

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
A M S T E R D A M

STATISTISCHE AFDELING

Rapport S 275

VERTROUWELIJK

Enkele voorbeelden uit de praktijk van de
statistische consultatie

door

Ir A.R. Bloemena

BIBLIOTHEEK MATHEMATISCH CENTRUM
 ANSTERDAM

Juni 1960

In elk jaarverslag van het Mathematisch Centrum kan men in het gedeelte dat betrekking heeft op de Statistische Afdeling een lijst aantreffen van onderwerpen, welke onderwerp van bespreking zijn geweest en waarover adviezen zijn verstrekt. Nu behoort het geven van adviezen aan wetenschappelijke instellingen, individuele wetenschappelijke onderzoekers en aan het bedrijfsleven tot een der belangrijkste taken van het Mathematisch Centrum. De onderwerpen, waarover men de Statistische Afdeling om advies vraagt, zijn evenwel zodanig, dat men zich - terecht - kan afvragen op welk punt met wiskunde over deze onderwerpen iets te zeggen is. Uiteraard wordt deze vraag voor die gevallen, waarover rapporten van de afdeling verschijnen, in die rapporten impliciet beantwoord. Veel consultaties worden echter ook mondeling of met brieven afgehandeld, zodat geen rapporten beschikbaar zijn. Van beide soorten worden hier enkele gevallen besproken. Met opzet zijn daarbij zowel succesvolle als minder geslaagde gevallen uitgezocht.

Een deel van de problemen, waarover adviezen worden gegeven, vereist geen werkelijke toepassing van wiskundige berekeningen; het advies kan nl. soms alleen gebaseerd worden op het principe van het scheiden van hoofd- en bijzaken, een principe dat vaak door wiskundigen scherper gehanteerd kan worden dan door minder exact geschoolden. Als typisch voorbeeld hiervan is een probleem beschreven onder de titel: Het efficiënte gebruik van goederenwagons.

In het algemeen krijgt de Statistische Afdeling zelden problemen voorgelegd, die reeds statistisch zijn geformuleerd, dus in termen van een wiskundig model. Indien dit wel het geval is, is het in de regel afkomstig van een wiskundige in de industrie. Altijd wordt dan toch onzerzijds naar de achtergrond van het probleem gevraagd, teneinde eventueel op zinvolle manier vereenvoudigingen in het wiskundig model te kunnen aanbrengen.

Medici en biologen komen in grote aantallen naar het Mathematisch Centrum. Helaas vaak in een te laat stadium (vlak voor de promotie bijv., als de tekst van het proefschrift al bij de drukker is). Als voorbeeld van goede samenwerking met een medische kliniek is de gang van zaken beschreven van het probleem

dat in de jaarverslagen voorkomt onder de titel: "Het effect van vitamine K bij pasgeborenen". Een ander voorbeeld: "Resorptie van glucose", beschrijft het geval van een arts, waarmee het contact zeer moeilijk was.

Bij een aantal bedrijven beschikt men niet over ervaren statistici. Als men een probleem dan wel onderkent, en er ook literatuur over kan vinden, wordt de Statistische Afdeling gevraagd een "vertaling" te geven van die literatuur en uiteen te zetten hoe die op het onderhavige probleem kan worden toegepast. Een karakteristiek voorbeeld is beschreven onder de titel: "Het nemen van steekproeven van transportbanden".

Niet bij alle voorgelegde problemen is het mogelijk een oplossing of advies te geven. Als het probleem zeer moeilijk is, en indien er in de literatuur niets over bekend blijkt te zijn, dan kan een dergelijk probleem op de nominatie komen om intensief te worden bestudeerd. Resultaten van een dergelijke studie zijn echter nooit op korte termijn te verwachten. Ook komt een probleem soms niet tot een bevredigend slot door allerlei andere oorzaken, zie bijvoorbeeld het korte overzicht over "Ouderdomsbepaling van gesteenten door middel van radioactieve elementen".

Er zijn ook problemen waar de Statistische Afdeling herhaaldelijk mee wordt geconfronteerd, bijv. die betreffende loterijmachines. Ook andere puzzles met juridische aspecten komen wel ter sprake: zie bijv. het korte overzicht over "de verspreiding van reclamefolders in Nederland".

De gang van zaken bij een bespreking op de Statistische Afdeling is door Prof. Dr D. van Dantzig beschreven in "Wiskundige consultatie in de praktijk", Euclides, 30, 1954, 53, en in "Wiskundige consultatie ten behoeve van medisch, biologisch en ander onderzoek", Proc. Kon. Ned. Ak. v. Wet. 61, 1952, 13 (Ind. Math., 14, 1952, 13).

Het efficiënte gebruik van goederenwagons

De afdeling van de Nederlandse Spoorwegen, welke zich bezig houdt met besliskunde, maakte zich bezorgd over het betrekkelijk groot aantal uren per dag, dat een goederenwagen geen nuttige functie vervult. Deze wagons staan nl. lang stil op stations- en

fabrieksemlacementsen.

De N.S. heeft Nederland verdeeld in een aantal districten, waarvan ieder een centrum bezit. Het goederentransport kan nu in 3 fasen worden ingedeeld en wel als volgt:

- 1) Aan het eind van iedere middag en aan het begin van elke avond worden in ieder district in locale goederentreinen goederenwagons naar het desbetreffende centrum gebracht. Deze goederenwagons, geleegd of geladen, zijn afkomstig van de stations- of fabrieksemlacementsen, welke de locale trein passeert op weg naar het centrale punt.
- 2) Na middernacht vertrekken van ieder centrum uit goederentreinen naar centra in de andere districten (interlocale treinen).
- 3) In de ochtenduren worden vanuit ieder centrum in locale goederentreinen de goederenwagons op de plaats van bestemming gebracht. In deze treinstellen bevinden zich ook lege wagons, welke later (zie 1) geladen mee teruggenomen worden.

De locale treinen, genoemd in 1) en 3) zijn dezelfde treinen. Op de heenweg worden nl. de wagons afgeleverd en op de terugweg dezelfde of andere goederenwagons naar het centrum gebracht.

Gevraagd wordt de vertrektijdstippen van de locale treinen in de laatste fase zodanig te bepalen, dat

- 1) zoveel mogelijk wagons op dezelfde dag na gelost te zijn in het centrum terugkeren;
- 2) er meer gelegenheid geboden wordt tot lossen op de diverse emplacementsen.

Vertrekt men nl. te vroeg, dan zijn nog niet alle treinen uit de andere districten binnen. Een aantal wagons, behorende tot de locale treinen in de derde fase, is dan afkomstig van de corresponderende interlocale treinen van de vorige nacht.

Vertrekt men daarentegen te laat, dan zijn weliswaar alle interlocale treinen binnen, maar dan is er op een aantal stations niet voldoende gelegenheid tot lossen en kunnen niet alle wagons dezelfde dag leeg terugkeren.

Als studie-object werd gekozen het centrum Zwolle en het goederenvervoer langs het baanvak Zwolle-Leeuwarden. Op grond van vele niet nader aangegeven factoren leek het probleem op het eerste gezicht erg ingewikkeld.

Men was echter, daar het personenvervoer niet gestoord mag worden, zo beperkt in de keuze van mogelijke vertrektijdstippen, dat slechts één vertrektijdstip zinvol was. Dit laatste kon na enige bestudering van het probleem met eenvoudige hulpmiddelen worden aangetoond.

De functionarissen van de N.S., die zich met dit probleem bezighielden, hadden geen ervaring op het gebied van het samenstellen van dienstregelingen en waren door de vele bijkomstigheden overmand. Dit laatste verschijnsel komt men bij vele besliskundige opdrachten tegen.

Het effect van vitamine K bij pasgeborenen

Aan pasgeborenen, die ofwel te vroeg geboren zijn, dan wel om een andere reden een te gering gewicht hebben, wordt in de regel een injectie vitamine K toegediend. Door deze vitamine K bewerkt men dat de na de geboorte altijd verhoogde stollingstijd van het bloed van een dergelijk kind zo snel mogelijk weer op het normale niveau terugkomt. Dit is zeer belangrijk, opdat, indien het pasgeboren kind een bloeding krijgt, deze door slechte stollingseigenschappen van het bloed niet nog gevaarlijker wordt dan hij toch al is. De dosis vitamine K welke wordt toegediend is in de loop der jaren geleidelijk gestegen tot 10 mg. De laatste jaren is men echter tot de conclusie gekomen dat deze hoge dosis niet ongevaarlijk is. Men verdenkt het vitamine K ervan de bloedafbraak, die bij elk pasgeboren kind toch al optreedt (het wordt geel na de geboorte tengevolge van een lichte vergiftiging door afvalstoffen), te stimuleren, en wel in die mate, dat ernstige beschadiging van hersenkernen kan optreden door te grote hoeveelheid afvalstoffen, in het bijzonder bilirubine. Een kind, waarbij dergelijke beschadigingen optreden, overlijdt of wordt imbeciel. Begrijpelijk dat men direct zodra dit bekend werd de dosis vitamine K drastisch verlaagde. Bij een pasgeboren kind, waarvan het bilirubinegehalte in het bloed stijgt boven 20 mg ^o/oo, wordt,

teneinde het gevaar voor hersenbeschadigingen te verminderen, het bloed vervangen door bloed van een gezonde donor. In de verloskundige kliniek van het Wilhelminagasthuis heeft men dergelijke wisseltransfusies niet meer behoeven uit te voeren na de verlaging van de dosis van het vitamine K. In de kliniek van Prof. S. van Creveld werd nu door de chef de clinique, Dr J.I. de Bruijne, een onderzoek ingesteld naar de optimale dosis vitamine K, d.w.z. een dosis, waarbij het gunstige effect van vitamine K wel optreedt, maar nog geen verhoging van de bilirubineconcentratie van het bloed plaatsvindt. In het eerste onderzoek probeerde men het met 1 mg, maar bij deze kleine hoeveelheid ging het gunstige effect van vitamine K verloren. Met een dosis van 2 mg werd dit gunstige effect nog wel bereikt.

De wijze waarop dit onderzoek door de betreffende medici is verricht moge tot voorbeeld strekken voor de onderzoeken in de medische wetenschap.

Aan de Statistische Afdeling werd gevraagd de waarnemingen statistisch te bewerken. Hoewel de vragen alle met bekende statistische toetsen te beantwoorden zijn, is in verband met het grote belang van het onderzoek het gehele materiaal op een aantal verschillende manieren geanalyseerd, teneinde zoveel mogelijk inzicht in de waarnemingen te verkrijgen. De resultaten zijn in rapport S 241 en S 243 ter kennis gebracht van Prof. Van Creveld en Dr De Bruijne en toonden ruwweg gesproken aan dat (afgezien van heel bijzondere gevallen) met toediening van 2 mg mag worden volstaan. Het gunstige effect is dan zeer duidelijk aanwezig en het ongunstige wordt vermeden.

De resorptie van glucose

De medicus H. Muntendam had in een kliniek van het Wilhelminagasthuis te Amsterdam een onderzoek ingesteld naar de snelheid waarmee een kunstmatig hoog niveau van glucose in het bloed (teweeggebracht door een injectie van glucose) daalt tot het normale niveau. Na de injectie werd om de 10 minuten bloed afgenomen en daarin het glucosegehalte bepaald. Het is uit de medische literatuur bekend dat indien men de glucoseconcentratie op logaritmische schaal uitzet tegen de tijd, waarop men deze

concentratie constateert, de waarnemingen bij redelijke benadering op een rechte lijn liggen.

Ook hier bleek dit bij de waarnemingen, die door de heer Muntendam verkregen waren, zeer goed te kloppen. De eerste vraag was nu welke lijn zo goed mogelijk aansloot bij de punten. Een vraag die direct te beantwoorden is met de methodiek van de regressie-analyse.

Op deze wijze krijgt men voor elke patiënt een helling van de lijn. Is deze helling afhankelijk van de leeftijd? Is voor een gegeven leeftijd deze helling verschillend bij "gewone" patiënten (d.w.z. patiënten, die een ziekte hebben welke deze z.g. glucose-belastingsproef niet kan beïnvloeden) en bij carcinoompatiënten? Deze vragen zijn direct te beantwoorden door toepassing van bekende statistische toetsen.

Daar de heer Muntendam met mondelinge adviezen alleen niet voldoende geholpen kon worden, zijn de berekeningen alle uitgevoerd door assistenten van de Statistische Afdeling. Door één van hen werden de resultaten kort beschreven in een rapport, dat als appendix aan het proefschrift "De intraveneuze glucosebelastingsproef" werd toegevoegd.

Het nemen van steekproeven van transportbanden

Door de heer R.B. Harteveld, econoom van de Koninklijke Nederlandse Zoutindustrie, werd de vraag voorgelegd hoe men een partij zout, die in een schip wordt geladen, tijdens het laden bemonsteren moet en welke conclusies men uit de bemonsteringsresultaten kan trekken. In de praktijk is er maar één zinvolle manier waarop men monsters kan nemen, nl. van de transportband, die het zout van de bunker naar het schip vervoert. Moet men nu monsters met regelmatige tussenintervallen nemen en zo ja, wat is de "beste" afstand tussen twee monsters.

Over deze kwestie bracht de heer Harteveld zelf een artikel mee, dat in een Duits statistisch tijdschrift (Metrika) verschenen is, maar dat hij, als econoom, niet kon lezen.

De kwestie is ook nog op andere plaatsen in de literatuur behandeld. In het algemeen kan men zeggen dat een systematische bemonstering nauwkeuriger is dan een bemonstering met z.g. aselechte

tussenintervallen, doch anderzijds een uitvoerigere analyse vergt.

De resultaten uit de literatuur werden in een voor de heer Harteveld leesbaar résumé medegedeeld.

Bij dit soort vragen bewijst het omvangrijke kaartstelsel van statistische literatuur, dat in de loop der jaren is opgebouwd en steeds wordt bijgehouden, zeer goede diensten.

Ouderdomsbepaling van gesteenten door middel van radioactieve elementen

Door Mejuffrouw Ir A.C.W.C. Bot van het Laboratorium voor Metallurgie en Docimasie te Delft werd, mede namens Dr Schürmann te 's Gravenhage, een vraag voorgelegd over de nauwkeurigheid van ouderdomsbepalingen van gesteenten. In Zwitserland had de fysicus H. Stauffer deze bepalingen uitgevoerd en in het verslag erover een berekening van de nauwkeurigheid gegeven. Zo was een van de resultaten $(720 \pm 90) \cdot 10^6$ jaar. De berekening van de grenzen ± 90 zag er vrij vreemd uit en ons werd een oordeel gevraagd of deze berekening verantwoord was.

Waar het verslag, dat de resultaten begeleidde, zeer duister was, werd Dr Stauffer per brief een groot aantal vragen gesteld over de technische procedure en over de numerieke berekeningen. Stauffer antwoordde eerst niet; na een kleine aansporing verschaftte hij enkele inlichtingen met verder de mededeling: "it would be best, if you once come to Berne, ...". Waar ook Dr Schürmann niet meer op een brief van ons reageerde, is de zaak terzijde gelegd.

Een van de theoretische problemen waarmee men bij dit onderzoek te maken heeft, is echter van algemeen belang en staat op de nominatie om nader te worden uitgewerkt.

De verspreiding van reclamefolders in Nederland

De uitgeverij Reidel te Dordrecht wil de verkoop van een van de door hen uitgegeven boeken stimuleren. De uitgeverij is niet erkend (de bona fide boekhandel kan dus niet ingeschakeld worden); het aangeprezen boek van weinig verheffend kaliber. Het is voor het voortbestaan van de uitgeverij zeer belangrijk dat er een redelijk aantal (circa 3000) verkocht wordt. Men besluit een reclamefolder huis aan huis te laten bezorgen, niet door de P.T.T.,

maar door middel van een reclamebureau Web te Haarlem, dat voor dit soort verspreiding de vereiste organisatie bezit. De helft van de kosten wordt à contant voldaan, de rest dient binnen twee maanden na de verspreiding te worden betaald. Web garandeert dat minstens 98% van alle gezinnen de folder zullen ontvangen.

De folders worden verspreid; reactie hierop minimaal (150 bestellingen). Reidel gaat zich dan afvragen of de verspreiding goed is geschied. Hij stuurt, circa zes weken na de dag waarop de folders zijn bezorgd, een vertegenwoordiger er op uit om op een groot aantal plaatsen in den lande bij een aantal huizen aan te bellen en te vragen of men het bewuste reclamepapier heeft gezien. Ruwweg beweert 75% van de ondervraagden niets van die aard te hebben ontvangen, circa 10% zegt het inderdaad gekregen te hebben, de rest twijfelt. Onmiddellijk wordt het bureau Web medegedeeld dat zij een wanprestatie hebben geleverd en dat er niet verder betaald zal worden. Web deelt mede dat de klacht ongegrond is en laat een gerechtelijke vordering naar het resterende bedrag instellen. De advocaat van Reidel, de heer Bijleveld te Dordrecht, dient namens zijn cliënt een tegenvordering in ter grootte van het reeds betaalde bedrag.

De heer Bijleveld stelt zich telefonisch in verbinding met het Mathematisch Centrum en vraagt of men kan nagaan of de gezinnen welke door de vertegenwoordiger van Reidel werden bezocht, een representatieve steekproef zijn. Een van de mensen van de uitgeverij zal dan op het M.C. komen en daar alle gegevens betreffende de steekproef meebrengen.

Aldus geschiedt. De heer Pels komt op het M.C. en brengt de lijsten mee waarop staat welke adressen zijn bezocht en hoe de reactie was. Hij vertelt dat in twee plaatsen van ieder circa 50.000^x) inwoners van de circa 100 ondervraagden iedereen ontkende de folder te hebben gekregen. Toen is door de uitgeverij zelf in deze plaats huis aan huis de folder bezorgd. Reactie: 7 bestellingen. De vraag die men in dit verband aan ons stelde is nu deze. Als de kans dat ieder gezin, dat een folder ontvangt, een bestelling doet, gelijk is in heel Nederland, kan men dan uit

x) Waar de werkelijke getallen ons wel zijn medegedeeld, maar niet zijn genoteerd, is hier een aantal fictieve getallen ingevoerd, die de orde van grootte aangeven.

deze gegevens concluderen dat de verspreiding van de folders door het bureau Web niet goed is geweest.

Een ruwe berekening, stante pede uitgevoerd, leert nu dat men dit niet kan concluderen. Op 1.000.000 folders 150 reacties zou bij correcte verspreiding overeenkomen met een kans op reageren per gezin van $\frac{150}{1.000.000}$. Bij de 100.000 inwoners, ≈ 30.000 gezinnen, mag men dan verwachten $\frac{30.000}{1.000.000} \cdot 150 = 4,5$ reacties. Nu is het wel redelijk te onderstellen dat het aantal reacties van de 100.000 mensen bij benadering een Poisson-verdeling heeft met parameter 4,5. De kans op 7 of meer reacties is dan 0,17, hetgeen zo hoog is dat er op grond van deze cijfers geen reden is te twijfelen aan de premisse van correcte verspreiding, waarvan bij deze berekening is uitgegaan.

De resultaten van de steekproef, welke door de vertegenwoordiger van de uitgeverij zijn onderzocht, zijn met de heer Pels uitvoerig besproken. De wijze waarop de steekproef genomen is kan de toets der kritiek niet doorstaan: als men bij een gezin hoorde dat men de folder niet had gekregen, belde men bijv. ook aan bij de burelen. Dit zal, gezien de wijze waarop de 98%-clausule door Web wordt gebruikt, zeker een onzuiverheid in de antwoorden hebben veroorzaakt.

Onzerzijds werd toen opgemerkt dat de resultaten van de steekproef weinig verschilden van plaats tot plaats. Het aantal "nee"zeggers was opmerkelijk constant. Zelfs al is de verspreiding zeer slecht geweest dan zal dit toch niet in alle plaatsen even slecht gebeurd zijn. Er zullen dus zeker steden zijn geweest, waarbij de verspreiding goed is geweest. Op dit punt verder redenerend kwamen wij tot de vraag wat "nee" zeggen betekende. Betekent dat "weet het niet, maar ik zeg maar nee" of is dit een pertinent "nee". Men moet verder wel degelijk rekening houden met de zes weken tussenruimte tussen verspreiding en enquête. Daarom werd geadviseerd door een neutrale instantie (Stichting voor Statistiek) in een stad (bijv. Alkmaar) een folder van het gebruikte soort te laten verspreiden en na zes weken een goed opgezette enquête te houden. Vindt men dan weer 70% neezeggers dan is het duidelijk wat er gebeurd is. Vindt men véél minder neezeggers (dus 30-40%) dan zou men voor de rechter zijn eisen kunnen verde-

digen.

Het advies is niet opgevolgd, daar de Stichting voor Statistiek niet overtuigd was met een bona-fide zaak te maken te hebben.

De rekening van het M.C., zijnde f 50.--, werd eerst na herhaalde aanmaningen betaald.