Statistiek voor wiskunde A en C

Van beschrijving van een steekproef naar conclusies over een populatie

Anne van Streun, Bert Zwaneveld en Paul Drijvers
met bijdragen van Lidy Wesker

Inhoud

There are three kinds of lies: lies, damned lies, and statistics

Disraeli (1804 – 1881)

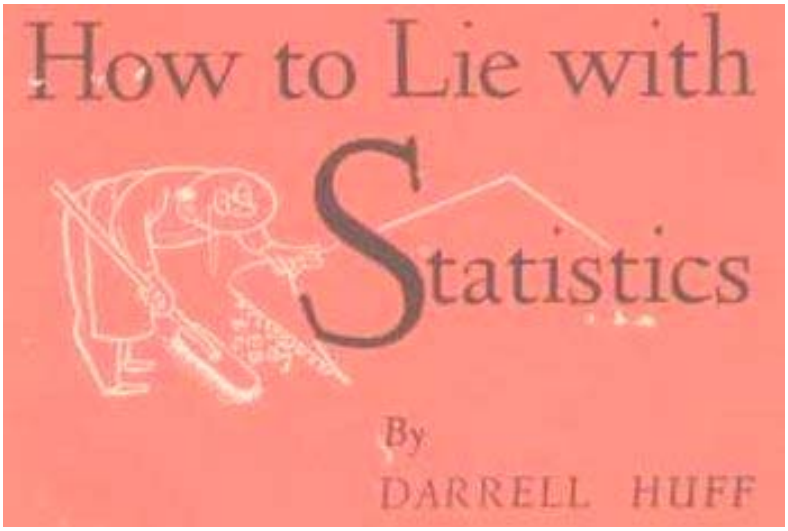
1. Oriëntatie

In de Volkskrant van 11 november 2006 stond het volgende bericht.

> Stel, u bent zwanger en u besluit een test te laten doen op het syndroom van Down. Zo'n test is uitstekend, maar niet feilloos. Ongeveer 1 procent van de baby's heeft het syndroom. Als het kind het syndroom heeft, is er 90 procent kans dat de testuitslag positief zal zijn. Als de baby niet is aangedaan, is er toch nog 1 procent kans dat de testuitslag positief is. U laat zich testen, en de test is positief. Hoe groot is de kans dat het kind werkelijk het syndroom van Down heeft?
> Dit soort berekeningen is altijd erg ingewikkeld — zeker voor zenuwachtige vrouwen en hun partners. Maar gelukkig zijn er dan deskundigen die een handje kunnen helpen.
> Nou ja, niet dus de verloskundige. Toen dit probleem aan 22 verloskundigen werd voorgelegd, kon geen enkele de juiste cijfers leveren. Gynaecologen deden het iets beter: van de 21 specialisten wist er 1 het juiste antwoord, even veel als in de groep zwangeren zelf. Ook toen het probleem wat makkelijker werd geformuleerd, stonden de zorgverleners nog met de mond vol tanden.

De vraag die volgens dit krantenbericht zowel betrokkenen als deskundigen voor raadsels stelt, luidt: hoe groot is de kans op een kind dat werkelijk het syndroom van Down heeft, als de test daarop positief is? Het feit dat slechts één van de 21 gynaecologen hierop het juiste antwoord weet, toont aan dat statistiek een lastig vak is, terwijl het in verschillende beroepspraktijken – en daarmee ook in veel vervolgopleidingen – een belangrijke rol speelt.

Ook het citaat bovenaan op deze pagina maakt duidelijk dat statistiek 'tricky' is: de combinatie van ogenschijnlijk eenvoudige en praktijknabije situaties enerzijds en ingewikkelde onderliggende vakkennis anderzijds maken het vak niet populair. Ook Huff (1954) benadrukt het gevaar van misbruik van de statistiek in zijn bekende boek 'How to lie with statistics'.



Figuur 1 Omslag van 'How to lie with statistics'

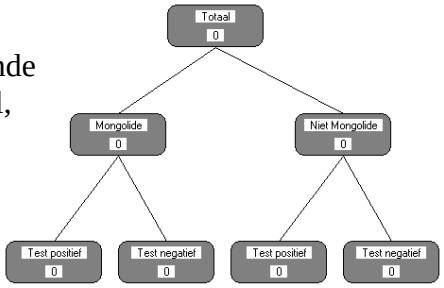
De statisticus Hemelrijk (1972), die begin jaren zeventig meewerkte aan een TV-serie over statistiek, spreekt van ‘een hachelijke zaak’:

> Het leven is vol onzekerheden en de wetenschap ook. De toekomst voorspellen is een hachelijke zaak, maar toch is het vaak nodig. De statistiek is een wetenschap die methoden, wiskundige methoden, tracht te ontwikkelen om het voorspellen op grond van waarnemingen beter mogelijk te maken. Een wetenschap die op allerlei terreinen kan worden gebruikt.

Aan de ene kant speelt statistiek dus een grote rol in onze hedendaagse maatschappij en in de moderne wetenschap. Op basis van verzamelde cijfermatige gegevens worden allerlei voorspellingen gedaan op basis waarvan belangrijke beslissingen worden genomen. Aan de andere kant wordt daarbij veelal gebruik gemaakt van onderliggende kanstheoretische en statistische modellen en toevalsprocessen, die de gebruiker niet kent en die ook niet eenvoudig te doorgronden of te beoordelen zijn. Een toegepast en ogenschijnlijk toegankelijk vak dus, statistiek, maar met diepe achtergronden en daardoor ook het risico van ten onrechte getrokken conclusies op basis van verkeerd begrepen data. Dit probleem speelt met name wanneer steekproefgegevens worden gegeneraliseerd en er op basis van een steekproef conclusies over een onderliggende populatie worden getrokken, of wanneer steekproefgegevens worden geëxtrapoleerd in de tijd om toekomstige trends te voorspellen.

Juist vanwege de maatschappelijke relevantie en de rol die het speelt in vervolgopleiding en beroepspraktijk heeft statistiek een belangrijke plaats in het wiskundeonderwijs in het VO. Ook voor de leerling die in de toekomst geen bèta-studierichting zal kiezen, is statistiek van belang. Daarom kennen de programma’s voor wiskunde A en C een aanzienlijke statistiek component. Echter, voor deze leerlingen is het lastig om diep op de kanstheoretische achtergronden in te gaan, zodat het gevaar bestaat dat het statistiekonderwijs ontaardt in een verzameling truukmatige technieken. Vanuit deze problematiek wordt gepleit voor een probleemgeoriënteerde invulling van de statistiekprogramma’s voor wiskunde A en C (cTWO, 2009; Van Streun en Van de Giessen, 2007ab). Deze aanpak vormt de achtergrond van dit katern. Speciale aandacht wordt besteed aan de denkstap van het beschrijven van een steekproef naar het trekken van conclusies over de onderliggende populatie. De kansrekening krijgt in dit katern weinig aandacht.

Opgave 1
Los het probleem in het krantenartikel verschillende manieren op. Gebruik bijvoorbeeld een kruistabel, het boomdiagram hiernaast, of een meer formele methode van kansrekenen.
Zie eventueel Tijms (2008) voor een bespreking van Bayesiaanse statistiek.



2. Probleemstelling

Het uitgangspunt van dit katern is de relevantie van de statistiek in het maatschappelijk verkeer, in beroepspraktijk, in vervolgopleidingen en in wetenschappelijke disciplines. Wellicht is er geen enkel ander deelgebied van de wiskunde dat zo veelvuldig wordt toegepast als de statistiek: het verzamelen, ordenen en vergelijken van datasets uit experimenten, en het trekken van conclusies daaruit.

Dit uitgangspunt legitimeert het onderwerp statistiek in de programma's van wiskunde A van havo en vwo en van wiskunde C van vwo. Omdat het hier leerlingen betreft die niet bètagericht zijn, kan de wiskundige en kanstheoretische onderbouwing van de statistische technieken voor moeilijkheden zorgen. Hoe dan toch ervoor te zorgen dat deze leerlingen statistiek op een verstandige manier kunnen gebruiken en de resultaten van statistisch onderzoek zinvol kunnen interpreteren? De antwoorden op deze vragen zijn vooral ingewikkeld als op basis van steekproefgegevens uitspraken worden gedaan die de steekproef overstijgen en de hele populatie betreffen. Deze problematiek leidt tot de volgende twee centrale vragen van dit katern.

1. *Hoe kunnen leerlingen van wiskunde A en C met inzicht de stap zetten van het beschrijven van steekproefgegevens naar het trekken van verantwoorde conclusies over de onderliggende populatie?*
2. *Welke instructiestrategieën kunnen deze ontwikkeling ondersteunen?*

In de volgende paragraaf wordt deze kwestie eerst in brede zin verkend door het vakgebied in kaart te brengen en een aantal valkuilen te beschrijven. In paragraaf 4 komt de zogeheten empirische cyclus aan de orde. In paragraaf 5 functioneert deze cyclus als leidraad voor het ontwerpen van onderwijs over hypothesetoetsing, als specifiek geval van de extrapolatie van steekproef naar populatie. Het katern eindigt met conclusies die concrete handvatten bieden voor het statistiekonderwijs bij wiskunde A en wiskunde C.

3. Probleemverkenning

Ter verkenning van het hierboven aangeduide probleem wordt hieronder eerste het vakgebied van de statistiek globaal beschreven. Dan gaan we kort in op de benadering van de exploratieve data-analyse. Ten slotte inventariseren we een aantal mogelijke valkuilen die in het statistisch denken veelal de kop kunnen opsteken.

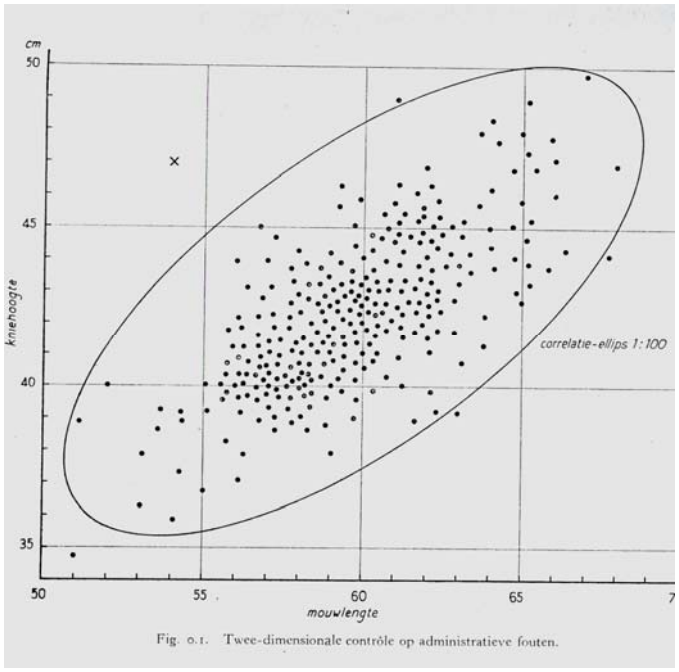
3.1 Globaal overzicht van statistiek als vakgebied

Alle statistiek begint met het verzamelen van gegevens, ook wel data genoemd. Moore (1990) spreekt over die data als 'getallen met een context' en betoogt dat statistiek wel een wiskundige discipline is maar geen tak van de wiskunde. Statistiek heeft zich immers ontwikkeld tot een eigen discipline met karakteristieke denkmethoden. Gal en Garfield (1997) beargumenteren dat in de statistiek de context de operaties motiveert en daarmee de bron is voor de betekenis en de interpretatie van de resultaten. De onzekerheid, die besloten ligt in de beperkte nauwkeurigheid van de data, onderscheidt statistisch onderzoek van de meer exacte signatuur van andere wiskundige vakgebieden. Zo hebben veel statistische problemen niet een eenduidige wiskundige oplossing. In plaats daarvan beginnen ze met een vraag en sluiten ze af met een conclusie, die wordt ondersteund door feiten en aannames maar die tevens onzekerheid bevat, bijvoorbeeld weergegeven door middel van een significantieniveau of onbetrouwbaarheid. Die conclusie moet worden geëvalueerd in termen van de kwaliteit van de redenering, de adequaatheid van de toegepaste methoden en de aard van de data en van de bewijsvoering die zijn gebruikt.

Het vakgebied van de kansrekening en statistiek (tegenwoordig ook 'stochastiek' genoemd) wordt veelal verdeeld in drie deelgebieden: de beschrijvende statistiek, de kansrekening en de verklarende statistiek (zie bijvoorbeeld het klassieke leerboek van Freedman et al., 2006). In de *beschrijvende statistiek* gaat het erom gegevens uit metingen, experimenten of uit een steekproef op een overzichtelijke manier samen te vatten. De frequentieverdeling kan worden weergegeven in tabellen, of kan worden gevisualiseerd met behulp van diagrammen zoals boxplot, polygoon, staafdiagram of histogram. Originele historische voorbeelden van grafische weergaven van data zijn te vinden in Neeleman & Verhage (1999). Ook het uiterst lezenswaardige onderzoek van Sittig en Freudenthal (1951, zie figuur 2) naar een nieuw systeem van confectiematen bevat mooie datavisualisaties. De keuze van de grafische representatie wordt mede bepaald door het meetniveau van de variabele: gaat het om 'echte' getallen zoals een snelheid of een reactietijd, dan wel om waarnemingen zoals kleur haar of geslacht, die zich niet in zinvol in een numerieke waarde laten vangen? Voor een nadere toelichting op het begrip 'meetniveau'zie bijlage 1. Karakteristieken van frequentieverdeling zijn centrum, uitgedrukt in centrummaten zoals modus, mediaan en gemiddelde, en spreiding, uitgedrukt in bijvoorbeeld tussenkwartielafstand of standaardafwijking.

In de *kansrekening* wordt het uitgangspunt niet gevormd door daadwerkelijk gevonden gegevens, maar door theoretische kansmodellen. In theorie weten we bijvoorbeeld dat de kans om met een dobbelsteen 5 te gooien gelijk is aan 1/6. Kansverdelingen, de theoretische tegenhangers van frequentieverdelingen, kunnen ook worden gevisualiseerd, bijvoorbeeld met een kansenhistogram. Daarnaast heeft ook een kansverdeling karakteristieken als centrum (de verwachting) en spreiding (de standaarddeviatie). Er zijn verschillende 'families' van kansverdelingen die veel worden gebruikt als theoretisch model voor een populatie of voor het trekken van een steekproef uit een populatie. Van een discrete kansverdeling is sprake als de uitkomsten bestaan uit 'losse' punten, zoals bijvoorbeeld het aantal meisjes in gezinnen met drie kinderen. Bekende discrete kansverdelingen zijn de binomiale kansverdeling en de

Poissonverdeling. Bij een continue kansverdeling bestaan de mogelijke uitkomsten in theorie uit een interval van getallen. Denk bijvoorbeeld aan lichaamslengte of reactietijd. De bekendste continue kansverdeling is de normale verdeling. Een veelgebruikt boek op het gebied van kansrekening is Feller (1968). Dat kansen niet altijd eenvoudig te interpreteren zijn, wordt helder beschreven in Van Lambalgen & Meester (2005).



Figuur 2 Spreidingsdiagram uit Sittig & Freudenthal (1951)

In de *verklarende statistiek*, ook wel wiskundige of mathematische statistiek genoemd, vormen net als bij de beschrijvende statistiek gegevens uit een steekproef of experiment het uitgangspunt. Het doel is nu echter om die te gebruiken voor uitspraken of voorspellingen die verder gaan dan de steekproef zelf, namelijk die de hele populatie betreffen waaruit de steekproef afkomstig is. Om dit te kunnen doen en vooral om iets te kunnen zeggen over de kwaliteit van de voorspellingen, komen de kansverdelingen uit de kansrekening van pas. Hoe betrouwbaar is mijn voorspelling, met welke onzekerheidsmarges moet ik rekening houden? De antwoorden op dergelijke vragen volgen uit de kansrekening. De voorspelling die in de verklarende statistiek worden gemaakt hebben veelal een van de drie volgende gedaanten:

- Een schatting van een karakteristiek van de populatieverdeling met een getal (puntschatting);
- Een schatting van de marges van een karakteristiek van de populatieverdeling (intervalschatting of betrouwbaarheidsinterval);
- Een uitspraak over een karakteristiek van een populatieverdeling (toetsen van hypotheses).

3.2 Wat is moeilijk aan statistiek?

Als we de hierboven geschetste driedeling van het vakgebied nader in ogenschouw nemen, vormt vooral de verklarende statistiek een probleem voor leerlingen wiskunde A of C en studenten van de gammawetenschappen of biomedische wetenschappen. Daarvoor zijn twee oorzaken aan te wijzen. Ten eerste is het fundament van de kansrekening en de kansverdelingen in veel gevallen zwak, waardoor de verklarende statistiek wordt gebouwd op

een drassige ondergrond. Ten tweede vraagt de overgang van beschrijvende naar verklarende statistiek om een aantal samenhangende maar subtiele blikwisselingen:

- van het beschrijven van praktische, concrete gegevens uit een steekproef naar het voorspellen of schatten van kenmerken van een onderliggende populatie, die vaak slechts theoretisch en abstract beschreven is;
- van een frequentieverdeling van steekproefgegevens, weergegeven door bijvoorbeeld een histogram, naar een kansverdeling van een toevalsvariabele, weergegeven door de formule of de grafiek van de kans(dichtheids-)functie;
- van de zekerheid van de waarnemingen die in het verleden zijn gedaan naar de onzekerheid van de voorspellingen of schattingen van ontwikkelingen in de toekomst.

Los van de specifieke probleemstelling van dit katern is in de oriëntatie aangegeven dat statistiek in het algemeen een vakgebied is met vele valkuilen. Zulke valkuilen liggen aan de basis van veel voorkomende misleidende conclusies in publicaties over verzamelde data, bijvoorbeeld in de media. Dergelijke misleidingen zijn in sommige gevallen eenvoudig als zodanig te ontmaskeren. Hieronder bespreken we een vijftal bekende misvattingen en fouten.

Valkuil 1: Is er wel samenhang tussen twee variabelen?
Een eerste valkuil is dat toevallige samenhang tussen twee variabelen in een dataset te snel wordt geïnterpreteerd als betekenisvolle samenhang. In veel gevallen waarin samenhang wordt gesuggereerd is de vraag of er inderdaad wel een werkelijke samenhang is. Vervolgens is het dan de vraag of je die samenhang uit de context kunt begrijpen en er een zinvolle betekenis aan kunt geven.

Voorbeelden

- Het verkeer in Drenthe is veiliger dan Utrecht want in 2007 vielen er in Utrecht tweemaal zoveel dodelijke slachtoffers in het verkeer dan in Drenthe.
- Vrouwen rijden voorzichtiger dan mannen, want zij zijn bij 25% van de ernstige verkeersongevallen betrokken en mannen bij 75%.
- Schlesinger onderzocht in 1950 of zich onder de presidenten die eerst gouverneur waren geweest naar verhouding meer een ‘goede’ presidenten bevonden. Na raadpleging van 50 historici ontstond onderstaande tabel (Hemelrijk, 1972). Is het gouverneurschap een goede leerschool voor het presidentschap?

	‘goede’ president	‘slechte’ president
eerst gouverneur	9	4
niet eerst gouverneur	6	10

- In het databestand van de Nationale Doorsnee 2000 staat een tabel van het aantal guldens per week dat leerlingen indertijd verdienden met een baan(tje) (zie paragraaf 5.1 en http://www.fi.uu.nl/archief/nationaledoorsnee/). De vraag is of jongens vaker een baantje hebben dan meisjes.

	Niet werken	Werken	Totaal
Jongens	593	389	982
Meisjes	694	324	1018
Totaal	1287	713	2000

Aanbeveling
Controleer statistische redeneringen uit de media met een 2 x 2 kruistabel. Ook kun je de sterkte van de samenhang inschatten door een puntenwolk te maken en door verschillende maten van samenhang met de rekenmachine of software te berekenen. Voor de interpretatie van de betekenis van een samenhang moet je terug naar de variabelen in de context.

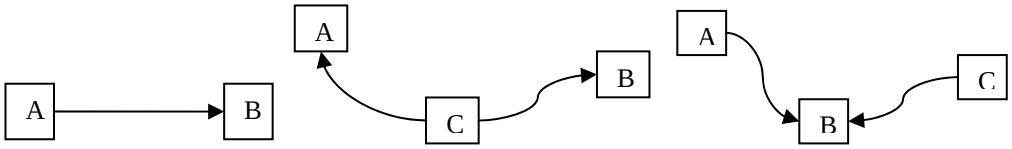
Opgave 2
Ga in de bovenstaande voorbeelden na of er sprake is van samenhang. Gebruik bijvoorbeeld een handige manier van percenteren.

Valkuil 2: Samenhang is nog geen causaliteit
De tweede valkuil is dat uit samenhang ten onrechte causaliteit wordt afgeleid. Als er betekenisvolle samenhang tussen twee variabelen bestaat, is de vraag of er ook sprake is van een oorzaak-gevolg relatie. De verwarring tussen samenhang en causaliteit is wijd verbreid en wekelijks te signaleren in kranten en politieke of maatschappelijke beschouwingen. Een veelvoorkomende statistische valkuil dus! Het onderzoek naar de relatie tussen roken en longkanker draaide bijvoorbeeld decennia lang om de vraag of hier sprake was van een oorzaak-gevolg relatie (Zie het internationaal bekende leerboek statistiek van Freedman et al., 2006). De tabaksindustrie heeft jarenlang andere verklarende variabelen dan het roken aangevoerd, zoals erfelijke aanleg die ook het roken stimuleert, geslacht en leeftijd.

Voorbeelden
- In 1987 promoveerde een huisarts op de aanwezigheid van longkanker onder zijn patiënten. Hij constateerde een hoge positieve correlatie tussen longkanker en het aantal stofdeeltjes in de longen bij patiënten die er in huis een volière op na hielden. Zijn dringend advies was de volières de deur uit te doen. Op grond van dezelfde data bleek het rookgedrag van die volièrehouders het hogere percentage gevallen van longkanker te kunnen verklaren.
- Voordat er een vaccin tegen polio was ontwikkeld, constateerde men in de Verenigde Staten een hoge correlatie tussen het aantal poliogevallen en de hoeveelheden verkochte frisdranken. Dus …
  Polio-epidemieën kwamen echter vooral voor in warme zomers.
- De correlatie tussen (te) hoge bloeddruk en de hoogte van het inkomen is sterk positief. Dus … Wellicht houdt de hoogte van het inkomen verband met de verantwoordelijkheden en werkdruk en zijn dat de oorzaken van stress.
- Uit een Amerikaans onderzoek bleek dat de drop-out van high schools sterk negatief correleerde met het salaris van de leraren. Dus …
  De leraren aan (elitaire) particuliere scholen werden echter het best betaald.
- De levensverwachting in de armere wijken in Nederland is lager dan in de meer welvarende wijken. Dus ….
  Hier spelen mogelijk ook andere factoren een rol, zoals eetgewoonten en levensstijl.

Aanbeveling
Vertaal samenhang niet vanzelfsprekend in causaliteit, maar onderzoek de samenhang tussen variabelen.

Opgave 3
Gebruik de onderstaande visualiseringen om de situaties in de genoemde voorbeelden in kaart te brengen.



Valkuil 3: De steekproef deugt niet
De derde valkuil betreft het trekken van conclusies op basis van gegevens uit een ongeschikte steekproef. Het steekproefonderzoek (onder andere het opinieonderzoek) kent een lange historie van mislukkingen, die onder meer veroorzaakt worden door de manier waarop de steekproef wordt getrokken. Een belangrijke valkuil die zich vaak aan het zicht van de waarnemer onttrekt!
Het eenvoudigste concept is dat van de *aselecte steekproef*, dat is de steekproef waarbij elk lid van de populatie via een perfecte loting dezelfde kans heeft om in de steekproef terecht te komen. Leerlingen kunnen zelf in hun eigen leefomgeving een steekproef bedenken, die aan dit criterium lijkt te voldoen. De discussie over die voorbeelden zal het begrip van een aselecte steekproef verhelderen. Het vaasmodel past goed bij de aselecte steekproef. Een intuïtieve kijk op de rol van het toeval in de overgang van populatie naar steekproef kunnen leerlingen ook ontwikkelen door met een computerprogramma het trekken van een aselecte steekproef te simuleren.
Een lastiger concept is dat van de *representatieve* steekproef. Lastiger, want dan moet je allereerst de vraag beantwoorden in welk opzicht die steekproef dan wel representatief moet zijn. Welke variabelen in de populatie kunnen van invloed zijn op de variabele waar het ons om gaat? En kennen we de verdeling van die variabelen in de populatie? Van een grote aselecte steekproef, groot vergeleken met de populatie, mag je hopen dat de relevante variabelen die van invloed kunnen zijn representatief in de steekproef voorkomen. Bij een landelijk onderzoek kun je daarnaast eerst een verdeling maken op bijvoorbeeld grote steden – kleinere steden – platteland, en op man – vrouw, en op leeftijdscategorieën, en op inkomen. Daarna neem je in elke deelgroep, bijvoorbeeld, platteland en vrouw en 60+ en drie keer een modaal inkomen, een aselecte steekproef. Uiteraard nemen de kosten van het onderzoek toe naarmate de steekproef groter is en/of meer combinaties van andere relevante variabelen noodzakelijk lijken.

Voorbeelden
- Beroemd in dit verband is het opinieonderzoek dat Gallup in 1948 deed naar de vraag wie de volgende president van de Verenigde Staten zou worden: Truman of Dewey. Omdat Gallup alleen maar kiezers via de telefoon had benaderd, was de (zeer grote) steekproef vertekend: er zaten geen kiezers zonder telefoon in. Het armere deel van de bevolking had geen telefoon. Gallup voorspelde dat Dewey ruim zou winnen, het werd echter Truman. In de jaren daarna voldeed die aanpak goed toen vrijwel iedereen een vaste telefoonaansluiting had. Bij de verkiezing van Obama tot president in 2008 realiseerde men zich dat er nu bij die eenvoudige methode weer een systematische vertekening in de steekproef gaat optreden. De jongere kiezers hebben vaak alleen een mobiele telefoon.
- Een ander storend effect bij het veel voorkomende onderzoek met vragenlijsten is het bewust onjuist beantwoorden van de vragen door de respondenten. In de V.S. is bijvoorbeeld het Bradley-effect bij opiniepeilingen bekend. Bradley was in 1982

volgens de peilingen voor de verkiezing van gouverneur van Californië sterk favoriet, maar zijn opponent won royaal. Bradley was de eerste zwarte kandidaat voor die post en de ondervraagden wilden niet toegeven dat ze geen zwarte gouverneur wilden. Bij de verkiezing van Obama in 2008 was men ook bang voor het Bradley-effect, maar dat bleek bij de stemming niet te zijn opgetreden. Bekende vertekeningen treden ook op bij vragen over inkomen, sexueel gedrag en andere 'gevoelige' onderwerpen.

- In de Volkskrant van 30 mei 2009 schrijft Maarten Keulemans hoe onderzoeksbureau TNS / Nipo de opinie van Nederlanders over het elektronisch patiëntendossier onderzocht. Dat deed men via een online vragenlijst, die blijkt te zijn ingevuld door panels van 'beroepsvrijwilligers'. Dat zijn bedreven internetters, terwijl je juist bij mensen die niet zoveel op hebben met internet twijfels mag verwachten over digitale nieuwigheden als een digitaal patiëntendossier…
- Bij enquêtes, bijvoorbeeld telefonisch afgenomen enquêtes, treedt ook vaak vertekening op doordat beoogde te enquêteren personen niet thuis zijn. Dit probleem van de non-respons maakt het extra moeilijk om uit de gevens conclusies te trekken.

Aanbeveling
Sta stil bij de steekproef die de statistische data genereert, en maak heldere keuzes ten aanzien van representativiteit en aselectheid en bedenk hoe om te gaan met non-respons.

Opgave 4
Bedenk hoe je in de klas, bijvoorbeeld naar aanleiding van een eigen onderzoekje, de aselectheid en representativiteit van steekproeven kunt bespreken als onderdeel van het beoordelen van een statistisch onderzoek.

Valkuil 4: Het onwaarschijnlijke gebeurt doorlopend
De vierde valkuil van de statistiek is dat we ermee moeten leven dat er doorlopend onwaarschijnlijke gebeurtenissen plaatsvinden, die, als ze een opvallende gedaante hebben, soms tegen de intuïtie ingaan. Dit staat bekend als de Black Swan Theory (Taleb, 2007). Daar staat tegenover dat iets heel normaals, zoals bij 20 keer een geldstuk gooien 10 keer kop en 10 keer munt, een vrij kleine waarschijnlijkheid heeft, namelijk 0.18.

Voorbeeld
Op school vindt een lotto plaats. In de aula staat een bak met 41 ballen, genummerd 1 tot 41. Een eersteklasser trekt hieruit vijf ballen zonder terugleggen. De eerste bal: nummer 1. De tweede: nummer 2. De derde: nummer 3. Het wordt onrustig in de zaal. Dan komt nummer 4. Geloei stijgt op. Ten slotte wordt bal 5 getrokken, en bestormen boze leerlingen het podium. Wat is hier aan de hand? De leerlingen vinden het rijtje 1-2-3-4-5 uiterst onwaarschijnlijk. Als daarentegen 37-13-4-3-8 was getrokken, had niemand raar opgekeken. Het paradoxale is echter, dat zowel 1-2-3-4-5 als 37-13-4-3-8 zeer zeldzame trekkingen zijn, elk met een kans van

$$\frac{1}{41\cdot 40\cdot 39\cdot 38\cdot 37}\approx 1{,}1\cdot 10^{-8}$$

Aanbeveling
Laat leerlingen ervaren dat ogenschijnlijk zeldzame gebeurtenissen toch plaats vinden, en dat de kansen op voor de hand liggende gebeurtenissen soms niet zo groot zijn als je zou denken.

Opgave 5
Reken de aan het begin van deze valkuil genoemde 0.18 na en verklaar de uitdrukking $\frac{1}{41 \cdot 40 \cdot 39 \cdot 38 \cdot 37}$ in de regel hierboven.

Valkuil 5: Verwarring van verleden en heden, van feiten en mogelijkheden
De laatste valkuil van de statistiek betreft de rol van de tijd. In statistiek en kansrekening speelt de tijd niet zelden een rol. Van gebeurtenissen in het verleden weten we of ze zijn opgetreden of niet. Of minstens weten we dat het toeval daarin geen rol meer speelt. Achteraf, a posteriori, is alles bepaald en liggen de gegevens vast. Van tevoren, a priori, is er sprake van een toevalsexperiment waarvan we de uitkomst nog niet kennen. De valkuil is dat men deze twee perspectieven met elkaar verwart.

Voorbeelden
- De uitspraak ‘de kans dat ik gisteren met een dobbelsteen een 5 gooide is 1/6’ onzin, zelfs al weet ik niet meer wat ik gisteren heb gegooid. Het was een 5 of niet, maar nu, achteraf, speelt het toeval daarin geen rol meer.
- Als iemand met pokeren onder de beker een dobbelsteen heeft gegooid, is de uitspraak ‘de kans dat hij een 5 heeft is 1/6’ niet terecht, want de dobbelsteen ligt inmiddels stil en het toeval heeft zijn invloed verloren. Wel is dit voor de andere spelers, die nog niet weten wat er onder de beker ligt, een werkbaar model op basis waarvan ze hun strategie kunnen bepalen.
- Bij een steekproef van 1000 leerlingen uit het hele land wordt onderdeel 1a van het CE wiskunde nagekeken. De maximale score voor het onderdeel is 4 punten, en op basis van de steekproef wordt een 95%-betrouwbaarheidsinterval gemaakt van het gemiddelde aantal scorepunten voor dit onderdeel van alle leerlingen in het land. Dit interval is <2,3 ; 2,6>. Wie nu zegt dat het landelijk gemiddelde met kans 95% in dit interval ligt, heeft ongelijk. Vooraf leidt de statistische methode waarmee het interval wordt geconstrueerd in 95% van de gevallen tot een interval dat het landelijk gemiddelde bevat. Echter, of we ons nu, bij dit concrete interval, in het 95%-geval bevinden, of in de 5% van de gevallen dat ons interval mis is, dat weten we niet.

Aanbeveling
Maak leerlingen bewust van de tijdsfactor. Achteraf, a posteriori, spreken we van feiten, van frequenties. Vooraf, a priori, kennen we slechts mogelijkheden en kansen.

Opgave 6
Maak als volgt een plaatje bij de situatie van de scores van een onderdeel, zeg 1a, van het CE wiskunde. Zet in een assenstelsel op de horizontale as het landelijke gemiddelde van de score op dit onderdeel en op de verticale as dat van de steekproef van 1000 leerlingen. Geef met een horizontaal lijnstuk voor elke waarde van het steekproefgemiddelde een intervalschatting voor het landelijk gemiddelde aan.

3.3 Exploratieve data-analyse
Door de technologische ontwikkelingen is het verzamelen van gegevens in onze samenleving eenvoudig en zelfs vanzelfsprekend geworden. Door bij de kassa de bonuskaarten te scannen, beschikt Albert Heijn bijvoorbeeld over gedetailleerde gegevens over het aankoopgedrag van klanten. Iets vergelijkbaars geldt voor webwinkels zoals Amazon of Bol, die het klikgedrag van hun bezoekers vastleggen. De aldus verkregen gegevens zijn vanuit het oogpunt van

marketing zeer interessant. Door zogeheten ‘data-mining’-technieken worden klantprofielen opgesteld, waarop marketingstrategieën kunnen worden afgestemd.
ICT-hulpmiddelen voor data-analyse maken de verwerking van dergelijke grote aantallen gegevens mogelijk. In de jaren zeventig is mede onder invloed hiervan de zogeheten exploratieve data-analyse ontstaan. Het uitgangspunt van exploratieve data-analyse (EDA) is dat realistische data het uitgangspunt vormen van de analyse en dat de analyse erop gericht is de data ‘te laten spreken’.

Die analyse is een ‘data-driven’ ontdekkingsreis door de gegevens. Tukey, de grondlegger van de exploratorieve data-analyse en bekend als bedenker van de boxplot, gebruikt de metafoor van de detective (Tukey, 1977). Als een detective doorzoekt de analist de gegevens op sporen in de vorm van structuren en patronen, ontdekt hij bijzonderheden, interpreteert hij die en ontwikkelt hij hypothesen. Die patronen kunnen onverwacht zijn; bij aanvang van de data-analyse liggen de hypothesen nog niet vast.

Om data zinvol en nuttig te kunnen gebruiken, worden ze samengevat en overzichtelijk weergeven. Opsplitsing, filtering en transformatie van ruwe data brengt onderliggende en niet direct zichtbare informatie naar boven. Reken- en tekentechnieken zijn dank zij de beschikbare ICT minder tijdrovend dan vroeger. Relaties tussen variabelen worden onderzocht, bijvoorbeeld met puntenwolken (spreidingsdiagrammen) en gekwantificeerd met correlatietechnieken.

Exploratieve data-analyse is geen eindstations, maar vormt de aanleiding voor de zogeheten confirmatieve data-analyse, waarin hypothesen die uit de exploratie naar voren zijn gekomen met nieuwe gegevens worden getoetst (Tukey, 1977). De EDA-aanpak staat de klassieke benadering dus niet in de weg; eerder vormt EDA een voortraject daarheen.
In het statistiekonderwijs in het voortgezet onderwijs wordt steeds meer gebruik gemaakt van echte data, die via internet bereikbaar zijn en geanalyseerd kunnen worden met software zoals Excel of VU-Stat (Konold & Van de Giessen, 2008). Het zwaartepunt ligt veelal bij de klassieke statistiek, waarin het steekproef-populatiemodel centraal staat. De vraag is dan ook of EDA kan leiden tot een vruchtbare aanpak van het statistiekonderwijs voor wiskunde A en wiskunde C, die intuïtiever en concreter is, en waarin de barrière van de kansrekening gedeeltelijk kan worden vermeden. Recente artikelen (Van Streun & Van de Giessen, 2007a, 2007b) pleiten hiervoor. Ook de concept-examenprogramma’s wiskunde A en wiskunde C voor 2014 ademen enigszins de geest van exploratieve data-analyse. Bijlage 3 bevat de globale eindtermen Statistiek en kansrekenen van de concept-examenprogramma’s wiskunde A voor havo en vwo en C voor vwo per 2014 (cTWO, 2009). Daarin vormen data en datavisualisatie het vertrekpunt en komt de verklarende statistiek voor als laatste eindterm bij vwo; bij havo blijft de verklarende statistiek op de achtergrond.

4. Wat weten we al?

Wat is er zoal bekend uit de onderzoeksliteratuur over het onderwijs in de statistiek? In 4.1 bespreken we eerst het beoogde inzicht in statistisch onderzoek in het algemeen. Vervolgens komt in 4.2 de stap van het beschrijven van steekproefgegevens naar het trekken van conclusies over de onderliggende populatie in een voorbeeld aan de orde.

4.1 De statistische cyclus

Statistici en statistiekdidactici pleiten ervoor dat leerlingen en studenten in het statistiekonderwijs de gelegenheid krijgen om alle fasen van het statistisch onderzoeksproces te doorlopen, analoog aan de manier waarop dit in de statistische praktijk gebeurt (Wild & Pfannkuch, 1999). De opzet en uitvoering van zo'n onderzoek doet een beroep op eerder opgedane statistische kennis, of is aanleiding nieuwe kennis te verwerven.
Uit welke fasen bestaat zo´n statistisch onderzoeksproces dan wel? In 2005 heeft de *American Statistical Society* het rapport *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) uitgebracht (Franklin et al., 2005). Daarin worden in het statistisch probleemoplosproces vier fasen onderscheiden (Scheaffer & Tabor, 2008): het formuleren van vragen, het verzamelen van gegevens, het analyseren van gegevens en het interpreteren van de resultaten. Als de leidende principes voor statistiekonderwijs noemt het GAISE-rapport:

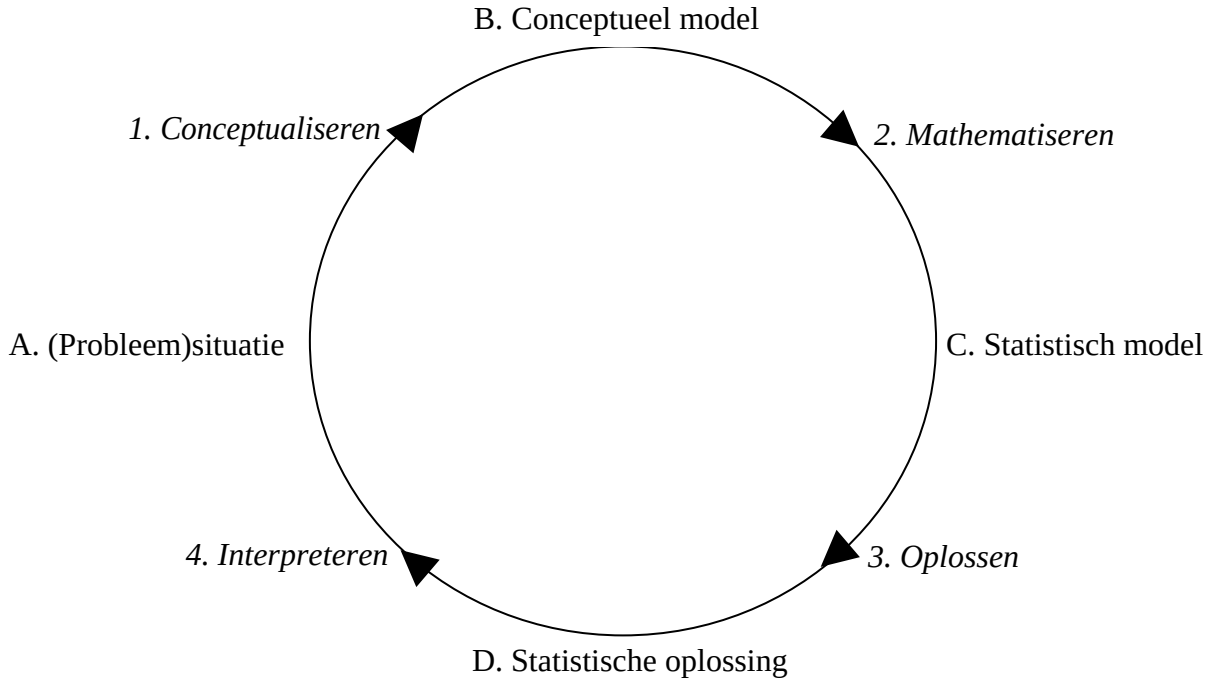
- Inzicht is belangrijker dan procedurele vaardigheden;
- Actief leren is de sleutel tot het ontwikkelen van inzicht;
- Zo mogelijk moeten gegevens uit de reële wereld in het statistiekonderwijs worden gebruikt;
- Het gebruik van geschikte software is essentieel voor de nadruk op begrippen boven berekeningen;
- Op alle onderwijsniveau's dienen alle vier de fasen van het statistische probleemoplossingsproces aan de orde te komen;
- In de onderzoeken die als voorbeelden dienen in het onderwijs moet het gebruik van statistiek essentieel zijn om de vraag te beantwoorden.

Moore (1997) stelt voor om de verschillende componenten van statistisch denken fasegewijs in een leerlijn statistiek te introduceren. Als mogelijke fasering ziet hij:

- Begin met grafische weergaven van de data en interpreteer wat je ziet.
- Zoek naar patronen en opvallende afwijkingen van die patronen.
- Kies op basis van dat onderzoek numerieke maten voor speciale aspecten.
- Als de verdeling voldoende regelmatig lijkt, zoek dan een mathematisch model.

In het rapport van de werkgroep SKACA van cTWO (2007) staat een probleemgeörienteerde aanpak van statistiek bij wiskunde A en C centraal. Aansprekende, bij de profielen passende problemen worden via de empirische onderzoekscyclus aangepakt vanuit reële data, zo mogelijk zelf op of bij de school verzameld (Nijdam, 2008). Het doorlopen van die empirische cyclus betekent, aldus SKACA, het probleem omschrijven (theoretiseren), een onderzoeksplan hierover opstellen, met operationalisering van variabelen, verwachtingen over de uitkomsten uitspreken, data verzamelen, data analyseren, conclusies trekken, evalueren en nieuwe problemen voorleggen. Deze empirische cyclus wordt met name in de gammawetenschappen heel veel gebruikt.
Op verschillende plaatsen wordt dus gepleit voor statistiekonderwijs dat de nadruk legt op het statistisch onderzoeksproces als geheel. Om dat handen en voeten te geven hebben we behoefte aan een model voor het uitvoeren van statistisch onderzoek. Een dergelijk model is de *statistische cyclus* (zie figuur 3). Deze is te beschouwen als een vereenvoudigde vorm van

de empirische cyclus, en is een aangepaste versie van de modelleercyclus uit het Elwierkatern Modelleren (Spandaw & Zwaneveld, 2008).



Figuur 3 De statistische cyclus

In figuur 3 staat deze statistische cyclus. De letters A, B, C en D staan voor de verschillende stadia van het oplossingsproces, en de cursieve *1, 2, 3* en *4* voor de acties die ons naar het volgende stadium moeten brengen. Hieronder worden de stadia en de acties kort toegelicht. Op het mathematiseren (2) en het statistisch model (C) wordt in 4.2 uitvoeriger ingegaan, omdat die de focus van het katern betreffen.

A (Probleem)situatie
De start van een statistisch onderzoek wordt bepaald door een probleemsituatie, waarin we iets te weten willen komen over een of meer aspecten, de *variabelen*, in een goed gedefinieerde verzameling, *de populatie*. Bijvoorbeeld gaat het erom op basis van gegevens een voorspelling te doen of een bepaalde beslissing af te wegen.

1 Conceptualiseren
Een fundamenteel probleem bij alle vormen van statistisch onderzoek is het zodanig formuleren van de probleemstelling en de probleemaanpak dat duidelijk wordt hoe onderscheid gemaakt kan worden tussen structurele en toevallige invloeden. Het is nodig de invloed van het toeval te kunnen herkennen, isoleren en kwantificeren. Een eerste stap is daarom het conceptualiseren van de probleemsituatie. Daartoe stel je je voor hoe de populatie er voor wat betreft de relevante variabele(n) uitziet en om wat voor soort variabele het gaat: welk meetniveau betreft het, is de variabele discreet of continu (zie bijlage 1)?
Vervolgens wordt een *steekproef* uit de populatie getrokken. Het doel daarbij is dat de verdeling van de variabele(n) in de steekproef de verdeling in de populatie goed weergeeft en bijvoorbeeld representatief en/of aselect is (zie valkuil 3 in paragraaf 3.2). In lijn met de EDA-benadering wordt de verdeling van de variabele in de steekproef gevisualiseerd (tabellen, histogrammen, boxplots, spreidingsdiagrammen) en worden steekproefgrootheden

berekend zoals centrummaten, spreidingsmaten en maten voor samenhang. Onderzoek van Bakker (2004) laat zien dat het begrip ‘verdeling’ een belangrijke rol speelt: leerlingen moeten een dataset als een geheel gaan beschouwen in plaats van als een aantal losse punten.

B Conceptueel model
Het conceptueel model bestaat dus uit een probleemstelling in termen van de te onderzoeken variabelen en de populatie, en uit gegevens uit een geschikte steekproef.

2 Mathematiseren
De tweede activiteit uit de statistische cyclus is die van het mathematiseren. Dat betekent ten eerste dat een veronderstelling gemaakt wordt over de kansverdeling van de variabele in de populatie. Dit is nodig om bij het oplossen de rol van het toeval te kunnen kwantificeren. Vervolgens wordt de vraagstelling geformuleerd in de vorm van een statistische uitspraak over een van de parameters van die kansverdeling. Denk aan uitspraken als ‘de verwachting van het verschil tussen de twee groepen is gelijk aan 0’ of ‘de kans op een bepaalde gebeurtenis ligt tussen 0.4 en 0.5’. Ten slotte wordt de toevalsvariabele die uit het trekken van de steekproef naar voren komt beschreven, en ook de bijbehorende kansverdeling, zoals die volgt uit de aanname over de populatieverdeling.

C Statistisch model
Het statistisch model bestaat dus uit een aanname over de kansverdeling van de variabele(n) in de populatie, een statistische uitspraak die moet worden onderzocht en de bepaling van de kansverdeling van de toevalsvariabele die bij het trekken van de steekproef naar voren komt.

3 Oplossen
Nu volgt de fase van het statistische rekenwerk. Op basis van de steekproefgegevens en de vastgestelde kansverdeling kan de statistische uitspraak worden beoordeeld. Dat leidt bijvoorbeeld tot marges bij schattingen van populatieparameters (*betrouwbaarheidsintervallen*) of tot bevestiging of weerlegging van statistische hypothesen. De betrouwbaarheid van het resultaat wordt in een kans uitgedrukt. In sommige gevallen wordt het – moeilijke of zelfs onmogelijke – statistisch rekenwerk vervangen door statistische simulatie.

D Statistische oplossing
Het resultaat is de statistische oplossing, die bijvoorbeeld bestaat uit een betrouwbaarheidsinterval of een al dan niet verworpen hypothese, waarbij je de betrouwbaarheid in een kans hebt uitgedrukt.

4 Interpreteren
Wat resteert, is de vertaling van de statistische oplossing naar de probleemsituatie: wat betekent het statistische resultaat in termen van het oorspronkelijke probleem? Het kan zijn dat deze vertaling eenvoudig is en dat het probleem daarmee is opgelost. Het is echter ook mogelijk dat op grond van de interpretatie een of meer acties van statistische cyclus moeten worden herhaald, of dat de conclusie aanleiding is tot nieuwe of aangepaste vragen.

Tot zover deze eerste bespreking van de statistische cyclus, waarvan varianten bestaan, en waarvoor geldt dat niet altijd elke stap bewust wordt doorlopen. Het hierboven genoemde GAISE model komt overeen met deze cyclus, behalve dat daarin het mathematiseren en het statistisch model niet apart worden genoemd. In de verklarende statistiek is het voor het trekken van conclusies over een populatie op basis van een steekproef nodig een statistisch

model te hebben van die populatie wat betreft de variabele onder beschouwing. Dat statistische model is de kansverdeling van die variabele in de populatie.
Bij het meten van een variabele in een steekproef treden twee soorten onzekerheden op. De eerste is dat maar een deel van de populatie wordt bekeken waardoor het mogelijk is dat die steekproef de populatie niet goed representeert. De tweede onzekerheid wordt veroorzaakt door het optreden van meetfouten. Met behulp van de kansrekening, uitgaande van het statistische model, kan de onzekerheid in de mate dat de steekproef de populatie representeert, worden gekwantificeerd.
In de praktijk neemt men aan dat men die kansverdeling kent. Soms suggereert het steekproefresultaat het statistisch model, soms wordt het statistisch model op eerder onderzoek of de literatuur gebaseerd. In de huidige onderwijspraktijk, en met name in de eindexamenopgaven, is dat statistisch model gegeven.

4.2 Van steekproef naar populatie: mathematiseren en het statistisch model

In deze paragraaf doorlopen we de statistische cyclus aan de hand van een voorbeeld, ontleend aan Scheaffer & Tabor (2008). De nadruk daarbij ligt op de overgang van de beschrijving van de steekproef naar het mathematiseren en het opstellen van een statistisch model voor de populatie, dus op activiteit 2 en stadium C van de cyclus van figuur 3.

Probleemsituatie
Leerlingen van een *high school* in de Verenigde Staten willen onderzoeken hoe strikt hun ouders ten opzichte van hun kinderen zijn. Zij realiseren zich dat ‘strikt’ een onduidelijk omschreven begrip is.

Conceptualiseren
Na onderling overleg besluiten ze hun onderzoek te richten op de vraag of de leerlingen een eindtijd hebben waarop ze thuis moeten zijn na een avondje stappen. Van daaruit formuleren ze als vraag: ‘Welk percentage leerlingen van hun school heeft zo’n eindtijd?’ Daarna hebben de leerlingen aselect 100 mede-leerlingen uit de ongeveer 1200 van hun school geselecteerd en gevraagd of ze een eindtijd hebben.

	wel eindtijd	*geen eindtijd*	*totaal*
meisjes	33	35	68
jongens	20	12	32
totaal	53	47	100

Tabel 1 Steekproefgegevens over eindtijd uitgesplitst naar geslacht

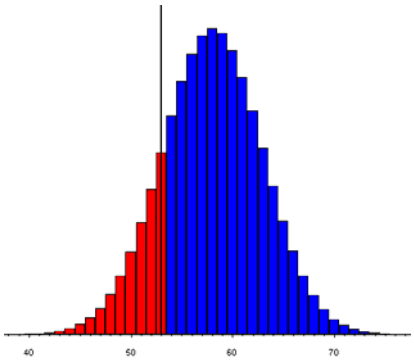
Conceptueel model
Merk op dat de populatie goed gedefinieerd is, namelijk de leerlingen van de school.
Het resultaat van de steekproef staat in tabel 1.
Kennelijk hebben 53 van de 100 leerlingen een eindtijd. De vraag is nu wat dit steekproefresultaat zegt over de populatie leerlingen van de school.

Mathematiseren
Er is nu een populatie (met de huidige ongeveer 1200 leerlingen) en een zekere, maar onbekende proportie p van leerlingen met een eindtijd. Zes jaar geleden is een dergelijk onderzoek ook uitgevoerd met als resultaat dat toen gold dat $p = 0{,}58$. We nemen aan dat dit destijds de populatieproportie was.
De vraag is of het steekproefresultaat voldoende argument is om te concluderen dat we nu te maken hebben met een andere populatie dan zes jaar geleden, namelijk een populatie met een proportie $p < 0{,}58$. Of is het zo dat nog steeds geldt $p = 0{,}58$, maar dat de steekproefproportie

van 0,53 door toeval tot stand is gekomen? Om deze laatste vraag te beantwoorden, is de kans op een steekproefresultaat van 53 bij een populatieproportie van 0,58 interessant. De onderzoekende leerlingen realiseren zich dat ze zich niet op de kans op een steekproefresultaat van precies 53 moeten richten, maar op steekproefresultaten van 53 of minder. De kans op precies 53 is immers vrijwel 0. De uiteindelijke vraag is dan volgende: is de kans op een steekproefresultaat van 53 of minder uit een populatie met $p = 0{,}58$ zo klein, dat het aannemelijk is dat de steekproef niet getrokken is uit een populatie met $p = 0{,}58$, maar uit een populatie met $p < 0{,}58$?

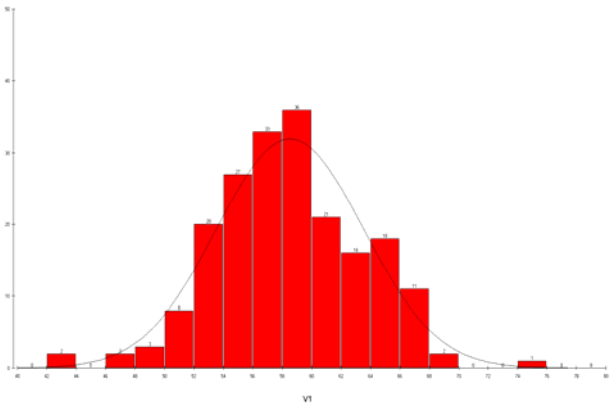
Statistisch model
Het statistische model ziet er dus als volgt uit. De (nul)hypothese is: de steekproef komt uit een populatie met $p = 0{,}58$. Daar staat als (alternatieve) hypothese tegenover dat de steekproef komt uit een populatie met een (onbekende) $p < 0{,}58$.
De leerlingen letten op de stochastische variabele X, het aantal leerlingen met een eindtijd in de aselecte steekproef van omvang 100. X heet de *toetsingsgrootheid*.
Als de hypothese $p = 0{,}58$ geldt, dan is X binomiaal verdeeld met $n = 100$ en $p = 0{,}58$.
Als de kans op een steekproefresultaat van 53 of minder klein is, zeg kleiner dan 0,05 (het *significantieniveau*), dan is het besluit dat er niet uit een populatie met $p = 0{,}58$ is getrokken, maar uit een populatie met een p die kleiner is dan 0,58.

Oplossen
Nu komt het wiskundig oplossen. Dat kan plaatsvinden op twee manieren: door berekening of door simulatie (zie Barbella et al., 1990). De eerste aanpak, die van de kansrekening, komt erop neer dat de linker overschrijvingskans $P(X \leq 53 \mid n = 100 \text{ en } p = 0{,}58)$ wordt berekend. De kansberekening kan aan ICT-gereedschap worden uitbesteed en dit levert een waarde van ongeveer 0,18 (zie figuur 4).



Figuur 4 Met VU-Stat berekende overschrijdingskans bij de gevonden waarde van 53

De tweede aanpak, die illustratief kan zijn voor leerlingen, is die van simulatie. In figuur 5 staat het resultaat van 200 simulaties uit een binomiale verdeling met parameters $n = 100$ en $p = 0{,}58$. Het steekproefgemiddelde ligt keurig bij 0,58 en de steekproefstandaarddeviatie bij 0,05, zoals ook theoretisch mocht worden verwacht. We zien dat 35 van de 200 simulaties 53 of minder successen opleveren. Dat leidt tot een geschatte overschrijdingskans van 0,18 (afgerond).
We merken nog op dat voor havoleerlingen de aanpak met simulatie de enige mogelijke is, want voor hen behoort kansrekening niet tot het programma.



Figuur 5 Met VU-Stat gesimuleerde frequentieverdeling

Statistische oplossing
De statistische oplossing is een kans van 0,18, die groter is dan het gestelde significantieniveau van 0,05. Dat betekent dat de nulhypothese wordt geaccepteerd.

Interpreteren
Ten slotte interpreteren de leerlingen dit statistische resultaat. De kans op het gevonden steekproefresultaat van 53 is dermate groot dat die best aan toeval is te wijten en er onvoldoende argumenten zijn om aan te nemen dat de populatie van nu afwijkt van die met $p = 0{,}58$. Het ouderlijk beleid wat betreft het stellen van een eindtijd lijkt dus niet veranderd te zijn.

Opgave 7
Verklaar de standaarddeviatie van 0.05 die onder het kopje mathematiseren als ‘theoretisch verwacht’ wordt beschreven.

Opgave 8
Worden jongens en meisjes door hun ouders verschillend behandeld qua eindtijd?

5. Ontwerpen

In het voorgaande is de statistische cyclus besproken en zijn twee suggesties voor onderwijs gedaan. De eerste was het idee dat leerlingen om gevoel voor statistisch onderzoek te krijgen ook een keer de hele cyclus moeten doorlopen. De tweede was de constatering dat met name de overgang van het beschrijven van een steekproef naar het trekken van conclusies over een populatie moeilijk is. In de cyclus zit dit probleem vooral bij activiteit 2 en stadium C. De vraag die in deze paragraaf centraal staat, is nu: hoe ontwerp je statistiekonderwijs dat deze twee ideeën recht doet? We beantwoorden deze vraag in paragraaf 5.1 aan de hand van een uitgebreid voorbeeld, waarin de statistische cyclus de leidraad vormt en waarbij elke stap ook vanuit didactisch perspectief wordt bekeken op een manier die de voorbeeldsituatie overstijgt. Dit leidt tot een aantal handvatten voor het te ontwerpen onderwijs.

5.1 De lichaamslengte van meisjes in Zeeland

In deze paragraaf is de probleemstelling verzonnen, maar gebruiken we echte data, namelijk die van de Nationale Doorsnee. Op 10 oktober 2000 is in Nederland onder deze naam een grote enquête gehouden onder leerlingen van klas 1 en 2 van het voortgezet onderwijs. De leerlingen kregen vragen voorgelegd over lichaamslengte, tijdsbesteding, zakgeld, enzovoorts. Deze gegevens zijn per e-mail naar het Centraal Bureau voor de Statistiek gestuurd, dat ’s avonds de resultaten van 50071 leerlingen in huis had! Figuur 6 geeft een indruk van de dataset zoals die eruit ziet in het computerprogramma VU-Stat. Elke rij stelt een leerling voor, een *case* of *record*. Elke kolom staat voor een kenmerk, een *variabele*. De gegevens van de ‘Nationale Doorsnee’ zijn te vinden op http://www.fi.uu.nl/archief/nationaledoorsnee/ .

LEERJAAR	GESLACHT	LEEFTIJD	LENGTE	SPORT	TV	COMPUTER	VAK	ZAKGELD	WERK	PROVINC	ONTBIJT
1	Meisje	12	160	1	4	6	Wiskunde	5	0	23	2
1	Meisje	11	162	3	4	4	Informatiekunde	3	0	23	3
1	Meisje	12	168	1	20	5	Ander vak	25	0	23	2
1	Jongen	13	149	0	6	2	Wiskunde	10	0	23	1
1	Meisje	12	170	3	30	1	Techniek	5	0	23	2
1	Meisje	11	164	2	3	0	Techniek	3	0	23	2
1	Jongen	12	163	4	21	14	Engels	5	0	23	2
1	Meisje	13	169	6	5	1	Biologie	7	0	23	9
1	Meisje	12	170	4	23	2	Geschiedenis	20	0	23	0

Figuur 6 Een blik op de dataset van de Nationale Doorsnee

Probleemsituatie

De dienst voor Jeugdgezondheidszorg in Zeeland heeft de indruk dat de meisjes uit Zeeland langer zijn dan die van de rest van Nederland. Op het moment dat men hoort van de Nationale Doorsnee ontstaat het idee om het initiatief te gebruiken voor het onderzoeken van dit vermoeden. Men schakelt een onderzoeksbureau in om na te gaan of de gegevens van de Nationale Doorsnee dit vermoeden kunnen bevestigen. Gelukkig waren de mensen van de Nationale Doorsnee al van plan ook gegevens over de lichaamslengte van de betrokken leerlingen te verzamelen.

Didactisch perspectief

Vanuit het perspectief van het onderwijsontwerp is het belangrijk dat de probleemsituatie herkenbaar en aansprekend is voor leerlingen, niet gekunsteld en niet te veel specifieke domeinkennis (in dit voorbeeld: van provinciale verschillen in lichaamskenmerken) vereist. De context van de Nationale Doorsnee is goed bruikbaar,

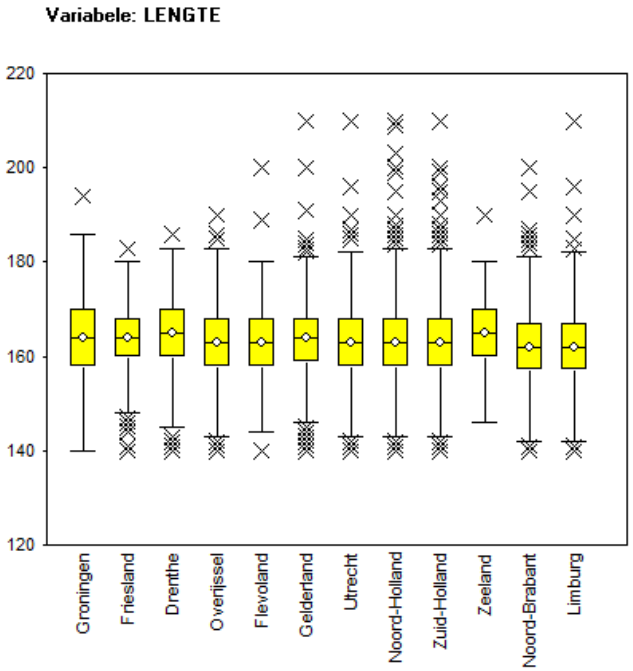
omdat de variabelen uit de leefwereld van leerlingen komen en het bestand vrij beschikbaar is en een groot aantal gegevens bevat.

Conceptualiseren

In overleg met de dienst voor Jeugdgezondheidszorg in Zeeland besluit het onderzoeksbureau na te gaan of de gegevens uit de Nationale Doorsnee aselect zijn. Dat blijkt inderdaad het geval te zijn. Exploratieve data-analyse leidt tot een tabel waarin de lengtes van de meisjes worden uitgesplitst per provincie. De gegevens worden gevisualiseerd in boxplots (zie figuur 7).

LENGTE

PROVINCIE	Groningen	Friesland	Drenthe	Overijssel	Flevoland	Gelderland	Utrecht	Noord-Holland	Zuid-Holland	Zeeland	Noord-Brabant	Limburg
Aantal waarnemingen	504	321	976	2014	635	3220	2687	4356	5442	83	3949	1412
Gemiddelde	163,8	163,3	164,6	163,0	162,7	163,3	162,7	163,1	162,9	165,0	161,9	161,9
Mediaan	164,0	164	165,0	163,0	163	164,0	163	163,0	163,0	165	162	162,0
Modus	160	165	165	160	160	165	165	160	165	160	160	160
Minimum	140	140	140	140	140	140	140	140	140	146	140	140
Maximum	194	183	186	190	200	210	210	210	210	190	200	210
SD	7,86	6,99	7,51	7,48	7,67	7,30	7,48	7,58	7,48	6,47	7,40	7,86



Figuur 7 Lengtes van meisjes in DND weergegeven in VU-Stat, uitgesplitst per provincie

	Gemiddelde lengte (cm)	*Aantal*
Meisjes uit Zeeland	165,0	83
Meisjes uit de rest van Nederland	162,8	25516

Tabel 2 Lichaamslengte van meisjes in DND voor Zeeland en de rest van Nederland

Tabel 2 vat de gegevens samen. De 83 meisjes uit Zeeland zijn dus gemiddeld 2,2 cm langer dan de 25516 meisjes uit de andere provincies. De standaardafwijking van de Zeeuwse meisjes is gelijk aan 6,47 cm. De vraag is nu of wat deze resultaten zeggen over de

onderliggende populaties. Kan het verschil van 2,2 cm door het toeval bepaald zijn of gaat het hier om een structureel verschil tussen Zeeuwse meisjes en meisjes uit andere provincies?

Didactisch perspectief
Vanuit didactisch perspectief is het van belang om met de leerlingen helderheid te scheppen over de variabelen en hun meetniveaus (zie bijlage 1). Ook de keuze van de manier waarop de steekproefgegevens worden gerepresenteerd vraagt om discussie met de klas. In dit geval voldoet een tabel, waar in andere gevallen een grafische weergave geschikter is.
Ten slotte is het verschil tussen steekproef en populatie van belang. In dit voorbeeld bestaat de populatie uit alle Zeeuwse meisjes die in klas 1 en 2 van het voortgezet onderwijs zitten. De vraag is of de populatie van Zeeland wezenlijk afwijkt van de populatie van meisjes uit klas 1 en 2 van de rest van het land wat betreft de variabele lichaamslengte en dat gaan we besluiten op basis van een steekproef van 83 Zeeuwse meisjes uit de Nationale Doorsnee. We beschouwen deze steekproef als aselect uit de populatie.
In deze fase van het onderwijs kan het goed zijn leerlingen gevoelig te maken voor de problemen van het nemen van steekproeven en de daarbij optredende variabiliteit. Laat hen bijvoorbeeld een aantal keren met een dobbelsteen gooien en men zal zien dat er ineens een onwaarschijnlijk lange reeks dezelfde waarnemingen optreedt, of dat zessen lang uitblijven of juist veel voorkomen. Eventueel kan het werkelijk uitvoeren van een toevalsexperiment worden vervangen door een simulatie met statistische software. Als leerlingen wordt gevraagd een serie van 10 toevalsgetallen van 0 tot 9 op te schrijven, zullen ze vermoedelijk nauwelijks een serie noteren waarin hetzelfde cijfer twee keer na elkaar voorkomt, terwijl dat in een simulatie waarschijnlijk wel zal gebeuren.
Ook kan het zinnig zijn leerlingen een onderzoek of experiment te laten uitvoeren, waarbij ze worden geconfronteerd met methodologische problemen of moeilijkheden met meetnauwkeurigheid, aselectheid en representativiteit.

Mathematiseren
Hoe gaan we deze situatie mathematiseren? Uitgangspunt is een steekproef (N=83) uit de populatie van Zeeuwse meisjes van klas 1 en 2. We willen weten of deze populatie qua lichaamslengte structureel afwijkt van die van de meisjes in de rest van Nederland. Voor de populatie van de meisjes in de rest van het land doen we net alsof de gegevens die uit steekproef van de Nationale Doorsnee komen, namelijk een gemiddelde lichaamslengte van 162,8, ook de werkelijke waarde van het populatiegemiddelde is. Uitgaande van deze aanname, hoe groot is dan de kans op een verschil van 2,2 cm of meer? Om daarover iets te kunnen zeggen, moeten we de kansverdeling van het steekproefgemiddelde weten. Eerst een afkorting:

X = de gemiddelde lichaamslengte van de meisjes uit de Zeeuwse steekproef.

Hoewel we al weten dat de gevonden waarde van X gelijk is aan 165,0, is X voorafgaand aan de meting een toevalsvariabele. Op grond daarvan gaan we beslissen en daarom heet X de toetsingsgrootheid.
Tijd voor nog een aanname: we gaan ervan uit dat X een normale verdeling heeft, en als schatting voor de standaarddeviatie van X gaan we uit van die van de steekproef, die we nog moeten delen door de wortel van de steekproefomvang:

$$\sigma = \frac{6,47}{\sqrt{83}} \approx 0,710$$

Omdat we van de Jeugdgezondheidsdienst aannemen dat de Zeeuwse meisjes in elk geval niet korter zijn dan die van de rest van Nederland, gaan we rechtseenzijdig toetsen.
Het statistische model ziet er dus als volgt uit:

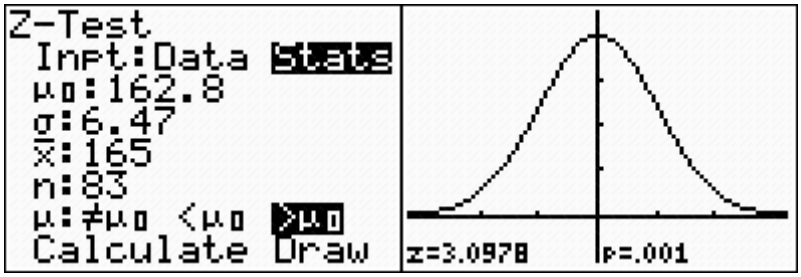
- De toetsingsgrootheid is het steekproefgemiddelde *X*
- *X* heeft een normale kansverdeling met $\sigma = 0,710$
- De nulhypothese is: de verwachting μ van *X* is gelijk aan 162,8.
- De alternatieve hypothese is: $\mu > 162,8$.

Als, aangenomen dat de nulhypothese waar is, de kans op de gevonden waarde of een nog afwijkender waarde van *X* klein is, dan besluit het onderzoeksbureau dat er niet uit een populatie met $\mu=162,8$ is getrokken, maar uit een populatie met $\mu > 162,8$. Als drempel voor wat we een kleine kans noemen kiezen we 0,05 (het *significantieniveau*).

Didactisch perspectief

Vanuit didactisch perspectief zijn er veel opmerkingen te maken bij deze fase van de statistische cyclus.
Ten eerste merken we op dat het belangrijk is dat leerlingen het type probleem kunnen identificeren. In dit geval gaat het bijvoorbeeld niet om een proportie, maar om een gemiddelde. Dat wijst op een Z-toets. Het is aan te raden om leerlingen te leren kiezen uit de beschikbare toetsen, ook al is het repertoire aan toetsen in het curriculum klein.
Ten tweede is het bij het bepalen van hypothesen van belang dat de nulhypothese zo wordt gekozen ‘dat ermee gerekend kan worden’ en dus enkelvoudig is: de nulhypothese mag maar één parameterwaarde bevatten. In praktijk komt dat erop neer dat de nulhypothese vaak inhoudt ‘dat er niets aan de hand is’. Dat kan contra-intuïtief zijn voor leerlingen, omdat we juist geïnteresseerd zijn in de mogelijkheid dat er wel iets bijzonders gebeurt, in dit geval dat de Zeeuwse meisjes wel degelijk langer zijn.
Een derde punt van aandacht is de schuivende verhouding tussen wat we al weten en wat we willen weten: hoewel we weten dat het gevonden steekproefgemiddelde 165,0 is, negeren we dat aanvankelijk, en hoewel we de waarde van μ niet kennen, doen we onder de nulhypothese alsof we weten dat die gelijk is aan 162,8. Ook dit kan vreemd overkomen op de leerling, en vraagt expliciete bespreking. Redeneringen in de stijl van ‘*als* de nulhypothese waar is, *dan* weten we dat de kans op dit resultaat of nog erger …’ verdienen nadrukkelijke aandacht.
Als vierde punt noemen we nog de onderliggende aannames: die van de normale verdeling van de toetsingsgrootheid, van de verwachting op basis van het steekproefgemiddelde van de rest van Nederland, en van de standaarddeviatie uit de Zeeuwse steekproef. Mag dat allemaal zomaar, al die aannames? Op schoolniveau hebben we helaas nauwelijks instrumenten om de redelijkheid van de aannames te onderzoeken. Het zal dus blijven bij een wat onbevredigend ‘dat mag voorlopig, maar in het echt moet je dat nagaan…’. Samengevat is het mathematiseren, samen met het oplossen hieronder, een complex samenspel van aannames en stappen in de statistische cyclus, waarin veel van perspectief gewisseld wordt tussen zekerheid en onzekerheid, tussen vooraf en achteraf, tussen frequentie en kans. Om dit proces overzichtelijk te maken voor leerlingen, kan een stappenplan gebruikt worden. Bijlage 2 bevat een voorbeeld van een dergelijk Plan van Aanpak.

Oplossen
Dan komt de fase van het wiskundig oplossen. De kans $p(X \geq 165,0 \mid \mu = 162,8, \sigma = 0,710)$ wordt met behulp van de normale kansverdeling berekend. Deze blijkt ongeveer 0,001 te zijn (figuur 8). De statistische oplossing is een kans van 0,001, die kleiner is dan het gestelde significantieniveau van 0,05. Dat betekent dat de nulhypothese wordt verworpen.



Figuur 8 Uitvoering van de toets met de grafische rekenmachine TI84

Didactisch perspectief
De didactische moeilijkheden van deze fase zijn vergelijkbaar met die van het mathematiseren. Bij het uitvoeren van de toets – of dit nu met pen-en-papier gebeurt, of met ICT-gereedschap zoals de grafische rekenmachine of VU-Stat – moet de leerling in staat zijn de verschillende gegevens op de juiste manier te interpreteren en te gebruiken in de toetsingsprocedure. Het stappenplan van bijlage 2 kan ook hierbij houvast bieden.
Verder kan het nuttig zijn om het oplossen ook via toevalssimulatie aan te pakken. Niet alleen omdat simulatie een methode is voor als de middelen voor de kansrekening te kort schieten, maar ook omdat leerlingen daarbij de rol van het toeval bij het uitvoeren van een toets beter ervaren. Er gaan dan ook stemmen op om toetsen door simulatie vooraf te laten gaan aan toetsen met kansrekening. Een voorbeeld van deze aanpak staat in Barbella et al. (1990). Statistische software biedt voor deze benadering veel mogelijkheden.

Interpreteren
De laatste activiteit van de statistische cyclus is die van het interpreteren. De kans op het gevonden steekproefresultaat van 165,0 of meer, uitgaande van de nulhypothese, is dermate klein dat die nauwelijks aan toeval te wijten kan zijn en er dus voldoende reden is om aan te nemen dat de Zeeuwse populatie afwijkt van die van de rest van Nederland. Het onderzoeksbureau zal dus aan de Jeugdgezondheidszorg rapporteren dat de gegevens uit de Nationale Doorsnee inderdaad steun geven voor het vermoeden dat de Zeeuwse meisjes uit klas 1 en 2 langer zijn dan die in de rest van Nederland.

Opgave 9
Zij bijlage 2. Zoek in een van de gangbare schoolboeken een opgave over hypothesetoetsing en analyseer deze volgens het bovenstaande stappenplan. Bedenk wat de didactische moeilijkheden zijn en hoe die aangepakt kunnen worden. Ga ook na of de methode een stappenplan aanbiedt en zo ja, hoe zich dat verhoudt tot de statistische cyclus.

Opgave 10
In het hierboven beschreven voorbeeld zegt een leerling: “Dus de kans dat H0 waar is, is gelijk aan 0,001”. Hoe reageer je daarop?

Opgave 11
Veronderstel dat je de vraag of de Zeeuwse meisjes significant gemiddeld langer zijn dan die in de rest van Nederland in je klas aanbiedt. Een leerling (of een groepje leerlingen) komt met het volgende statistische model.
Er zijn twee populaties: de Zeeuwse en de rest; uit elk is een steekproef getrokken waarvan de gegevens in de Nationale Doorsnee zijn opgeslagen. Bij de ene steekproef, de Zeeuwse, hoort de toevalsvariabele *X*, bij de andere steekprof, de rest van Nederland, hoort de toevalsvariabele *Y*. Als toetsingsgrootheid nemen *X* – *Y*. Deze toevalsvariabele is normaal verdeeld. De nulhypothese is dat beide populaties wat betreft de gemiddelde lichaamslengte gelijk zijn, dus dat beide steekproeven uit dezelfde populatie getrokken zijn. De verwachtingswaarde van *X* – *Y* is dus onder de nulhypothese gelijk aan 0. De standaarddeviatie van de toetsingsgrootheid is (onder de alleszins redelijke aanname dan *X* en *Y* onafhankelijk zijn) gelijk aan $\sqrt{\sigma_1^2 + \sigma_2^2}$ met σ_1 de standaarddeviatie van *X* en σ_2 de standaarddeviatie van *Y*. Deze laatste is met behulp van de standaarddeviaties van figuur 7 te berekenen. Als nu P($X - Y > 2{,}2 \mid H_0$) kleiner dan 0,05 is wordt de nulhypothese verworpen. Hoe reageer je hierop?

5.2 Van steekproef naar populatie: handvatten voor onderwijsontwerp

Uit het voorgaande zijn we ingegaan op de moeilijkheden bij de overgang van steekproef naar populatie. Daarbij hebben we ons geconcentreerd op het toetsen van hypothesen en is geconstateerd dat de grootste problemen zitten bij de stappen Mathematiseren en Oplossen. De moeilijkheden worden veroorzaakt door de soms contra-intuïtieve en impliciete blikwisselingen tussen vooraf en achteraf, tussen hypothesen en gegevens, tussen wat we willen weten en we al weten, populatie en steekproef, tussen kans en frequentie.

Bij het ontwerpen van onderwijs moet hiermee rekening worden gehouden. Uit het voorgaande vallen de volgende handvatten hiervoor te destilleren.

1. Aansprekende probleemsituatie
   Kies een aansprekende probleemsituatie, die voorstelbaar is en weinig specifieke voorkennis veronderstelt.
2. Variabiliteit in steekproeven
   Laat leerlingen de variabiliteit in steekproeven ervaren door toevalsexperimenten te doen, steekproeven te nemen of toevalssimulaties uit te voeren.
3. Populatie en variabelen
   Stel expliciet aan de orde wat de populatie is in een probleemsituatie en maak consequent onderscheid tussen populatie en steekproef. Stel vast welke variabelen een rol spelen, welk meetniveau ze hebben en hoe ze overzichtelijk kunnen worden weergegeven.
4. Een toets kiezen
   Bespreek met leerlingen het repertoire aan beschikbare toetsen, dat overigens beperkt is, en met name wanneer welke toets bruikbaar is.
5. Hypotheses opstellen
   Leer leerlingen dat de nulhypothese enkelvoudig moet zijn en dat dit in praktijk veelal neerkomt op een hypothese die stelt dat het verwachte effect *niet* optreedt.
6. Blikwisselingen
   Bespreek met de leerlingen de blikwisselingen die optreden tijdens het uitvoeren van een toets. De gevonden waarde van de toetsingsgrootheid weten we veelal, maar speelt

in eerste instantie geen rol. De parameterwaarde die we niet kennen, wordt juist bij de kansrekening bekend verondersteld, uitgaande van de nulhypothese. Doorloop de statistische cyclus vanuit het perspectief weten – niet weten, gegeven – gevraagd, steekproef – populatie. Besteed aandacht aan de 'verhaallijn' in het oplossingsproces en met name aan redeneringen van het type '*als …, dan…*'.

7. Simuleren en berekenen
   Laat zien dat in de oplossingsfase gebruik kan worden gemaakt van berekeningen, maar ook van simulatie.
8. Stappenplan
   Hoewel een stappenplan het karakter van een recept of keurslijf kan krijgen, lijkt het bij het onderwijs in hypothesetoetsen nuttig om iets dergelijks aan te bieden. In de gangbare wiskundemethoden staan hiervan voorbeelden. Zie ook bijlage 2.

6. Conclusie

In het begin van het katern stelden we onszelf de volgende twee vragen:

1. Hoe kunnen leerlingen van wiskunde A en C met inzicht de stap zetten van het beschrijven van steekproefgegevens naar het trekken van verantwoorde conclusies over de onderliggende populatie?
2. Welke instructiestrategieën kunnen deze ontwikkeling ondersteunen?

In een poging deze vragen te beantwoorden, hebben we allereerst het vakgebied in kaart gebracht en een aantal potentiële valkuilen van statistiek besproken. Vervolgens is de statistische cyclus beschreven, die te beschouwen is als een variant van de modelleercyclus uit het katern Modelleren.

De moeilijkheden van leerlingen, en zeker die van wiskunde A en C, lijken vooral te liggen in de overstap van steekproef naar populatie, en met name in de stap van het mathematiseren. Dat is een vrij complex proces, dat vele blikwisselingen vergt, die soms tegen de intuïtie ingaan. Voor het specifieke probleem van hypothesetoetsen is deze stap uitgewerkt in paragraaf 5, Ontwerpen. De analyse van didactische aandachtspunten heeft geleid tot een aantal concrete handvatten voor onderwijsontwerp.

Het antwoord op de eerste vraag komt er in de kern op neer dat leerlingen een ervaringsbasis nodig hebben, die ze kunnen opdoen in experimenten en simulaties. Deze basis moet worden uitgebouwd door een expliciete discussie van de stappen die in het proces van hypothesetoetsing worden gezet. Speciale aandacht daarin verdienen de rol van het toeval en de blikwisselingen tussen het perspectief van de gegevens en de steekproef enerzijds en dat van de kansmodellen en de populatie anderzijds. Een onderwijsstrategie om hiermee om te gaan kan niet anders dan deze conceptuele moeilijkheden serieus nemen door er de nodige aandacht aan te besteden. De eerder genoemde handvatten bieden daarvoor aanknopingspunten.

Referenties

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht: CD-ß Press

Barbella, P., Denby, L., & Landwehr, J.M. (1990). Beyond exploratory data analysis: the randomization test. *Mathematics Teacher, 83*, 144-149.

Broek, J. van den, & Kop, P. (2001). *Kattenaids en statistiek*. Utrecht: Epsilon.

cTWO (2009). Concept-examenprogrammas wiskunde 2014. Utrecht: cTWO. www.ctwo.nl.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. 1*. New York: John Wiley.

Franklin, C. , Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaeffer R. (2005). *A curriculum framework for PreK-12 statistics education*. Prepared by Guidelines for Assessment and Instruction in Statistics Education (GAISE). Alexandria MD: American Statistical Association. http://www.amstat.org/education/gaise/.

Freedman, D., Pisani, R., & Purves, R. (2006). *Statistics*. New York: Norton & Company Inc.

Sittig, J. & Freudenthal, H. (1951). *De juiste maat. Lichaamsafmetingen van Nederlandse vrouwen als basis van een nieuw maatsysteem voor damesconfectiekleding*. Leiden: Stafleu.

Gal, I., & Garfield, J. B. (Eds.). (1997). The assessment challenge in statistics education (First ed.). Amsterdam: IOS Press.

Gonick, L. & Smith, W. (2004). *Het stripverhaal van de statistiek*. Epsilon, http://www.epsilon-uitgaven.nl/E32.php

Hemelrijk, J. (1972). *Statistiek, te pas en te onpas*. Hilversum: Teleac.

Huff, D. (1954). *How to Lie with Statistics*. New York: Norton.

Konold, C., & Giessen, C. van de (2008). Data-analyse met behulp van educatieve software. *Euclides, 83*(4), 208-212.

Lambalgen, M. van, & Meester, R. (2005). Wat zeggen al die getallen eigenlijk? *Nieuwe Wiskrant, Tijdschrift voor Nederlands Wiskundeonderwijs, 24*(4), 9-14.

Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of the giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Moore, D. S. (1997). New pedagogy and new content: The case for statistics. *International Statistical Review, 65*, 123-165.

Neeleman, W., & Verhage, H.B. (1999). Historische hoogtepunten van grafische verwerking. *Nieuwe Wiskrant, Tijdschrift voor Nederlands Wiskundeonderwijs, 18*(3), 18-33.

Nijdam, B. (2008). Probleemgeoriënteerde statistiek en kansrekening binnen wiskunde A/C. *Euclides, 83*(4), 143-147.

SKACA (2007). *Aanbevelingen voor de inhoud van de statistiek en kansrekening binnen wiskunde A en C vwo en wiskunde A havo.* Utrecht: cTWO. www.ctwo.nl.

Scheaffer, R., & Tabor, J. (2008). Statistics in the High School Mathematics Curriculum. Building sound reasononing under uncertain conditions. *Mathematics Teacher, 102*(1), 56-61.

Spandaw, J., & Zwaneveld, B. (2008). Modelleren. In: A. van Streun, B. Zwaneveld & P. Drijvers (Eds.), *Handboek Vakdidactiek Wiskunde*. www.elwier.nl.

Streun, A. van, & Giessen, C. van de (2007a). Een vernieuwd statistiekprogramma deel 1: Statistiek leren met data-analyse. *Euclides, 82*(5), 176-179.

Streun, A. van, & Giessen, C. van de (2007b). Een vernieuwd statistiekprogramma deel 2: Data-analyse, een mogelijke opzet. *Euclides, 82*(6), 217-221.

Taleb, N.N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Tukey, J.W. (1977). *Exploratory Data Analysis*. New York: Addison-Wesley.

Tijms, H. (2008). Bayesiaanse kansrekening. *Euclides, 83*(4), 154-156.
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistics Review, 67*, 223-265.

Bijlage 1: Meetniveaus

In de statistiek worden vier meetniveau's van variabelen onderscheiden: nominaal, ordinaal, interval, en ratio. Het meetniveau van een variabele heeft gevolgen voor de geschikte grafische representatie en de passende centrum- en spreidingsmaten. De volgende tabel geeft een overzicht van kenmerken van deze typen, met voorbeelden, grafische weergave en centrum- en spreidingsmaten.

meetniveau	**kenmerken**	**voorbeelden**	**grafische weergave**	**centrum- en spreidingsmaat**
Nominaal	alleen namen volgorde speelt geen rol	Geslacht Roepnaam Merk	Cirkeldiagram Staafdiagram	Centrum: modus
Ordinaal	volgorde heeft betekenis, maar onderlinge afstand niet	Onderwijsniveau Rang bij krijgsmacht of politie	Cirkeldiagram Staafdiagram	Centrum: mediaan
Interval	getallen waarbij verschillen betekenis hebben, maar verhoudingen niet.	Temperatuur in ºC Schoenmaat	Histogram Polygoon Boxplot	Centrum: gemiddelde Spreiding: standaarddeviatie, tussenkwartielafstand
Ratio	getallen waarbij ook de verhoudingen betekenis hebben. Er is een (natuurlijk) nulpunt.	Lengte Gewicht Massa Temperatuur in ºK	Histogram Polygoon Boxplot	Centrum: gemiddelde Spreiding: standaarddeviatie, tussenkwartielafstand

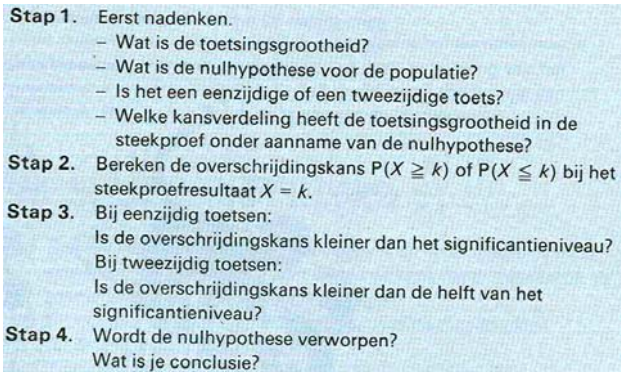
Natuurlijk helpt deze tabel niet om het onderscheid tussen meetniveaus betekenis te geven. Leerlingen kunnen wellicht aan de hand van criteria bij voorbeelden op het goede spoor komen en voor elk meetniveau zelf voorbeelden bedenken.

Opgave 12

1. Kun je zeggen dat Piet drie keer zo zwaar is als zijn broertje? Welk meetniveau heeft dus de variabele lichaamsgewicht?
2. Kun je zeggen dat Piet (schoenmaat 42) twee keer zo grote voeten heeft als zijn broertje met schoenmaat 21? Van welk meetniveau is hier sprake?
3. Kun je zeggen dat Piet vier keer zo brutaal is als zijn broertje? Meetniveau?

Bijlage 2: Stappenplan hypothesetoetsing

Eerder in dit katern is opgemerkt dat leerlingen wellicht houvast kunnen gebruiken voor het uitvoeren van een statistische toets, en met name voor de fase van het mathematiseren. Voor het toetsen van hypothesen zijn verschillende stappenplannen beschikbaar die in deze behoefte kunnen voorzien. Een aardig stappenplan in stripvorm is bijvoorbeeld te vinden in het beeldende boek 'Het stripverhaal van de statistiek' (Gonick & Smith, 2004). Ook Van den Broek en Kop (2001) geven een stapsgewijze aanpak van hypothesetoetsing. Figuur 9 toont een stappenplan uit een oudere editie van een van de Nederlandse wiskundemethoden.


Stap 1. Eerst nadenken.
- Wat is de toetsingsgrootheid?
- Wat is de nulhypothese voor de populatie?
- Is het een eenzijdige of een tweezijdige toets?
- Welke kansverdeling heeft de toetsingsgrootheid in de steekproef onder aanname van de nulhypothese?

Stap 2. Bereken de overschrijdingskans $P(X \geq k)$ of $P(X \leq k)$ bij het steekproefresultaat $X = k$.

Stap 3. Bij eenzijdig toetsen:
Is de overschrijdingskans kleiner dan het significantieniveau?
Bij tweezijdig toetsen:
Is de overschrijdingskans kleiner dan de helft van het significantieniveau?

Stap 4. Wordt de nulhypothese verworpen?
Wat is je conclusie?

Figuur 9 Stappenplan voor de uitvoering van de toets

Een stappenplan voor hypothesetoetsing, dat aansluit bij de statistische cyclus zoals in dit katern beschreven, zou er als volgt kunnen uitzien.

Conceptualiseren

1. Bedenk wat de vraag is, hoe die in statistische termen vertaald kan worden, welke variabele een rol speelt en van welk type en meetniveau deze is.
2. Verzamel gegevens en breng die overzichtelijk in beeld.

Mathematiseren

3. Definieer de toetsingsgrootheid, zeg *T*.
4. Bepaal het type verdeling van de toetsingsgrootheid of doe hierover aannames.
5. Formuleer de statistische hypothesen in termen van de toetsingsgrootheid. De nulhypothese H0 moet enkelvoudig zijn.
6. Ga na welke waarde voor *T* je verwacht als H0 waar is en in welke richting de afwijkingen liggen als de alternatieve hypothese H1 waar is.
7. Kies een significantieniveau.

Oplossen

8. Neem aan dat H0 waar is. Bepaal onder die aanname de kans dat de toetsingsgrootheid de gevonden waarde aanneemt die je onder H0 zou verwachten, of verder van deze waarde afwijkt.
9. Vergelijk deze overschrijdingskans met de onbetrouwbaarheidsdrempel.

Interpreteren

10. Interpreteer de conclusie in termen van de oorspronkelijke probleemstelling.

Bijlage 3: Concept-eindtermen Statistiek en kansrekening wiskunde A en C

De door cTWO opgestelde concept-examenprogramma's 2014 bevatten voor wiskunde A van havo en vwo en voor wiskunde C van vwo een domein Statistiek en kansrekening (cTWO, 2009). De eindtermen van wiskunde A worden hieronder opgesomd. Die van wiskunde C vwo zijn identiek aan die van wiskunde A vwo, zij het dat het subdomein Verklarende statistiek (E6) bij wiskunde C is vervallen.

Wiskunde A vwo Domein E: Statistiek en kansrekening

Subdomein E1: Probleemstelling en onderzoeksontwerp
11 De kandidaat kan bij een probleemstelling die zich leent voor een statistische aanpak een plan maken om antwoord op de probleemstelling te verkrijgen, waarbij geschikte variabelen worden gekozen.

Subdomein E2: Visualisatie van data
12 De kandidaat kan verkregen data verwerken in een geschikte tabel of grafiek en deze op waarde interpreteren.

Subdomein E3: Kwantificering
13 De kandidaat kan de verkregen data samenvatten in voor de probleemstelling geschikte maten en hieraan interpretaties verbinden.

Subdomein E4: Kansbegrip
14 De kandidaat kan het kansbegrip gebruiken om bij een toevalsproces de kans op een bepaalde uitkomst of gebeurtenis te bepalen aan de hand van een diagram, combinatoriek, kansregels en simulatie.

Subdomein E5: Kansverdelingen
15 De kandidaat kan aangeven in welke situatie een toevalsvariabele een bepaalde kansverdeling bezit en van die verdeling de karakteristieken verwachtingswaarde en standaardafwijking hanteren.

Subdomein E6: Verklarende statistiek
16 De kandidaat kan in een probleemsituatie op basis van steekproefgegevens een uitspraak doen over een populatie, de betrouwbaarheid daarvan kwantificeren en het resultaat duiden in termen van de context.

Wiskunde A havo Domein E: Statistiek en kansrekening

Subdomein E1: Presentaties van statistische data interpreteren
13 De kandidaat kan statistische data die op diverse manieren zijn gerepresenteerd en/of samengevat interpreteren en beoordelen op relevantie.

Subdomein E2: Statistische data verwerken
14 De kandidaat kan statistische data verwerken, organiseren, bewerken, weergeven in grafieken, tabellen en diagrammen, en samenvatten met geschikte centrum- en spreidingsmaten.

Subdomein E3: Data en kansen

15 De kandidaat kan bij een toevalsproces de waarschijnlijkheid (kans) van een bepaalde uitkomst of gebeurtenis bepalen of inschatten.

Subdomein E4: Data analyseren

16 De kandidaat kan bij een probleemstelling die zich leent voor een statistische aanpak het soort probleem herkennen en data verzamelen en analyseren om antwoord op de probleemstelling te verkrijgen.