

STATISTIEK IN STAPPEN

Nel Verhoeven

Vijfde editie

Boom



Met onderstaande unieke activeringscode krijg je via **www.boomstudent.nl** toegang tot de online leeromgeving. Deze code is persoonsgebonden en gekoppeld aan de 5^e editie. Na activering van de code is de online leeromgeving twee jaar toegankelijk. De code kan tot zes maanden na het verschijnen van een volgende editie geactiveerd worden. De code is eenmalig te gebruiken. Deze activeringscode is alleen bruikbaar voor een studentlicentie en geeft geen toegang tot de docentomgeving van het platform. Ben je docent? Dan heb je jouw activeringscode per e-mail ontvangen. Deze code activeer je op **www.boomdocent.nl**.

Omslagontwerp: Haagsblauw, Den Haag

© 2026 Nel Verhoeven | Boom

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever. Auteursrecht ten aanzien van tekst- en datamining en machinelearning is nadrukkelijk voorbehouden.

Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bijvoorbeeld een (digitale) leeromgeving of een reader in het onderwijs (op grond van artikel 16, Auteurswet 1912) kan men zich wenden tot Stichting Uitgeversorganisatie voor Onderwijslicenties, Postbus 3060, 2130 KB Hoofddorp, www.stichting-uvo.nl.

No part of this book may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher. No part of this publication may be reproduced in the context of text and data mining, machine learning or for any other purpose which is not expressly permitted by law without permission of Boom Uitgevers.

ISBN 978 90 2446 897 3

NUR 741

www.boomstudent.nl

www.boom.nl/hogeronderwijs

Inhoud

Voorwoord	6
0 Wanneer gebruik je statistiek?	9
1 Onderzoek doen	17
1.1 Inleiding	17
1.2 Fase 1 – De ontwerpfase van onderzoek	17
1.3 Fase 2 – Data verzamelen	21
1.4 Fase 3 – Gegevens analyseren	26
1.5 Fase 4 – Evaluatie en rapportage van je resultaten	28
1.6 Samenvatting en leestips	30
1.7 Bronnenlijst	31
2 Statistiek met AI	33
2.1 Inleiding	33
2.2 Wat is artificial intelligence?	33
2.3 Prompts schrijven	36
2.4 GenAI-tools bij statistische analyses	41
2.5 AI personaliseren	49
2.6 Samenvatting	52
2.7 Bronnenlijst	53
3 Beschrijvende statistiek – één variabele	55
3.1 Inleiding	55
3.2 Meetniveaus	56
3.3 Frequentieverdelingen	63
3.4 Grafieken van één variabele	68
3.5 Centrummaten	78
3.6 Spreidingsmaten	89
3.7 Bijzondere eigenschappen van de standaardafwijking	101
3.8 De AI-assistent	104
3.9 Samenvatting, leestips en opdrachten	104
3.10 Bronnenlijst	105

4	Beschrijvende statistiek – twee variabelen	107
4.1	Inleiding	107
4.2	Kruistabellen	108
4.3	Grafieken van twee variabelen	114
4.4	GenAI bij beschrijvende analyses van twee variabelen	123
4.5	Samenvatting en opdrachten	127
4.6	Bronnenlijst	128
5	Kansverdelingen	129
5.1	Inleiding	129
5.2	Wat is een kans?	129
5.3	Rekenregels voor kansen	132
5.4	Kansverdelingen	138
5.5	Hypergeometrische kansverdeling	141
5.6	Binomiale kansverdeling	146
5.7	Normale kansverdeling	150
5.8	GenAI inzetten bij kansverdelingen	163
5.9	Samenvatting, leestips en opdrachten	166
5.10	Bronnenlijst	167
6	Schatten en toetsen	169
6.1	Inleiding	169
6.2	Puntschatting	170
6.3	Intervalschatting	177
6.4	Waarom we toetsen	188
6.5	W8-toetsprocedure	195
6.6	Het onderscheidingsvermogen van een toets en fouten in de interpretatie	196
6.7	Inzet van GenAI bij toetsen en schatten	199
6.8	Samenvatting, leestips en opdrachten	202
6.9	Bronnenlijst	203
7	Verschiltoetsen	205
7.1	Inleiding	205
7.2	Toetsen voor één gemiddelde	206
7.3	Gepaarde t -toets	215
7.4	Onafhankelijke z -toets voor twee gemiddelden	219
7.5	Onafhankelijke t -toets voor twee gemiddelden	221
7.6	Correcties op de t -toets (Welch)	227
7.7	GenAI bij verschiltoetsen	229
7.8	Samenvatting, leestips en opdrachten	234
7.9	Bronnenlijst	236

8	Correlatie en regressie	237
8.1	Inleiding	237
8.2	Correlatie	238
8.3	Enkelvoudige regressieanalyse	248
8.4	Verklaarde variantie: hoe goed past het model op de data?	252
8.5	Hypothesen toetsen met regressieanalyse	258
8.6	Onderscheidingsvermogen van de toets	261
8.7	Correlatie en regressie met behulp van GenAI	263
8.8	Samenvatting en opdrachten	265
8.9	Bronnenlijst	266
9	Variantieanalyse	267
9.1	Inleiding	267
9.2	Hypothesen toetsen met variantieanalyse	267
9.3	Variantieanalyse in W8-stappen	270
9.4	Ondersteuning bij variantieanalyse	275
9.5	Variantieanalyse een stap verder	282
9.6	Variantieanalyse en GenAI	285
9.7	Samenvatting, leestips en opdrachten	288
9.8	Bronnenlijst	289
10	Verdelingsvrije toetsen	291
10.1	Inleiding	291
10.2	Chi-kwadraattoets	292
10.3	De rangsomtoets – verschillen tussen twee onafhankelijke groepen	302
10.4	Rangtekentoets – verschillen tussen twee afhankelijke groepen	309
10.5	Kruskal-Wallis-toets – het vergelijken van meer dan twee onafhankelijke groepen	314
10.6	Rangcorrelatietoets: Spearman	319
10.7	Ondersteuning van GenAI bij verdelingsvrije toetsen	325
10.8	Samenvatting, leestips en opdrachten	326
10.9	Bronnenlijst	328
	Bijlagen	329
	Over de auteur	337
	Index	338

Voorwoord

Op feesten en partijen ben je niet bepaald populair als je vertelt dat je ‘statisticus’ bent. Dat wordt deels veroorzaakt door de verwachting dat statistiek moeilijk is en saai, en deels doordat mensen denken dat statistiek iets is waar je in het dagelijks leven nauwelijks mee te maken hebt.

Toch vertel ik met plezier: ‘Ik ben statisticus!’ Ik geef les in statistiek en ik vind het heel leuk en interessant om statistiek te gebruiken. Waarom? Ik kan mijn ongebreidelde nieuwsgierigheid naar antwoorden op vragen in de samenleving bevredigen door middel van het analyseren en interpreteren van data. Bovendien, je beseft het misschien niet, maar iedereen gebruikt dagelijks statistiek, leest erover in de krant, ziet het op de tv en op internet en past het toe op het werk. Veel studenten zien het nut van statistiek niet in als ze op de universiteit beginnen, maar vaak krijg ik na een aantal jaren positieve reacties van mensen die constateren dat ze toch veel aan de cursussen statistiek hebben gehad.

Hoe komt dat nu, dat veel mensen toch zo tegen statistiek opzien? En ... wat is eraan te doen? Dat was zo ongeveer het onderwerp van mijn proefschrift, en vier jaar intensieve studie resulteerde in een aantal aanbevelingen om onderwijs in statistiek ‘aantrekkelijker’ te maken. Maar ja, hoe doe je dat dan? Minder saaie onderwerpen (noem maar eens een alternatief voor de ‘standaarddeviatie’), prettige interactie tijdens de lessen, vooral heel veel hulp en uitleg en bij voorkeur het afschaffen van tentamens? Tja, daar heb ik ook het antwoord niet op. Wel heb ik geprobeerd om met dit boek statistiek wat toegankelijker en duidelijker te maken. Ik hoop dat ik in die opzet geslaagd ben.

De meesten van jullie gebruiken dit boek omdat de opleiding het voorschrijft. Ik weet het: je ziet op tegen de colleges, tegen het moeilijke huiswerk, de oefeningen met SPSS en vooral het tentamen. Wat ik voor jullie kan doen is mijn ervaring met statistiek en met het geven van cursussen aan beginners en gevorderden omzetten in een heldere en functionele tekst. Ik ben sinds de jaren negentig bij methodologie- en statistiekonderwijs betrokken, niet alleen als ontwerper en auteur, maar juist ook als docent. Het zijn vooral de studenten die mij gedurende de afgelopen jaren hebben geïnspireerd en gestimuleerd om statistiek toegankelijk te maken voor een breder publiek en het uit de stoffige academiekamertjes te trekken. Statistiek is van alle tijden en van alle mensen. In de afgelopen jaren heb ik vele voorbeelden van statistische technieken verzameld, kritisch gekeken naar de uitleg van de vele verdelingen, transformaties en toetsen, en daaruit heb ik op basis van mijn ervaring en de leerdoelen in het hoger onderwijs een keuze gemaakt. Statistiek is een middel om een doel te bereiken, geen doel op zich. Statistiek moet toch vooral worden gebruikt om gegevens te ordenen, om goede resultaten te presenteren en om ze bruikbaar te maken voor de samenleving. Statistiek als instrument dus. Schrijven is een eenzaam vak. Achter de computer, tot in de late uurtjes, vele weekends was ik

er zoekt mee. Toch komt zo'n boek niet in eenzaamheid tot stand. Daarom wil ik graag een aantal mensen bedanken. Allereerst Saskia Haartsen, Astrid van der Schee, Esther den Hollander, Rianne de Klerk, Robby Peele, Thirza Anker, Loes Heiligers en Lenneke van der Mark, voor het meedenken, redigeren, becommentariëren en vooral het moreel steunen van het schrijfproces. Verder wil ik Joseph Resovsky, Alexander Whitcomb, Rob Erven en Rika Verhoef bedanken voor hun waardevolle adviezen en bijdragen. Rob Erven voorzag het manuscript van minutieus commentaar en hij liet me op een andere wijze naar de tekst kijken.

Een bijzonder woord van dank wil ik richten aan mijn studenten van de cursus 'multivariate statistiek' die mij behulpzaam waren bij het bedenken van voorbeelden om betrouwbaarheid en validiteit duidelijk te maken. De voorbeelden zijn van Thom, Lotte, Anna, Carina, Fleur, Sasja, Koen, Daida, Janneke, Bas en Merijn.

Het is al vaak gezegd en geschreven: zonder thuisfront ben je nergens. Inmiddels zijn ze mijn gebogen houding achter de computer zo gewend dat ze vreemd opkijken als ik dan eens een weekend besluit om niet te werken.

Ik draag dit boek op aan de toekomstige generatie, die in nog grotere mate met statistiek in aanraking zal komen dan nu al het geval is. Immers, de informatiestromen worden steeds sneller, cijfers spelen een steeds belangrijkere rol in allerlei processen. Statistiek, met al haar regels en voorwaarden, is daarbij onontbeerlijk. Over twintig jaar zullen studenten nog steeds statistiekonderwijs volgen. Ik hoop dat dit boek bijdraagt aan het verhelderen van de uitleg en het vereenvoudigen van de toepassing, en daarmee aan een grotere aantrekkingskracht van het vak.

Bij de vijfde druk

Hoe verandert kunstmatige intelligentie de manier waarop studenten leren, onderzoeken en denken? Sneller dan je denkt. In het hoger onderwijs zorgt generatieve AI voor een ware revolutie. Statistische formules blijven hetzelfde, maar de manier waarop we ermee werken verandert razendsnel. In deze vijfde druk van *Statistiek in stappen* laat ik zien hoe je AI-tools slim inzet: van ondersteuning bij het uitvoeren van analyses tot het interpreteren van resultaten – en vooral: het maken van onafhankelijke keuzes.

In deze vijfde druk speel ik in op die ontwikkeling. Je maakt kennis met statistische toepassingen van generatieve AI, en leert hoe je deze tools slim inzet. Niet alleen om berekeningen uit te voeren, maar ook om de juiste onderzoeksvragen te formuleren, gefundeerde keuzes te maken en resultaten te interpreteren.

Deze editie is niet alleen inhoudelijk vernieuwd met talloze AI-voorbeelden, ook de opdrachten zijn aangepast aan deze ontwikkelingen. De grootste vernieuwing? De leeromgeving zelf. Het boek is volledig gedigitaliseerd, met directe toegang tot extra materiaal, uitwerkingen, en uitgebreide handleidingen voor Excel en SPSS. Liever papier? De vertrouwde printversie is uiteraard nog steeds beschikbaar.

Het schrijven van deze nieuwe editie was een grote ontdekkingsstocht in de snel veranderende wereld van generatieve AI. Een leerzame reis, waarin ik veel inzichten heb opgedaan. Daarbij gaat mijn dank uit naar Jan Anton Brouwer en Loes Heiligers voor de begeleiding bij de totstandkoming van de tekst en naar Thirza Anker voor hulp bij checkpoints, figuren en opdrachten. Ik wil Esther den Hollander bedanken voor haar waardevolle suggesties bij de ontwikkeling van deze vijfde druk.

Ik wil jullie vragen om kritisch mee te blijven kijken. Zo kunnen we blijven werken aan de dynamiek en de reikwijdte van dit interessante vak.

Nel Verhoeven

Ovezande, november 2025

Wanneer gebruik je statistiek?

0

De titel van deze inleiding lijkt wat vreemd. Wanneer gebruik je statistiek?

Het antwoord is duidelijk: eigenlijk altijd! Het dagelijkse leven is er vol van:

- In de rij bij de kassa schat je de wachttijd in.
- Je berekent je gemiddelde eindcijfer voor het vak Economie.
- Je leest in de weer-app op je telefoon dat er morgen 50% kans is op zon.

Statistiek, wat is dat eigenlijk?

De eerste vraag die je kunt stellen is: wat is *statistiek* eigenlijk? Statistiek wordt gezien als de wetenschap die zich bezighoudt met het verzamelen, bewerken, interpreteren en presenteren van kwantitatieve gegevens, cijfers dus. Eigenlijk is statistiek ‘wetenschap-overstijgend’ omdat het niet beperkt blijft tot één discipline, maar multidisciplinair wordt toegepast. Dat wil zeggen dat statistiek in heel veel vakgebieden wordt toegepast. Het slaat een brug omdat het zowel in de natuur- als in de sociale wetenschappen wordt gebruikt, maar ook bij alfavetenschappen, zoals linguïstiek (onderzoek hoe ‘taal’ is opgebouwd). Het is een vakoverstijgende methode, een hulpmiddel om cijfermatige gegevens te verwerken. Dat gebeurt met grafieken, tabellen en met behulp van toetsen en modellen. Statistiek kan ook duiden op ‘statistieken’: een verzamelwoord voor de presentatie van cijfermatige resultaten. Dat komt eigenlijk op hetzelfde neer.

Statistiek gaat over cijfermatige analyse, de analyse van getallen, van numerieke gegevens. Die gegevens moet je eerst verzamelen. Wanneer doe je dat? Meestal tijdens de fase van dataverzameling in je onderzoek. In deze inleidende module geven we de statistiek een plek in onderzoeksprojecten. We laten zien wanneer je statistiek inzet en welke rol die in je onderzoek heeft.

Het nut van statistiek

Statistiek is geen populair onderwerp; ook is het geen gewild vak in de opleiding. De drempel om statistiekcursussen te volgen of om statistiek toe te passen is vaak hoog. Deels komt dat omdat veel mensen het nut van statistiek voor hun professionele loopbaan en voor de samenleving niet zien: ze zijn van mening dat ze statistiek toch niet in hun professionele loopbaan zullen gebruiken. Dat is niet terecht. In tegenstelling tot de verwachting wordt statistiek (bijna) dagelijks toegepast: in de media, tijdens je werk, en natuurlijk tijdens je studie.

Een andere reden waarom statistiek geen populair vak is, is dat men statistiek vaak moeilijk vindt. Is het nodig om formules te kunnen begrijpen zodat je ze kunt toepassen? Ik denk van niet. Ten eerste zijn formules weliswaar complex, maar met een goede uitleg en heldere voorbeelden kun je ze wel leren begrijpen. Ten tweede hoeft je niet voor iedere statistische toepassing formules (uit je hoofd) te kennen. Je kunt heel goed bewerkingen met cijfers uitvoeren, waarbij de formules op de achtergrond blijven. In deze uitgave willen we de formules op een laagdrempelige wijze uitleggen en toepassen en daarbij alleen formules gebruiken als het echt nodig is voor de toepassing, niet als doel op zich.

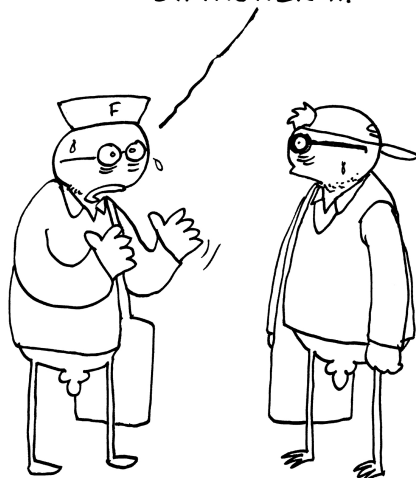
Wanneer statistiek vaker gebruikt wordt als middel om een doel te bereiken en niet als doel op zich, zal de drempel om statistiek toe te passen lager worden. Statistiek moet worden ingezet als *instrument*, om vragen te beantwoorden. Het moet vooral worden gebruikt om gegevens te ordenen, om goede resultaten te presenteren en om ze bruikbaar te maken voor de samenleving.

Met deze uitgave willen we de vooroordelen over statistiek wegnemen. Dat gebeurt door levendige, aansprekende en actuele voorbeelden te gebruiken die dicht bij de belevingswereld van de student liggen, door statistische technieken op een simpele en heldere wijze uit te leggen en door veel oefenstof aan te bieden. Geen lange theoretische verhandelingen over formules, maar eenvoudige en duidelijke toepassingen in de praktijk. En alleen gebruik van formules waar dat echt nodig is.

FOKKE & SUKKE

DURVEN DAT WEL TOE TE GEVEN

EVEN TUSSEN ONS:
IK VOND EEN CONDOOM
OMDOEN STUKKEN
MOEILIJKEER DAN
STATISTIEK !!



RGVT

Statistiek leren

Als je dan toch een cursus statistiek gaat volgen, is het goed om veel te oefenen. Je hebt geluk als je zelf aan een onderzoeksproject kunt meedoen: oefening genoeg. Is dat niet zo, dan adviseren we je om veel in groepjes te werken: een docent is niet altijd voorhanden, en onze ervaring is dat studenten elkaar heel goed kunnen helpen met het maken van opdrachten en het bestuderen van lastig materiaal. En dan is er de onvolprezen docent, dus: blijf niet met vragen zitten, maar stel ze ook echt (je bent heus niet de enige die iets niet begrijpt), vraag om extra uitleg waar dat nodig is.

Ondersteuning van de AI-assistent

In deze nieuwe druk van *Statistiek in stappen* komt daar nog een ‘assistent’ bij, de AI-assistent (AI = artificial intelligence). Met behulp van kunstmatige intelligentie, ook wel afgekort als AI, kun je leren hoe je statistische tools kunt toepassen om je data te verwerken, om met de resultaten de hoofdvraag van je onderzoek te beantwoorden. Deze assistent kan met behulp van ontelbare databases razendsnel informatie verzamelen en combineren tot een antwoord op je vraag.

Een waarschuwing is echter op haar plaats: gebruik AI als vraagbaak, checkpoint, ondersteuning en hulp bij uitleg, maar blijf vooral *zelf* je analysekeuzes maken, uitvoeren en de resultaten interpreteren.

Verschillende soorten statistiek

Je kunt statistiek op verschillende manieren gebruiken:

- Om gegevens te ordenen en te beschrijven: we spreken dan van *beschrijvende* statistiek. Zo kun je bijvoorbeeld een omschrijving geven van de woonkenmerken van de bewoners van de Provincie Utrecht.
- Om kansen mee te berekenen: *kansrekening* is een meer theoretische vorm van statistiek waarbij op basis van een theoretische verdeling de kans wordt berekend dat een bepaalde gebeurtenis plaatsvindt. Zo kun je bijvoorbeeld uitrekenen wat de kans op een sociale huurwoning is voor woningzoekenden in de Provincie Utrecht.
- Om gegevens te toetsen en de resultaten breder toe te passen: we spreken dan van **inferentiële statistiek** of toetsende statistiek. Zo kun je bijvoorbeeld nagaan in welke mate inkomen, leeftijd en opleiding invloed hebben op de huizenprijzen in de Provincie Utrecht.

Doelstelling

De hoofddoelstelling van *Statistiek in stappen* is om je als student een aantal instrumenten te geven om gegevens (bestanden) kwantitatief te analyseren. Het bevat een inleiding in statistische methoden van data-analyse. De bedoeling is dat je leert om deze statistische instrumenten in te zetten bij de kwantitatieve verwerking van gegevens. Het kan als studieboek dienen, maar het kan ook als naslagwerk worden gebruikt wanneer je kwantitatieve gegevens moet verwerken.

Een andere belangrijke doelstelling is om je statistiek ‘eigen’ te maken. De modules zijn zodanig opgebouwd dat op logische wijze wordt uitgelegd hoe een bepaalde techniek werkt, waar deze voor gebruikt wordt, en bij welke onderzoeksvragen deze techniek kan worden toegepast. De nadruk ligt niet zozeer op de formules achter de technieken (hoewel deze wel worden uitgelegd), maar meer op de toepassing ervan. Ten slotte is het de bedoeling om de informatie ‘just in time’ en ‘just enough’ aan te bieden. Dat wil zeggen: de juiste hoeveelheid uitleg en informatie op het moment dat het nodig is.

Deze methode is geschikt voor studenten in het hoger onderwijs, en ze dient als naslagwerk voor mensen die kwantitatieve onderzoeksgegevens willen verwerken.

Globale opzet van het boek

In *Statistiek in stappen* hebben we ervaringen met statistiek en met het geven van cursussen aan beginners en gevorderden samengebracht en omgezet in heldere en functionele tekst. In de afgelopen jaren zijn allerlei voorbeelden van statistische technieken verzameld, is kritisch gekeken naar de uitleg van de vele verdelingen, transformaties en toetsen, en daaruit is op basis van ervaring (en de leerdoelen) in het hoger onderwijs een keuze gemaakt. Daarnaast spelen we in op recente ontwikkelingen op het gebied van generatieve AI, waarbij we tools aanreiken die je kunnen ondersteunen met uitleg, stapsgewijze aanpak, double check en interpretatie van resultaten.

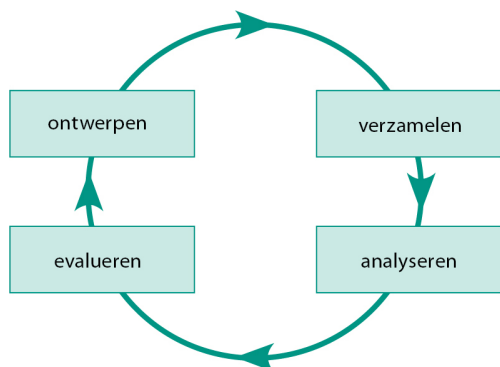
In veel modules hebben we formules gebruikt: waar nodig is dat gebeurd, waar dat niet nodig was, hebben we ze weggelaten. Dus: *zo min mogelijk, maar zo veel als nodig*. En wat denk je? Het aantal formules dat je nodig hebt om kennis te maken met de statistiek in deze uitgave, past op minder dan vier kantjes A4. Dat is nog eens een meevaller.

Opbouw per module

Voordat de ‘harde’ statistiek begint, moet je weten wat de plaats is van statistiek in onderzoeksprojecten: wanneer gebruik je statistiek en hoe verhoudt zich dat tot de rest van je onderzoek? Daarover gaat module 1. We besteden aandacht aan een aantal basisbegrippen en bespreken de fasen van onderzoek: *ontwerp*, *dataverzameling*, *analyse* en *evaluatie*. We laten zien waar en hoe je statistiek kunt inzetten en hoe je de gevonden resultaten presenteert. Deze korte samenvatting van de fasen van onderzoek is gebaseerd op de methode *Wat is onderzoek?*, die eerder in deze reeks is verschenen.

In module 2 maak je kennis met kunstmatige intelligentie en geven we een korte uitleg van de kernbegrippen op dit gebied. Daarnaast introduceren we een aantal innovatieve tools met AI, die je kunt inzetten ter ondersteuning van je analyse. Ook geven we aan hoe je deze AI-tools kunt gebruiken.

In module 3 komt beschrijvende statistiek aan de orde. We behandelen meetniveaus, centrum- en spreidingsmaten, figuren en tabellen. Waar nodig wordt de leerstof verdiept, zoals bij de positie van waarnemingen of eigenschappen van verdelingen. We beginnen met beschrijvingen van telkens één variabele, univariate statistiek. Dat wordt in module 4 gevolgd door beschrijvende statistiek met twee variabelen (bivariaat genaamd), onder andere grafieken van twee variabelen en kruistabellen.



Figuur 0.1 Fasen in onderzoek

In modules 5 en 6 maak je kennis met de meer theoretische kant van de statistiek: kansrekening, punt- en intervalschatting, hypothesetoetsing en significantie. Dit gebeurt stapje voor stapje. We gaan eerst in op een aantal kansverdelingen die nodig zijn om toetsen te begrijpen. Daarna worden in module 6 punt- en intervalschattingen behandeld. Ook besteden we aandacht aan toetstheorie; we leggen uit wat hypothesen zijn, waarom ze nodig zijn en hoe je met ‘significantie’ omgaat. Module 6 wordt afgesloten met de introductie van een methode voor het uitvoeren van toetsen, de *W8-procedure*.

Significantie

Het begrip ‘significantie’ is de afgelopen decennia vaak misbruikt om beleid te maken, producten te verkopen (‘echt’ significante verbeteringen ... nou, echt niet dus), en mensen anderszins over de streep te trekken. Kijk maar naar de reclame van een bekend wasmiddel dat ‘significant betere resultaten geeft dan zijn concurrenten’. Het is belangrijk te bekijken wat zo’n begrip betekent, hoe je ermee omgaat en welke alternatieven je hebt voor interpretatie, want die zijn er wel natuurlijk (vooral in de nuance van de interpretatie van resultaten naar beleidsmakers toe). Dit doen we in module 6. Zo komen ook begrippen als kwaliteit van een effect, de sterkte van verbanden, onderscheidingsvermogen (in het Engels ‘power’ genoemd) en fouten van zowel de eerste als de tweede soort aan de orde. Helaas leidden ze de afgelopen tijd tot misvattingen en fouten in interpretaties, en (erger nog) tot foute beslissingen en maatregelen. Waar zijn de echte statistici als je ze nodig hebt?



In *Statistiek in stappen* worden ook diverse toetsen uitgelegd, afgeleid van de leerdoelen in het hoger onderwijs. In module 7 bespreken we toetsen voor het analyseren van verschillen tussen twee populaties. Dat zijn de *z*- en de *t*-toets, voor afhankelijke en onafhankelijke metingen. Ook de toets voor één gemiddelde wordt behandeld. In module 8 komen correlatie- en regressieanalyse aan bod: het analyseren van verbanden tussen twee variabelen en het interpreteren van mogelijke effecten. Een uitbreiding van ‘verschiltoetsen’ vind je in module 9: de variantieanalyse toetst verschillen tussen meer dan twee populaties. Om een oplossing te geven voor de toenemende stroom van variabelen op laag meetniveau (zogenoemde ‘kwalitatieve variabelen’) behandelen we in module 10 ten slotte een aantal verdelingsvrije alternatieven voor correlatie- en verschiltoetsen. Met verdelingsvrije toetsen testen we verbanden of verschillen tussen groepen, zonder dat aan de voorwaarden van meetniveaus en verdeling is voldaan.

Maar, zoals aangekondigd, beginnen we met een kort overzicht (in module 1) van de onderzoeksfases waar het allemaal om draait en bespreken we een aantal sleutelbegrippen.

Leeswijzer

Ben je een beginner wat betreft statistiek, dan verdient het aanbeveling om alle modules van begin tot eind door te werken en de oefenopdrachten te maken. Overleg met medestudenten of -onderzoekers is daarbij zeker raadzaam. Heb je ervaring met kwantitatieve analyse, dan kan deze lesmethode prima als naslagwerk dienen, via de specifieke module-indeling of via de uitgebreide index.

Hulpmiddelen

Om je het leren gemakkelijker te maken bieden we bij deze uitgave verschillende hulpmiddelen aan:

- *Leerdoelen*: aan het begin van elke module worden de leerdoelen vermeld.
- *Checkpoints*: ieder onderwerp eindigt met een paar controlevragen, zodat je kunt nagaan of je het geleerde goed begrepen hebt.
- Per onderwerp worden, naast een stuk theorie, diverse *voorbeelden* gegeven. Deze voorbeelden staan in kaders en hebben in de kantlijn een icoon van een wijzende hand.
- *Formules* worden niet zomaar aangeboden; ze worden tot in detail uitgelegd.
- Deze uitgave bevat een samenvattend *overzicht van de gebruikte formules*, samengebracht in de bijlage.
- Verder zijn er vier *tabellen* van kansverdelingen als bijlage opgenomen.
- *Opdrachten*: aan het eind van iedere module is een aantal opdrachten opgenomen, zodat je kunt oefenen met de procedures, de technieken en de interpretatie van uitkomsten. Tenzij anders vermeld berusten de data voor deze opdrachten niet op daadwerkelijk verzamelde gegevens. Waar we gebruikmaken van bestaande resultaten, wordt de bron vermeld. Deze opdrachten zijn direct te maken in Boom Academie.
- *Samenvattingen*: aan het eind van iedere module wordt de behandelde stof nog eens kort voor je samengevat, zowel op het gebied van statistiek als op het gebied van AI-assistentie.

Daarnaast is het volgende materiaal beschikbaar:

- introductie van AI-tools met uitleg, links en instructievideo's;
- een uitgebreide inleiding van SPSS met een introductie en per module een groot aantal voorbeelden, bewerkingen, suggesties en extra opties;
- een overzicht van veelgebruikte onderzoekstermen, met hun definitie;
- een introductie in Excel;
- links naar interessante statistische informatie;
- suggesties voor aanvullende literatuur en links;
- keuzediagrammen om de keuze voor een statistische toets te vergemakkelijken.

Het interactieve materiaal is te bekijken op Boom Academie. In het boek hiernaar verwezen met een beeldschermicoon in de kantlijn.

SPSS

Je maakt ook langzaam kennis met de SPSS-software, zowel met de opdrachtstructuur, de datavelden als met de resultatenvensters. Aan het eind van deze uitgave ben je in ieder geval in staat om SPSS te gebruiken voor een reeks eenvoudige analyses.

De uitgebreide Inleiding SPSS kun je hier downloaden:

Inleiding SPSS



Alle statistische technieken in deze uitgave zijn ook in SPSS uitgewerkt.

SPSS brengt ieder jaar een nieuwe versie uit. De versies die voor deze uitgave gebruikt zijn, zijn 29 en 30 (Mac en Windows). Bij het verschijnen van deze uitgave is er ongetwijfeld weer een nieuwe versie onderweg. Onze ervaring is dat er meestal geen grote wijzigingen zijn en dat je met een handleiding van een eerdere versie van SPSS heel goed uit de voeten kunt.

Kunstmatige intelligentie

Ter ondersteuning van je analyses kun je gebruikmaken van digitale tools die werken met kunstmatige intelligentie oftewel artificial intelligence (AI). In iedere module laten we zien welke tools je het beste kunt gebruiken en op welke wijze je dat kunt doen. Daarbij speelt de W8-procedure voor toetsing een grote rol. Naast uitleg en voorbeelden, leer je hoe je AI kunt inzetten bij het uitvoeren van statistische opdrachten en je leert om zelf goede vragen (zogeneten 'prompts') te ontwikkelen in de tools. Door eerdere opdrachten zorgvuldig te bewaren, leg je een eigen statistisch archief van ondersteuning en advies aan, waarop je te allen tijde kunt terugvallen. Daarnaast kun je AI-tools gebruiken om je onderzoekslogboek bij te houden, zodat je gemakkelijk (digitaal) kunt bladeren in je aantekeningen en beslismomenten. Zo creëer je je eigen AI-assistent.

Ten slotte

Hiervoor staan toezeggingen over eenvoud van de uitleg, de stapsgewijze structuur, informatie die alleen wordt aangeboden wanneer en waar het nodig is. Verder hebben we ons ten doel gesteld geen ‘statistiekbijbel’ of ‘statistiekkookboek’ te schrijven. Lastig, want stapsgewijze analyses en uitgangspunten bij je toetsen zouden al snel het tegendeel doen vermoeden. Mocht je in deze uitgave toch tekenen van ‘verval’ ontdekken, schroom dan niet om ons hierover te berichten. We passen het in een volgende versie graag aan.

Onderzoek doen

1

1.1 Inleiding

Statistische analyse vindt meestal plaats tijdens de derde fase van je onderzoek: de analysefase. Dat betekent dat statistiek slechts een *onderdeel* is van je hele onderzoeksproject. Je hebt al een ontwerp gemaakt, gegevens verzameld en een deel van je verslag is ook al klaar. Om je statistische analyses goed te kunnen uitvoeren moet je dus ook kennis hebben van de andere fasen van je onderzoek. In deze eerste module bespreken we kort alle fasen van een kwantitatief onderzoeksproject. Naast een beschrijving van de inhoud van iedere fase geven we een overzicht van veelgebruikte begrippen in elke fase.

Ook presenteren we een casus, die we per fase voor je uitwerken. Mocht je net aan je onderzoek begonnen zijn, dan raak je bij het doornemen van deze module in ieder geval bekend met het onderzoeksjargon. Ook leer je wanneer tijdens je project statistische analyses aan de orde komen. Ten slotte raak je thuis in de wereld van dataverzameling en rapportage. Wil je meer weten over het hele onderzoeksproces, dan heb je baat bij de leeswijzers die zijn opgenomen. Wil je meer weten over de fasen van onderzoek, dan is het aan te raden om *Wat is onderzoek?* (Verhoeven, 2022) te lezen.

Leerdoelen

Aan het eind van deze module ben je bekend met:

- de fasen van kwantitatief praktijkonderzoek;
- een aantal veelgebruikte onderzoeksbegrippen;
- de mogelijkheden van ontwerp, dataverzameling, analyse en rapportage.

1.2 Fase 1 – De ontwerpfase van onderzoek

Je onderzoek start met het ontwerp. Bij het ontwerpen van je onderzoek komt van alles kijken. Je moet het onderwerp goed afbakenen en een doelstelling opstellen, de hoofd- en deelvragen formuleren, een voorstel uitwerken voor het verzamelen van gegevens en de analyse en een *tijdpad* maken dat ook haalbaar is. Oeps, geen gemakkelijke klus.



Foodtruckfestival 1



Sinds enkele decennia worden overal in Nederland zogenaamde foodtruckfestivals georganiseerd, een trend overgewaaid uit de Verenigde Staten. Deze trend is sinds de 19^e eeuw vanaf het Texaanse platteland naar stedelijke gebieden overgewaaid. In Nederland ontwikkelden simpele 'snackwagens' zich vanaf het begin van deze eeuw tot trendy mobiele keukens, als onderdeel van de hipstercultuur. Een foodtruckfestival is een evenement waarbij allerlei wereldgerechten vanuit een aantal mobiele keukens worden aangeboden in bijvoorbeeld een park. Waren deze festivals aanvankelijk kleinschalig, tegenwoordig zijn dit culturele evenementen van klein tot groot, gecombineerd met livemuziek en allerlei culturele activiteiten. De festivals zijn meestal gratis toegankelijk. Stel, je doet onderzoek naar de beleving van en mening over de organisatie van deze festivals.

Tijdens de ontwerpfase komt de opzet van je onderzoeksproject tot stand. Dat doe je in overleg met je begeleider, eventueel met de opdrachtgever of met je team. Doe je individueel onderzoek, dan is het toch verstandig om het ontwerp met iemand te bespreken.

Onderdelen van de start van je onderzoek

Simpel gezegd komt de start van je onderzoek hierop neer:

- Er is een **aanleiding tot onderzoek**: een medisch onderzoeker wil een nieuw medicijn testen, de provincie wil weten hoe haar bewoners brede welvaart ervaren, er moet een beleidsmaatregel worden ontwikkeld, men wil een product geëvalueerd hebben, er moet een nieuwe markt worden verkend, en ga zo maar door.
- Het kan ook zijn dat je een *beroepsproduct* ontwikkelt en daarvoor kwantitatief onderzoek doet. Zo kun je bijvoorbeeld een instrument testen met een experiment, een gemeentelijk plan testen op draagvlak, of een nieuwe maatregel evalueren.
- Er ligt een vraag of probleem; daaruit formuleer je de **centrale vraagstelling**, **deelvragen** en een **doelstelling**. De doelstelling is tweeledig: je geeft aan wat de opdrachtgever met de resultaten wil doen en je beschrijft de onderzoeksdoelen.

- d. Je bakent je onderwerp goed af: wat valt er wel onder en wat niet. Welke definities gebruik je, zijn er eerdere onderzoeksresultaten bekend, welke theoretische modellen kun je gebruiken? Dat doe je met een vooronderzoek.
- e. Je schrijft een voorstel op voor het uitvoeren van het onderzoek, met daarin:
- de aanleiding en de **probleemstelling**;
 - een voorstel tot een **dataverzamelmethode**, periode en populatie;
 - een voorstel voor de analyse;
 - een overzicht van te verwachten situaties die mogelijk de betrouwbaarheid/validiteit aantasten;
 - een tijdpad.

6W-methode

Het belangrijkste onderdeel bij de start van je onderzoek is dat je de juiste vragen stelt. Daarmee doelen we niet alleen op de juiste hoofdvraag en deelvragen, maar ook op vragen voor je dataverzameling en je analyse. Een handig hulpmiddel daarbij is de **6W-methode** (Verhoeven, 2022). Door middel van zes vragen krijg je zicht op de doelstelling van je onderzoek, de hoofdvraag en de deelvragen.

Deze vragen zijn:

- Wat wil je weten? Wat is de situatie of het probleem?
- Wie gaan aan het onderzoek deelnemen? Over wie gaat de vraag, de situatie of het probleem?
- Wanneer: welk tijdsbeeld past daarbij?
- Waar speelt de vraag zich af (afdeling, regio, locatie te bepalen)?
- Waarom wil je dit weten? Wat is de relevantie?
- Waartoe leidt het antwoord? Wat is het doel van je onderzoek? Wat wil je met de resultaten bereiken?

Foodtruckfestival 2

De 6W-vragen kunnen voor het festivalonderzoek als volgt worden ingevuld:

- **W1:** Onderzoek naar de beleving van het foodtruckfestival door bezoekers en deelnemers.
- **W2:** Het onderzoek betreft de bezoekers en deelnemers aan het festival.
- **W3:** Het onderzoek wordt gehouden in 2025, tijdens het festival.
- **W4:** Het onderzoek wordt gehouden in de gemeente waar het festival is georganiseerd.
- **W5:** Het is belangrijk dit onderzoek te houden, omdat er steeds meer publiek op het festival afkomt, met verschillende wensen. Dat betekent dat je moet letten op het aanbod van activiteiten en muziek (voor kinderen, gezinnen, jongeren en ouderen), veiligheid (verkeer, EHBO), voedsel (gezond), en financieel beheer (kosten).
- **W6:** Met de resultaten willen de organisatoren het festival nog beter en veiliger laten verlopen, en het aanbod beter op de bezoeker laten aansluiten.



Aanleiding voor het onderzoek naar beleving van foodtruckfestivals is dat de gemeenten die de festivals organiseren, willen weten of het aanbod aansluit bij de bezoekerspopulatie. En dat geldt niet alleen voor eten en drinken. Zo kan het bijvoorbeeld zijn dat overdag activiteiten voor (kleine) kinderen (en hun ouders) worden georganiseerd, terwijl de avonden meer gericht zijn op een volwassen festivalpubliek, bijvoorbeeld met muziek. Verder is het van belang dat de organisatie gladjes verloopt, veilig, gezond en in goede harmonie. Ten slotte willen de gemeenten goed voorbereid zijn als er meer mensen (dan verwacht) op zulke festivals afkomen. Daarom is het goed om zowel de bezoekers als de deelnemende foodtruckers, beveiliging, verkeersregelaars, sponsors en andere organisatoren bij het onderzoek te betrekken. De volgende vraag zou als hoofdvraag kunnen dienen: hoe beleven de bezoekers en deelnemers het foodtruckfestival in 2025 in hun gemeente?

Onderzoeksbegroting

Het is belangrijk dat je het uiteindelijke ontwerp met je opdrachtgever en/of je begeleider bespreekt en dat je precies afspreekt wat je wel, niet en wanneer gaat doen. In veel gevallen is er ook sprake van een onderzoeksbudget. Werk je als onderzoeker in opdracht, dan stel je van tevoren een begroting op met het aantal uren (en het uurtarief) waarvoor je denkt het onderzoek te kunnen uitvoeren. Ben je student, dan zal dit (op een stagevergoeding na) niet het geval zijn. Het kan wel zo zijn dat je budget krijgt om de kosten te betalen (bijvoorbeeld van een laboratorium, het printen van vragenlijsten of reiskosten).

Onderzoekslogboek

We sluiten deze paragraaf af met een tip: houd een **onderzoekslogboek** bij waarin je aantekeningen maakt over je onderzoek. Zo'n logboek (ook wel *audit trail* genaamd) is in iedere onderzoeksfase bijzonder functioneel: je kunt keuzemomenten noteren, gespreksverslagen opnemen, korte notities maken, een tijdpad bijhouden, en ga zo maar door. Het nut van een logboek wordt zeker ook tijdens de analysefase bewezen: je kunt de stappen in de analyse bijhouden, tussentijdse berekeningen uitvoeren, interpretaties opschrijven en notities maken over het verdere verloop van je analyse.

Voor zo'n logboek kun je een schrift gebruiken, maar Excel, OneNote, of zelfs Word zijn handige programma's om een log in bij te houden. Digitale tools om een logboek bij te houden zijn er ook, bijvoorbeeld Evernote, Remarkable of NotebookLM. De laatste is een tool die gebruikmaakt van kunstmatige intelligentie. Het voordeel van digitale tools is dat je actuele aantekeningen zichtbaar zijn op ieder device dat je gebruikt. Gebruik je toch een schrift, markeer dan belangrijke beslissingen met een gekleurde zelfklevende ruiter, zodat je deze terug kunt vinden.

1.3 Fase 2 – Data verzamelen

Nu je ontwerp vaststaat, ben je klaar voor de tweede fase van je onderzoek, het verzamelen van de gegevens. Daar valt veel over te zeggen, veel meer dan in een korte paragraaf is samen te vatten. We beperken ons tot de hoofdlijnen: wat houdt ‘data verzamelen’ in, welke hoofdgroepen van dataverzameling zijn er, welke **onderzoekstypen** zijn geschikt voor statistische analyse?

Bij het verzamelen van gegevens komt heel veel kijken. Allereerst maak je een onderscheid tussen kwantitatieve en kwalitatieve dataverzameling: het verzamelen van cijfermatige gegevens en van niet-cijfermatige gegevens. De hoofdvraag van je onderzoek is daarbij bepalend. Vervolgens bepaal je welke methode je gaat gebruiken voor je dataverzameling: wordt het een enquête, een experiment of een heranalyse van bestaande data, ga je interviews houden, observeren of literatuuronderzoek doen?

Kwantitatieve dataverzamelingsmethoden

Wil je gegevens verzamelen bij een grote groep personen, zijn de te verzamelen gegevens oppervlakkig, niet van gevoelige aard of doe je metingen bij mensen die een bepaalde behandeling ondergaan, dan kies je meestal voor een *kwantitatieve methode* van dataverzameling. Je hoeft immers niet ‘de diepte in’ met je vragen; je kunt oppervlakkig en snel iemands mening meten en je hebt in korte tijd de beschikking over (veel) cijfermatige gegevens. Welke methoden zijn er zoal? De hoofdgroepen:

- de **enquête**: een vragenlijst (**vragenlijstonderzoek**) voor een grote groep personen;
- het **experiment**: het meten van een effect van een behandeling of een test op een bepaalde uitkomst, in een laboratorium (**laboratoriumexperiment**; zuiver experiment) of in een natuurlijke omgeving (**veldexperiment**);
- de **secundaire analyse**: een heranalyse (met een nieuwe vraag) op kwantitatieve gegevens die al eerder voor een ander doel verzameld zijn.

Kwalitatieve dataverzamelingsmethoden

Het kan ook zijn dat je hoofdvraag gaat over achtergronden van bepaalde gebeurtenissen of gedrag, vanuit het perspectief van de deelnemende personen. Of je wilt meer inzicht in de leefgewoonten van een bepaalde groep mensen die met elkaar leven, dan alleen cijfermatige gegevens. Of je zoekt verklaringen voor situaties in het heden, met behulp van gebeurtenissen in het verleden. In al deze gevallen verdient een *kwalitatieve methode* de voorkeur. De hoofdgroepen zijn:

- het **interview**: een gesprek op hoofdlijnen met één persoon of een groep mensen waarin het perspectief van de onderzochte persoon of groep vooropstaat;
- de **observatie**: het systematisch *waarnemen* van het gedrag van een of meer personen, bijvoorbeeld in hun eigen omgeving; observatieonderzoek kent een aantal subtypen, zoals directe en indirecte observatie, of participierend onderzoek;
- **kwalitatief bureauonderzoek**, bijvoorbeeld archiefonderzoek, secundaire kwalitatieve analyse van bestaand tekstmateriaal, inhoudsanalyse, literatuuronderzoek, enzovoort.



Foodtruckfestival 3

De onderzoeker ontwerpt voor het foodtruckfestival twee methoden van dataverzameling: zowel een enquête onder bezoekers, als interviews met een aantal deelnemende gemeenten, foodtruckers en veiligheidsfunctionarissen. De onderzoeker kiest voor een enquête onder bezoekers, omdat dan een grote groep personen kan worden ondervraagd, op een vrij groot aantal vragen. De gegevens kunnen kwantitatief worden geanalyseerd. Van de deelnemende gemeenten, foodtruckers en veiligheidsfunctionarissen wordt meer specifieke informatie verwacht en een mening die vanuit een bepaalde achtergrond wordt gegeven. Dat is de reden dat daar voor diepte-interviews wordt gekozen: veel informatie in de diepte bij een vrij kleine groep personen. Het onderzoek is daarmee zowel kwantitatief als kwalitatief van opzet.

Meetbaar maken

Vervolgens ga je de begrippen meetbaar maken, met andere woorden, je maakt vragen voor de enquête, een experimentele behandeling, een interviewvraag of een observatiecategorie. Dit heet **operationaliseren**. Operationaliseren doe je zowel in **kwalitatief onderzoek** als **kwantitatief onderzoek**. Het is een lastige fase, omdat je maar één gelegenheid hebt om je vragen te stellen en dus moet het wel in één keer goed zijn. Je vraagt je voortdurend af: meet ik wel wat ik meten wil? Met deze vraag wordt ook wel de begripsvaliditeit aangeduid: de mate waarin je de juiste vragen stelt.

Populatie en steekproef

Ten slotte ga je de data ook werkelijk verzamelen. Dan dienen zich weer nieuwe beslismomenten aan:

- Welke **populatie** heb je voor ogen (dat heb je, als het goed is, ook al in het onderzoeksontwerp opgenomen)? Over welke groep personen wil je met je onderzoek uitspraken doen?
- Vaak is het niet mogelijk om iedereen uit je populatie te vragen deel te nemen. Je kiest er dan voor om een (willekeurig) deel van deze populatie te vragen. Maar, welke **steekproef** ga je trekken? Trek je wel een steekproef?
- Hoe ga je mogelijke deelnemers benaderen?
- Hoe zorg je ervoor dat je een zo hoog mogelijke **respons** krijgt? Zo veel mogelijk deelnemers dus.

Voordat je een statistische analyse op de verzamelde data loslaat, is het belangrijk om te weten wat het verschil is tussen een populatie en een steekproef, en welke gevolgen het trekken van een bepaalde steekproef heeft voor de statistische methoden die je kiest en voor je analysesresultaten. Elk beslismoment heeft consequenties.

Definitie populatie

De populatie is het **domein** van je onderzoek. Dat zijn alle **eenheden** (personen, cases, enzovoort) waarover je een uitspraak wilt doen. Zo kun je onderzoek doen naar studieresultaten van studenten Geneeskunde aan Nederlandse universiteiten, de mening van Facebook-gebruikers analyseren, of BMW-autobezitters onderzoeken.

Definitie steekproef

Het is bijna nooit mogelijk om iedereen in je populatie te bevragen. Daarom kiezen de meeste onderzoekers ervoor om een steekproef te trekken waarmee je bij een deel van de populatie gegevens verzamelt. Dat kan op twee manieren: *select* en *aselect* (ook wel willekeurig genoemd).

Soorten steekproeven

Bij een **aselecte steekproef** selecteer je willekeurig, met behulp van een bepaalde methode. Iedere persoon of eenheid in je populatie heeft een berekenbare kans om in de steekproef terecht te komen. Daarvoor heb je wel een lijst nodig waar alle personen in de populatie op voorkomen, een **steekproefkader**. Zo kun je een aselecte steekproef trekken uit alle ingeschreven medicijnenstudenten aan Nederlandse universiteiten, uit het klantenbestand van BMW aselect een steekproef trekken.

Methoden van aselecte steekproeftrekking zijn:

- **enkelvoudig aselect**: een simpele steekproef uit alle eenheden in de populatie;
- **stratificatie**: de populatie bestaat uit 'strata', bestaande deelpopulaties; uit ieder stratum wordt een aselecte steekproef getrokken;
- **systematische steekproef** met een aselect begin: uit de populatie wordt aselect een begin getrokken, daarna volgt iedere vijfde of tiende persoon, enzovoort; deze methode is populair bij onderzoek in woonwijken, dan wordt op huisnummer geselecteerd;
- **clustersteekproef**: in de populatie zijn clusters aanwezig; er wordt een steekproef van clusters getrokken en deze clusters worden in hun geheel onderzocht, bijvoorbeeld bepaalde klassen in een school;
- **getrapte steekproef**: een combinatie van meerdere steekproefmethoden, meestal met een cluster als basis.

Bij een **selecte steekproef** selecteer je personen op basis van specifieke kwaliteiten, of omdat ze aan bepaalde voorwaarden voldoen. Selecte steekproeven worden ook ingezet als de mogelijkheid ontbreekt om een aselecte steekproef te trekken, bijvoorbeeld omdat er geen steekproefkader voorhanden is. Zo kun je socialmediavrienden vragen mee te doen aan je onderzoek.

Methoden van selecte steekproeftrekking zijn:

- praktisch bruikbare steekproef: iedereen die wil, kan meedoen;
- **quotasteekproef**: je selecteert een aantal **respondenten** per kenmerk, bijvoorbeeld vijftig mannen en vijftig vrouwen;
- **sneeuwbalsteekproef**: je begint in je eigen netwerk en vraagt mensen of zij nog iemand kennen die wil meedoen aan je onderzoek; zo groeit je steekproef;
- doelgerichte steekproef: je ondervraagt mensen op specifieke ervaringen of expertise, bijvoorbeeld bij expertonderzoek;
- zelfselectiesteekproef: kandidaten voor onderzoek kunnen zich aanmelden als ze aan bepaalde voorwaarden voldoen; dit gebeurt vaak bij medische experimenten waarbij de proefpersonen aan bepaalde fysieke ingangseisen moeten voldoen.

Het belang van willekeurige steekproeven bij kwantitatief onderzoek

Het doel van kwantitatief onderzoek is meestal om met de resultaten uitspraken te doen over je populatie. Als je slechts over een steekproef beschikt, een deelverzameling dus, dan moet je ervoor zorgen dat deze op belangrijke kenmerken op deze populatie lijkt; de steekproef moet dus representatief zijn. Pas dan kun je de resultaten generaliseren! Je verbreedt dan de resultaten uit je steekproef naar je populatie. Bij aselechte steekproeven is die mogelijkheid het grootst. Dat neemt niet weg dat we dit toch moeten toetsen. Dat doen we door te analyseren in welke mate de steekproef op de populatie lijkt. We zullen deze generalisatie pas toepassen als we zeker zijn dat onze resultaten niet het gevolg zijn van de samenstelling van de steekproef. Dit is een van de belangrijkste functies van statistische analyses.

Wanneer is een experiment ‘willekeurig’?

Bij **experimenten** is over het algemeen geen sprake van aselechte steekproeven, want er worden geen steekproeven getrokken. Proefpersonen melden zichzelf aan. Toch wil een onderzoeker waarborgen dat de experimentele resultaten valide zijn. Dat gaat als volgt:

- Proefpersonen worden geselecteerd op bepaalde kenmerken.
- Vervolgens worden deze proefpersonen *willekeurig* aan de verschillende condities van een experiment toegewezen.
- Deze toewijzing wordt **randomisatie** genoemd.

In de meeste gevallen is het voor experimentele onderzoekers echter voldoende dat ze conclusies kunnen trekken over het effect van een experiment, en niet zozeer dat ze kunnen generaliseren naar een populatie.

Het veldwerk

Wil je een goede basis leggen voor je analyse, dan moet je goede gegevens verzamelen. Voor statistische analyse zullen dat meestal experimentele gegevens en enquêtegegevens zijn, vaak ook bestaande datasets. Aan die laatste kun je niet zoveel verprutsen, want die zijn al eens voor een ander doel verzameld. Het is echter van het grootste belang dat je experimentele data en enquêtedata goed van kwaliteit zijn. Wat betekent dat?

Voor *kwantitatieve* gegevens onder andere:

- voldoende grote steekproeven, dat maakt je data betrouwbaar;
- niet te veel **missende waarden**, anders is je dataset minder representatief;
- goede vragen en experimentele instrumenten, dat maakt je begripsvaliditeit goed (maar ook de betrouwbaarheid).

Voor *kwalitatieve* gegevens onder andere:

- goed voorbereide interviewers;
- interviewdeelnemers die ook daadwerkelijk iets over het onderwerp kunnen vertellen, zoals ervaringsdeskundigen of experts op een bepaald vakgebied;
- goede vragen en de juiste observatiecategorieën, voor het verhogen van de begripsvaliditeit.

Foodtruckfestival 4

Voor het onderzoek onder bezoekers en deelnemers aan het foodtruckfestival is gekozen voor een enquête en een aantal diepte-interviews. Voor de laatste neemt de onderzoeker een selecte steekproef van experts: twee gemeentelijke vertegenwoordigers, twee foodtrucks en twee veiligheidsexperts. De enquête onder bezoekers wordt uitgevoerd door studenten, die een even aantal van vijftig mannen en vrouwen ondervragen die het festival bezoeken. Een selectieve ‘quotasteekproef’ dus. De onderzoeker weet dat de resultaten daarmee niet kunnen worden gegeneraliseerd. Dat is ook niet zo belangrijk: de festivaldirecteur wil graag een indicatie van de evaluatie door de bezoekers; dan is een aselechte steekproef niet nodig en in dit geval niet zo handig ook.



Onderzoekskwaliteit

Bij het ontwerpen van je onderzoek houd je rekening met de onderzoekskwaliteit. Bij kwantitatief onderzoek is deze grofweg in twee soorten te verdelen:

- **betrouwbaarheid**, de mate waarin je onderzoek **toevallige fouten** bevat;
- validiteit, de mate waarin je onderzoek systematische fouten bevat; hier wordt ook weer een onderscheid in aangebracht:
 - **interne validiteit**: de mate waarin je de juiste conclusies kunt trekken, zoals bij experimenten;
 - **externe validiteit**: de mate waarin je resultaten generaliseerbaar zijn, zoals bij enquêtes;
 - **begripsvaliditeit**: de mate waarin je meet wat je wilt meten.

Deze kwaliteitsaspecten spelen een heel belangrijke rol gedurende je hele onderzoek. Je houdt er rekening mee als je de opzet van je onderzoek ontwerpt, bij de dataverzameling en analyse en je kijkt erop terug in de discussie en conclusie. We zullen er, waar nodig, speciale aandacht aan besteden.