

Introductie	9
--------------------	---

Hoofdstuk 1

VOORBEREIDING: DE EUROPEAN SOCIAL SURVEY	11
---	----

1.1 Toegang tot de ESS	11
1.2 De ESS downloaden	12
1.3 De ESS gebruiksklaar maken	13

Hoofdstuk 2

INLEIDING TOT DE BASISBEGRIPPEN VAN DE UNI- EN BIVARIATE STATISTIEK	15
--	----

2.1 Inleiding	15
2.2 Een classificatie van uni- en bivariate analysetechnieken	17
2.2.1 Criterium1: het belang van meetniveaus	17
2.2.2 Criterium2: de gewenste informatie	20
2.2.3 Criterium3: de gewenste presentatievorm	24
2.2.4 De keuze van techniek volgens drie criteria: een overzicht	24
2.3 Steekproef, populatie en representativiteit	26
2.4 Significantie en hypothesetoetsing	27
2.4.1 Stap 1: Het formuleren van een hypothese	30
2.4.2 Stap 2: De toetsstatistiek	32
2.4.3 De hypotheselogica toegepast op een voorbeeld	39
2.4.4 Vertaling naar de praktijk van significantietoetsen	42

Hoofdstuk 3

UNIVARIATE ANALYSE	43
---------------------------	----

3.1 Inleiding	43	5
---------------	----	---

3.2	Gegevens op Nominaal niveau	45
3.2.1	Frequentietabel	45
3.2.2	Grafische voorstelling	47
3.2.3	Kengetallen	53
3.2.4	Deze analyse in SAS	54
3.3	Gegevens op Ordinaal niveau	58
3.3.1	Frequentietabel	58
3.3.2	Grafische voorstelling	59
3.3.3	Kengetallen	62
3.3.4	Deze analyse in SAS	66
3.4	Gegevens op Interval & Ratio niveau	70
3.4.1	Frequentietabel	70
3.4.2	Grafische voorstelling	73
3.4.3	Kengetallen	76
3.4.4	Een voorbeeld van stijging in meetniveau	102
3.4.5	Deze analyse in SAS	106
3.5	Transformaties van variabelen	118
3.5.1	Z-transformatie	119
3.5.2	Het bereik van een index of somschaal wijzigen	121
3.5.3	Deze analyse in SAS	129
3.6	Speciale tabellen	135
3.6.1	Tabellen voor variabelen met gelijke antwoordcategorieën (Tables of frequencies)	135
3.6.2	Tabellen voor vragen met meerdere antwoordmogelijkheden (Multiple dichotomy)	137
3.6.3	Deze analyse in SAS	142
Hoofdstuk 4		
BIVARIATE ANALYSE		145
4.1	Inleiding	145
4.2	Associatie tussen gegevens op Nominaal niveau	149
4.2.1	Kruistabellen	149
4.2.2	Grafische voorstelling	157
4.2.3	Associatiematen	162
4.2.4	Deze analyse in SAS	171
4.2.5	Output	172
4.3	Associatie tussen gegevens op Ordinaal niveau	176
4.3.1	Kruistabellen	176
4.3.2	Grafische voorstelling	178
4.3.3	Associatiematen	181
4.3.4	Deze analyse in SAS	183

4.4	Associatie tussen gegevens op Interval & Ratio-niveau	188
4.4.1	Kruistabellen	188
4.4.2	Grafische voorstelling	190
4.4.3	Associatiematen	198
4.4.4	Deze analyse in SAS	207
4.5	Associatie tussen gegevens op verschillende meetniveaus	214
4.5.1	Kruistabellen	214
4.5.2	Grafische voorstelling	217
4.5.3	Associatiematen	218
4.5.4	Deze analyse in SAS	231
4.6	Speciale tabellen	236
4.6.1	Kruistabellen voor variabelen met gelijke antwoordcategoriegroepen	236
4.6.2	Kruistabellen voor vragen met meerdere antwoordcategoriegroepen	238
4.6.3	Deze analyse in SAS	240

Hoofdstuk 5

WETENSCHAPPELIJK RAPPORTEREN OVER UNI- EN BIVARIATE ANALYSE

		243
5.1	Inleiding	243
5.1.1	Vormvereisten voor grafieken en tabellen	243
5.1.2	Werken met percentages	245
5.1.3	Omgaan met weinig waarnemingen in categorieën	246
5.1.4	Omgaan met variabelen met veel categorieën	246
5.2	Voorbeelden van publiceerbare tabellen en grafieken	250
5.2.1	Voorbeelden van univariate tabellen	250
5.2.2	Voorbeelden van bivariate kruistabellen	253
5.2.3	Voorbeelden van grafieken	254
5.3	Tabellen en grafieken lay-outen in SPSS	254
5.3.1	Tabellen op maat met de wizard	255
5.3.2	Tabellen aanpassen in SPSS	260
5.3.3	Grafieken aanpassen in SPSS	264
5.3.4	Tabellen exporteren en bewerken in een ander programma	279
5.3.5	Grafieken exporteren en bewerken in een ander programma	282
5.4	Tot slot: een algemeen overzicht van technieken en eigenschappen	284

Introductie

Statistiek en multivariate onderzoekstechnieken zijn vaak een nachtmerrie voor studenten. Binnen de sociale wetenschappen zijn statistische vakken niet zelden vakken met lage slaagcijfers en hoge vervloekingsgraden. Minder evident maar even reëel, is de afkeer van vele onderzoekers van de technische statistische literatuur. Niet zelden laat men gevorderde multivariate technieken in het sociaalwetenschappelijk onderzoek links liggen omdat de gevorderde handboeken té wiskundig of té ontoegankelijk geschreven zijn. Binnen de faculteit Politieke en Sociale Wetenschappen was de Methodengroep jarenlang verantwoordelijk voor het adviseren en assisteren van studenten en onderzoekers bij het toepassen van eenvoudige en complexe multivariate technieken in het dagelijks onderzoek. Of het nu thesisonderzoek is of fundamenteel wetenschappelijk onderzoek, dezelfde vragen komen steeds terug. Daarom werd gestart met het ontwikkelen van een reeks MTSO-INFO. De basisfilosofie van deze reeks was statistische software en statistische analysetechnieken zodanig uit te leggen, dat wie van een absoluut nulniveau start, met een minimum aan inspanning in staat is om gevorderde analyses uit te voeren.

Toen stilaan duidelijk werd dat de reeks niet alleen binnen de eigen faculteit gebruikt werd maar ook daarbuiten, werd met Acco samengewerkt aan het publiceren van de handleidingen in de reeks *Stap In Statistiek en Onderzoek*. Het uitgangspunt van deze reeks is exact dezelfde als die van MTSO-INFO: een instap bieden voor het werken met statistische software, om zo te komen tot een begrijpen en uitvoeren van gevorderde analysetechnieken. We beperken bewust de hoeveelheid theorie. We geven kort een inleiding over een techniek en gaan dan onmiddellijk aan de slag. Daarbij wordt de wiskundige achtergrond van de technieken in sterke mate uit de weg gegaan. Wie nadien meer wil weten over de techniek, de achtergronden en wie de uitgebreidere toepassingen ervan wil leren kennen, wordt doorverwezen naar handboeken statistiek. De reeks *Stap In Statistiek en Onderzoek* wil bewust geen statistiek-

reeks zijn, maar een praktijkreeks met een lage instapdrempel. Daarom staat de output van de statistische procedures en de programma's die tot die output leiden, centraal. Het uiteindelijke doel van een analyse moet immers het begrijpen van de sociale werkelijkheid zijn. Dat begrijpen komt voort uit het correct lezen en interpreteren van statistische output. Toch willen we benadrukken dat het regelmatig consulteren van een handboek statistiek naast deze reeks noodzakelijk is om de finesses die in de reeks aan bod komen beter te begrijpen.

Basisbegrippen in de statistiek zoals steekproef en populatie, meetniveau, mediaan en modus, inferentie en hypothesetoetsen worden in vele inleidende cursussen en boeken uitgebreid toegelicht. Het gebruik van wiskundige formules en definities wordt daarin niet geschuwd. Het begrijpen van deze basistechnieken en het inzicht verwerven in de wiskundige achtergrond ervan maakt echter slechts het eerste deel uit van een lang leerproces. Na de theoretische kennis volgt het leren 'beheersen en toepassen van de technieken'. In de reeks *Stap In Statistiek en Onderzoek* vertrekken we van het standpunt dat er een meer toegepaste manier bestaat om analyses aan te leren. Door gelijktijdig in te gaan op enerzijds de doelen van de techniek en anderzijds de concrete aanpak in de statistische programma's SPSS en SAS, leiden we de gebruiker in tot de uni- en bivariate statistiek. De afwisseling van korte theoretische toelichting met praktijkvoorbeelden, stelt de lezer in staat onmiddellijk aan de slag te gaan.

Als auteurs willen we tot slot Inge Pasteels en Laurent Snoeckx bedanken voor het nalezen van eerdere versies van dit boek. Het eindresultaat blijft uiteraard de volledige verantwoordelijkheid van de auteurs.

We hopen dat u als lezer vooral *praktisch plezier* beleeft aan deze reeks. Het soort plezier dat ontstaat wanneer je voor de eerste keer een significantietoets meester wordt of een scatterplot aangemaakt hebt.

DIMITRI MORTELMANS
en BRITT DEHERTOGH

Vorbereiding: de European Social Survey

Voor alle voorbeelden in dit volume wordt gebruik gemaakt van de European Social Survey. Deze keuze is gemaakt om twee redenen. In de eerste plaats bevat de ESS gegevens die onderzoekers uit diverse disciplines kunnen interesseren. Bovendien biedt het internationale karakter van het onderzoek ook veel mogelijkheden om internationaal vergelijkende voorbeelden te geven. Een tweede reden waarom voor de ESS geopteerd werd, is het vrij beschikbare karakter van de gegevens. Iedere onderzoeker kan gratis de data van de ESS gebruiken voor het eigen onderzoek. Dat betekent dat iedereen de voorbeelden uit deze reeks zelf thuis of op het werk kan reproduceren op reële gegevens en bovendien met dezelfde resultaten tot gevolg.

In wat volgt leggen we uit waar de ESS te vinden is en hoe de gegevens die in dit boek gebruikt worden, gedownload kunnen worden.

1.1 Toegang tot de ESS

De ESS is terug te vinden op de Noorse website van het ESS-netwerk. Open bijgevolg een webbrower en surf naar volgend adres:

<http://www.europeansocialsurvey.org/>

Om de ESS te gebruiken, vraagt men om je eerst te registreren. Na de gratis registratie kan je op de site de gegevens rechtstreeks downloaden.

Om je te registreren, ga je naar het data-archief. Kies in het menu op de hoofdpagina voor “**Archive & Data**”. Vaak staat de link naar het archief ook apart op de hoofdpagina:

<http://ess.nsd.uib.no/>

Je kan de registratieprocedure starten door links in het menu te kiezen voor “Register New User”. Vul je gegevens in en druk op *Register*.

Onmiddellijk na het registreren, krijg je de boodschap dat je als nieuwe gebruiker opgenomen bent. Nu kan je de website binnengaan en de data downloaden.


1.2 De ESS downloaden

Om in te loggen, typ je vanaf het registratiescherm je emailadres in het login-vak (rechtsboven) in. Klik op **Log In**. Het vak verdwijnt en de boodschap verschijnt dat je ingelogd bent.

ESS is een zeer goed gedocumenteerd project. Onder het submenu “*Survey documentation*” en “*Fieldwork documents*” kan je alle methodologische documenten over het project bekijken en downloaden. Als je zelf met de ESS aan de slag wil gaan, bekijk dan zeker eens de documentatie over de weegvariabelen en de vragenlijst. We zijn nu geïnteresseerd om de geïntegreerde data van de eerste golf van de ESS te downloaden. Die zijn te vinden onder het submenu “**Data download**” bij het tabblad “**ESS Round 1**”.

ESS biedt de data aan in twee formaten: SAS en SPSS. Indien je het hele boek zal doornemen, heb je beide formaten nodig. We starten met het databestand in SAS-formaat. Klik¹ daarom op het SAS symbool (📄) achter “ESS1 – inte-

1. ESS brengt op regelmatige basis nieuwe versies uit van de data. Het is dus mogelijk dat er ondertussen een latere versie beschikbaar is op de website. Deze nieuwe versies geven doorgaans slechts minimale verschillen met de in dit boek gebruikte versie.

grated file, edition 5.1”. Vervolgens doe je hetzelfde met het SPSS-formaat (icoon: ).

Als de download succesvol was, staan er twee gecomprimeerde databestanden op je harde schijf met de naam: ESS1e05_1[1].sas.zip en ESS1e05_1[1].spss.zip.

Dit bestand kan je decomprimeren met Winzip of (als je Windows XP bezit) openen in Windows Verkenner.

Eens dat gedaan is, krijg je volgende lijst van bestanden:

1. De ESS-gebruiksovereenkomst:
ESSConditionsOfUse.txt
2. Het SPSS Bestand:
ESS1e05_1.por
3. De SAS Bestanden:
ESS1.sas
ESS1e05_1sas.por
ESS1formats.sas
ESS1miss.sas
ESS1standards.sas

Als je deze bestanden op je harde schijf hebt, ben je klaar om het boek aan te vatten. Voor de voorbeelden in het boek, gaan we er van uit dat deze bestanden allemaal in de directory C:\ESS staan. Uiteraard kun je de bestanden in andere directories plaatsen, maar dan moet je het pad steeds aanpassen aan je eigen situatie.

1.3 De ESS gebruiksklaar maken

Voor je in dit boek aan de slag kan met de ESS-bestanden, vereisen deze enige bewerking (zie ook Volume 1 Kennismaken met SPSS en SAS) van deze reeks voor een meer uitgebreide toelichting. Voor **SPSS** moet je enkel ESS1e05_1.por importeren en opslaan als een .sav bestand. Voor **SAS** zijn meerdere bewerkingen op de standaardbestanden van de ESS nodig. We geven een kort overzicht zodat je de bestanden kan importeren in SAS en gebruiksklaar kan maken voor de voorbeelden in dit boek. We geven de wijzigingen ten opzichte van de ESS-standaardbestanden aan in vet.

1. Wijzigingen aan ESS1.sas

```
LIBNAME TRANS SPSS "C:\ESS\ESS1e05_1sas.POR";
LIBNAME ESS "C:\ESS";

/*To create formats (value labels): edit and run ESSformats program*/
%include "C:\ESS\ESS1formats.sas";
run;

data ESS.ESS1e05_1;
set TRANS._first_;

*run;

/*To create user-defined missing values .a, .b, .c and .d*/
%include " C:\ESS\ESS1miss.sas";
```

2. Wijzigingen aan ESS1formats.sas

```
/*Permanent formats*/

LIBNAME LIBRARY "C:\ESS ";
PROC FORMAT LIBRARY = LIBRARY;

/*Temporary formats*/
*PROC FORMAT;

%include "C:\ESS\ESS1standards.sas";
```

Tot slot starten we elke sessie van SAS volgende autoexec.sas op:

```
options nocenter nodate linesize = 100 pagesize = 30000;
title;

%INCLUDE "C:\ESS\ESS1.sas";

libname ess "C:\ESS";

data ess1e05_1;
set ess.ess1e05_1;
run;
```

In deze autoexec.sas wordt een extra library *ESS* aangemaakt die verwijst naar de directory "C:\ESS". Vermits de ESS-bestanden in deze directory geplaatst werden, kan je ze aanspreken door naar de ESS-library te verwijzen. Tot slot wordt een tijdelijk bestand *ess1e05_1* aangemaakt op basis van het permanente databestand. Dat is aangewezen zodat je de originele bestanden nooit kwijtraakt, mocht je iets verkeerd doen. Je kunt dan steeds de autoexec.sas opnieuw uitvoeren om een nieuwe kopie van het databestand aan te maken.

HOOFDSTUK 2

Inleiding tot de basisbegrippen van de uni- en bivariate statistiek

2.1 Inleiding

In een volume over univariate en bivariate statistiek hoort allereerst het onderscheid thuis tussen de bestudering van *het aantal kenmerken*. Proberen we informatie weer te geven van één kenmerk of één variabele, dan beschouwen we dit als *univariate* statistiek. Twee kenmerken geven aanleiding tot *bivariate* statistiek en door middel van *multivariate* statistiek¹ kunnen we verbanden nagaan tussen meerdere variabelen. Een domein binnen de statistiek dat zich bezighoudt met de vraag of de gevonden resultaten van uni- of bivariate analyses van toepassing zijn op de hele bevolking of enkel op de steekproef is de **inferentiële statistiek** (zie 2.3).

Univariate statistiek is gericht op het *beschrijven* van een kenmerk. We hebben informatie over één eigenschap of kenmerk van de respondenten en de *verdeling* van dat kenmerk willen we aan de hand van een aantal kengetallen en frequenties beschrijven. Het doel van univariate statistiek kan zijn om de resultaten te integreren in een publicatie, maar kan evengoed zijn om een verkenning uit te voeren. Verkenning van data is erg belangrijk als we verdere analyses willen doen. Een eerste stap voor alle technieken is te weten met welke variabelen we te maken hebben. Welk verloop kent de variabele? Hoe is ze in het bestand genoteerd? ...

Bij **bivariate statistiek** kan zowel *beschrijvende* als *verklarende* kennis verzameld worden, vermits er twee variabelen met elkaar in verband worden gebracht. Een verband tussen twee variabelen veronderstelt dat er een ‘*samenhang*’ is. Samenhang wil daarbij zeggen dat een variabele wijzigt, van zodra een andere variabele van waarde verandert. Net als univariate statistiek kan bivariate statistiek verkennend zijn, in functie van andere analyses.

1. Multivariate statistiek komt niet aan bod in dit deel, wel in alle volgende volumes uit de reeks.

Uni- en bivariate technieken kunnen niet zomaar willekeurig op alle variabelen worden toegepast. De technieken zijn gebonden aan een reeks voorwaarden. We hanteren in dit volume een driedeling van criteria om methoden te categoriseren. Op elk van de criteria wordt eerst theoretisch ingegaan vooraleer de toepassing van de methoden aan te vangen. In 2.2.4 geven we een schematisch overzicht van de technieken en de criteria.

- (1) De belangrijkste voorwaarde is het **meetniveau** van de variabele(n). Beschikken we over nominale, ordinale, interval- of ratiovariabelen (zie 2.2.1)?
- (2) Een tweede criterium is de keuze van **informatie** die we wensen te verzamelen, gegeven de onderzoeksvraag of hypothese die wordt gesteld. Willen we een maat van spreiding, centraliteit, vorm, symmetrische of asymmetrische samenhang (zie 2.2.2)?
- (3) Als laatste moet men kiezen voor de **presentatievorm** om de kennis weer te geven. Bieden we een grafiek of een tabel aan (zie 2.2.3)?

We hebben er voor gekozen om geen klassieke indeling te maken. In vele statistiekboeken wordt een inhoudstafel opgemaakt door verschillende technieken achtereenvolgens op te sommen. De inhoudstafel bevat dan termen zoals t-test, mediaan, χ^2 enzovoort. Het nadeel hiervan is dat de lezer, die geen zicht heeft op de toepassingsmogelijkheden van technieken, alles moet overlopen om een keuze te maken. Onze indeling laat toe om te vertrekken van de eigenschappen van de variabelen en zo een techniek te selecteren. Omdat hierdoor onduidelijk is welke technieken we behandelen in het volume, geven we een opsomming (zie ook 5.4). Om de frequentieverdelingen te beschrijven gaan we in op frequentietabellen, maten van spreiding, centraliteit en vorm (mediaan, modus, gemiddelde, interkwartielafstand, standaardafwijking, variantie, kurtosis, scheefheid). Bij de grafische weergaven bieden we informatie over staafdiagrammen, histogrammen, lijndiagrammen, cirkeldiagrammen, boxplots en scatterplots. Verder geven we ook kruistabellen en associatiematen weer (χ^2 , Phi, Cramer's V, Kendall's Tau, correlatie, Anova). Zoals je kan merken, komen in dit volume slechts een beperkt aantal technieken aan bod. In de keuze hebben we ons laten leiden door eigen onderzoekservaring en het aanbieden van een basis, waarvan we denken dat ze zinvol is voor de onervaren gebruiker. Waar we bijvoorbeeld weinig op zullen ingaan zijn niet-parametrische testen² (enkel χ^2 komt aan bod).

De structuur van dit volume is enigszins anders dan die van de eerste twee volumes (Kennismaken met SPSS en SAS/Databeheer met SPSS en SAS). Om

2. Dit zijn technieken die gebruikt worden als de verdeling niet voldoet aan bepaalde eisen zoals normaliteit. Ook wanneer steekproeven niet onafhankelijk zijn (bijvoorbeeld bij een experiment met twee meetmomenten) is men aangewezen op niet-parametrische testen.

herhaling van theorie en interpretaties te vermijden, worden alle procedures uitvoerig uitgewerkt in SPSS en beperken we het SAS-deel tot een weergave van syntax en output. Vanaf dit volume zullen we bovendien vaker syntax gebruiken in het SPSS-deel en de invulschermen via menustructuren achterwege laten. De reden hiervoor is dat in de eerste twee volumes van de reeks, de menugestuurde aanpak telkens in functie van een beter begrip van de syntax werd weergegeven. In dit volume gaan we er van uit dat de lezer ervaring heeft met het schrijven en lezen van SPSS en/of SAS-syntax, zodat de nadruk ligt op het uitvoeren van analyses aan de hand van de programmeertaal.

2.2 Een classificatie van uni- en bivariate analysetechnieken

Dat er in de loop der jaren veel uni- en bivariate analysemethoden zijn ontwikkeld is gekend. Ongetwijfeld is kennis over het rekenkundig gemiddelde en de correlatie ruim verspreid en veelvuldig toegepast. Wanneer welke techniek van toepassing is (onder welke voorwaarden), heeft menig student al kopzorgen gegeven. We maken voor dit deel een classificatie van technieken volgens drie criteria.

2.2.1 Criterium1: het belang van meetniveaus

2.2.1.1 Eigenschappen van meetschalen en hun meetniveaus

Een van de belangrijkste eigenschappen van een variabele is het meetniveau. Het meetniveau heeft te maken met de ‘meetlat’ waarop de variabele gemeten of verzameld is. Om bepaalde kenmerken van respondenten in een databestand op te nemen worden ze vaak herleid tot getallen of codes (zie ook Stap in statistiek en onderzoek nr. 1). We geven een klein voorbeeld. Het kenmerk geslacht zal niet in een databestand neergeschreven worden als man en vrouw, maar wel in *codevorm*. Mannen krijgen dan bijvoorbeeld code 1 en vrouwen code 2 (zie ESS bestand). Om duidelijk te maken wat die codes betekenen, worden *labels* toegekend aan de codes. Het label voor code 1 is man en voor code 2 vrouw. We herhalen dit basisonderscheid tussen codes en labels, omdat het van belang is voor meetniveaus. Als er allemaal codes in het bestand staan, dan zou je kunnen denken dat de variabelen hetzelfde meetniveau hebben. Niets is minder waar. Het meetniveau van de variabelen heeft te maken met de initiële manier van meten. Is de variabele gemeten met categorieën (vb. geslacht) waaraan later codes worden gehangen, dan spreekt men van een categoriale variabele. Als de variabele gemeten is op een meetlat met diverse numerieke waarden (vb. lengte) dan hebben

we te maken met een continue variabele. Onderstaande voorbeelden zullen dit verduidelijken.

Er zijn vier meetniveaus die te verdelen zijn in de twee grote groepen:

Meetniveau	CATEGORIAAL (niet-metrisch)		CONTINU (metrisch)	
	nominaal	ordinaal	interval	ratio
Eigenschappen	Meestal beperkte categorieën, zonder ordening, geen berekeningen mogelijk	Meestal beperkte categorieën, met ordening, geen berekeningen mogelijk	Numerieke waarden, met een betekenisvolle afstand en ordening, berekeningen mogelijk	Numerieke waarden, met een betekenisvolle afstand, ordening en een nulpunt, berekeningen mogelijk
Voorbeeld	<i>Vb. geslacht</i>	<i>Vb. leeftijdsklassen</i>	<i>Vb. geboortejaar</i>	<i>Vb. lengte</i>
Voorbeeld ESS	<i>Vb. partijlidmaatschap (prtmbe)</i>	<i>Vb. subjectieve gezondheid (health)</i>	<i>Vb. jaartal laatste betaald werk (pdjbyr)</i>	<i>Vb. werkuren (wkhct)</i>

Categoriale variabelen zijn variabelen die meestal slechts een beperkt aantal meetpunten hebben en waarvan je niet kan aannemen dat deze meetpunten doorlopend zijn. Een categoriale variabele heeft categorieën, die in het bestand worden omgezet naar codes. Men meet ze niet ‘in getallen’, maar in inhoudelijke categorieën. Dit heeft als gevolg dat er nooit berekeningen, zoals optelling of aftrekking, kunnen worden gedaan op categoriale variabelen. Ze worden ook wel eens kwalitatieve variabelen genoemd. Door de ongelukkig gekozen benaming, opteren we voor de term categoriaal of niet-metrisch. Binnen de groep van categoriale variabelen zijn er de nominale variabelen en de ordinale variabelen. Nominale variabelen bevatten enkel kwalitatieve begrippen, zonder enige ordening. Geslacht is een voorbeeld van een nominale variabele. Er zijn 2 antwoordcategorieën (man en vrouw) en ze zijn niet hiërarchisch te plaatsen. Bij ordinale variabelen zit er wel een ordening in de waarden. Een voorbeeld is een variabele met leeftijdsklassen (twintigers, dertigers, veertigers, vijftigers). De leeftijdsklassen zijn gerangschikt, maar kennen nog geen continuïteit (het gaat om afgebakende klassen). **Continue** of metrische variabelen hebben deze eigenschap wel: hun verloop is gekenmerkt door een aaneengesloten reeks van meetpunten. Er kunnen berekeningen op gebeuren. Bij de continue variabelen onderscheiden we interval- en ratioschalen. Intervalschalen hebben een meeteenheid. Het gaat om numerieke gegevens, de meetlat is een geijkte schaal waarvan de verschillen tussen waarden een betekenis hebben. Een voorbeeld is

het geboortjaar. Er kan een groot bereik aan jaartallen (getallen) genoteerd worden met een vast verschil. Het verschil tussen 1991 en 1992 is bijvoorbeeld evengroot als tussen 2005 en 2006: één jaar. Ratiovariabelen hebben dezelfde eigenschappen als intervalvariabelen en kennen daarbovenop een absoluut nulpunt. De variabele kan de waarde nul aannemen en die waarde heeft een betekenis. Omdat het geboortjaar nul niet mogelijk is in hedendaagse surveys, beschouwen we dit niet als een intervalvariabele. De lengte is bijvoorbeeld wel een ratiovariabele. Het is mogelijk om lengte 0 te hebben.

2.2.1.2 Meetniveaus in de praktijk

De theoretische indeling in meetniveaus levert een mooi afgebakende vierdeling op. Als we naar de praktijk van het databestand keren, dan kan opgemerkt worden dat het merendeel van de verzamelde gegevens categoriaal zijn. Vele vragen zijn gesteld met een beperkt aantal antwoordcategorieën. Drie belangrijke opmerkingen dienen in dit opzicht gemaakt te worden.

- (1) In de eerste plaats kan gesteld worden dat de vierdeling geen rigied raamwerk mag worden. Vraagstellingsmethoden laten niet altijd toe om de meest ideale variabele te verzamelen. Er dient een afweging gemaakt te worden tussen enerzijds een vraagstelling die aanspoort tot invullen van de survey en anderzijds de eenvoud van de variabele in het bestand. Dit levert vaak een gesloten vraag met een beperkt aantal antwoordcategorieën op. Dat we daardoor beperkt zouden zijn tot technieken geschikt voor categoriale variabelen is niet volledig correct. Voor gesloten vragen met meerdere antwoordcategorieën (vanaf 6 of 7) die een ordinaal karakter kennen en waarbij we denken dat er gelijke afstanden zijn tussen de antwoorden, wordt in de sociale wetenschappen soms een continuïteit onder de antwoordpatronen verondersteld. Het gaat dus om ordinale kenmerken, waarvan we denken dat ze een eigenschap van intervalvariabelen bevatten. Niet alle ordinale kenmerken komen hiervoor in aanmerking.³ Meestal gaat het om antwoordschalen die gebruikt worden om aan te geven hoezeer men het eens is met een bepaalde uitspraak, van helemaal niet eens tot helemaal eens met voldoende tussencategorieën. We verwachten dat mensen in hun hoofd een continue balans hanteren om een keuze te maken tussen de antwoordcategorieën die worden aangeboden. Respondenten worden verplicht om binnen de mogelijkheden van de vraag te blijven. Daarvoor

3. Als we het opleidingsniveau vragen van mensen in 7 categorieën (geen opleiding, lager onderwijs, lager secundair, hoger secundair, hoger onderwijs korte type, hoger onderwijs lange type, doctoraat), dan hebben we uiteraard niet te maken met gelijke intervallen en kunnen we geen technieken voor metrische variabelen selecteren.

moeten ze echter een afweging maken tussen de categorieën, een afweging die vertrekt vanuit een beeld van continuïteit. Een hele discussie is ontstaan rond de fictieve grens in aantal antwoordcategorieën vanaf welke we een continuïteit aanvaarden.⁴ Vanuit de pragmatiek raden we aan om die grens op 6 (liefst 7) antwoordcategorieën vast te leggen voor uni- en bivariate analyses. Variabelen met minder dan 6 categorieën, zullen we als niet-metrisch behandelen. Minstens 6 categorieën en een veronderstelling van continuïteit zijn noodzakelijk om een metrische techniek toe te passen, als het inhoudelijk gewenst is. Een voorbeeld kunnen we aantreffen in ESS bij de variabele *stflife* (How satisfied with life as a whole). Het bereik loopt van 0 (extremely dissatisfied) tot 10 (extremely satisfied). In plaats van de variabele als een ordinaal meetniveau te bekijken, kan ze door het grote aantal antwoordcategorieën en de gelijke afstand tussen de categorieën beschouwd worden als een metrische variabele.

- (2) De tweede opmerking die we maken gaat in de andere richting. In plaats van het aannemen van continuïteit onder de antwoorden van een categoriale variabele (van categoriaal naar continu), maken we van een continue een categoriale variabele. Een interval of ratio-variabele kan omgevormd worden via een hercodering of een berekening tot een variabele met een beperkt aantal categorieën. In Volume 2 van de reeks Stap in Statistiek en Onderzoek gingen we reeds in op de werkwijze om continue tot categoriale variabelen te hercoderen. Een klassiek voorbeeld is de omzetting van de metrische variabele leeftijd (*age*) naar leeftijdsklassen. Bij deze werkwijze moet je wel in je achterhoofd houden dat je gegevens weggooit: je gaat immers meer gedetailleerde informatie reduceren tot een minder gedetailleerd analyseniveau.
- (3) Interval en ratiovariabelen vormen samen de continue variabelen. Voor de meeste technieken is het verschil tussen beide verwaarloosbaar. In dit deel worden ze samengenomen tot één groep. Dit impliceert dat we zullen werken met drie niveaus: nominaal, ordinaal en interval-ratio.

2.2.2 Criterium2: de gewenste informatie

Voor de bestudering van univariate verdelingen en bivariate verbanden zetten we een aantal begrippen en kengetallen kort op een rij. Bij de univariate verdelingen bespreken we frequenties, centrummaten, maten van spreiding en maten van vorm. Maten van samenhang maken deel uit van bivariate analyses.

4. We mengen ons niet in de discussie en gaan uit van een gangbare methode. Ook nu is er geen enkele dogmatische regel die de menselijke logica of het gezond onderzoeksverstand kan vervangen.

Een *verdeling* bestaat uit de waarden die een variabele kan aannemen. Er is een verschil tussen de mogelijke en reëel verkregen waarden. Stel dat we een vraag in een survey stellen over het opleidingsniveau met 4 categorieën (lager onderwijs, lager secundair onderwijs, hoger secundair onderwijs, hoger onderwijs). Als een variabele maar één waarde aanneemt, dan is er geen diversiteit in waarden en dan zeggen we dat een variabele niet varieert (anders gezegd: het is een constante). In het geval je de vragenlijst voorlegt aan universiteitsstudenten, dan vullen ze bijna allemaal ‘hoger secundair onderwijs’ in. Er is dus amper variatie. Laten we de vraag invullen door een willekeurige groep mensen, dan zal er veel meer variatie zijn (veel meer verschillende opleidingsniveaus worden aangeduid). Een telling (of turving) van het aantal keer dat waarden voorkomen, levert een *frequentieverdeling* op. Voor elke waarde krijgen we het aantal waarnemingen. In het ESS bestand zitten bijvoorbeeld 20048 mannen en 22256 vrouwen.

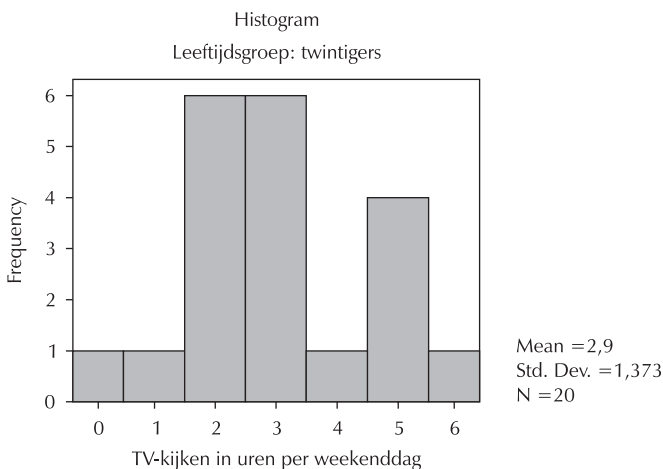
Elke verdeling heeft ook een bepaald *bereik*. Een verdeling bevat verschillende waarden die ergens starten en ergens eindigen. De afstand tussen de laagste en de hoogste waarde noemen we het bereik (de range in het Engels) van een verdeling. Bij nominale variabelen kunnen we niet van een bereik spreken omdat daar geen sprake is van een laagste en een hoogste waarde. Verwant met het bereik is het feit dat waarden van een verdeling kunnen *variëren* over het bereik. Dat wil zeggen dat respondenten verschillende antwoorden kunnen geven op een bepaalde vraag en dat de antwoorden anders kunnen zijn binnen het bereik van een verdeling. Niet iedereen antwoordt immers hetzelfde. Bij geslacht kunnen de antwoorden variëren tussen “man” en “vrouw”. Bij een houdingsschaal kunnen antwoorden bijvoorbeeld variëren van “helemaal niet eens” tot “helemaal eens”. In de statistiek is men voornamelijk geïnteresseerd in de wijze waarop antwoorden variëren. Het eindresultaat van variërende waarden is immers een verdeling van de antwoorden van alle respondenten en we willen als onderzoeker graag begrijpen hoe die verdeling er uit ziet, hoe ze tot stand is gekomen en welke andere kenmerken van invloed zijn op het tot stand komen van die variatie in antwoorden.

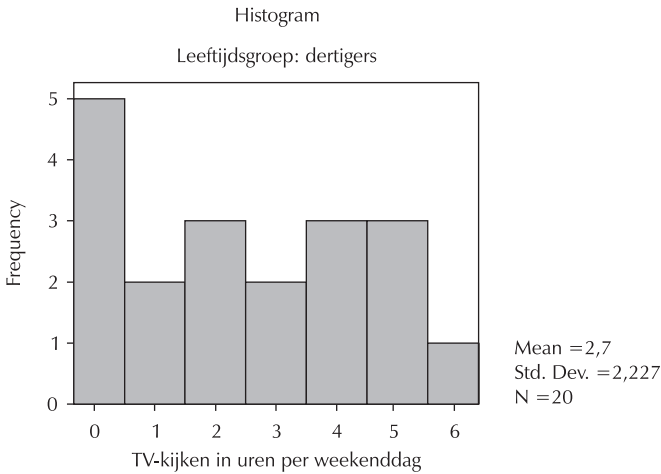
Verdelingen kunnen beschreven worden aan de hand van *kengetallen*. Dit zijn numerieke maatstaven die een samenvatting geven van een verdeling. Onvermijdelijk gaat hierdoor informatie verloren. Een van de belangrijkste doelen van statistiek, en van sociale wetenschappen in het algemeen, is echter eenvoud en systematiek aanbrengen in waarnemingen. Vandaar dat verdelingen regelmatig met hun kengetallen samengevat worden.

Een eerste groep univariate kengetallen zijn de *centrum- of positiematen*. Maten zoals de mediaan, modus, het rekenkundige gemiddelde en kwantielen geven aan rond welk getal de waarden variëren. Er is een minimum en een maximumgrens van het bereik en ergens tussenin ligt een waarde waar-

rond de andere waarnemingen opgetekend kunnen worden. De centrummaat wil een samenvattende maat zijn voor de volledige verdeling, door haar ‘middenpositie’. Een zogenaamd midden is niet voldoende om een samenvatting te geven van de verdeling. Naast de centrummaat op zich is het belangrijk om te weten hoe de waarden rond de centrummaat variëren. Liggen ze er dichtbij? Zijn ze erg verspreid over het bereik? *Spreidingsmaten* geven aan hoeveel diversiteit er in het voorkomen van de waarden is. Om een extreem voorbeeld te geven: als er maar één waarde wordt gevonden voor de variabele dan is de spreiding gelijk aan 0. In het geval vele waarden wel een keer worden aangetroffen hebben we een grotere diversiteit en dus een grotere spreiding. Tenslotte kan ook aan de hand van *maten van vorm* de verdeling bestudeerd worden. Hiermee bekijken we hoe symmetrisch een verdeling is (zijn er veel hoge waarden, veel lage waarden, een gelijke verdeling van hoge en lage waarden, ...). Bij de bespreking van de univariate statistieken voor continue variabelen zullen we dit verder toelichten (zie 3.4).

Een belangrijk gegeven waar we bij willen stilstaan is dat de uitkomsten van de kengetallen niet uniek zijn voor verdelingen. Het is goed mogelijk om bijvoorbeeld twee verdelingen met dezelfde maat van centraliteit te vinden, maar die een volstrekt ander verloop kennen. De maten dienen daarom altijd in combinatie met elkaar bekeken te worden. Een kort voorbeeld zal dit illustreren. Neem even aan dat we gegevens hebben over de tijdsbesteding aan tv-kijken, gerapporteerd door twintigers en dertigers. We willen nagaan of beide groepen verschillen in hun kijkgedrag in het weekend. Beide hebben het aantal uren per weekenddag opgegeven dat ze tv kijken. Bij de twintigers zien we dat ze vooral aantallen opgeven van 2 en 3 uur. De antwoorden van de dertigers daarentegen liggen verder uit elkaar. In een grafiek geeft dit de volgende verdelingen:





Beide grafieken vertonen een verschillende indeling. Bij de berekening van het gemiddelde als een maat voor centraliteit (zie 3.4) komt echter naar voor dat beide groepen quasi hetzelfde ‘gemiddeld aantal uren’ tv kijkt. De twintigers hebben een gemiddelde van 2,9 en de dertigers van 2,7. Als we ons louter zouden baseren op dit kengetal, kunnen we amper verschillen aantonen tussen de leeftijdsgroepen in hun kijkgedrag. Nochtans is de variatie in waarden erg afwijkend tussen de verdelingen. Bij de dertigers komen veel verschillende waarden voor. We zeggen dat er een grotere spreiding is bij de dertigers dan bij de twintigers, omdat de antwoorden minder geconcentreerd liggen rond enkele waarden. Het is dus belangrijk om diverse maten naast elkaar te leggen en vooral om ook een frequentietabel of grafiek op te vragen.

De *samenhang* (of associatie) tussen kenmerken omschrijven we als het al dan niet samen variëren van variabelen. Als de waarden van een variabele wijzigen, dan zullen de waarden van de andere variabele ook op een systematische manier wijzigen. Stel dat we ervan uitgaan dat er samenhang is tussen de leeftijd van respondenten en de mate waarin men zichzelf als religieus beschouwt (van helemaal niet religieus tot heel sterk religieus). Een verandering in leeftijd gaat gepaard met een verandering in religiositeit. We weten uit onderzoek dat ouderen sterker religieus zijn. De relatie tussen de twee variabelen kan omschreven worden als: een toename in leeftijd gaat gepaard met een toename in religiositeit.

2.2.3 Criterium3: de gewenste presentatievorm

We kunnen de weergave van verdelingen en verbanden opsplitsen in twee groepen: getallen en figuren. Onder getallen verstaan we frequenties en samenvattende maten die in getalvorm worden neergeschreven. Figuren zijn grafische voorstellingen van de verdeling.

Er is een verschil tussen de weergave die je zelf wenst om meer kennis te krijgen over de variabelen en verbanden en wat je toont aan het publiek. Als voorbereidende of verkennende analyse zal je diverse weergaven zoals een frequentietabel, een samenvattende maat en een grafiek laten genereren. Om inzicht te krijgen in de verdeling van variabelen vraag je dus best verschillende statistieken op. Wil je alles in een rapport opnemen, dan wordt dit veel te lang en onoverzichtelijk. Niet alles wat je opvraagt, is nuttig om weer te geven. Er dient steeds een selectie te gebeuren van kengetallen of grafische vormen.

Resultaten van een onderzoek begrijpelijk weergeven is essentieel om aansluiting te zoeken bij het doelpubliek. Kan het doelpubliek een grafiek interpreteren? Is een tabel meer dan een opeenstapeling van getallen voor hen? In wetenschappelijke tijdschriftartikels wordt vaak geopteerd voor een tabellenweergave. Soms prefereert men een grafische weergave, maar dit zien we vooral opduiken bij bivariate verbanden (oa. scatterplot). In een tabel is alle informatie gedetailleerd in de kolommen en rijen weergegeven. Grafieken zijn visueel aantrekkelijker, maar daardoor gaat ook informatie verloren die in een tabel wel getoond kan worden. Als we over te veel getallen beschikken, dan is een grafiek soms overzichtelijker. Maar zeker niet altijd. ESS is bijvoorbeeld zo'n groot bestand dat de grafieken te veel respondenten moeten weergeven, waardoor visuele inspectie moeilijk wordt. Bij de bespreking van de grafische vormen zal dit geregeld aangetoond worden. Populaire tijdschriften zullen doorgaans enkel mooi ogende en eenvoudige grafieken afdrukken. De keuze voor een tabel of grafiek is dus sterk afhankelijk van het doelpubliek en de gedetailleerdheid waarmee men de resultaten wil tonen. Beide methodes van weergave komen in het volume aan bod, we prefereren evenwel meestal de tabelvorm.

2.2.4 De keuze van techniek volgens drie criteria: een overzicht

De juiste techniek identificeren voor je onderzoeksprobleem ligt aan de basis van de analyse. Het is evident dat de keuze van een techniek cruciaal is. Uit een groot aanbod van mogelijke technieken een selectie maken is voor sommigen niet eenvoudig. Een structurering in tabelvorm van de technieken vol-

gens meetniveau van de variabelen maakt (hopelijk) de bomen door het bos zichtbaar. Niet alle mogelijke technieken voor de meetniveaus kenden hun weg naar de tabel. Onvermijdelijk vielen een aantal technieken met gelijkaardige doelen uit de boot. In andere statistische werken kunnen deze geraadpleegd worden.

2.2.4.1 Univariate technieken

De eerste dimensie waarop de technieken onderscheiden worden in de kolommen is het meetniveau van de variabelen. Elke rij bevat per meetniveau één of meerdere methodes. Telkens is een weergave in tabelvorm en een grafische voorstelling (cursief) in de cel gezet.

	Meetniveau		
	Nominaal	Ordinaal	Interval-Ratio
Frequentieverdeling	Frequentietabel <i>staafdiagram, cirkeldiagram</i>	Frequentietabel, cumulatieve frequentietabel <i>staafdiagram, histogram, lijndiagram</i>	Frequentietabel, cumulatieve frequentietabel <i>histogram, lijndiagram</i>
Centrum- en positiemaat	modus	modus, mediaan, kwantielen	modus, mediaan, kwantielen, rekendkundig gemiddelde <i>boxplot</i>
Spreidingsmaat	/	Variatiebreedte, interkwartielafstand	Variatiebreedte, interkwartielafstand Variantie, standaardafwijking <i>boxplot</i>
Vormmaat	/	/	Skewness, Kurtosis

2.2.4.2 Bivariate technieken

Het overzicht voor de bivariate technieken is anders opgevat. Omdat we met twee variabelen werken, hebben we te maken met een combinatie van twee meetniveaus. Ook de diversiteit aan maten van samenhang wordt immers ingegeven door het meetniveau van de variabelen. In elke cel staan de technieken die we in het volume bespreken. De grafische presentatievormen staan cursief.

Meetniveau var1	Meetniveau var2		
		Nominaal	Ordinaal
Nominaal	Chi ² , Phi, Cramer's V <i>geclusterd staafdiagram, gestapeld staafdiagram</i>	Chi ² , Phi, Cramer's V <i>geclusterd staafdiagram, gestapeld staafdiagram</i>	Anova <i>Boxpot voor meerdere variabelen</i>
Ordinaal	Chi ² , Phi, Cramer's V <i>geclusterd staafdiagram, gestapeld staafdiagram</i>	Chi ² , Kendall's Tau-b <i>geclusterd staafdiagram, gestapeld staafdiagram</i>	Anova <i>Boxpot voor meerdere variabelen</i>
Interval-Ratio	Anova <i>Boxpot voor meerdere variabelen</i>	Anova <i>Boxpot voor meerdere variabelen</i>	Correlatie <i>scatterplot</i>

2.3 Steekproef, populatie en representativiteit

Data verzamelen is een hele klus die start bij de afbakening van de onderzoeksgroep. De mensen waarover je gegevens wil verzamelen dienen op voorhand duidelijk te zijn bepaald. Gaat het om de hele bevolking? Of willen we over een deelgroep een onderzoek doen? De verzameling van waarnemingen die we willen bestuderen, noemen we de *populatie*. De populatie is met andere woorden de groep mensen, waarover we in termen van het onderzoeksprobleem iets willen weten. Dit zijn niet automatisch de *respondenten* in het onderzoek. Respondenten noemen we iedereen die bevestigd wordt. Vaak is het onmogelijk om de hele populatie te bevroegen. Zo zou het onbegonnen werk zijn om alle vrouwen in de Belgische bevolking op te nemen in een onderzoek naar hun arbeidspositie. Nochtans willen we wel iets kunnen zeggen over Belgische vrouwen (= de populatie). De oplossing voor dit probleem is het nemen van een *steekproef*. Een steekproef is een deelverzameling uit de populatie.

Diverse manieren zijn ontwikkeld om steekproefeenheden te weerhouden uit de populatie. Enerzijds is er de enkelvoudig aselechte steekproef, waarbij elk element in de populatie evenveel kans heeft om opgenomen te worden in de steekproef. Enkelvoudig aselechte steekproeven zijn vaak 'ideale' steekproeven, vanwege het op toeval berustte karakter van selectie. Anderzijds, en daar is men vaker op aangewezen in de praktijk, heeft men een aantal steekproefontwerpen ontwikkeld die respondenten selecteren op een bepaald kenmerk dat ze hebben. Respondenten komen dus niet toevallig in de steekproef

terecht. Op basis van een (aantal) kenmerk(en) zoals leeftijdsgroep, geslacht of gemeente worden respondenten uit de populatie genomen en alleen zij belanden in de steekproef. Voor een overzicht van steekproeftheorie en -methoden (getrapte steekproef, gestratificeerde steekproef, ...), waarbij vaste stramien in de selectie van respondenten worden gebruikt, verwijzen we naar andere statistische werken.

Een belangrijke voorwaarde van de meeste steekproefmethoden is dat de steekproef *representatief* is voor de populatie. Representativiteit is in theorie een eenvoudig begrip, maar sticht in de praktijk meer dan eens verwarring. Een steekproef is perfect representatief op een kenmerk als het kenmerk *in gelijke mate* voorkomt in de steekproef als in de populatie.

We nemen een voorbeeld om dit te verduidelijken. Als we uitspraken willen doen over tewerkgestelde vrouwen ten opzicht van niet-tewerkgestelde vrouwen, dan zou de steekproef in proportie evenveel van beiden moeten bevatten als de populatie. Wanneer 55% van de vrouwen in de populatie tewerkgesteld is (en bijgevolg 45% niet) dan zal dit percentage zich moeten weerspiegelen in de steekproef die uit deze populatie wordt getrokken. In een goede steekproef moet dan ook 55% van de werkende vrouwen geantwoord hebben. Deze redenering kunnen we verderzetten voor andere kenmerken (variabelen) uit die steekproef. Er kan bijvoorbeeld verondersteld worden dat naast de tewerkstelling ook de verhouding vrouwen met en zonder kinderen moet overeenkomen met de verdeling in de populatie of dat de leeftjidsverdeling overeen moet stemmen. Net in dat streven naar een “perfecte” steekproef komen de grootste problemen bovendrijven. We kunnen niet zomaar alle kenmerken tot in het oneindige vergelijken tussen de steekproef en de populatie. Indien we alle kenmerken van de populatie zouden bezitten, dan hoefden we geen onderzoek meer te doen. Vandaar dat het vergelijken van een steekproef met de populatie steeds beperkt is tot de kenmerken waarover we op populatieniveau gegevens hebben (vanuit volkstellingen of vanuit administratieve gegevens). In de sociale wetenschappen stellen we ons tevreden met een steekproef waarvan we de representativiteit op een aantal cruciale kenmerken kunnen nagaan (vaak zijn dat geslacht, leeftijd, burgerlijke staat).

2.4 Significantie en hypothesetoetsing

De vraag die daar onmiddellijk op volgt is de vraag naar de reikwijdte van de conclusies. Als we slechts een deel van de populatie bevragen over een ken-

merk waarvoor we op populatieniveau geen informatie hebben, kunnen we dan wel een uitspraak doen over de hele populatie? Is het mogelijk om een groep Belgische vrouwen te bevragen over aspecten van hun arbeidssituatie en daaruit conclusies te trekken voor *alle* Belgische vrouwen? Een antwoord vinden we in de *inferentiële of inductieve statistiek*. Deze tak van de statistiek heeft als doel resultaten te formuleren voor de volledige populatie op basis van de resultaten uit de steekproef. Veralgemening van steekproefresultaten staat hierbij centraal. We gebruiken het resultaat van de steekproef om een *schatting* te maken voor de populatie. Het spreekt voor zich dat de inferentiële statistiek geen absolute waarheid verkondigt, maar een kans berekent om voorspellingen af te leiden voor de populatie. Het is best mogelijk dat er fouten of onvolkomenheden in de voorspelling zijn geslopen. We spreken dan over de *foutenmarge* van de inferentie. Met andere woorden, inferentiële technieken trachten conclusies te trekken op niveau van de populatie op basis van verzamelde steekproefgegevens, mits een zekere foutenmarge. We vertrekken dus van een *steekproefkenmerk* om iets te kunnen vertellen over een onbekend kenmerk uit de populatie (dat is dan de *populatieparameter*).⁵ Daarbij gaan we er van uit dat we over de waarde van die populatieparameter nooit 100% zekerheid hebben. Meestal trachten we wel met een zekerheid van 95% of meer te werken (zie verder).

Er zijn twee prominente manieren om aan inferentiële statistiek te doen: schattingen (puntschattingen en intervalschattingen) en hypothesetoetsen. Bij het *schatten* van een parameter of een interval wordt getracht om op basis van de verkregen steekproefgegevens een raming of gissing te maken voor de populatieparameter (punt) of voor een bereik waarbinnen de populatieparameter zal liggen (interval). Men neemt het steekproefkenmerk als uitgangspunt, omdat dit de beste informatie is waarover men beschikt. *Puntschattingen* worden door middel van vaste ‘schattingsformules’ berekend. *Intervalschattingen* geven niet één getal, maar een begin- en eindwaarde waartussen de populatieparameter zich met een bepaalde ‘waarschijnlijkheid’ zal bevinden. Dit noemen we betrouwbaarheidsintervallen. Men streeft naar een zo hoog mogelijke kans dat de populatieparameter geschat kan worden binnen een bepaald interval. Stel dat we uit een aselechte steekproef van 500 volwassen Belgische mannen (+ 18j) weten dat de gemiddelde lengte 1,75m is. Het doel van de inferentie met een intervalschatting zal zijn om dit steekproefgemiddelde te nemen als schatting van het populatiegemiddelde. Het steekproefgemiddelde kan best in het midden van het interval geplaatst worden en het is mogelijk om aan de hand van bepaalde procedures na te gaan wat een boven- en een ondergrens is voor het interval: $1,75 \pm \text{interval}$.

28 | 5. Een kenmerk in de populatie krijgt de benaming parameter.



Zoals gezegd, werken de procedures steeds met een kansberekening. Zo kunnen we veronderstellen dat we er 95% zeker van willen zijn, dat de populatieparameter in het interval ligt. Als voorbeeld stellen we dat er met 95% kans aangenomen wordt dat de gemiddelde lengte van mannen (+18j) ligt in het interval van 1,70m tot 1,80m. Een klein interval, waarbij de boven- en ondergrens dicht bij elkaar liggen, is teken van een goede schatting van de parameter. Voor meer informatie over punt- en intervallschattingen, met alle mogelijke procedures, verwijzen we naar andere statistische werken.

We verdiepen ons voor dit volume alleen in de logica van **hypothesetoetsen**.⁶ Een hypothesetoets zal, zoals het woord zegt, nagaan of een idee dat je hebt over bepaalde uitkomsten klopt of niet. Het is een werkwijze die toelaat om een veronderstelling over een parameter te aanvaarden (ze klopt) of te verwerpen (ze klopt niet). We maken met andere woorden assumpties of gokken over de populatie, die we willen testen. Daarbij moeten steeds twee vragen beantwoord worden.

1. Wat is de inhoud van de hypothese? Wat veronderstellen we?
2. Hoe gaan we na of die hypothese klopt? Anders gezegd: hoe verwerpen of aanvaarden we de hypothese? Welke statistische test gebruiken we hiervoor?

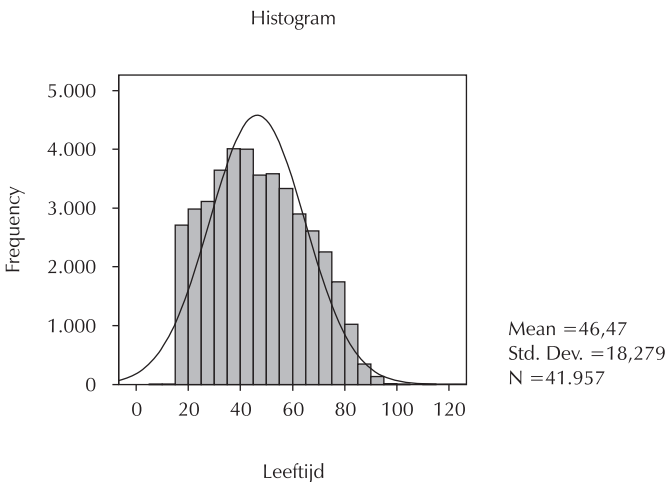
In feite zit de statistiek vol met hypothesetoetsen. Je kan bijvoorbeeld hypothesen formuleren voor statistische testen die op samenhang duiden (is er samenhang of niet?), over het vergelijken van verdelingen (lijkt de verdeling van een variabele op een gekende verdeling?) of over statistische modellen als regressies (heeft een variabele een invloed op een andere variabele?). Later, bij de behandeling van bivariate technieken, zullen we ingaan op de toepassing van hypothesetoetsen bij samenhang. Omwille van de eenvoud, bespreken we in de volgende paragraaf enkel de grote lijnen van de hypothesetoets, zodanig dat het resultaat van de toets zinvol kan worden geïnterpreteerd. Eens je de logica van een hypothese en de statistische toets van de hypothese begrijpt, kan je deze toepassen op alle mogelijke vormen van hypothesetoetsing.

6. De wiskundige achtergrond van hypothesetoetsen laten we achterwege, alsook een aantal begrippen en nuanceringen (oa. kansrekenen, functies van curves en vrijheidsgraden). Hierdoor zal de lezer met een grondige voorkennis van hypothesetoetsen dit deel mogelijk ontoereikend vinden.

2.4.1 Stap 1: Het formuleren van een hypothese

Waarom gebruiken we in de statistiek hypothesen? Net als in het dagelijkse leven willen we nagaan of bepaalde veronderstellingen kloppen of niet. Daarvoor vertrekken we van een vraag die we stellen. Om te bepalen of we volgend jaar op vakantie zullen gaan naar de Spaanse Costa, baseren sommigen zich bijvoorbeeld op weersvoorspellingen. De vraag luidt dan: zal het warm worden volgend jaar in juni aan de Spaanse kust? Is er een grote kans dat het warm wordt, dan boeken we een reis. Dat de weersvoorspellers al eens de bal misslaan, weten we allemaal. Men probeert echter om zo accuraat mogelijk te zijn in de voorspellingen. De kans dat het weer correct is voorspeld, moet zo hoog mogelijk zijn. Die redenering passen we toe op data die we hebben verzameld.

Een eerste stap is het formuleren van de hypothese. Wat willen we nagaan in de statistiek? Er zijn erg veel mogelijkheden. Zo kunnen we veronderstellen dat de gemiddelde lengte van mannen in de populatie gelijk is aan de verkregen steekproefwaarde (1,75m). Een tweede voorbeeld betreft het vergelijken van de gemiddelde lengte van mannen en vrouwen. Zijn mannen en vrouwen gemiddeld ongeveer even groot? Als derde voorbeeld zouden we kunnen veronderstellen dat de lengte van mannen verdeeld is volgens een gekende curve, namelijk de Gauss-curve of normaalverdeling (zie verder voor uitleg over de curve). Op onderstaande grafiek zien we de verdeling van de leeftijden en een klokvormige curve erover lopen. De vraag die we stellen is: lijkt het leeftijds patroon op de curve? In dit geval zijn er behoorlijk wat afwijkingen te zien op de grafische voorstelling.



Omdat we geen antwoord op hypothesen zoeken ‘met het blote oog’, doen we beroep op bepaalde statistische testen. We noemen dit de **toetsstatistieken**. We komen er straks nog op terug. Om de testen te kunnen uitvoeren, zal de hypothese volgens een bepaald stramen geformuleerd moeten worden. Hoe de hypothese wordt neergeschreven is bijgevolg geen vrije keuze, die zal bepaalde regels volgen. We keren even terug naar het voorbeeld van de weersvoorspelling. We kunnen de hypothese formuleren dat ‘het gemiddeld warmer dan 25°C zal worden in juni aan de Spaanse kust’. Maar ze kan ook in de andere richting geformuleerd worden: het zal gemiddeld niet warmer dan 25°C worden. Je kan met andere woorden assumpties in twee richtingen neerschrijven. Ook bij de statistische toepassing is dit het geval. We maken echter een vaste keuze in de statistiek voor een van beide en stellen alleen die hypothese voorop om ze te testen. De hypothese die we testen, noemen we in de statistiek de **nulhypothese (H_0)**. De andere (omgekeerde) hypothese wordt de **alternatieve hypothese (H_1)** genoemd.

De keuze in het voorbeeld valt op de tweede vorm: ‘de gemiddelde temperatuur (T) zal lager zijn dan 25°C ’.

$$H_0: T < 25^{\circ}\text{C}$$

$$H_1: T \geq 25^{\circ}\text{C}$$

Voor de voorbeelden in de statistiek wordt dit:

Nulhypothese	Alternatieve hypothese
H_0 : de gemiddelde lengte van mannen in de populatie = 1,75m	H_0 : de gemiddelde lengte van mannen in de populatie \neq 1,75m
H_0 : de gemiddelde lengte van mannen in de steekproef = de gemiddelde lengte van vrouwen in de steekproef (of het verschil tussen beide = 0)	H_1 : de gemiddelde lengte van mannen in de steekproef \neq de gemiddelde lengte van vrouwen in de steekproef (of het verschil tussen beide \neq 0)
H_0 : De leeftijdsverdeling van mannen in de steekproef = normaalverdeling	H_0 : De leeftijdsverdeling van mannen in de steekproef \neq normaalverdeling

Het zal je misschien al zijn opgevallen dat we de nulhypothese bij bovenstaande voorbeelden altijd op eenzelfde manier hebben neergeschreven. Namelijk in termen van gelijkheid. We hebben dit gedaan, omdat dit gebruikelijk is voor de meeste toetsstatistieken. De nulhypothese is deze van gelijkheid, de alternatieve hypothese van de ongelijkheid.

De voorbeelden stelen allemaal op ‘gelijkheid’ en ‘ongelijkheid’. Dit noemen we *tweezijdige testen*. Van het voorbeeld van de weersvoorspelling is af te leiden dat er ook hypothesen te formuleren zijn, die vertrekken vanuit een

‘groter dan’ of ‘kleiner dan’ principe. Deze worden aangeduid met de term *eenzijdige testen*. De nulhypothese is dan dat iets ‘kleiner dan’ is of ‘groter dan’ is. Bij de alternatieve hypothese worden de andere mogelijkheden neergeschreven: iets is groter dan of gelijk aan ... of iets is kleiner dan of gelijk aan Een voorbeeld is dat we willen weten of de gemiddelde lengte van mannen in de populatie kleiner is dan 1,75m:

Nulhypothese	Alternatieve hypothese
H_0 : de gemiddelde lengte van mannen in de populatie $< 1,75\text{m}$	H_0 : de gemiddelde lengte van mannen in de populatie $\geq 1,75\text{m}$

Tot dusver hebben we gesproken over de nulhypothese en de alternatieve hypothese. Deze worden altijd volgens vaste regels opgesteld. De *nulhypothese* gaat uit van een ‘gelijkheid’ of van een ‘groter dan’ of ‘kleiner dan’ veronderstelling. De *alternatieve hypothese* omvat alle andere gevallen, behalve wat in de nulhypothese is omschreven. Bovendien weten we ook dat alleen de nulhypothese getest wordt in de statistiek. Met deze informatie in het achterhoofd gaan we over tot de toetsstatistiek.

2.4.2 Stap 2: De toetsstatistiek⁷

De nulhypothese is de statistische uitspraak die we gaan testen. We gaan namelijk na of we de nulhypothese kunnen **aanvaarden** of **verwerpen**. Aanvaarden betekent dat we ze voor waar zullen aannemen, dat ze klopt. Als ze niet klopt, dan verwerpen we ze. Op de alternatieve hypothese wordt geen test uitgevoerd.⁸ De vraag die zich stelt, is dan: hoe ga je testen of zo’n nulhypothese verworpen moet worden of niet?

Het testen van een nulhypothese gebeurt met wat men een **toetsstatistiek** noemt. Omdat nulhypotheses niet lukraak mogen worden geformuleerd, hebben we vaak dezelfde ‘hypothesevormen’ (zoals het al dan niet gelijk zijn van gemiddelden). Het gevolg hiervan is dat we beroep kunnen doen op een reeks van toetsstatistieken die telkens geschikt zijn voor bepaalde hypothesevormen. Bij een toets gaat men een wiskundige bewerking uitvoeren, het resultaat ervan zal een **toetsingsgrootheid** zijn. Een toetsingsgrootheid is niets

7. Om dit stuk te kunnen begrijpen, moet de lezer al inzicht hebben in verdelingen en kenmerken van verdelingen. We leggen dit verder in dit boek uit.

8. We zijn wel geïnteresseerd in de alternatieve hypothese, maar dan alleen via de omweg van de nulhypothese en nooit rechtstreeks in een test. Als we de nulhypothese testen kunnen we uiteraard ook iets zeggen over de alternatieve hypothese. Stel dat de weerkaarten ons zeggen dat er een erg grote kans is dat het kouder dan 25°C zal worden, dan wordt de nulhypothese bevestigd en impliciet weten we ook dat de alternatieve hypothese niet uitkomt.

meer dan een getal (een uitkomst) van de berekening. Omdat er veel verschillende toetsstatistieken bestaan, gaan we niet in op de berekeningswijze. In dit volume is het belangrijk dat men begrijpt waarvoor een toetsstatistiek dient en wanneer die gebruikt wordt.

De toetsingsgrootte meet hoe goed de data bij de nulhypothese aansluit. Maar wat zegt die toetsingsgrootte ons? Een getal op zich zegt weinig. We hebben een standaard nodig om de uitkomst mee te vergelijken.

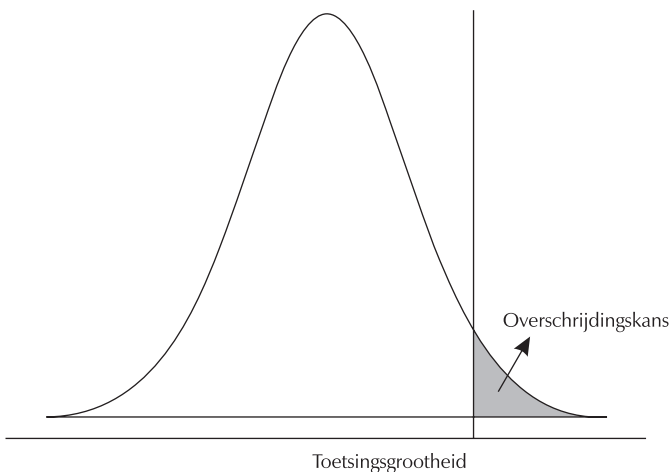
2.4.2.1 De theoretische verdeling als vergelijkingsbasis

We treffen een standaard om mee te vergelijken aan bij **theoretische verdelingen** van de toetsingsgrootte. De statistiek heeft namelijk beschikking over een waaier aan theoretische verdelingen. Die theoretische modellen geven aan welke verdeling van toetsingsgrootheden veel of weinig voorkomen, wanneer niet één maar een groot aantal steekproeven van dit type genomen worden. Je moet je die theoretische modellen voorstellen als een enorme hoeveelheid steekproeven, waarin men dezelfde berekening heeft uitgevoerd. Stel dus dat we niet één, maar alle mogelijke steekproeven trekken met een bepaalde N (= aantal eenheden). Binnen al deze steekproeven kan een steekproefkenmerk berekend worden (vb. het gemiddelde). Men kan met die gemiddelden opnieuw een verdeling aanmaken.⁹ Dit noemen we de **steekproevenverdeling**. Het blijft een theoretische verdeling, omdat ze bestaat uit alle mogelijke steekproeven die we in realiteit niet kunnen trekken.

We geven eerst een voorbeeld van een reële steekproevenverdeling, voor we het concept van theoretische verdelingen verder toelichten. Om te weten of het volgend jaar warm zal zijn in Spanje, kunnen we dit jaar een steekproef van alle dagen in juni nemen. Voor die steekproef van dagen verzamelen we informatie over de temperatuur op één plaats aan de Spaanse kust, telkens op hetzelfde uur. Als we elke dag op dezelfde plek de temperatuur noteren om 12u 's middags, dan hebben we 31 waarnemingen. Van die temperaturen kunnen we het gemiddelde berekenen. Dan hebben we één gemiddelde om ons oordeel te vellen of we een reis boeken of niet. Nog beter zou zijn om de gegevens van de vorige jaren te verzamelen en die gemiddelden allemaal te bekijken. We kunnen ze in een grafiek plaatsen en zien of er grote schommelingen zijn geweest voor de maand juni de afgelopen jaren. Op basis van de verdeling van de gemiddelde temperaturen van vorige jaren zouden we een voorspelling kunnen maken.

9. De steekproevenverdeling heeft net zoals andere verdelingen een gemiddelde en een standaardafwijking. Men noemt de standaardafwijking van de steekproevenverdeling de 'standaardfout'.

Dit principe wordt ook toegepast bij theoretische verdelingen. We maken een vergelijking met een theoretische verdeling, om dan een beslissing te kunnen nemen over de nulhypothese op basis van de toetsingsgrootheid. Er zijn vele mogelijke theoretische verdelingen. Die theoretische modellen worden in de statistiek grafisch voorgesteld door curven. Je hebt bijvoorbeeld een normaalverdeling, een binomiaalverdeling, een Poisson-verdeling, een F-verdeling, een chi-kwadraatverdeling, enzovoort. We spreken over het maken van een ‘vergelijking’, maar hebben nog niet verteld hoe dit wordt gedaan. Elke verdeling heeft een oppervlakte onder de curve.¹⁰ In de eerste fase van de significantietoets hebben we een toetsingsgrootheid berekend. Dan kunnen we nagaan op de curve welke kans er is dat een toetsingsgrootheid dezelfde waarde zou aannemen die even extreem of nog extremer is dan de waargenomen toetsingsgrootheid. De curve geeft met andere woorden de kans aan dat we de waargenomen toetsingsgrootheid volgens de theoretische verdeling kunnen aantreffen, onder de aanname dat de nulhypothese klopt. Die kans noemen we de **overschrijdingskans (p)**.¹¹ Wetende dat de toetsingsgrootheid berekend is op basis van de nulhypothese, kan je stellen dat de kans die je krijgt door je toetsingsgrootheid te vergelijken met het bijpassend theoretische model, eigenlijk dus een kans is dat je nulhypothese juist is. Krijg je een hoge kans, dan heb je veel kans dat de nulhypothese juist is en dus niet verworpen mag worden. Krijg je een kleine kans, dan is er een sterke aanwijzing dat ze verworpen mag worden. Onderstaande grafiek toont een theoretische curve, met een overschrijdingskans (grijze zone onder de curve).



10. Meer uitleg over functies van curves en de oppervlakte onder een curve kan je aantreffen in wiskundige of statistische werken met een wiskundige basis.

11. In dit volume spreken we vaak over de ‘significantiewaarde’ als we het over de overschrijdingskans hebben. Verwar dit niet met de significantiedrempel of het significantieniveau (zie verder).

Statistische programma's berekenen de toetsingsgrootheid en maken zelf een vergelijking met de theoretische verdeling. Werk je niet met een statistisch programma, dan kan je beroep doen op tabellen. Voor de theoretische verdelingen, die omschreven kunnen worden aan de hand van wiskundige functies, hebben wiskundigen bepaalde waarden uit de curves in tabellen gezet. Zo'n tabel beslaat vaak een of twee pagina's. Er staan dus behoorlijk veel waarden in, men noemt ze vaak 'kritische waarden'. Wiskundigen hebben met andere woorden een hele reeks waarden in tabellen geplaatst, zodat de eigen toetsingsgrootheid vergeleken kan worden met de waarden. Statistische programma's hebben echter het zoekwerk in deze tabellen overgenomen. We gaan er dan ook niet verder op in.

2.4.2.2 Hoe de nulhypothese aanvaarden of verwerpen?

De toetsingsgrootheid en het passende theoretische model worden gebruikt om na te gaan of we onze nulhypothese moeten verwerpen of niet. De toetsingsgrootheid is zoals gezegd een uitrekenen van een toetsstatistiek voor een nulhypothese. We krijgen dus een getal dat we kunnen vergelijken met dat theoretische model. In de theoretische verdeling zoeken we op wat de kans is dat de toetsingsgrootheid gelijk of extremer is dan de zelf berekende toetsingsgrootheid in de steekproef. Eenvoudig gesteld: het is de kans dat de nulhypothese verworpen mag worden of niet. Een hoge kans betekent een hoge kans om de nulhypothese niet te verwerpen en dus te aanvaarden. We nemen aan dat ze klopt als er een hoge kans is. Een lage kans geeft aan dat ze verworpen mag worden en dat ze dus niet klopt.

Blijft dan nog de vraag: wat is hoog en wat is laag? Op dat ogenblik komen we op het domein van de **significantie**. Significant wil zeggen "veelbetekend". Significant is dus *niet* "juist" of "bewezen" maar iets dat gegronde conclusies toelaat. Het zegt ons of de kans om de nulhypothese te verwerpen een zelf gestelde standaard heeft bereikt. In de statistiek gelden algemene afspraken over wanneer een kans zo klein is, dat de nulhypothese verworpen mag worden. Het aanvaarden of verwerpen van de nulhypothese hangt bijgevolg af van een eenheid (de afspraak) die we het **significantieniveau** (α) noemen. Een significantieniveau is een drempel die we zelf kiezen om uit te drukken hoe streng we willen zijn in het verwerpen of aanvaarden van de nulhypothese. De theoretische verdeling levert een overschrijdingskans op. Nu is het aan ons om te zeggen dat we die overschrijdingskans voldoende laag of hoog vinden of niet met het significantieniveau (zie 2.4.2.3).

Het significantieniveau kan uitgedrukt worden in een percentage (bv. 5% kans) maar wordt doorgaans in een proportie tussen 0 en 1 uitgedrukt.¹² De norm in de sociale wetenschappen is dat het significantieniveau wordt vastgelegd op minstens 0,05. Strengere standaarden zijn die van 0,01 en 0,001. Bij het rapporteren over significantietoetsen wordt vaak een compactere schrijfwijze gehanteerd. Wanneer de overschrijdskans kleiner is dan 0.05 wordt bijvoorbeeld één sterretje geplaatst. Bij een kans kleiner dan één honderdste, zet met twee sterretjes. Samengevat zijn dit de drie meest gebruikte significantieniveaus:

$$\begin{aligned} p < 0,05 & \rightarrow * \\ p < 0,01 & \rightarrow ** \\ p < 0,001 & \rightarrow *** \end{aligned}$$

In de vaktaal spreekt men van een significantie of betrouwbaarheid tot op het 0,05, het 0,01 en het 0,001 niveau. In een onderzoek dien je een beslissing te nemen welke van de drie grenzen je zal hanteren. Bij de keuze moet je echter heel goed opletten op de grootte van je steekproef, vermits de kans in de theoretische verdeling afhankelijk is van de steekproefgrootte. Grote steekproeven van bijvoorbeeld meer dan duizend observaties mogen niet te laks zijn. Je krijgt dan immers snel toetsingsgrootheden die binnen de 0.05-grens vallen. Je zal dan nulhypotheseën gaan verwerpen die eigenlijk niet fout zijn maar enkel fout lijken omwille van de grootte van de steekproef. Als je dan een grens van 0.05 hanteert, wordt ‘alles significant’ en kan je ‘alles aantonen’. Zulke onderzoeken zullen dus de strengste grens (één duizendste) moeten hanteren. Het kiezen van de significantiegrens volgt dus een eenvoudige regel: hoe groter de steekproef, hoe strenger de significantiegrens komt te liggen. In een bestand zoals ESS is het aangewezen om de 0,001-grens te hanteren.

2.4.2.3 Interpretatie van de overschrijdskans en het significantieniveau

Vooraleer we overgaan tot de interpretatie van de overschrijdskans aan de hand van het significantieniveau, geven we een bijkomende verduidelijking over de aard van het significantieniveau. Tot nu toe hebben we het significantieniveau beschreven als de gekozen grens om een beslissing te nemen over het al dan niet verwerpen van de nulhypothese. Vermits we met kansen werken, betekent het dat je *nooit* met 100% zekerheid een beslissing kan nemen. Er is altijd een (soms heel kleine) kans dat je verkeerd bent in de verwerping of de aanvaarding. We kunnen daarom de significantiegrens ook

36 | 12. Een aantal voorbeelden: 1% = 0,01; 10% = 0,1; 100% = 1.

uitdrukken in termen van mogelijke foute interpretaties die we maken over de nulhypothese. Het significantieniveau kan dan begrepen worden als een arbitraire grens die wordt getrokken om te bepalen hoeveel kans er is dat de nulhypothese correct wordt geëvalueerd. Er zijn een aantal fouten die we kunnen maken bij het verwerpen of aanvaarden van de nulhypothese. Zo kunnen we ze bijvoorbeeld aanvaarden, terwijl we ze eigenlijk moeten verwerpen. In het schema geven we vier types weer.

		Wat doe jij als onderzoeker?	
		Je aanvaardt de nulhypothese	Je verwerpt de nulhypothese
Wat is de aard van de “echte, doch ongekende” realiteit?	De nulhypothese is echt juist	JUISTE BESLISSING De nulhypothese was juist en dat neem jij ook aan als beslissing	FOUTE BESLISSING Je hebt een juiste nulhypothese onterecht verworpen. Dat noemt met een TYPE I fout in de statistiek.
	De nulhypothese is echt fout	FOUTE BESLISSING Je hebt een foute nulhypothese onterecht aanvaard. Dat noemt met een TYPE II fout in de statistiek.	JUISTE BESLISSING De nulhypothese was fout en jij hebt ze verworpen

Twee mogelijke beslissingen zijn juist (vak linksboven en rechtsonder). De nulhypothese moet respectievelijk verworpen/aanvaard worden en je doet dit. In twee andere situaties kan je echter een foute beslissing nemen. Ten eerste bestaat de fout van de onterechte verwerping (vak rechtsboven). De nulhypothese wordt verworpen, maar eigenlijk is ze correct. Dit noemt men in de statistiek een **type I fout**. Het tweede type fout doet zich voor als de nulhypothese wordt aanvaard, terwijl ze verworpen moet worden (**type II fout**, linksonder). Bekeken vanuit het uitgangspunt van verwerping is een type I fout de zwaarste fout. De nulhypothese wordt foutief verworpen, wat een onomkeerbare beslissing is. Het niet-verwerpen is minder definitief, we kunnen het eigenlijk niet met zekerheid stellen. Daarom richt de statistiek zich doorgaans op de kans om een type I fout te maken. Het significantieniveau is nu net de afspraak over het maken van zo'n Type I fout.

Of we de nulhypothese willen aanvaarden of verwerpen door het overschrijdingsniveau te vergelijken met de significantiedrempel, is afhankelijk van de gekozen toetstatistiek en de onderzoeksvraag. Bij de formulering van de nulhypothese volgen we de standaard dat er gelijkheid is (of groter dan of klei-

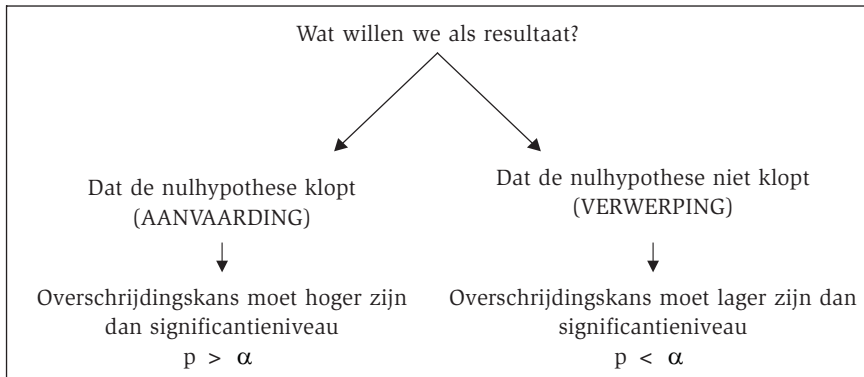
ner dan). De interpretatie van de aanvaarding of verwerping en terugkoppeling naar wat we wensen kan echter soms problematisch zijn. De termen aanvaarding en verwerping hebben namelijk in dagelijkse taal een positieve en negatieve connotatie, alsof we tevreden zijn met een aanvaarding en ontevreden met een verwerping. Niets is minder waar in de statistiek! We grijpen terug naar het eenvoudige voorbeeld van de weersverwachting voor de Spaanse kust, om uit te leggen wat we hiermee bedoelen. Het startpunt is dat je graag wil dat het warm zal zijn. De nulhypothese is echter in de andere richting geformuleerd.

$$H_0: T < 25^\circ\text{C}$$

$$H_1: T \geq 25^\circ\text{C}$$

Zal het kwik gemiddeld onder 25°C blijven of niet? Dit is wat getest wordt met de nulhypothese. Stel dat de hypothese wordt aanvaard met een lage kans op type I fout: de gemiddelde temperatuur zal waarschijnlijk onder 25°C blijven. Dit is het tegenovergestelde wat we wilden en toch spreken we van ‘aanvaarding’.

In een schema¹³ geven we weer wat dit betekent voor de interpretatie van de overschrijdingskansen aan de hand van het significantieniveau.



In de statistiek willen we bovendien meestal dat de nulhypothese verworpen wordt. Het interpreteren van de toetsstatistieken en overschrijdingskansen gaat meestal in die richting. De nulhypothese is geformuleerd in termen van gelijkheid (of groter dan of kleiner dan), maar onze hoop is dat dit niet klopt.

13. Met dit schema willen we zeker geen vrijgeleide geven om alleen overschrijdingskansen in output te interpreteren. Het overzicht dient enkel om tot een beter begrip van het aanvaarden en verwerpen van nulhypoteses te komen.

2.4.3 De hypotheselogica toegepast op een voorbeeld

Om een en ander nog eens samen te vatten, geven we een uitgewerkt voorbeeld in SPSS. We bekijken de vraag of de gemiddelde leeftijd van Belgische vrouwen die gehuwd zijn in ESS, gelijk is aan de gemiddelde leeftijd van gehuwde vrouwen in de Belgische bevolking. Het Nationaal Instituut voor de Statistiek in België publiceerde op zijn website dat de gemiddelde leeftijd van de Belgische getrouwde vrouwen in 2002 gelijk was aan 30,6 jaar. In de ESS vragen we de gemiddelde leeftijd van de Belgische gehuwde vrouwen op met behulp van volgende syntax¹⁴:

```
USE ALL.
COMPUTE filter = (cntry = "BE" and gndr = 2 and marital = 1).
FILTER BY filter.
EXE.

COMPUTE age = inwyr - yrbrn.
VARIABLE LABELS age 'leeftijd'.
EXECUTE.

DESCRIPTIVES
  VARIABLES = age
  /STATISTICS = MEAN SEMEAN.

FILTER OFF.
```

We krijgen volgende tabel:

Descriptive Statistics

	N	Mean	
	Statistic	Statistic	Std. Error
leeftijd	451	48,78	,661
Valid N (listwise)	451		

De gemiddelde leeftijd in België voor gehuwde vrouwen ligt dus op 30.6 jaar, terwijl die in de ESS 48.78 blijkt te zijn. We willen nu weten of deze waarde significant afwijkt van dat populatiegemiddelde. Samengevat:

14. De standaardfout van het gemiddelde (SEMEAN) vragen we in deze syntax op omdat we die verder in onze uitleg nog nodig zullen hebben.

Populatie	Gehuwde vrouwen in de Belgische bevolking
Steekproef	Gehuwde vrouwen in de ESS-steekproef
Populatieparameter	Gemiddelde leeftijd in de populatie
Steekproefkenmerk	Gemiddelde leeftijd in de ESS-steekproef
Onderzoeksvraag	Verschilt het gemiddelde dat we in de steekproef vinden significant van het gemiddelde in de populatie?

De onderzoeksvraag stelt dat we willen weten of dat steekproefgemiddelde *verschilt* van het populatiegemiddelde. We stellen een nulhypothese van gelijkheid op:

Nulhypothese	de verwachte populatieparameter voor het gemiddelde is gelijk aan de reële waarde die we reeds kennen OF H_0 : verwachte populatieparameter = populatiewaarde
Alternatieve hypothese	de verwachte populatieparameter voor het gemiddelde is niet gelijk aan de reële waarde die we reeds kennen OF H_1 : verwachte populatieparameter \neq populatiewaarde

De volgende stap is het bepalen van een statistische toets die er in slaagt om een steekproefgemiddelde te vergelijken met een populatiegemiddelde. Die blijkt te bestaan: de t-toets (in het Engels: one-sample t-test). De berekening van de toetsingsgrootheid voor deze test is betrekkelijk eenvoudig:

$$\begin{aligned} & \text{Toetsingsgrootheid t-test} \\ &= \frac{\text{gemiddelde steekproef} - \text{gemiddelde populatie}}{\text{Stan daarfout gemiddelde steekproef}} \end{aligned}$$

Of als we dat uitrekenen:

$$\text{Toetsingsgrootheid t-test} = (48,78 - 30,6)/0,661 = \mathbf{27,5}$$

De waarde 27,5 is de toetsingsgrootheid van de t-toets. We vragen om deze toets voor ons uit te voeren.

```

USE ALL.
COMPUTE filter=(cntry="BE" and gndr=2 and marital=1).
FILTER BY filter.
EXE.

T-TEST
  /TESTVAL = 30.6
  /VARIABLES = age.

FILTER OFF.
  
```

De t-toets wordt opgegeven, met inbegrip van de testwaarde (TESTVAL) voor het gemiddelde. Als resultaat geeft SPSS:

One-Sample Test

	Test Value = 30.6					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence interval of the Difference	
					Lower	Upper
leeftijd	27,521	450	,000	18,178	16,88	19,48

We zien in de tabel de door ons berekende waarde 27,5 in de kolom t. Deze kolom geeft de toetsingsgrootte t. Belangrijk is de vierde kolom “Sig. (2-tailed)”. In statistische output vinden we vaak de exacte overschrijdskans onder een kolomhoofding “Signif.” of “Prob.”. Daar vinden we de kans terug dat de nulhypothese in de populatie op zal gaan. Die kans wordt uitgedrukt in een proportie. Dat wil zeggen een kans tussen 0 en 1. Het getal 0.000 wil zeggen dat de kans dat onze nulhypothese juist is, gelijk is aan 0.000.

Dan komen we bij het significantieniveau. Zoals gezegd kunnen we kiezen tussen drie significantieniveaus: een kans kleiner dan 0.05, een kans kleiner dan 0.01 en een kans kleiner dan 0.001. De ESS is een grote steekproef waardoor we streng moeten zijn over ons significantieniveau. We kiezen dus één duizendste (kleiner dan 0.001). We vergelijken de waarde die SPSS geeft als kans nu met die grens:

Overschrijdingswaarde	Significantiegrens
0.000	0.001

Het is duidelijk dat de overschrijdskans die SPSS geeft *kleiner* is dan de grens die we vooropgesteld hadden. Om die reden verwerpen we dus onze nulhypothese (beide gemiddelde zijn gelijk aan elkaar) en aanvaarden we de alternatieve hypothese en stellen we dat *de gemiddelde leeftijd van de Belgische gehuwde vrouwen in de ESS significant verschilt van de gemiddelde leeftijd in de populatie Belgische gehuwde vrouwen.*

2.4.4 Vertaling naar de praktijk van significantietoetsen

We zetten de verschillende stappen nog even schematisch op een rij. Hierbij verlaten we even de theorie en gaan we kijken naar de manier waarop significantietoetsen in de praktijk uitgevoerd worden. Het veelvuldig gebruik van significantietoetsen heeft immers geleid tot een heleboel standaardtoetsen die in statistische software is voorzien. Bij elk van deze toetsen hoort automatisch een bepaalde vorm van nulhypothese.

STAPPEN	Uitgewerkt in het voorbeeld
1. Stel je onderzoeksvraag.	<i>Ik wil weten of de gemiddelde leeftijd van Belgische gehuwde vrouwen in de ESS verschilt van de gemiddelde leeftijd van gehuwde vrouwen in België.</i>
2. Formuleer de nulhypothese en bepaal op basis van deze vraag de toets die je moet uitvoeren.	<i>Ik heb een toets nodig die een populatiegemiddelde en een steekproefgemiddelde vergelijkt. Dat gebeurt met behulp van een t-toets. De nulhypothese bij die t-toets is dat het verschil tussen beide gemiddelden gelijk is aan nul (dat ze dus gelijk zijn aan elkaar).</i>
3. Voer de toets uit (vraag de toetsingsgrootte en significantiewaarde op in een statistisch programma).	<i>T-TEST /TESTVAL = 30.6 /VARIABLES = age.</i>
4. Vergelijk de significantie van de output met één van de drie significantieniveaus.	<i>De overschrijdingskans is 0.000 wat wil zeggen dat de toets significant is op het 0.001-niveau.</i>
5. Aanvaard of verwerp de nulhypothese en trek conclusies.	<i>We verwerpen de nulhypothese dat beide gemiddelden gelijk zijn aan elkaar en komen tot de conclusie dat de gemiddelde leeftijd van gehuwde vrouwen in de ESS significant verschillend is van de gemiddelde leeftijd in de populatie.</i>