



WETENSKAPLIKE BYDRAES VAN DIE PU VIR CHO
Reeks H: Inougcurele Redes Nr. 23

**PROGRAMMEERDERS OF INTERPRE-
TEERDERS? – DIE ROL VAN HEDEN-
DAAGSE STATISTICI**

J.C. Geertsema

**Potchefstroomse Universiteit vir CHO
1976**

PROGRAMMEERDERS OF INTERPRETEERDERS? – DIE ROL VAN HEDENDAAGSE STATISTICI

Daar is tans opnuut 'n besef van die noodsaaklikheid dat die wetenskap en die gewone man nader aan mekaar gebring moet word. Dit word deur die Amerikaanse denker en fisikus, Warren Weaver, soos volg gestel¹⁾: „Many scientists – and the public which they have over and falsely impressed – have created a horrid and dangerous gap between science and the rest of life ... What we must do – scientists and non-scientists alike – is close the gap. We must bring science back ... to take its place as a friendly and understanding companion to all the rest of life”.

In 'n onlangse nommer van *Science*²⁾, orgaan van die *American Association for the Advancement of Science*, word daar huis in die redaksionele artikel oor hierdie saak gehandel onder die opskrif *Understanding of science*. In dié artikel word 'n plan van aksie aangekondig om begrip van die wetenskap by die Amerikaanse publiek te bevorder. Die organiseerders van die plan beoog benutting van die massakommunikasiemiddels, veral televisie, 'n nuwe tydskrif, 'n lesingprogram en navorsing oor die oorsake en oorkoming van wanopvattings oor die wetenskap by die breë publiek.

Die artikel meld veral die volgende sake waarvan die publiek begrip behoort te kry:

Eerstens, 'n begrip van die wetenskap self, sy struktuur en sy werkswyse: „To understand science, one must acquire a taste, or at least a tolerance, for the beauty of mathematical models, of the structure of galaxies or crystals, of DNA. One must appreciate the process itself – data that fall into place after years of analysis, finding the missing element that completes a table, predicting an event on the basis of careful calculations and a flash of insight – as an end in itself”.

Tweedens moet die publiek insig kry in die nut van die wetenskap: „One must understand that the scientist, while not a rainmaker who can deliver solutions to specific problems on short order, nevertheless addresses himself to relevant problems of the day – that, in ways difficult to specifically foresee and almost impossible to control, scientific efforts do “pay off” and are not just idle games. The public must learn that scientific findings are always tentative and may prove erroneous or obsolescent, but that their tentative guide is more valid and safe than any other approach to the world”.

Die artikel sluit af deur die omvang en moeilike aard van die gestelde taak te beklemtoon en sê dan: „The future of science in America may well depend on this effort”. 'n Mens kan nie help om raak te sien dat hierdie laaste sin die hele motivering vir die bekendstellingspoging verraai nie, naamlik om onder-

Die statistiese wetenskap is seker een van die wat die meeste onder misverstande gely het. Die talle gesegdes en grappies ten koste van statistiek is berug, maar wat ontstellend is, is dat selfs media wat ingelig behoort te wees hulle aan growwe wanvoorstellings skuldig maak. 'n Voorbeeld hiervan is 'n artikel in die Amerikaanse nuustydskrif *Time* onder die opskrif *The science and snares of statistics*.⁶⁾ In 'n kritiese bespreking van die gebruik van statistiek word die hele statistiese wetenskap belaglik voorgestel deur middel van 'n aantal voorbeelde van skynbaar statistiese uitsprake wat gemaak is deur nie-statistici. So word verwys na 'n skatting van die skade tydens die berugte Detroit-onluste wat onder druk van koerantmanne deur die brandweerroof op \$500-miljoen gestel is. Later het dit geblyk aansienlik minder te wees. Terwyl so 'n verwysing seker nie te ernstig hoef opgeneem te word nie, aangesien dit met wetenskaplike statistiese werk weinig te doen het, is daar ander bewerings in die artikel van 'n ernstiger aard. Daar word verwys na die bekende Kinsey-verslae oor die geslagslewe van Amerikaners en beweer dat verreikende gevolgtrekkings gemaak is op grond van 'n steekproef wat veels te min mense ingesluit het, byvoorbeeld in een van die verslae oor mans is daar 5 300 mans ondervra wat dan té min sou wees. In verdere opmerkings word die hele proses van steekproefneming soos volg verdag gemaak: „One cherished statistical tool is the sample. Not even statisticians can agree on how big or good a sample can be relied upon as representing the whole”. Tereg wys 'n briefskrywer, Edwin Cox, in 'n volgende nommer van die tydskrif⁷⁾ daarop dat die Kinsey-verslae nie gekritiseer was vanweë die klein steekproef nie, maar omdat die steekproef nie getrek was uit die groep wat hulle moes verteenwoordig nie. Dit was naamlik 'n steekproef uit die Amerikaanse mans wat gewillig was om hulle intieme sake met navorsers te bespreek. Die geslagslewe van mens wat nie bereid is om dit met navorsers te bespreek nie, is heel moontlik heelwat anders en moes ook in ag geneem gewees het.

In ander opmerkings word die volgende gesê: „The computer has enshrined statistics”, says M.I.T.’s Professor Harold A. Freeman. „Without it, statistics would still be a grubby („smerige”) business. Where once all they had to do was count, and perhaps draw graphs, statisticians are now ‘programmers’, with a mystique all their own”. Hier het ons by die kern van die rede gekom — is statistici programmeerders? Die antwoord hierop is 'n besliste nee. Ons het hier te doen met een van die mees algemene misverstande oor statistiek — tans geklee in moderne gewaad. Die siening dat statistiek bloot die invloer van massas syfers in 'n rekenaar is, is 'n totaal skeue siening. Die opmerkings van briefskrywer Cox is weer reg in die kol: „Professionally trained statisticians are much more than number collectors and graph drawers. Sir Ronald Fisher’s (die eerste statistikus wat tot ridder geslaan is) work proves statistics was not a ‘grubby business’ before computers, and computer age statisticians like Tukey, Savage and Blackwell are far more than ‘programmers’”.

Dit word dus duidelik dat statistiek 'n besondere taak het om hulle weteen-
skap duidelik te omskryf en van die versameling, verwerkings en interpretasie van data
hou saamhang, wil ek tog die klem laat val op die interpretasie van data. As ons uit 'n oomblik terugkeer na die titel van hierdie rede, dan will ek statistici karakteriseer as interpreters. Waar die werkende siening van statistiek kan die volgende wees: Statistiek is die wetenskap van die versameling van data,
van die verwerkings van data en van die interpretasie van data.⁸

Houewel die drie aspekte van versameling, verwerkings en interpretasie van data benaminge interpreters hulle in huisserig. 'n Interpretere redeer is iemand wat die betekenis verduidelik, wat verstaan, wat vertolk. Die interpretere redeer is 'n tolk in die ware sin van die woord: hy moet as tussenmangier dien tussen data en die man wat wil weet wat die data sê. Die interpretere redeer is die man wat die data verstaan en die boodskap daarvan aan die belangstellende verdui-
delik. So gesien is dit duidelek dat die statistikuus veel meer is as 'n program-
merder. In dieselfde sameenval ek darem se dat die rekenaar beskou moet dat statistiek maar in rekenaarwerk kan opgaaan, terwyl die behoorlike teore-
tiese fundamente agterwee bly.

Daar moet net gewek word teen die opvatting, ook van sommige statistici, dat statistiek maar in rekenaarwerk kan opgaaan, terwyl die behoorlike teore-
tiese fundamente agterwee bly. Vercringing van wiskundeonderwyseers het die vereniging pas 'n bundel van 44 opstelle onder die titel *Statistics: a guide to the unknown*⁹ laat verskyn.

Die taak om statistiek bekend te stel geniet tans die besondere aandag van die Amerikaanse Statistiese Vereniging. In samewerking met 'n vereniging van wiskundeaardeleersers het die vereniging pas 'n bundel van 44 opstelle onder die titel *Statistics: a guide to the unknown*⁹ laat verskyn.

Die bundel is opgesette deur ekende deskundiges en beskryf in populaire taal moderne toepassings van statistiek. So is daar artikelsoor die rol van statistiek in die toetsing van statistiek. So is daar artikelsoor die moontlike gebruik van statistiek in die toetsing van statistiek. Daar was verskillende moontlike metodes om lik te toets vir effektiwiteit. Daar was verskillende moontlike metodes om likaansye navorsers Saak met 'n entstoel na voerkomlike interpretasieprobleem, aangesien sy-
fers getoon het dat die voorkeurs van polo soe egter nie tot 'n onoorkomlike interpretasieprobleem, aangesien sy-
voorkoms van polo soe in die vorige jaar nie. Hierdie wese van versameling van data sou egter nie tot 'n onoorkomlike interpretasieprobleem, aangesien sy-

(1) Versprei die entstoel so wyd moontlik en stel vas of dit nie lei tot 'n lager eksperiment te beplan:

Laat ons een van hierdie voorbeede in besondere beskou. In die vyftriger-
jaar het polo-epidemies baie ernstige armetings begin aanneem. Toe die Amerikaanse navorsers Saak met 'n entstoel na voerkomlike interpretasieprobleem, aangesien sy-

(2) Dien die entstoel toe aan kinders in een landstreep en vergelyk dan die liker was as in ander jarre om geen klarbalklike rede nie.

voorkoms van polio in dié landstreek met dié in die res van die land. Maar 'n soortgelyke probleem met die interpretasie van die data sou hier ondervind word – die voorkoms van polio verskil in ieder geval van landstreek tot landstreek.

(3) 'n Benadering wat baie sterk voorgestaan is, was om alle kinders in graad twee oor die hele land wat sou inwillig tot die proefneming, te ent en dan die voorkoms van polio onder hulle te vergelyk met die voorkoms onder die kinders in graad 1 en graad 3. Dié benadering het egter baie kritiek uitgelok, waaronder die volgende: By die diagnostering van 'n moontlike poliogeval onder die graad 2 kinders, sou 'n dokter navraag doen of die kind geënt is of nie. Dié kennis sou hom kon beïnvloed in geval van 'n moeilike diagnose. Daar was nog ernstiger besware. Die vrywillige kinders uit graad 2 vorm 'n spesiale groep – daar is later vasgestel dat hulle uit meer welgestelde en beter ontwikkelde huise kom. Daarom sou hulle as groep nie vergelykbaar wees met die groot massa graad 1- en graad 3-kinders nie. Die ander beswaar was dat dit byna onmoontlik sou wees om te beslis of 'n waargenoem verskil in voorkoms van polio by geënte en nie-geënte kinders die gevolg van die entstof is, of van toevallige variasie. Ondanks hierdie besware is daar tog voortgegaan met proefnemings op hierdie basis, maar in 'n groot deel van Amerika het plaaslike gesondheidswerkhede 'n sterk standpunt ingeneem en besluit om 'n vierde, statisties korrekte, benadering te volg.

(4) Die ondersoek is beperk tot graad 2-leerlinge en by elk wat ingewillig het om deel te neem is deur middel van 'n kansmeganisme besluit of die entstof of 'n kontrolemiddel aan hom toegedien word. Dit is so gedoen dat elke leerling 'n gelyke kans gehad het om die entstof of die kontrolemiddel te kry. Dit kon byvoorbeeld gedoen word deur by elke leerling 'n muntstuk op te skiet en die entstof toe te dien as kruis bo val. Daar is egter meer gesofistiekerde metodes om dit te doen. Die kontrolemiddel was 'n soutoplossing wat geen effek kon hê nie en wat presies soos die entstof daar uitgesien het. Die toediening is so gereël dat nòg die toediener nòg die ontvanger geweet het of die entstof of die kontrolemiddel toegedien word. Hiermee is die besware teen die ander metodes oorwin: by die diagnose van 'n moontlike poliogeval sou die dokter nie beïnvloed kon word deur kennis daarvan of die kind geënt is of nie. Verder is die leerlinge deur middel van die kansmeganisme op 'n objektiewe wyse in twee groepe verdeel. Dié metode van verdeling lewer gewoonlik twee redelik eenderse groepe, wat dan direk met mekaar vergelykbaar is. Baie belangrik is dat die resultate nou vatbaar is vir wiskundige ontleding gebaseer op die waarskynlikheidsteorie – die data kan nou wetenskaplik geïnterpreteer word.

Hierdie laaste metode is toe grootliks toegepas. Daar was omtrent 400 000 vrywilligers. Deur middel van 'n kansmeganisme is hulle verdeel in twee groepe wat toe uit ongeveer 200 000 elk bestaan het. Per honderduisend was daar 28 poliogevalle onder proefleerlinge en 71 onder die kontroleleerlinge. Wat

se hervide data? Dit kan byvoorbeelde beweer dat die enstot geen effek sou word dat dit bloot tot toevalle was dat uit die leerringe wat in elk gevall silek het nie en dat dit kontroleerde dat die kansmessigheid sou word die merderheid daarvan die eksperiment kontroleert sou word dat die kansmessigheid van die eksperiment kontroleert. Maar die statistikus kan die kansmessigheid by die kontrole van die enigevalle bereken omdat die eksperiment kontroleert van so 'n toelopgroep ingedeel is. Hierdie gevall word die verskill tussen 28 poliogevalle onder die proefleerlinge en 71 onder die kontroleerlinge statisties gevawuurter as baie modellik toe te skryf aan toevallike variasie. Die enigste alternatief is dat die enstot of effek dat die versameling en interpretasie van data delikate sake is wat strenge weten-tieel moet wese. So is die effektiwiteit van die Salik-enstot bervesigt. Ons sien dat die enigevalle gevall word dat kontroleerde statistiek van statistiese metodes te evaluer in die verhoedrome probleme, asook om bestanddele metodes te evaluer in die verhoedrome sterke kunde, dien die wiskunde as hulpmiddel in die statistiese navorsing.

In Area wat trans besondere onderwerp van navorsing in ons eie departement wat verband hou met die Engelse word *sequence*, diu aan: Dic sekvensie word dat sekvensieel. Sekvensieel word dat sekvensie bestuder dan statistiese prosedures wat daardeer gewoonweg almal tesame nie. Die voordel daarvan om die warnehings een vii een te verkry, is dat die statistikus dan nie voorbaai hoeft te besluit hoewel warnehings gesdean gaan word nie – hy kan soos die warne- hemings een-een verkry word ophou met warnehings doen wanmeer geneoeg informasie verkry is. Dikwels is dit tydswend of baie duur om 'n warne-

sokk 'n uutpyjl oopgeskiet moet word om 'n enkelle warnehing te verkry – 'n mak as wat nodig is – niks meer nie. Gestel byvoortbeel dat in 'n ander-

enklele warnehing is dan baie duur. 'n Sekvensiële metode lyk hier aanget- wes. Tydens die Tweede Wêreldoorlog is die eerste belangrike tegnologie in hierdie tegnologie met groot sukses in die geallieerde wapenbedryf toegespas en Tans is daar weer 'n nuwe oplewing in die navorsing in die rigting. Een van die vaders van die moderne statistiek, J. Neyman, bestempel 'n onlangse is dit as hoogsgehemme militêrigting besku.

resultaat van 'n navorser op hierdie gebied, H.E. Robbins, as „'n merkwaardige prestasie"¹¹⁾. Die rigting waarvan hier sprake is, wat insluit die teorie van betroubaarheidsrye, hou myns insiens groot moontlikhede vir die toekoms in. Ek vertrou dat u my sal toelaat om hierdie ontwikkeling kortlik te beskrywe. 'n Betroubaarheidsry is 'n ry van intervalle, sodanig dat die waarskynlikheid dat al die intervalle 'n onbekende parameter gelyktydig bevat, voorgeskryf kan word. In die praktyk word die intervalle van die ry een na die ander bereken en vir elke nuwe interval moet 'n addisionele waarneming gemaak word. Die voordeel hiervan is dat die proses op enige willekeurige punt gestaak mag word en nogtans bevat die laaste interval die parameter met voorgeskrewe waarskynlikheid. Hierdie punt waar die proses gestaak word, mag afhang van die data en ook van ander inligting. So stel 'n betroubaarheidsry ons in staat om, as dit blyk dat die waarnemings in 'n gegewe steekproef te min is vir behoorlike interpretasie, bloot voort te gaan en addisionele waarnemings te doen totdat dit blyk dat daar genoeg is, of totdat daar nie meer fondse beskikbaar is om verder te gaan nie. 'n Betroubaarheidsry is 'n kragtige hulpmiddel en besonder plooibaar. Die uitdaging is om *goeie* betroubaarheidsrye te konstrueer vir toepassing in verskillende situasies. 'n Begin is reeds hiermee gemaak, maar daar bly nog baie werk oor. Analoog aan 'n betroubaarheidsry, definieer ons 'n seleksiyer as 'n ry van versamelings van populasies wat gelyktydig die *beste* populasie met voorgeskrewe waarskynlikheid bevat. Sodoende kry ons 'n nuwe benadering tot die seleksieprobleem en die moontlike toepassings is baie.

Waar vroeër in hierdie rede gewys is op die belangrikheid daarvan om die wetenskap bekend te stel, geld hierdie belangrikheid in die besonder ook die statistiese wetenskap. Die insigte van die statistiek is van groot belang by alle wetenskaplike eksperimentering wat onvermydelik tot die interpretasie van data moet lei. Maar ook vir die breë publiek word dit al belangriker om 'n juiste beskouing van syfers te hê. Een of ander tyd kry elkeen met statistiese tegnieke te doen, of moet hy 'n standpunt inneem oor uitsprake wat op grond van statistiese metodes gemaak is. In sommige lande word daar byvoorbeeld al stellings soos die volgende oor die radio uitgesaai: „Die waarskynlikheid van reën mōre is 80 persent”.

Maar ook wat die toepassing van statistiek betref is daar baie moontlikhede as die potensiële benutter net weet waarom dit gaan. So het die voordele van statistiese steekproefneming blykbaar eers in die laaste paar jaar tot die rekenmeesters in ons land deurgedring en word dit nou met groot entoesiasme gebruik. Veral in die handel en die nywerheid is die moontlikhede legio.

Seker die blywendste bekendstellingsmetode van 'n wetenskap soos statistiek is deur middel van die onderwys. In etlike oorsese lande word statistiek reeds op skool aangebied, gewoonlik as deel van die wiskundeleerplan. Die statistici daar help ywerig om sodanige onderwys te bevorder en die probleme wat

Aan ons universiteit is daar 'n groeiende bewuswording van die noodsaaklikheid van minstens 'n elementêre kurssus in statistiek vir mense rigting. 'n Vraag wat in die verband belangrik word is die organisaasie van statistiek-kursses vir vryklike studielinge. Een moonlikheid is dat in elke departement van groep departemente iemand besas word met die dosering van statistiek vanaf die eerste tot die laaste studieling. So 'n stelsel is nodig dat dit leei tot isolering van die vryklike statistiek van makar by die universiteit. Die onderriging kontrak is uiters belangrik vir weder sydsie inspitasie en kennis. 'n Andere moonlikheid is dat die departement statistiek self alle kurssuse in vermenigvuldiging van statistiek-kurssuse wat dikwels dupliseer inghou.

'n Moonlikheid wat ek ook persoonlik wil ondersceu, is dat statistici wat in die verksillende toepassingsgriltes spesialiseer verbonde sal wees aan die departement. So kan biometriëtert virzelfdryd ook aan die betoekte ander deel van verbonde wees aan die departement Statistiek en Bodemkunde. Signometrie en die departemente Dierkunde, Plantkunde en Bodemkunde, Signometrie en die departementen Statistiek en Fisikologie was dan die departement Statistiek en Economics. So sal die statistici dan die universiteit deel uitmaak van die samtekunst hê, namarlik die departement Statistiek. Dit behoort ook te leei tot onderriging sameverkinig tussen die meer teoretiese statistici en die meer toegepaste statistici, tot almal se voordeel.

Die elemente van statistiek (en in ander disiplines) staan soms be-keend as dienskurusse met 'n effens onaangename assosiasie. Elk meen doestende in statistiek behoort kennis te neem van die opmerking van die voorraanstande Amerikaanse statistikus, John Tukey [3]: „Are we interested in getting the world to think at least a little like a statistician some of the time? And if so, then this ought to be one of our main businesses, not a service course“.

Om nou alles saam te vat: 'n Statistikus beoefen 'n wetenskap met 'n tie-

karakter — sy hooftaak is interpretasie van gegewens. Dit is 'n wanvoorstelling om 'n statistikus te sien as bloot 'n wiskundige, of bloot 'n programmeerdeur, of bloot 'n tabuleerdeur. Hierdie vermoë om te interpreteer is van groot belang vir die wetenskaplike sowel as die nie-wetenskaplike wêreld en behoort dus wyd bekend te wees. Daarom moet veral die onderrig van statistiek besondere aandag geniet.

VERWYSINGS

1. WEAVER, W. 1961. The imperfections of science. *American scientist*, 49:99-113.
2. SCIENCE, 177(4047) 4 Aug. 1972.
3. STOKER, H.G. 1968. Beginsels en metodes in die wetenskap. 2e dr. Kaapstad, Boekhandel de Jongh. p. 142-145.
4. STOKER, H.G. 1970. Oorsprong en rigting. Kaapstad. Tafelberg. Il: p. 70-72.
5. SPREUKE 3:13-15.
6. TIME, 8 September 1967, p. 31.
7. TIME, 22 September 1967, p. 7.
8. KEMPTHORNE, O. 1968. The future of departments of statistics. (*In Conference on the future of statistics, University of Wisconsin, 1967. The future of statistics; proceedings. New York, Academic Press.* p. 104).
9. TAUNER, J.M. red. et al. 1972. Statistics; a guide to the unknown. San Francisco, Holden-Day.
10. KENDALL, M.G. 1942. On the future of statistics. *Journal of the Royal Statistical Society*, 105:69-91.
11. NEYMAN, J. 1971. Foundations of behavioristic statistics. Foundations of statistical inference. New York, Holt, Rinehart and Winston. p. 4.
12. INTERNATIONAL CONFERENCE OF THE TEACHING OF PROBABILITY AND STATISTICS, Corbondale, Ill., 1969. 1970. The teaching of probability and statistics; proceedings of the first CSMP. Ed. by L. Rode. New York, Wiley Interscience.
13. CONFERENCE ON THE FUTURE OF STATISTICS, University of Wisconsin, 1967. 1968. The future of statistics; proceedings. New York, Academic Press.