

4.6 Compleet onderzoek

Inleiding

Het is nu tijd voor een compleet statistisch onderzoek. Bij een statistisch onderzoek zoek of verzamel je data zodat je een uitspraak kunt doen over je onderzoeksvraag en een betrouwbare en begrijpelijke onderbouwing kunt geven. Deze data verwerk je tot een overzichtelijk geheel waar je conclusies uit kunt trekken.



Figuur 1

Je leert in dit onderwerp

- een statistisch onderzoek opzetten.

Voorkennis

- alle kennis van het domein Statistiek.

Uitleg

Het is nu tijd voor een compleet statistisch onderzoek. Bij een statistisch onderzoek zoek of verzamel je data zodat je een uitspraak kunt doen over je onderzoeksvraag en een betrouwbare en begrijpelijke onderbouwing kunt geven. Bij de rapportage respecteer je altijd de privacy van deelnemers en verwerk je de gegevens altijd anoniem.

Bij een onderzoek is het essentieel dat je een goede **onderzoeksvraag** hebt.

Er zijn globaal drie soorten onderzoeksvragen:

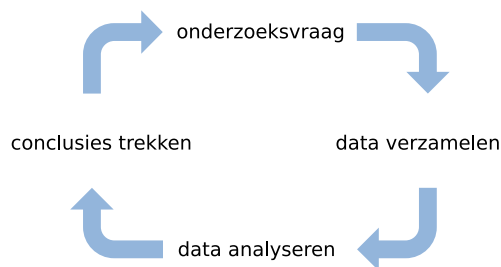
- de **frequentievraag**, hier gaat om 'hoeveel' en 'hoe vaak', het soort onderzoek dat er bij hoort is een beschrijvend onderzoek
- de **verschilvraag**, hier gaat het om het verschil tussen het ene en het andere, en het bijbehorende onderzoek is exploratief
- de **samenhangsvraag**, hier gaat het om de samenhang tussen aspecten, het bijbehorende onderzoek is hier ook exploratief.

Er wordt onderscheid gemaakt tussen **veldonderzoek** en **bureauonderzoek**. Met veldonderzoek verwerk je eigen **data**, bijvoorbeeld met een digitale enquête, zie het practicum **Data verzamelen met een enquête**. Bij een bureauonderzoek gebruik je bestaande data. Veel data zijn te vinden op internet, bijvoorbeeld <http://statline.cbs.nl/Statweb>.

Vervolgens ga je met de data aan de slag:

- Tabellen en grafieken maken en statistische maten voor centrum en spreiding berekenen.
- Proporties en bijbehorende betrouwbaarheidsintervallen berekenen.
- Verschil tussen twee variabelen onderzoeken en conclusies trekken.
- Samenhang tussen twee variabelen onderzoeken en conclusies trekken.

Voor de verwerking van data ben je aangewezen op tools, zoals Excel, VUStat, DWO of GeoEnZo. Soms kun je ook een grafische rekenmachine gebruiken. Elke **tool** heeft voor- en nadelen; bovendien willen ze nog weleens veranderen.



Figuur 2

Welke tool je ook gebruikt, je moet altijd zelf de uitspraken doen en onderbouwen. En bewaar altijd je (ruwe) data. Conclusies onderbouw je met tabellen, diagrammen, statistische maten en redeneringen. **Verwerk** alles in een mooi rapport, een heldere presentatie of een overzichtelijke poster. Beëindig het onderzoek met een kritische **reflectie**.

De tools vind je vaak op internet. Excel staat meestal al standaard op een pc of laptop.

VUStat vind je op <http://www.vusoft.be/vustat.html>

GeoEnZO vind je op <http://www.math4all.nl/pagina/geoenzo>

De DWO (nu Numworx) vind je op <http://www.fi.uu.nl/dwo/>

Het laden van grote databestanden kan soms best even duren. Vaak kun je alleen maar csv-bestanden inlezen. Sla je bestanden goed op, geef ze duidelijke namen en gooi niets weg.

Opgave 1

Wat is het verschil tussen een veldonderzoek en een bureauonderzoek?

Opgave 2

Wat voor werk kan een tool als Excel, DWO, VU-Stat of GeoEnZo voor je uit handen nemen?

Opgave 3

Conclusies onderbouw je met tabellen, diagrammen, en statistische maten.

- Wat is de toegevoegde waarde van tabellen?
- Waarom zou je diagrammen toevoegen?
- Waarom voeg je statistische maten toe?

Opgave 4

Wat is de bedoeling van een kritische reflectie na afloop van je onderzoek?

Theorie en voorbeelden

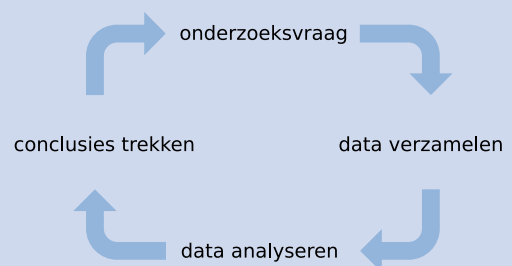
Om te onthouden

Bij een **statistisch onderzoek** zoek of verzamel je data zodat je een uitspraak kunt doen over je onderzoeksvraag en een betrouwbare en begrijpelijke onderbouwing kunt geven. Bij de rapportage respecteer je altijd de privacy van deelnemers en verwerk je de gegevens altijd anoniem.

Bij een onderzoek is het essentieel dat je een goede **onderzoeksvraag** hebt. Hierbij zoek je betrouwbare **data**. Deze data verwerk je tot een overzichtelijk geheel. Ze vormen de onderbouwing van het antwoord op je onderzoeksvraag.

Je trekt **conclusies** die je onderbouwt met tabellen, grafieken, statistische maten en redeneringen. **Verwerk** alles in een mooi rapport, een heldere presentatie of een overzichtelijke poster.

Het is gebruikelijk een goed onderzoek te beëindigen met een **kritische reflectie**.



Figuur 3

Voorbeeld 1

Op 28 juli 2014 vielen grote hoeveelheden neerslag in Nederland, op sommige plaatsen ontstond wateroverlast.

De KNMI vroeg zich af: "Hoe vaak komt extreme neerslag zoals op 28 juli tegenwoordig voor, en is dat anders dan vroeger?"

Bekijk het artikel op de [KNMI](#) website.

Conclusie van KNMI na statistisch onderzoek:

"We hebben op vier verschillende manieren uitgerekend hoe veel vaker intense buien, zoals die op 28 juli 2014 waargenomen zijn, nu voorkomen ten opzichte van het midden van de twintigste eeuw. Het aantal dagen met 50 mm per dag of meer per zomer of per jaar is nu twee keer zo groot als rond 1950 volgens een eenvoudige trendanalyse. We beschouwen ook de kans op zo'n hoge waarde als op 28 juli binnen de verdeling van waarnemingen met de hoogste etmaalsom van het jaar op één van de ongeveer 325 neerslagstations. Een extreme-waarden analyse geeft dat die kans minstens 1,5 keer groter geworden is. Binnen de verdeling van de hoogste dagsommen van het jaar op alle stations is de kans op zo'n waarde nu een factor 2,0 tot 2,6 groter onder wat aannames. Tenslotte geeft het verband met de Clausius-Clapeyron relatie een factor 1,3 tot 2,4 meer kans. Alles wijst er dus op dat deze buien nu ongeveer twee keer vaker voorkomen dan rond 1950."

Opgave 5

Bekijk [Voorbeeld 1](#) en lees het bijbehorende artikel.

- Welke soort onderzoeksvraag hoort bij dit voorbeeld?
- Wat is de conclusie van het KNMI naar aanleiding van het onderzoek?
- Naar welke statistische variabele hebben ze met name gekeken?
- Waarom hebben ze niet alleen maar gekeken naar de gemiddelde neerslag op een dag?
- Wat maakte het complex bij de verwerking van de meetgegevens?
- Tot welke periode van het jaar hebben ze zich beperkt en waarom?

Voorbeeld 2

In de jaren 2008-2014 is de de verkoop van woningen erg moeizaam verlopen. Maar heeft dit gevolgen gehad voor de vraagprijs van een woning?

Een journalist vroeg zich af: "Is er een samenhang tussen vraagprijzen en woningaanbod?"

Lees het artikel op de website van [Follow the Money](#).

Conclusie van de journalist na statistisch onderzoek:

"Natuurlijk is het vooral interessant of er ook nog enige samenhang is tussen bijvoorbeeld een afname in woningaanbod een toename in vraagprijzen. Met heel veel goede wil kun je een heel klein verband zien, maar dit lijkt me statistisch niet erg significant."

Opgave 6

Bekijk [Voorbeeld 2](#) en lees het bijbehorende artikel

- Welke soort onderzoeksvraag hoort bij dit voorbeeld?
- Welke conclusie trekt de journalist?
- Naar welke statistische variabelen is gekeken in dit onderzoek?
- Waarom zou de gemiddelde verkoopprijs per m² woonruimte als statistische variabele niet in het onderzoek bekeken zijn?
- Hoe kun je aan de grafiek zien met welk percentage de gemiddelde vraagprijs daalde in de periode?
- Waarom heeft de onderzoeker onderscheid gemaakt per gemeente?
- In welke gemeente nam volgens de onderzoeker de gemiddelde vraagprijs zelfs nog met meer dan 5% toe? Hoe zie je dat?

Voorbeeld 3

Vlak voor het WK in 2014 in Brazilië kwam de vraag naar boven: “Wie is de beste keeper van de eredivisie?”

Een journalist onderzocht de topkeepers. Zijn onderzoeksvraag was: “Wat zijn de verschillen tussen de keepers in de eredivisie in het tegenhouden van ballen?”

Bekijk ook het artikel [De beste keeper...](#)

Conclusie van de journalist na statistisch onderzoek:

“Jasper Cillessen als beste keeper van Nederland, dat strookt met onze intuïtie. Maar Kristoffer Nordfeldt van Heerenveen als een na beste keeper? Klopt dat ook? Het is in ieder geval contra-intuïtief, maar wellicht tonen de cijfers iets wat we nog niet eerder wisten: dat Kristoffer Nordfeldt een ondergewaardeerd supertalent is.”

Interessant was de toegevoegde reflectie op het statistisch onderzoek naar ballen tegenhouden:

“Maar het kan ook zijn dat toeval hier een grotere rol speelt dan we denken. Als je een munt tien keer opgooit, heb je meestal vijf keer kop of munt, maar soms ook acht keer kop of zelfs negen keer munt. Precies zo kun je na een seizoen eredivisie uitkomen op Cillessen, Nordfeldt en Van Dijk als keepers 1, 2 en 3. De steekproef – 34 wedstrijden, ergens tussen de 300 en 600 schoten – is wellicht niet groot genoeg.”

En tot slot:

“... het grootste onderscheid tussen een professionele keeper en een amateur is dat de amateur veel minder ballen stopt. Maar een professionele keeper onderscheiden van andere professionele keepers, dat is een stuk moeilijker – en lukt in elk geval *niet* aan de hand van hun reddingen, want die blijken vrij willekeurig te zijn en per seizoen te verschillen.”

Opgave 7

Bekijk [Voorbeeld 3](#) en lees het artikel.

- Welke soort onderzoeksvraag hoort bij dit voorbeeld?
- Welke conclusie trekt de journalist?
- Naar welke statistische variabelen is gekeken in dit onderzoek?
- Wat zegt het spreidingsdiagram in dit artikel?
- Welke statistische variabelen zouden volgens de onderzoeker ook bekeken moet worden om het verschil tussen topkeepers te kunnen bepalen?
- Wat is er lastig aan een onderzoek naar de statistische variabelen die bij onderdeel e genoemd zijn?

Toepassen

Je gaat nu een eigen statistisch onderzoek uitvoeren. Belangrijk is dat je over een goede onderzoeksvraag en over genoeg data beschikt. Bespreek dit met je docent.

Hieronder staan voorbeeldonderwerpen voor onderzoek via ‘CBS in de klas’, naar hardlooptijden en een vergelijking van wielerrondes. Het kan natuurlijk goed zijn dat je een ander onderwerp veel geschikter vindt voor een onderzoek. De opgaven hieronder kun je dus als opstapje/voorbeeldvragen gebruiken voor je eigen onderzoek. Er staan geen antwoorden en uitwerkingen bij die vragen.



Figuur 4

Opgave 8: Statistisch onderzoek: CBS in de klas

Via **CBS in de klas** kun je een aantal grote databestanden vinden om statistisch onderzoek mee te doen.

- a Kies één van de beschikbare databestanden en bedenk daarbij een goede onderzoeksvraag.
- b Voer het onderzoek uit, maak geschikte tabellen en diagrammen, bereken passende centrum- en spreidingsmaten en trek conclusies. Formuleer een antwoord op je onderzoeksvraag.

Opgave 9: Chicago marathon 2016

Elke grote stad kent een marathon. Zo ook de stad Chicago. De resultaten van de Chicago Marathon 2016 (ruim 40000 atleten) staan in **dit Excel-document**. Je kunt nu aan de slag met grafieken maken, statistische maten berekenen, groepen vergelijken.

- a Welke statistische variabelen zie je in dit document?
- b Kijk naar de tijden op de finish. Bereken gemiddelde en mediaan van de finishtijden van de lopers.
- c Bereken gemiddelde snelheden van alle lopers.
- d Er is een kolom in het document die aangeeft in welke leeftijdsklasse de loper hoort. Geef een histogram van de frequentieverdeling van de leeftijden van de lopers. Verwacht je de vorm van een normale verdeling? Motiveer je antwoord.
- e Welke vragen zou je allemaal nog kunnen beantwoorden met behulp van deze gegevens?
- f Kies een eigen onderzoeksvraag en voer een statistisch onderzoek uit.

Opgave 10: Wielrennen

Er zijn drie grote wielerrondes in Europa: De Tour de France, De Giro en De Vuelta. Het is een grote eer om een van deze rondes te winnen. De winnaar van de Tour de France krijgt meestal de meeste aandacht. Maar welke ronde is nu het zwaarste? Waar kijk je dan naar? Naar de lengte? Naar de eindtijd? Naar de gemiddelde snelheid? Naar het aantal uitvallers? Of naar het percentage wielrenners dat de finish haalden?

Op de website **Coureur locale** wordt de informatie over de rondes bijgehouden. In een Excel-document staat de kerninformatie per jaar van de drie rondes.

- a Welke statistische variabelen zie je in dit document?
- b Tussen welke kolommen in dit overzicht verwacht je sterke samenhang (of hoge correlatiecoëfficiënt)?
- c Vergelijk de drie rondes op gemiddeld aantal kilometers die de wielrenners per jaar af moesten leggen. Zet de drie rondes in volgorde van hoog naar laag.
- d Vergelijk de drie grote rondes op gemiddeld percentage wielrenners dat de finish haalde. Zet de drie rondes in volgorde van hoog naar laag.
- e Je kunt nog meer vergelijkingen maken tussen de rondes als het op zwaarte gaat en dan de drie rondes rangschikken. Zo krijg je veel vergelijkingen en dan kun je een rondezwaarte-index samenstellen. Welke ronde is volgens jouw systeem het zwaarste?



© 2021

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.

Email: f.spijkers@math4all.nl

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@xs4all.nl een gratis inlog voor de maatwerkdienst aanvragen.
