

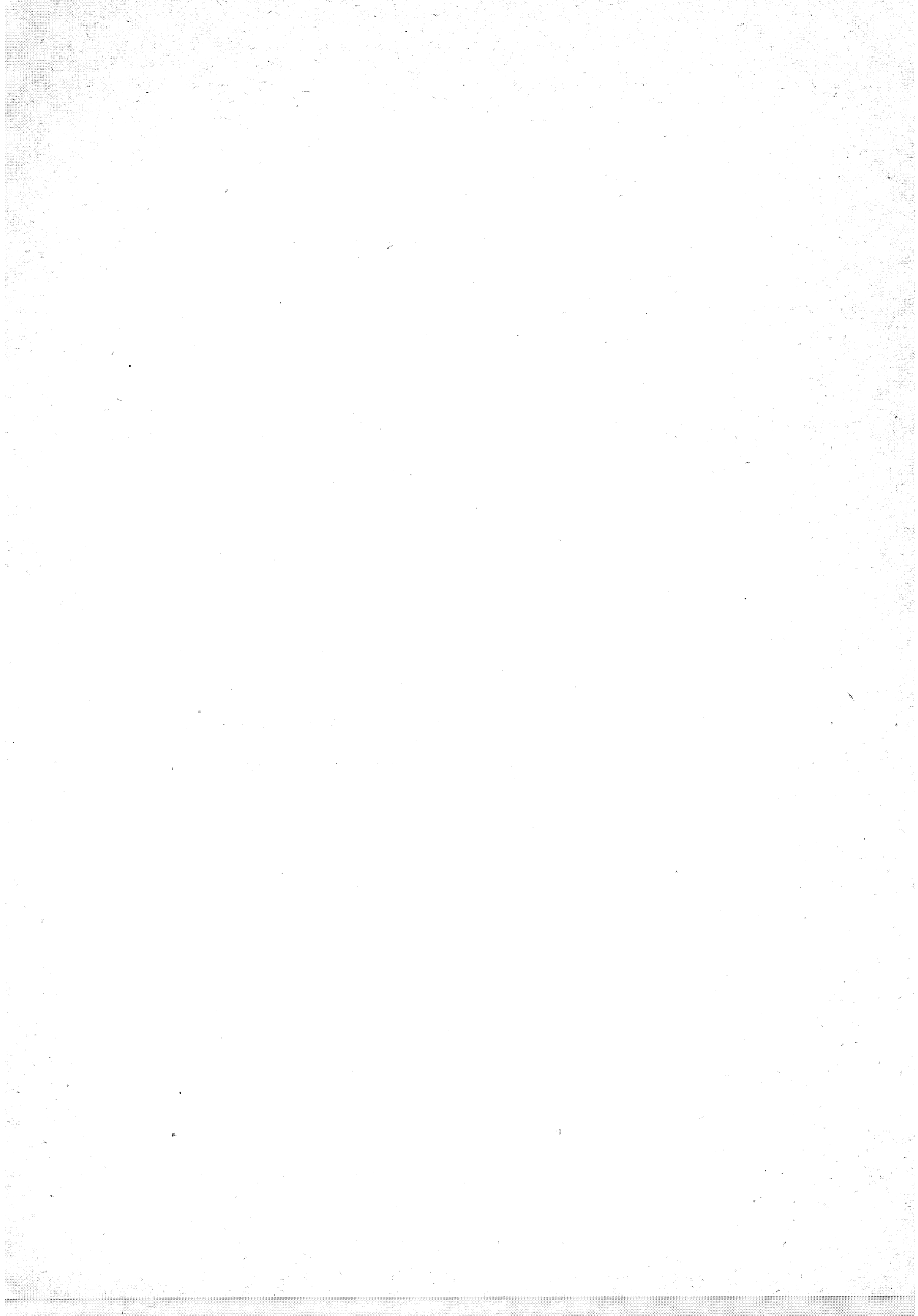
SECTIE VOOR ANALYTISCHE CHEMIE
VAN DE
KONINKLIJKE NEDERLANDSCHE CHEMISCHE VERENIGING

Symposium over

De toepassing van statistiek bij de chemische analyse

gehouden te Amsterdam
op 31 maart 1960

- I. A. R. BLOEMENA, Enige statistische beschouwingen
- II. M. J. MAURICE, Nauwkeurigheid en juistheid van chemische analyse-methodes
- III. J. AGTERDENBOS, Het splitsen van de totale fout in verschillende componenten



Some statistical procedures are described, which arise in considerations about accuracy of analytical-chemical methods.

A discussion is given of the troubles involved in the interpretation of duplicate observations, obtained by carrying out the two analyses simultaneously.

31 : 543

Van het eerste contact af dat U had met de analytische scheikunde zult U reeds bekend zijn geweest met (misschien ook wel: doordrongen zijn geweest van) het feit dat bij een chemische bepaling, in duplo uitgevoerd, de beide resultaten in het algemeen niet exact gelijk zijn. Bij het beoordelen van dergelijke resultaten rijst dus direct de vraag wat men nu kan concluderen ten aanzien van de kwantiteit, welke men wilde leren kennen. In de regel zal men het gemiddelde van de beide waarnemingen als een „juister” getal beschouwen dan ieder der waarnemingen apart. Toch blijft men terecht twijfelen of ook dit gemiddelde *de* kwantiteit voorstelt. Deze twijfel wordt in zeer vele publikaties en rapporten over analytisch-chemische onderwerpen tot uitdrukking gebracht door bij het gemiddelde grenzen op te geven: $\bar{x} \pm a$.

Dit is een eenvoudig voorbeeld van een geval waarbij men een oordeel moet geven in een onzekere situatie. De praktijk van de analytische chemie biedt evenwel talloze gecompliceerde situaties van soortgelijke aard. Ik noem U bijvoorbeeld het geval, dat men van twee analysemethodes wil onderzoeken of de ene nauwkeuriger is dan de andere. Doet men met beide methodes een aantal waarnemingen van eenzelfde grootte, dan zal in het algemeen de variabiliteit van de waarnemingen volgens methode 1 verschillen van die volgens methode 2. Is dit verschil nu toevallig, of is er een systematisch verschil in nauwkeurigheid tussen beide methodes? In deze situatie gebruikt men in de praktijk veelal de *hm-methode*. Ik citeer U uit een voordracht van Prof. van Dantzig¹, mijn kortgeleden overleden leermeester:

De hm-methode bestaat daarin, dat men de rechterhand aan de kin brengt, de duim rechts, de andere vingers links, dan de lippen oopenperst en de wenkbrauwen samentrekt, vervolgens „Hm” zegt, en ten slotte een oordeel ten beste geeft.

En dan even verder zegt hij:

Indien de methode wordt toegepast door iemand met een grondige kennis en uitgebreide ervaring omtrent het te beoordelen object, kan de bedoelde keelklank een snelle innerlijke samenvatting van deze kennis en ervaring tot uitdrukking brengen en tot goede resultaten leiden. Bovendien kan wel geen onzer nalaten van tijd tot tijd eens te hummen, daar niemand in staat is, al zijn onderwerpen te systematiseren en aan een grondige analyse te onderwerpen. Als methode echter heeft zij het bezwaar, dat zij bij haar beste beoefenaren een onfeilbaarheid van oordeel suggereert, dat zij geen garantie biedt tegen stem-

* Rapport S269 van de Statistische Afdeling van het Mathematisch Centrum.

** Voordracht gehouden op het symposium over „De toepassing van statistiek bij de chemische analyse”, op 31 maart 1960 te Amsterdam.

mings- en aandachtsfluctuaties bij degenen die haar toepassen en dat haar betrouwbaarheid in sterke mate afhangt van de kwaliteiten der beoordelaars.

Kortom, met de hm-methode zullen behalve juiste conclusies ook onjuiste conclusies verkregen worden. Het is niet bekend welke conclusies onjuist zijn. De frequentie waarmee onjuiste conclusies worden verkregen is eveneens niet bekend, is ook bij een beoordelaar niet constant in de tijd en varieert sterk van beoordelaar tot beoordelaar.

Het doel van de mathematische statistiek is nu methodieken te verschaffen die ook leiden tot conclusies die soms juist, soms onjuist zijn, maar waarbij de frequentie van de onjuiste conclusies bekend is, zodat men weliswaar nog steeds niet weet welke conclusies onjuist zijn, maar wel weet *hoeveel* er onjuist zijn. Voorts zal op grond van gegeven waarnemingsuitkomsten en bij een juiste toepassing van dezelfde statistische methodes een ieder tot eenzelfde uitspraak komen, dit in tegenstelling tot de hm-methode.

De statistiek bedient zich van methodes uit de kansrekening. Kansrekening is weer een onderdeel van de maattheorie, die op zijn beurt weer een onderdeel van de zuivere wiskunde is. In de theoretische publikaties over statistiek vindt U dan ook dat ruim gebruik gemaakt wordt van het wiskundig formalisme, maar voor de grondgedachte van de statistiek en voor het hanteren van een aantal methodes op dit gebied is het niet nodig veel wiskunde te kennen.

Bezien we eerst eens het geval dat men n onafhankelijke waarnemingen heeft gedaan van een onbekende grootte ξ . Men beschikt dus over n getallen x_1, \dots, x_n , die men ieder op kan vatten als een waarneming van een grootte x : een waarneming van ξ met de gegeven waarnemingsmethode. Men opent nu de weg voor de kansrekening door te zeggen dat x een kansgrootte is, en dat x dus een (zij het onbekende) kansverdeling bezit.

Symbolen van kansgrootheden worden aangegeven door onderstreping, om duidelijk te laten uitkomen dat men hier niet met een getal te doen heeft.

Nu kan men dan ook de waarnemingsmethode beschrijven door de kansverdeling. In de praktijk heeft men evenwel altijd met een eindig aantal waarnemingen en dus met een frequentieverdeling te maken. In figuur 1 is b.v. een frequentieverdeling van n waarnemingen aangegeven.

Indien r van de n waarnemingen tussen de grenzen 103 en 104 liggen dan heeft het desbetreffende staafje dus een hoogte evenredig aan r/n . Een kansverdeling

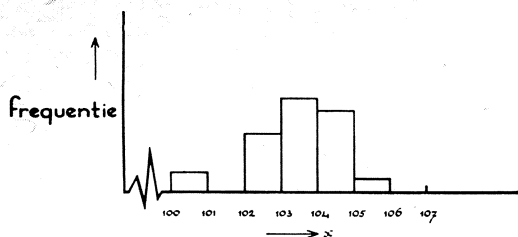


Fig. 1. Een frequentieverdeling.

is op te vatten als een frequentieverdeling voor on-eindig veel waarnemingen.

Verreweg de bekendste kansverdeling is de *normale* verdeling, die men in het algemeen afgebeeld vindt door de klokvormige kromme (figuur 2), en die wordt gekenmerkt door twee parameters, nl. de *verwachting* (of het *gemiddelde*) μ en de *standaardafwijking* (of *spreiding*) σ . Zoals in figuur 2 is aangegeven is σ de afstand van het gemiddelde tot een van de buigpunten van de kromme. Men noemt σ^2 de variantie. De spreiding σ is nu een bijzonder geschikte maat om de nauwkeurigheid van een waarnemingsmethode te

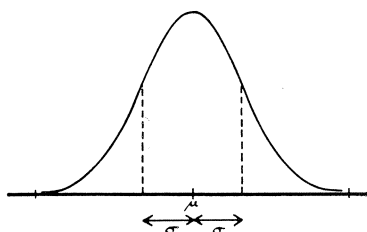


Fig. 2. De kansdichtheid van een normale verdeling.

karacteriseren, hoe groter σ hoe kleiner de nauwkeurigheid. Heeft men een normale verdeling met parameters μ en σ , dan is de kans dat een waarneming uit die verdeling tussen twee gegeven constanten a en b zal liggen, gelijk aan het oppervlak onder de kromme tussen de verticale lijnen door a en b : figuur 3. Het totale oppervlak onder de kromme is gelijk aan 1. Kiest men voor a de waarde $\mu - 2\sigma$ en voor b $\mu + 2\sigma$, dan ziet men direct dat de kans zeer gering is (kleiner dan 0.05) dat een waarneming buiten deze grenzen zal liggen. Er zijn uitgebreide tabellen waarin dergelijke kansen af te lezen zijn, die dus alleen kunnen worden gebruikt als μ en σ bekend zijn.

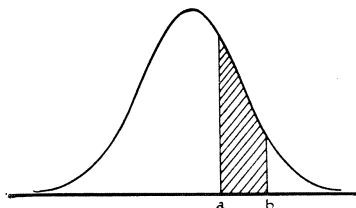


Fig. 3. De kans dat een waarneming tussen a en b zal liggen is gelijk aan het gearceerde oppervlak.

Meestal kent men μ en σ niet. Beschikt men wel over waarnemingen x_1, \dots, x_n van die normaal verdeelde grootheid, dan kan het probleem gesteld worden μ en σ te leren kennen. Nu kan men nooit de exacte waarde van de μ en σ berekenen uit een reeks waarnemingen, men kan ten hoogste *schattingen* van μ en σ berekenen. Als schattingen van μ neemt

men in de regel het *gemiddelde* van de waarnemingen,

$$\text{dus } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ en als schatting van } \sigma^2$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left\{ \sum_i x_i^2 - \frac{(\sum x_i)^2}{n} \right\} \quad (1)$$

De tweede uitdrukking voor s^2 is de vorm waarin deze op de meest efficiënte wijze kan worden berekend; bij gebruik van een rekenmachine kan de berekening achter elkaar doorgaan op de machine. De eerste uitdrukking geeft meer inzicht in wat s^2 is, maar is ongeschikt voor rekendoeleinden.

Als men een s^2 berekent dan zal de nauwkeurigheid waarmee σ^2 geschat wordt afhangen van het aantal waarnemingen. In het statistisch jargon spreekt men vaak over *vrijheidsgraden*: berekent men een s^2 uit n waarnemingen, dan is $n-1$ het aantal vrijheidsgraden.

Dat het hier om *schattingen* gaat is direct in te zien: als men nog eens n waarnemingen uitvoert, krijgt men in het algemeen een andere \bar{x} en een andere s^2 .

Een schatting van een parameter is wel heel belangrijk, maar meestal wil men meer weten. Dat kan. Stel eens dat σ^2 bekend is, en men over één waarneming x beschikt. Die x is dan tevens de schatting van μ . Op grond van kantheoretische overwegingen kan men nu de uitspraak doen, dat *behoudens een onbetrouwbaarheid van ten hoogste 0.05* geldt

$$x - 2\sigma < \mu < x + 2\sigma. \quad (2)$$

Hoe moet dit worden geïnterpreteerd? Eerst moet geconstateerd worden dat μ en σ constanten zijn, resp. onbekend en bekend, en dat als de waarneming x is verricht, deze ook bekend is, en dat dus dan de uitspraak (2) of juist is, of onjuist. Er is dus geen sprake meer van een kans dat deze onjuist zal zijn. Wat nu de zinsnede „behoudens een onbetrouwbaarheid van 0.05” betekent, is, dat als men uitspraken als (2) vele malen doet bij steeds andere problemen en waarnemingen, dat dan de fractie foute uitspraken gemiddeld ≤ 0.05 is. Op deze wijze verdisconteert de kansrekening de onzekerheid in de waarnemingen, waarvan men uitging, in uitspraken, waarvan men niet weet welke juist en onjuist zijn, maar wel hoe vaak men tot foute uitspraken komt.

Een interval, aangegeven in de formule (2) noemt men een *betrouwbaarheidsinterval*, met een onbetrouwbaarheid van 0.05. Doet men n onderling onafhankelijke waarnemingen en is σ bekend dan kan men ook een dergelijk interval opstellen. Daar de spreiding van een gemiddelde van n onderling onaf-

hankelijke waarnemingen $\frac{\sigma}{\sqrt{n}}$ is, krijgt men

$$\bar{x} - \frac{2\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{2\sigma}{\sqrt{n}} \quad (3)$$

terwijl als ook σ onbekend is en men de schatting s gebruikt, rekening moet worden gehouden met de onzekerheid in deze schatting. Men heeft dan

$$\bar{x} - \frac{ts}{\sqrt{n}} < \mu < \bar{x} + \frac{ts}{\sqrt{n}} \quad (4)$$

waarbij t een constante is, die afhangt van n en van de gekozen onbetrouwbaarheid. In tabel I zijn enkele waarden (z.g. *kritieke waarden*) van t gegeven.

Tabel I. Enige kritieke waarden van t .

Aantal vrijheids- graden	Tweezijdige onbetrouwbaarheid	
	0.01	0.05
1	31.8	12.7
2	7.0	4.3
3	4.5	3.2
4	3.7	2.8
5	3.4	2.6
7	3.0	2.4
9	2.8	2.3
19	2.5	2.1
∞	2.3	2.0

Voor de onbetrouwbaarheid van een betrouwbaarheidsinterval kan men iedere gewenste waarde kiezen, maar in elk geval mag er geen misverstand heersen, welke waarde in een bepaald geval gekozen is. Tegen deze regel wordt in veel technische publikaties nogal eens gezondigd!

Als laatste in deze reeks betrouwbaarheidsintervallen noem ik U nog die voor een verschil tussen twee gemiddelden. Stel dat er n onderling onafhankelijke waarnemingen x_1, \dots, x_n gegeven zijn van een normaal verdeelde kansgrootheid met onbekend gemiddelde μ_1 en onbekende σ , en evenzo m onderling onafhankelijke waarnemingen van een normaal verdeelde kansgrootheid met onbekend gemiddelde μ_2 en met dezelfde, onbekende σ . Dan kan men dus met behulp van formule 1 een schatting s_1^2 van σ^2 berekenen uit de x -steekproef, en een schatting s_2^2 van σ^2 uit de y -steekproef. Combineren van deze schattingen geeft een schatting:

$$s_c^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2},$$

nu met $n+m-2$ vrijheidsgraden. Het betrouwbaarheidsinterval voor $\mu_1 - \mu_2$ is nu

$$\bar{x} - \bar{y} - t_{sc} \sqrt{\frac{m+n}{mn}} < \mu_1 - \mu_2 < \bar{x} - \bar{y} + t_{sc} \sqrt{\frac{m+n}{mn}} \quad (5)$$

waar nu weer de kritieke waarde t voor een onbetrouwbaarheid 0.05 of 0.01 kan worden afgelezen uit tabel I, bij $n+m-2$ vrijheidsgraden.

Stel dat men in de volgende situatie verkeert: men heeft n waarnemingen x_1, \dots, x_n van een grootheid, die een normale verdeling bezit met onbekende μ en σ . Verder is er een getal μ_0 gegeven, waarbij men zich afvraagt wat op grond van de waarnemingsresultaten gezegd zou kunnen worden over de premisse dat $\mu = \mu_0$. Men noemt dit het stellen van een hypothese. Een statistische techniek, waarmee men de hypothese onderzoekt, noemt men een *toets*. Zo is in dit geval de hypothese $\mu = \mu_0$ en de relevante toets de toets van *Student*.

Student is de schuilnaam van *W. S. Gosset* (1876-1937), auteur van een aantal zeer belangrijke artikelen over statistische onderwerpen.

Deze toets wordt als volgt uitgevoerd. Bereken uit de waarnemingen \bar{x} -en s^2 en bereken dan het betrouwbaarheidsinterval (4). Er kunnen zich twee gevallen voordoen:

I: μ_0 ligt in het betrouwbaarheidsinterval. In dit

geval is de conclusie: er is geen reden te concluderen dat μ verschilt van μ_0 .

N.B.: Een onjuiste formulering van deze conclusie is hier: het blijkt dat $\mu = \mu_0$. Immers het kan best zo zijn dat $\mu \neq \mu_0$, maar dat dit verschil $\mu - \mu_0$ zo klein is, dat dit geheel wordt gemaskeerd door de spreiding in de waarnemingen.

II: μ_0 ligt buiten het betrouwbaarheidsinterval. Men kan nu de conclusie trekken: μ verschilt van μ_0 . Deze conclusie heeft nu een onbetrouwbaarheid gelijk aan de onbetrouwbaarheid α van het betrouwbaarheidsinterval. Men moet dit als volgt interpreteren. Stel dat men in zeer vele problemen en met steeds andere waarnemingen deze toets uitvoert in gevallen, waarbij exact aan de hypothese is voldaan, dus waar steeds $\mu = \mu_0$, dan is het aantal keren dat men een foute uitspraak doet, en waar men dus tot de conclusie komt: $\mu \neq \mu_0$, ongeveer gelijk aan α maal het aantal uitspraken.

Naast deze toets, die op één serie waarnemingen betrekking heeft, treft men in beschouwingen over nauwkeurigheid en juistheid van analysemethodes ook vrij veelvuldig toetsen aan voor het vergelijken van de gemiddelden en de spreidingen van twee reeksen waarnemingen. De eerste reeks waarnemingen, bijvoorbeeld met analysemethode A verkregen,

$$x_1, \dots, x_n$$

onderstellen we te bestaan uit onderling onafhankelijke waarnemingen van eenzelfde kansgrootheid x , welke een normale verdeling heeft met onbekend gemiddelde μ_1 en onbekende spreiding σ_1 . De tweede reeks, bij voorbeeld de waarnemingen verkregen bij analyse van dezelfde oplossing of hetzelfde preparaat met analysemethode B:

$$y_1, \dots, y_m$$

onderstellen we evenzo te bestaan uit onderling onafhankelijke waarnemingen van eenzelfde kansgrootheid y welke een normale verdeling bezit met onbekend gemiddelde μ_2 en onbekende spreiding σ_2 .

Interesseert U zich ervoor om te onderzoeken of $\mu_1 = \mu_2$ dan kan dit met de toets van *Student* alleen als óók de beide spreidingen σ_1 en σ_2 gelijk zijn. In dat geval berekent men het betrouwbaarheidsinterval (5). Twee gevallen kunnen zich nu voordoen:

I: De waarde 0 ligt in het betrouwbaarheidsinterval. Dan is de conclusie: er is geen reden te onderstellen dat de beide gemiddelden verschillen.

II: De waarde 0 ligt *niet* in het betrouwbaarheidsinterval. Men trekt dan de conclusie: er bestaat een systematisch verschil tussen de beide gemiddelden. Deze conclusie heeft dan een onbetrouwbaarheid gelijk aan die van het betrouwbaarheidsinterval. De interpretatie van deze onbetrouwbaarheid is dus weer als volgt. Pas de toets een groot aantal malen toe bij steeds andere waarnemingen in gevallen, waarin steeds exact geldt dat $\mu_1 = \mu_2$. De fractie uitspraken welke luiden: $\mu_1 \neq \mu_2$ (dus foute uitspraken) is nu gemiddeld ten hoogste gelijk aan de opgegeven onbetrouwbaarheid.

Een andere toets is die voor de hypothese, dat

$$\sigma_1 = \sigma_2,$$

dus dat de beide methoden even nauwkeurig zijn

(ongeacht het feit of $\mu_1 = \mu_2$ of $\mu_1 \neq \mu_2$). Bij de hierboven genoemde onderstellingen is de relevante toets de F -toets, genoemd naar *Sir Ronald Fisher*, een van de grondleggers van de moderne mathematische statistiek.

De uitvoering is als volgt: bereken s_1^2 volgens formule (1) uit de x -waarnemingen, bereken s_2^2 volgens dezelfde formule uit de y -waarnemingen, en bereken

$$F = \frac{s_1^2}{s_2^2}.$$

Als de hypothese $\sigma_1 = \sigma_2$ waar is, zal F ongeveer de waarde 1 bezitten. Verschilt F veel van 1, dan vat men dit als een aanwijzing op dat de hypothese niet vervuld is. De kansrekening biedt een middel om kritieke waarden F_α te berekenen, waarmee de gevonden waarde F vergeleken wordt. Weer kunnen zich twee gevallen voordoen.

I: F ligt in het interval met grenzen $\frac{1}{F_\alpha}$ en F_α .

Dan kan alleen geconcludeerd worden dat er geen reden is de hypothese te verwerpen.

II: F ligt buiten het interval met grenzen $\frac{1}{F_\alpha}$ en F_α .

In dit geval kan men de conclusie trekken dat $\sigma_1 \neq \sigma_2$ is, en wel dat $\sigma_1 < \sigma_2$ als $F < 1$ en $\sigma_1 > \sigma_2$ als $F > 1$. De totale onbetrouwbaarheid van deze conclusie is nu 2α .

Tabel II. Enkele waarden van $F_{0,025}$.

n \ m	2	3	4	5	10	16	∞
2	648	800	864	800	963	984	1018
3	38.5	39.0	39.1	39.3	39.4	39.4	39.5
4	17.4	16.0	15.4	15.1	14.5	14.3	13.9
5	12.2	10.7	10.0	9.6	8.9	8.7	8.3
10	7.2	5.7	5.1	4.7	4.0	3.8	3.3
15	6.3	4.9	4.2	3.9	3.2	3.0	2.5
25	5.7	4.3	3.7	3.4	2.7	2.4	1.9
∞	5.0	3.7	3.1	2.8	2.1	1.8	1.0

Er zijn zeer uitgebreide tabellen F_α , b.v. Tabel 18 van *Pearson and Hartley*². Voor enkele waarden van m en n en $\alpha = 0.025$ is F_α gegeven in Tabel II.

Behalve het toetsen van hypothesen en betrouwbaarheidsintervallen, omvat de statistiek ook de analyse van een onnauwkeurigheid of van een spreiding, zoals in het volgende voorbeeld.

Stel dat er een analyse, b.v. het bepalen van een gehalte, moet worden uitgevoerd en dat in het desbetreffende laboratorium een aantal analisten zijn, die wat deze analysemethode betreft, allen even vaardig en nauwkeurig werken. Aan een van deze analisten wordt nu de opdracht verstrekt en na zekere tijd produceert deze de gevraagde waarnemingsuitkomst(en). Een dergelijke uitkomst, x , zullen we ons als volgt opgebouwd denken

$$x = \mu + \varepsilon + \delta \quad (6)$$

waarin μ het gehalte is. De ε en de δ zijn samen de afwijking van de uitkomst van de analyse van het werkelijke gehalte. Nu is ε dat deel dat de invloed weerspiegelt van de persoon van de analist, en δ is de invloed van de waarnemingsonnauwkeurigheid die

inherent is aan de analysemethode. Als dezelfde analist de bepaling herhaalt dan zal ε dezelfde blijven, maar δ een andere waarde aannemen. Men kan δ dus als een kansgrootte zien, hetgeen we in het model (6) aangegeven door δ te onderstrepen.

$$\underline{x} = \mu + \varepsilon + \underline{\delta}. \quad (6')$$

We zullen verder nog onderstellen dat $\underline{\delta}$ een normale kansverdeling heeft met gemiddelde 0 en spreiding σ_w .

Indien een andere analist nu hetzelfde gehalte bepaalt zal deze analist door minuscule verschillen in techniek en in zijn handelingen een andere ε -component introduceren. We onderstellen nu dat de ε voor de verschillende analisten een normale verdeling heeft met gemiddelde 0 en spreiding σ_a , dus

$$\underline{x} = \mu + \underline{\varepsilon} + \underline{\delta}. \quad (6'')$$

Als I analisten de bepaling ieder J keer uitvoeren, dan is de spreiding in het gemiddelde van de IJ waarnemingen gelijk aan

$$\sqrt{\frac{\sigma_a^2}{I} + \frac{\sigma_w^2}{IJ}} \quad (7)$$

waaruit dus al direct volgt dat als σ_a^2 niet veel kleiner is dan σ_w^2 het geen zin heeft elke analist de bepaling vele malen te laten herhalen.

Kijkt men naar de J waarnemingen van één van de analisten en berekent men hieruit volgens formule (1) s^2 , dan is deze s^2 een schatting van σ_w^2 . Over de σ_a^2 krijgt men geen informatie door de waarnemingen van één analist te bestuderen. Men krijgt een te optimistische kijk betreffende de waarnemingsonnauwkeurigheid, als men alleen de onderlinge verschillen van de waarnemingen van één analist in de beschouwing betreft, daardoor schakelt men wat men noemt de spreiding-tussen-analisten, σ_a , uit, terwijl met deze spreiding volgens (7) wel degelijk rekening gehouden moet worden.

Hoe kan men iets over de grootte van σ_a en σ_w te weten komen? Een eenvoudig experiment kan hier het antwoord verschaffen. Ieder van I analisten herhaalt J maal de bepaling, dan geldt dus voor de j -de waarneming van de i -de analist:

$$x_{ij} = \mu + \underline{\varepsilon}_i + \underline{\delta}_{ij}. \quad (8)$$

Men kan nu met als basis het model (8) de volgende schattingen s_w^2 en s_a^2 van σ_w^2 en σ_a^2 opstellen:

$$s_w^2 = \frac{1}{I(J-1)} \left\{ \sum_{i,j} x_{ij}^2 - \frac{\left(\sum_i \left(\sum_j x_{ij} \right)^2 \right)}{I} \right\}$$

$$s_a^2 = \frac{1}{I-1} \left\{ \frac{\sum_i \left(\sum_j x_{ij} \right)^2}{J} - \frac{\left(\sum_{i,j} x_{ij} \right)^2}{IJ} \right\} - J s_w^2 \quad (9)$$

Men kan ook nog toetsen of $\sigma_a = 0$ is, wat dus zou inhouden dat er geen verschillen tussen analisten optreden.

Dit voorbeeld betreft een eenvoudige situatie. De statistiek kan veel ingewikkelder situaties aan, bijvoorbeeld waar gevraagd wordt de invloed van een aantal factoren op de grootte van de uitkomst van een pro-

cédé te onderzoeken. Deze en soortgelijke problemen worden meestal besproken onder de naam *variantie-analyse*.

Zojuist heb ik iets gezegd over een te rooskleurig beeld dat men van de nauwkeurigheid zou kunnen krijgen, door een bron van spreiding (in dit geval waren het de verschillen tussen analisten) uit te schakelen en alleen het effect van de overblijvende factoren te bestuderen. Deze situatie treffen wij ook aan bij het beoordelen van duplo-waarnemingen. Er zijn maar weinig gevallen, waarin duplo-waarnemingen, verkregen door eenzelfde bepaling *gelijktijdig* tweemaal uit te voeren, een enigszins betrouwbaar beeld geven van de werkelijke aan de analyse verbonden onnauwkeurigheden. In het algemeen wordt door het *gelijktijdig* (of het vlak na elkaar) uitvoeren van de bepalingen een aanzienlijk aantal bronnen van spreiding op vrij efficiënte manier geëlimineerd, terwijl men deze uitschakeling niet kan bereiken, indien geruime tijd verloopt tussen de beide waarnemingen.

Een ander aspect van duplo-waarnemingen is de zeer voor de hand liggende mogelijkheid van beïnvloeding van de tweede waarneming door de eerste uitkomst. Men stelt bij alle statistische methoden de eis, dat de waarnemingen, waarop deze worden toegepast, onderling onafhankelijk zijn. Het is onze ervaring, dat duplo-waarnemingen in deze te wensen overlaten.

Er zijn nog meer moeilijkheden bij het beoordelen van duplo-uitkomsten. Ieder van U kent de uitdrukking „de duplo's kloppen”. In het algemeen betekent dit dat de beide waarnemingen minder dan $a\%$ van elkaar verschillen. In zeer vele bedrijven en laboratoria is het regel, dat bepalingen, waarbij de duplo's niet kloppen, opnieuw moeten worden gedaan. Echter in zeer weinig gevallen kent men de consequenties van deze regels en weet men bij voorbeeld hoe groot de kans is dat de duplo-uitkomsten meer dan $a\%$ verschillen, als beide uitkomsten toch goede en gave waarnemingen zijn, verkregen bij een bepaling, waarbij geen misstappen zijn begaan. Is deze kans klein, bijvoorbeeld 0.05, dan kan tegen deze regel weinig bezwaar bestaan, doch is deze kans groot, bijvoorbeeld 0.25, dan betekent dit dat of relatief veel waarnemingen worden herhaald zonder dat hiervoor een dringende reden aanwezig is, of men dwingt degenen, die onderworpen zijn aan deze regel, tot beïnvloeding van de waarnemingen. Indien men uit dergelijke waarnemingen een schatting van de spreiding berekent, zal men deze onderschatten, zodat men een vertekend beeld van de waarnemingsnauwkeurigheid zal krijgen. Een statisticus zal met een dergelijk, geselecteerd, materiaal vaak minder kunnen beginnen dan met de oorspronkelijke, dus ook de niet kloppende, duplo-uitkomsten.

Er zijn gevallen denkbaar waar een regel dat duplo-waarnemingen slechts minder dan $a\%$ van elkaar mogen verschillen, ook *systematische* fouten introduceert, waar deze er oorspronkelijk niet waren. Tot nu toe is alleen de normale kansverdeling ter sprake gekomen, maar er zijn daarnaast ook zeer veel andere kansverdelingen. Zo o.a. die van het type dat in figuur 4 is weergegeven door de kansdichtheid.

Indien de beide duplo-waarnemingen uit een dergelijke verdeling komen, is de kans dat zij een verschil groter dan een gegeven getal vertonen, groter wanneer

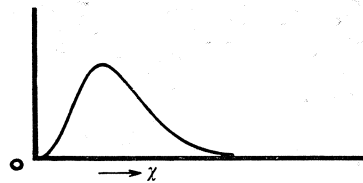


Fig. 4. Kansdichtheid van χ^2 -verdeling met tien vrijheidsgraden.

het gemiddelde van de beide waarnemingen hoog is, dan wanneer dit gemiddelde laag is. Dat wil dus zeggen dat in het algemeen een hoog (resp. laag) gemiddelde van de duplo's samengaat met een groot (resp. klein) verschil tussen de beide waarnemingen. Bij toepassing van meergenoemde regel is dus de kans dat een waarneming „acceptabel” is, groter voor een lage dan voor een hoge waarneming, hetgeen inhoudt dat men het gemiddelde systematisch *onderschat*, en men dus een systematische fout geïntroduceerd heeft.

In het licht van deze beschouwingen gezien, hebben duplo-waarnemingen (gelijktijdig verricht) alleen zin om te waarschuwen voor grove misstappen, die op de soms lange en moeizame weg, waarop tot het resultaat wordt gekomen, gemaakt kunnen zijn. Dit brengt ons dus direct in de problematiek van de „uitschieters” of „uitbijters”.

Over dit moeilijke onderwerp zal ik niet veel zeggen. Statistisch gezien is er geen alleszins bevredigende oplossing. Er zijn wel toetsingsmethodes om na te gaan of een bepaalde waarneming als uitbijter aan te merken is, maar er zijn daarbij ook de nodige moeilijkheden. Bijvoorbeeld indien een dergelijke toetsingsmethode gebaseerd is op de onderstelling dat de waarnemingen uit een *normale* verdeling afkomstig zijn, zijn deze methodes in de regel zeer gevoelig voor afwijkingen van deze onderstelling. Als Uw waarnemingen afkomstig zijn uit de verdeling van figuur 4, zult U vanwege de dikkere „staart” moeten verwachten dat af en toe een schijnbare uitbijter zal optreden.

In het algemeen zal daarom een statisticus er alleen vrede mee hebben dat een uitbijter weggelaten wordt, indien aanwijsbare vergissingen of aanwijsbare grove fouten gemaakt zijn. In alle andere gevallen zal hij bij de aanwezigheid van „verdachte” waarnemingen een statistische analyse tweemaal uitvoeren. Eénmaal mét en éénmaal zónder de „verdachte” waarnemingen. Zijn de conclusies in beide gevallen vrijwel dezelfde, dan is het compromis duidelijk. Komen er zeer verschillende antwoorden uit deze analyses, dan is er maar één weg: een herhaling, niet van één of twee waarnemingen, maar van een redelijk gedeelte van het experiment. Dan nog dient men in een publikatie of rapport deze gang van zaken te vermelden, en niet zonder nadere toelichting bij de waarnemingen alleen de tweede serie op te geven.

Om nog even op de duplo-waarnemingen terug te komen, het is niet mijn bedoeling om U ervan te overtuigen dat duplo-waarnemingen overbodig zijn. Duplicering van de waarnemingen kan een bijzonder gunstige invloed hebben op de scherpte, waarmee de conclusies getrokken kunnen worden. Het is dikwijls de enige manier om voldoende informatie te krijgen over de waarnemingsnauwkeurigheid. Maar dan moeten de duplo-waarnemingen niet *gelijktijdig* uitgevoerd worden. De beste en meest efficiënte methode is de

waarnemingen te herhalen, zodanig dat toch elke waarneming voor „vol” aangezien kan worden, hetgeen o.a. bereikt wordt door de volgorde van de proeven (die dus alle tweemaal voorkomen) aselekt te kiezen. Dit is evenwel niet altijd mogelijk, in verband met eisen die de methodiek in kwestie stelt, of door organisatorische moeilijkheden. In dat geval kunnen door de keuze van een geschikte proefopzet deze moeilijkheden toch overwonnen worden.

Discussie

Dr. A. E. de Vries vraagt:

Gebruikt men bij het beschrijven van een reeks waarnemingen, welke alleen positieve waarden kunnen hebben, de normale verdeling, dan is er volgens die normale verdeling een kans, dat ook negatieve waarden optreden. Welke betekenis hebben deze negatieve waarden?

Antwoord:

Een normale verdeling kan een modelvoorstelling zijn om de waarnemingsresultaten te beschrijven. In de regel zal, indien het gemiddelde van de kansverdeling groot, en de spreiding niet al te groot is, de kans op een negatieve waarde zo klein zijn, dat deze discrepantie tussen de werkelijkheid en de modelvoorstelling praktisch gezien toch acceptabel is. Is deze discrepantie niet acceptabel, dan dient men een ander model te kiezen, bijvoorbeeld met een afgeknotte normale kansverdeling of met een andere kansverdeling waarbij de kansgrootheid alleen positieve waarden aanneemt, bijv. een χ^2 -verdeling.

Drs. J. J. A. Lammerts van Bueren vraagt:

Zijn er regels te geven voor het doorgeven van analyse-resultaten verkregen uit een duplo-analyse zonder gebruik te maken van het statistisch vakjargon?

Antwoord:

De beste methode is altijd de waarnemingen zelf op te geven. Wil men beslist één getal, dan het gemiddelde, maar dan ook — bijvoorbeeld tussen haakjes — het aantal waarnemingen waarop dit gemiddelde is gebaseerd. Ook kan men, indien de waarnemingen altijd eenzelfde aantal malen herhaald worden, het verschil tussen hoogste en laagste waarneming opgeven.

Ir. R. de Vletter vraagt:

Er is vrijwel alleen gesproken over methodes gebaseerd op de normale verdeling. Komen niet-normale verdelingen vaak voor en wordt daardoor niet een nieuwe onzekerheid geïntroduceerd, die alleen door uitvoerig onderzoek is te elimineren?

Antwoord:

De hier besproken methodes zijn inderdaad alle gebaseerd op de normaliteitsonderstelling. Waar deze onderstelling niet vervuld is, dus bijvoorbeeld indien men met een scheve verdeling te maken heeft, dan kan men trachten door een transformatie van de waarnemingen deze meer „normaal” te maken. Bekend is in dit verband de logaritmische transformatie. Maar toch is dit veelal een weinig bevredigende methode. Een principiële beter uitgangspunt is dan de normaliteitsonderstelling niet in te voeren. Er is een heel arsenaal van methodes ontwikkeld, die niet zijn gebaseerd op de onderstellingen van de normale verdelingen. Men noemt dit de „verdelingsvrije methodes”, een naam die niet geheel correct is, want men heeft ergens toch met een al dan niet normale kansverdeling te maken. Het verdelingsvrije analogon van de toets van *Student* voor twee steekproeven is bijv. de tweesteekproeven-toets van *Wilcoxon* (zie o.a. *Constance van Eeden* en *Doraline Wabeke*, Handleiding voor de toets van *Wilcoxon*, rapport S 176 van het

Mathematisch Centrum). Het verdelingsvrije analogon voor de toets van *Student* voor het gemiddelde van een normale verdeling is de tekentoeets of ook de symmetrie-toets van *Wilcoxon* (zie o.a. *A. Benard* en *Constance van Eeden*, Handleiding voor de symmetrie-toets van *Wilcoxon*, rapport S 208 van het Mathematisch Centrum).

Of niet-normale verdelingen veelvuldig voorkomen in analytisch-chemisch werk is mij niet bekend. Wel bekend is dat zij voorkomen.

Drs. H. A. van Vucht vraagt:

Zijn de door U beschreven methodieken ook toepasbaar op analyseresultaten die uitgedrukt zijn in procenten? Immers hiervan weten wij a priori dat zij geen normale verdeling vormen daar waarden beneden 0 % en boven 100 % nooit kunnen voorkomen?

Antwoord:

Een van de modellen, die in de statistiek gebruikt worden om gevonden fracties (percentages $\cdot 100^{-1}$) te bestuderen is het eenvoudigst te beschrijven aan de hand van een kruis en munt spel. Stel dat men n maal een munt opwerpt, waarbij men iedere keer een kans p heeft op kruis, en een kans $1-p$ op munt. Zijn de worpen voorts onderling onafhankelijk, dus wordt het resultaat van één worp niet beïnvloed door dat van de andere worpen, dan is het aantal keren kruis dat gegooid wordt een kansgrootheid, welke een *binomiale* kansverdeling bezit. De kans op k keer kruis in n worpen is daarbij

$$\frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}.$$

Voor niet al te kleine waarden van n , p en $(1-p)$, bijvoorbeeld voor $np > 5$ en $n(1-p) > 5$ is deze kansverdeling weer te benaderen door een normale kansverdeling met verwachting np en spreiding $\sqrt{np(1-p)}$. Theoretisch behoort bij de normale verdeling een positieve kans op een waarneming < 0 en $> n$, maar in het gebied van de praktische toepasbaarheid van de benadering van een binomiale door een normale kansverdeling zijn deze kansen in de regel klein en zullen maar zelden tot moeilijkheden in de praktische interpretatie leiden.

Dr. A. E. de Vries vraagt:

Bij het samenvoegen van 3 reeksen van ieder 8 waarnemingen vindt men de standaard-deviatie met 21 vrijheidsgraden. Eén reeks van 24 waarnemingen heeft 23 vrijheidsgraden. Wat is het essentiële verschil?

Antwoord:

Zijn de waarnemingen alle onderling onafhankelijke waarnemingen uit een normale verdeling, met eenzelfde σ , maar is het gemiddelde van de normale verdeling voor de eerste serie μ_1 , de tweede μ_2 en de derde μ_3 , dan is de schatting s^2 van σ^2 , berekend door combineren van schattingen uit iedere serie, niet afhankelijk van wat de gemiddelden μ_1 , μ_2 en μ_3 zijn. Berekent men uit het gehele materiaal van 24 waarnemingen een s^2 met 23 vrijheidsgraden, dan is dit alleen dan een redelijk te achten schatting van σ^2 als $\mu_1 = \mu_2 = \mu_3$. U wint dan 2 vrijheidsgraden, maar voert een extra-onderstelling in.

Dr. A. H. Holtz vraagt:

Er zijn twee mogelijkheden om uitkomsten van een chemische analyse te berekenen:

- met behulp van een ijkcurve (bij het opstellen waarvan men rekening moet houden met de variantie van de uitkomsten van de bepalingen op de standaardoplossingen);
- met behulp van in iedere serie analyses opgenomen bepalingen op standaardoplossingen.

Is er iets te zeggen wanneer men welke werkwijze moet volgen en waarom?

Antwoord:

In geval a zal men de ijklijn en een eventueel betrouwbaarheidsinterval berekenen uit een vrij groot aantal waarnemingen, zodat de ijklijn vrij nauwkeurig bekend is.

In geval b beoordeelt men de analyse-uitkomsten tegen de achtergrond van een of enkele waarnemingen, die tegelijkertijd gedaan zijn, hetgeen dus neerkomt op ijkken aan de hand van een vrij onnauwkeurig bekende ijklijn.

Daarom zal de werkwijze b alleen worden gekozen, als men er niet op kan vertrouwen dat de ijklijn gedurende enige tijd ongewijzigd blijft. Soms treft men situaties aan waar een organisch reagens, dat om de paar dagen opnieuw moet worden klaargemaakt, verschillen introduceert in de uitkomsten, zodat men bij elke nieuwe portie reagens weer een ijkkromme moet bepalen. Meestal zal men dan overgaan tot werkwijze b. Een voorbeeld van deze situatie is de bepaling van de prothrombinetijd van bloed, waarbij als boosdoener een organische verbinding, de thromboplastine, fungeert. Men moet steeds elke dag bloed van een „normaal” persoon afnemen, daarvan de prothrombinetijd bepalen, en alle waarnemingen op die dag met deze waarnemingen vergelijken.

Dr. A. H. Holtz vraagt:

Is het geoorloofd uitkomsten van analyses van eenzelfde materiaal, welke door verschillende laboratoria verricht zijn, statistisch te bewerken? Dit impliceert toch dat men aanneemt dat alle fouten die door de laboratoria gemaakt worden (waarbij ook systematische fouten) en die soms zelfs zeer ernstig kunnen zijn, zoals Ir. Bloemena zelf toegeeft, stochastisch verdeeld zijn? Heeft men deze garantie? En geldt eigenlijk niet hetzelfde voor uitkomsten van verschillende analisten?

Antwoord:

Uiteraard is het mogelijk om waarnemingen die in verschillende laboratoria zijn verkregen, statistisch te bewerken. Wil men een indruk hebben van de verschillen die optreden bij de kwantitatieve analyse van bijvoorbeeld alcohol in een bloedmonster, dan kan men door verschillende laboratoria een aantal gelijke bloedmonsters laten analyseren. Welk model men nu voor de variantie-analyse moet kiezen, hangt er van af hoe men de laboratoria uitkoos. Zijn de laboratoria die meedoen willekeurig gekozen uit een groot aantal laboratoria, dan moet men inderdaad in het model de invloed van een laboratorium als een kansgrootheid opnemen. Zijn de laboratoria niet willekeurig gekozen, dus doen bijv. alle laboratoria die er zijn mee, of alleen de drie of vier best georganiseerde, dan moet men de invloed van een laboratorium niet als een kansgrootheid, maar als een constante introduceren.

Dat enkele laboratoria (of analisten) bij een dergelijk experiment resultaten vinden die sterk afwijken van de rest, betekent alleen dat de onderstelling dat de laboratoriuminvloed een normale kansverdeling heeft, mogelijk niet gerechtvaardigd was.

Een statistische analyse van een experiment, waaraan 34 laboratoria deelnamen is beschreven in *W. J. Youden, Evaluation of chemical analysis of two rocks, Technometrics 1, 421 (1959).*

¹ *Dantzig, D. van, Enkele historische betrekkingen tussen de mathematische en de verzamelende statistiek, Statistica, 4, 1950, 233—248.*

² *Pearson, E. S. and Hartley, H. O., Biometrika Tables for Statisticians, Cambridge University Press, 1958 (2nd impr.).*

Nauwkeurigheid en juistheid van chemische analyse- methodes *

(Precision and accuracy of
analytical methods)

543

M. J. MAURICE

Centraal Analytisch Laboratorium N.V. Onderzoekingsinstituut „Research” (A.K.U. en daarmee verbonden ondernemingen), Arnhem

A review is given of some methods for establishing the precision and the accuracy of analytical methods, whereby use is made of some statistical tests.

In reporting the precision of an analytical method the variance is used. Two cases will be distinguished:

- a the variance is constant at varying amounts of the compound to be determined.
- b the variance increases with increasing amounts of the compound.

Examples of both cases are given. Further, the calculation of the variance of a test method requiring a blank determination is discussed.

There are essentially two ways in which the accuracy of an analytical method can be established:

- a comparison of the results obtained with the known content of a standard preparation;
- b comparison of test results with those obtained by one or more independent methods.

In both cases the t-test is used. Some examples are given as illustrations.

1. Inleiding

Wanneer men een uitspraak wil doen over de bruikbaarheid van een analysemethode, dient men o.a. te weten of de methode juist is en hoe nauwkeurig deze is. De begrippen juistheid en nauwkeurigheid dienen scherp van elkander te worden onderscheiden, daar zij op geheel verschillende zaken betrekking hebben. Juistheid heeft betrekking op de systematische afwijkingen van de analyseresultaten t.o.v. de werkelijke waarde, terwijl nauwkeurigheid iets zegt over de toevallige schommelingen van deze resultaten om hun gemiddelde. Over de nauwkeurigheid zal eerst gesproken worden. Daarna zal de vraag behandeld worden hoe de juistheid van een analysemethode kan worden vastgesteld. Hierbij zal gebruik worden gemaakt van statistische methodes, die in de voorgaande voordracht toegelicht zijn. (Chem. Weekblad 465).

2. Nauwkeurigheid

2.1. Algemene beschouwingen

Wanneer een analyse enige malen wordt herhaald, blijkt, dat uitkomsten worden verkregen, die een zekere spreiding vertonen. Deze spreiding is steeds groter dan de som van de bijdragen van de controleerbare spreidingsbronnen, zoals wegen, aflezen van een buret e.d. Dit is een algemeen verschijnsel, dat tot de conclusie leidt, dat er naast een aantal bekende en meetbare oorzaken ook factoren een rol spelen, die (nog) niet bekend zijn en dikwijls moeilijk of in het geheel niet meetbaar zijn, bijv. fluctuaties ten gevolge van het chemisme van de reactie, die aan de analyse ten grondslag ligt, en de geofendheid van de waarnemer.

Ook de tijd speelt een belangrijke rol¹ en wel omdat er verschillende ongecontroleerde invloeden zijn, die langzaam met de tijd veranderen, bijv. de kwaliteit van een reagens. Deze invloed bemerkt men niet, wanneer een aantal bepalingen binnen korte tijd, bijv. een dag, wordt uitgevoerd, doch wel wanneer met het uitvoeren van deze bepalingen een langere tijd, bijv.

* Voordracht, gehouden op het symposium over „De toepassing van statistiek bij de chemische analyse” op 31 maart 1960 te Amsterdam.

10 dagen gemoeid is. Dit betekent dus, dat de spreiding in het laatste geval groter is dan in het eerste.

Bij het beoordelen van de gevonden spreiding in analyse-uitkomsten zal derhalve steeds in aanmerking moeten worden genomen of werd gewerkt volgens een genormaliseerd voorschrift, d.w.z. een voorschrift, waarin de invloed van onbekende factoren is gereduceerd, of volgens een niet-genormaliseerd voorschrift.

De tot nu toe genoemde spreidingsbronnen zijn alle inherent aan de gevolgde methode. Er zijn echter nog meer factoren, die de grootte van de spreiding van de analyse-uitkomsten beïnvloeden. Wordt een bepaling een aantal malen herhaald door een aantal analisten binnen één laboratorium, dan is de spreiding doorgaans groter dan wanneer deze analyses door één analist op hetzelfde laboratorium worden uitgevoerd. De spreiding wordt nog groter, wanneer men de bepalingen op verschillende laboratoria door verschillende analisten laat uitvoeren. Dit zijn dus spreidingsbronnen, die niet inherent zijn aan de analysemethode.

De nauwkeurigheid van een methode heeft nu uitsluitend betrekking op de spreiding van de uitkomsten ten gevolge van de inherente spreidingsbronnen. Indien ook één of meer bijdragen van niet-inherente spreidingsbronnen in de opgegeven spreiding worden opgenomen, kan niet meer worden gesproken over de nauwkeurigheid van de methode.

Wanneer men de beschikking heeft over de resultaten, die verkregen zijn door volgens één voorschrift een monster op verschillende laboratoria door een aantal analisten te laten analyseren, is het mogelijk afzonderlijk de bijdragen van de niet-inherente spreidingsbronnen en de nauwkeurigheid van de methode te berekenen. De hierbij gebruikte statistische techniek is die van de variantie-analyse.

2.2. Enige voorbeelden van het bepalen van de nauwkeurigheid van analysemethodes

Bij het bepalen van de nauwkeurigheid wordt met voordeel gebruik gemaakt van de variantie als maat voor de spreiding. We beperken ons tot diè analysemethodes, waarbij geen ijkcurve gebruikt wordt, d.w.z. tot de gravimetrische en volumetrische methodes.

Het is nu allereerst noodzakelijk vast te stellen

welk variatiegebied van hoeveelheden c.q. concentraties men wenst te onderzoeken. In een aantal gevallen zal dit variatiegebied klein zijn, nl. wanneer men met de ontworpen analysemethode een parameter van een productieproces wil controleren, die binnen enge grenzen fluctueert. Vervolgens dient te worden nagegaan of binnen dit variatiegebied de variantie een functie is van de hoeveelheid te bepalen stof of niet. De volgende gevallen kunnen zich dan voordoen:

- de variantie is constant bij variërende hoeveelheid te bepalen stof,
- de variantie neemt toe bij toenemende hoeveelheid te bepalen stof.

Het kan voorkomen, dat de variantie toeneemt bij afnemende hoeveelheid stof. Dit betekent, dat de ondergrens van het variatiegebied, waarbinnen de onderzochte methode kan worden toegepast, is overschreden. Dit geval, dat verder buiten beschouwing gelaten zal worden, doet zich voor bijv. in de gravimetrie en bij neerslagtitraties, wanneer de bepaling van kleine hoeveelheden van een verbinding wordt gestoord door een verminderde vormingssnelheid van het neerslag.

Met welke van de twee bovengenoemde gevallen men in een bepaalde situatie te maken heeft, kan worden uitgemaakt door uitgaande van verschillende hoeveelheden van de te bepalen stof, die liggen binnen het vooraf vastgestelde variatiegebied, een reeks bepalingen uit te voeren en deze een aantal malen te herhalen. De volgorde, waarin deze bepalingen worden uitgevoerd, moet dan een toevallige zijn. Voor het vaststellen van deze toevallige volgorde kan een lotingstabel worden gebruikt. Men krijgt op deze wijze voor iedere uitganghoeveelheid een serie uitkomsten. Van iedere serie worden de variantie en de standaardafwijking berekend en vervolgens wordt de standaardafwijking grafisch uitgezet tegen de bijbehorende hoeveelheid stof. Uit deze grafiek verkrijgt men dan een indruk aangaande het al dan niet bestaan van een functioneel verband tussen de standaardafwijking, σ , en de hoeveelheid stof, μ .

Wanneer uit deze grafiek blijkt, dat er waarschijnlijk geen functioneel verband bestaat tussen σ en μ , worden ter bevestiging van de verkregen indruk de varianties, behorend bij de grootste en de kleinste hoeveelheid stof, met elkaar vergeleken m.b.v. de F-toets. Indien met deze toets op een vooraf vastgesteld niveau, waarvoor meestal 0.05 wordt genomen, geen verschil kan worden aangetoond, mag worden aangenomen, dat de gevonden varianties schattingen zijn van dezelfde populatie-variantie. (Indien men met nog meer zekerheid wenst vast te stellen, dat de variantie binnen het onderzochte variatiegebied constant is, verdient het aanbeveling een controlekaart te tekenen.)

De verkregen schattingen kunnen hierna worden gecombineerd teneinde een betere schatting van de populatiewaarde te verkrijgen. Deze schatting is dan een maat voor de nauwkeurigheid.

In het geval, dat uit de grafiek blijkt, dat σ wel een functie van μ is, is het veelal ook mogelijk bij benadering vast te stellen welke de aard is van het functionele verband, $\sigma = f(\mu)$. Om te komen tot een binnen het onderzochte variatiegebied constante variantie is het noodzakelijk de waarnemingen te transformeren en van de getransformeerde gegevens

de variantie en de standaardafwijking te berekenen. De wijze van transformeren is afhankelijk van de aard van het functionele verband². In het algemeen moet zodanig worden getransformeerd, dat de getransformeerde grootheid, μ' , gelijk is aan $\int \frac{1}{f(\mu)} d\mu$.

Enige voorbeelden hiervan zijn weergegeven in tabel I.

Tabel I. Toe te passen transformaties bij enige veel voorkomende relaties tussen μ en σ .

Functioneel verband	Transformatie
σ evenredig met μ^2	$\mu' = \frac{1}{\mu}$
σ evenredig met μ	$\mu' = \log \mu$
σ evenredig met $\sqrt{\mu}$	$\mu' = \mu^2$

Het aantal uit te voeren bepalingen zal niet te klein mogen zijn. In vele gevallen verkrijgt men met 20—30 bepalingen zeker een redelijke indruk van de bereikbare nauwkeurigheid. Het is mogelijk, dat men wel eens met een geringer aantal zal moeten volstaan, bijv. wanneer het zeer langdurige bepalingen betreft.

Van beide genoemde gevallen, nl. constante variantie en toenemende variantie, zal een voorbeeld gegeven worden.

Het eerste voorbeeld (constante variantie) betreft de reductie van elementaire zwavel, opgelost in pyridine, met een mengsel van onderfosforigzuur, jodwaterstofzuur en zoutzuur. De gevormde zwavelwaterstof wordt uitgedreven en geabsorbeerd in een azijnzure koperacetatoplossing, waarna de overmaat koper wordt getitreerd met EDTA op murexide als indicator³. Het onderzochte variatiegebied is 2-8 mg zwavel. Uitgaande van 5, 10 en 15 ml van een oplossing van zwavel in pyridine werden door één ervaren analiste de in tabel II vermelde resultaten verkregen.

Allereerst wordt nagegaan of het verkregen waarnemingsmateriaal uitschieters bevat. Dit komt hierop neer, dat m.b.v. een tabel de grootte wordt bepaald van de kans, dat een extreme waarde of een grotere nog behoort tot dezelfde verdeling, waartoe de overige waarden behoren. Is de gevonden kans klein, bijv. kleiner dan 0.05, dan wordt de nulhypothese verworpen en aangenomen, dat de extreme waarde niet tot dezelfde verdeling behoort als de overige waarden⁴. Voorwaarde voor het toepassen van de in⁴ genoemde toets is, dat de waarnemingen althans bij benadering normaal verdeeld moeten zijn, hetgeen in het onderhavige geval verondersteld mag worden⁶.

De in tabel II vermelde uitkomsten blijken geen uitschieters te bevatten. Dan wordt voor ieder der drie series de standaardafwijking berekend. De grafiek, die s_1 weergeeft als functie van x_1 wekt de indruk, dat de standaardafwijking in dit geval geen functie is van de hoeveelheid stof. Deze indruk wordt nog bevestigd door s_1^2 en s_3^2 met elkaar te vergelijken m.b.v. de F-toets:

$$F = \frac{73.4 \cdot 10^{-4}}{61.0 \cdot 10^{-4}} = 1.20 \text{ met } \nu_1 = \nu_3 = 7 \text{ vrijheids-}$$

graden. Uitgaande van de nulhypothese, dat beide schattingen zijn van dezelfde populatievariantie wordt de kans bepaald, dat de gevonden F-waarde of een

grotere zal optreden ten gevolge van toevalsfluctuaties. In dit geval is deze kans groter dan 0.10, zodat er geen reden is de nulhypothese te verwerpen. Aangenomen mag dus worden, dat de drie variantieschattingen zijn van dezelfde populatiewaarde. De drie schattingen kunnen nu gecombineerd worden volgens:

$$s^2 = \frac{\sum v_i s_i^2}{\sum v_i} \quad (I)$$

$$s^2 = \frac{(7 \cdot 73.4 + 7 \cdot 58.1 + 7 \cdot 61.0) 10^{-4}}{21} = 64.2 \cdot 10^{-4} \text{ (mg)}^2 \quad (II)$$

$$\text{zodat } s = \sqrt{64.2 \cdot 10^{-4}} = 0.080 \text{ mg} \quad (III)$$

Deze gecombineerde standaardafwijking, s , heeft nu 21 vrijheidsgraden.

Het tweede voorbeeld (toenemende variantie bij toenemende hoeveelheid) betreft de amperometrische titratie van anorganisch materiaal in petroleum-sulfo-

zal dus ook de standaardafwijking toenemen met toenemende x . Deze indruk wordt nog bevestigd door s^2 te vergelijken met s^2 m.b.v. de F-toets. De gevonden overschrijdingskans bedraagt ca. 0.02, zodat de nulhypothese, inhoudende, dat tussen s_2^2 en s_1^2 geen verschil bestaat, moet worden verworpen. Om te komen tot een constante variantie binnen het onderzochte variatiegebied moeten de waarnemingsuitkomsten dus worden getransformeerd. Er werden verschillende transformaties beproefd, waarbij bleek, dat een logaritmische transformatie het beste voldoet. Hierbij werd gebruik gemaakt van een controlekaart⁵. De duplo-verschillen tussen de logaritmen van de individuele waarden, w' , werden berekend (tabel III). Uit het gemiddelde van deze duplo-verschillen, \bar{w}' , werd s' berekend volgens (IV):

$$\begin{aligned} \bar{w}' &= 0.0142 \\ s' &= 0.0142 \cdot 0.8862 = 0.0126 \end{aligned} \quad (V)$$

Het verband tussen de getransformeerde standaard-

Tabel II. Resultaten van de bepaling van zwavel door reductie tot H_2S .

	5 ml opl. x_1	10 ml opl. x_2	15 ml opl. x_3
	2.30 mg	4.48 mg	6.73 mg
	2.40	4.54	6.83
	2.24	4.47	6.80
	2.37	4.61	6.88
	2.24	4.58	6.92
	2.42	4.61	6.91
	2.20	4.42	6.84
	2.39	4.62	6.98
gemiddelde, \bar{x}_i	2.32 mg	4.54 mg	6.86 mg
variantie, s_i^2	$73.4 \cdot 10^{-4} \text{ (mg)}^2$	$58.1 \cdot 10^{-4} \text{ (mg)}^2$	$61.0 \cdot 10^{-4} \text{ (mg)}^2$
stand. afw. s_i	0.086 mg	0.076 mg	0.078 mg

naten met loodnitraat⁷. Dit anorganische materiaal bestaat hoofdzakelijk uit sulfaat en sulfiet. Enige van de verkregen resultaten, uitgedrukt in procenten natriumsulfaat, zijn vermeld in tabel III.

Zowel uit tabel III als uit de grafiek, die w weergeeft als functie van de gemiddelde x -waarde, blijkt, dat w toeneemt met toenemende x . Aangezien het

afwijking, σ' , en de niet-getransformeerde standaardafwijking, σ , wordt gegeven door:

$$\sigma_x = \frac{\sigma' \mu_x}{0.4343}, \text{ in de praktijk geschat door:}$$

$$s_x = \frac{s' \mu_x}{0.4343} \quad (VI)$$

Tabel III. Resultaten van de bepaling van anorganisch materiaal in petroleum-sulfonaten.

Monster	Percentage Na_2SO_4		Duplo-verschil, w	$w' = \log x_1 - \log x_2$
1	0.36	0.35%	0.01%	0.0122
2	0.37	0.38	0.01	0.0116
3	0.37	0.39	0.02	0.0229
4	0.44	0.46	0.02	0.0184
5	0.46	0.47	0.01	0.0093
6	2.63	2.61	0.02	0.0034
7	4.33	4.22	0.11	0.0112
8	4.95	5.10	0.15	0.0130
9	5.73	6.08	0.35	0.0257
				$\bar{w}' = 0.0142$

duplo-verschil, w , samenhangt met de standaardafwijking, s , volgens:

$$s = wA \quad (IV)$$

waarin $A = 0.8862$

waarin s_x = standaardafwijking bij een hoeveelheid $x = \mu_x$

μ_x = een bepaalde hoeveelheid Na_2SO_4 .

(VI) wordt gevonden door toepassing van de foutenvoortplantingswet op de transformatievergelijking $\mu'_x = \log \mu_x$. Substitutie van de in (V) genoemde waarde voor s' levert:

$$s_x = \frac{0.0126 \mu_x}{0.4343} = 0.029 \mu_x \quad (\text{VII})$$

De variatie-coëfficiënt, v , die gedefinieerd is als de standaardafwijking als fractie van het gemiddelde

(dus: $v = \frac{\sigma_x}{\mu_x}$), bedraagt in dit geval 2.9 %.

Een afzonderlijk probleem vormt het vastleggen van de nauwkeurigheid van methodes, waarbij één of meer van de gebruikte hulpstoffen eveneens de te bepalen component bevat en dus een blanco-bepaling moet worden uitgevoerd. In het eerste voorbeeld zou dit bijv. betekenen, dat nog een aantal blanco-bepalingen moet worden uitgevoerd en ieder der drie gevonden gemiddelden moeten worden verminderd met het gemiddelde van de uitkomsten van de blanco-bepalingen. Om de variantie te berekenen moeten de variantie van de bepaling en die van de blanco worden opgeteld. Immers, wanneer de uitkomst y gelijk is aan het verschil tussen de uitkomst van een bepaling, x , en die van een blanco, z , en x en z zijn onderling onafhankelijk, geldt:

$$y = x - z \quad (\text{VIII})$$

$$\text{en } \sigma_y^2 = \sigma_x^2 + \sigma_z^2 \quad (\text{IX})$$

(IX) wordt in de praktijk geschat door

$$s_y^2 = s_x^2 + s_z^2 \quad (\text{X})$$

De aldus berekende varianties kunnen niet worden vergeleken m.b.v. de F-toets, omdat ze onderling afhankelijk zijn. In dit geval dienen nu de varianties van de eerste en de laatste serie vóór aftrekken van de blanco te worden vergeleken m.b.v. de F-toets. Wordt daarbij geen significant verschil gevonden, dan kunnen de varianties worden gecombineerd. Deze gecombineerde variantie vermeerderd met die van de blanco-bepaling levert dan een schatting van de variantie van de onderzochte methode.

Het volgende voorbeeld moge een en ander verduidelijken. Volgens Morris en medewerkers⁸ kan zwavel worden bepaald door dit te koken met natriumsulfiet en het hierbij gevormde thiosulfaat — na maskeren van de overmaat sulfiet met formaldehyde —

met jodium te titreren. Naar onze ervaringen is het in dit geval noodzakelijk een blanco-bepaling uit te voeren. Een mogelijke verklaring hiervoor is, dat de maskering van de overmaat sulfiet niet kwantitatief is⁹. Uitgaande van 2, 4 en 6 ml van een oplossing van colloïdale zwavel werden de in tabel IV vermelde resultaten in mg zwavel verkregen. In deze tabel zijn de gemiddelde uitkomsten per serie, alsmede de varianties per serie opgegeven.

De in tabel IV genoemde uitkomsten bevatten geen uitschieters. Uit de grafiek, die s_1 weergeeft als functie van x_1 , krijgt men niet de indruk dat s_1 toeneemt met toenemende x_1 . Deze indruk wordt nog bevestigd door s_1^2 en s_3^2 te vergelijken m.b.v. de F-toets. De varianties worden nu gecombineerd volgens (I).

Voor de gecombineerde variantie, s_x^2 , wordt dan gevonden:

$$s_x^2 = \frac{(235 + 80.1 + 100) \cdot 10^{-4}}{3} = 138.4 \cdot 10^{-4} (\text{mg})^2 \quad (\text{XI})$$

De variantie na correctie voor de variantie van de blanco, s_z^2 , wordt berekend volgens (X). Gevonden wordt:

$$s_y^2 = 138.4 \cdot 10^{-4} + 18.3 \cdot 10^{-4} = 156.7 \cdot 10^{-4} (\text{mg})^2 \quad (\text{XII})$$

$$\text{zodat } s_y = 0.125 \text{ mg} \quad (\text{XIII})$$

Het aantal vrijheidsgraden van s_y , ν'_y , wordt bij benadering gegeven door:

$$\nu'_y = \frac{(s_x^2 + s_z^2)^2}{\frac{s_x^4}{\nu_x} + \frac{s_z^4}{\nu_z}} = 25.6 \approx 25 \quad (\text{XIV})$$

3. Juistheid.

3.1 Algemene beschouwingen

Het begrip juistheid heeft betrekking op systematische afwijkingen t.o.v. de werkelijke waarde. Of een analysemethode al dan niet juist is, zou nu min of meer eenvoudig zijn vast te stellen, wanneer die werkelijke waarde bekend was, nl. door de procedure toe te passen op een standaardpreparaat van de te bepalen stof, waarvan men uit andere overwegingen de ga-

Tabel IV. Resultaten van de bepaling van zwavel door omzetting in thiosulfaat.

2 ml opl. x_1	4 ml opl. x_2	6 ml opl. x_3	blanco z
2.15 mg	4.34 mg	6.51 mg	0.17 mg
2.05	4.53	6.46	0.20
2.38	4.43	6.35	0.21
2.10	4.24	6.39	0.16
2.34	4.29	6.59	0.18
2.43	4.41	6.30	0.24
2.18	4.38	6.32	0.13
2.42	4.41	6.38	0.11
Gemiddeld 2.26 mg	4.38 mg	6.41 mg	0.18 mg
Variantie $235 \cdot 10^{-4} (\text{mg})^2$	$80.1 \cdot 10^{-4} (\text{mg})^2$	$100 \cdot 10^{-4} (\text{mg})^2$	$18.3 \cdot 10^{-4} (\text{mg})^2$

rantie heeft, dat het gehalte ervan 100% is. Worden dan volgens de onderzochte methode resultaten gevonden, die niet meer dan toevallig van 100% afwijken, dan mag worden aangenomen, dat de methode juist is. Als voorbeeld kan worden genoemd het gebruik van metalen, welke met een gehalte van 99.999% ter beschikking staan bij het onderzoek van een bepalmethode voor het desbetreffende metaal. In dit geval zal in het algemeen de spreiding van de analyse-resultaten veel groter zijn dan de spreiding van dit gehalte, zodat de laatste kan worden verwaarloosd t.o.v. de eerste.

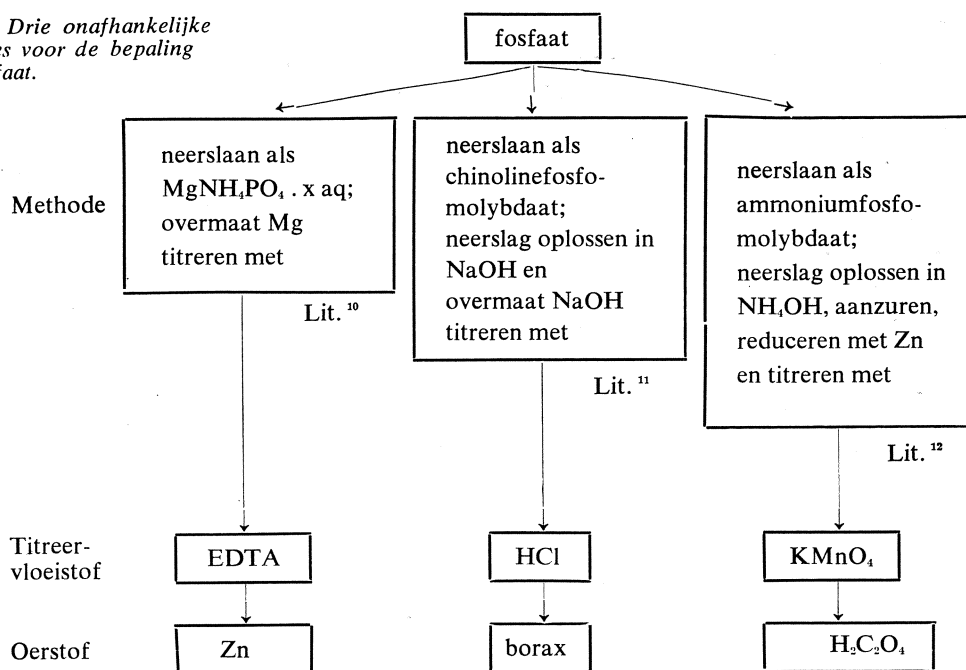
In zeer veel gevallen heeft men deze garantie echter niet, bijv. bij het onderzoek van vele technische che-

principe van de vergelijking van onafhankelijke methodes worden toegepast, nl. door aan meer monsters volgens ieder van een aantal verschillende methodes gelijktijdig één of — indien mogelijk — meer dan één bepaling uit te voeren.

Resumerend komen we tot de volgende gevallen:

1. standaardpreparaat is beschikbaar: vergelijking van de uitkomsten met de standaardwaarde, hetgeen statistisch gezien neerkomt op het toetsen van een gemiddelde t.o.v. de standaardwaarde.
2. er is geen standaardpreparaat beschikbaar, doch wel een stabiel monster van de te analyseren stof: vergelijking van de resultaten van onafhankelijke

Fig. 1. Drie onafhankelijke methodes voor de bepaling van fosfaat.



micaliën, welke niet eenvoudig te zuiveren zijn. Het is dan mogelijk de juistheid vast te stellen door de resultaten, verkregen volgens de onderzochte methode, te vergelijken met die, verkregen volgens een of meer andere methodes, die zodanig worden gekozen, dat alle gevolgde methodes onderling onafhankelijk zijn. Hierbij worden onder onafhankelijke methodes verstaan methodes, die op geheel verschillende principes zijn gebaseerd, terwijl ook de gebruikte titreervloeistoffen en oerstoffen verschillend zijn. Ter adstructie van dit begrip onafhankelijke methodes moge fig. 1 dienen, waarin schematisch drie onderling onafhankelijke methodes voor de bepaling van fosfaat zijn weergegeven.

Wanneer tussen de resultaten, verkregen volgens de verschillende methodes, geen verschil kan worden aangetoond, mag worden aangenomen, dat alle methodes dezelfde resultaten leveren. Het is dan waarschijnlijk, dat deze methodes alle juist zijn.

Het kan voorkomen, dat de stof, die men volgens verschillende methodes wil bepalen, niet stabiel is. Het is dan vrijwel onmogelijk een groot aantal bepalingen te verrichten zonder in de uitkomsten een sterke tijdsinvloed te introduceren. Toch kan ook in dit geval het

methodes; statistisch gezien betekent dit het toetsen van de verschillen van twee of meer gemiddelden t.o.v. nul.

3. er is geen standaardpreparaat beschikbaar en de te bepalen stof is instabiel: vergelijking van de uitkomsten van onafhankelijke methodes, toegepast op verschillende monsters en toetsen van de gemiddelde verschillen tussen de resultaten t.o.v. nul.

Deze beschouwingen over het vaststellen van de juistheid van een methode hebben echter alleen zin, wanneer deze methode niet geijkt is op een andere methode. Is dit laatste wel het geval, dan is de juistheid van de eerste methode uiteraard afhankelijk van de juistheid van de methode, waarop geijkt wordt. Een voorbeeld hiervan is de spectrofotometrische bepaling van sulfide. De juistheid van de methode is dan afhankelijk van de juistheid van de methode, die wordt gebruikt om het sulfidegehalte te bepalen van de oplossing, waarmee de ijkgrafiek wordt opgesteld.

Uiteraard is het ook bij het controleren van de juistheid noodzakelijk vooraf vast te stellen voor welk variatiegebied van hoeveelheden c.q. concentraties men over de juistheid een uitspraak wil doen.

3.2. *Enige voorbeelden van het controleren van de juistheid van analysemethodes.*

Het eerste voorbeeld, waarin we met het eerste van de drie bovengenoemde gevallen te maken hebben, betreft een door *Cimerman* en *Ariel*¹³ uitgewerkte micro-volumetrische bepaling van lood met thionalide (mononaftylthioglycolzuuramide). Met dit reagens wordt lood geprecipiteerd. Het neerslag wordt gewassen en opgelost in zuur, waarna het vrijgekomen thionalide met jodide-jodaatmengsel wordt geoxideerd tot dithionalide en de overmaat jodium met thiosulfaat wordt getitreerd. Het ontworpen voorschrift werd toegepast op oplossingen van loodnitraat van bekende concentratie. Hierbij verkregen de auteurs de volgende resultaten (tabel V):

Tabel V. Resultaten van de bepaling van lood m.b.v. thionalide.

Monster	Pb aanwezig x	Pb gevonden y	Verskil z = x-y	Relatieve verschil $d = \frac{x-y}{x} \cdot 100\%$
1	1.000 mg	1.000 mg	0 mg	0 %
		0.995	+0.005	+0.50
2	2.000	1.999	+0.001	+0.05
		1.994	+0.006	+0.30
3	4.000	3.989	+0.011	+0.28
		3.994	+0.006	+0.15
4	6.000	5.988	+0.012	+0.20
		5.978	+0.022	+0.37
5	8.000	7.977	+0.023	+0.29
		8.008	-0.008	-0.10
6	10.000	9.987	+0.013	+0.13
		9.966	+0.034	+0.34
Gemiddelde, $\bar{d} = +0.21 \%$				
Standaardafwijking van d, $s_d = 0.171 \%$				

De conclusie van de auteurs is: „Results are exact (maximum relative error $\pm 0.5 \%$)”. De vraag is echter of deze conclusie gerechtvaardigd is. Het absolute verschil, z, neemt toe met toenemende x. Dit geldt niet voor het relatieve verschil, d. Derhalve moet een verder onderzoek worden uitgevoerd met de relatieve verschillen en niet met de absolute verschillen. Wanneer de gevonden hoeveelheden, y, niet systematisch afwijken van de aanwezige hoeveelheden, x, mag het gemiddelde relatieve verschillen, \bar{d} , dus niet systematisch van nul afwijken. Als nulhypothese stellen we dus: \bar{d} wijkt niet meer dan toevallig van nul af. Vervolgens wordt de kans bepaald, dat de gevonden \bar{d} -waarde of een grotere kan optreden, wanneer de nulhypothese juist is. Hiertoe berekenen we de grootte t volgens:

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{\bar{d}}{s_d} \sqrt{n} \quad (\text{XV})$$

Substitutie van de gevonden waarden levert dan:

$$t = \frac{0.21}{0.171} \sqrt{12} = 4.25 \quad (\text{XVI})$$

Het aantal vrijheidsgraden bedraagt hier 11. Van deze t-waarde is de overschrijdingskans 0.002. De kans, dat de gevonden d-waarde niet systematisch van nul afwijkt is dus uitermate gering. De nulhypothese moet worden verworpen en er mag worden aangenomen, dat er een systematisch verschil bestaat. Dit

betekent dus, dat tevens een systematisch verschil tussen x en y mag worden aangenomen. De onderzochte methode kan dus niet juist worden genoemd; de resultaten zijn te laag.

Een voorbeeld van het tweede geval, i.c. de vergelijking van twee of meer onafhankelijke methodes, uitgaande van een stabiel monster, betreft de bepaling van stikstof in nitrocellulose¹⁴. Volgens de eerste methode wordt gereduceerd met een sterk zure FeCl_2 -oplossing tot NO, dat bepaald wordt (methode a). Bij de tweede methode wordt nitrocellulose opgelost in aceton, neergeslagen met verdunde loog en vervolgens door verwarmen daarin opgelost, waarna nitraat wordt gereduceerd tot ammoniak, dat titrimetrisch wordt bepaald (methode b). Ter wille van de eenvoudigheid worden in tabel VI alleen de aan één monster verkregen resultaten vermeld uitgedrukt in procenten stikstof.

Tabel VI. Resultaten van de N-bepaling in nitrocellulose volgens twee methodes.

	Methode a x	Methode b y	Verskil $d = \bar{x} - \bar{y}$
	12.06 %	12.05 %	
	12.05	12.05	
	12.09	12.04	
	12.07	12.04	
	12.05	12.03	
	12.07	12.03	
	12.11	12.00	
	12.08	12.01	
	12.10	12.01	
		12.02	
		12.00	
Gemiddelde	12.076 %	12.025 %	+0.051 %
Standaardafwijking	0.021 %	0.019 %	
Aantal waarnemingen	9	11	

De in tabel VI vermelde waarnemingsuitkomsten bevatten geen uitschieters. Tussen de varianties s_x^2 en s_y^2 , kan m.b.v. de F-toets geen verschil worden aangetoond, zodat deze varianties kunnen worden gecombineerd tot s^2 volgens (I). Gevonden wordt dan: $s = 0.0197 \%$. Vervolgens wordt de grootte t be-

rekend volgens:

$$t = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}} \quad (\text{XVII})$$

waarin n, m = aantal bepalingen per methode.

Substitutie van de gevonden waarden in (XVII) geeft:

$$t = \frac{12.076 - 12.025}{0.0197 \sqrt{\frac{1}{9} + \frac{1}{11}}} = 5.99 \quad (\nu = 18) \quad (\text{XVIII})$$

Van deze t-waarde bedraagt de overschrijdingskans minder dan 0.001, zodat een systematisch verschil tussen de resultaten van methode a en die van methode b mag worden aangenomen. Dit betekent, dat over de juistheid van de methodes geen uitspraak kan worden gedaan.

Het derde voorbeeld betreft het vergelijken van twee methodes voor de bepaling van totaal zwavel in viscose¹⁵. Viscose is een produkt, waarvan de samenstelling met de tijd verandert. Een aantal monsters werd volgens twee methodes geanalyseerd, waarbij per monster volgens elk der beide methodes één bepaling werd uitgevoerd. Bij de eerste methode wordt viscose geoxideerd met waterstofperoxide, waarna wordt drooggedampt en het residu wordt gesmolten met natriumperoxide of natriumcarbonaat. Hierbij worden de zwavelhoudende verbindingen omgezet in natriumsulfaat. Sulfaat wordt geprecipiteerd als bariumsulfaat en de overmaat barium wordt getitreerd met EDTA (methode a). Volgens de tweede methode wordt viscose behandeld met een mengsel van onderfosforigzuur, joodwaterstofzuur en zoutzuur, waardoor alle zwavelverbindingen worden omgezet in zwavelwaterstof en zwavelkoolstof. Deze beide verbindingen worden door verhitten uitgedreven, geabsorbeerd en elk afzonderlijk bepaald (methode b). De volgens beide methodes verkregen resultaten, uitgedrukt in procenten zwavel, zijn vermeld in tabel VII.

Tabel VII. Resultaten van twee methodes voor de bepaling van totaal zwavel in viscose.

Monster	Methode a x	Methode b y	Vershil d = x-y
1	2.50 %	2.53 %	-0.03 %
2	2.27	2.26	+0.01
3	2.12	2.15	-0.03
4	2.08	2.10	-0.02
5	2.29	2.32	-0.03
6	2.17	2.18	-0.01
7	2.20	2.18	+0.02
8	2.29	2.26	+0.03
9	1.93	1.96	-0.03
10	1.94	1.97	-0.03
11	1.94	1.94	0
12	2.26	2.25	+0.01
13	2.26	2.28	-0.02
14	2.24	2.24	0
15	2.00	2.06	-0.06
Gemiddelde, $\bar{d} =$			-0.0127 %
Standaardafwijking van d, $s_d =$			0.0240 %

Wanneer er tussen de resultaten der beide methodes geen systematisch verschil bestaat, moet het gemiddelde verschil, \bar{d} , niet meer dan toevallig van nul afwijken. Als nulhypothese wordt dus weer gesteld: \bar{d} wijkt niet systematisch van nul af. De kans wordt nu bepaald, dat de gevonden \bar{d} -waarde of een grotere kans optreden t.g.v. toevalsfluctuaties. Daartoe wordt de grootheid t weer berekend volgens (XV). Gevonden wordt:

$$t = \frac{-0.0127}{0.0240} \sqrt{15} = -2.05 \quad (\text{XIX})$$

De overschrijdingskans van deze t-waarde is bij 14 vrijheidsgraden 0.06. De nulhypothese behoeft niet te worden verworpen en een systematisch verschil tussen de resultaten der beide methodes kan niet worden aangetoond. Aangenomen mag nu worden, dat beide methodes juist zijn.

In het algemeen is het wenselijk bij het onderzoek naar de juistheid van een methode meer dan twee onafhankelijke methodes in het onderzoek te betrekken, zoals in¹⁵ is gedaan. Immers, wanneer twee methodes dezelfde resultaten geven, bestaat altijd nog de mogelijkheid, dat deze resultaten in beide gevallen op dezelfde wijze systematisch afwijken van de werkelijke waarde. Dit risico wordt kleiner naarmate het aantal onafhankelijke methodes, dat gelijke resultaten levert, groter wordt. De toetsingsprocedures worden dan echter iets ingewikkelder.

4. Slotopmerkingen

In het bovenstaande werd getracht enige wegen aan te geven, die kunnen dienen om de nauwkeurigheid en de juistheid van een analysemethode vast te stellen. Hierbij werd o.a. gebruik gemaakt van de F-toets en de t-toets. Het moge duidelijk geworden zijn, dat het van belang is statistische methodes toe te passen bij het interpreteren van analyseresultaten. Opgemerkt moet worden, dat het wenselijk is reeds bij het opzetten van experimenten rekening te houden met de statistische methodes, waarmee men het verkregen waarnemingsmateriaal wil analyseren. Zo zou bijv. de berekening van de gecombineerde standaardafwijking in het tweede sub 3.2 genoemde voorbeeld (tabel VI) vereenvoudigd kunnen worden, indien per methode een gelijk aantal bepalingen zou zijn uitgevoerd. Een intensief contact tussen statisticus en analyticus is derhalve zeer gewenst. Dit contact dient dan bij het begin van een onderzoek reeds te worden opgenomen.

Tot slot moge hier de aandacht worden gevestigd op de mogelijkheid de noodzakelijke berekeningen te vereenvoudigen door het toepassen van de zg. „quick and dirty” methodes¹⁶.

¹ Schmidt, R., *Metalen* 4, 227 (1950).

² Davies, O. L., „The design and analysis of industrial experiments”. — Oliver and Boyd, London-Edinburgh, 1956, p. 45.

³ Maurice, M. J., Diss. Amsterdam, V.U. 1956, p. 27 e.v.

⁴ Dixon, W. J. en Massey, F. J., „Introduction to statistical analysis”. — Mc Graw Hill Book Company, Inc., New York, 2e uitgave, 1957, p. 275 e.v.

⁵ *ibid.* p. 130 e.v.

⁶ Ontwerpnorm V1047, weergegeven van waarnemingsreeksen, p. 13 e.v.

⁷ Gordon, B. E. en Urner, R. S., *Analyt. Chem.* 25, 897 (1953).

- ⁸ Morris, H. E., Lacombe, R. E. en Lane, W. H., *ibid* 20, 1037 (1948).
- ⁹ Maurice, M. J., *Anal. Chim. Acta* 16, 578 (1957).
- ¹⁰ Bennowitz, R. en Tánzer, I., *Mikrochim. Acta* 1959, 835.
- ¹¹ Wilson, H. N., *Analyst* 76, 65 (1951).
- ¹² Riemann, W., Neuss, J. D. en Naiman, B., „Quantitative analysis, a theoretical approach”. — Mc Graw Hill Book Company, Inc., New York, 3e uitgave, 1951, p. 335 e.v.
- ¹³ Cimerman, Ch. en Ariel, M., *Anal. Chim. Acta* 12, 13 (1955).
- ¹⁴ Verschragen, P., *ibid* 12, 227 (1955).
- ¹⁵ Maurice, M. J., *Z. Anal. Chem.* 158, 410 (1957).
- ¹⁶ Wiggers, B. G., *Sigma* 5, 2 (1959) en 5, 62 (1959).

Discussie

Drs. P. Boumans vraagt:

Mag het verband tussen σ en μ worden berekend met behulp van de methode der kleinste kwadraten, indien het verband lineair is of hierin kan worden getransformeerd?

Antwoord:

Strikt genomen is dit niet geoorloofd, aangezien s_x en \bar{x} beide stochastische variabelen zijn en bovendien s_x niet normaal verdeeld is. In de praktijk maakt men echter geen grote fout, wanneer men de methode wel toepast. Het is dan wel noodzakelijk, dat men de beschikking heeft over een behoorlijk aantal schattingen van σ en van μ .

Anderzijds moet echter het belang van een exacte kennis van het verband tussen σ en μ niet worden overschat. De benadering van dit verband, die men verkrijgt door grafisch vast te stellen met welk van de drie in tabel I genoemde gevallen men te maken heeft, is ruim voldoende.

Ir. R. de Vletter vraagt:

Waarom zijn de analistinvloed en de laboratorium-

invloed buiten beschouwing gelaten? Deze invloeden zijn vaak groot en ze zijn van veel belang bij handelsanalyses en bij vergelijking van door verschillende bedrijven verkregen resultaten.

Antwoord:

Er zijn twee redenen te noemen waarom deze invloeden niet werden besproken. Deze beide invloeden dragen niet bij tot de nauwkeurigheid van een methode, zoals deze in het begin van de voordracht werd gedefinieerd. In de tweede plaats is het voor het behandelen van deze invloeden noodzakelijk de beginselen van de variantie-analyse te bespreken, hetgeen in het kader van dit symposium niet wel uitvoerbaar is.

Drs. E. K. Duursma vraagt:

Waarom is in tabel VII het gemiddelde verschil in vier decimalen opgegeven, terwijl de individuele verschillen slechts in twee decimalen zijn opgegeven?

Antwoord:

Volgens de gangbare afrondingsregels moet een uitkomst x op $\frac{1}{2}s_x$ tot $\frac{1}{4}s_x$ en een gemiddelde \bar{x} op $\frac{1}{2}s_{\bar{x}}$ tot $\frac{1}{4}s_{\bar{x}}$ worden afgerond. In het geval van uiteindelijke presentatie moet in tabel VII d op 0.01 en d op 0.002 worden afgerond. In verband echter met de nog uit te voeren berekening van t uit \bar{d} is deze laatste waarde niet in drie, doch in vier decimalen opgegeven. In het algemeen moet gesteld worden, dat het aanbeveling verdient eerst de noodzakelijke berekeningen uit te voeren en pas daarna tot afronding over te gaan.

Drs. H. A. van Vucht merkt op:

Inderdaad is een correcte toepassing van de afrondingsregels noodzakelijk. Veelal wordt door een te grove afronding van de afzonderlijke analyseresultaten een te gunstig of beter gezegd onjuiste standaardafwijking verkregen.

Het opsporen van fouten- bronnen in de analytische chemie *

J. AGTERDENBOS

(Analytisch Chemisch Laboratorium der Rijksuniversiteit, Utrecht)

Many errors in analytical chemistry may be classified into four different types. First there are systematic and accidental errors. Further in both classes, there are errors, the magnitude of which is independent of the quantity of the substance determined, and errors of which the magnitude is proportional to this quantity.

From very simple statistical analysis of results of suitable experiments, it may often be concluded that one or two types of errors dominate and other ones are negligible. In this way the improvement of analytical methods may be simplified.

Illustrations are given, of this application of statistics and of some other ones.

542.004.64

1. Inleiding.

We hebben vandaag al enkele malen gehoord dat er in de analytische chemie allerlei bronnen van fouten of storingen kunnen optreden, die het resultaat van de analyse beïnvloeden. Nu zal het soms voorkomen dat U direct aan de resultaten ziet dat Uw analyse-methode heel goed bruikbaar is voor het doel waarvoor U hem toepast. Dit kan bijvoorbeeld het geval zijn als U de resultaten niet beter hoeft te weten dan op 10 % nauwkeurig, terwijl U zo ziet, zonder enige statistiek toe te passen, dat U beslist geen enkele waarneming zult krijgen die meer dan 3 % afwijkt. In zo'n geval heeft het toepassen van statistiek meestal weinig zin.

In veel gevallen moeten we de bruikbaarheid van onze methode echter nog nader onderzoeken of eventueel verbeteren. Bij pogingen de methode te verbeteren kunnen we soms eenvoudig ons analytisch gevoel laten spreken en dat kan natuurlijk ook best nog succes hebben.

In andere gevallen echter kan het efficiënter zijn om eens na te gaan of we te weten kunnen komen welke foutenbronnen er in hoofdzaak optreden. Hebben we die dan min of meer gelokaliseerd, dan kunnen we aan de hand hiervan onze pogingen tot verbetering opzetten.

Een eenvoudig geval treedt hierbij op indien de grootte van de foutenbronnen onafhankelijk is van de hoeveelheid (concentratie) van het bepaalde produkt, althans binnen het bereik waarin onze gehalten of hoeveelheden liggen. Vaak echter blijkt de grootte van de storing af te hangen van de hoeveelheid stof die is bepaald. We zullen in deze voordracht zowel van een eenvoudig als van een meer gecompliceerd geval een voorbeeld bespreken, en nagaan hoe het opsporen van de foutenbronnen hierdoor kan worden vergemakkelijkt. Tot slot wilde ik U dan nog iets vertellen over de mate waarin bij de spectrofotometrie het optimale extinctiegebied beïnvloed wordt door foutenbronnen in dat deel van de analyse dat aan de meting van de extinctie vooraf gaat.

* Voordracht, gehouden op het symposium over „De toepassing van statistiek bij de chemische analyse” op 31 maart 1960 te Amsterdam.

2. Eenvoudig geval: constante standaard-deviatie.

2. 1. Voorbeeld.

We zullen hiervan een voorbeeld bespreken waarbij het mogelijk bleek de totale standaard-deviatie die bij een analyse optrad te splitsen in componenten die aan afzonderlijke bewerkingen konden worden toegeschreven. Het betrof hier een zwavelbepaling waarbij eerst H_2S wordt ontwikkeld in een zure oplossing van een sulfide (bewerking 1). Dit H_2S wordt geabsorbeerd in een oplossing van overmaat koperacetaat (bewerking 2). Het niet neergeslagen koper werd complexometrisch getitreerd (bewerking 3). De totale bewerking werd een aantal malen uitgevoerd waarbij de totale standaard-deviatie $s_{1, 2, 3}$ 0.059 mg bleek te bedragen. Verder bleek het mogelijk de standaard-deviatie van bewerking 3 afzonderlijk te bepalen ($s_3 = 0.013$ mg) en de standaard-deviatie van de gezamenlijke uitgevoerde bewerkingen 2 en 3 ($s_{2, 3} = 0.034$ mg). Door gebruik te maken van de regel dat de varianties (dus de kwadraten van de standaard-deviaties) additief zijn vinden we:

$$s_1^2 + s_2^2 + s_3^2 = 0.059^2$$

$$s_2^2 + s_3^2 = 0.034^2$$

$$s_3^2 = 0.013^2$$

en hieruit is te berekenen: $s_1 = 0.048$ mg en $s_2 = 0.031$ mg.

Hieruit ziet men bijvoorbeeld dat pogingen om de analyse-methode minder spreiding te geven zich vooral op de beide eerste bewerkingen zullen moeten richten. Voor nadere bijzonderheden wordt naar de oorspronkelijke literatuur van Maurice¹ verwezen.

2. 2. Opmerkingen.

1. Men moet er rekening mee houden dat de waarde voor s , die uit een bepaald aantal waarnemingen wordt berekend, slechts een ruwe schatting is van σ , de waarde die bij zeer veel waarnemingen zou worden gevonden. Bij meer dan 25 à 30 waarnemingen kan men zeggen dat s een normale verdeling heeft met standaard-deviatie $s/\sqrt{2n}$. Omdat 95 % van de populatie dus tussen $\bar{x} + 2s$ en

$\bar{x} - 2s$ ligt, vinden we hier dat bijv. bij $n = 32$ en dus $s_s = s/\sqrt{2} \times 32 = s/8$, de waarde van σ tussen $s + 2s/8$ en $s - 2s/8$ zal liggen, dus tussen $3/4$ en $5/4 \cdot s$. Voor kleine aantallen waarnemingen kan een betrouwbaarheidsinterval van σ gevonden worden met behulp van een z.g. χ^2 -tabel.

2. Men moet goed bedenken dat van de waarden van s_1 en s_2 , die boven berekend zijn, geen vrijheidsgraden bekend zijn. Men kan hierop de F-toets dus niet rechtstreeks toepassen.

3. Gecompliceerd geval. De standaard-deviatie hangt af van de te bepalen hoeveelheid stof en er zijn systematische afwijkingen.

In het eerste voorbeeld werd verondersteld dat steeds met ongeveer eenzelfde hoeveelheid te analyseren component gewerkt werd, zodat een eventuele afhankelijkheid van de grootte van de storing van de hoeveelheid van die component niet van belang was. Verder werd alleen met toevallige, en niet met systematische afwijkingen rekening gehouden.

In dit hoofdstuk zullen we de verschillende typen foutenbronnen statistisch onderzoeken en in hoofdstuk 4 de chemische interpretatie van de mogelijke conclusies nagaan.

3. 1. Classificatie van de afwijkingen.

Het komt in de analyse nogal eens voor dat de gemeten waarde waaruit de hoeveelheid te analyseren stof wordt berekend, niet uitsluitend wordt veroorzaakt door de te analyseren stof, maar dat deze met een „blanco” waarde moet worden gecorrigeerd. Deze correctie kan van tweeërlei aard zijn: soms moeten de gevonden waarden met een bepaald bedrag worden vermeerderd of verminderd, in andere gevallen moet met een bepaalde factor worden vermenigvuldigd. In beide gevallen zou zonder het aanbrengen van de correctie een fout ontstaan in het gevonden resultaat. De grootte van de afwijking is in het eerste geval onafhankelijk van de te bepalen hoeveelheid stof, in het tweede geval echter daarmee evenredig. In het eerste geval treedt een systematische fout op waarvan de absolute waarde constant is, en die we a zullen noemen, in het tweede geval is de relatieve (of procentuele) fout constant; deze duiden we met r aan.

Een duidelijk voorbeeld waarin beide foutenbronnen optreden is van *Benedetti-Pichler*². Hij bepaalde Al gravimetrisch als Al_2O_3 en gebruikte daarbij voor de precipitatie NH_4OH , die met SiO_2 was verontreinigd. De resultaten geeft tabel I. U ziet in de derde kolom dat min of meer onafhankelijk van de hoeveelheid Al_2O_3 die aanwezig was ruim 20 mg Al_2O_3 te veel werd gevonden. Meestal kan men een dergelijke fout corrigeren door het uitvoeren van een blanco-bepaling. In dit geval echter niet: bij de blanco werd helemaal geen neerslag gevonden, en we kunnen dit ook best begrijpen: het Al_2O_3 dat als drager voor het SiO_2 fungeert is bij de blanco niet aanwezig. En hiermee zien we dan meteen dat er een belangrijk geval is waarin we niet voor een blanco kunnen corrigeren door een afzonderlijke bepaling uit te voeren: nl. het geval, waarin de aanwezigheid van de te bepalen component essentieel is voor het optreden van die blanco. Dit komt niet alleen bij de gravimetrie voor.

In de titrimetrie kennen we het verschijnsel van de geïnduceerde reacties, die alleen optreden als de te analyseren component ook aanwezig is. Zo vindt de oxidatie van chloride-ionen door $KMnO_4$ vooral plaats in aanwezigheid van ijzer-ionen. Hiermee moet bij de ijzer-bepaling volgens *Zimmermann-Reinhardt* rekening gehouden worden. Soms gebeurt dit door bij de blanco ferri-ionen toe te voegen. Ook bij de titratie van ijzer met $K_2Cr_2O_7$ met als indicator het difenylamine, of één van de derivaten hiervan, is het uitvoeren van een blanco op de gebruikelijke wijze niet uitvoerbaar. Op deze laatste titratie komen we in een voorbeeld terug (zie 3. 2.).

Als we tabel I nog eens bekijken dan zien we dat ook het gebruik van goede NH_4OH resultaten gaf die te hoog lagen. Hier was echter de afwijking in grammen niet constant, maar de procentuele fout. De afwijkingen bleken toe te schrijven aan adsorptie van vocht aan de Al_2O_3 -neerslagen: de hoeveelheid ge-adsorbeerd water is blijkbaar min of meer evenredig aan de hoeveelheid Al_2O_3 .

Tabel I. Gravimetrische Al-bepaling als Al_2O_3 .

Al_2O_3 aanwezig	Al_2O_3 gevonden met oude NH_3		Al_2O_3 gevonden met vers gedest. NH_3			
	fout		fout			
mg	mg	mg	%	mg	mg	%
107.7	128.8	21.1	16.4	108.7	1.0	0.9
215.4	238.4	23.0	9.7	217.8	2.4	1.0
323.1	348.9	25.8	7.4	325.8	2.7	0.8
430.8	458.8	28.0	6.1	435.2	4.4	1.0
0.0	0.0	0.0		0.0	0.0	

Zowel bij de absolute als bij de relatieve constante systematische afwijking zijn de hierdoor veroorzaakte afwijkingen van het juiste resultaat niet precies constant. Er zijn nog toevallige fouten a.h.w. overheen gesuperponeerd. Die kunnen bijvoorbeeld veroorzaakt zijn omdat niet altijd precies evenveel NH_4OH is toegevoegd, omdat het oppervlak van het gegloeide Al_2O_3 per gewichtseenheid niet bij ieder experiment precies hetzelfde was, enz..

Behalve de toevallige fouten die optreden als fluctuaties in systematische fouten, zijn er andere die zelfstandig kunnen voorkomen, zoals afleesfouten en weegfouten.

Samenvattend komen we er dus toe onderscheid te maken tussen systematische en toevallige fouten, waarbij in beide gevallen de grootte van de fout onafhankelijk van de gemeten hoeveelheid kan zijn, of daarmee evenredig.

De grootte van de relatieve systematische fout r bedraagt in bovenstaand geval 0.9 %, met standaard-deviatie 0.1 %. De grootte van de absolute systematische fout a bedraagt (na correctie voor de vocht-adsorptie) in de vier experimenten resp. 20.1, 20.6, 23.1 en 23.6 mg, gemiddeld dus 21.8 mg met standaard-deviatie van 1.8 mg. (De waarnemingen geven de indruk dat de absolute afwijking ook enigszins afhankelijk is van de hoeveelheid Al_2O_3 , we gaan hier niet nader op in).

(Zie ook nog opm. 3, bij 3.2.)

3. 2. Berekening van de afwijkingen.

In het voorbeeld van de Al_2O_3 -bepaling kon de grootte van de NH_4OH -fout worden nagegaan door ook experimenten met zuivere NH_4OH uit te voeren, dus door een foutenbron weg te nemen. Meestal is dit echter niet mogelijk. Het leek daarom nuttig eens na te gaan hoe men ook bij de analyses, waarbij de foutenbronnen niet eenvoudig kunnen worden vermeden, toch de waarden van a en r , en zo mogelijk de standaard-deviaties in deze waarden, kan berekenen.

Het bleek niet eenvoudig in de literatuur uitkomsten van experimenten te vinden aan de hand waarvan een dergelijke berekening kan worden geïllustreerd. Daarom werden in ons laboratorium enkele series experimenten uitgevoerd, waarvan verwacht kon worden, dat ze de gewenste foutenbronnen in geschikte grootte zouden bevatten. Ons voorbeeld betreft de titratie van $Fe(II)$ met $K_2Cr_2O_7$ met als indicator difenylamine.

Bij deze titratie treedt een titratiefout op, die vóór al veroorzaakt wordt doordat de redox-indicator ook geoxideerd moet worden en misschien ook ten dele omdat de omslag-potentiaal van de indicator niet precies samenvalt met de equivalentie-potentiaal van de titratie. We zullen ons met de verklaring verder niet bezig houden.

Om de grootte van deze systematische fout a te bepalen bevelen *Belcher* en *Nutten*³ aan om twee verschillende hoeveelheden ijzer te titreren van eenzelfde oplossing, bijv. 25 ml en 10 ml. Stel we gebruiken daarbij resp. 37.70 en 15.20 ml $K_2Cr_2O_7$ -oplossing. We berekenen dan:

$(37.70 - a)/(15.20 - a) = 25/10$, dus $a = 0.20$ ml. De „blanco” komt dus overeen met 0.20 ml van de $K_2Cr_2O_7$ -oplossing. Verder zien we dat 1.5 ml van de $K_2Cr_2O_7$ -oplossing met 1 ml van de ijzeroplossing overeenkomt. We kennen dus ook de relatieve sterkten van de oplossingen. Hiervan is in dit geval gebruik gemaakt om na te gaan of het gebruikte *Mohr's* zout (p.a.) het theoretische gehalte aan $Fe(II)$ bevatte, aangenomen dat het gedroogde $K_2Cr_2O_7$ (p.a.) een gehalte van 100.0 % had. We zullen aannemen dat een onjuist gehalte aan $Fe(II)$ de enige oorzaak is van een procentueel constante storing (zie ook hoofdstuk 4).

Bij onze bepalingen werd het *Mohr's* zout niet in een voorraadoplossing klaar gemaakt en gepipetteerd, om de kans op oxidatie tegen te gaan, maar werd voor ieder experiment afzonderlijk afgewogen. De beide series A en B (tabel II) werden in de volgorde 1, 2, 3, 16 uitgevoerd.

Vooraf werd d.m.v. een lotingstabel vastgesteld welke nummers bij serie A (ca. 1600 mg) en welke bij serie B (ca. 200 mg) zouden behoren. Dit alles diende om tijdsinvloeden zoveel mogelijk te vermijden.

Uit kolom d krijgt men de indruk dat de bepaling slechts zeer geringe afwijkingen heeft, tenminste voor zover het serie A betreft. Serie B geeft de indruk dat of een blanco optreedt die met ca. 5 mg *Mohr's* zout overeenkomt, of dat het $Fe(II)$ gehalte van het *Mohr's* zout ca. 2.5 % te hoog is, of dat beide foutenbronnen optreden. We zien uit dit verschil tussen de resultaten van series A en B hoe gevaarlijk het is om conclusies te trekken uit één serie waarnemingen die alle met ongeveer dezelfde hoeveelheid te analyseren component zijn uitgevoerd.

Om nadere gegevens over de beide mogelijke foutenbronnen te krijgen werden de waarnemingen uit beide series op de boven reeds beschreven wijze gecombineerd. Hieruit is a uit de eerste rij dus berekend uit

$$\frac{40.94 - a}{5.35 - a} = \frac{1606.5}{204.6}$$

waaruit volgt $a = 0.16$. Dat betekent dat 0.16 ml 0.1 N $K_2Cr_2O_7$ nodig is om de indicator te oxideren (afgezien dus van andere oorzaken, zie 4. 1.).

De berekening van r ging als volgt: $(1606.5 - 204.6)$ mg *Mohr's* zout werd getitreerd met $40.94 - 5.35 = 35.59$ ml 0.1 N $K_2Cr_2O_7$. Deze hoeveelheid is equivalent met 1394.8 mg *Mohr's* zout, er is dus 0.5 % te weinig teruggevonden. Het $Fe(II)$ -gehalte van het *Mohr's* zout is dus 0.5 % te laag.

Tabel III geeft de resultaten van alle — volgens een lotingstabel gecombineerde — paren waarnemingen. Men vindt $\bar{a} = 0.154$ met s.d. = 0.032 en $\bar{r} = -0.41$ met s.d. = 0.19. De betrouwbaarheidsintervallen zijn bij $\alpha = 0.05$:

$$0.154 + \frac{2.36}{\sqrt{8}} \cdot 0.032 > a > 0.154 - \frac{2.36}{\sqrt{8}} \cdot 0.032 \text{ of}$$

$$a = 0.13 \text{ à } 0.18 \text{ ml } 0.1 \text{ N } K_2Cr_2O_7 \text{ en}$$

$$-0.41 + \frac{2.36}{\sqrt{8}} \cdot 0.19 > r > -0.41 - \frac{2.36}{\sqrt{8}} \cdot 0.19 \text{ of}$$

$$r \text{ is } -0.25 \text{ à } -0.57 \%$$

De blanco bedraagt dus 0.13 à 0.18 ml 0.1 N $K_2Cr_2O_7$ het $Fe(II)$ gehalte is 0.25 à 0.57 % te laag.

Tabel III. Berekening van de systematische fout uit series A en B.

Waarnemingen	Abs. fout a ml. 0.1 N	Rel. fout r %
1- 7	0.16	-0.5
2-16	0.15	-0.5
10- 3	0.14	-0.3
15- 4	0.11	-0.3
5-14	0.22	-0.8
12- 6	0.14	-0.3
8- 9	0.15	-0.2
13-11	0.16	-0.4

Opmerkingen:

1. Men kan \bar{r} ook berekenen uit de waarde $\bar{d} = -0.64$ en $\bar{d} = +5.2$ uit tabel II. Men vindt dan uiteraard ook 0.41 %, maar geen betrouwbaarheidsinterval.

2. De berekende betrouwbaarheidsintervallen gelden slechts als alleen fouten van het absoluut constante type en fouten van het relatief constante type optreden.

3. Men bedenke dat de waarden 0.032 en 0.19 voor de standaarddeviaties ieder te wijten zijn aan de spreidingen in a en r beide. Zo is niet steeds exact evenveel indicator toegevoegd; deze storing in a heeft ook spreiding in r tot gevolg.

Een schatting van de invloed van s_a en s_r ieder krijgt men uit de standaard-deviaties die uit tabel II, kolom d berekend zijn: 1.8 mg voor serie A, 0.93 mg voor serie B.

Tabel II. Titratie met 0.1 N $K_2Cr_2O_7$.

Serie	Waarneming nr.	Mohr's zout	Kal. bichr.	Mohr's zout	c — a	c — a
		afgewogen mg a	ml b	uit b ber. mg c	mg d	% e
A	1	1606.5	40.94	1604.4	—2.1	—0.13
A	2	1591.7	40.57	1589.9	—1.8	—0.11
A	5	1577.3	40.16	1573.8	—3.5	—0.22
A	8	1603.3	40.97	1605.6	2.3	0.14
A	10	1599.2	40.83	1600.1	0.9	0.06
A	12	1603.0	40.92	1603.6	0.6	0.04
A	13	1586.3	40.48	1586.4	0.1	0.01
A	15	1608.5	41.03	1607.9	—0.6	—0.04
B	3	202.7	5.30	207.7	5.0	2.5
B	4	204.4	5.31	208.1	3.7	1.8
B	6	211.3	5.51	215.9	4.6	2.2
B	7	204.6	5.35	209.7	5.1	2.5
B	9	197.2	5.17	202.6	5.4	2.7
B	11	210.9	5.52	216.3	5.4	2.6
B	14	207.4	5.47	214.4	7.0	3.4
B	16	207.4	5.42	212.4	5.0	2.4

Noemen we de gezochte waarden s_a (mg) en s_r (%)

$$\text{dan is bij 1600 mg: } 1.8^2 = s_a^2 + \left(\frac{1600}{100} \cdot s_r\right)^2$$

$$\text{en bij 200 mg: } 0.93^2 = s_a^2 + \left(\frac{200}{100} \cdot s_r\right)^2$$

Hieruit volgt $s_a = 0.9$ mg en $s_r = 0.1$ %. Omdat de waarden 1.8 en 0.93 uit slechts weinig waarnemingen bekend zijn is een grote fout in deze waarden zeer wel mogelijk.

De splitsing van de totale s.d. in een absoluut en een relatief constant deel, in een geval waarin bij veel concentraties gemeten wordt, is in ⁴ besproken.

4. Foutenbronnen.

In het voorgaande werd a steeds aan de blanco toegeschreven en r aan een onjuist Fe(II)-gehalte van het Mohr's zout. Hoewel in het beschreven geval deze oorzaken wellicht een belangrijke rol hebben gespeeld, moeten we ons toch realiseren dat in het algemeen ook andere foutenbronnen kunnen optreden. Daarom is hieronder een overzicht van veel voorkomende foutenbronnen gegeven. Men bedenke dat niet iedere foutenbron altijd in dezelfde groep thuis hoort, en dat sommige fouten een gedeeltelijk absoluut, en een gedeeltelijk relatief karakter hebben. Zie ook de voorgaande voordracht ⁵, tabel I.

4. 1. Systematische fouten.

a. Absolute fouten.

1. Verontreiniging in reagentia (in constante hoeveelheid toegevoegd). Hierbij kan zowel de aanwezigheid van de te bepalen component tot fouten leiden als de aanwezigheid van andere stoffen die in essentiële opzichten soortgelijke eigenschappen hebben. Die „andere” stoffen kunnen ook in het monster zitten.
2. Verontreiniging uit apparatuur (glas, filterpapier).
3. Fouten door oplosbaarheid.
4. Indicatiefouten.

b. Relatieve fouten.

1. Onjuiste samenstelling van de standaard, bijv. door onzuiverheid of onjuist kristalwater-gehalte.
2. Onjuiste maatkolf, of onjuiste gebruikstemperatuur. Voor organische oplosmiddelen is de invloed ca. 0.1 %, voor water ca. 0.02 % per °C.
3. Absorptie van water aan neerslagen.
4. Niet-kwantitatief verlopende processen (uitschudden).

4. 2. Toevallige fouten.

1. De meeste systematische fouten zijn niet precies constant, zodat er een toevallige fout overheen gesuperponeerd wordt.
2. Sommige systematische fouten gaan over in toevallige fouten als een blanco wordt afgetrokken, of tegen een blanco wordt gemeten. De invloed van verschillende methodes waarmee in de colorimetrie de blanco in rekening gebracht wordt, op de nauwkeurigheid van de analyse, wordt elders besproken ⁶.
3. Weegfouten en afleesfouten. Het is niet altijd eenvoudig te zeggen of deze tot absolute of relatieve toevallige fouten zullen leiden.

4. 3. Interpretatie van analyse-resultaten.

Een literatuur-voorbeeld moge verduidelijken dat ook een betrekkelijk gering aantal waarnemingen reeds tot nuttige conclusies kan leiden. *B. R. Sant* ⁷ vond bij een nitriet-bepaling door oxidatie met $K_3Fe(CN)_6$ in aanwezigheid van $ZnSO_4$ de in tabel IV gegeven resultaten. Het verschil tussen de aanwezige en de gevonden hoeveelheid nitriet, uitgedrukt in mg lijkt toe te nemen met de hoeveelheid nitriet. De (hier niet behandelde) rangcorrelatie-toets bevestigt dit. Het in procenten uitgedrukte verschil lijkt beter constant. De gemiddelde waarde bedraagt 0.6% met standaarddeviatie 0.2%. Deze waarden leiden tot het betrouwbaarheidsinterval met grenzen $0.6 + (0.2 \cdot 2.57)/\sqrt{6}$ en $0.6 - (0.2 \cdot 2.57)/\sqrt{6}$ dus 0.2 en 1.0 %.

Men moet goed bedenken dat dit interval is be-

rekend onder de veronderstelling dat er helemaal geen absolute fout zou zijn, en dat hoeft niet helemaal juist te zijn. De conclusie dat er een relatief constante systematische fout zou zijn, lijkt echter zeker gerechtvaardigd en bij nader onderzoek van de methode zou het dus voor de hand liggen te beginnen met de onder 4. 1. b genoemde foutenbronnen.

Tabel IV. De bepaling van nitriet⁷

NO ₂ aanwezig mg	NO ₂ gevonden mg	Vershil mg	Vershil %
2.281	2.277	-0.004	-0.2
4.580	4.544	-0.036	-0.8
11.50	11.39	-0.11	-1.0
22.81	22.77	-0.04	-0.2
34.30	34.16	-0.14	-0.4
46.00	45.54	-0.46	-1.0

Wellicht komt U er niet direct toe om zich te verdiepen in de statistische fundering van de conclusies die uit tabel IV getrokken zijn. Ook dan kan het naar ik meen nuttig zijn om een dergelijke serie waarnemingen eens op de hier beschreven wijze te bekijken. Immers ook zonder dat U een betrouwbaarheidsinterval berekent zult U de indruk krijgen dat er in hoofdzaak een systematische fout optreedt die evenredig is met de geanalyseerde hoeveelheid. En U zult daaruit concluderen dat U bij pogingen tot het verbeteren van Uw analysemethode met bepaalde foutenbronnen niet of nauwelijks rekening hoeft te houden, met andere foutenbronnen echter wel. Hierkunt U zich uiteraard veel tijd en moeite besparen.

De oorzaak van de spreiding in de resultaten van tabel IV is hierbij nog niet ter sprake gekomen. Het lijkt ook niet goed mogelijk dit op grond van zo weinig waarnemingen te doen. Door het verrichten van een aantal parallel-waarnemingen bij ieder van de toegepaste hoeveelheden zouden hierover wellicht conclusies te trekken zijn.

(Op de mogelijke gevolgen van het feit dat Sant voor verschillende hoeveelheden nitriet, werkt met oplossingen van verschillende concentratie gaan we hier niet in).

5. Het optimale extinctie-gebied in de spectrofotometrie.

Tot slot wilde ik U dan nog iets vertellen over het optimale gebied bij spectrofotometrische analyses. Het blijkt nl. dat de chemische fouten die gemaakt worden vóór de extinctie-meting plaats vindt, belangrijk van invloed kunnen zijn op de ligging van dit gebied. Ook hier blijkt het zin te hebben deze chemische storingen te splitsen in die welke over het gehele concentratiegebied constant zijn, en die welke evenredig zijn met de onderzochte hoeveelheid.

5. 1. De extinctie-meting.

Het mag bekend verondersteld worden dat Ringbom⁸ uit de wet van Lambert-Beer afleidde dat de fout in een concentratie-meting afhangt van de fout in de transmissie volgens $dc/c = (0.43 \cdot 10^6)/E \cdot dT/T$ ($E =$ extinctie); voor een aantal metingen is dus $s_c = (0.43 \cdot 10^6)/E \cdot s_T$.

Nu is s_T over het gehele gebied min of meer constant voor ieder toestel en we kunnen dan ook voor een bepaalde waarde van s_T het verband tussen s_c en E berekenen. De waarde van s_T bedraagt veelal tussen enkele tienden procenten en 1%, bij zeer goede toestellen echter veel minder, bij slechte meer. Voor $s_T = 0.2\%$ is het verloop als in figuur 1 aangegeven. Bij $E = 0.43$ heeft de curve een minimum van $2.7 s_T$ dus hier is s_c dan 0.5% . Tussen $E = 0.1$ en 1.2 bedraagt s_c steeds minder dan het dubbele van deze waarde.

Men schrijft hierom veelal voor, spectrofotometrische analyses uit te voeren onder omstandigheden waarbij extincties tussen genoemde grenzen worden gemeten. Soms worden de grenzen ook iets nauwer genomen, bijvoorbeeld $E = 0.2$ tot 0.8 , waarbij de waarde van s_T niet meer dan ca. 20% boven het minimum stijgt.

Er moge op gewezen worden dat deze beschouwing niet voor alle types apparaten geldt; soms wordt de curve iets verschoven. Zie o.a. Kortüm⁹.

5. 2. Chemische invloeden.

Bij bovengenoemde aanbeveling in een bepaald extinctie-gebied te meten, wordt dus geen rekening gehouden met de chemische storingen die de meting vooraf gaan. We hebben hier meestal niet te maken met systematische fouten omdat die door het opstellen van de ijkcurve wegvallen. De andere storingen splitsen we weer in die welke evenredig zijn met de hoeveelheid te analyseren component (s_r) en die welke daarvan onafhankelijk zijn (s_a). Ook de invloed van ieder van deze storingen is in fig. 1 aangegeven. Hierin is voor $s_r = 2\%$ aangenomen; voor s_a kan de waarde het best in extinctie-eenheden worden uitgedrukt. In de figuur is $s_a = 0.01$ genomen.

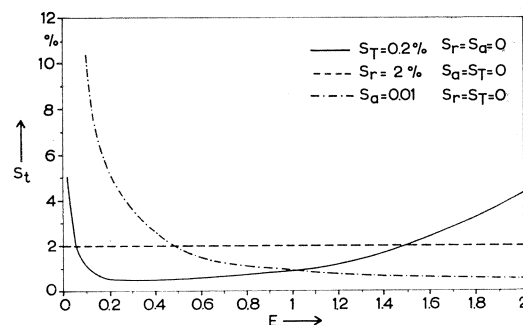


Fig. 1.

De totale standaard-deviatie s_t zal via de regel van de additiviteit van varianties kunnen worden berekend. Figuur 2 geeft de curve voor het geval s_T met s_a is gecombineerd en voor de combinatie van s_T met s_r . In het eerste geval wordt de curve naar rechts verschoven. Het optimum komt bij $E = 1.0$ te liggen, en het optimale gebied ligt niet meer tussen 0.1 en 1.2 , maar tussen 0.4 en 1.7 . Er treedt dus een aanzienlijke verschuiving op. Door s_T en s_r te combineren wordt de curve veel vlakker. Het gebied waarbinnen de waarde niet boven het dubbele van het minimum stijgt ligt nu tussen $E = 0.03$ en $E = 1.9$, dus dit is veel groter geworden.

In gevallen waarin s_a en s_r kleiner zijn of s_T groter is, zijn de verschillen met het door Ringbom aangegeven

gebied uiteraard niet zo groot. Vaak zal het echter de moeite waard zijn met een en ander rekening te houden. Voor verdere getallen-waarden zie men¹⁰. In deze publikatie moet op blz. 403 de term s_u^2/E^2 door $(100s_u)^2/E^2$ vervangen worden.

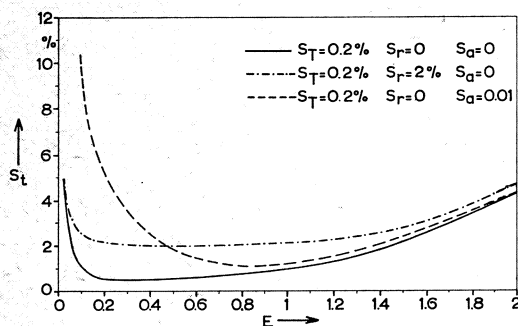


Fig. 2.

6. Samenvatting.

Bij analytische bepalingen kunnen systematische en toevallige fouten worden onderscheiden. In beide groepen kan men veelal nog onderscheid maken tussen die fouten welke evenredig zijn aan de geanalyseerde hoeveelheid component (relatieve fouten)

en die welke daarvan onafhankelijk zijn (absolute fouten).

Op grond van waarnemingen, die bij verschillende hoeveelheden te analyseren component worden uitgevoerd, is het veelal mogelijk een indruk te krijgen welke foutenbronnen tot de totale fout bijdragen. Hierdoor kan het verbeteren van analyse-methodes vergemakkelijkt worden. In de spectrofotometrie kan men er rekening mee houden bij het gebruiken van een optimaal extinctiegebied.

De schrijver dankt Ir. A. R. Bloemena voor een waardevolle discussie, en de heer N. Martinus voor het uitvoeren van een aantal experimenten.

- ¹ Maurice, M. J., Z. anal. Chem. **158**, 271, (1957).
- ² Benedetti-Pichler, A. A., Ind. Eng. Chem. Anal. Ed. **8**, 373 (1936).
- ³ Belcher, R., and Nutten, A. J., Quantitative inorganic analysis, London, Butterworth, 1960.
- ⁴ Agterdenbos, J., Z. anal. Chem. **159**, 202 (1958).
- ⁵ Maurice, M. J., Chem. Weekblad.
- ⁶ Agterdenbos, J., Z. anal. Chem. **157**, 161, (1957).
- ⁷ Sant, B. R., Anal. Chim. Acta **19**, 523, (1958).
- ⁸ Ringbom, A., Z. anal. Chem. **115**, 332, (1938/39).
- ⁹ Kortüm, G., Kolorimetrie, Photometrie und Spektrometrie. Göttingen en Heidelberg, Springer, 1955.
- ¹⁰ Agterdenbos, J., Z. anal. Chem. **154**, 401, (1957).

