

Computerstatistiek

S. L. Kemme
Lettelbert

Samenvatting

In het dagelijks leven worden we regelmatig geconfronteerd met ingewikkelde verzamelingen van gegevens. Om inzicht te krijgen in een verzameling proberen we de belangrijkste karakteristieken daarvan op te sporen. Daarbij wordt gebruik gemaakt van allerlei statistische technieken zoals: verzamelen, selecteren, samenvatten en representeren.

In het statistiekonderwijs spelen de statistische technieken doorgaans een dominante rol. Vaak worden leerlingen geconfronteerd met een kunstmatige verzameling van gegevens waarin het proces van verzamelen en selecteren al heeft plaatsgevonden. Door deze voorselectie krijgen statistische vragen een voorafgegeven interpretatie. Deze onevenredig grote aandacht voor technieken geeft leerlingen een eenzijdige kijk op de statistiek. Het idee van statistiek als een modellerende en ordenende activiteit om een beter inzicht te krijgen in de werkelijkheid, gaat vrijwel verloren.

In dit artikel wordt een methode besproken waarmee leerlingen kunnen werken met ruwe verzamelingen van gegevens. Dat dwingt leerlingen na te denken over statistische vragen, een interpretatie aan die vragen te geven en een geschikte manier te kiezen om de data te selecteren. Op deze manier kan het onderwijs in de statistiek geënt worden op realistische situaties. Bij een dergelijke aanpak is het gebruik van de computer onmisbaar.

Inleiding

“Zijn meisjes beter in wiskunde dan jongens?”

Hoewel dit een veelgestelde en realistische vraag is over de werkelijkheid, kan die niet zonder meer beantwoord worden. Daarvoor is de vraag te open. Er zullen keuzen moeten worden gemaakt. Zoals: welke wiskunde, welke leerlingen, naar welke wiskundige resultaten kijk je, wat versta je onder ‘beter’, enzovoorts. Dit proces van toespitsing hoort bij de methodologie van statistisch onderzoek. In het statistiekonderwijs moet dat dan ook op een fundamentele manier aanwezig zijn.

Er bestaan twee bekende manieren van data-analyse. De eerste is de zogenaamde ‘klassieke’ inductieve statistiek. Hierbij wordt pas naar de data gekeken nadat een heldere operationaliseerbare vraag is gesteld. Bij de tweede vorm, de ‘exploratory data analysis’ (EDA) wordt een gegeven data-set onderzocht door er geschikte representaties bij te zoeken.

In dit artikel schetsen we een derde manier van data-analyse: de ‘logische gegevens-analyse’. Een realistische vraag wordt eerst op de betekenis onderzocht en er wordt voor een specifieke interpretatie gekozen. Deze interpretatie resulteert in een geschikte selectie van de data. Logische gegevens-analyse kan een brug zijn tussen de inductieve statistiek en EDA.

Het idee van logische gegevens-analyse kan op een vanzelfsprekende manier worden geïmplementeerd in een computerprogramma. Met behulp van de computer kunnen gemakkelijk selecties uit ruwe dataverzamelingen gemaakt worden. Iedere selectie hoort bij een specifieke onderzoeksvraag. Dat betekent dat het computerprogramma een vraagtaal moet bevatten zodat leerlingen gemakkelijk met nieuwe vragen en selecties kunnen experimenteren. Als dit gedeelte van het programma gecombineerd wordt met mogelijkheden voor grafische en numerieke representaties van data, ontstaat een krachtig instrument waarmee leerlingen op een actieve en constructieve manier de principes van de statistische methode kunnen aanleren.

Inductieve statistiek

De hamvraag van de statistiek is: gegeven een ingewikkelde verzameling van data, hoe kunnen we deze gegevens zó organiseren dat we een helder overzicht en inzicht kunnen krijgen in de verzameling?

In de klassieke inductieve statistiek is de vraag: “Zijn meisjes beter in wiskunde dan jongens?” een vraag over de meeste algemene denkbare populatie (wat dat dan ook mag zijn). Het probleem is om een procedure te vinden waarmee het wiskundig gerechtvaardigd is om conclusies te trekken over deze populatie vanuit een beperkte dataverzameling. Daarvoor zijn allerlei praktisch

uitvoerbare technieken ontworpen. Bijvoorbeeld door een nul-hypothese, een alternatieve hypothese, te formuleren en een betrouwbaarheidsinterval te bepalen. Wiskundig gezien is er sprake van de constructie van een kansmodel. Pas na deze theoretische analyse mag de onderzoeker beginnen met het verzamelen van data. Op deze manier wordt een 'objectieve' methode van onderzoek gegarandeerd, waarin persoonlijke keuzen van de onderzoeker zichtbaar zijn gemaakt en vastgelegd. Het gevaar van een zichzelf waarmakende voorspelling, het grote gevaar van onzorgvuldig statistisch onderzoek, is daarmee omzeild.

Inductieve statistiek is een einddoel in het vwo. Het is een voorbereiding op de methoden van inductieve statistiek in de wetenschap. Leerlingen vinden de procedures voor het testen van hypothesen maar moeilijk en omslachtig. De methode wordt vaak gepresenteerd als een algoritmische procedure die je kunt uitvoeren zonder dat je echt hoeft te snappen wat de belangrijke gedachten zijn achter de principes van de inductieve statistiek. Die gedachten hebben ook een behoorlijk abstract karakter van zichzelf. Wat is bijvoorbeeld een 'algemene populatie'? Moet je daarbij denken aan alle mensen op aarde? Een dergelijk beeld is niet terzake en zelfs misleidend. Het aantal mensen doet er niet toe. Belangrijk zijn de theoretische kenmerken die de populatie verondersteld wordt te hebben. Dit is een manier van hypothetisch denken waar leerlingen weinig vertrouwd mee zijn. Daar komt dan nog het kansmodel overheen. Simpel gezegd komt het op het volgende neer. Stel dat men wil weten hoe groot kastanjes *in het algemeen* zijn. Neem de populatie kastanjes in gedachten. Hoe zou je van die populatie aan de weet kunnen komen wat hun grootte is? Naar een kastanjeboom in Lettelbert gaan en de kastanjes meten die toevallig onder de boom liggen? Zo werkt dat niet in de statistiek. Men zal zich allereerst moeten afvragen of diè kastanjes onder diè boom in Lettelbert representatief zijn voor de hele populatie. Maar hoe kun je dat nagaan, als je nog niet weet hoe die populatie eruit ziet?

Het redeneren vanuit kansen is één van de steunpilaren van de inductieve statistiek. Ook hier speelt het hypothetisch redeneren een belangrijke rol. Veel argumenten hebben immers een 'Wat als ...' gedaante. Bijvoorbeeld: in een vaas zitten twintig rode en dertig witte ballen, *wat* is de kans om een rode én een witte te pakken *als* je er twee tegelijk pakt. Deze manier van redeneren wordt doorgaans aangeleerd vanuit concrete ervaringen met systematisch tellen in kunstmatige situaties. Het is eveneens belangrijk om zich daarbij te realiseren dat het begrip 'kans' een theoretisch en mentaal begrip is. De bewering dat de kans op het gooien van een zes bij een dobbelsteen een zesde is, is iets heel anders dan de bewering dat morgen de kans op regen 40% is. In het eerste geval gaat het om de situatie dat er met een fictie-

ve zuivere dobbelsteen wordt geworpen waarbij na een groot aantal keer gooien ongeveer een zesde gedeelte van de worpen een zes zal zijn. In het tweede geval gaat het om een bewering over een gebeurtenis die niet herhaalbaar is waarbij het percentage alleen een mate van onzekerheid aangeeft. Ondoordachte associaties met het woord 'kans' uit de omgevingstaal, kunnen leerlingen dus totaal op het verkeerde been zetten.

Samengevat

In de inductieve statistiek wordt de werkelijkheid op een hoog wiskundig niveau gemodelleerd. Het onderwijs zal een paar moeilijke klippen moeten zien te omzeilen. Dat betreft het abstracte karakter van het begrip 'populatie', het hypothetisch redeneren en de verwarring in taalgebruik rondom het woord 'kans'. In het verleden werden deze problemen veelal opgevangen door de leerlingen te confronteren met kunstmatige situaties. In onderwijs dat uitgaat van realistische situaties en de pretentie heeft toepasbaar te zijn, zal er naar andere wegen gezocht moeten worden.

Exploratory Data Analysis (EDA)

De door Tukey e.a. gepropageerde visie van 'exploratory data analysis' kan een beginpunt zijn voor een wiskundige opbouw vanuit de realiteit. Bij EDA is de vraag: Wat laten de data zien? Het onderzoek is een soort detectivewerk binnen een bak met ruwe gegevens. Vooral grafische technieken zijn belangrijk gereedschap bij dit onderzoek. Een prachtig voorbeeld van deze situatie is gegeven door Engel (1982) in zijn onderzoek naar het toevalskarakter van het selectiesysteem voor de dienstplicht in Amerika. EDA biedt veel mogelijkheden om statistiek op een concreet niveau te leren. Een aantal grafische technieken, zoals stengel-blad diagrammen en boxplots, hebben hun weg al gevonden in het wiskunde-onderwijs.

Hoewel EDA een veelbelovende introductie lijkt te zijn voor het statistiekonderwijs, is er toch een fundamenteel verschil met de inductieve statistiek. Inductieve statistiek start met een statistisch probleem *waarna* naar de gegevens wordt gekeken. EDA *begint* met een verzameling gegevens waarover allerlei vragen gesteld worden. EDA geeft dus een beeld van de statistische methode die haaks staat op die van de inductieve statistiek.

In de dagelijkse praktijk is het echter nooit zo dat er een statistisch probleem is zonder een bijbehorende gegevensverzameling. Omgekeerd is het ook zo dat er aan een bepaalde gegevensverzameling automatisch al een stel vragen hangen. Om bij het voorbeeld van de kastanjes te blijven. De situatie zou als volgt kunnen zijn. Bij het rapen van kastanjes in Lettelbert, vraagt de raper zich af of alleen de grote en zware kastanjes onder de boom liggen. Er is dus een concrete vraag die rechtstreeks gekoppeld is aan de gegevensverzameling, de kastanjes van de kastanjeboom.

Logische gegevens-analyse

In het algemeen zijn dus bepaalde gegevensverzamelingen direct gekoppeld aan statistische vragen. De statistische vraag is echter meestal niet in overeenstemming met het karakter van de gegevensverzameling. De vraag of meisjes beter zijn in wiskunde dan jongens, is een generaliserende vraag die niet past bij een gegevensverzameling bestaande uit de wiskundecijfers van drie scholen. In een dergelijke situatie zal de vraag moeten worden ingeperkt. Dat vraagt om 'logische gegevens-analyse'. Het gaat om vragen als: hoe specifiek/representatief is de gegevensverzameling; wat is de strekking van de vraag; hoe kunnen beide bij elkaar passend worden gemaakt?

In het volgende zal dit worden uitgewerkt aan de hand van het voorbeeld over de wiskundige capaciteiten van meisjes en jongens.

In de tabel hieronder staan de resultaten van het vwo-examen van een school. Alleen een klein gedeelte van de tabel is weergegeven. De kolommen we en wt bevatten de cijfers op wiskunde 1, respectievelijk wiskunde 2. De vraag of meisjes beter in wiskunde zijn dan jongens kunnen we met deze gegevens alleen in beperkte zin beantwoorden. We kunnen hooguit iets beweren over het verschil in wiskundige capaciteiten tussen jongens en meisjes op deze school, in dit examenjaar. Daarmee wordt de vraag aanzienlijk ingeperkt. We zullen dus niet gaan proberen om een algemene uitspraak aan de gegevens te ontlocken.

In de volgende stap wordt een beslissing genomen over de criteria waarop we zullen besluiten wanneer we meisjes beter vinden in wiskunde dan jongens. Kijken we naar de gemiddelden? Of zullen we de frequentie-grafiek erbij halen zodat we ook naar de spreiding van de resultaten kunnen kijken? Zullen we besluiten dat meisjes beter zijn omdat 66% van de meisjes het moeilijke

wiskunde 2 hadden gekozen en maar 30% van de jongens? De uiteindelijke keuze van het criterium zal afhangen van de situatie en het belang van de vraag. Bij ieder criterium past een eigen gegevensverzameling en een eigen statistische aanpak. De keuze van het criterium komt hier dus vóór de keuze van het statistische gereedschap.

De realistische vraag of meisjes beter zijn in wiskunde dan jongens, leidt dus tot een 'logische' analyse en daarmee tot inperkingen van vragen en aanpassingen van de gegevensverzameling. Deze *logische gegevens-analyse* vormt de brug tussen EDA en de inductieve statistiek en kan daarmee een belangrijke bijdrage leveren aan het onderwijs in de statistische methode. Bij deze logische analyse kan geschikt gebruik gemaakt worden van de mogelijkheden van de computer.

Logische gegevens-analyse en de computer

In realistische situaties zien gegevensverzamelingen er groot en ingewikkeld uit. Die complexiteit is tevens een voorwaarde om op een geschikte manier kennis te maken met de statistische methode. Fundamentele activiteiten als 'verzamelen', 'deelverzamelingen nemen', 'representeren' en 'samenvatten' kunnen alleen een functionele betekenis krijgen vanuit een onoverzichtelijke situatie. Deze activiteiten geven inzicht en overzicht aan de gegevensverzameling. De onoverzichtelijkheid wordt hiermee het hoofd geboden.

Verzamelen, selecteren, tellen in grote gegevensverzamelingen zijn handelingen die veel tijd vragen als het met de hand moet gebeuren. Het gebruik van een computer lijkt niet te vermijden. Het is dus de moeite waard op zoek te gaan naar een computeromgeving waarin deze activiteiten op een gemakkelijke en heldere manier kunnen worden uitgevoerd. Het gebruik van relationele

naam	seks	ne	fr	du	eng	we	wt	na	sc	bi	ee	et
Sepp	m	6	6	•	6	5	3	5	5	6	•	•
Arie	m	7	7	•	8	6	•	6	6	6	•	•
George	m	5	•	•	5	5	•	5	4	•	5	•
Tine	v	7	•	6	7	8	7	6	8	7	•	•
Marc	m	8	•	•	9	9	•	8	•	8	8	•
Chiel	m	7	4	•	8	10	9	9	9	•	8	•
Robert	m	8	•	•	6	7	•	7	6	•	7	•
Pieter	m	6	8	•	7	7	•	7	8	7	•	•
Jacky	v	7	•	7	7	8	6	6	5	7	•	•
Nicky	v	6	•	•	5	7	•	6	5	7	•	•

De tabel EXAMEN

database-omgevingen is per definitie voor dit soort situaties gemaakt.

In relationele database-omgevingen worden de gegevens weergegeven in één of meer tabellen. Dit is een vanzelfsprekende en krachtige manier. Een tabel is een verzameling rijen en kolommen. Een rij bevat de gegevens van een 'individu' uit de gegevensverzameling. Iedere kolom bevat de informatie van één van de gegevens van alle individuen. Kolommen kunnen worden geselecteerd door ze op te noemen. Rijen worden geselecteerd door voorwaarden aan de gegevens te stellen.

Stel dat we in het voorbeeld van de tabel EXAMEN het volgende criterium hadden gekozen: "Meisjes zijn beter als het percentage meisjes met een voldoende voor wiskunde 1 groter is dan dat van de jongens". We hadden dan de kolom 'we' moeten kiezen en eerst alleen naar de meisjes met voldoende voor 'we' moeten kijken en hetzelfde vervolgens moeten doen voor de jongens. In de vraagtaal van DataStat¹ ziet een dergelijke selectie er als volgt uit:

Selecteer uit tabellen: EXAMEN

De kolommen; naam, we

Waarvoor geldt: sekse = 'm' en we > 5

Het resultaat is een nieuwe tabel met de kolommen 'naam' en 'we' en met rijen de meisjes met een cijfer hoger dan 5 voor wiskunde 1.

De vraagtaal ondersteunt op een vanzelfsprekende manier de keuze van een beslissingscriterium. Het dwingt tot een uitvoerbare keuze van dat criterium.

Relationele database-omgevingen kunnen zeer complexe gegevensverzamelingen de baas. Niet alleen kan een tabel een grote omvang hebben, een gegevensverzameling kan ook bestaan uit meerdere tabellen. De gegevensverzameling *Het weer in de zomer van 1990* bestaat bijvoorbeeld uit vijf tabellen. Iedere tabel bevat de gegevens van een plaats in Nederland. Iedere tabel heeft acht kolommen met de dagelijkse weergegevens van dertig weken over temperatuur, luchtdruk, hoeveelheid regen, tijdsduur van de regen op een dag, tijdsduur van zonneschijn, windsnelheid. Iedere rij bevat de gegevens van een dag uit die dertig weken. De tabellen hangen met elkaar samen doordat ze dezelfde kolomnamen hebben. Het verschil tussen een zomer aan de kust of meer in het land laat zich nu onderzoeken door twee tabellen van een plaats aan zee en een plaats ver van zee te koppelen en de gegevens samen te vatten.

Een dergelijke grote klus laat zich alleen maar met een computer uitvoeren. In een relationele database omgeving is dit op een eenvoudige manier te realiseren.

Naast het snelle en accurate werk dat een computer kan verrichten, zijn ook de interactieve mogelijkheden een belangrijk aspect van het gebruik van de computer in de klas. In een interactieve omgeving kunnen leerlingen op

een constructieve en creatieve manier aan het werk zijn. Voor de software heeft dit vergaande gevolgen. De software moet een 'open' karakter hebben. Niet het programma stelt de vragen, maar de leerling. De computer voert uit wat de leerling heeft opgedragen. De computer is de knecht van de leerling en niet de leerling de knecht van de computer. Alleen dan is te verwachten dat leerlingen voldoende worden uitgedaagd om zo hun eigen ervaringen op te kunnen bouwen.

Het toetsen van statistiekonderwijs met de computer

Het computerprogramma DataStat en het bijbehorende lesmateriaal voor klas 5 en 6 wiskunde A is in nauwe samenwerking met een aantal proefscholen ontwikkeld. Iedere nieuwe versie van het programma en het lesmateriaal is direct in de klas uitgetoetst en bijgesteld op basis van de ervaringen.

De toetsing had een belangrijke plaats in de experimenten. Niet alleen werd ieder gedeelte afgesloten met een toets of een proefwerk, ook zijn er schoolonderzoeken gemaakt en afgenomen. Bij deze schoolonderzoeken is de computer ingezet. De schoolonderzoeken vonden plaats in de week waarin ook andere individuele schoolonderzoeken (taalvaardigheid, natuurkundepracticum) werden afgenomen. Dit schoolonderzoek wiskunde leek het meest op een praktisch schoolonderzoek bij natuurkunde. De leerlingen kregen een opdracht en kregen de beschikking over een computer om de opdracht te verwerken. Na afloop van het practicum was er tijd om de resultaten schriftelijk uit te werken.

In één van de schoolonderzoeken werden de examenresultaten gegeven van twee verschillende scholen. Bij iedere leerling was niet alleen het cijfer maar ook de groep vermeld. Zo kon een koppeling worden gelegd met de groep en de docent.

Eén van de vragen luidde:

- > *Onderzoek de waarheid van de volgende bewering:
Een goed cijfer in wiskunde B levert ook een goed cijfer in wiskunde A.
Beargumenteer je antwoord vanuit de database.*

Om de vraag te kunnen beantwoorden moeten de leerlingen een correcte selectie maken uit de rijen en uit een combinatie van variabelen. De meeste leerlingen waren in staat om een korte en overzichtelijke werktabel te produceren en/of een regressielijn of regressievergelijking op te stellen. 40% van de leerlingen gaf een beargumenteerd antwoord. Enkele reacties waren:

Raoul: Ik vind dat een hoog cijfer een 8 of meer is, en dan is de bewering waar. Kijk maar naar de nummers 3 en 4 van de tabel.

Janneke: Als je een hoog cijfer op wiskunde B hebt dan

heb je ook een hoog cijfer op wiskunde A (maar als je geen hoog cijfer hebt op wiskunde B, dan kun je toch nog een hoog cijfer op wiskunde A hebben).

Marco: Ja, als je naar de grafiek kijkt (hij maakte een regressielijn tussen de cijfers) dan zie je dat het cijfer voor wiskunde A altijd hoger is dan dat voor wiskunde B.

Edwin: Twee cijfers op wiskunde B zijn hoger dan een 8 en bij deze twee zijn de cijfers zelfs hoger dan 8,5. Van de andere cijfers hoger dan een 7 op wiskunde B was er slechts één hoger dan een 7 op wiskunde A. Toch geloof ik (ondanks dat ene cijfer lager dan een 7) dat de bewering waar is.

Timon: Ik denk dat de bewering globaal waar is. (Hij gaf een regressievergelijking.) Maar als je naar de grafiek kijkt, zie je slechts twee punten waar het cijfer op wiskunde A hoger is dan dat op wiskunde B.

In deze reacties zien we dat leerlingen kunnen leren om een geschikte methode te ontwerpen voor een statistisch probleem. Ze kunnen keuzen maken die praktisch uitvoerbaar zijn. Opvallend is het contrast tussen de globale interpretatie (de regressielijn) en de lokale interpretatie zoals onderstreept door Timon. De globale interpretatie steunt de algemene opvatting dat leerlingen die wiskunde B doen altijd goed zijn in wiskunde A. In de lokale interpretatie klinkt twijfel door over de waarheid van die bewering.

Geen enkele leerling merkt op dat de populatie te klein is om een algemene uitspraak te kunnen doen. We moeten onze verwachtingen over het begrijpen van een statistische methode dus niet te hoog stellen. Dit is maar een begin. In het nieuwe programma voor de onderbouw is het gebruik van de computer bij de statistiek opgenomen in één van de kerndoelen. Misschien dat er dan via een langere leerstoflijn meer bereikt kan worden op dit punt.

Tot slot

Het gebruik van logische gegevens-analyse in het statistiekonderwijs opent de mogelijkheid tot een meer realistische cursus in de richting van de statistische methode. Een schets van een leergang voor het hele leerplan zou er als volgt uit kunnen zien. Men zou kunnen beginnen met EDA-activiteiten, waarbij het accent ligt op het kiezen van geschikte grafische representaties vanuit een realistische vraag over een voorgeselecteerde gegevensverzameling. In de volgende stap maken de leerlingen kennis met de logische gegevensanalyse in combinatie met EDA. De gegevensverzamelingen hebben een uit-

gebreid en complex karakter. Daardoor worden de leerlingen gedwongen om na te denken over selectie-criteria. Dit kan functioneren als het begin van het leren van een statistische methode. Tot slot kan begonnen worden met kansrekening, uitmondend in de inductieve statistiek. In deze opzet wordt pas in een laat stadium met kansrekening begonnen.

Het idee hier achter is tweeledig. De eerste twee fasen leveren een concreet en realistisch ervaringsfundament waarop het denken en redeneren over toevalssituaties gestalte kan krijgen. Bovendien vraagt het theoretische en mentale karakter van de kansrekening om voldoende formele en abstraherende vaardigheden van leerlingen. Het is beter om daar niet te vroeg mee te beginnen.

Noot

- [1] Het programma DataStat is ontwikkeld aan de universiteit van Groningen met het doel informatica-aspecten te integreren binnen wiskunde A op het vwo. Vanuit de bovengeschetste optiek leek een combinatie van een relationele database-omgeving met een statistiekpakket een succesvolle keuze. DataStat bevat een vraagtaal (DataStat 0) en een statistische gereedschapskist (DataStat 1). Met DataStat 0 kan een selectie uit andere tabellen worden gemaakt. Dat levert een werktabel op die kan worden ingevoerd in DataStat 1 waarna de statistische representatie kan plaatsvinden. DataStat is een volledig menugestuurde 'open' werkomgeving.

Literatuur

- Bakx, G.J.Th. (1990). De conferentie Wiskundeles en Informatie-Technologie. *Nieuwe Wiskrant*, 9 (2), 3-9.
- Biehler, R. (1990). Daten analysieren mit dem Computer: Unterstützung von Begriffsbildung und Anwendungsorientierung in der Stochastik. *Mathematik Unterricht*, 36 (6), 50-71.
- Engel, A. (1982). Statistik auf der Schule: Ideen und Beispiele aus neuerer Zeit. *Mathematik Unterricht*, 28, (1), 57-85.
- Kemme, S.L. (1991). Effekten van computergebruik. In: Broekman, H.G.B (et.al.): *Algoritmen en heuristieken in contextrijk reken-wiskundeonderwijs*. OW&OC, Utrecht.
- Kemme, S, J. Helder, K. Visser, M. Bos (1992). *DataStat, leerlingenboek, docentenhandleiding en programma*. Wolters Noordhoff, Groningen.
- Tall, D. (1986). *Building and Testing a Cognitive Approach to the Calculus using Interactive Computer Graphics*. The University of Warwick, Warwick.
- Tukey, John W. (1977). *Exploratory Data Analysis*. Reading, Mass.