

Beschrijvende en toetsende statistiek

Het begrijpen en interpreteren van
tabellen en statistiek met spss

Boom

BREGJE VAN GRONINGEN
EN CONNIE DE BOER



Beschrijvende en toetsende statistiek

Beschrijvende en toetsende statistiek

Het begrijpen en interpreteren
van tabellen en statistiek met SPSS

**Bregje van Groningen
Connie de Boer**

Boom

inclusief website!

Met onderstaande unieke activeringscode krijg je via www.boomstudent.nl toegang tot de website met extra materiaal.

Deze code is persoonsgebonden en gekoppeld aan de 1^e druk. Na activering van de code is de website twee jaar toegankelijk. De code kan tot zes maanden na het verschijnen van een volgende druk geactiveerd worden. De code is eenmalig te gebruiken.

Omslagontwerp: Cunera Joosten, Amsterdam

Opmaak binnenwerk: Nu-nique grafische vormgeving, Goor

© 2022 Bregje van Groningen & Connie de Boer | Boom uitgevers Amsterdam

Behoudens de in of krachtens de Auteurswet gestelde uitzonderingen mag niets uit deze uitgave worden vervoelvoudigd, 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 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.

ISBN 978 90 2444 218 8

ISBN 978 90 2444 219 5 (e-book)

NUR 916

www.boomstudent.nl

www.boomhogeronderwijs.nl

Woord vooraf

Veel studenten hebben moeite met statistiek. Om het hen zo gemakkelijk mogelijk te maken hebben we in dit boek veel aandacht besteed aan een duidelijke uitleg en een zo leesbaar mogelijke tekst. Ter wille van de leesbaarheid hebben we nauwelijks verwijzingen naar literatuur opgenomen. In de literatuurlijst achter in het boek staan de publicaties waarop wij onze kennis hebben gebaseerd. Deze publicaties gaan op sommige statistieken veel uitgebreider in. De literatuurlijst kun je daardoor ook beschouwen als aanbevolen literatuur voor verdere studie.

Een eerder boek van onze hand beperkte zich tot de beschrijvende statistiek. Dat boek hebben we uitgebreid met toetsende of inferentiële statistiek. Het uitrekenen van de formules met de hand is komen te vervallen. We volstaan in dit boek met uitleg van het gebruik van SPSS voor het uitvoeren van de analyses. De instructies hiervoor staan in aparte kaders. In de tekst van de hoofdstukken leggen we uit hoe je de gecreëerde SPSS-tabellen kunt interpreteren en hoe je vervolgens de bevindingen rapporteert.

Eerdere teksten voor dit boek zijn gebruikt in het onderwijs over onderzoeksmethoden en statistiek aan studenten Communicatiewetenschap. De reacties van studenten en docenten van deze cursussen hebben we gebruikt om de tekst iedere keer weer te verbeteren.

De volgende personen willen we noemen als dank voor hun inhoudelijke en tekstuele bijdragen: Wouter de Nooy, Hoàn Le en Gert van Driel.

Bregje van Groningen
Connie de Boer

Inhoud

Woord vooraf	5
Inleiding	13
1 Basiselementen	17
1.1 Datamatrix	17
1.2 Frequentietabellen	20
1.2.1 Hoe ziet dit eruit in SPSS?	21
1.2.2 Missing values	23
1.2.3 Grafieken	24
1.3 Variabelen	25
1.3.1 Afhankelijke en onafhankelijke variabelen	27
1.4 Meetniveaus	28
1.4.1 Nominaal meetniveau	28
1.4.2 Ordinaal meetniveau	29
1.4.3 Interval meetniveau	30
1.4.4 Ratio meetniveau	31
1.4.5 Criteria	32
1.5 Waarden van variabelen	32
1.5.1 Continue en discrete meetschalen	32
1.6 Univariate, bivariate en multivariate analyses	33
1.6.1 Univariate analyses	34
1.6.2 Bivariate analyses	34
1.6.3 Multivariate analyses	35
1.7 Centrummaten	36
1.7.1 Modus	36
1.7.2 Mediaan	38
1.7.3 (Rekenkundig) gemiddelde	41
1.7.4 Keuze tussen centrummaten	42
1.7.5 Centrummaten: samenvattend	42
1.8 Spreiding	43
1.8.1 Spreiding voor variabelen op minimaal ordinaal niveau	44
1.8.2 Spreiding voor variabelen op interval- en rationiveau	47
1.9 Centrum- en spreidingsmaten in SPSS	51
1.10 Standaardiseren (z-scores)	54
1.11 Normale en scheve verdelingen	57
1.11.1 (Standaard)normale verdeling	58
1.11.2 Scheve verdelingen	63

1.12	Rapportage	66
1.13	Samenvatting	66
2	Van steekproef naar populatie	69
2.1	De steekproef	69
2.2	Kansverdelingen	73
2.2.1	(Standaard)normale verdeling	73
2.3	Steekproevenverdeling	75
2.3.1	Steekproevenverdeling van een discrete variabele	75
2.3.2	Steekproevenverdeling van een continue variabele	78
2.3.3	Standaardfout	81
2.4	Inferentiële statistiek	81
2.4.1	Betrouwbaarheidsintervallen	82
2.4.2	Toetsen van hypothesen	87
2.5	Representativiteitstoetsen	102
2.6	Samenvatting	107
3	Bewerken en controleren van je data	109
3.1	Syntax	109
3.2	Missing values	112
3.3	Compute	117
3.4	Hercoderen	121
3.5	Select Cases	125
3.6	Samenvatting	132
4	Kruistabelanalyse	133
4.1	Samenhang in kruistabellen	137
4.2	Chi-kwadraat	138
4.2.1	Geobserveerde en verwachte frequenties	139
4.2.2	Voorwaarden voor een kruistabelanalyse	142
4.2.3	Toetsen van de nulhypothese	143
4.3	Wat zijn associatiematen?	146
4.3.1	Meetniveau van de variabelen	147
4.3.2	Symmetrische en asymmetrische relaties	147
4.4	Nominale associatiematen	148
4.4.1	Cramers V	148
4.4.2	Phi	152
4.4.3	Goodman en Kruskals tau	155
4.5	Ordinale associatiematen	161
4.5.1	Gamma	164
4.5.2	Somers' d	167
4.6	Het kiezen van een associatiemaat	170
4.6.1	Formulering van uitspraken op basis van je onderzoek	171

4.7	Tabelsplitsing	173
4.7.1	Spurieuze samenhang	175
4.7.2	Specificatie	181
4.7.3	Versluiting	183
4.8	Samenvatting	184
5	T-toetsen	187
5.1	T-toets op één steekproefgemiddelde (one sample t-test)	188
5.1.1	T-toets op één steekproefgemiddelde in SPSS	189
5.1.2	Rapportage vanuit SPSS	193
5.2	T-toetsen op twee gemiddelden: afhankelijk en onafhankelijk	194
5.2.1	Afhankelijke t-toets/gepaarde waarnemingen (paired samples t-test)	196
5.2.2	Onafhankelijke t-toets	202
5.3	Samenvatting	212
6	Variantieanalyse	213
6.1	Eenwegsvariantieanalyse (one-way ANOVA)	217
6.1.1	Interpretatie vanuit SPSS	221
6.1.2	Effectgrootte bij variantieanalyse	225
6.1.3	Post hoc-toets bij variantieanalyse	226
6.1.4	Conclusie aan de hand van de tabellen	228
6.2	Tweewegsvariantieanalyse	230
6.2.1	Tweewegsvariantieanalyse in SPSS	235
6.2.2	Tweewegsvariantieanalyse in wetenschappelijke artikelen	241
6.3	Samenvatting	242
7	Correlatie en regressie	243
7.1	Pearsons correlatiecoëfficiënt	243
7.1.1	Grafische weergave	244
7.1.2	Pearsons correlatie in SPSS	247
7.1.3	Spearman's rho correlatiecoëfficiënt	251
7.1.4	Partiële correlaties	254
7.2	Enkelvoudige regressie	257
7.2.1	Interpretatie van de SPSS output	263
7.3	Meervoudige regressieanalyse	272
7.3.1	Dummyvariabelen	272
7.3.2	Interpretatie meervoudige regressieanalyse	274
7.3.3	Schijnsamenhang in een meervoudige regressie	282
7.3.4	Regressie- en correlatieanalyses in wetenschappelijke tijdschriften	283
7.4	Samenvatting	287

8	Schaalconstructie	289
8.1	Validiteit van een meting	289
8.1.1	Latente en manifeste variabelen	290
8.2	Betrouwbaarheid van een meting	292
8.3	Schalen	292
8.3.1	Factoranalyse	294
8.3.2	Betrouwbaarheidsanalyse: interne consistentie	302
8.3.3	Maken en beschrijven van de schaal	303
8.4	Meerdere factoren	306
8.4.1	Het roteren van de assen	310
8.4.2	Het interpreteren van de factoren en de noodzaak van rotatie	311
8.5	Gebruik en presentatie van de resultaten	316
8.6	Overige vormen van betrouwbaarheid	317
8.6.1	Stabiliteit en equivalentie	318
8.7	Samenvatting	320
	Bijlage 1 Formules	321
	Bijlage 2 Tabellen met overschrijdingskansen en kritieke waarden	325
	Literatuur	335
	Register	337
	Over de auteurs	341

Kaders

Kader 1.1	Invoeren van gegevens en het maken van een frequentietabel	22
Kader 1.2	Het maken van grafieken	25
Kader 1.3	Gedrag van mensen is bepalend voor koffiekeuze	25
Kader 1.4	Centrummaten	40
Kader 1.5	Centrum- en spreidingsmaten	52
Kader 1.6	Berekenen van z-scores	56
Kader 2.1	Representativiteitstoets χ^2 voor één variabele	103
Kader 2.2	Representativiteitstoets t-toets voor één variabele	106
Kader 3.1	Missing maken van waarden	114
Kader 3.2	Waarden in output tonen	115
Kader 3.3	Nieuwe variabele maken door middel van Compute	119
Kader 3.4	Herocoderen van variabelen	123
Kader 4.1	Kruistabellen maken	135
Kader 4.2	Berekenen van nominale associatiematen	154
Kader 4.3	Berekenen van gamma en Somers' d	170
Kader 4.4	Tabelsplitsing	180
Kader 5.1	Uitvoeren van een t-toets op één steekproef-gemiddelde in SPSS	192
Kader 5.2	Uitvoeren van een afhankelijke-toets in SPSS	201
Kader 5.3	Uitvoeren van een onafhankelijke-toets in SPSS	208
Kader 6.1	Eenwegsvariantieanalyse	229
Kader 6.2	Berekenen van een tweewegsvariantieanalyse via GLM	232
Kader 7.1	Het maken van een spreidingsdiagram	246
Kader 7.2	Het berekenen van de correlatie	253
Kader 7.3	Het berekenen van partiële correlaties	256
Kader 7.4	Het uitvoeren van een regressieanalyse	269
Kader 8.1	Uitvoeren van een factoranalyse	301
Kader 8.2	Berekenen van Cronbachs alfa	304

Inleiding

Sociale wetenschappers willen kennis vergaren over de sociale werkelijkheid. Dat kan nieuwe kennis zijn, maar ook kennis waaruit blijkt dat wat we ‘wisten’ niet of niet helemaal klopt. Kennis kun je verkrijgen door het uitvoeren van een onderzoek. Wanneer je een onderzoek hebt uitgevoerd, wil je de resultaten van het onderzoek zo duidelijk mogelijk weergeven. Dat kan op een korte en overzichtelijke manier gebeuren door middel van kengetallen, tabellen of grafieken. Dit is waar het in de beschrijvende statistiek om gaat: het samenvattend beschrijven van de kenmerken van een groep onderzoekseenheden. Veelal gebruiken we een steekproef om de data te verzamelen om tot die kennis te kunnen komen. Maar in hoeverre zeggen de steekproefgegevens iets over de hele populatie? Dat is waar het in de toetsende statistiek over gaat: een schatting doen over populatiekenmerken op basis van de steekproefgegevens.

Voor statistiek heb je cijfers nodig. In de basis gaat statistiek namelijk over tellen: hoe vaak komt iets voor, en hoe verhoudt zich dat tot andere kenmerken? Stel dat je wilt weten welke sociale mediakanalen het beste zijn voor een influencer marketingcampagne. Je zou dan eerst kunnen kijken welke platforms die influencers voornamelijk gebruiken. Als onderzoeker kun je dan een inhoudsanalyse uitvoeren onder sociale mediaplatforms en gaan tellen hoe vaak daar een influencer op voorkomt. Of je voert een enquête uit onder een aantal mensen en vraagt hen op welke platforms zij hoofdzakelijk influencers volgen. Laten we uitgaan van dit laatste voorbeeld. De data die je door de enquête verkrijgt zouden er als volgt uit kunnen zien:

Tabel 1 Beschrijving van het kenmerk favoriete sociale mediakanaal voor het volgen van influencers door middel van een frequentieverdeling.

Favoriete platform	Absolute frequentie	Percentage
YouTube	25	14,7
Instagram	60	35,3
Snapchat	10	5,9
Facebook	20	11,8
TikTok	55	32,3
Totaal	170	100

Je hebt dan een beschrijving gegeven van een van de kenmerken van je onderzoekseenheden, namelijk: op welk sociale mediakanaal zij voornamelijk influencers volgen. Maar je kunt ook andere kenmerken beschrijven. Je zou kunnen vragen naar de leeftijd van de respondenten, hoeveel influencers ze

volgen, waarom ze dat doen, in welke mate ze zich door deze influencers laten beïnvloeden, enzovoort.

Stel dat uit de enquête blijkt dat de gemiddelde leeftijd van de respondenten 20,3 is. Dan geef je met dat kengetal een samenvattende beschrijving van een specifiek kenmerk (leeftijd) van die groep. Beschrijvende statistieken zeggen dus iets over de samenstelling van de steekproef. Ze zeggen echter niet direct iets over de samenstelling van de *populatie* waaruit de steekproef afkomstig is. De sociale wetenschapper zal dus moeten nagaan in hoeverre de gevonden kengetallen in de steekproef voor de hele populatie gelden. Door gebruik te maken van de toetsende of inferentiële statistiek kun je op basis van steekproefgegevens een schatting maken van de populatiewaarden. Je kunt dan met een bepaalde mate van waarschijnlijkheid uitspraken doen over die populatie. Stel dat de eerder genoemde groep van 170 personen aselekt getrokken is uit het publiek van een festival. In dat geval kun je met een bepaalde mate van waarschijnlijkheid de gemiddelde leeftijd van het publiek schatten, of je kunt een betrouwbaarheidsinterval berekenen waarbij je na kunt gaan wat naar schatting de minimale en maximale gemiddelde leeftijd van de hele populatie van festivalgangers is.

In dit boek bespreken we de beschrijvende en toetsende (of inferentiële) statistiek. In de hoofdstukken over de verschillende statistische analyses bespreken we eerst de theoretische achtergrond. Wanneer heb je een bepaalde analysetechniek nodig en wat betekenen de gevonden kengetallen? Voor het uitvoeren van de analyses maken we gebruik van het computerprogramma SPSS, een programma dat in de sociale wetenschappen veel wordt gebruikt. In dit boek is gebruikgemaakt van SPSS 27, maar eerdere versies zijn wel te gebruiken. Het is mogelijk dat de lay-out er dan iets anders uitziet, maar de inhoud van de SPSS-tabellen is hetzelfde. In aparte kaders leggen we bij elke analyse uit hoe je die analyse zelf met SPSS kunt uitvoeren. In wetenschappelijke teksten mag je nooit SPSS-tabellen uit de SPSS-output kopiëren. Het is de bedoeling dat je zelf de relevante gegevens selecteert en presenteert in je eigen tabellen of verwerkt in de tekst van je onderzoeksverslag. Wij gebruiken in dit boek *wel* de SPSS-output om te laten zien hoe deze eruitziet en om uit te leggen welke informatie je waar in de SPSS-tabellen kunt vinden. In de komende hoofdstukken leggen we bij elke analyse uit welke gegevens uit de SPSS-output van belang zijn en hoe je die moet interpreteren. Ten slotte laten we aan de hand van voorbeelden zien hoe je de statistische gegevens rapporteert. Hiervoor bestaan in de sociale wetenschappen regels, conventies. Wij laten zien hoe de rapportage er volgens de APA-regels (versie 7) uit hoort te zien.

Voordat je kunt beginnen met het creëren van een SPSS-output en het uitvoeren van analyses moeten eerst gegevens verzameld worden en die data moet je invoeren in een datamatrix. De eerste drie hoofdstukken beschrijven de basisprincipes en het begrippenkader dat je je eigen moet maken voor je aan de analyses kunt beginnen.

In hoofdstuk 1 (*Basiselementen*) bespreken we de basisbegrippen, de datamatrix, soorten variabelen en de verschillende meetniveaus. In dit hoofdstuk komen ook eenvoudige beschrijvende statistieken aan de orde zoals centrum- en spreidingsmaten en het maken van frequentietabellen.

In hoofdstuk 2 (*Van steekproef naar populatie*) gaan we in op de begrippen die van belang zijn voor de toetsende (inferentiële) statistiek. Hier komen kansverdelingen, steekproevenverdeling, betrouwbaarheidsintervallen en het toetsen van hypothesen aan de orde. Ten slotte bespreken we in dit hoofdstuk ook representativiteitstoetsen waarmee je (soms) kunt nagaan in hoeverre je steekproef – met betrekking tot een specifieke variabele – representatief is voor de populatie.

Hoofdstuk 3 (*Bewerken en controleren van je data*) is volledig gericht op het werken met SPSS. Je leert hier hoe je een syntax maakt en gebruikt. Een syntax is de besturingstaal van SPSS, die je per SPSS-handeling of -analyse apart kunt opslaan en al of niet met (kleine) variaties kunt hergebruiken voor analyses van je databestand. In dit hoofdstuk worden ook SPSS-opdrachten behandeld waarmee je aanpassingen in je databestand kunt doorvoeren of met je variabelen kunt rekenen.

De in de sociale wetenschappen meest gebruikte analysetechnieken behandelen we in hoofdstuk 4 tot en met 8. Dit zijn *kruistabelanalyse* (hoofdstuk 4), *t-toetsen* (hoofdstuk 5), *variantieanalyse* (hoofdstuk 6), *correlatie en regressie* (hoofdstuk 7) en *schaalconstructie* (hoofdstuk 8).

Bij dit boek hoort een website met opdrachten en databestanden waarmee geoefend kan worden. De opdrachten zijn gerangschikt naar hoofdstuk en zijn gericht op het creëren van de juiste SPSS-tabellen, op het correct interpreteren van de statistieken en op het volgens de APA-regels rapporteren van de informatie.

Website

Bij dit boek hoort een website met aanvullend materiaal. Je komt op deze site door in te loggen op www.boomstudent.nl. Op pagina 4 van dit boek staat een persoonlijke inlogcode voor toegang tot de website.



Basiselementen

1

Je doet onderzoek om iets over de werkelijkheid te weten te komen. Op basis van dat onderzoek kun je dan uitspraken doen over de werkelijkheid. Daarbij moet duidelijk zijn over wie of wat je op basis van het onderzoek een uitspraak doet. Dat zijn de objecten of *onderzoekseenheden*, de personen of zaken waarover je iets zegt.

Als je op basis van een onderzoek bijvoorbeeld de conclusie trekt dat de gemiddelde leeftijd van de eerstejaarsstudenten van een universiteit 19,7 jaar is, zeg je op basis van dat onderzoek iets over eerstejaarsstudenten. Dat zijn dan de onderzoekseenheden. Van deze studenten beschrijf je een kenmerk, namelijk de leeftijd. In het onderzoek kun je meer kenmerken van de studenten hebben verzameld, zoals geslacht, vooropleiding en studiekeuze. Deze kenmerken (leeftijd, geslacht, vooropleiding en studiekeuze) zijn de *variabelen* in het genoemde onderzoek.

Onderzoekseenheden hoeven niet altijd personen te zijn. Je kunt op basis van een onderzoek ook uitspraken doen over de lengte van voorpagina-artikelen in dagbladen. In dat geval zijn de voorpagina-artikelen de onderzoekseenheden, de objecten waarover je een uitspraak doet. De lengte is hier een kenmerk van de onderzochte artikelen en is daarom een variabele in het onderzoek. Met behulp van statistiek kun je precieze uitspraken doen over de kenmerken van onderzoekseenheden, zoals ‘de gemiddelde leeftijd van eerstejaarsstudenten is 19,7 jaar’ en ‘de meeste voorpagina-artikelen zijn korter dan twee kolommen’. Een ander voorbeeld: ‘televisiekijkers met een hoge opleiding kijken vaker naar het nieuws dan televisiekijkers met een lage opleiding’. In deze uitspraak wordt iets gezegd over televisiekijkers. In het onderzoek zijn televisiekijkers de onderzoekseenheden. In het voorbeeld zijn de kenmerken van die televisiekijkers: opleiding en de frequentie waarmee naar het nieuws wordt gekeken. ‘Opleiding’ en ‘frequentie nieuws kijken’ zijn de variabelen in het onderzoek. Meer voorbeelden van onderzoekseenheden en variabelen staan in paragraaf 1.3.

1.1 Datamatrix

Variabelen, de kenmerken van onderzoekseenheden, kunnen verschillende waarden hebben. Bij sommige kenmerken zijn de waarden al een getal, bij andere kenmerken zou je voor de voorkomende categorieën een getal kunnen

verzinnen. De waarden van bijvoorbeeld het kenmerk leeftijd zijn getallen die direct gerelateerd zijn aan de werkelijkheid. Als een persoon 21 jaar oud is, is het logisch dat deze persoon de waarde 21 krijgt voor de variabele 'leeftijd in jaren'. Maar de variabele geslacht heeft geen vaststaande numerieke waarde. Om in de statistiek toch op een geordende wijze iets te kunnen zeggen over de onderzoekseenheden, krijgen de categorieën 'man', 'vrouw' en 'anders' waarin de variabele 'geslacht' bijvoorbeeld kan worden onderverdeeld, wel een numerieke waarde om de dataverwerking te vergemakkelijken. Je zou kunnen besluiten vrouwen de waarde 1 te geven, mannen de waarde 2 en anders de waarde 3. Op die manier kun je alle onderzoekseenheden voorzien van een numerieke waarde voor het kenmerk 'geslacht'.

Al deze kenmerken van onderzoekseenheden kun je onderbrengen in een *datamatrix*. Een datamatrix is een spreadsheet waarin per onderzoekseenheid alle kenmerken als afzonderlijke variabelen worden beschreven. De onderzoekseenheden staan in de rijen van de datamatrix en de variabelen in de kolommen (tabel 1.1).

Stel dat je lekker op een terrasje zit met je vrienden. Je vraagt je af of je vrienden naar dezelfde televisiezenders kijken als jij. Om dat uit te vinden, maak je een lijstje met een aantal televisiezenders waarnaar je zelf regelmatig kijkt en gaat iedereen vragen hoeveel uur zij ongeveer per week naar die zender kijken. Daarbij schrijf je ook op welke leeftijd die persoon heeft en of het een man of een vrouw is. Omdat je alles in getallen wilt uitdrukken, stel je dat als iemand vrouw is zij de waarde 1 krijgt, en als iemand man is de waarde 2 (we hadden ook vrouw de waarde 2 kunnen geven en man de waarde 1, of een symbool kunnen toevoegen, waarbij vrouw = ♀ en man = ♂). De datamatrix ziet er dan als volgt uit.

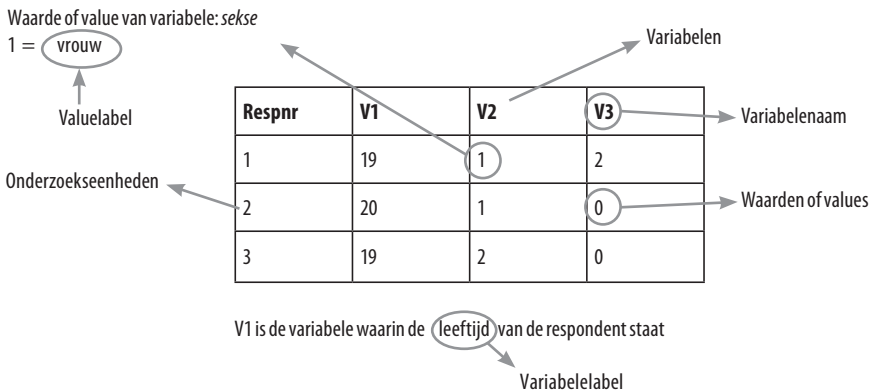
Tabel 1.1 Voorbeeld van een datamatrix

Persoon	Leeftijd	Sekse	NPO1	NPO3	RTL4	RTL5	Net5	BBC
1	19	1	2	0	2	2	4	0
2	20	1	0	0	1	2	2	1
3	19	2	0	1	1	1	0	3
4	22	1	3	2	2	3	3	2
5	24	2	1	0	3	1	1	1
6	21	2	0	3	2	0	1	1
7	20	1	2	2	1	2	3	2

In de eerste rij staan per kolom de namen van de variabelen die je in je onderzoek hebt gemeten. In de cellen daaronder staan de waarden die de respectievelijke onderzoekseenheden hebben op die variabelen. Persoon 1 is dus een 19-jarige vrouw, die twee uur naar NPO1, RTL4 en RTL5 kijkt, vier uur naar Net5, en niet naar NPO3 en de BBC.

Voor elke afzonderlijke variabele kun je een *frequentieverdeling* maken om uitspraken te doen over de percentuele verdeling van de onderzoekseenheden over de waarden van die variabele. In dit voorbeeld, waarbij we ons hebben beperkt tot zeven onderzoekseenheden, is 42,9% man (drie van de zeven personen) en 57,1% vrouw (vier van de zeven personen). Voor elke variabele (elke kolom in de datamatrix) kun je een dergelijke frequentieverdeling maken.

Om met de terminologie wat vertrouwd te raken laten we nogmaals een klein gedeelte van de datamatrix zien, en zullen we stilstaan bij de verschillende begrippen die hierbij horen. We hadden al gezien dat in de rijen de onderzoekseenheden staan, en dat deze variëren op een aantal kenmerken (vandaar ook de naam: variabelen). In de datamatrix in tabel 1.1 is duidelijk te zien wat er met de variabelen bedoeld wordt. Met 'leeftijd' wordt de leeftijd van de respondent bedoeld, en met 'seks' het geslacht van de respondent. Met 'NPO1' kunnen echter verschillende dingen worden bedoeld. Is de onderzoekseenheden gevraagd hoeveel uur ze per week kijken? Of per dag? Of misschien is hun wel een schaal voorgelegd waarop ze konden antwoorden van 0 = nooit tot 5 = heel vaak. De namen die boven de kolommen van een datamatrix staan, zijn de *variabelenamen*. We zouden ook voor heel andere variabelenamen kunnen kiezen, zoals voor V1, V2 en V3, omdat deze informatie bijvoorbeeld correspondeert met respectievelijk vraag 1, vraag 2 en vraag 3 in een vragenlijst (zie figuur 1.1).



Figuur 1.1 Terminologie bij datamatrix

V3 is een variabelenaam. Wat je bedoelt met die variabelenaam, maak je kenbaar in het *variabelelabel*. Het label van V1 is in voorstaand voorbeeld leeftijd in jaren. Het label van V2 is seks, en van V3 is het label aantal uur dat naar NPO1 wordt gekeken. De getallen die in de matrix staan, noemen we *waarden* of, in het Engels, *values*. De waarden die bij variabele V1 horen (leeftijd), zijn gemakkelijk te interpreteren: 19 betekent dat deze respondent 19 jaar oud is. De waarde 1 van de variabele V2 is niet direct te interpreteren. Daarom worden ook de waarden voorzien van een label: het *valuelabel*. Daarin geef je aan wat je bedoelt met de waarden. In ons voorbeeld is het valuelabel van de waarde 1 van

de variabele V2: vrouw. Het valuelabel van de waarde 2 van de variabele V2 is man. Om te begrijpen wat er met de verschillende variabelenamen, variabelenlabels en valuelabels wordt bedoeld, is het nodig om deze informatie ergens te vermelden. Dit kan bijvoorbeeld in een codeboek. In SPSS kun je deze informatie zelf toevoegen in het tabblad 'Variable View' (zie kader 1.1).

1.2 Frequentietabellen

Een variabele (kenmerk) met de daarbij behorende waarden kun je op een overzichtelijke manier presenteren in een *frequentietabel*. Stel dat je in de zomer met wat vrienden op een terrasje zit en jij bent aangewezen om de drankjes te gaan halen. Je kunt proberen alles te onthouden, maar op een bierviltje de drankjes turven is gemakkelijker. Door te turven maak je een overzicht van het aantal keer dat een waarde voorkomt. Het tellen van de streepjes brengt je op de *absolute frequentie*.

Tabel 1.2 Turven en tellen van drankjes

Drankje	Aantal (geturfd)	Absolute frequentie	Percentage
Bier		8	47,1
Rosé		5	29,4
Cola Light		1	5,9
Cappuccino		3	17,6
Totaal		17	100

Dit is de basis voor het opstellen van een frequentietabel. Uit tabel 1.2 blijkt dat van de zeventien mensen acht een biertje willen, vijf een rosé enzovoort.

Een absolute frequentie kan moeilijk te interpreteren zijn, zeker wanneer je meerdere frequentieverdelingen met elkaar wilt vergelijken. Daarom is het handig om naast de absolute frequenties percentages te geven. De percentages bereken je door de absolute frequentie waarmee een specifieke waarde voorkomt te delen door het totaal aantal eenheden. In vorenstaand voorbeeld heeft

$8/17 = 0,471 = 47,1\%$ van je vrienden een biertje besteld.

De week daarop zit je weer op een terras, maar nu met 22 vrienden. Het aantal bier, rosé en cola light is hetzelfde, maar in plaats van drie, worden nu acht cappuccino's besteld.

Absoluut gezien worden dezelfde hoeveelheden bier, rosé en cola light besteld, maar relatief gezien (kijkend naar de percentages, rekening houdend met het totale aantal drankjes) wordt er minder bier, rosé en cola light besteld.

Uit tabel 1.3 blijkt dat nu 36,4% een biertje bestelt: $8/22 = 0,364$.

Tabel 1.3 Aantal drankjes (absoluut en in percentages)

Drankje	Aantal (geturfd)	Absolute frequentie	Percentage
Bier		8	36,4
Rosé		5	22,7
Cola Light		1	4,5
Cappuccino		8	36,4
Totaal		22	100

1.2.1 Hoe ziet dit eruit in SPSS?

Stel, je wilt van de zeventien vrienden tijdens het eerste terrasbezoek weten hoe oud ze zijn. Je pakt weer je bierviltje, vraagt ieders leeftijd en gaat turven. Later die dag voer je je gegevens in SPSS in, en je laat SPSS een frequentietabel maken (zie kader 1.1). Zoals is te zien in tabel 1.4, geeft SPSS naast de absolute frequentie (*Frequency*) en het percentage (*Percent*) ook het geldige percentage (*Valid Percent*) en het cumulatieve percentage (*Cumulative Percent*).

Tabel 1.4 Frequentieverdeling van de variabele leeftijd (SPSS-output)

		leeftijd			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	19	4	23,5	23,5	23,5
	20	5	29,4	29,4	52,9
	21	3	17,6	17,6	70,6
	22	4	23,5	23,5	94,1
	24	1	5,9	5,9	100,0
Total		17	100,0	100,0	

In de eerste kolom (*Valid*) zie je de waarden die de variabele leeftijd in ons voorbeeld heeft. Je vrienden zijn 19, 20, 21, 22 of 24 jaar oud. In de tweede kolom (*Frequency*) staan de absolute frequenties, het aantal keer dat een bepaalde waarde voorkomt. In dit geval zijn de percentages in de vierde kolom (*Valid Percent*) identiek aan de percentages in de derde kolom (*Percent*). In paragraaf 1.2.2 zullen we bespreken in welke situaties dit niet het geval is. In deze kolommen kunnen we aflezen dat 23,5% van de onderzoekseenheden 22 jaar

is. In de kolom *Cumulative Percent* worden de percentages van elke volgende waarde bij de voorgaande opgeteld ($23,5 + 29,4 = 52,9$ enzovoort). Je zou aan de hand van deze kolom kunnen concluderen dat 52,9% van de onderzoekseenheden (in dit geval je vrienden op het terrasje) 20 jaar of jonger is.



SPSS

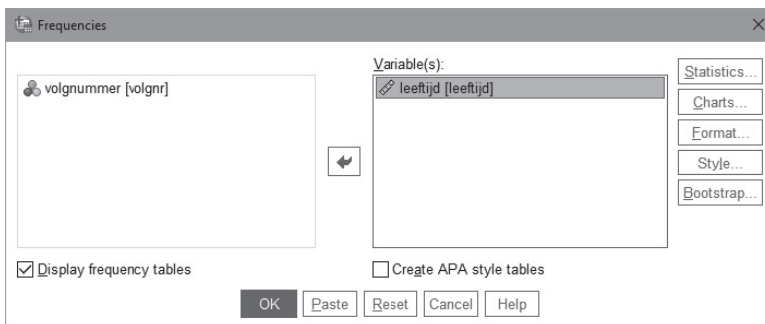
Invoeren van gegevens en het maken van een frequentietabel

Het invoeren van gegevens gebeurt in de Data View van SPSS. Per onderzoekseenheid kun je bijvoorbeeld na een uniek volgnummer voor elke onderzoekseenheid de waarde voor de variabele leeftijd intikken (zie figuur A). De namen van de variabelen (hier: 'volgnr' en 'leeftijd') kun je invoeren op de pagina die achter deze datamatrix ligt (klik linksonder op Variable View).

	volgnr	leeftijd	var	var	var
1	1,00	19,00			
2	2,00	20,00			
3	3,00	22,00			
4	4,00	21,00			
5	5,00	24,00			
6	6,00	22,00			
7	7,00	19,00			

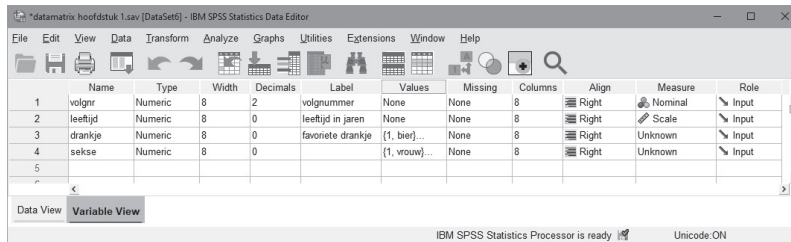
Figuur A Datamatrix

Als de data in SPSS zijn ingevoerd, kun je om een frequentietabel vragen. Ga via *Analyze* → *Descriptive Statistics* → *Frequencies* om vervolgens in het *Frequencies*-venster (zie figuur B) de variabelen te selecteren waar je een frequentietabel van wilt. SPSS maakt zelf het outputbestand waarin je deze tabel kunt vinden.



Figuur B Frequencies-venster

Door in Data View van SPSS links onderin op 'Variable View' te klikken, krijg je een overzicht van jouw variabelen te zien.



Figuur C Variable View

Kader 1.1

1.2.2 Missing values

Je vrienden willen je best vertellen hoe oud ze zijn. Wanneer je echter willekeurig mensen op een terras gaat vragen hoe oud ze zijn, kan het gebeuren dat ze je dat niet willen vertellen. Dan heb je voor die personen (onderzoekseenheden) dus geen informatie over het kenmerk leeftijd. Deze ontbrekende waarden noem je *missing values*. Je neemt deze mensen wel mee in je onderzoek naar de kenmerken van de personen op een terras. Maar als je wilt weten met welk percentage elke leeftijd vertegenwoordigd is, wil je soms niet dat deze onderzoekseenheden met onbekende leeftijden meetellen bij de berekening van de percentages. In tabel 1.5 is te zien hoe dit er in SPSS uitziet.

Tabel 1.5 Frequentieverdeling naar leeftijd met missing values (SPSS-output)

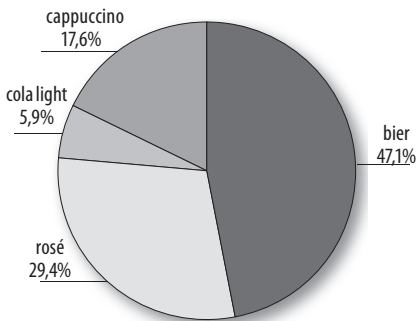
		leeftijd			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	19,00	4	20,0	23,5	23,5
	20,00	5	25,0	29,4	52,9
	21,00	3	15,0	17,6	70,6
	22,00	4	20,0	23,5	94,1
	24,00	1	5,0	5,9	100,0
	Total	17	85,0	100,0	
Missing	System	3	15,0		
Total		20	100,0		

In totaal zitten twintig personen op het terras. Daarvan hebben drie mensen geen antwoord gegeven op de vraag naar hun leeftijd. Er is nu een verschil tussen Percent en Valid Percent. Bij Percent worden deze mensen namelijk wel meegerekend (15% van de ondervraagden heeft geen antwoord gegeven). Het percentage 19-jarigen van alle mensen op het terras is 20%. Of dat een zinnig

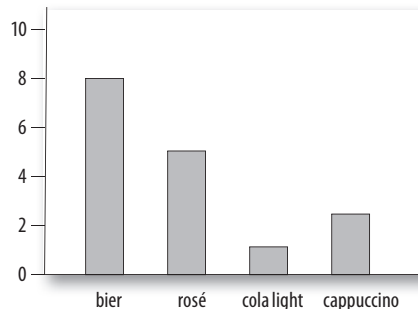
percentage is, is nog maar de vraag, want de drie die geen antwoord hebben gegeven zouden ook 19 jaar kunnen zijn, maar informatie daarover ontbreekt. Daarom kun je in dit geval beter kijken in de kolom met Valid Percent, het geldige percentage. Hierin worden de mensen die geen antwoord hebben gegeven niet meegerekend. Dan kun je constateren dat 23,5% van de mensen op het terras die deze vraag hebben beantwoord 19 jaar is. In paragraaf 3.2 wordt verder ingegaan op hoe je in SPSS waarden missing kunt maken en wat daar de consequenties van kunnen zijn in je onderzoek.

1.2.3 Grafieken

Een frequentietabel kun je ook grafisch weergeven. Een grafiek geeft geen extra informatie, maar kan visueel snel duidelijk maken wat de verdeling is van de waarden van een variabele. Bij een frequentieverdeling is het mogelijk om een taart- of staafdiagram te gebruiken. In figuren 1.2 en 1.3 wordt visueel duidelijk gemaakt welke dranken door veel en welke dranken door weinig van je vrienden zijn besteld. De diagrammen geven de verdeling van de percentages of de frequenties weer.



Figuur 1.2 Taartdiagram drankjes



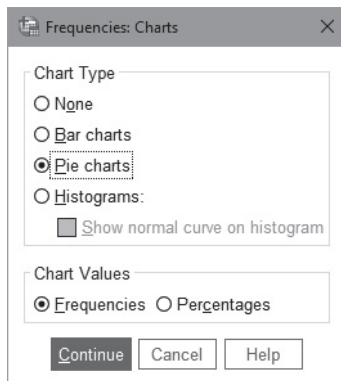
Figuur 1.3 Staafdiagram drankjes

In het taartdiagram kun je zien dat de meeste mensen een biertje hebben besteld (47,1%) en dat cola light door het kleinste aantal mensen is besteld (5,9%). Ditzelfde is af te lezen in het staafdiagram: acht mensen bestelden een biertje, en één persoon een cola. Een taart- of cirkeldiagram wordt meestal gebruikt om aan te geven wat relatief vaak voorkomt (percentages), terwijl staafdiagrammen meestal worden gebruikt om absolute aantallen weer te geven. Beide worden alleen gebruikt wanneer een variabele niet zo heel veel waarden heeft. Wanneer je in een enquête 200 onderzoekseenheden naar hun leeftijd vraagt en daar vijftig verschillende leeftijden uitkomen, is het niet erg overzichtelijk om daar een cirkeldiagram van te maken.



SPSS kan op verschillende manieren een grafiek voor je maken. De eenvoudigste manier is om dit te doen bij het maken van een frequentietabel. Via *Analyze* → *Descriptive Statistics* → *Frequencies* geef je eerst aan dat je een frequentieverdeling wilt maken. In het betreffende scherm (zie kader 1.1 figuur B) kun je nu *Charts* aanklikken. Vervolgens kun je in het *Charts*-venster (zie figuur A) aangeven of je een staafdiagram (*Bar charts*), een taartdiagram (*Pie charts*) of een histogram (*Histograms*) wilt. Histogrammen zullen besproken worden in paragraaf 1.11. Onder *Chart Values* geef je aan of dit in absolute frequenties of in percentages moet worden weergegeven.

Door in de output van SPSS dubbel te klikken op het figuur is het mogelijk om de tekst te bewerken of de arceringen te veranderen.



Figuur A Charts-venster

Een andere manier om grafieken te maken is via *Graphs*. Deze optie geeft veel verschillende soorten grafieken, die hier niet nader worden besproken.

Kader 1.2

1.3 Variabelen

We hebben gezien dat variabelen de kenmerken van onderzoekseenheden zijn. Laten we om dit nog verder te verduidelijken eens kijken naar het artikel in kader 1.3.



Cappuccinodrinkers vertonen eerder obsessief gedrag dan mensen die café latte drinken. Lattedrinkers zijn eerder geneigd anderen tevreden te stellen, terwijl mensen die hun koffie zwart drinken sneller humeurig, direct en in zichzelf gekeerd zijn.

Dit blijkt uit onderzoek waarbij duizend koffiedrinkers werden geobserveerd. Tijdens de observaties en de enquête werd gekeken naar zowel persoonlijke als psychologische eigenschappen waaronder introversie/extraversie, geduld, perfectionisme en gevoeligheid.

Tijdens de enquête moesten de ondervraagden zich inleven in tal van verschillende situaties. Achteraf werd gevraagd wat voor soort koffie zij het liefst dronken. Op basis hiervan kon Dokter Ramani Durvasula een link leggen tussen het gedrag van mensen en de koffie die zij drinken.

Bron: nu.nl, 20 september 2013, lifestylepagina

Kader 1.3

Zoals gesteld hebben onderzoekseenheden verschillende kenmerken die we variabelen noemen. In het nieuwsartikel ‘gedrag van mensen is bepalend voor koffiekeuze’ zijn de onderzoekseenheden koffiedrinkers. Mensen die geen koffie drinken worden in dit onderzoek namelijk niet meegenomen in de resultaten. Er zijn duizend mensen (we zeggen ook wel: $n = 1000$) geobserveerd (een van de methodes van onderzoek) en er is bij hen een enquête afgenomen (een andere methode van onderzoek). De eerste variabele waarop de koffiedrinkers van elkaar verschillen, is ‘het soort koffie dat zij het liefst drinken’. Deze variabele heeft in dit geval drie categorieën, drie waarden, zij zijn namelijk cappuccinodrinkers, lattedrinkers, of zij drinken hun koffie zwart. Deze informatie kan zowel door de observatie zijn verkregen, als door de enquête die zij hebben ingevuld.

In het artikel worden meerdere variabelen genoemd, namelijk de mate van obsessief gedrag (sommige koffiedrinkers zijn meer obsessief dan anderen), de mate van anderen tevreden willen stellen (bijvoorbeeld gemeten op een schaal van in zeer kleine mate tot en met in zeer grote mate), de mate van humeurigheid, de mate van directheid en de mate van in zichzelf gekeerd zijn.

De manier waarop je de variabelen meet, is bepalend voor de rest van je onderzoek. Meestal baseer je dit op voorgaand (wetenschappelijk) onderzoek, en/of op theorieën die je toepast op het onderwerp. Het meetbaar maken van je variabelen in één of meerdere vragen wordt *operationalisatie* genoemd. In het methode-deel van je onderzoek neem je een gedeelte op waarbij je verantwoordt welke vragen je stelt om welke variabelen te meten en welke antwoorden (waarden) je opneemt. Sommige variabelen (zoals demografische kenmerken) hoeven niet altijd helemaal onderbouwd te worden. Een variabele als ‘leeftijd’ kan immers niet op veel andere manieren uitgelegd worden dan ‘het aantal jaren dat iemand leeft’. Maar bij sommige begrippen is dat moeilijker. Wanneer je een inhoudsanalyse doet naar de mate van seksisme in tijdschriften, moet je eerst duidelijk maken wat je onder seksisme verstaat, en vervolgens moet je daar een of meer items bij maken waarmee je de mate van seksisme wilt vaststellen. Je noemt dit dan een *latente variabele*: een variabele die je niet direct met één variabele kunt meten, maar waarbij je meerdere variabelen nodig hebt om dit vast te stellen. Voor de operationalisatie van seksisme in tijdschriften gebruik je bijvoorbeeld twee items (variabelen) waarmee je aangeeft of er in het tijdschrift:

- vrouwen in een ‘typisch vrouwelijk’ beroep zijn afgebeeld (0 = nee, 1 = ja);
- mannen in een ‘typisch mannelijk’ beroep zijn afgebeeld (0 = nee, 1 = ja).

NB: Voor de codeur die dit moet gaan verwerken is het dan ook handig een lijst met ‘typisch vrouwelijke’ en ‘typisch mannelijke’ beroepen te hebben.

1.3.1 *Afhankelijke en onafhankelijke variabelen*

In wetenschappelijk onderzoek waarbij kwantitatieve methoden worden gebruikt, wordt veelal nagegaan wat de verbanden zijn tussen kenmerken van onderzoekseenheden. Soms is er niet alleen sprake van een verband of samenhang, maar oefenen de kenmerken invloed uit op elkaar. Als je twee variabelen met elkaar in verband brengt, kan de één afhankelijk en de ander onafhankelijk zijn.

Je zou je kunnen voorstellen dat welke tijdschriften je leest, afhankelijk is van je leeftijd, of dat het geloof dat je hebt invloed uitoefent op de krant die je leest, of de partij waarop je stemt. De variabele die invloed uitoefent, is de *onafhankelijke variabele* en wordt in de beschrijvende statistiek meestal aangegeven door een x . Als de waarde van deze x verandert, heeft dat gevolgen voor de andere variabele. De variabele die wordt beïnvloed, is de *afhankelijke variabele* (meestal aangegeven door een y). In het voorbeeld van leeftijd en welk tijdschrift je leest, is leeftijd dus de onafhankelijke variabele en het type tijdschrift de afhankelijke variabele. De keuze voor een tijdschrift wordt (mede) beïnvloed door de leeftijd. In dit geval is het moeilijk voor te stellen dat het andersom zou kunnen zijn. De keuze voor een tijdschrift kan immers nooit je leeftijd beïnvloeden. Maar je kunt wel beredeneren dat kinderen niet *Elsevier* en wel *Donald Duck* lezen, terwijl dat bij ouderen eerder andersom is.

Als we nog eens kijken naar het voorbeeld van de koffiedrinkers, zien we dat in dit onderzoek ‘soort koffiedrinker’ de onafhankelijke variabele is, en dat de andere variabelen (obsessief gedrag, humeurigheid enzovoort), de afhankelijke variabelen zijn. Er wordt immers gesteld dat als iemand een cappuccinodrinker is, iemand eerder obsessief gedrag vertoont dan wanneer iemand een lattedrinker is. De variabele ‘mate van obsessief gedrag’ wordt hier dus beïnvloed door het soort koffiedrinker dat een persoon is.

Er is niet altijd een onderscheid in onafhankelijk en afhankelijk te maken bij de analyse van het verband tussen variabelen. Als je kijkt of er een verband is tussen het aantal uren dat iemand achter de computer zit en het aantal uren dat iemand televisiekijkt, is niet duidelijk wat nu wat beïnvloedt. Wanneer je veel achter de computer zit, heb je minder tijd om televisie te kijken. Maar andersom is het ook waar: als je veel televisie kijkt, heb je minder tijd om achter de computer te zitten.

Sommige variabelen zijn bijna altijd onafhankelijk, zoals leeftijd en sekse. Er zijn namelijk maar weinig factoren die je leeftijd kunnen beïnvloeden, of je geslacht. Hoe vaak je ook achter de computer zit, je leeftijd of geslacht zal er immers nooit door veranderen.

Wat de afhankelijke en wat de onafhankelijke variabele is, zal vaak blijken uit wat de onderzoeker wil weten. De volgende voorbeelden van onderzoeksvragen maken dat duidelijk:

- In welke mate heeft woonplaats invloed op het inkomen dat iemand verdient?
- In hoeverre wordt de krant die iemand leest bepaald door zijn inkomen?
- Is er een verband tussen iemands favoriete televisieserie en zijn favoriete boekgenre?

Bij de eerste vraag ga je ervan uit dat woonplaats invloed heeft op het inkomen dat iemand verdient. Het inkomen is hoger of lager voor mensen met een verschillende woonplaats. Je gaat daarbij impliciet uit van een theorie die de hoogte van het inkomen verklaart door de woonplaats. Woonplaats is hier dan ook de onafhankelijke variabele (x), en inkomen de afhankelijke variabele (y). In de tweede vraag is het inkomen juist de onafhankelijke variabele (x). De vraag is of de krant die mensen lezen anders is voor de verschillende inkomensgroepen. In dit geval heeft de onderzoeker kennelijk overwegingen die de invloed van het inkomen op de keuze voor een krant aannemelijk maken. In de laatste vraag is er geen (on)afhankelijke variabele. Er is alleen sprake van een verband tussen favoriete televisieserie en boekgenre. De onderzoeker zag geen aanleiding in de vraagstelling een richting aan te geven in het verband tussen televisieserie en boekgenre. De favoriete serie van iemand zou het favoriete boekgenre kunnen beïnvloeden, maar het omgekeerde kan ook het geval zijn. Mensen die een voorkeur hebben voor een bepaald boekgenre zullen televisieseries kijken die daarbij aansluiten.

1.4 Meetniveaus

De manier waarop je een kenmerk meet, bepaalt ook het meetniveau van de variabele. Het meetniveau van de variabele bepaalt welke analyses wel en welke analyses niet mogelijk zijn. Hoe hoger een meetniveau, hoe meer mogelijkheden er zijn. Er zijn vier meetniveaus: nominaal, ordinaal, interval en ratio.

1.4.1 Nominaal meetniveau

Het meest elementaire meetniveau kenmerkt zich doordat je niet kunt rekenen met de waarden die je aan de variabelen hebt gegeven. De numerieke waarde is slechts een naamgeving en heeft als getal geen betekenis. De drankjes die geturfd zijn (tabel 1.2) zijn een voorbeeld van een nominaal meetniveau. Je kunt voor de verschillende drankjes een waarde kiezen (1 = bier, 2 = rosé, 3 = cola light, 4 = cappuccino), maar je had net zo goed andere getallen, letters of een symbool kunnen gebruiken: ☕ = cappuccino, = 🍷 rosé enzovoort). De gekozen waarde-aanduidingen onderscheiden de verschillende soorten drankjes. Je

kunt niet spreken van een rangordening in die drankjes. Bier (1) is niet meer of minder dan rosé (2), cola light (3) of cappuccino (4). De volgorde in deze getallen kwam toevallig zo uit, omdat bier het eerste drankje was waarvoor je een waarde moest kiezen.

Andere voorbeelden van een nominaal meetniveau zijn geslacht, politieke partij, beroep, religie, woonplaats, favoriete televisiezender, type koffiedrinker en de krant die je leest.

1.4.2 *Ordinaal meetniveau*

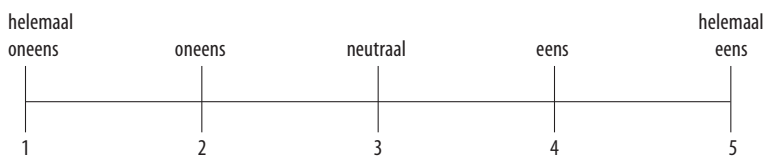
Bij een ordinaal meetniveau is wel sprake van een rangordening. De intervallen tussen de waarden hebben bij ordinale variabelen echter geen betekenis. Een voorbeeld van een ordinale variabele is opleiding, waarbij de respondenten de hoogste opleiding opgeven die ze hebben afgerond.

Je hebt ervoor gekozen om de variabele opleiding de volgende waarden toe te kennen:

- 1 vmbo;
- 2 havo;
- 3 vwo;
- 4 hbo;
- 5 wo.

Iemand met een vmbo-opleiding heeft een lagere opleiding genoten dan iemand met een havo-opleiding, die weer een lagere opleiding heeft dan iemand die vwo heeft gedaan. De volgorde is bij een ordinaal meetniveau dus wel van belang. De intervallen zijn echter niet gelijk. De afstand tussen vmbo (1) en havo (2) is niet net zo groot als de afstand tussen havo (2) en vwo (3). Het is ook niet zo dat wanneer je een wo-diploma hebt, je vijf keer hoger opgeleid bent dan iemand met een vmbo-diploma, omdat wo hier toevallig de waarde 5 heeft en vmbo de waarde 1. In plaats van de waarden 1, 2, 3, 4 en 5 had je ook de waarden 1, 4, 5, 8 en 9 kunnen kiezen voor de respectievelijke opleidingsniveaus. Wel is het belangrijk dat de waarden op blijven lopen: een hogere opleiding moet ook een hogere waarde hebben.

Wanneer je in een vragenlijst voor de beantwoording van een vraag een schaal gebruikt (zie figuur 1.4), heeft ook deze variabele een ordinaal meetniveau. Ook hier hebben de afstanden tussen de waarden geen betekenis. In dit geval geldt: hoe hoger je scoort, hoe meer je het met de vraag eens bent. Maar het verschil tussen 'oneens' en 'neutraal' is niet net zo groot als tussen 'eens' en 'helemaal eens', bijvoorbeeld. Je kunt ook niet zeggen dat als je op deze schaal '4' scoort, je het dan twee keer zoveel eens bent met de stelling in vergelijking met iemand die op deze schaal '2' scoort. Dat komt omdat de afstanden tussen de waarden betekenisloos zijn.



Figuur 1.4 Ordinale variabele: schaal

Andere voorbeelden van variabelen die op ordinaal niveau gemeten zijn, zijn: inkomensklassen (minder dan 1000, 1000 tot en met 2000, en meer dan 2000 euro per maand), leeftijdsgroepen (jonger dan 25, 25 tot 44, 45 tot 64 en 65 jaar en ouder), frequentie bioscoopbezoek (als dit niet in absolute aantallen maar als volgt is gemeten: eenmaal per week, twee- à driemaal per maand, eenmaal per maand, minder vaak dan eenmaal per maand).

Variabelen met een nominaal of ordinaal meetniveau noemen we ook wel ‘categorisch’. Het gaat hierbij namelijk vooral om de verschillende categorieën die je kunt onderscheiden door middel van verschillende cijfers, maar je kunt met deze cijfers niet rekenen.

1.4.3 Interval meetniveau

Als variabelen op intervalniveau gemeten zijn, is er niet alleen sprake van rangordening, maar hebben de intervallen tussen de verschillende waarden die een variabele aan kan nemen ook een betekenis. Een veelgebruikt voorbeeld is temperatuur. Het verschil tussen 5 en 10 °C is even groot als het verschil tussen 10 en 15 °C (namelijk 5 °C). De vaste meeteenheid voor graden geeft betekenis aan de verschillen tussen de waarden.

Wat je echter niet kunt zeggen is dat 20 °C twee keer zo warm is als 10 °C. Dit komt door het ontbreken van een natuurlijk (of absoluut) nulpunt. Het nulpunt bij graden Celsius is namelijk arbitrair. Er zijn meer manieren om temperatuur te meten, zoals door middel van graden Fahrenheit. Bij meting in graden Fahrenheit is er een ander nulpunt en zijn de intervallen tussen de graden anders dan bij graden Celsius. Wanneer het in Amerika tien graden warmer wordt, is die temperatuur in de regel gemeten in Fahrenheit. Wanneer in Nederland de temperatuur met tien graden stijgt, is dit niet dezelfde warmtestijging, omdat wij hier in graden Celsius rekenen.

Variabelen als inkomensklassen en leeftijdsgroepen kun je ook op intervalniveau meten als je ervoor zorgt dat de afstanden tussen de waarden altijd even groot zijn. Stel, je kiest de waarden voor de variabele leeftijdsgroepen als volgt:

Beschrijvende en toetsende statistiek heb je vaak nodig bij het opzetten en uitvoeren van onderzoek, en bij het interpreteren en presenteren van je resultaten. Met beschrijvende statistiek beschrijf je de kenmerken van een groep onderzoekseenheden. Bij toetsende statistiek doe je een schatting over populatiekenmerken op basis van de steekproefgegevens. Kennis van deze statistieken is ook nodig om wetenschappelijke artikelen goed te kunnen begrijpen.

In dit boek leggen de auteurs op heldere wijze uit hoe je de verschillende statistische analyses met SPSS kunt uitvoeren en hoe je de gevonden resultaten kunt interpreteren. Aan de hand van voorbeelden zie je hoe je de statistische gegevens vervolgens rapporteert. In de eerste drie hoofdstukken staan de basisprincipes, het begrippenkader en de uitleg over het gebruik van SPSS centraal. In hoofdstuk 4 tot en met 8 komen de meest gebruikte analysetechnieken in de sociale wetenschappen aan bod. Dit zijn kruistabelanalyse, t-toetsen, variantieanalyse, correlatie en regressie, en schaalconstructie (waaronder factoranalyse).

Op de website bij dit boek vind je opdrachten en databestanden waarmee je kunt oefenen met het maken van SPSS-tabellen, het interpreteren van de statistieken en op het rapporteren van de informatie.

Beschrijvende en toetsende statistiek is geschreven voor hbo- en wo-studenten onderzoeksmethoden en statistiek.

