

Statistiek in huidig vwo wiskunde A-onderwijs (2)

Heeft de kritische keizer kleren aan?

B. van Putten

Vakgroep Wiskunde, Landbouwniversiteit Wageningen

Inleiding

In het eerste deel van dit artikel [1] werd een begin gemaakt met een beschouwing over de behandeling van het onderwerp Statistiek in enkele veelgebruikte schoolboeken voor vwo wiskunde A. Met name werd gekeken naar het aspect van de verwerving van 'kritisch inzicht', dat in wiskunde A een centrale rol speelt.

In dit tweede (laatste) deel vervolgen we de bespreking van misvattingen, ook hier weer aan de hand van voorbeelden. De voorbeelden die thans worden besproken, lijken in het algemeen een meer aan de wetenschappelijke onderzoekspraktijk ontleende achtergrond te hebben dan de acht voorbeelden die in deel 1 het licht zagen.

Voorbeeld 9

De triathlon is een uitputtingsslag, die bestaat uit 3,8 km zwemmen, 180 km fietsen en het lopen van een marathon (42,2 km).

A beweert dat slechts 20% van de deelnemers aan een triathlon de eindstreep haalt. B meent dat het er meer zijn.

Om tot een beslissing te komen, nemen ze een triathlon met vijftig deelnemers als steekproef.

>> Wordt de bewering van A verworpen als dertien deelnemers de eindstreep halen met $\alpha = 0,05$?

Commentaar

Wanneer het meningsverschil van A en B betrekking zou hebben op een en dezelfde triathlon, zeg de triathlon die morgen gehouden staat te worden te P, dan zou B gelijk krijgen wanneer elf of meer deelnemers over de eindstreep gaan; in dat geval finisht immers meer dan 20%, de inzet van het meningsverschil. Gebruik maken van statistische toetsingstheorie zou dan zelfs geheel overbodig zijn.

Vanaf nu gaan we er echter vanuit dat A en B in algemene termen over 'triathlons' praten en dus geen specifieke wedstrijd op het oog hebben. In dat geval fungeert het cijfer 20% overigens weer als *spookgetal* (zie voorbeeld 6). Heeft die 20% betrekking op alle triathlons die waar

ook ter wereld en op welke verre toekomstdatum dan ook gehouden worden, of heeft men het eigen land met het specifieke klimaat voor ogen en daarbij wellicht nog een beperkte tijdshorizon, of slaat deze 20% op wellicht nog weer wat anders?

Goede nota dient er verder van te worden genomen dat in de toetsingsprocedure die hier wordt gesuggereerd, persoon A in de asymmetrische beslissingsprocedure het voordeel van de twijfel krijgt, omdat zijn/haar bewering tot nulhypothese wordt verheven. Stemt met name B daarmee ten volle in? De klip die de hele voorgestelde toetsenrij echter alsnog schipbreuk doet leiden is de volgende. Terwijl A en B over de een of andere (super-)populatie een meningsverschil hebben, doen ze slechts aan één triathlon een waarneming. De uitkomst hiervan kan over het 'geheel' van triathlons (wat dat dan ook moge zijn) een vertekend beeld geven, denk aan een triathlon in storm en hagel, of een triathlon analogon van de Elfstedentocht van 1963.

Een stap in de goede richting ware dus geweest het nemen van een aselechte steekproef uit een welgedefinieerde populatie van triathlons.

Voorbeeld 10

Enige kenners zitten gezellig bij elkaar en gaandeweg ontstaat er een discussie over de smaak van cola. Er zijn er die beweren dat PB beter smaakt dan B&N, anderen beweren het tegendeel. De kantinebeheerder stelt voor om de proef op de som te nemen.

Elk krijgt twee glazen voorgezet: één met PB en één met B&N. Geblinddoekt proeft iedereen en geeft zijn voorkeur. Uiteraard kent men de volgorde waarin men moet proeven niet! Er blijkt dat negen van de elf personen B&N prefereren boven PB.

>> Wat concludeer je?

Oplossing

Tekentoets. We toetsen tweezijdig de hypothese 'geen verschil'.

Overschrijdingskans kan bij 9 'plussen' ($n = 11, p = \frac{1}{2}$)

$$P(X \geq 9) = 0,0327 > 2,5\%$$

dus bij 5% significantieniveau nog geen significant verschil.

Mede gezien de geringe steekproeflengte dus geen overtuigend verschil.

Commentaar

Het probleem heeft iets weg van het onderwerp 'Statistische rangschikkings- en selectietheorie', dat ingaat op problemen van het selecteren van 'de beste behandeling'. Dit thema komt terecht steeds meer in de statistische schijnwerpers te staan. Hier echter laten proefopzet en technische uitwerking nogal wat te wensen over. Zo is de populatie waarover kennelijk een uitspraak wordt verlangd volslagen onduidelijk; men krijgt de indruk dat hier op de een of andere wijze populatie en steekproef samenvallen. Maar zelfs wanneer men alsnog de mededeling zou ontvangen dat deze groep van klaarblijkelijke *kenners* beschouwd mag worden als zijnde het resultaat van een aselechte trekking uit bijvoorbeeld de algehele populatie van cola-experts, treden ernstige problemen op. Als voorbeeld voor deze problemen volgt hieronder de beschrijving van een mogelijk verloop van het experiment.

Nadere informatie leert dat er in de gezellige discussie voorafgaande aan het experiment maar twee personen waren die PB het lekkerst vonden, de negen anderen prefereerden B&N. De kantinebeheerder, die al geen beste beurt maakte toen hij na afloop van het experiment laconiek meedeelde dat hij de zojuist geschonken cola niet wenste te beschouwen als een rondje van de zaak, werd vervolgens met kracht van redelijke (statistische) argumenten geprest om *per persoon* mee te delen welke cola als lekkerste werd aangewezen. Geheel in overeenstemming met het gevoerde predikaat 'kenner' bleek iedereen zijn al eerder opgegeven favoriete cola te hebben aangewezen. Daarop ontbrandde in volle hevigheid de discussie opnieuw; al spoedig kon deze niet meer als 'gezellig' worden bestempeld, al voorkwam het inhoudende alcoholpercentage vervelender taferelen. Een scenario dat hier diametraal tegenover staat is ook mogelijk: maar liefst *alle* kenners wezen 'de verkeerde cola' als beste aan. Terwijl het aantal identiteitscrises met elf toenam, droop iedereen verslagen af. Het experiment liet een vervelende bijmaak na.

Voorbeeld 11

Iemand wil nagaan, welke van de twee supermarkten in zijn omgeving de goedkoopste is. Daartoe kiest hij willekeurig tachtig artikelen en gaat na in welke van de twee bedrijven deze artikelen het goedkoopst zijn. Het blijkt dat vijftig van de gekozen artikelen in supermarkt A goedkoper zijn en dertig in B.

>>Is er bij een significantieniveau van 5% reden om aan te nemen dat A goedkoper is dan B?

Commentaar

Noodzakelijkerwijze hebben de beide supermarkten exact hetzelfde assortiment, heel kras!

Wanneer al toepassing van wiskundige statistiek vereist was, zou een 'eindigheidscorrectie' in verband met steekproeftrekking *zonder teruglegging* op z'n plaats geweest zijn. Er lijkt, gelet op het doel van het onderzoek, geen reden om A en B asymmetrisch te behandelen; toch gebeurt in de voorgestelde beslissingsprocedure dit laatste wel degelijk. Verder wordt de alternatieve hypothese 'gestuurd' in de richting dat A goedkoper is dan B. Is dit soms een produkt van (in supermarkten onbekende) *data snooping*? De *grootten* van de prijsverschillen per artikel spelen kennelijk geen rol, een merkwaardigheid.

Tot daadwerkelijke toetsing zal het echter allemaal niet komen. Wat is immers goedkoper? Even afgezien van het gegeven dat het antwoord op de vraag welke supermarkt goedkoper is *persoonsgebonden* zal zijn – immers: zoveel hoofden, zoveel zinnen – zal het natuurlijk van groot belang zijn te weten hoe het gesteld is met de prijzen van de *courante* artikelen. Het hoeft op grond van verschillen in *bedrijfsstrategische keuze* zeker geen verbazing te wekken wanneer de dertig artikelen die bij B goedkoper bleken, precies alle *courant* zijn, en de vijftig artikelen die bij A goedkoper bleken, *incourant*. Het doorsnee publiek zal, onbekend met de zegeningen der statistiek, in dit geval B bestempelen als de goedkoopste van de twee. Diegene die, door bittere armoede gedreven, overweegt het omschreven experiment uit te voeren, zou erop gewezen kunnen worden dat het in uiteindelijke besluitvorming betrekken van luxe-artikelen, als hier geschiedt, een vertekening kan geven wanneer men slechts wenst te letten op primaire levensbehoeften. In uiterste noodgevallen zou men trouwens kunnen overwegen per artikel daar te kopen waar de prijs het laagst is. Een *revival* dus voor het aloude zogeheten boerenkoopmanschap.

Voorbeeld 12

Bij een metaalallieerbedrijf bevat het eindprodukt 30% tin indien het productieproces goed verloopt. In monsters vindt men dan bij analyse dat het percentage tin normaal verdeeld is met gemiddelde 30,00 en standaarddeviatie 1,5. Er wordt regelmatig gecontroleerd. Men neemt aan dat bij storingen de standaarddeviatie gelijk blijft en alleen het gemiddelde verandert.

Op een dag vindt men bij een aselechte steekproef van acht monsters de volgende percentages tin:

29,2 29,2 29,4 29,5 29,5 29,6 29,8 29,9

>>Zijn deze data voldoende reden om te twifelen aan het goede verloop van het productieproces?

Oplossing

Tweezijdig toetsen van het steekproefgemiddelde \bar{X}

met $E(\bar{X}) = 30$ en $\sigma(\bar{X}) \approx 0,53$.

Overschrijdingskans circa 18% groter dan elk redelijk significantieniveau. Dus onvoldoende reden tot twijfel.

Commentaar

Hoe zou het toch komen dat je, terwijl je onbevangen de verkregen data op je in laat werken, toch steeds, ondanks alle statistiek, hoe langer hoe meer de indruk krijgt dat *wel degelijk* aan het goede verloop van het productieproces moet worden getwijfeld omdat maar liefst alle acht monsters minder dan 30% tin bevatten? Dat komt omdat zelfs nog veel meer moet worden getwijfeld aan de juistheid van de opgegeven standaarddeviatie ter waarde van 1,5. Enige statistische berekening, die hier echter achterwege wordt gelaten, toont aan dat het optreden van een steekproef die een zo grote discrepantie tussen geschatte en werkelijke standaarddeviatie vertoont, bezijden alle realiteit staat.

(De oplettende lezer constateert dat de auteur zich hier nu zelf aan *data snooping* schuldig maakt. Als verweer brengt de auteur hier tegenin dat een zodanig onwaarschijnlijke gebeurtenis in de praktijk niet voor *zal* komen en in voorbeelden met fictieve data niet voor *mag* komen).

Niet onmogelijk is dat dit voorbeeld aldus een afname van mogelijk bij de leerling aanwezig gevoel voor *common sense* teweegbrengt.

Voorbeeld 13

Twee merken autolampen schelen veel in prijs. A is het duurste, maar als we de reclame mogen geloven heeft A ook de meeste branduren. Een consumentenorganisatie wil dat onderzoeken en neemt een proef met honderd automobilisten, die eerst merk A en vervolgens merk B gebruiken. In negenenzeventig gevallen blijkt de lamp van A meer branduren te hebben dan die van B. In de overige eenentwintig gevallen wint merk B het.

>> *Geeft dit resultaat aanleiding te veronderstellen dat merk A een langere levensduur heeft? Neem een significantieniveau van 5%.*

Commentaar

De consumentenorganisatie zou er goed aan doen om, afgezien van eventueel te trekken statistische conclusies, ook de prijzen van A en B in de uiteindelijke aanbevelingen te betrekken. Het zou best kunnen dat op basis van correct uitgevoerd statistisch onderzoek blijkt dat A een langere levensduur heeft. Eén van de sterke punten van de statistische wetenschap is dat zelfs minime verschillen als 'gemiddeld brandt A zeveneneenhalve seconde langer dan B', bij voldoende steekproefomvang onherroepelijk worden gedetecteerd. In situaties met een zodanig minimaal verschil in levensduur zal de doorsnee automobilist kiezen voor de goedkoopste van de twee, dus voor B.

Zou overigens 'langere gemiddelde levensduur' altijd

voor ieder auto-onderdeel te prefereren zijn? Bedenk wat u liever in uw nieuwe auto zou willen hebben: *of* een remsysteem dat het – u kunt uw horloge erop gelijk zetten – tien jaar uithoudt, *of* een variant die met gelijke kansen er ofwel na zes weken de brui aan geeft, ofwel twintig jaar ononderbroken diensten bewijst? Het laatste type heeft wel de langste gemiddelde levensduur!

De volgende proefopzet met statistische *blokvorming* zou hier veel meer voor de hand gelegen hebben: wijs op basis van loting een lampje van merk A toe aan de ene kant van de auto en een lampje van merk B aan de andere kant; pas daarna een statistische analyse toe overeenkomstig deze proefopzet.

Buiten dit alles: de in dit vraagstuk gekozen proefopzet is van dien aard dat het trekken van statistische verantwoorde conclusies een hachelijke onderneming is. Immers, *eerst* wordt A gebruikt, *daarna* B. Deze werkwijze kan allerlei systematische fouten introduceren; om niet meer te noemen zou het bijvoorbeeld zo kunnen uitpakken dat de proefnemingen, te beginnen met A, in de lente een aanvang nemen. Tegen de winter was zo ongeveer iedereen aan het B-lampje toe. Dit laatste lampje liet zijn licht al spoedig niet meer schijnen omdat net die winter de pekel weer eens overvloedig moest worden gestrooid en vochtinwerking zijn tol eiste. De consumentenorganisatie maakte vervolgens bekend dat op grond van uitgebreid onderzoek tot de conclusie moest worden gekomen dat het met de kwaliteit van B beroerd is gesteld.

Voorbeeld 14

Bij een Aa × Aa-kruising is de kans p dat een nakomeling van het type aa is, gelijk aan $\frac{1}{4}$. Bij andere kruisingen, bijvoorbeeld Aa × aa of Aa × AA, heeft p een andere waarde. In een bepaalde situatie was het type van de ouders niet duidelijk. Er was alleen het vermoeden dat beiden van het type Aa zouden zijn.

>> *Welke twee hypothesen kan men hier tegenover elkaar zetten? Wat is de toetsingsgrootte X? Hoe is X onder H_0 verdeeld? Waarom heb je hier te maken met een tweezijdige toets?*

Van de zesendertig nakomelingen bleken er vijftien van het type aa te zijn. Hoe groot is de hierbij behorende overschrijdingskans?

Commentaar

Zonder verlies aan algemeenheid zou men het probleem enig vlees en bloed kunnen geven door te stellen dat het een muizen-kruising betreft waarbij AA en Aa zich manifesteren door een grijze kleur vacht, terwijl aa een witte vacht vertoont. (Het probleem waarin AA, Aa en aa alle fenotypisch onderscheidbaar zijn, is zo eenvoudig oplosbaar dat we er hier geen uitwerking aan geven). De tekst stelt dat (in eerste instantie) het genotype van de

ouders niet duidelijk was. Een beetje onderzoeker zou in het geval van een witte kleur niet al teveel moeite met raden hebben gehad. Dus waren beide ouders van grijs fenotype en beschikten zodoende elk over tenminste één A-gen.

In plaats van maar liefst vijftien witte nakomelingen zou slechts één wit exemplaar al voldoende zijn geweest om te kunnen concluderen dat beide ouders elk over tenminste één a-gen beschikten.

Toen de onderzoeker de voorwaarden 'tenminste één A-gen' en 'tenminste één a-gen' combineerde, bleek de aanvankelijke onzekerheid over de genetische achtergrond (zelfs zonder inschakeling van het statistische toetsingsapparaat) te zijn opgeheven: beide waren inderdaad van genotype Aa.

Een lastiger probleem zou zijn opgetreden wanneer geen der jonge muizen wit was geweest. In dat geval had de statistische wetenschap te hulp geroepen kunnen worden. Deze zou in dat geval echter nog steeds niet door middel van de techniek van het statistisch toetsen van hypothesen te werk zijn gegaan. Geschetst vraagstuk hoort namelijk thuis in de klasse van *statistische toewijzingsproblemen*, in voorbeeld 1 reeds aangestipt.

Voorbeeld 15

Een placebo is een namaakmedicijn dat objectief gezien geen werking kan hebben en de goede werking van zo'n medicijn is het placebo-effect. Een pas aan een ziekenhuis benoemde psycholoog beweert tijdens een stafvergadering dat bij 40% van de lijdens aan hevige hoofdpijn deze klachten van psychosomatische aard zijn. Zijn collegae klinici wensen deze bewering te controleren en stellen met zorg pillen samen uit tarwemeel en water, die zij aan een aantal patiënten geven met het verzoek dit nieuwe geneesmiddel te gebruiken en te becommentariëren. De resultaten zijn als volgt:

Acht patiënten vonden het nieuwe middel stukken beter dan de andere pijnstillers.

Acht patiënten vonden het nieuwe middel even goed als hun vorige pijnstiller.

Eén patiënt zei dat het nieuwe middel hem soms pas na twee uur ging helpen, terwijl andere pijnstillers direct werkten.

Drieëndertig patiënten vonden het nieuwe middel volstrekt waardeloos.

>> *Toets de bewering van de psycholoog met behulp van deze gegevens.*

Oplossing

Stel het onbekende gedeelte hoofdpijnlidens met psychosomatische klachten is p .

Dan beweert de psycholoog: $p = 0,40$.

Stel X is het aantal lijdens met psychosomatische klachten in de steekproef van 50, dan stellen we als uitkomst vast: $X = 16$ (die ene patiënt tellen we maar niet mee, na

twee uur kan de hoofdpijn ook vanzelf overgegaan zijn). De overschrijdingskans is $P(X \leq 16 | p = 0,40) = 0,1561$. De klinici zullen vermoedelijk eenzijdig willen toetsen omdat zij van mening zijn dat $p < 0,40$, maar ook dan is dit resultaat niet significant op 10% niveau.

Commentaar

De zorg waarmee de pillen worden vervaardigd staat in schril contrast met de zorg die aan de opzet van het experiment is besteed. Zijn de patiënten uit de een of andere populatie (aselect) getrokken? Zouden de sceptische klinici, die het onderzoek uitvoeren, hun vragen niet subtiel kunnen sturen in de richting van het antwoord 'waardeloos', zodat de mogelijk juiste bewering van de psycholoog wordt weggehoond en de tijdelijke aanstelling van deze bekwame man geen vervolg krijgt?

Bij de aanvang van het experiment werd het net lente; de hoofdpijn verdween als sneeuw voor de zon en de nieuwe tabletten kregen een gunstig onthaal. Of de winter was net in aantocht en...

'Acht patiënten vonden het nieuwe middel even goed als hun vorige pijnstiller.' Ook waardeloos?

'Drieëndertig patiënten vonden het nieuwe middel volstrekt waardeloos.' De twee uitersten die hierbij als conclusie zouden kunnen worden getrokken, zijn:

- Het betreft hier de echte psychosomaten. Niets helpt ze, en dat is hun goed recht, want daar zijn ze immers psychosomaat voor.
- Het betreft hier integendeel de 'echte' hoofdpijnlidens. Ze merken dat ze gefopt zijn, hetgeen ook al weer niet bevorderlijk werkt op het genezingsproces.

Inmiddels moge duidelijk zijn dat het geheel enige ruimte laat inzake het interpreteren van de onderzoeksgegevens.

Voorbeeld 16

Bij een groot bedrijf met vestigingen in het hele land werken duizenden mensen. Er is nogal wat ziekteverzuim: gemiddeld 8,3 dagen per werknemer per jaar (met een standaardafwijking van 3,7 dagen).

Dit jaar heeft de directie besloten iedere werknemer vijf extra snipperdagen te geven. Aan het eind van dit jaar zal nagegaan worden of dit tot een lager ziekteverzuim heeft geleid. Daartoe zal een aselechte steekproef van honderd personen uit het totale personeelsbestand genomen worden.

>> *T is het totaal aantal dagen ziekteverzuim van deze honderd personen. Bereken $E(T)$, $Var(T)$ en $SD(T)$ onder de aanname dat het ziekteverzuim niet veranderd is.*

Neem nu als toetsingsgrootte het gemiddeld aantal dagen ziekteverzuim X van de honderd personen in de steekproef.

>> Formuleer de hierbij behorende H_0 - en H_1 -hypothesen.

Bereken, aannemend dat H_0 juist is, $E(X)$, $\text{Var}(X)$ en $SD(X)$. (Neem aan dat de SD niet veranderd is).

Na een jaar blijkt het totaal aantal dagen ziekteverzuim van die honderd personeelsleden 742 dagen te zijn, gemiddeld 7,42 dagen per werknemer. Een vermindering met bijna een dag per werknemer.

>> Wat is de bij $X = 7,42$ behorende z -waarde (aannemend dat H_0 juist is en dat de SD niet veranderd is)? Hoe groot is de hierbij behorende overschrijdingskans?

Welke hypothese wordt geaccepteerd ($\alpha = 0,05$)?

Commentaar

Zonder nog nader op het verhaal in te gaan, wordt geconstateerd dat een 'dimensiefout' is gemaakt: het gemiddelde wordt gemeten in dagen per werknemer per jaar, terwijl de bijbehorende standaardafwijking genoteerd is in dagen zonder meer. Hoe zijn dit gemiddelde en deze standaardafwijking tot stand gekomen? Zo te zien is er dus gemiddeld over (vele) jaren: jaren waarin griep epidemieën voorkwamen en jaren waarin het weer er prima toe bijdroeg dat veel werknemers zich gezond voelden. Deze zogeheten *tussen-jaren variatie* zal in de grootte van de standaardafwijking tot uitdrukking komen. Ook de *tussen-werknemers variatie* zal in deze standaardafwijking zijn verwerkt. Inzicht in de grootten van de verschillende componenten binnen de uiteindelijk gegeven standaardafwijking van het ziekteverzuim wordt niet nader gegeven.

De standaardafwijking die echter in het vervolg van het vraagstuk zou moeten worden gebruikt 'klikt' in geen deele met de hierboven besproken standaardafwijking; immers de steekproef laat geen ruimte voor een *tussen-jaren effect*. Al met al wordt op deze wijze getracht om onvergelijkbare zaken met elkaar in verband te brengen. Het vraagstuk is daarom statistisch niet aan te vatten. Daarenboven wordt nog gesuggereerd dat men met statistische technieken een *causaal* verband zou kunnen aantonen, omdat men wenst na te gaan of het geven van extra snipperdagen tot een lager ziekteverzuim heeft geleid. Helaas is het aantonen van causale verbanden de statistiek niet gegeven. In een goed opgezet onderzoek zou verder nog geworsteld zijn met de kwestie dat ziekteverzuim op extra gegeven snipperdagen wellicht niet mogelijk is; waar zou dit gegeven in de statistische analyse een plaats moeten krijgen?

Het zou weleens kunnen zijn dat, gelet op de wens van een dalend ziekteverzuim, de kosten van het cadeautje van vijf snipperdagen meer geweest zijn dan de baten.

Voorbeeld 17

In een medisch tijdschrift werd bij een artikel over hoge bloeddruk het volgende tabelletje afgedrukt dat de rela-

tie tussen therapie-trouw (doen wat de dokter zegt) en therapie-succes (verlaging van de te hoge bloeddruk) weergeeft:

		therapie-succes		
		ja	nee	
therapie-trouw	hoog	23%	34%	57%
	laag	12%	31%	43%
		35%	65%	100%

De vermelde getallen zijn percentages van de onderzochte groep patiënten. (...)

>> Stel eens dat er in totaal veertien patiënten aan het onderzoek deelnamen. Zet de percentages om in aantallen en (...)

Commentaar

In de gegevens schuilt een onjuistheid die van onvergelijkbaar minder gewicht is in vergelijking met eerder gesignaleerde onvolkomenheden, maar die duidelijk maakt dat de kritische aandacht nimmer mag verslapen. Bij veertien patiënten kunnen slechts percentages in de tabel voorkomen die op hele cijfers afgeronde uitkomsten zijn van $\frac{100i}{14}$ met i lopend van 0 tot en met 14. De eerste waarde van de onderzoekspopulatie die past bij de in de tabel gegeven percentages is gelijk aan 65; men zou verder kunnen inzien dat 96 de grootste populatieomvang is die zich met de opgegeven percentages niet verdraagt.

Voorbeeld 18

In de kantine van een grote fabriek staan vierkante tafeltjes met vier stoelen er omheen. Een psycholoog observeert het gedrag van de mensen die zich daar in de middagpauze ophouden. In het bijzonder kijkt hij naar de tafeltjes waaraan twee mensen zitten. Die twee kunnen tegenover elkaar zitten of naast elkaar aan een hoek van het tafeltje. In de loop van enkele dagen ziet hij 87 keer twee mensen aan een tafeltje: 26 keer tegenover elkaar en 61 keer naast elkaar. De psycholoog concludeert hieruit dat er een uitgesproken voorkeur is voor de hoekopstelling, mogelijk ook vanwege angst voor het zogenaamde oogcontact dat niet te vermijden is als twee mensen tegenover elkaar zitten.

Statistisch verdedigt hij zijn bevinding als volgt:

Wanneer er geen voorkeur zou zijn, zal met gelijke kansen de ene en de andere opstelling gekozen worden.

Laat p de kans zijn op de hoekpositie en neem als toetsingsgrootte X het aantal keren dat de hoekopstelling wordt gekozen.

>> Formuleer de opvatting van de psycholoog (voordat hij met zijn observaties begint) als de H_1 -hypothese en de daarbij behorende hypothese H_0 .

Hoe is onder H_0 de stochast X verdeeld?

Bereken de overschrijdingskans van het gevonden resultaat $X = 61$.

Wat is de conclusie bij $\alpha = 0,10$?

Waarom is hier sprake van een tweezijdige toets?

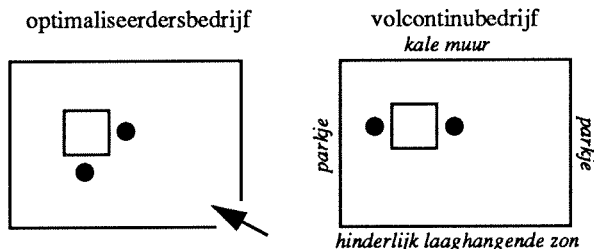
Ben je het eigenlijk wel eens met $p = 0,5$ als je ervan uitgaat dat de twee mensen zonder een bepaalde voorkeur aan zo'n tafeltje gaan zitten?

Commentaar

De psycholoog observeert gedurende enkele dagen; het is niet onwaarschijnlijk dat hij daarbij herhaalde malen een zelfde tweetal personen gadeslaat die in de loop der jaren een vast 'zitpatroon' hebben aangenomen. Het werken met een binomiaal model moet dientengevolge worden ontraden. Verder is het zeer wel denkbaar dat de kansverdeling van het zitgedrag voor de zijkanten en vooral voor de hoeken van de kantine anders is dan voor het midden ervan, zulks ook alweer, ditmaal in verband met de eis van een constante succeskans, ten detrimente van een binomiaal model.

De psycholoog kijkt slechts in deze bedrijfskantine. Extrapolatie naar alle mogelijke bedrijfskantines is een hachelijke onderneming; extrapolatie zelfs naar alle situaties waarbij mensen weleens in oogcontactuele problemen zouden kunnen geraken, ligt natuurlijk geheel buiten bereik.

De psycholoog dient terdege te beseffen dat er vele platvloerse redenen kunnen zijn waarom mensen zo gaan zitten zoals ze gaan zitten. Onderstaande figuren geven schematisch de indeling weer van twee bedrijfskantines.



De linker bedrijfskantine is die van een optimaliseringsbedrijf waar de ingang van de kantine met een pijltje is aangegeven. Men luncht steeds in tweetallen en men loopt nooit één stap meer dan nodig is. Eén tafeltje is slechts getekend (de andere tafeltjes worden op dezelfde manier bezet); de plaats van de aanzittenden is met een rondje aangegeven.

De rechter bedrijfskantine is die van een volcontinubedrijf. De werknemers komen, ook hier in tweetallen, aan hun broodnodige verpozing door gedurende de maaltijd te genieten van het uitzicht op fraai aangelegde parkjes. Ook voor dit geval is slechts één tafeltje met aanzittenden getekend, aan de andere tafeltjes is men op analoge wijze aangeschoven. Men neemt het daarbij voor lief dat het hoofd van de collega onvermijdelijk in beeld komt. Zelfs het toch wel door de werknemers als hinderlijk ervaren oogcontact wordt met het oog op de schoonheid der parkjes op de koop toegenomen. Het is niet te wensen dat de psycholoog ooit nog eens op grond van zijn

onderzoekingen in deze bedrijfskantine tot de conclusie komt dat oogcontact in het maatschappelijk verkeer als zeer gewenst wordt ervaren.

Conclusie

De lezer zal hebben ingezien dat alleen al op het punt van kritisch inzicht waar en wanneer statistiek niet en soms ook wel kan worden toegepast, de huidige schoolboeken op het gebied van vwo wiskunde A vatbaar zijn voor enige verbetering. Met spijt moet verder geconstateerd worden dat nu en dan toch wel een karikuraal beeld van de werkwijze van de statistiek wordt gegeven. Bovendien wordt de leerling regelmatig geconfronteerd met beschrijvingen van wat voor degelijk wetenschappelijk onderzoek zou moeten doorgaan maar in feite op geen enkele wijze de toets der kritiek kan doorstaan; in voorbereidend wetenschappelijk onderwijs lijkt een dergelijke werkwijze niet op z'n plaats.

Over de vraag hoe de verbeteringen moeten worden gerealiseerd, zal nog diepgaand moeten worden nagedacht. Naar het gevoelen van de auteur zou het een aardig leerdoel van een vak als vwo wiskunde A zijn in voorkomende gevallen kritiek te leveren op conclusies die op vermeend statistische gronden zijn verkregen. Het is de vraag of realisering van een afgeronde cursus bij het huidige examenprogramma zonder problemen mogelijk is. In de algemene statistische gereedschapenkist lijken nog voldoende aardige mogelijkheden aanwezig om leerlingen op passende wijze met de statistiek in kennis te brengen. Hopelijk is als bijproduct van de hier geleverde kritiek ook enig inzicht in en begrip voor de wiskundige statistiek ontstaan. In deze wetenschap probeert men experimenten zodanig op te zetten en te analyseren dat op praktische vragen een antwoord kan worden geformuleerd; de conclusies worden uit verzameld cijfermateriaal getrokken, dwars door allerlei ongecontroleerde storende invloeden heen.

Literatuur

- [1] Deel 1 van dit artikel is verschenen in Nieuwe Wiskrant jrg 9 nr 4, juli 1990, pp. 31 e.v.
- [2] Bos, D.J.P. e.a.: *Wiskunde lijn, wiskunde A*, Jacob Dijkstra, Groningen, z.j.
- [3] Broek, L. van den e.a.: *De Wageningse Methode, deel 56, Hypothese toetsing*, Stichting De Wageningse Methode, 1988.
- [4] Dijkhuis, J.H. e.a.: *Getal en ruimte. 5/6 V ~ A1 en ~ A2*, Educaboek BV, Culemborg, 1985.
- [5] Doove, F.M.W. e.a.: *Sigma 5A, 6A*, Wolters-Noordhoff B.V., Groningen, 1986.
- [6] Groen, W.E. e.a.: *Moderne wiskunde bovenbouw vwo 5A en 6A*, Wolters-Noordhoff BV, Groningen, 1987.
- [7] Putten, B. van: *Enige kanttekeningen bij de vwo-examens Wiskunde A in 1987*, Rhenen, februari 1988.