Data en toeval

Prof. dr. Herman Callaert

Statistiek voor de eerste graad:

een statistisch onderzoek

Inhoud

In de eerste graad van het secundair onderwijs komen leerlingen in contact met statistiek. Statistiek is geen onderdeel van wiskunde, het is een apart vak zoals economie of aardrijkskunde.

Niet alle volwassen mannen zijn even groot, niet iedereen kan even snel lopen, treinen rijden niet allemaal even stipt en in zakjes M&M's zitten niet altijd evenveel snoepjes. Variabiliteit is overal. Statistiek is een wetenschap die methoden en technieken aanreikt om met variabiliteit om te gaan. Hierbij gaat abstract redeneren samen met de kunst om data en context juist te interpreteren.

A/B-stroom > Wiskunde – natuurwetenschappen – technologie – STEM > Onderwijsdoel 6.16 / 6.8
"De leerlingen voeren ***een beschrijvend statistisch onderzoek*** *uit met* 20 à *25 zelf verzamelde, niet gegroepeerde gegevens van 1 grootheid"*

Een statistisch onderzoek

Zoeken naar een zinvol antwoord op een concrete vraag leidt in de statistiek tot een statistisch onderzoek. Daarbij doorloop je stapsgewijs een volledig proces. Hieronder zie je de grote onderdelen van zo'n proces.

1. ***Je formuleert een onderzoeksvraag***.
   In statistiek is dat een vraag die kan aangepakt worden met data en waarbij je een antwoord verwacht waarin variabiliteit een rol speelt.
2. ***Je verzamelt data***.
   Je maakt een plan om relevante data te verzamelen en je voert dat plan uit. Je houdt hierbij de context van de onderzoeksvraag in het oog.
3. ***Je analyseert de data***.
   In de eerste graad werk je met methoden van de exploratieve statistiek. In de latere jaren komen ook kansmodellen en verklarende statistiek aan bod. Overal gebruik je ICT bij rekenen en tekenen.
4. ***Je formuleert een conclusie***.
   Een conclusie is zelden "zwart-wit". In statistiek "interpreteer" je de resultaten in de context van de onderzoeksvraag. Je verwoordt ook de variabiliteit in je conclusie.

1. Een onderzoeksvraag formuleren

In dit deel belichten we een drietal aspecten van een onderzoeksvraag. We beginnen met voorbeelden die het onderscheid tussen een wiskundige vraag en een statistische vraag illustreren. Verder tonen we dat je moet starten met een authentieke vraag als je een authentiek onderzoek wil voeren. In een laatste punt geven we aan dat een initiële vraag dikwijls verduidelijking nodig heeft vooraleer je op een goede manier data kan verzamelen. We eindigen met concrete voorbeelden van onderzoeksvragen die je kan gebruiken in de klas.

1.1. Een statistische vraag

Voorbeeld.

Op de leeftijd van 6 à 7 jaar beginnen kinderen hun melkgebit te wisselen voor een blijvend gebit. Lucas zit in het eerste leerjaar en je vraagt hem:

1. hoeveel melktanden ben je al kwijt?
2. hoeveel melktanden zijn de kinderen in je klas al kwijt?

Dit zijn 2 vragen over "een aantal uitgevallen melktanden". Wat is het verschil?

1. Bij de eerste vraag verwacht je een precies en eenduidig antwoord, bijvoorbeeld 4 (Lucas heeft al 4 uitgevallen melktanden). Dit type antwoord verwacht je bij ***een wiskundige vraag.***

2. De tweede vraag is totaal anders van aard. Uit de context weet je dat niet alle kinderen melktanden verliezen op hetzelfde ogenblik. Je kan hier geen "precies en eenduidig" antwoord geven zoals in wiskunde. Het ene kind is al 5 tanden kwijt, een ander nog maar 2 en een derde misschien nog helemaal geen. De tweede vraag is ***een statistische vraag***.

Een statistische vraag verwacht een antwoord waarin variabiliteit een rol speelt.

Voorbeeld.

Als je een aantal getallen hebt en men zegt: "bereken het gemiddelde", dan moet je de som maken van die getallen en de uitkomst delen door het aantal getallen.
Wat is het gemiddelde van de getallen 10, 20, 30?

Is dit een statistische vraag? Waarom?

Dit is geen statistische vraag.
Dit is een wiskundige vraag. Het is een oefening in optellen, tellen en delen. Een statistische naam geven aan een wiskundige procedure maakt er nog geen statistiek van.

Voorbeeld.

Met “de lengte van een woord” bedoelen we “het aantal letters in dat woord”.

Opdracht: welk type vragen staan hieronder (wiskundig of statistisch)?

1. Hoe lang is het woord statistiek?
2. Hoe lang zijn de woorden op dit blad?
3. Gebruiken leerboeken van het vijfde leerjaar langere woorden dan leerboeken van het eerste leerjaar?

De eerste vraag is een wiskundige vraag met uniek antwoord 10 (het woord “statistiek” telt 10 letters).

De tweede vraag is een statistische vraag (exploratieve statistiek). Er is geen eenduidig antwoord. Je moet hier op zoek gaan naar methoden en technieken om de variabiliteit in de lengte van de woorden op dit blad te beschrijven en helder voor te stellen.

De derde vraag is een statistische vraag (verklarende statistiek) waar je op basis van steekproeven uitspraken zal doen over “de populatie van woorden” in boeken van die leerjaren.

Voorbeeld.

In een klas zitten 14 leerlingen. Zij hebben elk een zakje M&M snoepjes. Die snoepjes hebben verschillende kleuren: blauw, bruin, geel, groen, oranje en rood. De leerlingen tellen hoeveel rode snoepjes zij hebben. Bob heeft er 11, Liam heeft er 10, enz. Hiernaast zie je hoeveel rode snoepjes elke leerling heeft.

BOB	11
LIAM	10
EMMA	8
ADAM	6
NOOR	12
STAN	8
LUCAS	7
LARS	6
LENA	11
MATS	7
FIEN	13
KOBE	11
LIEN	12
MILA	7

Opdracht.

1. Stel een “wiskundige vraag” over de data die je hier ziet.
2. Stel een “statistische vraag” over de data die je hier ziet.

Wiskundige vragen kunnen zijn:
- Hoeveel rode snoepjes heeft Lars?
- Wie heeft er de meeste rode snoepjes?
- Hoeveel rode snoepjes zijn er in totaal?

Statistische vragen kunnen zijn:
- Hoeveel rode snoepjes zitten er in een M&M zakje?
- Als je opnieuw een zakje M&M’s zou krijgen, verwacht je dan hetzelfde aantal rode snoepjes te vinden? Waarom?
- Lars zegt dat er geen “gewone” zakjes zijn, bijna iedereen heeft ofwel veel ofwel weinig rode snoepjes. Geef jij Lars gelijk? Waarom?

Leerlingen weten wat een statistische vraag is en kunnen zelf statistische vragen formuleren.

1.2. Een authentieke vraag

Een goede statisticus, net zoals een goede musicus,
ben je niet, als je alleen maar techniek beheerst.

David Moore (1998)

Tegenvoorbeeld.

Een centraal examen wiskunde in Engeland bevat vragen over statistiek. Per vraag wordt aangegeven welke competentie er getoetst wordt. Er is ook een scoresleutel opgesteld als leidraad voor iedereen die meewerkt bij het verbeteren van al die examens.

Vraag: "Het gemiddelde aantal goals bij 20 matchen was 4 goals per match. Hoeveel goals zijn er in totaal gescoord bij deze 20 matchen?".
Bedoeling: toetsen van: "*de leerling begrijpt en gebruikt het gemiddelde van discrete data*".
Scoresleutel: "correct antwoord" als de leerling het getal 80 in het antwoordvakje heeft ingevuld.

Nota.
Deze vraag toetst helemaal niet of de leerling het gemiddelde "*begrijpt*". Men toetst enkel of de leerling het gemiddelde kan "berekenen". Om in het antwoordvakje 80 te kunnen invullen, moet je alleen maar weten dat "gemiddelde" *een synoniem is* voor "de som gedeeld door het aantal". Je hoeft ook niets te kennen van voetbal. Een gelijkaardige vraag maar dan zonder context zoals: "Het gemiddelde van 20 getallen is 4, bereken de som." is even efficiënt om niet te weten te komen dat de leerling met het juiste antwoord misschien wel denkt dat het gemiddelde een maat voor spreiding is.

Voorbeeld. [Arm aan statistische ideeën. Technisch oefenen van "definitie–>formule–>berekening".]

In de voorbije 75 jaar zijn er 8 topscorers in geslaagd om in de Belgische Eerste Klasse (Jupiler Pro League) meer dan 30 goals te scoren in één seizoen. Deze 8 voetballers hadden de volgende topscore: 40, 48, 47, 35, 35, 35, 39, 31. Voor deze scores kan je berekenen dat het gemiddelde 38.8 is en de mediaan 37. Nadat deze berekeningen waren uitgevoerd zag men dat er bij het overschrijven een fout stond in die scores. Het getal 47 moet 37 zijn. Welk kengetal zal door deze fout het meest beïnvloed worden, het gemiddelde of de mediaan?

Nota. Voor de juiste scores is het gemiddelde 37.5 en de mediaan 36. Beide kengetallen dalen, het gemiddelde het meest. Zoals bij het vorige voorbeeld heb je ook hier niets aan de context. Van de leerling wordt verwacht dat hij zo snel mogelijk overstapt op definitie $\rightarrow$ formule $\rightarrow$ rekenen.

> *Bij een authentieke onderzoeksvraag dient de context niet om weg te gooien.*

Voorbeeld. [Vraag die stimuleert om zelfstandig aan de slag te gaan met statistische ideeën en technieken.]

Bij een wedstrijd tussen scholen wordt de 100 meter voor vrouwen gelopen.
In jouw school zijn er 3 leerlingen die in deze afstand uitblinken. Je ziet hier de tijden (in sec) die zij recent in 7 oefenwedstrijden haalden. Jij mag maar één leerling naar die wedstrijd sturen. Wie selecteer je en waarom?

Amber	14.49	14.71	15.26	15.68	14.75	15.14	14.36
Emma	14.98	14.84	15.17	14.62	14.69	14.41	14.49
Fiebe	14.41	15.44	14.78	15.61	14.98	15.83	14.61

Nota. Context, data, rekenen, tekenen… en redeneren komen hier aan bod. Je doorloopt een heel proces dat uiteindelijk leidt tot een conclusie waarbij je motiveert wie je selecteert.

1.3. Een precieze vraag

Soms kan je in een statistisch onderzoek gebruik maken van data die door anderen zijn verzameld. Als je weet hoe dat is gebeurd en als je de data kan vertrouwen, dan kan je hierop verder bouwen. In de meeste gevallen is "zelf data verzamelen" onderdeel van het onderzoek. Zelfs met veel ervaring kom je hierbij toch nog onvoorziene situaties tegen. Om dan toch op een goede manier verder te kunnen werken moet je dikwijls de vraag meer preciseren.

> *"Preciseren van de onderzoeksvraag"* en *"data verzamelen"*
> is dikwijls een heen–en–weer proces.

Voorbeeld.

Een vraag zoals: "Hoeveel boeken sleuren leerlingen mee naar school?" is te breed voor leerlingen van de eerste graad. Zij zijn nog niet vertrouwd met steekproeven om daarna uitspraken te doen over "leerlingen" in het algemeen. Je kan hier de onderzoeksvraag beperken tot "Hoeveel boeken hebben de leerlingen van mijn klas vandaag meegebracht?". Als je daarbij afspreekt wat meetelt als "boek" (handboek, ringmap, schrift, atlas, agenda…) dan kan je van start gaan. Nu weet je hoe je de data zal verzamelen voor de meer gepreciseerde onderzoeksvraag.

Voorbeeld.

Je krijgt de opdracht om "het geluksgevoel" te onderzoeken bij de bewoners van een woonzorgcentrum in Brugge.
Hoe pak je dat aan?

- Vraag je om op een schaal van 0 tot 10 aan te geven hoe gelukkig men zich voelt?
- Is het de bedoeling om meerdere criteria (voeding, omgeving, eenzaamheid…) te bevragen?
- …

Hier moet je samen met de opdrachtgever komen tot een meer precieze formulering van de vraag.

Voorbeeld.

Bij enquêtes kan al bij het stellen van de vraag heel wat fout gaan. Wist je dat je tot een verschillende conclusie komt als je in je enquête de vraag stelt "Is het verkeer een grotere luchtvervuiler dan de industrie?" dan als je de vraag formuleert als "Is de industrie een grotere luchtvervuiler dan het verkeer?".
Het opstellen en afnemen van enquêtes is op zichzelf al een uitgebreide topic in de statistiek.

1.4. Voorbeeldvragen

Zelf onderzoeksvragen formuleren is voor veel leerlingen nieuw.

In het begin kan het goed zijn dat onderzoeksvragen de kans krijgen om te "groeien" in klasverband:
- laat leerlingen zelf vragen *formuleren*
- laat leerlingen een geformuleerde vraag *beoordelen*, *verfijnen*, *aanvullen,…*
- ….

Het kan handig zijn dat leerlingen wat ideeën aangereikt krijgen. Je kan bijvoorbeeld starten met een lijstje van brede thema's zoals:

- sport - muziek - milieu - pesten - verkeer - kunst	- vriendschap - reizen - voeding - biologie - klimaat - stress	- verdriet - gamen - feesten - vakantie - mode - ……..

Wat wil je bij zo'n thema te weten komen? Er zijn heel wat mogelijkheden.

- Wil je dingen opmeten en beschrijven?
  - hoeveel?
  - hoe dikwijls?
  - hoelang?
- Wil je voorkeuren en meningen te weten komen?
  - wat is je favoriete…?
  - ben je akkoord met…?
  - wat vind je de beste manier om…?
- Wil je dingen vergelijken?
  - is er een verschil tussen… en…?
  - is er een verandering voor… en na …?
  - is er een verband tussen … en…?
- Wil je veralgemenen?
  - wat is hier typisch …?
  - kan je voorspellen hoe …?
  - is er hier een algemene trend …?

Voorbeeld.

Een reclamestunt van een pop-up ijssalon zegt dat je bij de opening gratis een potje ijs krijgt. Je mag daarbij kiezen uit: vanille, chocolade, aardbei of banaan.
Vraag: "Welke ijssmaak verkiezen de leerlingen van mijn klas bij deze reclamestunt?".

Thema = voeding, vakantie,.. (ijsjes).
Bedoeling = voorkeuren van mijn klasgenoten noteren en beschrijven.

Voorbeeld.

In België heeft 46 % van de mensen bloedgroep O, 42 % heeft bloedgroep A, 9 % heeft B en slechts 3 % heeft AB. Je wil weten wat de bloedgroep van je medeleerlingen is.
Vraag: "Welke bloedgroep hebben de leerlingen van mijn klas?".

Thema = biologie, geneeskunde
Bedoeling = bloedgroep van mijn klasgenoten noteren en beschrijven.

Voorbeeld.

Je verjaardag ken je wel, maar weet je ook op welke dag van de week je geboren bent? Was dat een dinsdag of een vrijdag of…? Je denkt waarschijnlijk dat er ongeveer evenveel geboorten zijn op eender welke dag van de week. Is dat waar?
Vraag: "Op welke dag van de week zijn de leerlingen van mijn klas geboren?".

Thema = verjaardagen, feesten.
Bedoeling = geboortedag van mijn klasgenoten noteren en beschrijven.

Voorbeeld.

Op het internet heb je een app gevonden om je reactiesnelheid te testen. Je denkt dat je daar uitzonderlijk goed in bent. Maar is dat echt zo? Hoe snel zouden je medeleerlingen zijn?
Vraag: "Welke score halen de leerlingen van mijn klas als zij met die app hun reactiesnelheid testen?".

Thema = sport of verkeer of gamen (reactiesnelheid).
Bedoeling = reactietijden opmeten en beschrijven.

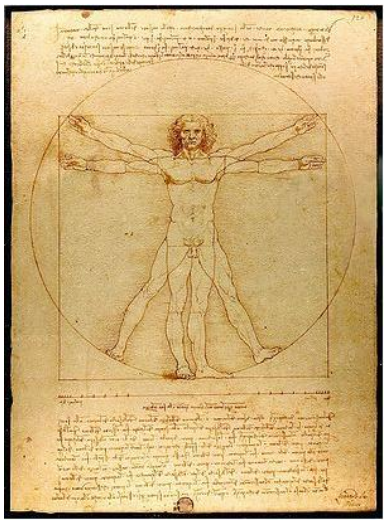
Bij het formuleren van onderzoeksvragen is het niet de bedoeling dat alle vragen opgelost kunnen worden met de exploratieve statistische technieken van de eerste graad. Ook in de eerste graad mogen onderzoeksvragen gesteld worden die breed zijn en aandacht hebben voor vergelijken, voorspellen, veralgemenen… zoals:

- stijgt de "hartslag per minuut" na een sprintje van 100 meter? [veranderlijken vergelijken]
- als in een zakje met 50 M&M snoepjes er 10 rode zitten, worden er dan 20 % rode snoepjes gemaakt in het totale productieproces van M&M's? [van steekproef naar populatie]

Voorbeeld.

Leonardo Da Vinci was een kunstenaar en een wetenschapper. Hij kende veel van de bouw van het menselijk lichaam en beschreef allerlei soorten verhoudingen. Die kennis is handig als je mensen wil schilderen of beeldhouwen. Hij beweerde:

- dat je geknield nog drie kwart bent van je totale lichaamslengte
- dat de lengte van je hand een negende is van je lichaamslengte
- dat de spanwijdte van je volledig uitgestrekte armen gelijk is aan je lichaamslengte.



Da Vinci baseerde zich voor die verhoudingen op een "ideaal lichaam" van een volwassen man. Zou het bij leerlingen van de eerste graad ook waar zijn dat je uit de spanwijdte van je volledig uitgestrekte armen kan afleiden hoe groot je bent?

Thema = kunst, anatomie.
Bedoeling = vergelijken, voorspellen.

2. Data verzamelen

Om te weten welke data je moet verzamelen kijk je naar de onderzoeksvraag. Maar dat is niet genoeg. Je moet ook weten op welke manier je die data moet verzamelen. Het heen–en–weer proces tussen "verder preciseren van de onderzoeksvraag" en "een plan opstellen om data te verzamelen" komt hier terug aan bod.

2.1. Een plan opstellen

Voorbeeld.

Een reclamestunt van een pop-up ijssalon zegt dat je bij de opening gratis een potje ijs krijgt. Je mag daarbij kiezen uit: vanille, chocolade, aardbei of banaan.
Vraag: "Welke ijssmaak verkiezen de leerlingen van mijn klas bij deze reclamestunt?".

Een plan opstellen om hier data te verzamelen lijkt eenvoudig: "elke leerling zegt welke ijssmaak hij/zij verkiest en de leerkracht noteert het antwoord".

Plannen om data te verzamelen lijken soms eenvoudig… tot je eraan begint. Dan ontdek je dat je toch wat extra afspraken nodig hebt zoals:

- moet de klas stil zijn of mogen medeleerlingen commentaar geven? De eerste leerling zegt "banaan" waarop medeleerlingen roepen "bah, dat meen je niet, zo slecht!". De volgende leerling ging ook "banaan" zeggen maar durft dat nu niet meer. Zij zegt "aardbei".
- wat doe je met een leerling die niet wil antwoorden omdat hij alleen straciatella lust? Voorzie je "geen antwoord" ook als een mogelijk antwoord in je onderzoek?

Voorbeeld.

"Hoe goed kunnen leerlingen de tijdsduur van een minuut schatten?"

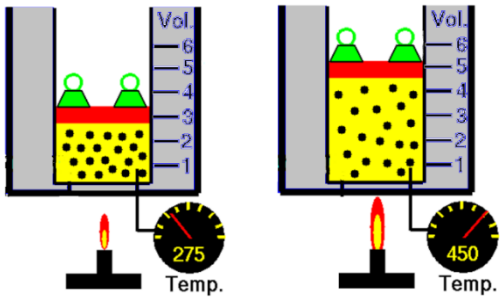
Zonder verdere precisering kan je hier geen plan opstellen.

- hoe moeten de leerlingen de tijdsduur van een minuut schatten: rechtstaand, geblinddoekt,…?
- wie gaat dit opmeten en hoe?

Meer info over dit onderzoek vind je in "Een statistisch onderzoek naar het schatten van de tijdsduur van 1 minuut" op https://www.uhasselt.be/lesmateriaal-statistiek (klik op "*Werkteksten*", scrol naar "1.*Exploratieve statistiek*" en klik op "*Exploratieve statistiek. Werktekst voor de leerling.*").

Voorbeeld.

De volumewet van Gay-Lussac zegt dat het volume van een (ideaal) gas recht evenredig is met de temperatuur wanneer je de druk en de massa constant houdt. Op bijgaande figuur zie je een illustratie van deze gaswet waarbij een vaste massa gas gevangen zit in een container waarop een vaste druk wordt uitgeoefend. Voor dit labo–experiment werd een volume van 280 milliliter opgetekend bij een temperatuur van 275 Kelvin en vergrootte het volume tot 460 ml wanneer de temperatuur steeg tot 450 K.



Bij laboratoriumproeven in de natuurwetenschappen hoort meestal een hele procedure: hoe je de proef moet opstellen, welke producten je moet gebruiken, welke meettoestellen je moet aankoppelen…
Het plan om data te verzamelen is hier al (grotendeels) vastgelegd in de labo-handleiding waarbij het de bedoeling is om de wet van Gay-Lussac experimenteel te verifiëren.

2.2. De dataset

Bij een statistisch onderzoek is een dataset niet zomaar een hoop gegevens. De onderzoeksvraag samen met een plan om de data op te meten, zeggen wat je moet opmeten en op welke manier je dat moet doen. Die opmetingen schrijf je dan neer in een schema met een duidelijke structuur. Dat noemen we een dataset.

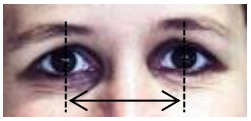
Hoe je dat schema moet opstellen, leer je in de tekst: “Soorten data en de structuur van een dataset” op: https://www.uhasselt.be/lesmateriaal-statistiek (klik op *Werkteksten* en scrol naar *4.Methoden en technieken bij een statistisch onderzoek – Soorten data en de structuur van een dataset*).

2.3. Data cleaning

Data cleaning (controle van de data) hoort ook thuis in een statistisch onderzoek. Als je weet dat bepaalde waarden niet mogelijk zijn, zoals 516 voor de lengte (in cm) van een baby, dan moet je dit getal controleren vooraleer je aan je onderzoek begint. Misschien is er bij het intikken een decimaal punt vergeten en moest er 51.6 in die databank staan.

Voorbeeld.

De pupilafstand is de afstand (in mm) tussen de pupillen van je ogen wanneer je rechtdoor in de verte kijkt. Die afstand wordt gebruikt om een bril aan te passen. Bij volwassenen is de pupilafstand ongeveer 62 à 63 mm maar er zit heel wat variabiliteit op, zowat van 51 mm tot 77 mm.



Je wordt gevraagd om mee te werken aan een onderzoek over de pupilafstand. Men zal je daarvoor 17 opgemeten pupilafstanden elektronisch doorsturen. In het Excel bestand vind je de volgende data (in mm): 62 58 63 67 71 62 64 62 59 15 57 62 61 69 59 63.
Zal je hier onmiddellijk beginnen met de analyse van deze data?

Er zit een onmogelijk getal tussen. Er is geen enkele volwassene met een pupilafstand van 15 mm, dat bestaat niet! Je vermoedt dat iemand 15 heeft getikt in plaats van 51. Fouten in een databank, die door anderen is opgesteld, mag je zomaar niet op eigen houtje aanpassen. Je moet contact opnemen met de onderzoekers en vragen om dat getal te controleren. Je kan tegelijkertijd ook zeggen dat zij je een dataset van 17 pupilafstanden hadden beloofd en dat er maar 16 zijn toegekomen.

2.4. Voorbeelden van een dataset

Hieronder zie je verschillende voorbeelden. Er is een dataset met 1 veranderlijke, een dataset met 2 veranderlijken en een zogenaamde "afgeleide" dataset.

Bloedgroepen: een dataset met 1 veranderlijke.

Om te weten wat de bloedgroep van je medeleerlingen is noteer je gewoon bij iedereen de bloedgroep. Je hebt dan 22 resultaten.

Je kan de data opschrijven als:

- A A O B A A O O O A A A O O A OB A O O A A

Als er in deze dataset een tikfout zit (zoals bloedgroep OB die niet bestaat) dan mag je daar zomaar niet zelf iets anders van maken. Je denkt misschien dat het AB moest zijn, maar hoe weet je dat? Als je correct wil zijn, dan zit er niets anders op dan te herbeginnen en opnieuw bij alle 22 leerlingen hun bloedgroep op te vragen.

Een volledige dataset geeft meer informatie en ziet er als volgt uit (BLG = bloedgroep):

Naam	BLG
BOB	O
LIAM	A
EMMA	A
ADAM	O
NOOR	B
STAN	A

Naam	BLG
JEF	A
LARS	O
LENA	O
MATS	O
FIEN	A
KOBE	A

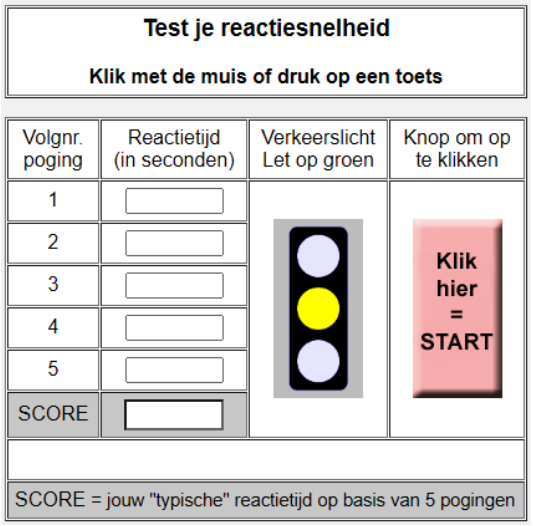
Naam	BLG
LIEN	A
MILA	O
JAN	O
LUCAS	A
ELLA	OB
PIET	A

Naam	BLG
ANN	O
KRIS	O
DRIES	A
YVES	A

Als er nu in de dataset een tikfout zit (zoals bloedgroep OB) dan kan je die snel herstellen. De dataset vertelt je dat de fout bij Ella zit. Zij zegt dat ze bloedgroep B heeft en dus verander je OB in B.

Reactiesnelheid: een dataset met 2 veranderlijken.

Je wil een beeld krijgen van de reactiesnelheid van de leerlingen in je klas. Je spreekt af dat je op een laptop de app installeert: "Test je reactiesnelheid" (app met verkeerslichten). Als het licht van rood op groen springt moet je zo snel mogelijk "klikken". Dat "klikken" kan zowel een klik met de muis zijn als een druk op een toets. Na 5 pogingen berekent de app jouw "typische" reactietijd. Dat is je score.



Terwijl je deze app bekijkt kom je op het idee om de oorspronkelijke onderzoeksvraag uit te breiden. Je wil de reactiesnelheid van je medeleerlingen onderzoeken maar je wil ook weten of de reactietijd anders is bij "klikken met de muis" dan bij "drukken op een toets".

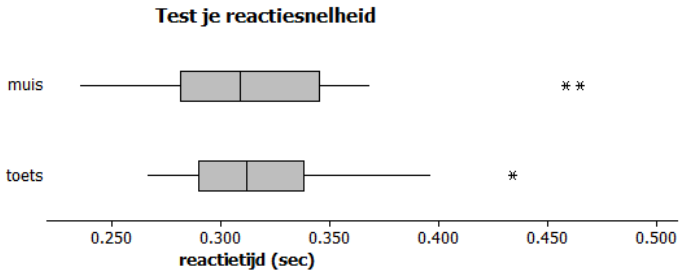
Je vraagt aan alle leerlingen om beide testen te doen. De resultaten noteer je in één dataset. Die zou er als volgt kunnen uitzien:

Naam	Muis	Toets
BOB	0.316	0.302
LIAM	0.368	0.354
EMMA	0.285	0.319
ADAM	0.268	0.266
NOOR	0.253	0.270
STAN	0.368	0.305
JEF	0.465	0.331
LARS	0.311	0.396

Naam	Muis	Toets
DRIES	0.343	0.327
YVES	0.307	0.285
LENA	0.292	0.319
MATS	0.327	0.341
FIEN	0.282	0.267
KOBE	0.280	0.329
ANN	0.284	0.291
KRIS	0.319	0.292

Naam	Muis	Toets
LIEN	0.235	0.269
MILA	0.270	0.374
JAN	0.294	0.305
LUCAS	0.458	0.303
ELLA	0.349	0.337
PIET	0.344	0.434

Nota.
De reactiesnelheid is een numerieke veranderlijke die continu is. In de eerste graad kunnen leerlingen wel al kengetallen zoals een gemiddelde of een mediaan berekenen. Een verdere analyse met bv. een boxplot (zoals hiernaast) of met een histogram kunnen leerlingen nog niet doorvoeren.



In dit onderzoek is (alles in sec) $\bar{x} = 0.3190$ en $Me = 0.309$ bij klikken met de muis en $\bar{x} = 0.3189$ en $Me = 0.312$ bij drukken op een toets.

Nota over het gebruik van de app "Test je reactiesnelheid".
Ga naar de website https://www.uhasselt.be/lesmateriaal-statistiek klik op *Werkteksten*, scrol naar "1.*Exploratieve statistiek*" en klik op *Reactietijd*. Het bestand reactietijd.zip wordt dan gedownload naar je PC. Een rechterklik op reactietijd.zip samen met "Alles uitpakken…" levert 10 bestanden. Zorg ervoor dat die 10 bestanden in eenzelfde map staan. Vanaf nu kan je de app gebruiken op je PC zonder verbinding met het internet: dubbelklik "reactietijd.html".

Referentie: De app is gebaseerd op: The Online Reaction Time Test, © 2002 by Jim Allen, gywh.com.

Dag van de week waarop kinderen geboren worden: een "afgeleide" dataset.

Voor een statistisch onderzoek moet je niet altijd zelf de data verzamelen. Er zijn heel veel data beschikbaar op het internet. Die data zijn meestal al op een of andere manier bewerkt en "afgeleid" van een oorspronkelijke dataset.

Bij de geboorte van een kind worden heel veel dingen genoteerd zoals datum, geslacht, bloedgroep…. Hieronder links zie je een voorbeeld van een stukje uit zo'n dataset.

ID	DATUM	GESLACHT	BLOEDGROEP	LENGTE	GEWICHT	….
….	….	….	….	….	….	….
….	12/04/2018	J	AB	55.9	3720	….
….	12/04/2018	M	O	50.8	3180	….
….	13/04/2018	M	O	50.8	2990	….
….	13/04/2018	J	A	50.8	2900	….
….	13/04/2018	M	A	55.9	4350	….
….	13/04/2018	M	AB	49.5	2770	….
….	….	….	….	….	….	….

Geboorten

Datum	Aantal
8/04/2018	192
9/04/2018	360
10/04/2018	359
11/04/2018	346
12/04/2018	370
13/04/2018	359

Je kan deze dataset gebruiken om te tellen hoeveel keer een bepaalde datum (zoals 12/04/2018) voorkomt. Zo weet je hoeveel kinderen er op die dag geboren zijn. Op die manier krijg je een frequentietabel die eruitziet zoals hierboven rechts. Daar zie je een stukje uit een Excel bestand dat je vindt op de volgende website van de Federale Overheidsdienst (FOD) Economie: https://data.gov.be/nl/dataset/d2843e3731ffda68e8001ada663d4627634fc586 .

Je kan nu van "datum" overstappen op "dag van de week". Zo levert de tabel van de FOD Economie de volgende frequentietabel voor de geboorten in België over een periode van 20 jaar:

"Op welke dag van de week zijn kinderen geboren in België tussen 1/1/2000 en 31/12/2019?" is een onderzoeksvraag die je met de "afgeleide" dataset hiernaast (het is eigenlijk een frequentietabel) kan aanpakken.

Geboorten in België [1 jan 2000 – 31 dec 2019]

Geboortedag	Aantal geboorten op deze dag
maandag	379 633
dinsdag	409 310
woensdag	387 202
donderdag	393 495
vrijdag	394 982
zaterdag	228 914
zondag	224 827

Als je alleen maar met de leerlingen van je klas werkt, dan begin je met een klassieke dataset:

Naam	Dag
Alexander	dinsdag
Louis	zondag
Audric	woensdag
Remi	zondag
Charles	maandag
Augustin	zaterdag

Naam	Dag
Arnaud	maandag
Walter	dinsdag
Gilles	woensdag
Nicholas	woensdag
Mathias	vrijdag
Henri	zaterdag

Naam	Dag
Maurice	zondag
Mathias	woensdag
Amélie	vrijdag
Orige	woensdag
Giulia	woensdag
Maité	maandag

Naam	Dag
Tine	woensdag
Lizz	zaterdag
Anna	maandag
Manon	woensdag

3. De data analyseren

3.1. Soorten veranderlijken

Welke statistische methoden je allemaal kan gebruiken, hangt voor een deel af van het soort veranderlijke waarover je beschikt.
In het secundair onderwijs werk je met 2 soorten veranderlijken: categorische en numerieke.

Categorische veranderlijken hebben waarden die in categorieën terechtkomen.
Deze waarden hoeven zich niet te lenen tot "wiskundige bewerkingen".
Voorbeeld: de kleur van M&M-snoepjes.

Numerieke veranderlijken hebben waarden die numeriek (= getallen) zijn.
De waarden zijn getallen en daarop zijn "wiskundige bewerkingen" mogelijk.
Voorbeeld: de lengte (in cm) van een kind bij de geboorte.

Een bespreking van de soorten data vind je in de tekst: "Soorten data en de structuur van een dataset" op: https://www.uhasselt.be/lesmateriaal-statistiek (klik op *Werkteksten* en scrol naar *4.Methoden en technieken bij een statistisch onderzoek – Soorten data en de structuur van een dataset*).

3.2. De gereedschapskist

Statistiek reikt, ten behoeve van andere wetenschappen, methoden aan om op een juiste manier data te verzamelen en daaruit zinvolle informatie te halen. Statistiek is een experimentele discipline die geen eigen data heeft maar werkt met data van andere disciplines.

Met data werken is *zowel een kunst als een wetenschap*: abstract redeneren gaat hier samen met de kunst om data en context juist te interpreteren.

Statistiek heeft eigen methoden en technieken en maakt bovendien gebruik van tools uit:
- **wiskunde**: abstract redeneren, rekenvaardigheden, …
- **ICT**: datamanipulatie, statistische pakketten, …
- **visuele perceptie**: tabellen, grafieken, …
- **communicatie**: vanaf het formuleren van de vraag tot aan de conclusie en het verslag
- **domein–specifieke basiskennis uit**: economie, geneeskunde, pedagogie, biologie…

In de eerste graad beginnen leerlingen met een beperkt aantal tools. Om een statistische analyse te kunnen uitvoeren (zoals hieronder in de voorbeelden 3.3 en 3.4) moeten leerlingen, op een eenvoudig niveau, over tools beschikken zoals:
- rekentechnieken (bewerkingen, ordenen en tellen, proporties, …)
- tekentechnieken (staafdiagram, cirkeldiagram, dotplot, lijndiagram, …)
- ICT-vaardigheden bij tekenen en rekenen
- communicatievaardigheden (formuleren van een conclusie, van onderzoeksstappen, …)

3.3. Analyse van een categorische veranderlijke (nominaal)

Bij de vraag “Welke bloedgroep hebben de leerlingen van mijn klas?” werk je met de gecorrigeerde dataset die er als volgt uitziet:

Naam	BLG
BOB	O
LIAM	A
EMMA	A
ADAM	O
NOOR	B
STAN	A

Naam	BLG
JEF	A
LARS	O
LENA	O
MATS	O
FIEN	A
KOBE	A

Naam	BLG
LIEN	A
MILA	O
JAN	O
LUCAS	A
ELLA	B
PIET	A

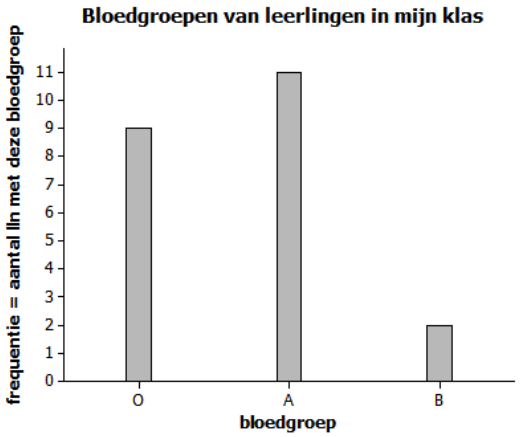
Naam	BLG
ANN	O
KRIS	O
DRIES	A
YVES	A

Om een beter zicht te krijgen op deze dataset kan je een frequentietabel opstellen die toont hoeveel leerlingen een bloedgroep van een bepaalde soort hebben.

Deze frequentietabel is veel overzichtelijker dan de oorspronkelijke dataset. Hij geeft je een goed leesbaar antwoord op de vraag hoe de bloedgroepen verdeeld zijn bij de leerlingen van je klas.

Bloedgroep	Frequentie = aantal leerlingen met deze bloedgroep
O	9
A	11
B	2

Je kan de data ook grafisch voorstellen met een staafdiagram zoals hieronder.



Dit onderzoek heeft ook een bredere context. De oorspronkelijk vraag begon met: “In België heeft 46 % van de mensen bloedgroep O, 42 % heeft bloedgroep A, 9 % heeft B en slechts 3 % heeft AB”.

Bloedgroep AB komt dus voor in België. Je zou in een klas kunnen zitten waar een leerling die bloedgroep AB heeft. In jouw klas is dat niet het geval en dat kan je expliciet weergeven. Hiervoor maak je een frequentietabel die toont wat er is maar ook wat er niet is. Zoiets doe je alleen maar omdat de context zegt dat AB mogelijk is. Je doet dit bijvoorbeeld niet bij de kleuren van M&M-snoepjes. Als je weet dat alleen blauw, bruin, geel, groen, oranje en rood mogelijke kleuren zijn, dan maak je geen frequentietabel waar je ook paars aan toevoegt om dan te zeggen dat er 0 paarse snoepjes zijn.

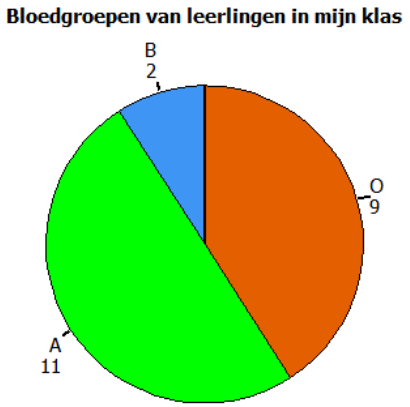
Bloedgroep is een categorische veranderlijke met waarden A, B, AB en O. Deze waarden hebben geen natuurlijk volgorde. In veel teksten over bloedgroepen zie je dat men de volgorde O, A, B, AB gebruikt. Dit "gebruik" nemen we over in de frequentietabel en bij het tekenen van grafieken.

Bloedgroep	Frequentie = aantal leerlingen met deze bloedgroep
O	9
A	11
B	2
AB	0

Nota.
Categorische veranderlijken waarbij de waarden geen natuurlijke volgorde hebben, worden nominale categorische veranderlijken genoemd.

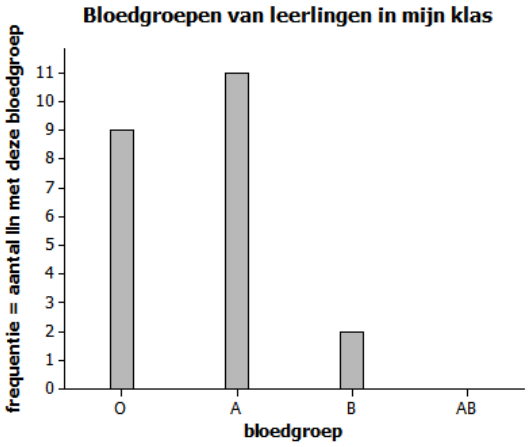
Om bovenstaande frequentietabel grafisch voor te stellen, kies je een grafiek die zo duidelijk mogelijk is.

Studies uit het domein van de perceptiepsychologie zeggen dat een cirkeldiagram (of taartdiagram = pie chart) zelden een goede keuze is om data grafisch voor te stellen. Hoeken van sectoren vergelijken is voor het menselijk oog moeilijker dan hoogteverschillen zien bij staafjes. Bovendien is er in een cirkeldiagram geen plaats om categorieën met frequentie nul voor te stellen. Dat zie je hiernaast.



Het is wel belangrijk dat leerlingen cirkeldiagrammen kunnen lezen want ze komen nog veel voor. Wanneer je echter zelf een keuze kan maken om je data grafisch voor te stellen, dan is een cirkeldiagram meestal af te raden.

Hiernaast zie je een staafdiagram voor de verdeling van de bloedgroepen. Op deze figuur is het duidelijk dat bloedgroep AB een mogelijke bloedgroep is maar dat niemand in je klas die bloedgroep heeft.



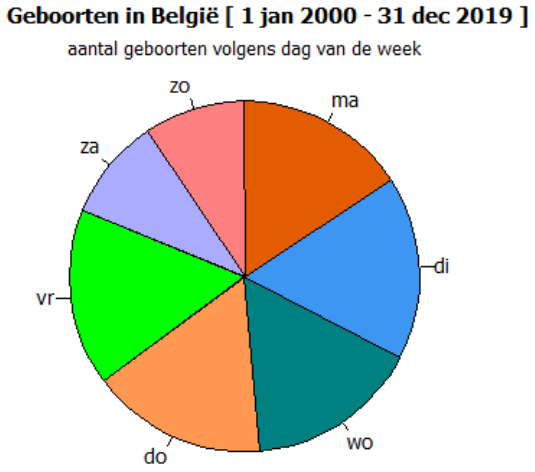
3.4. Analyse van een categorische veranderlijke (ordinaal)

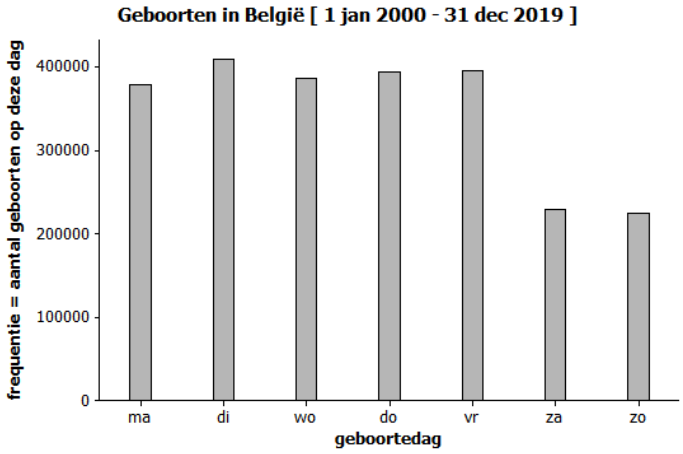
De vraag “Op welke dag van de week worden kinderen geboren?” heb je opgesplitst in 2 delen:
- een vraag over de kinderen in België tussen 1/1/2000 en 31/12/2019
- een vraag over de leerlingen in je klas.

Bemerk dat de dagen van de week een natuurlijke volgorde hebben. Die gebruik je bij het opstellen van een frequentietabel en bij grafische voorstellingen van de data.

Nota.
Categorische veranderlijken waarbij de waarden een natuurlijke volgorde hebben, worden ordinale categorische veranderlijken genoemd.

Voor de kinderen in België heb je hierboven al een frequentietabel opgesteld. De informatie in die tabel kan je ook grafisch voorstellen. Het is hier terug duidelijk dat een cirkeldiagram moeilijker te lezen is dan een staafdiagram. Ter informatie staan hieronder beide figuren, maar zelf kies je voor het staafdiagram.

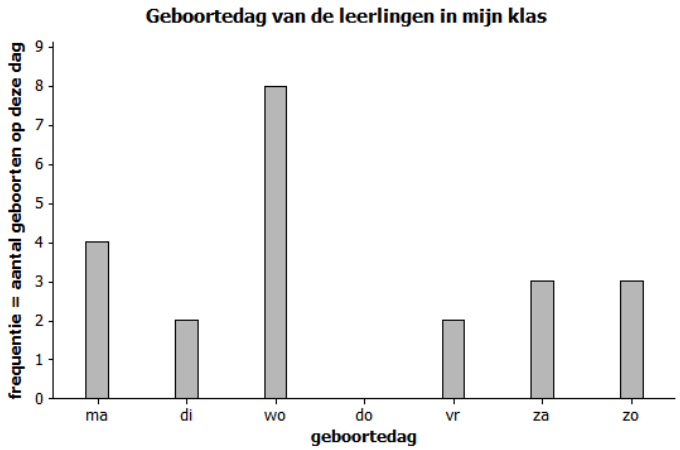




De dataset die je hebt opgesteld voor de leerlingen van je klas kan je samenvatten in een frequentietabel en voorstellen met een staafdiagram.

Geboortedag van de leerlingen in mijn klas

Geboortedag	Frequentie = aantal geboorten op deze dag
maandag	4
dinsdag	2
woensdag	8
donderdag	0
vrijdag	2
zaterdag	3
zondag	3



3.5. Analyse van een numerieke veranderlijke

“Hoe lang zijn de namen van de leerlingen in mijn klas” is een statistische vraag. Daarbij voorzie je dat je antwoord rekening zal moeten houden met variabiliteit want… niet alle namen zijn even lang. Een volledig uitgewerkt onderzoek bij deze vraag vind je op KlasCement in de tekst “Een statistisch onderzoek: Uitgewerkte voorbeelden”. Ga naar https://www.klascement.net en vul daar de zoekterm “herman callaert” in.

Nota.
Met “lengte van een naam” bedoel je het aantal letters in die naam. Aantallen zijn getallen. “Lengte van een naam” is een numerieke veranderlijke.

“Hoe kiezen leerlingen in mijn klas lukraak een getal tussen 1 en 10?” is een andere vraag. Het is een statistische vraag. Je voorziet immers dat niet iedereen eenzelfde getal kiest en over die variabiliteit zal je iets moeten zeggen in je antwoord. “Het lukraak gekozen getal” is een numerieke veranderlijke met waarden die getallen zijn.

Om dit onderzoek uit te voeren spreek je af dat leerlingen onafhankelijk van elkaar een getal opschrijven (een geheel getal, tussen 1 en 10, met 1 en 10 inbegrepen). In jouw klas krijg je dan bv. de volgende dataset:

Naam	Getal
BOB	9
LIAM	3
EMMA	8
ADAM	1
NOOR	5
STAN	2

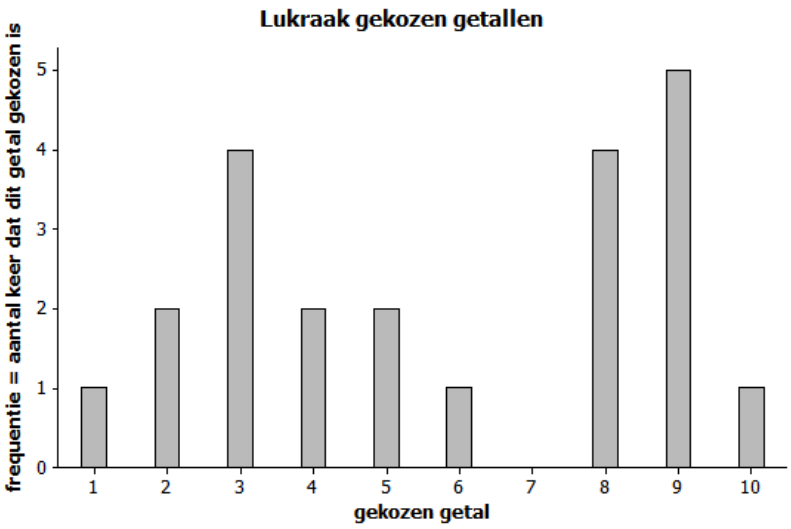
Naam	Getal
JEF	3
LARS	4
LENA	8
MATS	4
FIEN	9
KOBE	8

Naam	Getal
LIEN	9
MILA	5
JAN	2
LUCAS	9
ELLA	3
PIET	8

Naam	Getal
ANN	3
KRIS	6
DRIES	10
YVES	9

Deze dataset kan je samenvatten in een frequentietabel en grafisch voorstellen met een staafdiagram.

Gekozen getal	Frequentie = aantal keer dat dit getal gekozen is
1	1
2	2
3	4
4	2
5	2
6	1
7	0
8	4
9	5
10	1



De 22 leerlingen van je klas hebben lukraak een getal gekozen. Het gemiddelde van die 22 getallen is gelijk aan 5.8 en de mediaan is 5.5.

4. Een conclusie formuleren

4.1. Het grotere kader

Een conclusie gaat over het uitgevoerde onderzoek. Bij grote statistische studies hoort een volledig rapport, inclusief de dataset. Samen met de besluiten verwacht je daar ook antwoorden op de contextvragen:

1. *Waarom* is dit onderzoek uitgevoerd? (Wie wil wat weten?)
2. *Waar* is dit onderzoek uitgevoerd? (In het buitenland? In mijn gemeente?)
3. *Wanneer* is dit onderzoek uitgevoerd? (Vorige eeuw? Dit jaar?)
4. *Wie of wat* is er onderzocht? (Wat zijn de "elementen" in de studie?)
5. *Wat* is er opgemeten? (Wat zijn de "veranderlijken"?)
6. *Hoe* is men te werk gegaan? (Hoe is de steekproef getrokken? Hoe is er gemeten?)

De conclusie van een statistisch onderzoek staat of valt bij de manier waarop de data tot stand zijn gekomen.

Zo'n rapport zorgt ervoor dat anderen, ook later, de studie kunnen beoordelen.

In de eerste graad, bij een eerste kennismaking met een statistisch onderzoek, volstaat het dat leerlingen een bondige conclusie formuleren. In zo'n conclusie zeggen zij wat ze in de data gevonden hebben, wat er typisch lijkt of wat speciaal in het oog springt. Het is daarbij nuttig dat leerlingen ook nu en dan een onderzoek wat breder kaderen zoals: "in onze klas hebben we dit gevonden, maar in een andere klas verwachten we iets anders te zien" of "ons groepje is te klein om al een of ander patroon te ontdekken, met een grotere groep verwachten we een staafdiagram dat niet zo erg op en neer springt" enz.

4.2. Voorbeeld: bloedgroepen

Bij de leerlingen van onze klas hebben we de bloedgroep genoteerd. We zien dat de meerderheid (20 van de 22 leerlingen) bloedgroep O of A heeft. Dat is niet onverwacht want O en A zijn ook de bloedgroepen die in België het meest voorkomen.

In België komt O (46 %) iets meer voor dan A (42 %) maar in onze klas is het anders: 9 leerlingen hebben O en 11 hebben A. Dat is niet zo eigenaardig als je bedenkt dat onze klas een toevallig groepje van 22 inwoners van België is. Bij een ander groepje van 22 leerlingen vinden we waarschijnlijk iets anders. We verwachten niet alleen variabiliteit binnen ons groepje (niet iedereen heeft dezelfde bloedgroep) maar ook variabiliteit tussen verschillende groepjes.

Bloedgroep B komt veel minder voor in België (9 %). In onze klas zijn er maar 2 leerlingen met deze bloedgroep. Bloedgroep AB is in België echt zeldzaam (3 %) en in onze klas heeft niemand AB.

Bij de grafische voorstelling van wat we gevonden hebben, kiezen we voor een staafdiagram. We tonen daarbij niet alleen hoe de aanwezige bloedgroepen verdeeld zijn, maar we tonen ook dat bloedgroep AB in onze klas niet voorkomt.

4.3. Voorbeeld: dag van een geboorte

Als je in de frequentietabel voor de Belgische geboorten alle frequenties samentelt dan zie je dat het over 2 418 363 kinderen gaat die in die 20 jaar geboren zijn. In het bijhorende staafdiagram bemerk je een bijzonder patroon. Er is een duidelijk verschil tussen "de werkdagen" en "het weekend".

- In het weekend zijn er veel minder bevallingen dan op de werkdagen. Dat verschil is echt groot. Dat zie je aan de kortere staafjes die boven za (zaterdag) en zo (zondag) staan.
- De staafjes boven de werkdagen (ma, di, wo, do, vr) zijn veel langer. Op die dagen worden meer kinderen geboren. Ook daar is er wat variabiliteit en dinsdag is blijkbaar de topdag voor bevallingen.

Nota.
Je kan de context van dit onderzoek (het gaat hier over geboorten) gebruiken om een verklaring te zoeken voor het patroon in dat staafdiagram. Lang niet alle bevallingen gebeuren "spontaan", er zijn er ook die medisch worden "ingeleid". In die gevallen kan men zelf plannen wanneer de bevalling plaats heeft en dan kiest de arts (of de moeder of de materniteit) liever niet voor een weekend.

Bij je klasgenoten zie je geen patroon in het staafdiagram. Het staafdiagram toont "de toevalligheid" van 22 geboorten met, voor dit groepje, een piek van 8 geboorten op woensdag en geen enkele geboorte op donderdag.

4.4. Voorbeeld: lukraak een getal kiezen

Dit onderzoek gaat over "lukraak kiezen". Je verwacht daarbij dat er geen enkel getal "bevoordeligd" is zodat elk getal "ongeveer evenveel keer" gekozen wordt.

Bij een klein groepje van 22 leerlingen die kunnen kiezen uit 10 verschillende getallen, verwacht je niet dat elk getal ongeveer evenveel keer zal optreden. Inderdaad, in dit onderzoek springt het staafdiagram op en neer. Het toont de variabiliteit in het aantal gekozen getallen. Bij die 22 leerlingen kwam 5 keer het getal 9 voor terwijl het getal 7 helemaal niet opdook.
Je kan hier het gemiddelde $(\overline{x} = 5.8)$ en de mediaan $(Me = 5.5)$ van de 22 gekozen getallen berekenen, maar veel informatie over hoe deze leerlingen een getal hebben gekozen haal je daar niet uit.

Als je datzelfde onderzoek zou herhalen met een grotere groep (zoals 200 of 2000 leerlingen) dan verwacht je niet dat daar geen enkele 7 zou tussen zitten. Je verwacht dan een staafdiagram te vinden waar boven de gekozen getallen staafjes staan die "ongeveer" even lang zijn.

Nota.
Mensen kunnen niet zo goed "lukraak kiezen". Als men aan een groep mensen vraagt om een getal tussen 1 en 10 te kiezen, dan gebeurt het dikwijls dat het getal 7 zeer veel voorkomt. Er bestaat blijkbaar zoiets als een "lievelingsgetal". Als je echt lukrake getallen wil hebben, dan gebruik je beter een toevalsgenerator (random number generator) in een rekentoestel.

4.5. Uitbreiding: 100 m vrouwen

Deze uitbreiding is een voorbeeld van een statistisch onderzoek waar leerlingen kunnen ontdekken dat "werken met data" veel meer is dan "rekenen en tekenen". Zij mogen daarbij zelfstandig op exploratie gaan met de tools die zij zich intussen al eigen hebben gemaakt.

Bij een wedstrijd tussen scholen wordt de 100 meter voor vrouwen gelopen.

In jouw school zijn er 3 leerlingen die in deze afstand uitblinken. Je ziet hier de tijden (in sec) die zij recent in 7 oefenwedstrijden haalden.

Amber	14.49	14.71	15.26	15.68	14.75	15.14	14.36
Emma	14.98	14.84	15.17	14.62	14.69	14.41	14.49
Fiebe	14.41	15.44	14.78	15.61	14.98	15.83	14.61

Jij mag maar één leerling naar die wedstrijd sturen. Wie selecteer je en waarom?

Bij dit onderzoek beschik je al over de vraag en de dataset. Dat dacht je toch.

De vraag "Wie selecteer je en waarom?" zegt dat je moet motiveren waarom je een bepaalde leerling selecteert om jouw school te verdedigen op de interscholenwedstrijd. In de opgave is niet gespecifieerd welk criterium je daarvoor moet gebruiken. De context veronderstelt wel dat je je baseert op de sportprestaties die in de dataset staan en niet op iets anders (zelfs niet als Emma de dochter van de directrice is).

Er zijn hier verschillende mogelijkheden en wat je kiest moet je motiveren.

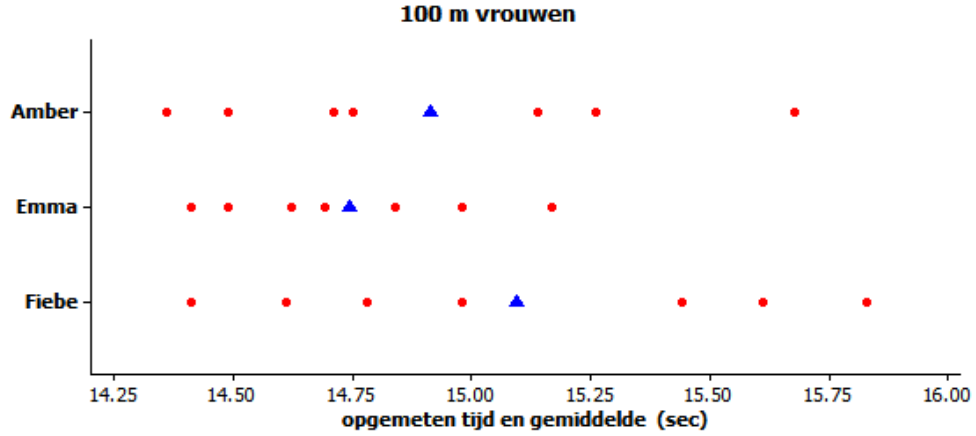
A. Selecteer je Amber?

Als je naar de data kijkt dan zie je dat 14.36 de allerbeste tijd is. Dat is het minimum van die 21 genoteerde tijden. Amber is dit jaar de beste van haar school. Op een schoolfeest heeft zij de "beker 100 m vrouwen" gekregen.
Hoe kan je nu zeggen dat je Amber niet selecteert voor die interscholenwedstrijd? (En hoe leg je dat uit aan haar ouders nadat Amber de beker van de school gekregen heeft?)

Als je het minimum (= de beste tijd) als criterium neemt, dan is Amber de juiste selectie.

B. Selecteer je Emma?

Een grafiek vertelt meestal veel meer dan een dataset. Hieronder zie je op eenzelfde figuur de individuele prestaties van elke kandidaat (de bolletjes) samen met hun gemiddelde (het driehoekje).



Je merkt dat Emma de beste gemiddelde tijd heeft. Bovendien is zij een regelmatige atlete met tijden die niet veel van dat gemiddelde afwijken. Haar gemiddelde is 14.743 $(\overline{x}_E = 14.743)$.

Fiebe heeft het slechtste gemiddelde met bovendien veel variabiliteit (zowel tijden die veel beter zijn als tijden die veel slechter zijn). Haar gemiddelde is 15.094 $(\overline{x}_F = 15.094)$.

Het gemiddelde van Amber ligt tussen dat van Emma en dat van Fiebe. In vergelijking met Emma heeft Amber een grotere variabiliteit in haar prestaties. Haar gemiddelde is 14.913 $(\overline{x}_A = 14.913)$.

Als je als criterium neemt: “een regelmatige atlete met weinig variabiliteit en die bovendien het beste gemiddelde kan voorleggen”, dan selecteer je Emma.

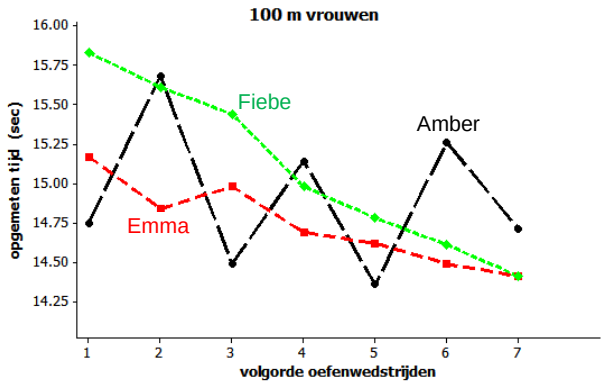
C. Selecteer je Fiebe?

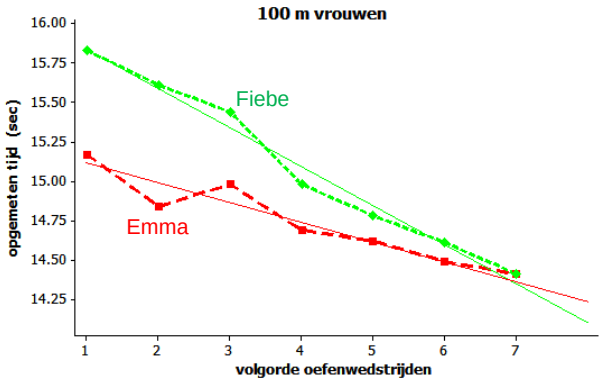
Kan je, zonder de getallen te veranderen, ook voor Fiebe gaan?
Ja, want context is belangrijk. Dezelfde getallen in een andere context kunnen een ander beeld geven.

Niet lang voordat de interscholenwedstrijd plaats vindt, volgen Amber, Emma en Fiebe een identiek trainingsschema. Met telkens ongeveer evenveel dagen tussentijd wordt zeven keer in de school een oefenwedstrijd georganiseerd. Het is daar dat Amber de beker van haar school won. De gelopen tijden ken je al, die staan in de dataset hierboven. Wat de leerlingen niet wisten is dat zij hun tijden moesten doorgeven in de volgorde van die oefenwedstrijden. In de juiste volgorde zien diezelfde tijden eruit als:

	Opgemeten tijden volgens volgorde oefenwedstrijden						
	1	2	3	4	5	6	7
Amber	14.75	15.68	14.49	15.14	14.36	15.26	14.71
Emma	15.17	14.84	14.98	14.69	14.62	14.49	14.41
Fiebe	15.83	15.61	15.44	14.98	14.78	14.61	14.41

De evolutie in de tijd kan je grafisch voorstellen met een lijndiagram. Dat zie je hieronder.





Op de linkse grafiek zie je de evolutie van elke leerling, vanaf de eerste oefenwedstrijd tot de zevende.

Voor Amber is die evolutie een ramp. Zij gaat van beste naar slechtste naar beste naar…. Haar lijndiagram springt op en neer. Dat helpt niet veel om te voorspellen wat er de volgende keer zou kunnen gebeuren. Haar "beste-van-de-school tijd" lijkt eerder een gelukkig toeval dan een bevestiging van stabiele topkwaliteit.

Bij Emma en Fiebe is de lijngrafiek helemaal niet zo wispelturig. Er zijn schommelingen maar beide grafieken tonen een duidelijke trend naar steeds betere (= steeds kortere) tijden.

Bij Emma kan je opmerken dat zij het telkens beter deed dan Fiebe tot op de laatste oefenwedstrijd. Daar liepen beiden dezelfde tijd.

Op de rechtse grafiek zie je bij Emma een trend die naar kortere tijden gaat. Dat is ook zo bij Fiebe maar bij haar is die trend nog groter. Als je die trend illustreert met op zicht een "trend-lijn" door de punten van Fiebe te tekenen, dan zie je dat die lijn sterker daalt dan de "trend-lijn" van Emma. Op basis van deze trend verwacht je dat Fiebe het beter zal doen dan Emma op de interscholenwedstrijd want die komt kort na de zevende oefenwedstrijd waar Fiebe Emma heeft ingehaald. Dus selecteer je Fiebe.

> *Leerlingen ontdekken dat een statistisch onderzoek niet noodzakelijk leidt tot een éénduidig antwoord. Statistische problemen starten met een vraag en eindigen met een antwoord dat vanuit data en context rekening houdt met variabiliteit en toeval.*

Nota.
Onderdelen uit een andere tekst met 4 volledig uitgewerkte onderzoeken kan je zeker ook gebruiken in de eerste graad. Hoe je dat doet, vind je op https://www.uhasselt.be/lesmateriaal-statistiek. Klik daar op *Werkteksten*, scrol naar *1.Exploratieve statistiek* en klik op *Exploratieve statistiek: structuur van de uitgewerkte onderzoeken.*