

# Statistiek in 20 stappen

voor management en economie

Arie Buijs

Eerste druk



Noordhoff Uitgevers



# Statistiek in 20 stappen

voor management en economie

**Arie Buijs**

---

Eerste druk

Noordhoff Uitgevers Groningen / Houten

*Ontwerp omslag:* Rocket Industries, Groningen

*Omslagillustratie:* Getty Images

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan: Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB Groningen, e-mail: [info@noordhoff.nl](mailto:info@noordhoff.nl)

*Aan de totstandkoming van deze uitgave is de uiterste zorg besteed. Voor informatie die desondanks onvolledig of onjuist is opgenomen, aanvaarden auteur(s), redactie en uitgever geen aansprakelijkheid. Voor eventuele verbeteringen van de opgenomen gegevens houden zij zich aanbevolen.*



0 / 14

© 2014 Noordhoff Uitgevers bv Groningen/Houten, The Netherlands.

Behoudens de in of krachtens de Auteurswet van 1912 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. Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (postbus 3060, 2130 KB Hoofddorp, [www.reprorecht.nl](http://www.reprorecht.nl)). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, [www.stichting-pro.nl](http://www.stichting-pro.nl)).

*All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.*

ISBN (ebook) 978-90-01-85547-5

ISBN 978-90-01-81442-7

NUR 916

# Woord vooraf

Het Hoger Onderwijs is volop in beweging. De overheid speelt hierbij een belangrijke rol. Dan gaat het bijvoorbeeld over de financiering en het toezicht op universiteiten en hogescholen. Enerzijds wordt er gekeken naar onderwijsrendementen waarmee wordt gemeten hoeveel studenten in staat zijn een opleiding succesvol af te ronden. Anderzijds zijn er discussies over het kennisniveau van afgestudeerden, zoals dat onder andere zou moeten blijken uit de kwaliteit van de afstudeerscripties. Binnen dat spanningsveld worden maatregelen bedacht, zoals het bindend studieadvies, selectie bij toelating en diverse vormen van bijscholing. Ook krijgen studenten strakere tijdschema's opgelegd waarbinnen zij hun opleiding moeten voltooien.

Een van de discussiepunten betreft het kennisniveau van studenten op het gebied van wiskunde en statistiek. Dat niveau loopt erg uiteen voor diverse opleidingen. Eigenlijk is dat logisch, want sommige vakgebieden vereisen een grondige parate kennis van statistische methoden en technieken, terwijl bij andere studies wellicht een beperkter pakket kan volstaan.

Uit diverse gesprekken die ik had met collega's uit het onderwijsveld werd mij duidelijk dat het wenselijk is dat iedere afgestudeerde van een hogeschool of universiteit op z'n 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 beperkt aantal oefenopgaven. De antwoorden van deze oefenopgaven zijn achterin het boek terug te vinden.

Bij dit boek is de website [www.statistiekin20stappen.noordhoff.nl](http://www.statistiekin20stappen.noordhoff.nl) beschikbaar. Hier staan voor studenten uitwerkingen en extra oefenmateriaal.

In de voorbereidingsfase van dit boek had ik het voorrecht om te kunnen brainstormen met dr. ir. G. Judith Boertjens van uitgeverij Noordhoff en met drs. Gert-Jan Reus van de Fontys Hogeschool in Tilburg. Collega Reus heeft tevens een aantal ideeën aangeleverd voor praktijkvoorbeelden. Verder kreeg ik van Kai Strohmeyer assistentie bij een aantal statistische toepassingen van Excel. Ook kreeg ik enkele nuttige adviezen van dr. Bo van der Rhee van de Nyenrode Business Universiteit. Ik ben hen daarvoor zeer dankbaar.

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, voorjaar 2014  
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 17
  
- 2 Werken met grafieken 19**
  - 2.1 Grafieken maken 20
  - 2.2 Vier grafieken 20
  - 2.3 Het histogram 24
  - 2.4 Twee aanvullende opmerkingen 26
  - Opgaven 27
  
- 3 Centrummaten en spreidingsmaten 29**
  - 3.1 Centrummaten 30
  - 3.2 Spreidingsmaten 32
  - 3.3 Aandachtspunten 35
  - Opgaven 36
  
- 4 Maatstaven voor frequentieverdelingen 39**
  - 4.1 Centrummaten bij gegroepeerde gegevens 40
  - 4.2 Spreidingsmaten bij gegroepeerde gegevens 43
  - 4.3 Enkele opmerkingen 44
  - Opgaven 46
  
- 5 Rekenen met kansen 49**
  - 5.1 Waar komen kansen vandaan? 50
  - 5.2 Volgordeproblemen 52
  - 5.3 Nog meer volgordeproblemen 53
  - Opgaven 55
  
- 6 Voorwaardelijke kansen 57**
  - 6.1 Wat zijn voorwaardelijke kansen? 58
  - 6.2 Kansregels met voorwaardelijke kansen 60
  - 6.3 Twee toepassingen 61
  - Opgaven 63
  
- 7 Kansvariabelen 65**
  - 7.1 Wat zijn kansvariabelen? 66
  - 7.2 Een discrete kansvariabele 66
  - 7.3 Verwachtingswaarde en variantie 69
  - Opgaven 71

- 8 De normale verdeling 75**
- 8.1 Continue variabelen 76
  - 8.2 De normale verdeling 77
  - 8.3 Werken met andere normale verdelingen 81
  - 8.4 Gemiddelden van waarnemingen 82  
Opgaven 83
- 9 De binomiale verdeling 85**
- 9.1 Is het ja of is het nee? 86
  - 9.2 Gemiddelde en spreiding 88
  - 9.3 Kansen berekenen als  $n$  groot is 89
  - 9.4 Twee aanvullende punten 90  
Opgaven 92
- 10 Steekproeven nemen 95**
- 10.1 Populatie of steekproef? 96
  - 10.2 Van populatie naar steekproef 97
  - 10.3 Hoe neem je een steekproef? 98
  - 10.4 Drie bekende steekproefgrootheden 99  
Opgaven 102
- 11 Schatten van het gemiddelde 105**
- 11.1 Populatie en steekproef nogmaals 106
  - 11.2 Nu andersom: van steekproef naar populatie 108
  - 11.3 Hoe breed is het interval? 109  
Opgaven 111
- 12 Fractieschattingen 113**
- 12.1 De steekproeffractie 114
  - 12.2 Een steekproeffractie is een kansvariabele 114
  - 12.3 Van steekproef naar populatie 116  
Opgaven 118
- 13 Een toets voor het gemiddelde 121**
- 13.1 De nulhypothese en de alternatieve hypothese 122
  - 13.2 Toetsen met kritieke gebieden 123
  - 13.3 Toetsen met een  $p$ -waarde 126  
Opgaven 128
- 14 Standaarddeviatie onbekend 131**
- 14.1 Schatten van de standaarddeviatie 132
  - 14.2 Op weg naar de  $t$ -verdeling 133
  - 14.3 Het betrouwbaarheidsinterval voor  $\mu$  135  
Opgaven 137
- 15 Associatie 139**
- 15.1 Kijken naar de kruistabel 140
  - 15.2 De maatstaf van Yule 143
  - 15.3 Relatief risico 144  
Opgaven 146



<b>16</b>	<b>De regressielijn</b>	<b>149</b>
16.1	Verbanden tussen variabelen	150
16.2	Wat is de beste lijn?	151
16.3	Voorspellen met de regressielijn	154
	Opgaven	156
<b>17</b>	<b>Correlatie</b>	<b>159</b>
17.1	Een verkenning van samenhang	160
17.2	De correlatiecoëfficiënt	161
17.3	Enkele aandachtspunten	164
	Opgaven	166
<b>18</b>	<b>Prijsindexcijfers</b>	<b>169</b>
18.1	Absoluut en relatief	170
18.2	Indexcijfers	171
18.3	Samengestelde indexcijfers	172
18.4	Koppelen van reeksen	175
	Opgaven	176
<b>19</b>	<b>Voorspellen met tijdreeksen</b>	<b>179</b>
19.1	Voortschrijdende gemiddelden	180
19.2	Trendlijnen	182
19.3	Voorspellen met trend-en seizoenbewegingen	184
	Opgaven	187
<b>20</b>	<b>Werken met Excel</b>	<b>189</b>
20.1	Werken met databestanden	190
20.2	Maatstaven berekenen	192
20.3	Kansen opzoeken met Excel	193
20.4	Relaties tussen variabelen	194
	Opgaven	197
	<b>Antwoorden</b>	<b>198</b>
	<b>Register</b>	<b>205</b>

# Occasions bij Ben de Beun

1

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	1999	345.000	Slecht	Ja	1.800
2	BMW	Rood	1996	320.000	Goed	Nee	3.900
3	Mazda	Geel	2010	56.000	Uitstekend	Ja	11.800
4	Opel	Grijs	2003	130.000	Redelijk	Ja	3.500
5	Peugeot	Wit	2007	165.000	Matig	Nee	4.700
6	Volkswagen	Zwart	2012	18.000	Goed	Nee	17.900

Bron: administratie De Beun BV, april 2014

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.

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 op pagina 16 staat het bestand met de 30 auto's van garage Ben de Beun.

Het onderzoek gaat over een aantal kenmerken van tweedehands auto's. De populatie bestaat uit alle tweedehands auto'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. Dat 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 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, dan kun je zo'n uitkomst niet zinvol interpreteren.

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 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 ratioschaal.

#### 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 tweedehands auto'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 intervallschaal?

Bij de gegevens van het voorbeeld is bouwjaar typisch een kenmerk gemeten met een intervallschaal. 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 intervallschaal 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 delen. Die groepen noemen we *klassen*. Bij iedere 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 2014

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

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

Het maken van klassen bij een ratio- of interval variabele 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 gaven.

**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 2014

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.

**TABEL 1.4** Staat van onderhoud versus kilometers

<b>Kilometers x 1000</b>					
<b>Staat van onderhoud</b>	<b>0 -&lt; 80</b>	<b>80 -&lt; 120</b>	<b>120 -&lt; 200</b>	<b>200-&lt; 400</b>	<b>Totaal</b>
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
<b>Totaal</b>	<b>6</b>	<b>4</b>	<b>11</b>	<b>9</b>	<b>30</b>

Bron: steekproef april 2014

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

<b>Nr.</b>	<b>Merk</b>	<b>Kleur</b>	<b>Bouwjaar</b>	<b>Aantal kilometers</b>	<b>Staat van onderhoud</b>	<b>Automaat ja/nee</b>	<b>Vraagprijs</b>
1	Volvo	Blauw	1999	345.000	Slecht	Ja	1.800
2	BMW	Rood	1996	320.000	Goed	Nee	3.900
3	Mazda	Geel	2010	56.000	Uitstekend	Ja	11.800
4	Opel	Grijs	2003	130.000	Redelijk	Ja	3.500
5	Peugeot	Wit	2007	165.000	Matig	Nee	4.700
6	Volkswagen	Zwart	2012	18.000	Goed	Nee	17.900
7	Volvo	Grijs	2004	163.000	Matig	Nee	7.400
8	BMW	Grijs	2002	225.000	Redelijk	Nee	5.300
9	Fiat	Wit	2007	92.000	Goed	Nee	4.900
10	Mazda	Grijs	2003	197.000	Matig	Ja	3.500
11	BMW	Zwart	2005	173.000	Goed	Nee	6.800
12	Peugeot	Rood	2004	144.000	Matig	Nee	2.900
13	Opel	Blauw	2006	88.000	Uitstekend	Ja	10.500
14	Opel	Grijs	2007	146.000	Redelijk	Ja	8.200
15	Toyota	Wit	1997	267.000	Slecht	Nee	1.200
16	Volkswagen	Blauw	2000	279.000	Slecht	Nee	1.800
17	Toyota	Wit	2001	162.000	Matig	Nee	1.900
18	Opel	Zwart	2004	210.000	Redelijk	Ja	4.700
19	Fiat	Rood	2008	63.000	Uitstekend	Nee	4.800
20	Volkswagen	Grijs	2005	173.000	Goed	Ja	5.200
21	Opel	Grijs	2004	228.000	Goed	Ja	2.900
22	Volvo	Rood	2009	135.000	Uitstekend	Ja	14.600
23	BMW	Grijs	2002	293.000	Matig	Nee	4.600
24	Volvo	Blauw	2003	255.000	Redelijk	Nee	4.800
25	Volkswagen	Zwart	1998	173.000	Slecht	Nee	1.300
26	Volkswagen	Wit	2009	46.000	Goed	Nee	9.800
27	Toyota	Blauw	2006	92.000	Goed	Ja	4.100
28	Volkswagen	Rood	2002	116.000	Matig	Nee	2.300
29	Fiat	Wit	2010	35.000	Uitstekend	Nee	5.900
30	BMW	Grijs	2011	46.000	Goed	Ja	16.200

Bron: administratie De Beun BV, april 2014

# Opgaven

- 
- 1.1** 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.
- 1.2** 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.
- 1.3** Maak een kruistabel waarbij de waarnemingen worden ingedeeld op basis van de kenmerken 'vraagprijs' en 'automaat'. Kies dezelfde klassen als bij vraag 1.2.
- 1.4** 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.

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

---