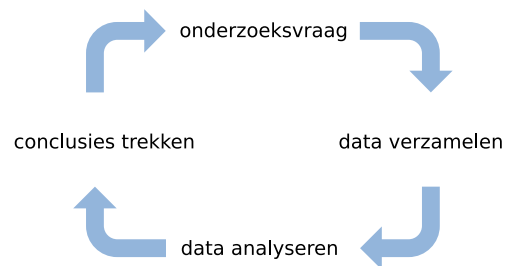


6.1 De statistische cyclus

Inleiding

Bij statistisch onderzoek gaat het om het vinden van verklaringen voor bepaalde verschijnselen met behulp van statistiek en grote datasets. Een voorbeeld is de set met gegevens van de Chicago Marathon in 2016. Er deden meer dan 40000 mensen aan mee en je kunt vragen stellen als: 'Is er bij de lopers een verband tussen de tijd halverwege en de eindtijd?', of 'Lopen mensen jonger dan 30 jaar sneller dan mensen boven de 30?', of 'Zijn de looptijden normaal verdeeld en kun je de verdeling ervan verklaren?'. Allemaal vragen die je met behulp van statistische methoden kunt aanpakken. Daarbij hanteer je een vaste onderzoekscyclus.



Figuur 1

Je leert in dit onderwerp

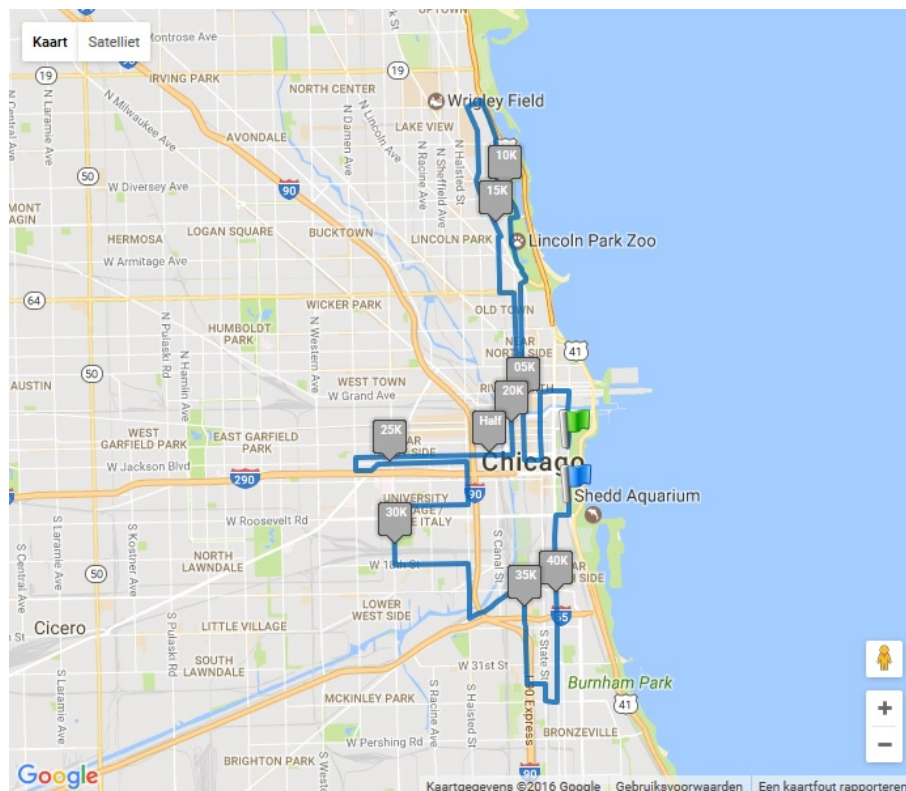
- de opzet van een statistisch onderzoek herkennen in bestaande onderzoeken;
- zelf een statistisch onderzoek opzetten.

Voorkennis

- de beschrijvende statistiek: het maken van (som)frequentietabellen, diagrammen en het berekenen van centrummaten en spreidingsmaten;
- werken met de normale verdeling, met normaal waarschijnlijkheidspapier en met de wortel-n-wet.

Verkennen

Opgave V1



Figuur 2

Bekijk de [gegevens van de Chicago Marathon 2016](#).

- Hoe heette de vrouwelijke deelnemer die het eerst is gefinisht? De hoeveelste eindtijd had zij?
- Hoe heette de vrouwelijke deelnemer die het laatst is gefinisht?
- Bekijk de gegevens en formuleer er een onderzoeksvraag bij. Bedenk hoe je dit onderzoek zou willen uitvoeren en hoe je de antwoorden zou willen formuleren.

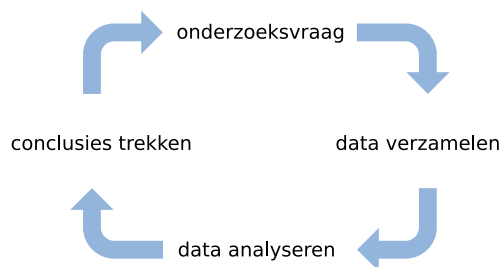
Uitleg 1

Bij statistisch onderzoek zoek je antwoorden op een vraag die een grote groep betreft en die je alleen kunt krijgen door gegevens (data) te verzamelen van een deel van die groep. Door die data te analyseren, te ordenen in tabellen, centrum- en spreidingsmaten en diagrammen probeer je die antwoorden te vinden en te onderbouwen. Je doorloopt de statistische onderzoekscyclus die je hiernaast aantreft.

De grote groep noem je 'populatie' en de kleinere groep waarvan je de data verzamelt heet de 'steekproef'. De moeilijkheid is nu vooral dat je niet zeker weet of wat voor de steekproef geldt, ook automatisch voor de hele populatie geldt. Geeft de steekproef wel voldoende de verschillen in de populatie weer, is hij wel representatief genoeg?

Bij zo'n statistisch onderzoek kun je verschillende stappen onderscheiden:

- de onderzoeksvraag formuleren
- de steekproef kiezen en data verzamelen
- de data analyseren, dus ordenen middels tabellen, diagrammen en centrum- en spreidingsmaten
- conclusies trekken



Figuur 3

Stel je voor dat je wilt onderzoeken op welke leeftijd je het snelst bent op een marathon (42,195 km hardlopen) en of het verschil maakt of je man bent of vrouw. Je gebruikt hiervoor de gegevens van de marathon van Chicago uit 2016, daaraan deden ruim 40000 mannen en vrouwen van over de hele wereld mee. Dat lijkt toch een voldoende grote steekproef, maar is hij ook representatief?

Deze vraag gaat over verschillen tussen deelgroepen.

Je kunt ook vragen naar percentages (proporties) binnen de populatie: "Hoeveel procent van de marathonlopers is ouder dan 50 jaar?"

Of naar verbanden: "Is er een verband tussen de tussentijd halverwege en de behaalde eindtijd?"

Elk van deze vragen vereist een andere onderzoeksopzet.

Opgave 1

Bekijk in **Uitleg 1** het verhaal van de Chicago Marathon in 2016. Open de **gegevens van de Chicago Marathon 2016** in Excel.

- Hoeveel deelnemers waren er? Hoeveel vrouwen waren daarbij?
- Is de steekproef van ruim 40000 marathonlopers die deelnamen aan de Chicago Marathon 2016 groot genoeg om iets te kunnen zeggen over alle marathonlopers? Of zelfs over mensen in het algemeen?
- Is de steekproef van ruim 40000 marathonlopers die deelnamen aan de Chicago Marathon 2016 representatief voor alle marathonlopers?

Opgave 2

Je wilt een onderzoek opzetten rond de **gegevens van de Chicago Marathon 2016**.

- Welke stap in de onderzoekscyclus is al voor je gedaan?
- Probeer de daar als eerste gestelde onderzoeksvraag wat preciezer te formuleren.
- Beschrijf nu hoe je het onderzoek zou gaan opzetten, maak een plan van aanpak. (In de volgende onderdelen ga je hiermee verder.)
- Er wordt in de uitleg ook een onderzoeksvraag geformuleerd waarbij het verband tussen twee variabelen wordt gezocht. Welke variabelen zijn dit en voor welke stappen in het onderzoek betekent dit iets?

Uitleg 2

Rond het thema 'verkiezingen' vinden ook veel interessante statistische onderzoeken plaats.

Zo is door **KiesKompas** een groot onderzoek gehouden voor de Tweede Kamer verkiezingen van september 2012 in Nederland. In het bestand **TK-Verkiezingen-NL-2012** vind je de gegevens van een aselechte steekproef van 1550 personen uit hun enorme databestand. Het gaat hierbij vooral om kwalitatieve variabelen.

De basisgegevens zijn verzameld via internet, zie **Figuur 4** tk2012.kieskompas.nl.

Ook bij dit onderzoek is de statistische onderzoekscyclus gehanteerd. Je kunt er allerlei vragen bij stellen, zoals:

"Vinden mannen veiligheid belangrijker dan vrouwen?"

"Hoe zit het per partij met de zekerheid van de mensen die van plan zijn erop te stemmen?"

"Hoe sterk zal per partij de uitslag van 2012 anders zijn dan die in 2010?"

"Hoe sterk is het verband tussen jezelf links of rechts vinden en jezelf conservatief of progressief vinden?"

Maar je kunt zelf natuurlijk ook nog wel wat vragen bedenken.



Opgave 3

Bekijk het Excel-bestand met de gegevens van een steekproef uit het KiesKompas onderzoek voor de verkiezingen in 2012. Om de betekenis van deze getallen te begrijpen heeft KiesKompas dit **co-deboek TK-Verkiezingen-NL-2012** gemaakt.

- Welke partij heeft nummer 100?
- Hoeveel mensen hebben deze lijst ingevuld en mochten in 2010 nog niet stemmen?
- Hoe zou je er achter kunnen komen in hoeverre mensen tussen 18 jaar en 30 jaar anders gaan stemmen dan mensen boven de 60 jaar?

Opgave 4

In **Uitleg 2** kun je lezen dat de dataset die je kunt gebruiken een aselechte steekproef is uit een veel grotere groep mensen die toen via internet de enquête via tk2012.kieskompas.nl hebben ingevuld.

- Betekent het woord 'aselect' automatisch dat de steekproef representatief is?
- Is een enquête via internet niet bij voorbaat een niet representatieve steekproef?
- Hoe kun je er voor zorgen dat die representativiteit zo goed mogelijk wordt?

Theorie en voorbeelden

Om te onthouden

Er is sprake van **statistisch onderzoek** als je probeert om een onderzoeksvraag die een grote groep betreft (de populatie) te beantwoorden door het verzamelen van gegevens van een deel van die groep (de steekproef). Die steekproef moet dan representatief zijn voor de populatie.

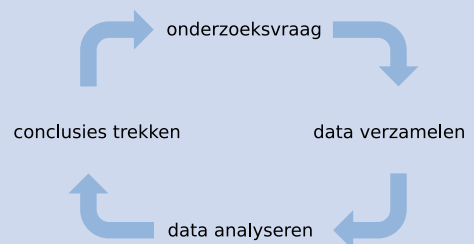
Bij statistisch onderzoek doorloop je de stappen die in de figuur zijn weergegeven.

Het is in dit onderdeel de bedoeling dat je deze **statistische onderzoekscyclus** vertaalt in een plan van aanpak voor een eigen onderzoek.

Er bestaan in grote lijnen drie soorten vragen die elk een eigen aanpak vereisen:

- vragen naar percentages (proporties);
- vragen naar verschillen tussen deelgroepen met betrekking tot dezelfde statistische variabele;
- vragen naar verbanden tussen twee statistische variabelen.

Je hebt kennis nodig betreffende het werken met Excel en grotere datasets. Zie daarvoor het practicum **Data presenteren en vergelijken**.



Figuur 5

Voorbeeld 1

Bij vakken als l.o. en b.s.m. worden vaak sportgegevens bijgehouden, bijvoorbeeld over de Cooper-test, de shuttle-run-test, hoogspringen, discuswerpen, en dergelijke. Bij een vak als n.l.t. worden bij diverse modules ook gegevens bijgehouden waarmee je statistisch onderzoek zou kunnen doen. En wellicht kun je op je eigen school ook aan examenresultaten komen in de loop van de jaren. Met dergelijke gegevens van de eigen school kun je heel goed een statistisch onderzoekje opzetten.

Hoe zou je bijvoorbeeld onderzoeken of de jongens beter in de exacte vakken zijn dan meisjes?

Antwoord

Eerst ga je natuurlijk even nadenken over wat ook weer precies exacte vakken zijn, wat doe je bijvoorbeeld met biologie en n.l.t.? Of ga je liever per vak kijken? En welke gegevens gebruik je erbij: rapportcijfers, cijfers van het schoolexamen of die van het centraal examen, of werk je liever met het gemiddelde van die laatste twee? En wat doe je dan met vakken die geen centraal examen kennen, zoals wiskunde D?

Tijd voor een statistisch onderzoekje. In een goed plan van aanpak horen:

1. Inleiding:

Hierin zeg je iets over waarom je dit onderzoekje wilt gaan uitvoeren. Formuleer een vermoeden, een uitspraak die je wilt testen.

Verder beschrijf je wat je onder 'exacte vakken' verstaat en hoe je in grote lijnen gaat werken.

2. De onderzoeksvraag formuleren:

Bijvoorbeeld: "Zijn jongens beter in de exacte vakken dan meisjes?"

Wellicht maak je ook deelvragen bijvoorbeeld over afzonderlijke vakken.

3. De steekproef kiezen en data verzamelen:

Kies een jaargang en een schooltype waarvan je gegevens kunt krijgen, bijvoorbeeld de eindexamensresultaten van verleden jaar. En kies je dan voor tijdvak 1 of voor de gegevens na tijdvak 2? Je plaatst enkele kanttekeningen bij de representativiteit van die steekproef voor alle jongeren van dat schooltype. Bedenk welke gegevens je allemaal wilt hebben, nu in ieder geval: m/v, profiel, cijfers SE, CE1.

Je zoekt uit hoe je aan de gegevens kunt komen en beschrijft hoe je die dan gaat verwerken. Bekijk bijvoorbeeld het Excel-practicum: **Data presenteren en vergelijken** als je dit nog nooit eerder in Excel hebt gedaan.

4. De data analyseren:

Dat betekent ordenen middels tabellen, diagrammen en centrum- en spreidingsmaten. Denk er wel om dat je relatieve frequenties gebruikt. In het vervolg van dit onderwerp leer je diverse diagrammen gebruiken.

5. Conclusies trekken:

Hier ga je mee verder in het vervolg van dit onderwerp.

Opgave 5

Bekijk **Voorbeeld 1** van het onderzoek naar de resultaten van de exacte vakken.

- a Welke vakken zou jij als exacte vakken nemen? Wat is een nadeel van het vragen naar de resultaten van de exacte vakken tegenover een afzonderlijk vak?
- b Hoe zit het met de representativiteit van de steekproef als je leerlingen van jouw school hiervoor gebruikt?
- c Waarom wordt er met relatieve frequenties gewerkt?

Opgave 6

Bedenk de opzet van een statistisch onderzoek dat je op je eigen school kunt uitvoeren (met gegevens van je eigen school). Stel een plan van aanpak op en verzamel de data. Denk eerst goed na over welke gegevens je wilt hebben. In het vervolg van dit onderwerp zul je dit onderzoek gaan uitwerken.

Voorbeeld 2

Bekijk het bestand **TK-Verkiezingen-NL-2012**.

Hoe zou je op je eigen school een vergelijkbaar onderzoek kunnen opzetten?

Antwoord

Er zijn verschillende mogelijkheden:

- Je kunt een enquête opzetten met vragen zoals je die in het onderzoek van KiesKompas aantreft en de leerlingen van je school die vragenlijst laten invullen. Je moet er dan wel flink reclame voor maken, want niet alle leerlingen nemen dit serieus. Verstandig is om ook docenten (bijvoorbeeld van geschiedenis, maatschappijleer, wiskunde) in te schakelen. Wellicht kunnen zij er in hun lessen wat mee. Een mooi schoolbreed project?
- Je kunt in een jaar dat er verkiezingen zijn (zoals 2023) scholierenverkiezingen organiseren op je eigen school. Daarbij kun je dan ook enquêtes laten invullen.

Opgave 7

Bekijk **Voorbeeld 2**. Stel dat je een schoolbrede enquête wilt opzetten. Bekijk het practicum **Data verzamelen met een enquête** hoe dit bijvoorbeeld via Google-Drive kan.

- a Bedenk een geschikte set enquêtevragen, vergelijkbaar met die van de dataset die je in het voorbeeld aantreft.
- b Bedenk of je de hele school wilt laten meedoen, of een steekproef wilt trekken. Aan welke voorwaarden moet zo'n steekproef dan voldoen om iets te kunnen zeggen over de leerlingen van jouw school?
- c Verzamel je data in een Excel-bestand.

Opgave 8

Een andere manier om aan data over verkiezingen te komen is door ze te organiseren op de eigen school.

- a Bekijk de website <https://www.scholierenverkiezingen.nl>.
- b Organiseer scholierenverkiezingen en houd een enquête om de gewenste gegevens te verzamelen.
- c Bedenk een manier om de vooraf gepeilde uitslag te vergelijken met de daadwerkelijke uitslag op jullie school na de verkiezingen.

Verwerken

Opgave 9

Je wilt een eigen onderzoek opzetten. Kijk naar de statistische cyclus.

- a Bedenk een onderzoeksvraag bij een statistisch probleem dat jou interesseert.
- b Maak een plan van aanpak en laat dit door je docent controleren.
- c Verzamel je gegevens in een Excel-bestand.



© 2024

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt. Klik op  in de marge bij de betreffende opgave. Uw mailprogramma wordt dan geopend waarbij het emailadres en onderwerp al zijn ingevuld. U hoeft alleen uw opmerkingen nog maar in te voeren.

Email: f.spijkers@math4all.nl

Met de Math4All Foliostraat kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@math4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
