

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

S 153 (V 10)

Statistiek en Practijk

door

Prof.Dr J. Hemelrijk

1954

1. LAPLACE heeft de waarschijnlijkheidsrekening genoemd: "het gezonde verstand teruggebracht tot een berekeningsmethode; het exact maken van zaken, die verstandige mensen intuïtief aanvoelen, zonder dat zij zich er volledig rekenschap van kunnen geven." Daar statistiek toegepaste waarschijnlijkheidsrekening is, slaat deze opmerking ook op de statistiek. Dit terugbrengen van (natuurlijk slechts een gedeelte van) het gezonde verstand tot een berekeningsmethode heeft echter niet alleen een exactisering, maar ook een verscherping tot gevolg. De precisering, die in een quantitative beoordeling van waarnemingen gelegen is in vergelijking met een kwalitatieve, geeft uitkomst in gevallen, waarin het "gezonde verstand" alléén blijft twijfelen.

Voorbeeld Laten A en B twee middelen zijn om infecties te bestrijden. Ter vergelijking van de genezende werking gebruikt men middel A in 100 gevallen en B in 100 andere gevallen. Vindt men nu als resultaat:

(1)

	succes	mislukking
A	95	5
B	20	80

dan zal men, zonder daarbij berekeningen te hulp te roepen, concluderen, dat A beter werkt dan B. Men begrijpt zonder meer, dat een dergelijk resultaat niet verkregen zou zijn, indien B even goed of beter was geweest dan A.

Indien men echter met een minder uitgesproken resultaat te maken heeft, wordt het moeilijker om tot een beslissing te komen. Stel b.v., dat men met beide middelen 15 proeven heeft gedaan en men vindt

(2)

	succes	mislukking
A	15	0
B	9	6

of

(3)

	succes	mislukking
A	14	1
B	9	6

dan kan men zeker niet zonder meer zeggen, dat een dergelijk resultaat niet verkregen zou zijn, indien B even goed als A was. Immers door wat men gewoonlijk "toevallige omstandigheden" noemt, kan het aantal successen bij B wel wat kleiner uitgevallen zijn dan bij A.

In dergelijke gevallen is een statistische precisering van belang. Deze leidt hier tot het volgende resultaat: een uitkomst als (2), of een nog groter verschil tussen A en B, zal, als A en B even goed zijn, slechts in minder dan 1 op de 100 experimenten van deze omvang voorkomen. Op grond hiervan kan men (2) dus beschouwen als een duidelijke aanwijzing van de superioriteit van A t.o.v. B. Voor geval (3) vindt men echter, dat een dergelijk (of een groter) verschil in resultaat tussen A en B 1 op de 12 keer gevonden zal worden (als A en B even goed zijn) en dit is dus eigenlijk niets bijzonders; men kan op grond hiervan niet besluiten, dat A beter dan B zou zijn.

Hier komen dus twee verschijnselen naar voren. Ten eerste kan men statistische hulpmiddelen soms tot beslissingen komen in gevallen, waarin men zonder die hulpmiddelen zou blijven twijfelen. Anderszijds echter kan een kleine wijziging in de resultaten soms voldoende zijn, om de conclusie sterk te beïnvloeden.

2. Het feit, dat kleine veranderingen in de resultaten (b.v. één succes meer of minder, zoals boven) de conclusie sterk kan beïnvloeden, is een reden, om zorgvuldig te experimenteren, om te voorkomen dat dergelijke veranderingen op zouden treden door oorzaken, die met de te onderzoeken vraag niet te maken hebben. Om te zien, op welke valstrikken wij verdacht meoten zijn, zullen wij een ietwat precieze omschrijving van de waarschijnlijkheidsrekening moeten geven dan de bovenstaande van LAPLACE afkomstige. Wij citeren er daarom (in vrije vertaling) één van J. NEYMAN (blz. 156 van zijn "Lectures and Conferences on Mathematical Statistics and Probability", Washington 1952). Hij zegt: "de waarschijnlijkheidsrekening houdt zich bezig met de algemene vraag, hoe vaak het één of andere verschijnsel voor zal komen in experimenten van een gespecificeerde aard, waarbij het toeval een rol speelt. Wil men dan hierover iets kunnen zeggen, dan moet men iets weten over de "werkwijze van het toeval" bij deze experimenten. Is hierover niets gegeven, dan

zijn er geen conclusies (behalve triviale) mogelijk. Heeft men b.v. gegevens over 10.000 boerderijen in de V.S., maar is het niet bekend op welke wijze deze 10.000 uit alle boerderijen gekozen zijn, dan heeft men aan deze gegevens vrijwel niets. Zijn zij echter op een welomschreven wijze, b.v. door loting met gelijke kansen ("aselect") uit alle boerderijen gekozen, dan kan men statistische bewerkingen op de gegevens toepassen. Om terug te keren tot ons voorbeeld van de twee middelen A en B: indien er niets gegeven is over de wijze van keuze der twee groepen patiënten, waarop A resp. B is toegepast, dan kan men zelfs aan (1) geen betekenis toekennen. Immers dan is het mogelijk, dat A alleen is toegepast op patiënten met een zwakke of gemakkelijk geneesbare infectie en B alleen op patiënten met een hardnekkige infectie, of dat A in één land en B in een ander land is gebruikt, zodat de groepen patiënten in het geheel niet te vergelijken zijn. Dergelijke grove fouten in een proefopzet worden tegenwoordig niet zo vaak meer gemaakt, maar het zijn de minder in het oog lopende - die echter van dezelfde aard zijn - waar men nog voor op moet passen. Zo kan het b.v. voorkomen, dat de volgorde van een aantal quantitative biochemische bepalingen invloed op de uitkomsten heeft en daar is men soms niet op bedacht. Dergelijke moeilijkheden kan men steeds voorkomen - ook als men ze zich niet realiseert - door het "toeval" in een bepaald keurslijf dwingen, waardoor het toeval zich op een bepaalde, de statisticus bekende, wijze moet gedragen. Men bereikt dit b.v., door de volgorde der proeven door loting te bepalen. Bij het behandelde voorbeeld kan men b.v. de 30 patiënten door loting in twee groepen verdelen en vervolgens door loting bepalen welke groep men met A en welke men met B zal behandelen. Gebruikt men daarbij een lotingsmechanisme, waarvan de statistische eigenschappen bekend zijn, dan kan men het systematisch optreden van foute conclusies voorkomen. Er kan nauwelijks voldoende aangedrongen worden op het volgen van dit soort nauwkeurige methoden bij het opzetten van proeven, daar anders het voordeel van de statistiek: het verkrijgen van conclusies in gevallen, waarin men op het oog geen beslissing kan treffen, verloren gaat door het nadeel: dat deze conclusies vaker fout zijn dan men op grond van de statistische berekeningen vermoedt.

3. Het feit, dat bij kleine waarnemingsreeksen kleine verschillen in de waarnemingen tot een aanzienlijk verschil in de conclusie kunnen leiden, geeft wel eens aanleiding tot verkeerde opvattingen over de praktische betekenis van de statistiek in dergelijke gevallen. Wij zullen dit toelichten aan een weinig bekend voorbeeld, dat ook op andere punten interessant is. Zo zeggen b.v. DEEN, HESSELINK, MARGADANT en TIRION ("De gerechtelijke schriftvergelijking", T.v. Strafrecht 60 (1951) 314-319): "Men mag dus kansberekeningen zeker niet toepassen op een beperkt materiaal, want dan hebben kleine verschillen een veel te grote invloed op het percentage", d.w.z. op de uitkomst. Hier treedt een typische misvatting op, die - daar het de gerechtelijke schriftvergelijking betreft - gemakkelijk tot ernstige gerechtelijke dwalingen kan leiden. De zaak is nl. deze, dat de genoemde auteurs optreden als deskundigen voor de rechtbank bij allerlei zaken, waarbij valsheid in geschrifte of anonieme brieven e.d. in het geding zijn en waarbij het wenselijk is vast te stellen, of een bepaalde brief of een (eventueel vervalste) handtekening van een bepaalde persoon afkomstig is of niet. De beoordeling van deze vraag op de onder schriftdeskundigen naar het schijnt gebruikelijke wijze, berust - volgens kritiek daarop van Prof. Dr C.J.F. BÖTTCHER ("De gerechtelijke schriftvergelijking", T.v. Strafrecht 59 (1950) 238-265), een kritiek, waar ik mij grotendeels bij aansluit - op onwetenschappelijke, subjectieve, methoden. Zonder hier diep op in te gaan, kunnen enkele punten, die min of meer van statistische aard zijn, genoemd worden.

Om na te gaan, of een betwist handschrift van een bepaalde persoon afkomstig is, vergelijkt men dit betwiste handschrift met een proeve van het handschrift van die bewuste persoon. Men zoekt dan naar punten van overeenstemming tussen de beide handschriften en indien deze "voldoende in aantal" zijn, verklaart men, dat het betwiste handschrift inderdaad van deze persoon afkomstig is. De basis van deze methode - het vergelijken van de handschriften - is natuurlijk op zichzelf gezond en voor de hand liggend, maar hij is zeer onvolledig. Om enkele van de het meest in het oog lopende gebreken van de gebruikelijke methode te noemen:

a. Men zoekt vaak ijveriger naar de punten van overeenkomst dan naar punten van verschil, welke laatste soms in het geheel niet vermeld worden.

b. Er is geen nauwkeurige omschrijving van wat men als punten van overeenkomst kan laten tellen.

c. Men vergelijkt het betwiste handschrift alleen met dat van die ene persoon, waarvan men vermoedt, dat het de schrijver is, en niet met dat van andere personen.

Deze gebreken in de methode leiden ertoe, dat men er geen idee van heeft, hoe groot de kans op een foute conclusie is. In eenvoudige gevallen, b.v. als men weet, dat geen van beide handschriften verdraaid zijn, kan deze kans, ondanks de gebrekkigheid van de methode, nog wel vrij klein zijn, maar het is niet ~~onaannemelijk~~ te onderstellen, dat hij in moeilijke gevallen heel groot is. Indien men zich nl., om maar één der vele mogelijkheden ter sprake te brengen, niet tot een bepaalde bekende groep van kenmerken van overeenkomst beperkt, maar er een onbekend aantal toelaat, heeft men geen idee over de waarschijnlijkheidsverdeling van het aantal punten van overeenkomst, dat tussen twee willekeurige (verschillende) handschriften uit een bepaalde groep van handschriften bestaat. Daar men bovendien het betwiste handschrift slechts met één of hoogstens enkele andere vergelijkt, heeft men er evenmin enig idee van, van hoeveel andere personen dan de onderzochte men óók nog zou kunnen concluderen, dat het betwiste stuk door hen geschreven is. Onenigheden tussen de deskundigen behoort dan ook niet tot de uitzonderingen.

Het is duidelijk, dat men in deze toestand, die inhoudt, dat de betrouwbaarheid van de gevolgde methode volkomen onbekend is, slechts verbetering kan brengen door een statistisch onderzoek en door de ontwikkeling van een objectievere en wetenschappelijkere methode dan de nu gebruikte. Een voorstel in deze richting is gedaan door Prof. BÖTTCHER in het boven geciteerde artikel. Hier kunnen wij niet op ingaan. Vermeld zij slechts, dat daarbij aan de bovengenoemde bezwaren tegemoetgekomen wordt door te werken met een niet te groot aantal zeldzaam voorkomende schriftkenmerken, die objectief waarneembaar zijn, en dat daardoor de bewijsvoering gedeeltelijk in de vorm van een statistische toetsing (met zeer kleine onbetrouwbaarheidsdrempel) gebracht kan worden. Typisch is nu, dat de tegenwerping van DEEN, enz., "Men mag de kansberekening zeker niet toepassen op een beperkt materiaal, want dan hebben kleine verschillen een veel te grote invloed op de uitkomst", een volledige miskennis van het principe van wetenschappelijk onderzoek inhoudt.

Immers de waarschijnlijkheidsrekening en de statistiek zijn **juist** bedoeld, om een duidelijk beeld te verkrijgen van de in het beschikbare materiaal aanwezige onzekerheden; het is van groot belang zich te realiseren, dat kleine verschillen bij weinig waarnemingsmateriaal van groot belang zijn, iets wat men op het oog niet zou vermoeden. Alleen door exact te werk te gaan, kan men te weten komen welke conclusies men nog wel en welke men niet mag trekken. Deze genoemde onzekerheden worden niet - zoals door de genoemde tegenwerping gesuggereerd wordt - door de waarschijnlijkheidsrekening in het materiaal naar binnen gebracht, maar zij worden er alleen door aangewezen en zij verdwijnen niet als men er de ogen voor sluit. Als men de exacte methoden vervangt door subjectieve en intuïtieve, betekent dit, dat men terwille van het bereiken van een conclusie de betrouwbaarheid van die conclusie opoffert.

4. Het voorbeeld der gerechtelijke schriftvergelijking is betrekkelijk uitvoerig besproken, omdat het een aantal voor de statistiek belangrijke aspecten vertoont, die ook bij andere toepassingen in analoge vorm optreden. Tevens is het een geval, waarbij het bereiken van verkeerde conclusies tot ernstige praktische gevolgen kan leiden, nl. tot gerechtelijke dwalingen. Tevens blijkt eruit van hoeveel belang het is, bij statistische problemen een zeer kritische denkwijze te gebruiken. Dit geldt niet alleen bij het opzetten van een proef of waarnemingsschema, maar ook bij de interpretatie der resultaten. Eén der meest bedenkelijke feiten in de praktijk van het wetenschappelijke onderzoek, voor zoverre dit op onzekere verschijnselen betrekking heeft, is b.v. het niet publiceren van de meeste negatief uitgevallen experimenten, maar wel van de positieve resultaten. Dit komt er, in zijn meest grove vorm, op neer, dat men een experiment net zo lang herhaalt, tot het slaagt en dan dit geslaagde experiment aan de openbaarheid prijsgeeft met verzwijging van de voorafgaande mislukkingen. Hoewel deze methode voor circusartisten als de juiste beschouwd moet worden - daar het experiment dan bij verdere herhalingen gewoonlijk weer lukt, omdat het op een bij het oefenen ontwikkelde behendigheid berust - moet het voor alle onderzoekingen, waarbij de kans op gelukken van het experiment niet met de tijd toeneemt, afgewezen worden.

Toch wordt deze fout b.v. in de praktijk van het zoeken naar nieuwe geneesmiddelen herhaaldeijk gemaakt, vandaar dat soms nieuwe middelen na enige tijd weer van de markt verdwijnen. Ook bij andere onderzoekingen, b.v. die van paranormale verschijnselen, is het een veel voorkomende fout. Vele "sterke verhalen", die de basis van allerlei bijgeloof vormen, komen op deze wijze te voorschijn uit de grauwe massa van dagelijkse gebeurtenissen, die iedereen snel vergeet. De oorzaak van het optreden van deze fout ligt in het verwarren van twee verschillende toepassingsmogelijkheden van de statistiek: 1. als ontdekkingsmiddel, 2. als bewijsmiddel. Indien men een onderzoek doet, waarbij niet van tevoren vaststaat, wat men precies wil trachten aan te tonen, mag men in de regel geen definitief resultaat verwachten, maar slechts aanwijzingen voor verder experimenteel onderzoek. Er kunnen nl. bij een willekeurig onderzoek zoveel bijzonderheden optreden (ook toevalligerwijze), dat men niet kan nagaan, hoe groot de kans is op het toevallig optreden van het een of andere schijnbaar systematische effect. Beschouwt men nu de waarnemingsresultaten met een nieuwsgierig oog en toetst men de bijzonderheden, die men erin ontdekt met Statistische methoden, dan kan met behulp van deze toetsen een indruk verkrijgen van de "graad van bijzonderheid" van de opgemerkte verschijnselen, maar alleen in extreme gevallen zal men de realiteit ervan bewezen kunnen achten. In de regel zal een herhaling van het onderzoek met een van tevoren nauwkeurig omschreven doelstelling nodig zijn, om na te gaan of deze bijzondere verschijnselen de eerste keer toevallig opgetreden zijn of niet. Kortom: wenst men de statistiek te gebruiken bij een bewijsvoering, dan dient men de waarnemings - of experimenteermethode in overeenstemming te brengen met een van te voren geformuleerde vraagstelling en een van tevoren gekozen wijze van statistische verwerking. Het plan voor het gehele bewijs, zowel voor het experimentele als het numerieke deel daarvan, dient van tevoren uitgestippeld en naderhand ook werkelijk uitgevoerd te worden, indien men wil bereiken, dat de gebruikelijke interpretatie van de statistische uitkomsten verantwoord is.

5. Kritisch denken is dus één der eerste vereisten voor een statistisch onderzoek. Dat het niet altijd overbodig is hierop te hameren moge blijken uit het volgende, wel wat extreme, voorbeeld.

In het "Rapport van de commissie tot onderzoek van het wichelroede - en aardstralenprobleem" van de Koninklijke Nederlandse Akademie voor Wetenschappen (Verslag van de gewone vergadering der Afdeling Natuurkunde 63 (1954) 52-85) wordt verteld, dat Prof. GLRLACH uit München proeven heeft verricht met wichelroedelopers, die beweerden van afschermkastjes te kunnen aangeven of zij af- of aanstonden. Bij de proeven gaven deze wichelroedelopers in ongeveer 50% van de gevallen het goede en in ongeveer 50% het foute antwoord. Welnu, dit beschouwden zij als een vrij goed resultaat, blijkbaar in de overtuiging, dat een gewoon mens wel altijd het verkeerde antwoord zou geven. Dit is een ietwat grof voorbeeld van een foutieve denkwijze en men behoeft geen statisticus te raadplegen, om zulke fouten te vermijden. Er zijn echter fijner kneepjes in het vak, waarvoor dit wel nuttig kan zijn, vooral waar het de opzet van een proef betreft. Men verlieze nl. niet uit het oog, dat de statistische analyse van waarnemingsresultaten bepaalde eisen stelt aan de wijze, waarop deze verzameld zijn en dat deze eisen in de regel aan de statisticus beter bekend zijn dan aan anderen. Voldoet men niet aan deze eisen, dan kan men ernstige afbreuk doen aan de doeltreffendheid van het onderzoek.

6. Dit brengt ons tot het laatste punt van deze voordracht: de organisatie van de statistiek in Nederland. Hierover zal ik kort en onvolledig zijn. De mogelijkheid van Statistische Consultatie bestaat voor alle T.N.O.-instituten in de vorm van de Afdeling Bewerking Waarnemingsuitkomsten, die desgewenst gesteund wordt door de Statistische Afdeling van het Mathematisch Centrum. Van deze twee instituten is het eerste uiteraard vooral ingericht op het toepassen van statistische methoden - al is ook het daar verrichte theoretische onderzoek verre van onbelangrijk - terwijl het tweede gericht is op theoretische studie en opleiding - al worden ook daar veel praktische opdrachten uitgevoerd. Tezamen kunnen deze instituten veel doen voor onderzoekers in andere wetenschappen en in de industrie. Eén van de beste vormen van samenwerking met instellingen, die vaak statistiek moeten toepassen, is volgens onze ervaring, dat de betrokken instelling één van zijn personeelsleden opleidt of op laat leiden tot statisticus in nauwe samenwerking met de genoemde instituten, terwijl deze samenwerking ook later wordt voortgezet.

Mogelijkheden voor deze opleiding bestaan b.v. bij het Mathematisch Centrum.

Dat in de industrie en bij verschillende laboratoria en T.N.O.-instituten reeds een vrij groot aantal statistici werkzaam zijn, waarvan sommige reeds sinds lang voor de oprichting van T.N.O.-A.B.W. en M.C., mag bekend ondersteld worden. Dat er verschillende adviesbureaux in Nederland bestaan, die statistische onderzoeken uitvoeren ten behoeve van de industrie is misschien minder bekend en wellicht voor de T.N.O.-instituten ook van minder belang. Wel van belang is ongetwijfeld, dat de belangstelling voor de statistiek in de industrie sterk groeiende is, zo sterk, dat aan de vraag naar statistici met een grondige opleiding bij lange na niet voldaan kan worden. Een gevolg van deze groeiende belangstelling is ook de oprichting van de Kwaliteitsdienst voor de Industrie, die door cursussen, coördinatie, het vormen van werkgroepen enz., de toepassing van o.a. statistische methoden in de industrie ter verhoging van de kwaliteit bevordert. Statistische werkgroepen op theoretisch gebied, colloquia en cursussen worden ook door anderen georganiseerd en de kristallisatie van een net van statistische posten in wetenschap en industrie en van regelmatig overleg tussen hen, die die posten bezetten, vindt zienderogen plaats. Daarbij mag de Vereniging voor Statistiek niet ongenoemd blijven. Evenals bij andere gelegenheden van deze aard besluit ik daarom met de aansporing aan allen, die statistische methoden gebruiken, om lid van deze vereniging te worden en om, indien zij niet zelf statistische specialisten zijn, te profiteren van de vrij uitzonderlijke mogelijkheden voor statistische consultatie, die in Nederland aanwezig zijn, door - vooral ook vóór zij aan een experiment beginnen - de raad van een vakman in te roepen.
