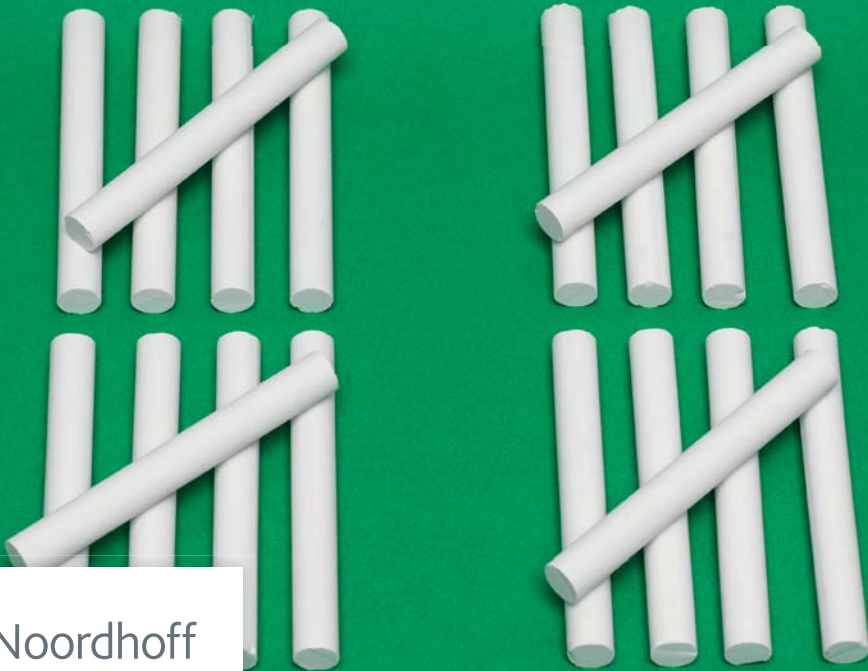


Statistiek in 20 stappen



Noordhoff

Arie Buijs

2^e druk

Statistiek in 20 stappen

Arie Buijs

Tweede druk

Noordhoff Groningen / Utrecht

Ontwerp omslag: Shootmedia

Omslagillustratie: Shootmedia

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB
Groningen of via het contactformulier op www.mijnnoordhoff.nl.

De informatie in deze uitgave is uitsluitend bedoeld als algemene informatie. Aan deze informatie kunt u geen rechten of aansprakelijkheid van de auteur(s), redactie of uitgever ontleen.

0 / 20



© 2020 Noordhoff Uitgevers bv, Groningen/Utrecht, Nederland.

Deze uitgave is beschermd op grond van het auteursrecht. Wanneer u (her)gebruik wilt maken van de informatie in deze uitgave, dient u vooraf schriftelijke toestemming te verkrijgen van Noordhoff Uitgevers bv. Meer informatie over collectieve regelingen voor het onderwijs is te vinden op www.onderwijsauteursrecht.nl.

This publication is protected by copyright. Prior written permission of Noordhoff Uitgevers bv is required to (re)use the information in this publication.

ISBN (ebook) 978-90-01-57538-0

ISBN 978-90-01-57537-3

NUR 916

Woord vooraf

De studieprogramma's bij instellingen van hoger onderwijs worden permanent aan een kritisch oordeel onderworpen. Dat leidt vaak tot zinvolle inhoudelijke bijstellingen van het curriculum. Ook de overheid speelt een belangrijke rol. Dan gaat het over de financiering en het toezicht op universiteiten en hogescholen en ook wordt gekeken naar onderwijsredermenten. Regelmatig zijn er discussies over het kennisniveau van afgestudeerden, soms op basis van de kwaliteit van de afstudeerscripties.

Een van de discussiepunten betreft de kennis die studenten hebben van methoden op het gebied van wiskunde en statistiek. Het niveau daarvan loopt erg uiteen voor diverse opleidingen. Eigenlijk is dat logisch, want sommige vakgebieden vereisen nou eenmaal meer kennis van statistische methoden en technieken dan andere. Uit diverse gesprekken met collega's uit het onderwijsveld werd mij duidelijk dat het wenselijk is dat iedere afgestudeerde van een hogeschool of universiteit op zijn minst beschikt over een soort basispakket aan kennis en vaardigheden. Ongeacht de studierichting eigenlijk. Voor wat betreft het vakgebied statistiek wil dit boek in die behoefte voorzien.

De titel *Statistiek in 20 stappen* weerspiegelt dat dit boek is opgebouwd uit twintig compacte hoofdstukken. Elk van die hoofdstukken zou in een lesmodule van een of twee uur kunnen worden behandeld.

Binnen elk hoofdstuk komt het hoofdonderwerp aan de orde, worden studenten met korte tussenvragen geprikkeld om mee te denken en is er een aantal oefenopgaven. De antwoorden van deze oefenopgaven zijn achterin het boek terug te vinden.

Bij dit boek is ook de website www.statistiekin20stappen.noordhoff.nl beschikbaar. Hier staan voor studenten uitwerkingen en extra oefenmateriaal.



Bij de eerste editie van dit boek hebben dr. ir. Judith Boertjens van uitgeverij Noordhoff en drs. Gert-Jan Reus van de Fontys Hogeschool in Tilburg mij geholpen. Verder kreeg ik assistentie van Kai Strohmeyer.

In deze tweede editie zijn enkele kleine wijzigingen en toevoegingen gedaan. Per hoofdstuk zijn twee extra vraagstukken toegevoegd en mede dankzij de oplettendheid van collega Willem Oldenmenger van de Hanze-hogeschool in Groningen konden enkele storende drukfouten worden hersteld. De hoofdstukkenindeling is dezelfde gebleven.

Mijn wens is dat het boek voor de studenten prettig leesbaar en toegankelijk zal blijken. Daarbij hoop ik dat het veel studenten zelfs enthousiast zal maken voor het vakgebied statistiek. Als er opmerkingen, aanmerkingen of commentaren zijn, verneem ik die graag.

Bilthoven, najaar 2019
Arie Buijs

Inhoud

- 1 Een eerste stap 9**
 - 1.1 Wat is statistiek? 10
 - 1.2 Typen variabelen 12
 - 1.3 Tabellen maken 14
 - Opgaven 18

- 2 Werken met grafieken 21**
 - 2.1 Grafieken maken 22
 - 2.2 Vier grafieken 22
 - 2.3 Het histogram 26
 - 2.4 Twee aanvullende opmerkingen 28
 - Opgaven 30

- 3 Centrummaten en spreidingsmaten 33**
 - 3.1 Centrummaten 34
 - 3.2 Spreidingsmaten 36
 - 3.3 Aandachtspunten 39
 - Opgaven 41

- 4 Maatstaven voor frequentieverdelingen 45**
 - 4.1 Centrummaten bij gegroepeerde gegevens 46
 - 4.2 Spreidingsmaten bij gegroepeerde gegevens 49
 - 4.3 Enkele opmerkingen 50
 - Opgaven 52

- 5 Rekenen met kansen 55**
 - 5.1 Waar komen kansen vandaan? 56
 - 5.2 Volgordeproblemen 58
 - 5.3 Nog meer volgordeproblemen 60
 - Opgaven 62

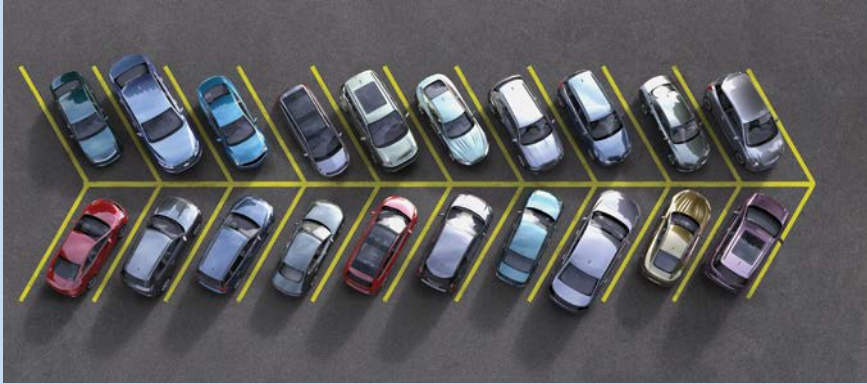
- 6 Voorwaardelijke kansen 65**
 - 6.1 Wat zijn voorwaardelijke kansen? 66
 - 6.2 Kansregels met voorwaardelijke kansen 68
 - 6.3 Twee toepassingen 69
 - Opgaven 72

- 7 Kansvariabelen 75**
 - 7.1 Wat zijn kansvariabelen? 76
 - 7.2 Een discrete kansvariabele 76
 - 7.3 Verwachtingswaarde en variantie 79
 - Opgaven 82

- 8 De normale verdeling 87**
- 8.1 Continue variabelen 88
 - 8.2 De normale verdeling 89
 - 8.3 Werken met andere normale verdelingen 93
 - 8.4 Gemiddelden van waarnemingen 94
Opgaven 96
- 9 De binomiale verdeling 99**
- 9.1 Is het ja of is het nee? 100
 - 9.2 Gemiddelde en spreiding 102
 - 9.3 Kansen berekenen als n groot is 103
 - 9.4 Twee aanvullende punten 104
Opgaven 106
- 10 Steekproeven nemen 109**
- 10.1 Populatie of steekproef? 110
 - 10.2 Van populatie naar steekproef 111
 - 10.3 Hoe neem je een steekproef? 112
 - 10.4 Drie bekende steekproefgrootheden 114
Opgaven 116
- 11 Schatten van het gemiddelde 119**
- 11.1 Populatie en steekproef nogmaals 120
 - 11.2 Nu andersom: van steekproef naar populatie 122
 - 11.3 Hoe breed is het interval? 123
Opgaven 126
- 12 Fractieschattingen 129**
- 12.1 De steekproeffractie 130
 - 12.2 Een steekproeffractie is een kansvariabele 130
 - 12.3 Van steekproef naar populatie 132
Opgaven 135
- 13 Een toets voor het gemiddelde 139**
- 13.1 De nulhypothese en de alternatieve hypothese 140
 - 13.2 Toetsen met kritieke gebieden 141
 - 13.3 Toetsen met een p -waarde 144
Opgaven 146
- 14 Standaarddeviatie onbekend 149**
- 14.1 Schatten van de standaarddeviatie 150
 - 14.2 Op weg naar de t -verdeling 151
 - 14.3 Het betrouwbaarheidsinterval voor μ 153
Opgaven 155
- 15 Associatie 159**
- 15.1 Kijken naar de kruistabel 160
 - 15.2 De maatstaf van Yule 163
 - 15.3 Relatief risico 164
Opgaven 166

16	De regressielijn	171
16.1	Verbanden tussen variabelen	172
16.2	Wat is de beste lijn?	174
16.3	Voorspellen met de regressielijn	176
	Opgaven	179
17	Correlatie	183
17.1	Een verkenning van samenhang	184
17.2	De correlatiecoëfficiënt	185
17.3	Enkele aandachtspunten	188
	Opgaven	190
18	Prijsindexcijfers	193
18.1	Absoluut en relatief	194
18.2	Indexcijfers	195
18.3	Samengestelde indexcijfers	196
18.4	Koppelen van reeksen	199
	Opgaven	201
19	Voorspellen met tijdreeksen	205
19.1	Voortschrijdende gemiddelden	206
19.2	Trendlijnen	209
19.3	Voorspellen met trend- en seizoenbewegingen	210
	Opgaven	213
20	Werken met Excel	217
20.1	Werken met databestanden	218
20.2	Maatstaven berekenen	220
20.3	Kansen opzoeken met Excel	221
20.4	Relaties tussen variabelen	222
	Opgaven	225
	Antwoorden	226
	Illustratieverantwoording	235
	Register	236

Occasions bij Ben de Beun



Een medewerker van een autotijdschrift doet een onderzoek naar de prijzen en eigenschappen van gebruikte auto's. Voor acht bekende modellen werd het aanbod verzameld bij diverse aanbieders, zowel van autohandelaren als van websites waarop particulie-

ren hun auto aanbieden. Voor garage Ben de Beun werd een lijst van 30 aangeboden auto's samengesteld. In de volgende tabel staan de gegevens van de eerste zes auto's uit dat bestand.

TABEL 1.1 Het aanbod van occasions bij garage Ben de Beun

Nr.	Merk	Kleur	Bouwjaar	Aantal kilometers	Staat van onderhoud	Automaat ja/nee	Vraagprijs
1	Volvo	Blauw	2005	345.000	Slecht	Ja	1.800
2	BMW	Rood	2002	320.000	Goed	Nee	3.900
3	Mazda	Geel	2016	56.000	Uitstekend	Ja	11.800
4	Opel	Grijs	2009	130.000	Redelijk	Ja	3.500
5	Peugeot	Wit	2013	165.000	Matig	Nee	4.700
6	Volkswagen	Zwart	2018	18.000	Goed	Nee	17.900

Bron: administratie De Beun BV, april 2020

Na het verzamelen van de gegevens vraagt de medewerker zich af wat hij allemaal kan doen met die gegevens. Hoe moet je die

grote hoeveelheid informatie overzichtelijk maken? Wat voor conclusies kun je daaruit trekken?

1

Een eerste stap

In dit hoofdstuk beantwoorden we de volgende vragen:

- Wat houdt het vakgebied van de statistiek eigenlijk in?
- Waarom is het belangrijk een onderscheid te maken tussen diverse typen van variabelen?
- Hoe kan een tabel op een correcte manier worden opgesteld en hoe kunnen op basis daarvan uitspraken worden gedaan?

In veel situaties komen we in aanraking met informatie in de vorm van gegevensbestanden. Zoiets ontstaat onder andere omdat we mensen hun mening vragen via een enquête, of omdat we kans zien via internet allerlei gegevens te verzamelen. Hoe je dat verzamelen op een verantwoorde manier doet is op zichzelf al een belangrijke vraag. Wat je daarna met die gegevens kunt doen is nog veel belangrijker. In dit boek besteden we aandacht aan het vakgebied van de statistiek.

In dit inleidende hoofdstuk geven we in paragraaf 1.1 eerst een beeld van dat vakgebied. In paragraaf 1.2 bespreken we de verschillen tussen diverse typen van variabelen en in paragraaf 1.3 laten we zien hoe tabellen een belangrijk hulpmiddel zijn bij het presenteren van waargenomen uitkomsten.

1.1 Wat is statistiek?

De term statistiek kennen we in allerlei betekenissen. Soms zegt men: volgens de statistieken moet dit of dat gelden. Zoals: volgens de statistieken is het aantal strandbezoekers altijd het hoogst op zondagen. Of: volgens de statistieken wordt Ajax vaak kampioen door het winnen van de laatste wedstrijd van de competitie.

Statistiek is ook de naam van een breed vakgebied. Dat omvat onderdelen zoals kansrekening, maar ook het doen van volkstellingen of het uitvoeren van opiniepeilingen. Er wordt soms gezegd dat je met een handig gebruik van statistieken de mensen van alles kunt wijsmaken, of het nou waar is of niet. De gegevens vormen altijd het centrale punt en zijn de basis van alle inzichten. Het doel van dit boek is om te laten zien op welke manier je gegevens kunt verzamelen, ordenen, samenvatten en analyseren, met als doel om meer te begrijpen van de wereld om je heen. Hiertoe bekijken we eerst welke soorten statistiek er zijn en introduceren we enkele basisbegrippen.

Soorten statistiek

Je kunt statistiek voor verschillende doeleinden gebruiken. Je kunt proberen om bepaalde patronen in een collectie verzamelde gegevens te vinden. Zo iets noemen we ook wel het analyseren van een databestand. Om de gegevens te beschrijven maak je soms gebruik van tabellen, grafieken en soms ook van bepaalde maatstaven zoals de gemiddelde uitkomst, of de mate waarin uitkomsten blijken te variëren. Je wilt dan een grote verzameling gegevens op een doelgerichte manier beschrijven om een kort en krachtig beeld te schetsen. We noemen dit *beschrijvende statistiek*.

Je kunt ook verder gaan. Als je eenmaal gedegen kennis hebt van bepaalde onderwerpen, dan wil je soms voorspellingen doen voor de toekomst of je wilt jouw analyse gebruiken om beslissingen te nemen. Daarbij speelt kansrekening een belangrijke rol. Het gaat er vaak om dat je ondanks allerlei onzekerheden tot verantwoorde conclusies wilt komen.

Er zijn tal van voorbeelden te geven van het toepassen van statistiek binnen de vakgebieden economie en management. We noemen hier:

- Een onderneming wil graag een beeld krijgen van het ziekteverzuim onder haar werknemers. Na het verzamelen van de gegevens kunnen allerlei vragen worden bestudeerd. Is het verzuim lager dan vorig jaar? Verzuimen ouderen meer dan jongeren? Zijn er bepaalde afdelingen met een opvallend hoog verzuim?
- De overheid is bezig met de hoogte van de pensioenpremies. Om een goede schatting te kunnen maken van de toekomstige betalingen is het belangrijk dat er gegevens beschikbaar zijn over de verwachte levensduur van mannen en vrouwen. Ook moeten schattingen worden gemaakt van het aantal werkenden nu en in de toekomst, want zij moeten de premies betalen. Dat alles vereist het verzamelen en analyseren van veel gegevens.
- Een supermarkt is bezig met de inkoop van voorraden voor de komende kerstperiode. Dat vereist een inzicht in de te verwachte drukte en ook van de mogelijk verschuivende voorkeuren van klanten.
- De minister van Economische Zaken wil graag de hoogte van de inflatie weten. Daarvoor worden prijzen van heel veel producten verzameld. Deze worden omgerekend tot een prijsindexcijfer.

- Een fabrikant eist voor onderdelen die worden geproduceerd dat de afmetingen daarvan tussen een hoogste en een laagste toegestane grens vallen. Met statistische methoden kan de kwaliteit van de producten periodiek worden gemeten.

Dit zijn maar enkele voorbeelden. Vast staat dat bij heel wat vakgebieden en beroepen – variërend van marketing tot productiemanagement en van accountant tot beleidsmaker bij de overheid – een beheersing van een aantal statistische technieken erg belangrijk is. Ook is in veel gevallen kennis van een aantal statistische methoden een noodzaak om afstudeerprojecten succesvol te kunnen voltooien.

Enkele basisbegrippen

Onderzoek doe je om meer te weten te komen over een bepaald onderwerp. Dat kan van alles zijn, bijvoorbeeld een opiniepeiling bij een groep personen, een keuring van producten in een fabriek of het bepalen van het percentage bedrijven dat met export bezig is. Voordat je begint moet je vaak goed afspreken over welke eenheden je uiteindelijk een uitspraak wilt doen. Doe je een opiniepeiling onder alle volwassen Nederlanders? Bekijk je de export van alle bedrijven, of alleen van bedrijven die minstens 25 mensen in dienst hebben?

De totale verzameling elementen waarop het onderzoek betrekking heeft noemen wij de *populatie*. Dus als we een onderzoek doen naar de tevredenheid over de studiefinanciering, dan bestaat de populatie waarschijnlijk uit alle Nederlandse studenten die gerechtigd zijn dit te ontvangen. En bij een politieke peiling bestaat de populatie uit alle kiesgerechtigde Nederlanders. Vaak is het bezwaarlijk om de totale populatie in je onderzoek te betrekken. Het is misschien veel te veel werk om iedereen te ondervragen en soms is ook niet alle informatie beschikbaar. Daarom maak je vaak gebruik van een *steekproef*, dat betreft een onderzoek bij een deel van de populatie. Je hoopt dan dat de steekproef toch een goed beeld geeft van het grote geheel.

Populatie en steekproef

Telkens weer moet je goed het onderscheid weten tussen populatie en steekproef.

Een element van de populatie kan allerlei kenmerken hebben. Die kenmerken noemen we *variabelen*. Deze variabelen kunnen vaak allerlei waarden aannemen.

- ?? Kijk eens naar het praktijkvoorbeeld in de tabellen 1.1 en 1.5 over gebruikte auto's. Wat zou je daar de populatie noemen? En wat is dan de steekproef? Welke variabelen werden verzameld?

Aan het begin van dit hoofdstuk staan in tabel 1.1 de eerste zes auto's vermeld. In tabel 1.5 staat het bestand met de 30 auto's van garage Ben de Beun. Dus zowel tabel 1.1 als tabel 1.5 is te beschouwen als een steekproef. Ze geven een beeld van de populatie van alle tweedehandsauto's, maar dat beeld

hoeft natuurlijk niet precies overeen te komen met de werkelijke samenstelling van de populatie.

Het onderzoek gaat over een aantal kenmerken van tweedehandsauto's. De populatie bestaat uit alle tweedehandsauto's die op de Nederlandse markt worden aangeboden. Hieruit is een steekproef samengesteld voor garage Ben de Beun, waarbij men zich heeft beperkt tot acht gangbare modellen. De kenmerken waar men op let zijn prijs, merk, kilometerstand, staat van onderhoud, bouwjaar, kleur en het wel of niet hebben van een automaat. Let erop dat deze variabelen onderling sterk verschillen van karakter.

?? Welke waarden kan de variabele 'merk' aannemen? En welke waarden kan de variabele 'kilometerstand' aannemen?

De variabele 'merk' heeft als mogelijke uitkomsten Fiat, BMW, Mazda, Opel, Volkswagen, enzovoort. Dus dat zijn allemaal aanduidingen. Dat noemen we ook wel *kwalificaties*.

De variabele kilometerstand laat als mogelijke uitkomsten juist allemaal getallen zien. Misschien wel tussen 1.000 en 400.000 kilometer. Zo'n variabele is prettig, want met getallen kun je rekenen.

We maken onderscheid tussen *kwalitatieve* en *kwantitatieve* variabelen. Als de uitkomst van de variabele wordt bepaald door een getal, dan heb je te maken met een kwantitatieve variabele. Als de uitkomsten die verschijnen aanduidingen zijn, maar geen getallen, dan heet dat een kwalitatieve variabele. Met kwalitatieve variabelen kan niet (zinvol) gerekend worden, met kwantitatieve variabelen kan dat wel.

?? Kun je van een kwalitatieve variabele een kwantitatieve variabele maken en kun je een kwantitatieve variabele omzetten in een kwalitatieve variabele?

Als je een databestand invoert in een computerprogramma, dan is het soms handig om een kwalitatieve variabele om te zetten in codenummers. Voor de variabele 'merk' zou je ervoor kunnen kiezen om de uitkomst Fiat aan te duiden met code 1, BMW met code 2, Mazda met code 3, enzovoort. Je moet echter wel opletten blijven. Na het omzetten van kwalificaties in een getal kun je met die getallen nog steeds geen zinvolle berekeningen uitvoeren. Denk je maar eens in dat de gemiddelde waarde van de codes 3,27 zou bedragen. Heb je ooit gehoord van merk 3,27? Zo'n getal is zonder betekenis. Andersom kan het wel. Je kunt een kwantitatieve variabele omzetten in een kwalitatieve variabele. Bijvoorbeeld bij variabele kilometerstand zou je ervoor kunnen kiezen om de kilometerstanden van 0 t/m 79.999 aan te duiden als 'heel weinig' gereden, 80.000 t/m 119.999 als 'weinig' gereden, 120.000 t/m 199.999 als 'veel' gereden en 200.000 en hoger als 'heel veel' gereden. Zo'n omzetting is soms handig om aan anderen snel een beeld te geven van de waargenomen kilometerstanden. Je moet wel voorzichtig zijn, want in principe levert dit veel grovere informatie op dan wanneer je met de oorspronkelijke gegevens werkt.

1.2 Typen variabelen

Men kan allerlei soorten variabelen onderscheiden. We zagen al de begrippen kwantitatieve en kwalitatieve variabelen. Binnen de categorie kwantitatieve variabelen maken we een onderscheid tussen *discrete* en *continue*

variabelen. Bij een discrete variabele is er in principe een beperkt aantal mogelijke uitkomsten. Hoeveel passagiers zitten er in een autobus? Dat zal waarschijnlijk een getal zijn tussen 0 en de maximale capaciteit van bijvoorbeeld 80 passagiers. Wat zijn de prijzen per fles van vier soorten cola? Dat zijn misschien de bedragen €0,95, €1,28, €1,45 en €1,68. Dus uitkomsten mogen ook wel cijfers achter de komma hebben. Van een discrete variabele kun je in principe alle mogelijke uitkomsten opnoemen.

Dat is anders bij een continue variabele omdat daarbij in principe oneindig veel denkbare uitkomsten mogelijk zijn. Je moet dan denken aan de wachttijd totdat de volgende trein langskomt, het gewicht van een pompoen of de afstand tussen twee bomen langs een weg. Als je maar goede meetapparatuur hebt, kun je heel nauwkeurig meten en dan zijn er dus heel veel denkbare uitkomsten. Een handige hulpregel is dat je bij een continue variabele tussen twee mogelijke uitkomsten altijd nog wel een extra uitkomst kunt bedenken die daar tussenin ligt.

Je moet je wel realiseren dat het verschil tussen een discrete en een continue variabele in de praktijk kleiner is dan in de theorie, omdat de meetapparaten een eindige nauwkeurigheid hebben. Afgelegde afstand is een voorbeeld van een continue variabele, maar de kilometerteller gaat meestal niet verder dan één getal achter de komma. Hierdoor worden continue variabelen eigenlijk omgezet in discrete variabelen.

Vier meetniveaus

Een ander onderscheid tussen variabelen betreft de schaalindeling. Dat wordt ook wel het meetniveau genoemd. Er zijn vier bekende meetniveaus. Dat zijn de nominale schaal, de ordinale schaal, de intervalschaal en de ratschaal.

1 *Nominaal meetniveau*

Bij een nominaal meetniveau kunnen we de denkbare uitkomsten niet op een logische manier op een vaste volgorde zetten. Je kunt het kenmerk van de variabele weliswaar weergeven door een codenummer, maar daarmee hebben die cijfers nog niet een logische volgorde.

?? Bij welke variabelen van de tweedehandsauto's is er sprake van een nominaal meetniveau? Kijk hiervoor in tabel 1.1.

Bij de gegevens in de tabel zijn merk en kleur voorbeelden van kenmerken gemeten met een nominale schaal.

2 *Ordinaal meetniveau*

Bij een ordinale schaal heb je wel te maken met een vaste volgorde. Er is dan een duidelijke manier om vast te stellen wat de logische volgorde moet zijn.

?? Welke variabelen in het voorbeeld hebben een ordinaal meetniveau?

De staat van onderhoud is een voorbeeld van een kenmerk gemeten met een ordinaal meetniveau. Want 'uitstekend' is het beste, 'goed' volgt daarna, en vervolgens 'redelijk', 'matig' en 'slecht'. Een logische opeenvolging dus. Als je die aanduidingen door elkaar zou husselen, dan ontstaat een niet-logisch rijtje.

Andere bekende voorbeelden van ordinaal meetniveau zijn rangen bij de landmacht (van soldaat derde klas tot en met generaal) en antwoordmogelijkheden bij een enquête (van zeer slecht tot zeer goed).

3 Interval meetniveau

Bij een interval meetniveau hebben verschillen tussen twee uitkomsten een eenduidige betekenis. Een intervalschaal heeft geen natuurlijk nulpunt. Het is eigenlijk vrij willekeurig gekozen. Temperatuur is hier een typisch voorbeeld van. Het nulpunt is het vriespunt van water, maar men had evengoed voor iets anders kunnen kiezen.

?? Welke variabelen uit de tabel worden gekenmerkt door een intervalschaal?

Bij de gegevens van het voorbeeld is bouwjaar typisch een kenmerk gemeten met een intervalschaal. Het tijdsverschil tussen bouwjaar 1995 en 1998 is even groot als tussen 2009 en 2012, namelijk 3 jaar. De intervallen zijn dus gelijk.

Andere bekende voorbeelden van intervalschaal zijn temperatuur en tijds-aanduiding.

4 Ratio meetniveau

Bij een ratio meetniveau is er wel een natuurlijk nulpunt. Dus als we van de auto's de leeftijd zouden berekenen, dan bestaat een auto van tien jaar oud twee keer zo lang als een auto van vijf jaar.

?? Bij welke variabelen van de auto's is er sprake van een ratio meetniveau?

Bij de gegeven tabel zijn 'kilometerstand' en 'prijs' voorbeelden van een kenmerk gemeten met een ratioschaal. Nul is hier een logisch begin van de schaal.

Andere bekende voorbeelden van een ratio meetniveau zijn gewicht en verstreken tijd.

?? Waarom is indeling in meetniveaus van belang?

Omdat het meetniveau bepalend is voor de rekenkundige bewerkingen die je mag toepassen, is het belangrijk daaraan aandacht te geven.

1.3 Tabellen maken

Een verzameling gegevens kan omvangrijk zijn. Daarom vormt dat vaak een onoverzichtelijk geheel. Het is belangrijk om daarin orde te scheppen. Dat doen we door tabellen en grafieken te maken waarmee we de informatie op een heldere manier kunnen weergeven. Het gaat er in principe om dat we de gegevens in groepen indelen. Die groepen noemen we *klassen*. Bij elke klasse gaan we vaststellen hoeveel waarnemingen daartoe behoren. Die aantallen heten *frequenties*. Bij tabellen moet wel aan enkele voorschriften worden voldaan:

- Er moet een opschrift zijn, zodat de lezer weet waarover de tabel informatie geeft.
- Er moet een kopregel zijn, zodat je kunt zien wat er in de kolommen staat vermeld.

- De klassen moeten zodanig zijn gekozen dat alle waarnemingen een plek kunnen krijgen. Die klassen moeten zo mogelijk in een logische volgorde staan.
- Er moet een totaalregel zijn.
- Als het kan moet er een bronvermelding zijn.

Als we het bestand met de 30 auto's (tabel 1.5) bekijken, kunnen we bijvoorbeeld de staat van onderhoud in beeld brengen, zoals in tabel 1.2 gedaan is.

TABEL 1.2 Staat van onderhoud van 30 auto's

Staat van onderhoud	Aantal
Uitstekend	5
Goed	9
Redelijk	5
Matig	7
Slecht	4
Totaal	30

Bron: steekproef april 2020

In tabel 1.2 vormen alle vijf antwoordmogelijkheden een eigen klasse. Bij een andere variabele kan het gebeuren dat men eerst klassen wil gaan vormen.

?? Hoe zou je een klasse-indeling maken van de aantallen gereden kilometers?

Het maken van klassen bij een ratio- of intervalvariabele is in feite een willekeurige keus. Graag maken we klassen van een gelijke breedte, maar soms is het handiger enkele extra brede klassen te maken als er in een gebied weinig waarnemingen zijn. Dit doen we in de tabel 1.3 voor de gereden kilometers, waarbij we de indeling volgen die we in paragraaf 1.1 als voorbeeld noemden.

TABEL 1.3 Kilometerstand van 30 auto's

Kilometerstand	Aantal
0 - < 80.000	6
80.000 - < 120.000	4
120.000 - < 200.000	11
200.000 - < 400.000	9
Totaal	30

Bron: steekproef april 2020

Let op de notatie bij de klasse 0 - < 80.000. Dit geeft aan dat uitkomsten van minstens 0 tot hoogstens 79.999 in die klasse zitten. Anders gezegd: deze notatie vertelt dat de ondergrens wél tot die klasse behoort en de bovengrens niet.

- ?? Stel dat een auto precies 120.000 kilometer op de teller heeft staan. In welke klasse moet je die dan indelen?

De waarde 120.000 is het laagste getal dat in de klasse 120.000 - < 200.000 geplaatst hoort te worden.

Met tabellen kun je allerlei vormen van presentatie kiezen. Een bekende vorm is de zogenoemde *kruistabel*. Daarin worden de waarnemingen ingedeeld aan de hand van twee kenmerken tegelijk. We laten dat zien aan de hand van de variabele 'kilometerstand' ten opzichte van de variabele 'staat van onderhoud'. Dat levert tabel 1.4.

Kruistabel

Met een kruistabel kun je samenhang tussen twee variabelen zichtbaar maken.

TABEL 1.4 Staat van onderhoud versus kilometers

Kilometers x 1000					
Staat van onderhoud	0 - < 80	80 - < 120	120 - < 200	200 - < 400	Totaal
Uitstekend	3	1	1	0	5
Goed	3	2	2	2	9
Redelijk	0	0	2	3	5
Matig	0	1	5	1	7
Slecht	0	0	1	3	4
Totaal	6	4	11	9	30

Bron: steekproef april 2020

We zien hier de frequenties van combinaties van de twee variabelen verschijnen. Aan de randen van de tabel zien we rechts een totaalkolom en onderaan een totaalregel. Grappig is dat je in die totaalkolommen weer dezelfde informatie als in tabel 1.2 en 1.3 ziet verschijnen.

TABEL 1.5 Het aanbod van occasions bij garage Ben de Beun

Nr.	Merk	Kleur	Bouwjaar	Aantal kilometers	Staat van onderhoud	Automaat ja/nee	Vraagprijs
1	Volvo	Blauw	2005	345.000	Slecht	Ja	1.800
2	BMW	Rood	2001	320.000	Goed	Nee	3.900
3	Mazda	Geel	2016	56.000	Uitstekend	Ja	11.800
4	Opel	Grijs	2009	130.000	Redelijk	Ja	3.500
5	Peugeot	Wit	2013	165.000	Matig	Nee	4.700
6	Volkswagen	Zwart	2018	18.000	Goed	Nee	17.900
7	Volvo	Grijs	2010	163.000	Matig	Nee	7.400
8	BMW	Grijs	2008	225.000	Redelijk	Nee	5.300
9	Fiat	Wit	2013	92.000	Goed	Nee	4.900
10	Mazda	Grijs	2009	197.000	Matig	Ja	3.500
11	BMW	Zwart	2011	173.000	Goed	Nee	6.800
12	Peugeot	Rood	2010	144.000	Matig	Nee	2.900
13	Opel	Blauw	2012	88.000	Uitstekend	Ja	10.500
14	Opel	Grijs	2013	146.000	Redelijk	Ja	8.200
15	Toyota	Wit	2003	267.000	Slecht	Nee	1.200
16	Volkswagen	Blauw	2006	279.000	Slecht	Nee	1.800
17	Toyota	Wit	2007	162.000	Matig	Nee	1.900
18	Opel	Zwart	2010	210.000	Redelijk	Ja	4.700
19	Fiat	Rood	2014	63.000	Uitstekend	Nee	4.800
20	Volkswagen	Grijs	2011	173.000	Goed	Ja	5.200
21	Opel	Grijs	2010	228.000	Goed	Ja	2.900
22	Volvo	Rood	2015	135.000	Uitstekend	Ja	14.600
23	BMW	Grijs	2008	293.000	Matig	Nee	4.600
24	Volvo	Blauw	2009	255.000	Redelijk	Nee	4.800
25	Volkswagen	Zwart	2004	173.000	Slecht	Nee	1.300
26	Volkswagen	Wit	2015	46.000	Goed	Nee	9.800
27	Toyota	Blauw	2012	92.000	Goed	Ja	4.100
28	Volkswagen	Rood	2008	116.000	Matig	Nee	2.300
29	Fiat	Wit	2016	35.000	Uitstekend	Nee	5.900
30	BMW	Grijs	2017	46.000	Goed	Ja	16.200

Bron: administratie De Beun BV, april 2020

Opgaven

1

- 1.1** Geef voor de volgende variabelen aan op welk meetniveau zij worden gemeten.
- a** de provincie waarin een student is geboren
 - b** de afstand tussen het woonhuis en de school van een leerling
 - c** het aantal uren per week dat een docent werkt
 - d** het unieke nummer van een paspoort
 - e** de kleurcodes waarmee de weersituatie wordt aangegeven door het KNMI
 - f** de klanttevredenheid gemeten op een vijfpuntsschaal
 - g** het bouwjaar van een auto
 - h** het landnummer dat hoort bij een telefoonnummer
- 1.2** Maak voor het bestand van de 30 auto's (tabel 1.5) een tabel waaruit blijkt hoeveel auto's van de diverse merken er voorkomen.
- 1.3** Maak een tabel waaruit de verdeling van de auto's op basis van de vraagprijs blijkt. Maak klassen van €5.000 breed. Kies als laatste klasse €15.000 en hoger.
- 1.4** Maak een kruistabel waarbij de waarnemingen worden ingedeeld op basis van de kenmerken 'vraagprijs' en 'automaat'. Kies dezelfde klassen als bij vraag 1.2.
- 1.5** Onderzoek de volgende beweringen:
- a** Zwarte auto's hebben doorgaans een niet zo goede staat van onderhoud.
 - b** Duitse auto's hebben doorgaans een hogere vraagprijs dan de rest.
 - c** Als je de auto's op basis van bouwjaar onderverdeelt in oude en jonge occasions, dan blijkt dat jonge occasions vaker een automaat hebben.
- 1.6** Een onderzoeker wil bekijken of Duitse auto's vaker een automaat hebben in vergelijking met niet-Duitse auto's. Maak een kruistabel op basis van de gegevens in tabel 1.5.

Zie verder hoofdstuk 20 voor enkele vragen met een groot databestand van 200 auto's.

Occasions in beeld gebracht



Bij garage Ben de Beun worden 30 auto's aangeboden. Van alle auto's zijn het merk, de kleur, de kilometerstand, het bouwjaar en de vraagprijs bekend. Een medewerker van Ben de Beun wil graag op een overzichtelijke manier het aangeboden assortiment weergeven om daarmee snel een beeld te krijgen. Is het mogelijk om in één oogopslag

een beeld te krijgen van de verschillende merken die worden aangeboden? Ook zou hij graag willen laten zien hoe de verdeling is van de vraagprijzen van de aangeboden occasions. Daarom wil hij proberen enkele grafische voorstellingen te maken van diverse kenmerken. Het bestand van de 30 auto's is in tabel 1.5 te vinden.