tot dalen, de gemiddelde phosphaatuitscheiding in de vitamine-perioden lager worden dan in de perioden zonder vitamine-toediening. Daar zij in de vitamine-perioden inderdaad het laagst was bevonden, kan het gevonden significante verschil zeer wel het gevolg zijn geweest van deze opeenvolging der proefperioden en niet van het al dan niet toedienen van het vitamine-preparaat.

Een nadere nauwkeurige beschouwing van het eerstverkregen resultaat wees hier dus op een zwakke plek in het onderzoek, n.l. op een tevoren niet verwacht verband tussen de te onderzoeken eigenschap en een der factoren, die de grenzen van de steekproefklassen hadden bepaald. Een tweede statistische bewerking, waarbij wel rekening was gehouden met de aan het licht gekomen onafhankelijkheid, leerde dat de metingen geen argument inhielden om een invloed aan te nemen van het vitamine-preparaat op de phosphaatstofwisseling.

Toch was het statistisch onderzoek in dit geval niet uitsluitend afbrekend: het toonde aan, dat de proef in andere jaargetijden diende te worden voortgezet, wilde er zulk een verband uit te voorschijn kunnen komen en deze aanwijzing was even onverwacht als vruchtbaar!

Werd in dit voorbeeld een circumscripte suggestie verkregen voor

het bestaan van een bepaalde, de draagwijdte der statistische conclusie beperkende factor, in vele andere gevallen is de situatie minder doorzichtig doordat een verkregen aanwijzing er slechts op duidt, dat er iets met de opzet van het onderzoek niet in orde is, zonder dat er uit volgt, welke de storende factor zou zijn. Een aanwijzing in deze geest wordt vaak verkregen door na te gaan, of het gegeven materiaal mag worden beschouwd als, zoals dat heet, homogeen. Wat wordt hiermee bedoeld?

Volgens de besproken cardinale eis dient een statistisch onderzoek uit te gaan van gegevens omtrent een materiaal, dat althans in bepaalde opzichten uitsluitend door het toeval is begrensd. Het statistisch onderzoek maakt hiervan immers gebruik door aan te nemen, dat in deze opzichten de gegeven getallen zouden gehoorzamen aan bepaalde wetten of kansdistributies, die juist op dit toeval berusten. De vraag, of de gegevens inderdaad zoo toevallig zijn begrensd als dit door de cardinale eis wordt voorgeschreven, kan nu vaak afzonderlijk worden onderzocht, n.l. door statistisch na te gaan of de theoretisch te verwachten kansdistributie wel degelijk voor de gegeven meet- of telgetallen geldt. Deze uiterst nuttige controle stelt natuurlijk de eis, dat eerst de invloed, die de te onderzoeken factoren zouden hebben uitgeoefend, uit de gegeven getallen uitgeschakeld wordt: de controle dient alleen de z.g. ,,niet nader geanalyseerde factoren'' te betreffen, omdat juist van die factoren was aangenomen, dat zij volgens het toeval zouden zijn begrensd.

Behalve de opmerking, dat het dus heel nuttig is, zo mogelijk na te gaan, of de verkregen meetgetallen wel een normale kansdistributie hebben, zoals dat in den regel maar zonder meer wordt aangenomen, moge een concreet voorbeeld zulk een situatie nog toelichten.

In het jaar 1930 werd in de Leidse vrouwenkliniek een wijziging ingevoerd in de behandeling van de geboorten, resulterend in een meer aktieve leiding van de baring door de behandelende arts, waarvan o.a. een daling werd verwacht van de kans op doodgeboren kinderen. Om na te gaan of deze verwachting was verwezenlijkt, werd in 1940 door H. C. Koch Jr. een statistisch onderzoek ingesteld naar de aantallen in de jaren 1924 tot 1939 in die kliniek levend en doodgeboren kinderen. De conclusie van dat onderzoek luidde, dat na het invoeren van de aktieve leiding de kans op dood geboren worden inderdaad overtuigend kleiner was geworden: het percentage dood in die kliniek ter wereld gekomen kinderen was in de jaren na de invoering (1930—1939) significant kleiner dan in de jaren ervoor (1924—1929).

Bij dit statistisch onderzoek was aangenomen, dat elk dezer twee steekproefklassen (de geboorten in de jaren na en in de jaren voor de

invoering der aktieve leiding) was begrensd door toevallige factoren, althans door factoren, die overigens onafhankelijk waren van de te onderzoeken kans op dood geboren worden. Zo moest dus in elk dezer twee jaargroepen het percentage telkens per jaar dood geboren kinderen overeenkomen met de opvatting, dat de kans, dat een voldragen vrucht dood ter wereld zou worden gebracht, in die jaren dezelfde zou zijn geweest. De jaren 1924—1929 dienden dus te kunnen worden opgevat als even zo vele steekproeven uit een „homogeen" materiaal. Het berekenen van de simultane excentriciteit (*g*) leverde echter een veel te kleine overschrijdingskans, zodat niet kon blijven vastgehouden worden aan het vermoeden, dat het materiaal in dit opzicht homogeen zou zijn zoals de cardinale eis dat vergt.

Nadere beschouwing der gegevens bracht aan het licht, dat het eerste jaar van de onderzochte periode een bijzonder hoog percentage dood geborenen had opgeleverd. Volgens de simultane excentriciteit mocht dit dus niet met voldoende betrouwbaarheid worden beschouwd als een „toevallig" zeer ongunstig jaar in een periode, waarin de kans op het dood geboren worden ener voldragen vrucht constant zou zijn geweest: hoogst waarschijnlijk was in dit jaar een bijzondere oorzaak voor het dood geboren worden aan het werk en dus was dit jaar ten onrechte in de bestudeerde steekproef opgenomen.

Opnieuw de gegeven getallen statistisch bewerkende, maar nu na uitschakeling van het zo ongunstige jaar 1924 — dus uitsluitend beschouwende de jaren 1924 tot 1939 — bleken deze getallen geen argument te bevatten voor een gunstige invloed van de aktieve leiding der baring.

De controle op homogeniteit van het gegeven cijfermateriaal bleek dus in dit geval nuttig: het toonde een fout aan in de eerste statistische bewerking en wel door aan te tonen, dat de gegeven telgetallen waarschijnlijk niet in overeenstemming waren met wat wij de cardinale eis noemden. Waardoor zij daar niet mee in overeenstemming waren, dat bleek echter bij deze controle niet.

Moge uit al deze overwegingen en voorbeelden blijken, dat de betekenis van een statistische conclusie, behalve door de bereikte betrouwbaarheid, ook nog wordt bepaald door haar draagwijdte en dat deze draagwijdte wordt beperkt door de cardinale eis, dat de onderzochte steekproef en zijn klassen uitsluitend begrensd mogen zijn door factoren, die onafhankelijk variëren van de te onderzoeken factoren. De conclusie draagt zo ver als aan deze cardinale eis is voldaan en geen stap verder.

Beschikken wij dan over een hulpmiddel om helder te zien, in hoeverre in elk gegeven geval aan deze cardinale eis is voldaan? Ja: *ons gezonde verstand* en *alleen dat*. Het nuchtere gezonde verstand moge

daarom, bevrijd van alle normen of geleerdheid, op elk statistisch onderzoek worden losgelaten, ongebreideld aangezweept door de meest kritisch ingestelde wens om de fundamenten van het statistisch te beoordelen onderzoek te slopen. Een conclusie, die, door zulk een louterend vuur heengegaan zijnde, toch gehandhaafd bleef, kan althans in zoverre vertrouwen inboezemen, dat zij met de nodige kritiek is bekeken.

DE MEEST ECONOMISCHE OPZET VAN EEN ONDERZOEK, WAARVAN DE UITKOMSTEN LATER STATISTISCH MOETEN WORDEN ONDERZOCHT

Als in de een of andere branche van praktijk of wetenschap een onderzoek zal worden ondernomen, is het vaak bij het begin reeds duidelijk, dat de daarmee verkregen meet- of telgetallen later statistisch zullen moeten worden onderzocht teneinde een zo juist mogelijke conclusie met bekende betrouwbaarheid te verkrijgen. In zulk een geval kan het zeer nuttig zijn, reeds bij de opzet van dat onderzoek rekening met die latere statistische bewerking te houden. Daardoor kan n.l. vaak *de meest economische opzet* worden uitgestippeld, dat is die gang van het onderzoek, waarbij de verhouding tussen de te verkrijgen conclusie en het daartoe op te sporen aantal gegevens zo gunstig mogelijk is, dus waarbij het verlangde inzicht met zo weinig mogelijk meetgetallen of individuen wordt verkregen of waarbij met een gegeven aantal metingen of individuen het nauwkeurigste resultaat wordt bereikt. Dit impliceert, dat de meest economische opzet tevens die is, die bij een gegeven omvang van het onderzoek de meeste kans oplevert om een positieve uitspraak te bereiken.

Om de meest economische opzet te kunnen aanwijzen is meestal een gedetailleerde kennis omtrent de statistische werkmethoden vereist, zodat het niet in de lijn van dit boekje kan liggen om de richtsnoeren daarvoor te behandelen. In elk voorkomend geval roepe men daartoe de raad van een statisticus in. En daar wil dit boekje toe opwekken, want hoe vaker deze raad inzake de meest economische opzet wordt ingeroepen, des te vruchtdragender zal al het meet- en telwerk zijn, dat in praktische en theoretische onderzoekingen wordt verricht en des te rijker zal dus al dit moeizame werk worden beloond. Met rijker wordt hier meer in het bijzonder bedoeld dat een zelfde hoeveelheid technisch werk bij een statistisch economischer opzet meer blijvende, n.l. meer betrouwbare conclusies zal opleveren. Een tweetal voorbeelden, samenhangende met statistische methoden die hier behandeld werden, moge deze opwekking meer kracht bij zetten.

Eerst een zeer simpel geval.

Indien gevraagd wordt of een bepaalde afmeting door een bepaalde ingreep wordt veranderd, dan wordt deze afmeting zowel vóór als nà die ingreep een aantal malen gemeten. De beste schatting voor het verschil is dan het verschil tussen de twee gemiddelden. De standaardfout van dat verschil hangt dan zowel van de nauwkeurigheid der enkele metingen af als ook van het aantal metingen vóór de ingreep en van het aantal metingen erna (zie blz. 50). Als de nauwkeurigheid van de meetmethode niet wordt veranderd, hoe worden de metingen dan het nuttigst verdeeld over vóór en nà de ingreep? Of, anders uitgedrukt: afgesproken dat de afmeting in totaal een bepaald aantal malen zal worden gemeten, hoeveel metingen daarvan moeten dan vóór de ingreep vallen en hoeveel er na, wil de schatting van de verandering dier afmeting het nauwkeurigst zijn? Het antwoord van de statistica luidt: als vóór en na de ingreep even veel malen wordt gemeten.

Volgens blz. 50 volgt de standaardfout van de bedoelde beste schatting van het verschil uit $s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$ (als x de meetgetallen na de ingreep voorstelt en y die er voor), waarvoor wij kunnen schrijven

$$\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y} \text{ of } \frac{n_y\, s_x^2 + n_x\, s_y^2}{n_x\, n_y}.$$

Als de omstandigheden voor het meten vóór en na de ingreep praktisch dezelfde waren, zal hun nauwkeurigheid ook ongeveer dezelfde zijn, zodat s_x ongeveer gelijk zal zijn aan s_y en beide dus, zonder dat wij een grove fout maken, kunnen worden vervangen door een gemeenschappelijk symbool s_m, zodat

$$s^2_{\bar{x}-\bar{y}} = \frac{n_y\, s_m^2 + n_x\, s_m^2}{n_x\, n_y} = \frac{n_x + n_y}{n_x\, n_y}\, s_m^2.$$

Bij diverse verdelingen van ons constante aantal metingen over vóór en na de ingreep blijft $n_x + n_y$ zowel als s_m constant en varieert alleen de noemer $n_x\, n_y$. Nu is het product van twee getallen, die een constante som hebben, het grootst als zij gelijk aan elkaar zijn, zodat $n_x\, n_y$ het grootst is als $n_x = n_y$. In dat geval is dus de standaardfout van het verschil het kleinst en de gevoeligheid en de nauwkeurigheid van het onderzoek het grootst. De meest economische opzet is dus, dat de bedoelde afmeting even vaak vóór als na de ingreep wordt gemeten.

De invloed van deze verdeling der metingen is niet gering. Als er in totaal 100 metingen kunnen worden verricht, dan is het het meest economisch, dat er 50 voor en 50 na de ingreep worden uitgevoerd; $n_x\, n_y$ is dan $50 \times 50 = 2500$. Wordt 20 maal voor en 80 maal na de ingreep gemeten, dan is $n_x\, n_y = 20 \times 80 = 1600$, wordt 10 maal en 90 maal gemeten, dan is $n_x\, n_y = 10 \times 90 = 900$, zodat de variantie $s^2_{\bar{x}-\bar{y}}$ zich in die drie gevallen verhoudt als $\frac{1}{25} : \frac{1}{16} : \frac{1}{9}$ en de standaardfout als de wortels $\frac{1}{5} : \frac{1}{4} : \frac{1}{3}$. Stel dat het tussen de twee gemiddelden gevonden verschil bij de middelste verdeling (20 en 80 metingen) 2 maal de standaardfout bedroeg, dan was de excentriciteit $t = 2$ en de overschrijdingskans 0.05, zodat het onderzoek een resultaat zou hebben opgeleverd dat juist op de grens van de significatie viel.

Bij de meest economische verdeling (50 en 50 metingen) was de standaardfout $\frac{1}{4} : \frac{1}{5}$ of $\frac{5}{4}$ maal kleiner, zodat de excentriciteit $\frac{5}{4} \times 2$ of $2\frac{1}{2}$ bedroeg en de overschrijdingskans ongeveer 0.01 was en het verschil dus zeer overtuigend is aangetoond. Bij 10 en 90 metingen daarentegen zou de standaardfout $\frac{1}{3} : \frac{1}{4} = \frac{4}{3}$ maal groter, de excentriciteit dus $\frac{3}{4} \times 2 = 1.5$ en de overschrijdingskans 0.15 zijn, zodat de meetgetallen geen aanwijzing voor een werkelijk verschil zouden vormen.

Het tweede voorbeeld betreft de methode van de meervoudige variantie-analyse. De directie van een fabriek wenste gedurende een winter aan het personeel een voorbehoedmiddel tegen verkoudheid te verstrekken, maar dan op een zodanige manier, dat na afloop zou kunnen worden geconstateerd of het zou hebben geholpen. Daar het bedrijf in twee gebouwen werd uitgeoefend, stelde zij voor, het middel slechts aan het personeel van het ene gebouw uit te delen, omtrent welk voorstel het oordeel van het Instituut voor Praeventieve Geneeskunde werd gevraagd.

Bij een bezoek ter plaatse bleek, dat het ene gebouw oud, rommelig gebouwd en zeer tochtig was, het andere nieuw, regelmatig en op uitstekende wijze geventileerd. Bij de voorgestelde klasse-indeling zou dus niet zijn voldaan aan de cardinale eis, dat de grens tussen de klassen dient te berusten op het toeval of althans op een eigenschap, die onafhankelijk is van de te onderzoeken eigenschap, hier de kans op kouvatten: de kans op verkoudheid is in het oude gebouw allicht veel groter dan in het nieuwe en als dan na de winter zou blijken, dat het verkoudheidspercentage in de oude fabriek anders zou zijn dan in de nieuwe, dan zou zulks nog geen aanwijzing inhouden om dat speciaal aan het verstrekte voorbehoedmiddel toe te schrijven (juist zoals dat ook niet kon, indien de schoolmelkproef zoals dat op blz. 115 was gesteld, op twee scholen was uitgevoerd). De verhouding tussen resultaat en aantal gegevens zou dus wel heel ongunstig zijn, doordat het nagestreefde resultaat onmogelijk zou kunnen worden bereikt. Het tegenvoorstel omtrent de meest economische opzet hield niet alleen rekening met het feit, dat de twee gebouwen allicht zeer uiteenlopende kans op verkoudheid bieden, maar ook dat de uiteenlopende plaats, die een persoon in zulk een gebouw inneemt, die kans beïnvloedt. Dit geldt niet alleen voor het verschil tussen een werkman en een kantooremployé, maar ook b.v. voor een groep werklui, die regelmatig bij elkaar op een zeer tochtige plaats van het oude gebouw rondom een hevig brandend vuur staan, tegenover een andere groep, die in een goed geventileerde kamer van datzelfde oude gebouw een sorteermachine rustig staat te controleren. Besloten werd het gehele personeel in betrekkelijk kleine groepen in te delen, zó, dat binnen elke groep de kans op kouvatten tijdens het werk kon worden geacht weinig

uiteen te lopen; in elke groep zouden de namen in alfabetische volgorde worden genoteerd en dan genummerd, waarna alle personen met een oneven nummer tabletten met het te bestuderen voorbehoedmiddel zouden ontvangen en allen met een even nummer blanco tabletten (dit laatste om suggestieve invloeden en het onderling verdelen van de tabletten tegen te gaan).

Zonder nauwkeurig in te gaan op de statistische bewerking van de resultaten — de aantallen verzuimdagen wegens „verkoudheidsziekten" van elk lid van het personeel van dit en van vele andere bedrijven, die mede in deze proef werden opgenomen — kan een nadere aanwijzing worden gegeven over het nut van deze kleine, ten opzichte van de verkoudheidskans zo veel mogelijk homogene groepen.

Het alfabetisch rangschikken van de namen, het in deze volgorde nummeren en het uitkiezen van de oneven nummers als klasse met het voorbehoedmiddel is een manier om de klassegrens te doen berusten op een eigenschap, die onafhankelijk is van de te onderzoeken eigenschap, de kans op kouvatten. Indien dit middel zou zijn toegepast op het complex van alle proefpersonen (d.w.z. zonder dat dit complex eerst in kleine homogene groepjes zou zijn verdeeld) dan zou het resultaat moeten worden afgelezen als het verschil van het gemiddeld aantal verzuimdagen wegens verkoudheidsziekten in de klasse met het voorbehoedmiddel en het gemiddelde in de klasse zonder dat middel en de grootte van de standaardfout van dat verschil — uit $s^2_{\bar{x}-\bar{y}} = s_{\bar{x}}^2 + s_{\bar{y}}^2$ — zou een gevolg zijn van de mate, waarin dat aantal verzuimdagen zou uiteenlopen *bij de verschillende personen van zulk een grote klasse.* Er werd al gewezen op het feit, dat in de verschillende gedeelten dier fabrieksgebouwen de kans op kouvatten kon worden geacht sterk uiteen te lopen. Door het personeel echter eerst in kleine, meer homogene groepen in te delen en pas binnen elk van die groepen de „toevallige" indeling tussen de voorbehoedmiddel-klassen tot stand te brengen, kan de methode van de meervoudige variantie-analyse worden toegepast, waardoor, zoals dat op blz. 65 beschreven werd, de gevoeligheid wordt vergroot. In dit geval wordt die grotere gevoeligheid bereikt doordat nu de verschillen in verkoudheidskansen tussen de verschillende groepen afzonderlijk kunnen worden geanalyseerd en dan als het ware kunnen worden geëlimineerd. Daar de groepen zelf juist ten opzichte van de kans op kouvatten zo homogeen mogelijk zijn gekozen, is te verwachten, dat het aantal verzuimdagen *bij de verschillende personen van telkens één zo'n kleine groep* veel minder uiteen zal lopen dan dat voor het gehele personeel het geval zal zijn. En de mate van nauwkeurigheid van het onderzoek hangt nu vrijwel uitsluitend van dit uiteenlopen binnen deze homogene groepen af.

In de mate dat de homogene groepen, waarin het personeel eerst werd verdeeld, werkelijk ten opzichte van de verkoudheidskans homogener zijn dan het onverdeelde, totale personeel, in die mate is het onderzoek dus gevoeliger geworden.

De kans, dat een invloed van het voorbehoedmiddel door een onderzoek als het besprokene aan het licht zal worden gebracht, hangt van twee factoren af: van de grootte dezer invloed en van de gevoeligheid der onderzoekingsmethode. De meest economische opzet van het onderzoek brengt mee, dat dit onderzoek zo gevoelig mogelijk is als bij een personeel van de gegeven omvang kan worden bereikt en daarmee is de kans op het constateren van de eventuele invloed van het voorbehoedmiddel ook zo groot mogelijk geworden.

Vragen wij ons ten slotte nog af, welk hulpmiddel wij bezitten om voor een onderzoek de statistisch meest economische opzet af te leiden, dan kan daarop het antwoord slechts zijn: een zo goed mogelijke speciale kennis van en inzicht in de statistica. Uiteindelijk blijve dit dus het terrein van de specialist-statisticus, al kan iemand, die een zuiver inzicht heeft in de gedachtengang, waarop de statistica berust, in vele gevallen een gelukkige greep doen in de richting van de meest economische opzet. In elk geval: hij zal het nut van de meest economische opzet hoger apprecieren dan degene, voor wie de statistica een gesloten boek is van mystisch-wiskundig karakter, en hij zal dus ook reeds bij de opzet van zijn werk meer profiteren van de door hem ingeroepen steun der statistiek.

HET NUT VAN EEN GLOBALE KENNIS DER STATISTICA

Samenvattend blijkt, dat alle onderzoekers, die uit metingen of steekproeftellingen conclusies over de werkelijkheid wensen te trekken, bewust of onbewust de statistica moeten toepassen. Een globale kennis, zoals dit boekje kan verschaffen omtrent de gedachtengang, waar die statistica op berust, stelt hen geenszins in staat, de eigen uitkomsten exact statistisch te analyseren, maar kan hun daarbij in de volgende opzichten toch van groot nut zijn:

1. zij kunnen daardoor beter aanvoelen, waar en wanneer zij de hulp van de statisticus voor het beoordelen van de door hen verkregen meet- of telgetallen met behoorlijke kans op succes kunnen inroepen, en zullen de gunstige gelegenheden daartoe dus minder vaak ongebruikt laten voorbijgaan;
2. zij zijn beter in staat, met de statisticus samen tot de gunstigste eindconclusie te komen;
3. zij kunnen beter aanvoelen, wanneer zij de hulp van de statisticus met succes kunnen inroepen voor de meest economische opzet van hun voorgenomen onderzoekingen;
4. zij kunnen de draagwijdte van hun conclusies beter beoordelen;
5. zij zullen door de hulp van de statisticus vaker worden behoed voor het trekken van onverantwoorde conclusies uit hun metingen of tellingen;
6. zij zullen, als uit hun onderzoek geen verantwoorde conclusie volgt, door zijn hulp vaker een duidelijke aanwijzing krijgen voor de richting, waarin hun onderzoek met meer kans op succes opnieuw kan worden opgezet.

En dit alles is mogelijk, doordat de statistica de weg wijst voor het meten van

DE BETROUWBAARHEID VAN EEN CONCLUSIE UIT ONNAUWKEURIGE GETALLEN.

Een overzicht van de in dit boek gebruikte

SYMBOLEN en FORMULES

en ook de

NOMOGRAMMEN,

nodig voor het aflezen van

de overschrijdingskans,

vindt men na het aanhangsel op blz. 146 tot 155.

AANHANGSEL

(DE BECIJFERING VAN ENKELE DER BESPROKEN VRAAGSTUKKEN)

Deze appendix is in hoofdzaak bedoeld om de methoden te beschrijven, die worden toegepast voor het becijferen van

1. de kwadraatjessom van een reeks van getallen: $[x_,^2]$,
2. de kwadraatjessom tussen en die binnen enige klassen van getallen: $[f\hat{x}_,^2]$ en $[x_.^2]$,
3. de productensom van een reeks getallenparen: $[x_, y_,]$.

I. *Het vraagstuk omtrent de invloed van het toevoegen van een vitamine-preparaat aan schoolmelk op de vermeerdering in gewicht, besproken op blz.* 114 *e. v.*

De meetgetallen, gevonden voor de gewichtsvermeerdering in de winter 1941/42 van 825 schoolkinderen aan wie gevitamineerde schoolmelk was verstrekt, worden kortheidshalve x genoemd, die voor de 726 kinderen met niet-gevitamineerde schoolmelk y. De eerste staan in de eerste twee kolommen van tabel XV weergegeven, de tweede op dezelfde wijze in tabel XVI.

Deze twee kolommen vormen telkens een frequentietabel:

1. *kolom* x bevat een reeks regelmatig afdalende waarden van de meetschaal der gewichtsvermeerdering,

2. *kolom* f vermeldt telkens het aantal malen, dat een in kolom x genoteerde maat was aangetroffen. (Het is duidelijk, dat de som van deze frequentiekolom gelijk is aan het aantal gegeven x-waarden — dus het aantal kinderen — of n.)

Om hieruit het gemiddelde $\bar{x}$ en in het bijzonder de kwadraatjessom $[x_,^2]$ — nodig voor de berekening van de strooiïng s_x — met zo weinig mogelijk moeite te becijferen, wordt een bepaald rekenschema gevolgd, dat op drie rekenkundige stellingen berust, n.l.:

a. Als van enige getallen eenzelfde bedrag wordt afgetrokken, dan is de kwadraatjessom van de nieuwe getallen gelijk aan die van de oorspronkelijke. (De kwadraatjessom hangt immers uitsluitend van de afwijkingen der getallen ten opzichte van hun aller gemiddelde af; door van elk getal eenzelfde bedrag af te trekken wordt ook hun gemiddelde datzelfde bedrag kleiner, zodat bedoelde afwijkingen — $x_,$ — dezelfde blijven.)

b. Als enige getallen elk door a gedeeld worden, dan is de kwadraatjessom van de nieuwe getallen a^2 maal kleiner dan die van de oorspronkelijke. (De verschillen met het gemiddelde zijn immers elk a maal kleiner geworden, hun kwadraten dus elk a^2 maal.)

c. Voor de kwadraatjessom van een groep getallen, die wij x noemen, geldt

$$[x_,^2] = [x^2] - \bar{x}\,[x]$$

want:

$$\begin{aligned}[x_,^2] &= [(x-\bar{x})^2] = [x^2 - 2x\bar{x} + \bar{x}^2] = \\ &= [x^2] - 2\bar{x}\,[x] + n\bar{x}^2 = \\ &= [x^2] - 2\bar{x}\,[x] + \bar{x}\,[x] = \\ &= [x^2] - \bar{x}\,[x], \text{ q. e. d.}\end{aligned}$$

De berekening, die op deze drie stellingen berust, begint met het kiezen van één der waarden in kolom x als z.g. werkgemiddelde. Meestal wordt daartoe een waarde gekozen, die vermoedelijk dicht bij het rekenkundig gemiddelde ligt of wel de waarde, die de hoogste frequentie heeft. Hier werd gekozen $x_w = 1$ kg. Dan wordt aangelegd:

3. *kolom d*, bevattende een reeks hele getallen, het volgende steeds een kleiner dan het er boven staande, met het getal 0 ter hoogte van de zoëven gekozen x-waarde. Deze d-waarden hangen dus zodanig van de erbij behorende x-waarden af, dat

a. van elke x-waarde het werkgemiddelde x_w is afgetrokken,

b. elk verkregen getal door de „klassebreedte" b gedeeld is (de klassebreedte is het verschil tussen twee op elkaar volgende waarden in kolom x (dit verschil is over de gehele kolom x hetzelfde).

Tengevolge van *a*. is het gemiddelde een bedrag x_w kleiner geworden en veranderd, tengevolge van *b*. is het gemiddelde b maal kleiner en de kwadraatjessom b^2 maal kleiner geworden, zodat

$$\bar{\bar{x}} = x_w + b\bar{d}$$
$$[x_,^2] = b^2\,[d_,^2]$$
$$s_x^2 = b^2\,s_d^2.$$

Van de d-waarden worden nu de gemiddelde en kwadraatjessom berekend, hetgeen minder cijferwerk kost dan bij de oorspronkelijke x-waarden, doordat het kleine en hele getallen zijn [1]).

Dit geschiedt in de volgende twee kolommen:

4. *kolom fd*, verkregen door telkens de twee laatste getallen van éénzelfde regel (dus een van kolom f en een van kolom d) met elkaar te vermenigvuldigen. Het zal duidelijk zijn, dat de som van deze kolom fd gelijk is aan de som van alle d-waarden, dus $[d]$.

5. *kolom* fd^2, weer verkregen door telkens de twee laatste getallen van éénzelfde regel (dus nu een van kolom d en een van kolom fd) met elkaar te vermenigvuldigen. Het is duidelijk, dat de som van deze kolom fd^2 gelijk is aan de som van de kwadraten van alle d-waarden, dus $[d^2]$.

Deze becijferingen kosten weinig tijd, doordat het hele en meest kleine getallen betreft. Om te controleren of rekenfouten zijn gemaakt, kunnen nog 3 kolommen worden aangelegd:

6. *kolom* $d + 1$, bevattende een reeks hele getallen als die van kolom d maar de 0 één regel lager, zodat ze elk 1 groter zijn dan het getal in kolom d op hun regel.

7. *kolom* $f\,(d + 1)$, verkregen door vermenigvuldigen van een getal f en een getal $d + 1$ op telkens dezelfde regel.

8. *kolom* $f\,(d + 1)^2$, verkregen door vermenigvuldigen van de twee laatste getallen — $(d + 1)$ en $f\,(d + 1)$ — van telkens eenzelfde regel. De som dezer laatste kolom is $[(d + 1)^2]$.

De controle bestaat hieruit, dat de kolomsommen behoren te voldoen aan $[d^2] + 2\,[d] + n = [(d + 1)^2]$. Deze controle van CHARLIER verloopt hier gunstig: $3223 - 2 \times 19 + 825 = 4010$.

Het rekenkundig gemiddelde, de kwadraatjessom en de strooiïng der d-waarden, zo nodig ook de standaardfout van dat gemiddelde, worden nu gemakkelijk uit

---

[1]) Het invoeren der d-waarden geeft in dit speciale vraagstuk weinig voordeel doordat het werkgemiddelde klein is ($x_w = 1$ kg) en de klassebreedte een eenvoudige waarde heeft (0.5 kg). Het werd hier uitsluitend ingevoerd om het principe te demonstreren.

TABEL XV

Berekening van het gemiddelde en van de kwadraatjessom

x = gewichtsvermeerdering in kg.
f = frequentie.

x	f	d	fd	fd^2	$d+1$	$f(d+1)$	$f(d+1)^2$
8.0	1	14	14	196	15	15	225
7.5	0	13	0	0	14	0	0
7.0	0	12	0	0	13	0	0
6.5	0	11	0	0	12	0	0
6.0	1	10	10	100	11	11	121
5.5	1	9	9	81	10	10	100
5.0	3	8	24	192	9	27	243
4.5	4	7	28	196	8	32	256
4.0	2	6	12	72	7	14	98
3.5	6	5	30	150	6	36	216
3.0	10	4	40	160	5	50	250
2.5	32	3	96	288	4	128	512
2.0	77	2	154	308	3	231	693
1.5	136	1	136	136	2	272	544
1.0	216	0	+553	0	1	216	216
0.5	184	—1	—184	184	0	0	0
0	105	—2	—210	420	—1	—105	105
—0.5	27	—3	— 81	243	—2	— 54	108
—1.0	11	—4	— 44	176	—3	— 33	99
—1.5	4	—5	— 20	100	—4	— 16	64
—2.0	3	—6	— 18	108	—5	— 15	75
—2.5	1	—7	— 7	49	—6	— 6	36
—3.0	1	—8	— 8	64	—7	— 7	49
			—572				
	825 = n		— 19 = $[d]$	3223 = $[d^2]$			4010 = $[(d+1)^2]$

de kolomsommen gevonden (de kwadraatjessom volgens de derde der zoëven genoemde stellingen):

$$\bar{d} = [d] : n = -19 : 825 = -0.023$$
$$[d_{,}^2] = [d^2] - \bar{d}\,[d] = 3223 - 0.023 \times 19 = 3223 - 0.437 = 3222.6$$
$$s_d^2 = [d_{,}^2] : (n - 1) = 3222.6 : 824 = 3.915.$$

Van dit getal wordt vaak $^1/_{12}$ of 0.08333 . . . afgetrokken om de invloed te corrigeren, die het telkens samenvoegen van enige getallen tot één klasse waarschijnlijk zal hebben gehad. Deze „correctie van Sheppard" mag niet worden toegepast, indien een overschrijdingskans moet worden bepaald.

Voor de oorspronkelijke x-waarden volgen het gemiddelde, de strooiing en de standaardfout van het gemiddelde nu uit:

$$\bar{x} = x_w + b\bar{d} = 1 - 0.5 \times 0.023 = 0.988 \text{ kg}$$
$$s^2_x = b^2 s^2_d = 0.25 \times 3.915 = 0.9783$$
$$(s_x = \sqrt{0.9788} = 0.990 \text{ kg})$$
$$s^2_{\bar{x}} = s^2_x : n = 0.9788 : 825 = 0.001186$$
$$(s_{\bar{x}} = \sqrt{0.001186} = 0.034 \text{ kg}).$$

Voor de kinderen met de niet-gevitamineerde schoolmelk zijn de berekeningen in en onder tabel XVI uitgevoerd. Wat voor de x-waarden d werd genoemd, heet voor deze y-waarden e.

TABEL XVI

(als tabel IX, maar voor de niet-gevitamineerde schoolmelk)

y	f	e	fe	fe^2	$e + 1$	$f(e + 1)$	$f(e + 1)^2$
8.5	1	15	15	225	16	16	256
⋮							
5.5	1	9	9	81	10	10	100
5.0	1	8	8	64	9	9	81
4.5	0	7	0	0	8	0	0
4.0	3	6	18	108	7	21	147
3.5	3	5	15	75	6	18	108
3.0	7	4	28	112	5	35	175
2.5	20	3	60	180	4	80	320
2.0	36	2	72	144	3	108	324
1.5	109	1	109	109	2	218	436
1.0	159	0	+334	0	1	159	159
0.5	134	—1	—134	134	0	0	0
0	119	—2	—238	476	—1	—119	119
—0.5	84	—3	—252	756	—2	—168	336
—1.0	23	—4	— 92	368	—3	— 69	207
—1.5	21	—5	—105	525	—4	— 84	336
—2.0	2	—6	— 12	72	—5	— 10	50
—2.5	1	—7	— 7	49	—6	— 6	36
⋮							
—7.0	1	—16	— 16	256	—15	— 15	225
—15.0	1	—32	— 32	1024	—31	— 31	961
			—888				
	726 $= n$		—554 $= [e]$	4758 $= [e^2]$			4376 $= [(e+1)^2]$

Controle op rekenfouten:

$$[e^2] + 2[e] + n = [(e + 1)^2]$$
$$4758 - 2 \times 554 + 726 = 4376$$

$\bar{e} = [e] : n = -554 : 726 = -0.763$
$[e_{,}^2] = [e^2] - \bar{e}\,[e] = 4758 - 0.763 \times 554 = 4758 - 422.70 = 4335.30$
$s_e^2 = [e_{,}^2] : (n - 1) = 4335.30 : 725 = 5.9800$
$\bar{y} = y_w + b\bar{e} = 1.0 - 0.5 \times 0.763 = 0.618$ kg
$s_y^2 = b_y^2\, s_e^2 = 0.5^2 \times 5.9800 = 1.4950$
$s_{\bar{y}}^2 = s_y^2 : n = 1.495 : 726 = 0.002059$
$(s_{\bar{y}} = \sqrt{0.002059} = 0.045$ kg)

De beste schatting voor het verschil in gemiddelde gewichtsvermeerdering van de kinderen met de gevitamineerde schoolmelk en van die met de niet-gevitamineerde melk bedraagt

$$\bar{x} - \bar{y} = 0.988 - 0.618 = 0.370 \text{ kg}$$

en de standaardfout hiervan volgt uit:

$$s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}} =$$
$$= 0.001186 + 0.002059 = 0.003245$$
$$\text{of } s_{\bar{x}-\bar{y}} = 0.057 \text{ kg}$$

zodat de excentriciteit is

$$t = d : s = 0.370 : 0.057 = 6.5$$

met de valentie $N = N_1 + N_2 = 824 + 725 = 1549$.

Volgens het t-N-P-diagram is bij $t = 6.5$ en $N = \infty$ de dubbelzijdige overschrijdingskans P_2 vele malen kleiner dan 0.0001, zodat het verschil zeer significant is.

II. *Loopt de phosphaatuitscheiding bij verschillende personen uiteen?* (blz. 62.)

Voor de enkelvoudige variantie-analyse, die hier wordt toegepast, moet een kwadraatjessom tussen en een binnen de klassen worden berekend. Om met kleinere getallen te werken werden de meetgetallen, die in tabel XVII vermeld staan,

TABEL XVII.

Dagelijkse phosphaatuitscheiding in de urine, gemeten bij 5 proefpersonen A—E.

A	B	C	D	E	
84	139	122	98	115	558
92	149	140	88	127	596
87	147	141	115	111	601
107	146	131	92	102	578
79	158	150	90	96	573
92	150	160	92	111	605
105	174	143	74	136	632
77	183	152	94	121	627
111	141	121	89	133	595
73	134	115	110		432
103		133			236
94		111			205
105					105
1209	1521	1619	942	1052	6343 = $[x]$

in tabel XVIII vervangen door d-waarden, die elk 115 kleiner zijn (m.a.w. er werd een werkgemiddelde $x_w = 115$ ingevoerd). Deze nieuwe getallen worden bij elkaar opgeteld, ter controle op rekenfouten zowel horizontaal als verticaal. Aan de totaalsom — $[x] = 133$ — wordt gecontroleerd of bij het opschrijven der d-waarden rekenfouten zijn gemaakt: de totaalsom moet voldoen aan $[d] + nx_w = [x]$, hetgeen hier klopt: $133 + 54 \times 115 = 6343$ (zie de totaalsom van tabel XVII, die op dezelfde manier was berekend).

TABEL XVIII

Dezelfde meetgetallen als in de vorige tabel, maar verminderd met het werkgemiddelde $x_w = 115$

	A	B	C	D	E	totalen		
	—31	24	7	—17	0	—17		
	—23	34	25	—27	12	21		
	—28	32	26	0	—4	26		
	— 8	31	16	—23	—13	3		
	—36	43	35	—25	—19	— 2		
	—23	35	45	—23	— 4	30		
	—10	59	28	—41	21	57		
	—38	68	37	—21	6	52		
	— 4	26	6	—26	18	20		
	—42	19	0	— 5		—28		
	—12		18			6		
	—21		— 4			—25		
	—10					—10		
$[d]_{kl}$	—286	+371	+239	—208	17	= 133	=	$[d]$
$[d]^2_{kl}$	81796	137641	57121	43264	289	17689	=	$[d]^2$
f	13	10	12	10	9	= 54	=	n
$[d]^2_{kl} : f = f\hat{d}^2$	6292.0	13764.1	4760.1	4326.4	32.1	= 29175	=	$[f\hat{d}^2]$
					$[d]^2 : n$	= 328	=	$\bar{d}\,[d]$
						28847	=	$[f\hat{d}_{,}^2]$

Contrôle op rekenfouten:
$[d] + nx_w = 133 + 54 \times 115 = 6343 = [x]$ (zie tabel XVII).

De getallen op de 1e regel onder de rechthoek van tabel XVIII werden in het kwadraat verheven en deze kwadraten werden daaronder op de 2e regel genoteerd. De eerste vijf zijn de kwadraten der klassesommen: $[d]^2_{kl}$. Doordat voor een klassesom geldt $[d]_{kl} = f\hat{d}$ zijn deze klassekwadraten gelijk aan $f^2\hat{d}^2$, zodat deling door f telkens de waarde van $f\hat{d}^2$ leert kennen. Daarom zijn op de 3e regel de klassefrequenties f genoteerd (het aantal getallen in elke kolom van de rechthoek). Deze werden elk op het daarboven staande getal gedeeld, de verkregen uitkomsten staan op de 4e regel, getiteld $f\hat{d}^2$. Hun som (29175), die rechts op deze regel staat, is dus $[f\hat{d}^2]$.

Volgens de stelling $[a_{,}^2] = [a^2] - \bar{a}\,[a]$ moet deze kwadratensom worden verminderd met het product van 1. het gemiddelde van en 2. de som van de betrokken $\hat{d}$-waarden. Dit gemiddelde is gelijk aan het totale gemiddelde $\bar{d}$ en deze som is

gelijk aan de totaalsom $[d]$, zodat de reeds meer gebruikte correctie $\bar{d}\,[d]$ er van afgetrokken dient te worden:

$$[f\hat{d}_{,}{}^2] = [f\hat{d}^2] - \bar{d}\,[d].$$

De waarde van deze correctie wordt hier het gemakkelijkst gevonden uit $\bar{d}\,[d] = [d]^2 : n$. Daartoe werd ook de regel met de klassefrequenties f gesummeerd en werd deze som (het totale aantal meetgetallen n) gedeeld op het daarboven staande getal $[d]^2$. De uitkomst — $328 = \bar{d}\,[d]$ — staat onder $[f\hat{d}^2]$ genoteerd, zodat een aftrekking de kwadraatjessom $[f\hat{d}_{,}{}^2]$ levert. Doordat de klassebreedte der gegeven meetgetallen gelijk is aan die van de d-waarden (er is uitsluitend een werkgemiddelde toegepast, een andere eenheid werd niet ingevoerd) is dit ook de gevraagde kwadraatjessom tussen de klassen $[f\hat{x}_{,}{}^2]$ (anders was $[f\hat{x}_{,}{}^2] = b^2\,[f\hat{d}_{,}{}^2]$ als b de klassebreedte der x-waarden voorstelt).

TABEL XIX

Kwadraten van de d-waarden uit tabel

A	B	C	D	E	totalen
961	576	49	289	0	1875
529	1156	625	729	144	3193
784	1024	676	0	16	2500
64	961	256	529	169	1979
1296	1849	1225	625	361	5356
529	1225	2025	529	16	4324
100	3481	784	1681	441	6487
1444	4624	1369	441	36	7914
16	676	36	676	324	1728
1764	361	0	25		2150
144		324			468
441		16			457
100					100
8172	15933	7385	5524	1507	38521 = $[d^2]$

$$[d^2] = 38521$$
$$\bar{d}\,[d] = 328$$

totale kw.som = $[x_{,}{}^2] = [d_{,}{}^2] = 38193$ $\quad N_{\text{tot.}} = 54 - 1 = 53$

kw.som tussen de kl. = $[f\hat{x}_{,}{}^2] = [f\hat{d}_{,}{}^2] = 28847$ $\quad N_{\text{t.kl.}} = 5 - 1 = 4$

kw.som binnen de kl. = $[x^2_{.}] = 9346$ $\quad N_{\text{b.kl.}} = 49$

$G^2 = N_{\text{b.kl.}}\,[f\hat{x}_{,}{}^2] : [x_{.}{}^2] = 49 \times 28847 : 9346 = 152$ met $N = N_{\text{t.kl.}} = 4$

$P_1 < 0.0001$

of

$s_1{}^2 = 28847 : 4 = 7212 \qquad s_2{}^2 = 9346 : 49 = 191$

$s_1{}^2 : s_2{}^2 = 7212 : 191 = 38$

$s_1 : s_2 = 6.2$ met $N_1 = 4$ en $N_2 = 49$

$P < 0.0001$

De kwadraatjessom binnen de klassen wordt gevonden door die tussen de klassen van de totale af te trekken, dus van $[x_{,}^2]$ of $[d_{,}^2]$ die becijferd wordt uit $[d_{,}^2] = [d^2] - \bar{d}\,[d]$. De som van de kwadraten van alle d-waarden — $[d^2]$ — wordt becijferd door een tweede tabel aan te leggen (tabel XIX), waarin alle d-waarden van tabel XVIII door hun kwadraten zijn vervangen en door deze dan alle te summeren (ter contrôle op rekenfouten weer zowel horizontaal als verticaal).

Tenslotte kan de overschrijdingskans worden bepaald met behulp van de simultane excentriciteit G of door middel van de strooiingsverhouding $s_1 : s_2$.

De beschreven cijfermethode bestaat in hoofdzaak uit het opzoeken van kwadraten uit een kwadratentabel, het verdere cijferwerk is gering en eenvoudig.

III. *Verandert zware lichamelijke arbeid het vitamine-C-gehalte in het bloed?* (blz. 63 e. v.).

Om de kwadraatjessommen voor de toe te passen dubbele variantie-analyse te becijferen werd niet uitgegaan van de in tabel XI gegeven meetwaarden zelf, maar, om met kleinere getallen te kunnen werken, van getallen die elk 18 kleiner zijn dan die meetwaarden, m. a. w. van d-waarden bij een werkgemiddelde van $x_w = 18$ (gekozen omdat het werkelijke gemiddelde iets hoger ligt: 18.52, zie tabel XI). Deze nieuwe getallen zijn binnen de rechthoek van tabel XX op dezelfde manier gerangschikt als de gegeven meetgetallen in tabel XI.

Onder de rechthoek staan de kolomsommen $[d]_{\text{arb.}}$ en rechts van de rechthoek de regelsommen $[d]_{\text{pers.}}$. Deze behoren beide dezelfde totaalsom te leveren. Aan deze — $[d] = 31$ — kan gemakkelijk worden gecontroleerd of bij het opschrijven van de d-waarden geen rekenfouten zijn gemaakt: nagegaan wordt daartoe of voldaan is aan $[d] + nx_w = [x]$, hetgeen hier het geval blijkt te zijn: $31 + 60 \times 18 = 1111$ (verg. tabel XI).

Onder de kolomsommen staan hun kwadraten en zo staan de kwadraten der regelsommen rechts van deze. Ook de door deze kwadraten gevormde regel en kolom worden elk gesummeerd. Hun sommen — 1669 en 3327 — zijn betiteld $f[f\hat{x}^2]_{\text{arb.}}$ en $f[f\hat{x}^2]_{\text{pers.}}$. Daar voor een klassesom geldt $[d] = f\hat{d}$, is immers elk genoemd kwadraat van een klassesom gelijk aan $f^2\hat{d}^2$ en is hun som gelijk aan $[f^2\hat{d}^2]$, waarvoor ook $f\,[f\hat{d}^2]$ mag worden geschreven, omdat in ons vraagstuk bij elke klasse-indeling alle klassen hetzelfde aantal meetgetallen bevatten (de kolommen elk 20 en de regels elk 3) en de klassefrequenties f dus voor alle klassen eenzelfde getal zijn. Onder de tabellen leerde deling door de betrokken klassefrequentie ($f_{\text{arb.}} = 20$ en $f_{\text{pers.}} = 3$) de waarde van $[f\hat{d}^2]_{\text{arb.}}$ en $[f\hat{d}^2]_{\text{pers.}}$ kennen [1]).

Van elk dezer twee kwadratensommen moet, om de kwadraatjessommen te verkrijgen, $\bar{d}\,[d]$ worden afgetrokken (zie blz. 129) welk bedrag rechts van de zoëven genoemde delingen uit $\bar{d}\,[d] = [d]^2 : n$ werd berekend op 16.02. Met dit bedrag werden de verkregen kwadratensommen verminderd, waarmee de kwadraatjessommen tussen de klassen — [arbeid] en [personen] — gevonden zijn. Doordat de klassebreedte der d-waarden gelijk is aan die van de x-waarden (alleen een werkgemiddelde is toegepast, een andere eenheid werd niet uitgevoerd) zijn de kwadraatjessommen der d-waarden gelijk aan die van de x-waarden (anders waren die van de x-waarden b^2 maal die van de d-waarden als b de klassebreedte der x-waarden voorstelt).

---

[1]) In het vorige voorbeeld kon deze kortere manier voor het berekenen van $[f\hat{d}^2]$ niet worden gevolgd, doordat de klassen daar niet eenzelfde aantal getallen bevatten en dus elk klasse-somkwadraat afzonderlijk door de frequentie van de eigen klasse gedeeld moest worden.

Nadat een nieuwe rechthoek was aangelegd, waarin alle d-waarden door hun kwadraten waren vervangen (tabel XXI) en de som dezer kwadraten was bepaald (door hen ter contrôle op rekenfouten in twee richtingen op te tellen) werd deze waarde van $[x^2]$ met $\bar{x}\,[x]$ verminderd om $[x_,^2]$ te verkrijgen. Van deze totale kwadraatjessom [totaal] werden [arbeid] en [personen] afgetrokken om [residu] te berekenen en op overeenkomstige wijze werden de valenties behandeld.

De verdere berekeningen behoeven geen toelichting meer [1]).

Ook hier bestaat het cijferwerk in hoofdzaak uit het opzoeken van kwadraten.

TABEL XX

Vitamine-gehalte van bloed (in mgr/L) verminderd met het werkgemiddelde $x_w = 18$

Personen	0	8	50 min. na arbeid	$[d]_{\text{pers.}}$	$[d]^2_{\text{pers.}}$
A	6	7	13	26	676
B	5	5	4	14	196
C	0	2	3	5	25
D	0	1	0	1	1
E	2	5	11	18	324
F	—5	—2	—1	—8	64
G	0	3	4	7	49
H	—1	—1	3	1	1
I	4	12	1	17	289
J	—1	—1	—3	—5	25
K	—5	—9	—9	—23	529
L	4	8	8	20	400
M	—4	2	4	2	4
N	—2	—1	6	3	9
O	1	3	3	7	49
P	—7	—5	—3	—15	225
Q	—4	—2	0	—6	36
R	—1	—3	—2	—6	36
S	—3	—4	—3	—10	100
T	—7	—4	—6	—17	289
$[d]_{\text{arb.}}$	—18	16	33	31	$3327 = f\,[f\hat{d}^2]_{\text{pers.}}$
$[d]^2_{\text{arb.}}$	324	256	1089	1669	$= f\,[f\hat{d}^2]_{\text{arb.}}$

Controle op rekenfouten:

$[d] + nx_w = 31 + 60 \times 18 = 1111 = [x]$ (zie tabel XI).

---

[1]) Indien het niet de bedoeling was geweest een algemene methode voor het becijferen van een meervoudige variantie-analyse te geven, zou voor dit speciale vraagstuk een kortere weg zijn gevolgd door gebruik te maken van het feit, dat de kwadraatjessom van 3 getallen — a, b en c — gemakkelijk kan worden gevonden door hun 3 onderlinge verschillen op te schrijven — a—b, b—c, c—a, controleer of hun som 0 bedraagt — en in het kwadraat te verheffen: de som dezer drie kwadraten is gelijk aan 3 maal de gezochte kwadraatjessom. Nadat deze bewerking op de 20 regels meetgetallen van tabel XI is toegepast (d-waarden behoeven hier-

TABEL XXI

Kwadraten van de d-waarden uit tabel XX

	0	8	50 min. na arbeid	$[d^2]_{pers.}$
	36	49	169	254
	25	25	16	66
	0	4	9	13
	0	1	0	1
	4	25	121	150
	25	4	1	30
	0	9	16	25
	1	1	9	11
	16	144	1	161
	1	1	9	11
	25	81	81	187
	16	64	64	144
	16	4	16	36
	4	1	36	41
	1	9	9	19
	49	25	9	83
	16	4	0	20
	1	9	4	14
	9	16	9	34
	49	16	36	101
$[d^2]_{arb.}$	294	492	615	1401 = $[d^2]$

arbeid	personen	
		$[d] = 31$
$f\,[f\hat{d}^2] = 1669$ =	= 3327	$[d]^2 = 961$
$f = 20$	= 3	$n = 60$
:	:	:
$[f\hat{d}^2] = 83.45$	= 1109	$\bar{d}\,[d] = 16.02$
$\bar{d}\,[d] = 16.02$	= 16.02	
$[f\hat{d},^2] = 67.4$	= 1093	
$[f\hat{x},^2] = 67.4$	= 1093	

voor niet te worden ingevoerd!) is de som der verkregen getallen 3 maal de kwadraatjessom binnen de personenklassen. Om de kwadraatjessom van het residu te verkrijgen, dient daar die tussen de arbeidsklassen van te worden afgetrokken en deze wordt op overeenkomstige wijze gemakkelijk uit de regel der kolomsommen gevonden.

De totale kwadraatjessom behoeft hierbij dus niet te worden bepaald, zodat ook het opzoeken van de kwadraten van alle meetgetallen en het becijferen van hun som vervalt.

$[d^2] = 1401$
$\bar{d}\,[d] = 16$

[totaal] = $[x_r^2]$ = $[d_r^2]$	= 1385	$N_{tot.}$	= 60 — 1 = 59
[arbeid]	= 67.4	$N_{arb.}$	= 3 — 1 = 2
	1317.6		57
[personen]	= 1093	$N_{pers.}$	= 20 — 1 = 19
[residu]	= 224.6	N_{residu}	= 38

Invloed arbeid:
$G^2 = N_{res.}$ [arbeid] : [residu] = 38 × 67.4 : 224.6 = 11.4 met $N = N_{arb.} = 2$
Overschrijdingskans $P_1 = 0.004$
of
$s^2_{arb.}$ = [arbeid]: $N_{arb.}$ = 67.4 : 2 = 33.7
$s^2_{res.}$ = [residu]: $N_{res.}$ = 224.6 : 38 = 5.91
$s^2_1 : s^2_2 = 33.7 : 5.91 = 5.70$
$s_1 : s_2 = \sqrt{5.70} = 2.4$ met $N_1 = 2$ en $N_2 = 38$
Overschrijdingskans $P_1 < 0.01$.

IV. *De invloed van een vitamine-C-preparaat op de phosphaatuitscheiding bij 4 proefpersonen* (zie blz. 68).

De tweevoudige variantie-analyse, die op dit vraagstuk wordt toegepast, eist weer het berekenen van twee kwadraatjessommen tussen klassen en van een residu, met deze bijzonderheid, dat pas de gemiddelden uit telkens 10 meetgetallen een „rechthoek" vormen.

De gegeven meetgetallen staan op blz. 69 in tabel XIII verdeeld over 8 kolommen, n.l. voor elk van de 4 proefpersonen A, B, C en D een kolom van getallen,

TABEL XXII

De phosphaatuitscheidingen, vermeld in tabel XIII, verminderd met het werkgemiddelde 115

A		B		C		D		Totalen
—	+	—	+	—	+	—	+	
—8	—18	24	44	26	22	—27	—39	24
—36	—26	34	11	16	0	0	—38	—39
—23	—4	32	32	35	9	—23	—35	23
—10	—13	31	—4	45	13	—25	3	40
—38	—18	43	13	28	—4	—23	—39	—38
—4	—18	35	22	37	19	—41	—37	13
—42	—45	59	49	6	17	—21	—12	11
—12	—21	68	22	0	22	—26	—26	27
—21	—21	26	27	18	19	—15	—43	—10
—10	—22	19	10	—4	16	—5	—26	—22
—204	—206	371	226	207	133	—206	—292	29 = [d]

Controle op rekenfouten:
$[d] + nx_w = 29 + 80 \times 115 = 9229 = [x]$.

verkregen in een periode zonder gebruik van het vitamine-preparaat (kolom —) en een, afkomstig uit de periode met dat preparaat (kolom +). Hun totaalsom $[x]$ bedraagt 9229.

In tabel XXII zijn zij vervangen door d-waarden, die elk 115 kleiner zijn (m.a.w. werkgemiddelde $x_w = 115$, zo gekozen omdat het totaal gemiddelde iets hoger ligt: $\bar{x} = 115.4$). Het totaal van deze 80 d-waarden, horizontaal en verticaal bepaald, bedraagt 29, de controle op rekenfouten valt gunstig uit: $[d] + nx_w = 29 + 80 \times 115 = 9229 = [x]$.

De 8 kolomsommen worden nu in een rechthoek gerangschikt (tabel XXIV), waarbij de 4 perioden zonder vitamine-preparaat in één kolom en de 4 perioden met dat preparaat in de andere kolom gerepresenteerd zijn en verder elke proefpersoon een regel bezet. Op dezelfde manier als bij het vorige vraagstuk werden hieruit de kwadratensom tussen de kolommen en die tussen de regels berekend, met dit verschil, dat de klasse-frequenties hier niet zijn de aantallen getallen in deze laatste rechthoek, maar het aantal daaraan beantwoordende getallen in de eerste d-rechthoek, zodat $f_{\text{vit.}} = 40$ en $f_{\text{pers.}} = 20$ (omdat elke kolomsom van de laatste rechthoek de som van 40 d-waarden is en elke regelsom die van 20 d-waarden).

TABEL XXIII

Kwadraten van de d-waarden uit tabel XXII

A		B		C		D		
—	+	—	+	—	+	—	+	
64	324	576	1936	676	484	729	1521	6310
1296	676	1156	121	256	0	0	1444	4949
529	16	1024	1024	1225	81	529	1225	5653
100	169	961	16	2025	169	625	9	4074
1444	324	1849	169	784	16	529	1521	6636
16	324	1225	484	1369	361	1681	1369	6829
1764	2025	3481	2401	36	289	441	144	10581
144	441	4624	484	0	484	676	676	7529
441	441	676	729	324	361	225	1849	5046
100	484	361	100	16	256	25	676	2018
5898	5224	15933	7464	6711	2501	5460	10434	59625 = $[d^2]$

TABEL XXIV

Rechthoek van de klassesommen van tabel XXII

	zonder vit.	met vit.	$[d]_{\text{pers.}}$	$[d]^2_{\text{pers.}}$
A	—204	—206	—410	168100
B	371	226	597	356409
C	207	133	340	115600
D	—206	—292	—498	248004
$[d]_{\text{vit.}}$	168	—139	$[d]$ = 29	888113 = $f\,[fd^2]_{\text{pers.}}$
$[d]^2_{\text{vit.}}$	28224	19321	47545	= $f\,[fd^2]_{\text{vit.}}$

	vitamine	personen		
$f\,[f\hat{d}^2]$	= 47545	= 888113	$[d]$	= 29
f	= 40	= 20	$[d]^2$	= 841
	——:	——:	n	= 80
$[f\hat{d}^2]$	= 1188.625	= 44405.65		——:
$\bar{d}[d]$	= 10.5125	= 10.5125	$\bar{d}\,[d]$	= 10.5125
$[f\hat{x}_{,}^2] = [f\hat{d}_{,}^2]$	= 1178	= 44395		

$[d^2] = 59625$
$\bar{d}\,[d] = 10$

[totaal] ..... $[x_{,}^2] = [d_{,}^2]$	= 59615	$N_{\text{tot.}}$	= 80 — 1 = 79
[vitamine]	= 1178	$N_{\text{vit.}}$	= 2 — 1 = 1
	58437		78
[personen]	= 44395	$N_{\text{pers.}}$	= 4 — 1 = 3
[residu]	= 14042	N_{residu}	= 75

Invloed van het vitamine-preparaat:
$t^2 = 79 \times 1178 : 14042 = 6.63$
$t = \sqrt{6.63} = 2.6$ met $N = 79$
$P = 0.01$

V. *Is bij zenuwcellen het celvolume evenredig aan de 1.5e macht van het kernvolume?* (blz. 83 e.v.).

Van de meetgetallen, verkregen voor het kernvolume en voor het celvolume, werd telkens de logarithme genoteerd, waarbij zij resp. x en y werden genoemd, zodat de theorie moet worden getoetst, dat de lineaire regressie-vergelijking $y = A + Bx$ zou passen met $B = 1.5$.

De uit de metingen gevonden x- en y-waarden zijn hieronder binnen de omlijnde rechthoek van tabel XXV in de vorm van een correlatie-tabel weergegeven, waarin voor beide de klassebreedte op 0.1 was gekozen. (Doordat de oorspronkelijke getallen op 0.01 waren afgerond, verkregen de klassemiddens, die links en rechts langs het vierkant vermeld staan, de enigszins vreemd aandoende waarden van 1.645, 1.745 enz.).

Om uit zulk een correlatie-tabel de waarden van de kwadraatjessommen $[x_{,}^2]$ en $[x_{,}^2]$ en van de productensom $[x_{,}\, y_{,}]$ te berekenen, wordt een vast rekenschema gevolgd. Een kleine uitbreiding daarvan leert verder de waarde van die kwadraatjessommen kennen, die nodig zijn om de lineaire regressie en de gestelde regressiecoëfficient (hier 1.5) te toetsen.

De 4 eerste kolommen rechts van de rechthoek dienen voor de gebruikelijke becijfering van $[e]$ en $[e^2]$ teneinde $[y]$ en $[y^2]$ en dus $\bar{y}$ en $[y_{,}^2]$ te vinden. Op de zelfde manier dienen de 4 eerste regels onder de rechthoek voor het becijferen van $[d]$ en $[d^2]$ en dus van $\bar{x}$ en $[x_{,}^2]$.

De 5e kolom rechts van de rechthoek draagt als titel $[fd]$. Met f is hier bedoeld de frequentie, die telkens door één getal binnen de rechthoek wordt vermeld: het aantal malen dat één combinatie, bestaande uit eenzelfde x-waarde en eenzelfde y-waarde, in de gegevens voorkomt. Een $[fd]$-waarde in deze 5e kolom is de som van de frequenties op dezelfde regel, elk vermenigvuldigd met de bij-

TABEL XXV

Correlatie van x (= log kernvolume) en y (= log celvolume)

$y =$ \ $x =$	1.645	1.745	1.845	1.945	2.045	2.145	2.245	2.345	2.445	2.545	2.645	2.745
3.745												1
3.645												
3.545												1
3.445										1	1	
3.345										3		
3.245									1	1		
3.145									3	3	2	
3.045									3	2		
2.945							2	6	3	1		
2.845						2	14	15	2	1		
2.745					2	2	18	6				
2.645			1	1		12	19	4				
2.545					3	8	7					
2.445				7	14	12	1					
2.345			1	5	11	5						
2.245		1	3	11	1							
2.145			9	6	1							
2.045			3	1	1							
1.945		2										
1.845	1											

f_x	1	3	17	31	33	41	61	31	12	12	3	2
d	—5	—4	—3	—2	—1	0	1	2	3	4	5	6
fd	—5	—12	—51	—62	—33	0	61	62	36	48	15	12
fd^2	25	48	153	124	33	0	61	124	108	192	75	72
$[fe]$	—8	—18	—78	—109	—77	—41	43	54	46	63	18	20
$d\,[fe]$	40	72	234	218	77	0	43	108	138	252	90	120
$[fe]^2$	64	324	6084	11881	5929	1681	1849	2916	2116	3969	324	400
$f\hat{e}^2$	64	108	357.9	383.3	179.7	41	30.3	94.1	176.3	330.75	108	200

f_y	e	ef_y	e^2f_y	$[fd]$	$e\,[fd]$
1	11	11	121	6	66
0	10	0	0	0	0
1	9	9	81	6	54
2	8	16	128	9	72
3	7	21	147	12	84
2	6	12	72	7	42
8	5	40	200	31	155
5	4	20	80	17	68
12	3	36	108	27	81
34	2	68	136	54	108
28	1	28	28	28	28
37	0	0	0	22	0
18	—1	—18	18	4	—4
34	—2	—68	136	—27	54
22	—3	—66	198	—24	72
16	—4	—64	256	—36	144
16	—5	—80	400	—40	200
5	—6	—30	180	—12	72
2	—7	—14	98	—8	56
1	—8	—8	64	—5	40
247		—87	2451	71	1392
=		=	=	=	=
247 = n		$[e]$	$[e^2]$	$[d]$	$[de]$

71 = $[d]$

1015 = $[d^2]$

—87 = $[e]$

1392 = $[de]$

)73.35 = $[fe^2]$

behorende d-waarde. Zo is het eerste getal dezer kolom $1 \times 6 = 6$, het tweede 0, het derde $1 \times 6 = 6$, het vierde $1 \times 5 + 1 \times 4 = 9$, het vijfde $3 \times 4 = 12$, het zesde $1 \times 4 + 1 \times 3 = 7$ en het zevende $2 \times 5 + 3 \times 4 + 3 \times 3 = 31$. De som van deze kolom is $[d]$ en moet dus gelijk zijn aan de som van regel fd.

De 6e kolom rechts van de rechthoek — $e\,[fd]$ — wordt gevonden door elk getal van de 5e kolom te vermenigvuldigen met dat, dat in kolom e op dezelfde regel staat. De som van deze kolom is $[de]$.

Ter controle op rekenfouten (die door de veelheid der kleine berekeningen hier licht optreden) wordt de waarde van $[de]$ op dezelfde manier nog eens in de 5e en 6e regel onder de rechthoek uitgerekend.

De waarden van $\bar{x}$ en $\bar{y}$ en van $[x,^2]$ en $[y,^2]$ worden nu op de reeds eerder beschreven manier berekend uit $\bar{d} = [d] : n$ en $[d,^2] = [d^2] - \bar{d}\,[d]$ en uit dezelfde formules met e. De waarde van $[x, y,]$ wordt gevonden uit de stelling

$$[d,e,] = [de] - n\overline{de}$$

of

$$[d,e,] = [de] - \bar{d}\,[e]$$

waarna

$$[x,y,] = b_x b_y\,[d,e,]$$

waarin b_x en b_y de klassebreedten van de x- en de y-schaal voorstellen.

De stelling $[x,y,] = [xy] - n\bar{x}\bar{y}$
wordt als volgt bewezen:

$$\begin{aligned}[x,y,] &= [(x - \bar{x})\,(y - \bar{y})] = [xy - x\bar{y} - \bar{x}y + \bar{x}\bar{y}] = \\ &= [xy] - [x]\,\bar{y} - \bar{x}\,[y] + n\bar{x}\bar{y} = \\ &= [xy] - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = \\ &= [xy] - n\bar{x}\bar{y}\end{aligned}$$

De beste schatting voor de regressie-coëfficient B volgt nu uit

$$b = [x,y,] : [x,^2]$$

en de standaardfout van b uit

$$s_b^2 = ([y,^2] : [x,^2] - b^2) : N \qquad \text{waarin } N = n - 2$$

Op de gebruikelijke manier werd hieruit de dubbelzijdige overschrijdingskans $P_2 = 0.07$ afgeleid.

Voor de toets op de rechtlijnigheid der regressie wordt een variantie-analyse toegepast, waarvoor

$$s_1^2 = ([f\hat{y},^2] - b\,[x,y,]) : N_1 \qquad \text{waarin } N_1 = k - 2$$

$$s_2^2 = ([y,^2] - [f\hat{y},^2]) : N_2 \qquad \text{waarin } N_2 = n - k$$

Hiervan is alleen de kwadraatjessom tussen de y-klassen $[f\hat{y},^2]$ nog niet berekend. Om deze te vinden worden nog twee regels onder de correlatie-tabel aangelegd, dus de 7e en 8e:

7. Regel $[fe]^2$ bevat de kwadraten van de getallen in regel $[fe]$.

8. Regel $f\hat{e}^2$ wordt verkregen door elk getal van de vorige regel te delen door het recht daarboven op regel fx staande getal. De som dezer laatste regel is $[f\hat{e}^2]$.

Door hiervan de reeds bekende waarde van $\bar{e}\,[e]$ af te trekken wordt $[f\hat{e},^2]$ gevonden, waarna de gezochte kwadraatjessom tussen de y-klassen volgt uit $[f\hat{y},^2] = b_y^2\,[f\hat{e},^2]$.

De rest van de berekening behoeft geen nadere uitleg.

Becijfering van de kwadraatjes- en productensommen:

$$\bar{d} = [d] : n = 71 : 247 = 0.2875$$
$$[d_,^2] = [d^2] - \bar{d}\,[d] = 1015 - 0.2875 \times 71 = 1015 - 20.4125 = 994.5875$$
$$\bar{e} = [e] : n = -87 : 247 = -0.352$$
$$[e_,^2] = [e^2] - \bar{e}\,[e] = 2451 - 0.352 \times 87 = 2451 - 30.624 = 2420.376$$
$$[d_,e_,] = [de] - \bar{d}\,[e] = 1392 + 0.2875 \times 87 = 1417.0125$$
$$[x_,^2] = b_x^2\,[d_,^2] = 0.1^2 \times 994.5875 = 9.945875$$
$$[y_,^2] = b_y^2\,[e_,^2] = 0.1 \times 2420.376 = 24.20376$$
$$[x_,y_,] = b_x b_y\,[d_,e_,] = 0.1 \times 0.1 \times 1417.0125 = 14.170125$$

Becijfering van de best passende rechtlijnige regressie-vergelijking:

$$b = [x_,y_,] : [x_,^2] = 14.170125 : 9.945875 = 1.425$$
$$\bar{x} = x_w + b_x\bar{d} = 2.145 + 0.1 \times 0.2875 = 2.174$$
$$\bar{y} = y_w + b_y\bar{e} = 2.645 - 0.1 \times 0.352 = 2.610$$
$$a = \bar{y} - b\bar{x} = 2.610 - 1.425 \times 2.174 \doteq -0.488$$
$$y = 1.425\,x - 0.488$$

Toets van de verwachte rechtlijnigheid der regressie:

$$[f\hat{e}_,^2] = [f\hat{e}^2] - \bar{e}\,[e] = 2073.35 - 30.624 = 2042.726$$
$$[f\hat{y}_,^2] = b_y^2\,[f\hat{e}^2] = 0.1^2 \times 2042.726 = 20.427$$
$$s_1^2 = ([f\hat{y}_,^2] - b\,[x_,y_,]) : (k-2) = (20.427 - 1.425 \times 14.17) : (12 - 2) = 0.0225\,^{1)}$$
$$s_2^2 = ([y_,^2] - [f\hat{y}_,^2]) : (n-k) = (24.204 - 20.427) : (247 - 12) = 0.0161\,^{2)}$$
$$s_1^2 : s_2^2 = 0.0225 : 0.0161 = 1.3975$$
$$s_1 : s_2 = \sqrt{1.3975} = 1.18 \text{ met } N_1 = k-2 = 10 \text{ en } N_2 = n-k = 235$$
$$P_1 > 0.1$$
$$\text{of } G^2_{N_1} = N_1\,(s_1^2 : s_2^2) = 10 \times 1.3975 = 14 \text{ met } N_1 = 10, \text{ zodat } P_1 = 0.2$$

Toets van de theoretisch verwachte regressie-coëfficient $b = 1.5$:

$$s_b^2 = ([y_,^2] : [x_,^2] - b^2) : (n-2) = (24.20376 : 9.945875 - 1.425^2) : 245 = 0.001644$$
$$s_b = 0.041$$
$$t = d : s = (1.425 - 1.500) : 0.041 = -0.075 : 0.041 = -1.83$$
$$P_2 = 0.07$$

Indien men de kansdistributie der afwijkingen ten opzichte der gevonden regressie-vergelijking grafisch wil vergelijken met de best passende normale kansdistributie, dient voor elk gegeven meetpaar afzonderlijk de waarde van die afwijking te worden berekend uit de op blz. 87 besproken formule $d = y - y_r$ (in ons voorbeeld, waar $y_r = 1.425x - 0.488$, dus $d = y - 1.425x + 0.488$) waarna hun frequenties worden geteld en in een tredenlijn worden afgebeeld. De best passende normale kanslijn wordt gevonden uit een T-U-tabel, dat is een tabel, die bij een reeks excentriciteiten (T) de kansintensiteit (U) in de standaard-normale kansdistributie vermeldt, dat is de hoogte, die de standaard-normale kanslijn boven die T-waarde heeft. De twee eerste kolommen (T en U) van onderstaande tabel XXVI vormen zulk een tabel. Het oppervlak onder de standaard-normale kanslijn bedraagt 1, n.l. de som van de kansen op alle mogelijkheden samen. Daarbij is de in beeld gebrachte variabele in zijn standaard-deviatie $T = D : S$ uitgedrukt; als die variabele zelf langs de horizontale as is uitgezet, is het oppervlak dus S maal groter, d.w.z. gelijk aan S. Het oppervlak onder de tredenlijn bedraagt bn, want de hoogte van elke trede is het aantal daartoe be-

---

[1]) $s_1 = 0.150$ [2]) $s_2 = 0.127$

horende „individuen" en haar breedte is gelijk aan de gekozen klassebreedte b. Om bij een gegeven tredenlijn de passende kanslijn te vinden, moeten de U-waarden dus met $bn : s$ worden vermenigvuldigd en telkens bij de bijhorende t-waarden uitgezet, hier dus bij die d-waarden, waarvan $d = \pm st$. Volgens deze aanwijzingen zijn de d- en h (= hoogte)-waarden in tabel XXVI berekend, nadat eerst de factor $bn : s$ was becijferd [1]). Voor het tekenen (b.v. op mm-papier) kunnen de verkregen waarden nog met geschikte getallen, samenhangende met de gewenste grootte der tekening, worden vermenigvuldigd ten einde hen in mm uit te drukken.

TABEL XXVI

Berekening van de best passende normale kanskromme voor de d-waarden

$$
\begin{aligned}
s_d^2 &= [d^2] : (n-1) = ([y_r^2] - a\,[x_r y_r]) : (n-1) = \\
&= (24.2034 - 1.425 \times 14.170) : 246 = \\
&= (24.2034 - 20.1923) : 246 = 4.0111 : 246 = \\
&= 0.0163 \\
s_d &= 0.128 \\
bn : s &= 0.1 \times 247 : 0.128 = 193
\end{aligned}
$$

T	U	d $= \bar{d} \pm st$ $= \pm st$	h $= (bn : s)U$ $= 193U$
0	0.399	0	77
± 0.25	0.387	± 0.032	75
± 0.50	0.352	± 0.064	68
± 0.75	0.301	± 0.096	58
± 1.00	0.242	± 0.128	46.7
± 1.25	0.183	± 0.160	35
± 1.50	0.130	± 0.192	25
± 1.75	0.086	± 0.224	16.6
± 2.00	0.054	± 0.256	10.4
± 2.25	0.032	± 0.288	6.2
± 2.50	0.0175	± 0.320	3.4
± 2.75	0.0090	± 0.352	1.7
± 3.00	0.0044	± 0.384	0.85

---

[1]) Indien het rekenkundig gemiddelde van de variabele grootheid niet, zoals hier, 0 bedraagt, worden voor de horizontale as twee kolommen aangelegd: het gemiddelde $-st$ en het gemiddelde $+st$, waarbij telkens één h-waarde behoort.

VI. *Hebben „enige" kinderen een andere correlatie tussen hypomanie en fantasie dan „niet-enige" kinderen?* (blz. 74)

Deze vraag vergt het vergelijken van twee frequentie-vierkanten:

I. *Enige kinderen*

	hypomanie wel	hypomanie niet	
fantasie wel	32	15	47
fantasie niet	37	86	123
	69	101	170

II. *Niet-enige kinderen*

	hypomanie wel	hypomanie niet	
fantasie wel	12	5	17
fantasie niet	17	16	33
	29	21	50

$r = (ad - bc) : \sqrt{pqrs}$ [1])

$$r_1^2 = \frac{(32 \times 86 - 15 \times 37)^2}{47 \times 123 \times 69 \times 101} = \frac{2197^2}{40287789} \qquad r_1 = \frac{2197}{6347} = 0.346$$

$$r_2^2 = \frac{(12 \times 16 - 5 \times 17)^2}{17 \times 33 \times 29 \times 21} = \frac{107^2}{341649} \qquad r_2 = \frac{107}{584} = 0.183$$

$r_1 - r_2 = 0.346 - 0.183 = 0.163$

$$s^2 = n^2 : pqrs\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)$$

$$s_1^2 = \frac{170^2}{47 \times 123 \times 69 \times 101 \times (\frac{1}{32} + \frac{1}{15} + \frac{1}{37} + \frac{1}{86})} = \frac{28900}{40287789 \times 0.1366} = 0.00525$$

$$s_2^2 = \frac{50^2}{17 \times 33 \times 29 \times 21 \times (\frac{1}{12} + \frac{1}{5} + \frac{1}{17} + \frac{1}{16})} = \frac{2500}{341649 \times 0.405} = 0.018$$

$s^2_{1-2} = s_1^2 + s_2^2 = 0.00525 + 0.018 = 0.0233$

$s_{1-2} = 0.153$

$t_{1-2} = d : s = 0.163 : 0.153 = 1.07$

$P_2 = {}^1/_3$

---

[1]) Indien voor elk vierkant reeds was nagegaan, of beide eigenschappen significant gecorreleerd zijn, was voor elk vierkant t_r reeds berekend volgens $t_r = (ad - bc)\sqrt{n : pqrs}$ (zie blz. 48). In dat geval wordt r gemakkelijker gevonden uit $r = t_r : \sqrt{n}$.

DE BETEKENIS VAN DE IN DIT BOEK GEBRUIKTE

SYMBOLEN

Hoofdletters: eigenschappen der kansdistributies.
Kleine letters: meet- en telgetallen en de daarvan afgeleide grootheden (zodat b.v. s een uit de meetgetallen afgeleide schatting voor S is).

x (of y): een gegeven individuele waarde van een variabele grootheid (b.v. een getal, dat door een meting of steekproeftelling verkregen is)

n: het aantal gegeven individuele waarden van een variabele grootheid (b.v. het aantal van de gegeven meetgetallen betreffende een n-maal gemeten afmeting)

$[x]$: (te lezen: „som-x") de som van de n gegeven x-waarden

$\bar{x}$: (te lezen: „x-gemiddeld") het rekenkundig gemiddelde van de n gegeven x-waarden:

$$\bar{x} = [x] : n$$

$x_{,}$: (te lezen: „x-accent") de afwijking van een gegeven x-waarde tot het rekenkundig gemiddelde van de n gegeven x-waarden:

$$x_{,} = x - \bar{x}$$

s_x: de strooiïng van de n gegeven x-waarden (ook genoemd de standaardfout van de enkele meting) waarvoor als definitie geldt:

$$s_x^2 = [x_{,}^2] : N \quad \text{met de valentie } N = n - 1$$

f: een klassefrequentie (het aantal individuen of meetgetallen, die tot eenzelfde klasse behoren)

$[x]_{\mathrm{kl}}$: de som van de f gegeven x-waarden van één klasse

$\hat{x}$: (te lezen: x-klassegemiddeld) het rekenkundig gemiddelde van de f gegeven x-waarden van één klasse:

$$\hat{x} = [x]_{\mathrm{kl}} : f$$

$x_.$: de afwijking van een gegeven x-waarde tot het rekenkundig gemiddelde van de klasse, waartoe die x-waarde behoort:

$$x_. = x - \hat{x}$$

$\hat{x}_/$: de afwijking van een klassegemiddelde tot het totale gemiddelde:

$$\hat{x}_/ = \hat{x} - \bar{x}$$

x_v, y_v, f_v: de theoretisch voor x, y of f verwachte waarde

y_r: de waarde voor y, die aan de best passende regressievergelijking voldoet

d: de afwijking van een gegeven x-, y of f-waarde tot die, die theoretisch verwacht wordt:

$$d = x - x_v$$

dus bij de best passende regressievergelijking, waar y_r de theoretisch verwachte waarde voor y is:

$$d = y - y_r$$

en bij frequenties:

$$d = f - f_v$$

M: het centrum van een kansdistributie

S: de standaard-deviatie van een normale kansdistributie

D: de deviatie van een gegeven individuele waarde, dat is haar afwijking tot het centrum:

$$D = x - M$$

T: de excentriciteit van een individuele waarde in een normale kansdistributie, dat is haar deviatie gedeeld door de standaard-deviatie:

$$T = D : S$$

G: de simultane excentriciteit van een combinatie van individuele waarden van evenveel variabele grootheden, die elk een normale kansdistributie hebben. Hiervoor geldt als definitie:

$G^2 = [T^2]$ met de valentie N = het aantal T-waarden

P_1: de eenzijdige overschrijdingskans

P_2: de dubbelzijdige overschrijdingskans

OVERZICHT VAN DE IN DIT BOEK GEBRUIKTE

FORMULES

I. Steekproeftellingen

A. *naar één kenmerk:*
voor de best passende normale kansdistributie voor het aantal individuen met dit kenmerk, die in een steekproef van N individuen voorkomen, geldt, als de kans op dat kenmerk P bedraagt en de kans op het afwezig zijn van dat kenmerk Q (zodat $P + Q = 1$) . . . . . $M = PN$
en $S^2 = PQN$

B. *in k klassen:*
a. als de te toetsen theorie de klassekansen aangeeft . . $G^2 = [d^2 : f_v]$ met $N = k - 1$
b. als daartoe eerst c constanten bepaald moeten worden $g^2 = [d^2 : f_v]$ met $N = k - c - 1$

C. *in een frequentie-vierkant:*

$$t = (ad - bc)\sqrt{n : pqrs} \quad \text{met } N = n - 3$$
$$r = (ad - bc) : \sqrt{pqrs}$$
$$s^2_r = n^2 : pqrs\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)$$

II. Meetgetallen

A. *Voor een n maal gemeten grootheid*
is de beste schatting $\bar{x} = [x] : n$
met standaardfout . . $s^2_{\bar{x}} = s^2_x : n$ met $N = n - 1$

B. *Voor het verschil tussen twee gemeten grootheden X en IJ*
is de beste schatting $\bar{x} - \bar{y}$
met standaardfout . . $s^2_{\bar{x}-\bar{y}} = s^2_{\bar{x}} + s^2_{\bar{y}}$ met $N = n_x + n_y - 2$

C. *Voor de vraag, of S_x groter is dan S_y*
wordt de strooiïngsverhouding $s_x : s_y$ beoordeeld:
a. als N_x en N_y elk zijn klein: met behulp van de strooiïngsdiagrammen
b. als $N_x = 1$ is, uit . . . . . $t = s_x : s_y$ met $N = N_y$
c. als N_x klein en N_y groot is, uit . . . . . $g^2_N = N_x\, s^2_x : s^2_y$ met $N = N_x$
of $g^2_N = N_y\, [x_,^2] : [y_,^2]$ met $N = N_x$
maar wordt het verschil $s_x - s_y$ beoordeeld:
d. als N_x en N_y beide groot zijn, uit . . . . . $s^2_{sx} = s^2_{\bar{x}} : 2$ met $N = N_x$

D. *Variantie-analyse:*

$$[\text{residu}] = [\text{totaal}] - [A] - [B] - \text{enz.} =$$
$$= [x_{,}^2] - [f\hat{x}_{,}^2]_A - [f\hat{x}_{,}^2]_B - \text{enz.}$$
$$N_{\text{residu}} = N_{\text{totaal}} - N_A - N_B - \text{enz.}$$

waarin

$$N_{\text{totaal}} = n - 1$$
$$N_A = \text{het aantal } A\text{-klassen minus } 1.$$

Invloed van A te beoordelen met behulp van

$$s^2_1 : s^2_2 = ([A] : N_A) : ([\text{residu}] : N_{\text{residu}})$$

E. *Best passende regressievergelijking van willekeurige gedaante:*
te toetsen met de verhouding $s_1 : s_2$ bij

$$s^2_1 = [f\hat{d}^2] : N_1 \qquad \text{met } N_1 = k - c - 1$$
$$s^2_2 = [y_{.}^2] : N_2 \qquad \text{met } N_2 = n - k$$

F. *Rechtlijnige regressie* ($y = A + Bx$)*:*
Als voldaan is aan de voorwaarden, genoemd op blz. 82 e.v., volgen

a. de best passende waarden voor de constanten uit:

$$b = [x_{,}y_{,}] : [x_{,}^2]$$
$$a = \bar{y} - b\bar{x}$$

b. de controle op rechtlijnigheid uit de verhouding $s_1 : s_2$ bij

$$s^2_1 = ([f\hat{x}_{,}^2] - b\,[x_{,}y_{,}]) : N_1 \qquad \text{met } N_1 = k - 2$$
$$s^2_2 = ([y_{,}^2] - [f\hat{y}_{,}^2]) : N_2 \qquad \text{met } N_2 = n - k$$

c. de standaardfout van b uit:

$$s_b^2 = ([y_{,}^2] : [x_{,}^2] - b^2) : N_b \qquad \text{met } N_b = n - 2$$

III. Voor het cijferwerk werd gebruik gemaakt van de rekenkundige stellingen:

$$[x_{,}^2] = [x^2] - \bar{x}\,[x]$$
$$[x_{,}y_{,}] = [xy] - \bar{x}\,[y]$$
$$[f\hat{x}_{,}^2] = [f\hat{x}^2] - \bar{x}\,[x]$$
$$f\hat{x}^2 = [x]^2_{\text{kl}} : f$$

DE IN DIT BOEK GEBRUIKTE

NOMOGRAMMEN

Het *t-N-P-nomogram*,
bedoeld om bij een uit de gegevens berekende schatting voor de excentriciteit, t, samen met de valentie, N, de dubbelzijdige overschrijdingskans, P_2, af te lezen.

Het G^2-N-P-*nomogram*,
bedoeld om bij een uit de gegevens berekende simultane excentriciteit, G, samen met de valentie N_1, de (eenzijdige) overschrijdingskans, P_1, af te lezen.

Het *eerste strooiïngsnomogram,*
bedoeld om bij de valenties N_1 en N_2 van twee uit de gegevens berekende strooiïngen s_1 en s_2 die strooiïngsverhouding $s_1 : s_2$ af te lezen, waarvan de eenzijdige overschrijdingskans op een grotere waarde van die verhouding gelijk is aan

$$\mathbf{P_1 = 0.20}$$

Het *tweede strooiïngsnomogram,*
bedoeld om bij de valenties N_1 en N_2 van twee uit de gegevens berekende strooiïngen s_1 en s_2 die strooiïngsverhouding $s_1 : s_2$ af te lezen, waarvan de eenzijdige overschrijdingskans op een grotere waarde van die verhouding gelijk is aan

$$\mathbf{P_1 = 0.05}$$

Het *derde strooiingsnomogram,*
bedoeld om bij de valenties N_1 en N_2 van twee uit de gegevens berekende strooiingen s_1 en s_2 die strooiingsverhouding $s_1 : s_2$ af te lezen, waarvan de eenzijdige overschrijdingskans op een grotere waarde van die verhouding gelijk is aan

$$P_1 = 0.01$$

Het *vierde strooiïngsnomogram*,
bedoeld om bij de valenties N_1 en N_2 van twee uit de gegevens berekende strooiïngen s_1 en s_2 die strooiïngsverhouding $s_1 : s_2$ af te lezen, waarvan de eenzijdige overschrijdingskans op een grotere waarde van die verhouding gelijk is aan

$$\mathbf{P_1 = 0.001}$$