Statistiek in huidig vwo wiskunde A-onderwijs (1)

Heeft de kritische keizer kleren aan? [1]

B. van Putten [2]
Vakgroep Wiskunde, Landbouwuniversiteit Wageningen

Inleiding

Een van de doelstellingen van vwo wiskunde A is het aankweken van een kritische instelling terzake de toepassing van wiskunde. Van huidig onderwijs in het onderdeel Statistiek wordt aan de hand van voorbeelden uit gebruikelijke schoolboeken nagegaan of dit tot aanwakkering dan wel tot uitdoving van kritisch inzicht leidt. In dit eerste deel wordt na een algemene inleiding een begin gemaakt met de presentatie van de voorbeelden, waarna een korte tussentijdse evaluatie wordt gegeven.

Statistiek en wiskunde A: een ideaal huwelijk (?)

Het is verheugend te kunnen constateren dat in vwo wiskunde A voor het onderwerp Statistiek een royale plaats is ingeruimd. Er is immers alle kans op dat de leerling in vervolgonderwijs nog verder met de (wiskundige) statistiek in aanraking komt omdat deze tak van wetenschap een cruciale rol speelt in empirisch onderzoek. En mocht de leerling in verder onderwijs de statistiek kunnen ontlopen, zulks kan beslist niet het geval zijn in het leven van alledag waar door moderne media een constante stroom statistiek over onze hoofden wordt uitgestort.
Ruwweg gesproken kunnen in deze stroom twee componenten worden onderscheiden, te weten de *presentatie* van gegevens door middel van de zogeheten beschrijvende statistiek en de *concludering* op basis van de wiskundige statistiek. Voor wat betreft de eerstgenoemde component kan worden opgemerkt dat de doorgewinterde manipulator de beschrijvende statistiek met de bijbehorende schijn van onbevooroordeeldheid kan misbruiken om de argeloze toeschouwer gevolgtrekkingen te laten maken die de presentator welgevallig zijn. De tweede component, die van de concludering op basis van de wiskundige statistiek, is nog ondoorzichtiger dan de eerste. In goede handen is de toegepaste wiskundige statistiek een onmisbaar werktuig in wetenschappelijke bewijsvoering; in verkeerde handen, zowel die van de goedwillende amateur als die van de zojuist beschreven manipulator, kan onder de paraplu van exactheid giftige waar schuilgaan die doorgaans voor zoete koek wordt geslikt. In deze sfeer lijkt het van enig belang te vermelden dat schrikbarend veel van wat als 'statistisch bewezen' de wereld wordt ingestuurd, in feite is gebaseerd op onderzoek dat de statistische toets der kritiek niet kan doorstaan.

Op grond van het voorgaande moge duidelijk zijn dat een kennismaking met statistiek, in het kader van wat men soms wel eens plechtstatig 'de toerusting op de maatschappij' noemt, in middelbaar onderwijs niet mag ontbreken. In het Voorbereidend Wetenschappelijk Onderwijs lijkt het, gelet op de rol die de statistiek in de wetenschap speelt, een voor de hand liggende gedachte dat deze kennismaking niet per se vluchtig behoeft te zijn. Daarbij komt dat de statistiek bij uitstek een niet aflatende kritische instelling van haar beoefenaren verlangt. Al met al lijkt de statistiek dus een ideaal onderwerp om in een vak met de doelstellingen van wiskunde A te onderwijzen.

Een kritische opleiding kritisch beschouwd

Het lijkt een zinvolle gedachte om het onderwerp Statistiek binnen vwo wiskunde A, zoals dat in schoolboeken en examens inmiddels een plaats heeft gekregen, vanuit een positieve grondhouding aan een kritisch onderzoek te onderwerpen.
Boven alle hieronder nog te geven kritiek is verheven de waardering die opgebracht wordt voor het duidelijk in alle facetten naar voren komende enthousiasme waarmee men getracht heeft het onderwerp gestalte te geven binnen de belevingswereld en het voorstellingsvermogen van de leerling. Aangezien alleen al het aankweken van enig 'gevoel' voor statistiek een zeer moeilijke en tijdrovende onderneming is – de statistiek houdt er ogenschijnlijk een geheel eigen en grillige denkwijze op na – kan aan het ontwikkelen van statistisch gereedschap, bijvoorbeeld het uitdenken van statistische toetsen, amper aandacht worden geschonken.
Geconstateerd moet worden dat, op verschillende terreinen, vanuit statistisch oogpunt beschouwd, ontsporingen hebben plaatsgevonden.

Voor wat betreft het terrein van de definiëring van begrippen en de introductie van grootheden moet worden opgemerkt dat in het onderwijsmateriaal nogal

wat onjuistheden voorkomen; zo krijgen bijvoorbeeld 'overschrijdingskans', 'zijdigheid van alternatieve hypothese' en 'histogram' niet altijd het juiste licht. Deze mankementen zijn te repareren, hoewel soms flinke gedeelten tekst daarvoor op de helling zouden moeten.
Op het terrein van de vraag welke statistische techniek in een zekere correct beschreven praktische situatie moet worden toegepast, ontdekt men verschillende misverstanden. Over het algemeen wordt onvoldoende beseft dat de statistische gereedschappenkist waarover de leerling de beschikking krijgt, uiterst spaarzaam is gevuld, zodat sommige problemen waarvoor wel degelijk statistische technieken beschreven zijn, op het niveau van de leerling onopgelost moeten blijven. In sommige gevallen is zo'n statistische techniek duidelijk vanwege complexe aard niet in het examenprogramma ingepast. Het komt echter ook voor dat schoolboeken een geheel eigen arbeidsintensieve techniek ontwikkelen voor problemen waarvoor een aandoenlijk eenvoudige statistische oplossingstechniek bestaat die echter niet in het examenprogramma is geformuleerd. Op statistische gronden moeten deze eigen technieken vervolgens dan ook nog eens als ondeugdelijk of inferieur worden afgewezen.

De eerlijkheid gebiedt te zeggen dat het hierboven gesignaleerde enthousiasme ook hier en daar tot excessen heeft geleid. In overmoed werd soms gemeend dat – in beeldspraak – met hout, spijkers en een nijptang de fabricage van maanlanders in zicht kwam.
Een niet op zichzelf staand voorbeeld is het 'Hamilton-Madison'-vraagstuk in het eerste tijdvakexamen van 1987, waarvan naar de mening van de auteur de conclusie moet luiden dat het niet beantwoorden van de (alle) onderdelen van een dieper inzicht in de stof kan getuigen, dan het beantwoorden ervan volgens de richtlijnen in het antwoordmodel. Wellicht zou men er beter aan doen dergelijke stukken bij te zetten in het museum van de ontstaansgeschiedenis van wiskunde A, in plaats van er nog generaties lang leerlingen als oefenvraagstuk mee te confronteren. De auteur heeft zijn bevindingen met betrekking tot onder andere dit vraagstuk neergelegd in een rapport dat aan de CEVO is aangeboden. De indruk bestaat dat inmiddels een heroverweging is gemaakt terzake de keuze van examenvraagstukken.

In dit artikel wordt in de problematiek verder globaal op slechts één aspect nader ingegaan, te weten het drielandenpunt waar statistische theorie, kritische instelling en praktijksituatie samenkomen. Daartoe worden in schoolboeken voorkomende vraagstukken (soms aangevuld met de in het schoolboek gepubliceerde oplossing) voor wat dit aspect betreft aan een kritisch onderzoek onderworpen.
Met nadruk wordt gesteld dat dit aspect, hoewel van zekere importantie, dus bepaald niet het hele spectrum van alle mogelijk te leveren kritiek omspant. Op het moment van voorbereiding beschikte de auteur (slechts) over de in de literatuurlijst opgenomen schoolboeken. Uit deze boeken is min of meer naar evenredigheid een keuze voor vraagstukken gemaakt. Bewust is geen aselecte steekproef getrokken, maar zijn vraagstukken uitgekozen waaraan een zeker type tekortkoming duidelijk kan worden gedemonstreerd. Doorgaans wordt in het onderstaande dan ook slechts een zeer beperkt aantal tekortkomingen besproken, terwijl in sommige gevallen andere in dat zelfde vraagstuk voorkomende mankementen onvermeld blijven. De hier gepresenteerde (delen van) vraagstukken zijn daarom meestal eenvoudiger van aard dan andere in de schoolboeken opgenomen vraagstukken.
Inmiddels opgebouwde ervaring heeft geleerd dat in deze sfeer de statistische onvolkomenheid van een vraagstuk over het algemeen een explosief toenemende functie is van de complexiteit van het vraagstuk.

Uit het bovenstaande moge duidelijk zijn dat de hieronder te geven lijst van voorbeelden tegen vrij weinig moeite met andere voorbeelden kan worden aangevuld. Er is in dit geval van afgezien om bij ieder van de besproken vraagstukken de precieze bron te vermelden in de hoop dat iedere gedachte in de richting van een heksenjacht uitgebannen zou worden. Aangenomen mag worden dat eenieder die de wiskunde en haar beoefenaren een warm hart toedraagt waar mogelijk eraan zal willen meewerken dat de kwaliteit van het onderwijs in wiskunde, en in dit geval met name in wiskunde A, bewaakt en verbeterd wordt. Dit artikel is vanuit die optiek geschreven, waarbij de hoop wordt uitgesproken dat de hier gegeven aanwijzingen van misverstanden een eerste stap vormen op de weg in een genezingsproces. Een nog niet beantwoorde vraag is echter of, gegeven het huidige examenprogramma, de geconstateerde tekortkomingen consequent kunnen worden opgevuld met aan de realiteit ontleende voorbeelden.

Een centraal thema is uiteraard het statistisch toetsen van hypothesen. Om duplicering in het vervolg van dit artikel te voorkomen worden hier enige karakteristieken van dit onderwerp besproken, zij het zeker niet uitputtend. De toelichting geschiedt aan de hand van een voorbeeld.
De Vereniging van Sociaal Werkers in een zeker land ging er tot nu toe steeds vanuit dat een fractie $p = 0{,}02$ van de volwassen inwoners analfabeet is. Er bereiken deze vereniging echter signalen dat van een geflatteerd beeld sprake is: in werkelijkheid zou p groter dan 0,02 zijn. Men wenst dit statistisch te toetsen. Daarbij wordt de 'traditionele' bewering die stelt dat $p = 0{,}02$ is, als nulhypothese (H_0) gekozen; het *statistisch te bewijzen* vermoeden, namelijk dat p in werkelijkheid groter is, wordt als alternatieve hypothese (H_1) gesteld.
In tegenstelling tot wat de schoolboeken die hieraan aandacht besteden vermelden, is in de technische uitwerking van het geheel van niet noemenswaard belang of H_1 enkelvoudig wordt gekozen (bijvoorbeeld $p = 0{,}03$) of samengesteld ($p > 0{,}02$). Ook alweer in tegenstelling met wat genoemde schoolboeken doen voorkomen, is het fundamentele uitgangspunt van het statistisch toetsen van hypothesen dat van symmetrie, in de zin van verwisselbaarheid van de beide hypothesen, *geen* sprake is: H_0 is de hypothese die men wenst te verwerpen, terwijl men H_1 op grond van experimentele resultaten wil bewijzen. Daartoe zal een aselecte steekproef worden getrokken, en wel uit de *populatie waarvoor een uitspraak wordt verlangd*.

In tal van in de schoolboeken beschreven situaties wordt deze koppeling niet gemaakt; in overdrachtelijke zin geeft men de leerling in zo'n geval het cijfermateriaal dat in een onderzoek in een van de wijken van een zekere stad van het land is verzameld, en men vraagt na te gaan of de nulhypothese op grond van de verkregen resultaten moet worden verworpen.
Vergeleken met de kunst om in een dergelijk geval een in vier decimalen met de antwoordenlijst overeenstemmende overschrijdingskans te kunnen produceren, is het in de visie van de auteur van zo mogelijk nog groter gewicht dat de kritische leerling, zelfs bij ogenschijnlijk zeer overtuigend bewijsmateriaal (bijvoorbeeld 37 van de 100 aselect uit deze wijk getrokken personen blijken analfabeet), zich afvraagt of de betrokken wijk niet, met verder alle respect, een sloppenbuurt is die bepaald niet een beslissende indruk kan geven omtrent de waarde van de *populatiegrootheid* p waarin men is geïnteresseerd.

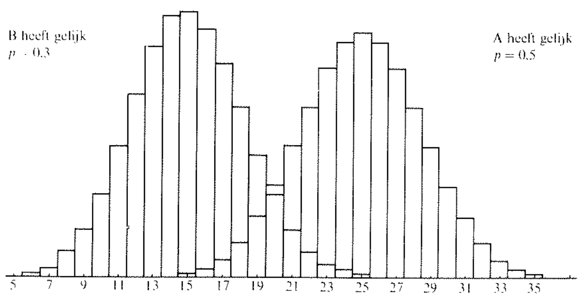
Terwijl we nu maar verder de analfabeten laten voor wat ze zijn, geven we hier nog een laatste karakteristiek van de statistische bewijsvoering via de methode van het statistisch toetsen van hypothesen. Wanneer de nulhypothese wordt verworpen, wordt de alternatieve hypothese, wanneer de proef verder correct is opgezet en het gebruikte model juist is, statistisch bewezen geacht. Een parameter in deze gehele bewijsvoering is de door de onderzoeker ingestelde onbetrouwbaarheidsdrempel α. De rol van α is de volgende.
De kans dat de zojuist op basis van statistische toetsing verworpen nulhypothese onterecht naar het rijk der fabelen is verwezen – een in de statistiek hoogst ernstig opgenomen voorval – is kleiner dan of gelijk aan de kans dat men bij aselecte trekking van een bal uit een vaas die zwarte en witte ballen in een verhouding van $\alpha : (1-\alpha)$ bevat, ongelukkigerwijze een zwarte opdiept (even afgezien van de flauwe beperking dat hiervoor α rationaal moet zijn). Het principe van de gevolgde werkwijze is daarmee onmiskenbaar dat van het leveren van een *bewijs uit het ongerijmde*: uitgaande van H_0, zijnde juist de hypothese die men wenst te verwerpen (sportiever kan het toch niet), hoopt men dat de in *werkelijkheid* verkregen onderzoeksresultaten een zodanige afwijking vertonen van wat men onder H_0 zou verwachten, dat vriend en vijand (oftewel u en uw criticus) in alle redelijkheid niet meer in de geldigheid van H_0 kunnen geloven. In dat geval wordt H_0 verworpen ten faveure van H_1.

Behalve statistische bewijsvoering in de enge zin via de zojuist aangestipte methode van het statistisch toetsen van hypothesen, zou men aan dit begrip ook een ruimere interpretatie kunnen geven als volgt. Iedere bewering die men met behulp van statistische gereedschappen poogt aannemelijk te maken, of dit nu een uitspraak over de kans op 'morgen zonneschijn' is, of 'meisjes minder mans in wiskunde', dient in de bijgeleverde argumenten kritisch tegemoet te worden getreden. Een aardige sport, zeker voor vwo-leerlingen, en tegelijk rakend in wat men zou kunnen aanwijzen als het hart van de statistiek, is nu om te onderzoeken waar de zwakke plekken in een bewijsvoering zitten. Voor men nu in woede zou ontsteken om een zo plompverloren uitgesproken negatieve houding jegens andermans werk: de ergste vijand (criticus) van de goede onderzoeker is de onderzoeker zelve; pas nadat hijzelf van alle wegen die van zijn conclusie zouden kunnen afvoeren, heeft geconstateerd dat ze doodlopend zijn, presenteert hij zijn ontdekkingen (wiskundig gesproken dan nog zo iets als 'modulo α').

De meeste van de in de schoolboeken gegeven voorbeelden schieten op het punt van correcte statistische bewijsvoering in deze brede zin ernstig tekort. In de hieronder te geven voorbeelden worden wegen aangegeven die leiden tot andere gevolgtrekkingen dan de gevolgtrekkingen die de leerling geacht wordt te maken.

Voorbeeld 1

Nederland is een waterrijk land. Er zullen dus betrekkelijk veel mensen kunnen zwemmen. Hoeveel volwassen Nederlanders inderdaad kunnen zwemmen, is niet precies bekend.
Twee mensen hebben er verschil van mening over.
A beweert dat het 50% is.
B beweert daarentegen dat dit minder, namelijk 30%, is. Zij willen door middel van een onderzoekje uitmaken wie gelijk heeft, althans wie ze gelijk zullen geven.
Of degene die ze gelijk zullen *geven*, echt gelijk *heeft*, is niet vast te stellen, want, zoals gezegd, het werkelijke percentage Nederlanders dat kan zwemmen is niet precies bekend.
Zij spreken af dat ze aan 50 willekeurige volwassenen zullen vragen of hij/zij kan zwemmen. Het aantal mensen dat deze vraag positief beantwoordt, is een stochast X. Zij gaan ervan uit dat X binomiaal verdeeld is. Ingeval A gelijk heeft is p gelijk aan 0,5. Ingeval B gelijk heeft, is p gelijk aan 0,3.
(...) Van de kansverdelingen die bij $p = 0{,}5$ en bij $p = 0{,}3$ horen staan hieronder de histogrammen getekend.



Commentaar

Terzijde: de onderzoeker dient er terdege rekening mee te houden dat er een discrepantie kan bestaan (en in de praktijk ook zeker zal bestaan) tussen de begrippen 'kunnen zwemmen' en 'menen te kunnen zwemmen'. Hoewel het verschil van mening over het eerste gaat, tracht men het geschil te beslechten via het tweede, waardoor mogelijk een vertekend beeld ontstaat. Misschien hebben A en B ook verschillende interpretaties van 'kunnen zwemmen'.
Hoewel de bijgevoegde figuur in dit geval weinig ruimte laat voor misverstanden, is de naamgeving 'histogram' niet erg gelukkig omdat deze slechts voorbehouden is aan continue stochasten. Beter kiest men in ons geval van een discrete stochast voor de term 'kolommendiagram'. In het algemene geval van een discrete stochast met uitkomsten die niet op gelijke afstanden liggen (uitkomsten bijvoorbeeld $1, 5, 6\frac{1}{2}, 9$)

houdt het verschil méér in dan alleen een verschil in naamgeving: in een kolommendiagram worden de diktes van de kolommen terecht constant gekozen; een 'histogram' met 'ongelijke klassebreedtes' kan in zo'n geval een compleet vertekend beeld geven.

Elk van de schoolboeken die op dit type toetsing (H_0: $p = p_0$ tegen H_1: $p = p_1$) ingaat, houdt er een eigen oplossingstechniek op na. Geen van deze schoolboeken onderkent dat hier in feite sprake is van een eenvoudig *statistisch toewijzingsprobleem*: geef die partij in het geschil gelijk die bij de verkregen uitkomst van het onderzoek de grootste kans op het optreden van die uitkomst heeft. In dit geval beslist men dat A gelijk heeft als X een uitkomst ter grootte 20 of meer oplevert, in het andere geval is het B die op basis van de koele cijfers meer het gelijk aan zijn kant lijkt te hebben. 'Toewijzing' van het gelijk aan A dan wel B geschiedt dus eenvoudig via de hoogste plaatselijke kans.
De methode via het statistisch toetsen van hypothesen is hier een onjuiste methode onder meer omdat de beschreven ruzie symmetrisch van opbouw is.

Voorbeeld 2
Piet gaat schijfschieten. Zijn kans om de roos te raken is $\frac{2}{5}$.
Hij schiet vijf keer.
X is het aantal malen dat hij raakt.
Geef de kansverdeling van X.

Oplossing
Het experiment is een binomiaal kansexperiment met parameters $p = \frac{2}{5}$ en $n = 5$.
Voor de kansverdeling van X vinden we dus
$P(X=0) = (\frac{3}{5})^5 = 0{,}0778$
(...)

Commentaar
De essentiële aannamen om inderdaad een binomiaal model te mogen aannemen zijn niet vermeld, namelijk dat de vijf door Piet te ondernemen pogingen *statistisch onafhankelijk* zijn en bovendien dat de kans op raakschieten een *constante succeskans* is. De hier gegeven kans ter waarde van $\frac{2}{5}$ kan een soort van over de dag gewogen gemiddelde zijn van Piet's kwaliteiten in deze. Het kan dus wat uitmaken op welk moment van de dag de avondmens Piet de bewuste vijf schoten lost.
Vanaf nu zullen we echter aannemen dat Piet's schietprestaties van een weliswaar stochastische, maar niettemin constante kwaliteit zijn.
Laat verder M_i de gebeurtenis zijn dat Piet's i^{de} schot een misser is.
Piet moet onmiskenbaar over bijzondere psychische vaardigheden beschikken wanneer hij zich niet door de uiterst deprimerende gebeurtenis $D = \bigcap_{i=1}^{4} M_i$ uit het lood laat slaan. Toepassing van een binomiaal model vergt in deze van Piet een zekere flegmatieke instelling, waardoor $P[M_5 | D]$ onveranderd gelijk is aan $\frac{3}{5}$.
Ook in de wat meer serieuze onderzoekspraktijk is het van uitermate groot belang om de beide genoemde voorwaarden voor de postulering van een binomiaal model op hun praktische meritus te beoordelen. Het gebeurt niet zelden dat beweringen die op basis van een onterecht aangenomen binomiaal model statistisch leken bewezen, moeten worden herzien. Soms is een heel andere verklaring van een waargenomen verschijnsel te geven dan de onderzoeker op basis van het binomiale model in eerste instantie voor ogen had. De praktische conclusie in Piet's geval is dat de in het schoolboek in vier decimalen berekende kansen voor doorsnee Pieten zeker niet opgaan; de bijzondere gaven van de schoolboek-Piet hadden niet onvermeld mogen blijven.

Voorbeeld 3
Volgens de VVV van het eiland Texel regent het daar in de zomer op 15% van de dagen.
Anja gaat daar een week kamperen. Vier van de zeven dagen regent het. Laat p de kans zijn dat het op een willekeurige zomerdag op Texel regent. X is het aantal dagen dat het regent in de week dat Anja daar kampeert.
Neem aan (H_0) dat de VVV gelijk heeft. Wat is dan E(X)?
Hoe groot is (onder H_0) de overschrijdingskans van $X = 4$?
Welke verdeling heb je (waarschijnlijk) gebruikt bij het beantwoorden van de vorige vraag? Was dat wel juist?

Commentaar
Stemt Anja's definitie van 'regendag' met die van de Texelse VVV overeen?
Genoemde 15% is een gemiddelde over de *gehele* zomerperiode en een groot aantal jaren, eventueel nog gecorrigeerd voor een broeikaseffect.
Anja's zomer was toevallig vrij buiig. Verder viel Anja's kampeervakantie in de late nazomer, waarvoor deze zelfde VVV, wanneer men vasthoudend bij deze instantie daarover doorvraagt, een over de jaren gemiddeld regenpercentage ter grootte van 23% opgeeft.
Niet te wensen is dat Anja, die men een betere vakantie had gegund omdat ze net met uitstekende cijfers een vwo-diploma in de wacht gesleept heeft, op de zevende dag van haar vakantie, *op basis van haar in het water gevallen vakantie* en enig binomiaal rekenwerk, de Texelse VVV de mantel heeft uitgeveegd omtrent een naar haar mening overtuigend bewezen onjuistheid met betrekking tot het Texels zomerregenpercentage dat vermeld staat in de overigens fraai en wolkenloos uitgevoerde kleurenfolder.
Anderzijds hadden natuurlijk twee (!) korte nachtelijke buitjes in een verder stralende week óók tot vier 'regendagen' kunnen leiden.
De leerling moet, nadat hij reeds binomiaal heeft zitten rekenen, nog een antwoord geven op de gestelde vraag: 'Was dat wel juist?' Hoewel op zich dit moment van reflexie weldadig aandoet, zijn er enige reminiscenties aan de officier die een soldaat het bevel tot schieten gaf en vervolgens van de soldaat rekenschap afeiste inzake zijn morele verantwoordelijkheid ten aanzien van enkele ethische vraagstukken samenhangend met het zojuist geloste schot.

Voorbeeld 4
Kandidaat-kosmonauten worden aan een zware (lichamelijke en psychologische) test onderworpen alvorens zij toegelaten worden tot een verdere opleiding.
De kans dat een kandidaat slaagt voor de eerste test is 10%.
Voor kandidaten die de eerste test niet gehaald hebben, volgen nog maximaal twee herkansingen, met iedere keer op-

nieuw dezelfde kans op slagen. Als de kandidaat de derde keer opnieuw niet aan de eisen voldoet, is hij definitief afgewezen.
Neem aan dat de testresultaten onafhankelijk van elkaar zijn.

a. Hoeveel % is de kans dat een willekeurige kandidaat wordt afgewezen?
b. Wat is de verwachtingswaarde van het aantal tests dat een willekeurige kandidaat zal moeten ondergaan?
c. Hoe groot is de kans dat bij een keuring van een groep van twintig kandidaten er bij de eerste test één slaagt, bij de tweede nul en bij de derde test weer één?
d. Uit een groep van vijftig kandidaten slaagde er slechts één bij de eerste test. Op grond van dit resultaat vermoedde iemand dat de kans op slagen kleiner was dan 10%. Is dit vermoeden terecht, wanneer men een significantieniveau van 2,5% aanneemt?

Commentaar

Men ziet dat in dit geval de onafhankelijkheid (keurig?) is gepostuleerd. Echter, zeer betreurenswaardig, zal in dit geval de aanname van onafhankelijkheid ervoor zorgen dat het vraagstuk goeddeels onoplosbaar wordt, zoals we hieronder zullen zien.
Terzijde: 'alvorens zij toegelaten worden tot een verdere opleiding' wekt de suggestie als zouden alle deelnemers aan de test uiteindelijk toch worden toegelaten.
Verder: (onderdeel c.) is het zo dat 'één slaagt, nul [*slaagt*], één slaagt'?
Een typische taalkronkel beslist dat er nul (in meervoud) *slagen* en één (in enkelvoud) *slaagt*. In de aanhef van het verhaal leidt het simultaan gebruiken van 'test', 'eerste test' en 'testresultaten' tot enige verwarring. Bij welwillend verder lezen kan men hieruit echter ontsnappen.

Wat betekent precies: 'de kans dat een kandidaat slaagt is 10%'?
De goedwillende lezer in wiens oren de woorden 'zware (lichamelijke en psychologische) test' nog resoneren, meent te begrijpen dat van de hordes die zich iedere keer weer aanmelden, door de bank genomen zo'n 10% slaagt. Voor deze goedwillende lezer is de kwaliteit van de af te leggen test uiteraard zodanig ('*zware*' test) dat deze op z'n minst enige selectieve waarde heeft en dus meer is dan een simpel geluksspel.
Mis! Om in de pas te lopen met de opstellers van dit vraagstuk moet wel degelijk slechts aan de vorm van een geluksspel worden gedacht, anders gaat de opgave zonder pardon de mist in. Maar zelfs wanneer men al in staat is gebleken de zware psychologische barrière te nemen om slechts te denken in 'geluk' en 'pech', dan gaat men uiteindelijk toch nog voor de bijl wanneer men de aanname dat de testresultaten *onafhankelijk* zijn, serieus wenst te nemen. Immers, als we met G_i aangeven de gebeurtenis dat voor een zekere met name genoemde kandidaat de i^{de} test gunstig afloopt, en met P_i de gebeurtenis dat de i^e test hem daarentegen juist parten speelt, dan zou de kans op $G_1 \cap P_2$ gelijk moeten zijn aan het produkt van de kansen op de beide afzonderlijke gebeurtenissen, een onjuistheid.

Pogingen om de zaak te redden door een gebeurtenis 'kandidaat niet bij test aanwezig wegens voor hem gunstige afloop van een der vorige tests' in het leven te roepen en in de berekeningen te verdisconteren, lossen het probleem niet op. Behoudens een nogal gekunstelde noodsprong is er slechts één duidelijke manier om uit de impasse te komen en de gepostuleerde onafhankelijkheid recht te doen, namelijk door alle kandidaten alle drie tests te laten ondergaan (bij reeds geslaagd: pro forma). Antwoorden op de vragen b. en c. zijn daarmee gemakkelijk te geven. Deze antwoorden komen echter niet met de gepubliceerde antwoorden overeen.
In onderdeel d. wordt een toetsingsprobleem aan de orde gesteld. Ongelukkigerwijze wordt de alternatieve hypothese geformuleerd *op basis van de bij het experiment verkregen data*, er wordt immers 'op grond van dit resultaat' een vermoeden geformuleerd. Deze manier van werken wordt in de statistiek wel *data snooping* genoemd; in feite werkt men naar een gewenst resultaat, i.c. het te bewijzen vermoeden, toe. Men dient echter, terwille van sportiviteit en bewijskracht, vermoedens te formuleren *voordat* de steekproefuitkomsten bekend zijn. Een voorbeeld zou dit nog aannemelijk kunnen maken. Gesteld dat de kandidaten in volgorde worden getest, en gesteld dat u op grond van nauwkeurige inspectie van de testresultaten hebt ontdekt dat volgnummers van geslaagden steeds een zevenvoud zijn. U verheft dit vermoeden tot alternatieve hypothese en hanteert een voor dit geval passende toets. Enigszins plastisch gesproken is de toets bij confrontatie met de door u aangeleverde data zeer onder de indruk van uw 'voorspellende' vermogens in een onvermoede richting en besluit op grond van de bijpassende resultaten dat uw vermoeden overtuigend is bewezen. U hebt echter uzelf (en mogelijk uw collega-wetenschappers) een slechte dienst bewezen: in feite was het natuurlijk allemaal puur toeval, een morgenwolk, maar ongetwijfeld bent u alweer bezig met een fijnzinnige ontrafeling van een nieuwe reeks van waarnemingen, waaraan u alweer zeer merkwaardige verschijnselen hebt ontdekt die nog wachten op uw ogenschijnlijk eenvoudig te leveren statistische 'bewijs'. Wanneer u op dit hellend vlak onvoldoende wrijving ondervindt zult u er, als ultiem geval, ongetwijfeld nog weleens toe komen om statistisch te 'bewijzen' dat de steekproef zoals die zich voor uw ogen heeft ontrold, noodzakelijkerwijze precies de resultaten zou geven zoals zij te zien gegeven *heeft*. Op dat moment zult u de euforie beleven van grenzeloos voorspellende vermogens.

Voorbeeld 5

Vroeger waren de meeste fietsen in Nederland voorzien van een terugtraprem. Tegenwoordig is dat anders: 60% van de Nederlandse fietsers gebruikt handremmen.
Bij een controle worden door de politie negentien willekeurig gekozen fietsers aangehouden.
Hoe groot is de kans dat een meerderheid van de aangehouden fietsers over een fiets met terugtraprem beschikt?

Commentaar

Een stap in de goede richting was wellicht het geven van het percentage per 'handrem' afgelegde fiets*kilometers* geweest.
Bekend zou kunnen zijn dat, vanwege nostalgische gevoelens, vooroorlogs Nederland zich immer per

'terugtraprem' verplaatst. De jeugd zit massaal en veel op de 'handrem'. Navraag leerde dat de controle plaatsvond nabij een grote scholengemeenschap en wel des morgens zeer vroeg. Geheel volgens het boekje werden op aselecte wijze door het aanhoudingsteam negentien fietsers tot staan gebracht. In eerste instantie wekte de vondst van een 'terugtraprem' bij de ijverige dienaren der wet enige bevreemding tot werd opgemerkt dat het hier een leraar betrof, die vanwege jarenlange financiële achterstelling, zich de automobiele weelde niet meer kon veroorloven en zich dientengevolge thans per rijwiel verplaatste.
Vermeldenswaardig is nog dat de commandant, een oude rot in het vak én Brigadier van politie, bij confrontatie met de door geleerde kansberekenaren geleverde kans ter waarde van 18,61% de juistheid van deze uitkomst meende te moeten betwisten: 'Meneer, op deze plaats en tijd controleer ik al jaren; een gebeurtenis zoals u die beschrijft heeft zich, voorzover ik mij herinner, nog nooit voorgedaan, dus is voor mij de kans erop afgerond nul.' Waarmee maar weer eens is aangetoond dat op straat weinig valt te leren.

Voorbeeld 6

Een kweker beweert dat minstens 90% van de door hem geleverde bloembollen zullen opkomen.
Mevrouw Roos poot in haar tuin 80 bloembollen, waarvan er 68 opkomen. Accepteer je de bewering van de kweker bij $\alpha = 0{,}05$?

Commentaar

Men kan zich afvragen hoe de kweker bij een opkomstpercentage van 90(%) komt. Noemt hij dit uit reclame-overwegingen, of is hier sprake van de een of andere niet-goed-geld-terug-garantie? In dat geval is er voor mevrouw Roos goede hoop dat sprake zal zijn van een tegemoetkoming van de kant van de kweker.
Het kan ook zijn dat de kweker genoemd percentage 'kent' op grond van jarenlange ervaring. Hij weet dan dat er goede en slechte jaren zijn, dat verschillende grondsoorten tot verschillende percentages aanleiding kunnen geven. Het zal duidelijk zijn dat in dat geval de 'berekening' van het percentage nogal schimmig is: zoveel procent van de klanten koopt dit type bloembollen en heeft de beschikking over zandgrond, zoveel procent heeft kleigrond, maar wacht eens, de kleigronders kopen van dit soort doorgaans wèl twee keer zoveel, enzovoort. Genoemd percentage zou men daarom kunnen kenschetsen als een *spookgetal*.
Verder is het bekend dat zekere besmettingen van de ene bloembol op de andere binnen een zelfde partij kunnen worden overgebracht, zodat met het oog op de statistische onafhankelijkheid een beroep op een binomiaal model weleens onrealistisch zou kunnen zijn.
Nu beschikte mevrouw Roos helaas over een cadmium-nikkel tuintje. Dankzij haar uitstekende zorgen gepaard gaande aan groeizaam weer, bleef de schade aan de bollentuin toch nog beperkt: van de 80 gepote bollen kwamen er 68 op.
Toen mevrouw Roos aan haar opkomende ontevredenheid een statistische onderbouwing wenste te geven, bleek de door haar berekende overschrijdingskans net te groot om de kweker van kwader trouw te kunnen betichten. Gelukkig maar; wellicht was er anders nog een onvruchtbare discussie met de tamelijk ongecijferde kweker uit voortgekomen.

Voorbeeld 7

Aan 100 personen wordt gevraagd of zij vertrouwen hebben in weersvoorspellingen.
30 van de 100 ondervraagden antwoordt dat zij geen enkele waarde hechten aan die voorspellingen.
Geef een 90%-betrouwbaarheidsinterval van het percentage Nederlanders, dat aan de voorspellingen geen waarde hecht, daarbij aannemende dat de ondervraagden een aselecte steekproef vormen van de Nederlandse bevolking.

Commentaar

Men dient er bij aselecte steekproeftrekking uit de Nederlandse bevolking op bedacht te zijn dat men voor het probleem kan komen te staan dat een zuigeling volgens de voorschriften zal moeten worden ondervraagd.
Bij het interpreteren van enquêtegegevens moet men verder niet uit het oog verliezen dat ze het resultaat kunnen zijn van een tendentieuze vraagstelling: 'vindt u ook niet dat ...?'
De enquêtes werden afgenomen nadat het KNMI (weer eens) drie dagen achter elkaar misgekleund had. De meeste interviews vonden in de stromende regen plaats en de meeste ondervraagden hadden net op grond van de weersvoorspelling hun paraplu thuisgelaten.

Voorbeeld 8

De atletes Ank en Bea doen mee aan een wedstrijd over 400 meter.
Uit ervaring weet Ank dat haar tijd X in sec., waarin zij deze afstand loopt, normaal verdeeld is met $E(X) = 53$ sec. en $\mathrm{SD}(X) = 3$ sec.
De 400-metertijd Y van Bea is een normaal verdeeld stochast met $E(Y) = 55$ sec. en $\mathrm{SD}(Y) = 4$ sec.
Benader de kans dat Bea sneller loopt dan Ank.

Commentaar

Omdat de dames aan één en dezelfde wedstrijd meedoen rijst de vraag wat de simultane, in dit geval bivariate, kansverdeling is die de dames erop nahouden. Hieruit zouden in principe rare dingen kunnen komen, bijvoorbeeld als gevolg van een sterke psychische druk die op Bea rust, omdat ze weet dat Ank als winnares gedoodverfd wordt.
Teneinde toch nog wat te kunnen rekenen, zullen we maar aannemen dat dergelijke merkwaardige fenomenen zich niet voordoen. In dat geval lijkt het niet onredelijk om voor de simultane kansverdeling een zogeheten Bivariaat Normaal model te postuleren dat naast de reeds aangegeven parameters ook de *correlatiecoëfficiënt* ρ bevat; er geldt $-1 \leqslant \rho \leqslant 1$. In deze laatste parameter is verdisconteerd hoe de tijden van de dames onderling van elkaar afhankelijk zijn. Aldus bevat het Bivariate Normale model vijf parameters: $E(X)$, $E(Y)$, $\mathrm{SD}(X)$, $\mathrm{SD}(Y)$ en ρ. Voor het gemak nemen we in het volgende aan dat (slechts) de *weersinvloed* het stochastische karakter geeft aan de af te leggen tijden. (Deze beperking is niet wezenlijk; het onderstaande verhaal vertelt alleen wat vlotter als je het maar over één variatiebron hebt).

Nu is Ank een tenger meisje en nogal zijwindgevoelig; een handicap is bovendien dat ze absoluut niet tegen

regen kan. Bij zonneschijn loopt ze echter als een lier. Bea is in al dit soort zaken een ware tegenpool van Ank. In dit geval lijkt een flink negatieve keuze voor ρ voor de hand te liggen. Wanneer de dames echter sterk op elkaar hadden geleken, zou ρ fors positief moeten zijn gekozen. Daar niets over de onderlinge wisselwerking van de tijden van de dames is bekendgemaakt, kan ρ in principe iedere waarde tussen -1 en 1 aannemen, inclusief de grenzen. Juist in deze beide uitersten bestaan er tussen de tijden X en Y zelfs *deterministische* relaties. In deze gevallen geldt $Y = (-4X + 377)/3$, respectievelijk $Y = (4X - 47)/3$, hetgeen men eenvoudig kan controleren door van linker- en rechterlid van deze gelijkheden de verwachting en de variantie te nemen. In het geval $\rho = -1$ wordt de gevraagde kans dan:
$P(Y < X) = P(-4X + 377 < 3X) =$
$P(X > 377/7) = 38{,}75\%$. Analoog vindt men in het geval $\rho = 1$ dat de gevraagde kans gelijk is aan 2,28%. De conclusie is dat de kans ergens tussen 2,28% en 38,75% zal uitkomen; de precieze uitkomst is bij gebrek aan informatie volstrekt onduidelijk.

Uitgaan van een Bivariaat Normaal model en onafhankelijke stochasten (zodat ρ de waarde 0 heeft) is een slag in de lucht. Zo te zien hebben de opstellers van het vraagstuk dit laatste model echter wel in gedachten gehad. De vraag om de kans te *benaderen* komt dan vreemd over, immers in dat geval is een *exacte* berekening mogelijk, en wel resulterend in 34,46% als uitkomst. Voor het juiste begrip zij nog meegedeeld dat een opmerking in de richting als zou de vraag om de kans te *benaderen* juist ingegeven zijn met het oog op het ontbreken van informatie over de correlatiecoëfficiënt dient, mede gelet op bovenvermelde *range* voor de gevraagde uitkomst bij variërende waarde voor ρ, als ongerijmd te worden afgedaan.
Bovenstaande berekeningen gelden voor een hypothetische wedstrijd in de (verre) toekomst. Bij een wedstrijd die bijvoorbeeld over een uur zal plaatsvinden, zou, als de vijf parameters van het model bekend zijn en de invloed van de weersomstandigheden op het loopgedrag van de beide dames eveneens bekend is, in principe een bijstelling van de berekende kans mogelijk moeten zijn op basis van de weersvoorspelling.

Tussentijdse evaluatie

De tot dusverre gepresenteerde voorbeelden laten reeds zien dat het huidige onderwijs in vwo wiskunde A voor wat betreft het onderwerp Statistiek tekortkomingen vertoont, mèt name op het terrein van de verwerving van 'kritisch inzicht', juist een der doelstellingen van wiskunde A. Wat dit laatste betreft moet nuchter worden vastgesteld dat de kritische keizer nu en dan in z'n hemd staat. Een grondige bezinning op het onderwerp is daarom absoluut noodzakelijk.
In het nog te verschijnen tweede (laatste) deel zullen nog andere typen van misvatting op dit gebied aan de hand van voorbeelden worden toegelicht.

Literatuur

Bos, D. J. P. e.a.: *Wiskunde lijn · Wiskunde A*, Jacob Dijkstra, Groningen, z.j.
Broek, L. van den e.a.: *De Wageningse Methode, deel 56, Hypothese toetsing*, Stichting De Wageningse Methode, 1988.
Dijkhuis, J. H. e.a.: *Getal en Ruimte. 5/6V ~A1 en ~A2*, Educaboek BV, Culemborg, 1985.
Doove, F. M. W. e.a.: *Sigma 5A, 6A*, Wolters-Noordhoff, Groningen, 1986.
Groen, W. E. e.a.: *Moderne wiskunde bovenbouw Vwo 5A en 6A*, Wolters-Noordhoff BV, Groningen, 1987.
Putten, B. van: *Enige kanttekeningen bij de Vwo-examens Wiskunde A in 1987*, Rhenen, februari 1988.

Noten

[1] Dit artikel heeft een ruime overlap met de voordracht 'Statistiek in Wiskunde A: nieuwe, gebruikte en gemiste kansen', die de auteur op uitnodiging heeft gehouden op 6 januari 1990 te Amersfoort tijdens het Wintersymposium van het Wiskundig Genootschap.
[2] Correspondentieadres: vakgroep Wiskunde, Landbouwuniversiteit, Dreijenlaan 4, 6703 HA Wageningen.