



# Basisboek Statistiek met SPSS

Handleiding voor het verwerken en analyseren van  
en rapporteren over (onderzoeks)gegevens



Noordhoff Uitgevers

Ben Baarda en Cor van Dijkum

6<sup>e</sup> druk



# Basisboek Statistiek met SPSS

**Ben Baarda**

**Cor van Dijkum**

---

Zesde druk

Noordhoff Uitgevers Groningen/Utrecht

Ontwerp omslag: G2K (Groningen-Amsterdam)

Omslagillustratie: Shutterstock - 231007846

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

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



0 / 19

© 2019 Baarda & Van Dijkum, vertegenwoordigd door Noordhoff Uitgevers bv  
Groningen/Utrecht, The Netherlands.

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

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

ISBN (ebook) 978-90-01-89581-5

ISBN 978-90-01-89580-8

NUR 916

# Woord vooraf bij de zesde druk

In het *Basisboek Statistiek met SPSS* vind je aanwijzingen voor het verwerken en analyseren van onderzoeksgegevens met behulp van SPSS. Oorspronkelijk staat SPSS voor 'Statistical Products for the Social Sciences'. Sinds SPSS in 2009 een onderdeel van IBM Business Analytics is geworden, kent het een veel bredere toepassing en wordt het niet alleen door sociale wetenschappers, maar ook door andere onderzoekers gebruikt zoals economen, marktonderzoekers, medische wetenschappers en technici. In dit boek maken we je al doende vertrouwd met het SPSS-pakket, maar ook met de statistiek. We nodigen je uit om aan de hand van een databestand over de relatie tussen geld, geluk en gezondheid aan de slag te gaan.

Bij het schrijven van dit boek hebben wij gebruikgemaakt van SPSS versie 25.0.

Voor de zesde druk is het gefingeerde voorbeelddatabestand waarmee in het boek wordt gewerkt, uitgebreid met een variabele over gezondheid en zijn de gegevens aangepast. Ook zijn de schermafdrucken afkomstig van de op dit moment laatste versie van SPSS, te weten versie 25.0. De behandelde onderwerpen zijn hetzelfde als in de vorige druk.

Het *Basisboek Statistiek met SPSS* komt optimaal tot zijn recht in samenhang met het *Basisboek Methoden en Technieken of Dit is onderzoek!*, waarin wij een praktische handleiding geven voor het opzetten en uitvoeren van onderzoek. Aanvullend kan ook het *Statistiekkwartet* gebruikt worden, waarin de belangrijkste statistische begrippen en toetsen op speelse wijze worden uitgelegd.

Het *Basisboek Statistiek met SPSS* is uiteraard ook los daarvan goed te gebruiken.

Ben Baarda  
Cor van Dijkum  
Najaar 2018

Helaas moeten wij vanwege het overlijden van Martijn de Goede zijn belangrijke didactische inbreng missen.



# Inhoud

Effectief studeren 8

## 1 Hoe bereid ik mij voor op statistiek met SPSS? 11

- 1.1 Inleiding 12
- 1.2 Geld, geluk en gezondheid; toelichting op de casus 13
- 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer! 14
- 1.4 Enkele algemene statistische begrippen 16
- 1.5 Hoe werkt SPSS onder Windows? 19  
Samenvatting 23

## 2 Hoe breng ik mijn gegevens in? 25

- 2.1 Hoe maak ik de opzet voor het SPSS-bestand met Variable View? 26
- 2.2 Hoe voer ik mijn gegevens in? 32
- 2.3 Hoe bewaar ik mijn ingevoerde gegevens? 35
- 2.4 Hoe controleer ik of ik fouten heb gemaakt bij de invoer van mijn gegevens? 36
- 2.5 Hoe roep ik de ingevoerde gegevens weer op? 38  
Samenvatting 40

## 3 Hoe verander en combineer ik gegevens? 43

- 3.1 Hoe hercodeer ik gegevens? 44
- 3.2 Hoe bewaar ik gehercodeerde gegevens? 46
- 3.3 Hoe kan ik een variabele indelen in klassen? 46
- 3.4 Hoe kan ik gegevens combineren? 48
- 3.5 Hoe kan ik bij het combineren van gegevens rekening houden met ontbrekende waarden? 49  
Samenvatting 53

## 4 Hoe analyseer ik mijn gegevens bij een frequentie-onderzoeksvraag? 55

- 4.1 Wat is een frequentieverdeling? 56
  - 4.2 Wanneer gebruik ik een frequentieverdeling? 58
  - 4.3 Hoe maak ik een frequentieverdeling met SPSS? 59
  - 4.4 Hoe geef ik frequenties grafisch weer? 60
  - 4.5 Hoe moet ik de uitvoer van de opdracht Frequencies lezen? 63
  - 4.6 Hoe kan ik een frequentieverdeling presenteren? 64
  - 4.7 Hoe moet ik een frequentieverdeling interpreteren en erover rapporteren? 64
  - 4.8 Hoe bereken ik frequenties voor subgroepen? 65
  - 4.9 Hoe vergelijk ik (sub)groepen? 66
  - 4.10 Wat doe ik met multiple responsen? 70
  - 4.11 Hoe maak ik variabelen vergelijkbaar? 72
- [Samenvatting 74](#)

## 5 Hoe analyseer ik mijn gegevens bij een verschil-onderzoeksvraag? 77

- 5.1 Wat is een verschilvraag? 78
  - 5.2 Verschilvraag bij een nominale test- en splitsingsvariabele. Wat is een kruistabel en chikwadraat? 79
  - 5.3 Verschilvraag bij ordinale testvariabelen en nominale splitsingsvariabelen. Wat zijn niet-parametrische toetsen? 86
  - 5.4 Verschilvraag bij interval/ratio testvariabele en nominale splitsingsvariabele; t-toets of variantieanalyse? 98
- [Samenvatting 112](#)

## 6 Hoe analyseer ik mijn gegevens bij een samenhang-onderzoeksvraag? 115

- 6.1 Wat is een samenhangonderzoeksvraag? 116
  - 6.2 Samenhang tussen twee nominale variabelen: Cramér's V 118
  - 6.3 Samenhang tussen twee ordinale variabelen: Spearman's rangcorrelatie! 121
  - 6.4 Samenhang tussen twee interval- of ratiovariabelen: Pearson's productmoment-correlatiecoëfficiënt! 125
  - 6.5 Samenhang tussen twee interval- of ratiovariabelen: enkelvoudige regressieanalyse 130
  - 6.6 Samenhang bij twee of meer variabelen op interval- of rationiveau: multiple correlatie- en regressieanalyse 134
  - 6.7 Samenhang tussen twee interval- of ratiovariabelen, gecontroleerd voor het effect van een derde variabele: partiële correlatie 138
- [Samenvatting 141](#)



- Bijlage I Hoe controleer ik de homogeniteit van samengestelde scores? [142](#)
- Bijlage II De samenhang tussen meerdere nominale variabelen [148](#)
- Bijlage III De samenhang van meerdere onafhankelijke variabelen met één afhankelijke variabele [154](#)

Illustratieverantwoording [159](#)

Register [161](#)

Over de auteurs [163](#)

# Effectief studeren

In dit *Basisboek Statistiek met SPSS* willen we je leren al doende vertrouwd te raken met het pakket. Na een algemene inleiding en kennismaking met de basisregels van SPSS kun je direct beginnen met het invoeren van gegevens in de computer en het analyseren daarvan.

We nodigen de lezer uit om aan de hand van een databestand over de relatie tussen geld, geluk en gezondheid aan de slag te gaan. Dit databestand omvat gefingeerde vragenlijstgegevens van 500 mannen en 490 vrouwen. Daarbij volgen we de gebruikelijke gang van zaken in een onderzoek:

- We beginnen bij het invoeren van de gegevens (hoofdstuk 2).
- Vervolgens leggen we uit hoe je gegevens kunt veranderen en combineren met SPSS (hoofdstuk 3).

Na dit voorbereidende werk volgt de analyse. Welke analysetechniek gebruikt moet worden, is onder meer afhankelijk van je specifieke onderzoeksvraag. Je vraagstelling bevat altijd een of meer specifieke onderzoeksvragen waarop je via onderzoek een antwoord probeert te vinden.

We maken daarbij een onderscheid tussen:

- frequentie-onderzoeksvragen (hoe vaak of in welke mate komt iets voor; hoofdstuk 4)
- verschil-onderzoeksvragen (vaststellen en toetsen van verschillen op één kenmerk tussen twee of meer groepen; hoofdstuk 5)
- samenhang-onderzoeksvragen (vaststellen en toetsen van een samenhang tussen twee kenmerken; hoofdstuk 6)

In de bijlage bespreken wij hoe je kunt nagaan of het verantwoord is om gegevens te combineren en bespreken we meer geavanceerde analysetechnieken als de loglineaire analyse en de univariate analyse waarbij de invloed van combinaties van variabelen wordt geanalyseerd.

Om verantwoord met SPSS te kunnen werken is niet alleen kennis van SPSS, maar ook statistische kennis nodig. In het eerste hoofdstuk bespreken we basisbegrippen als meetniveau, normaalverdeling, kans, significantie, een- en tweezijdige toetsing, power en effectgrootte. In de hoofdstukken 4, 5 en 6 wordt uitgelegd dat de keuze voor een statistische techniek niet alleen afhankelijk is van het type onderzoeksvraag maar ook van het meetniveau van de betreffende kenmerken, en of het om een populatie of om een steekproef gaat.

In de praktijk komt het nogal eens voor dat een respondent bij een vraag meer dan één antwoordmogelijkheid aangeeft of aanvinkt. Bij open vragen maak je achteraf antwoordcategorieën op basis van de 'open' antwoorden. Meestal geven respondenten in die gevallen uitgebreide antwoorden, waardoor ook meer antwoordcategorieën van toepassing zijn. De multipleresponsanalyse maakt het mogelijk om deze verschillende antwoorden geïnte-

greerd te analyseren en presenteren, in plaats van analyse en presentatie van elke afzonderlijke vraag of variabele (paragraaf 4.10).

De stof wordt verder gepresenteerd aan de hand van de volgende vragen die je je als gebruiker stelt:

- Welke statistische analysetechniek moet ik gezien mijn onderzoeksvraag gebruiken?
- Wat houdt die techniek in, wat zijn de voorwaarden en hoe voer ik de techniek in SPSS uit?
- Als het gelukt is de gekozen analysetechniek met SPSS uit te voeren, hoe lees ik dan de computer-output? Wat staat er precies?
- Hoe moet ik de resultaten interpreteren? Wat betekenen de resultaten voor mijn onderzoeksvraag?
- Hoe presenteer ik de resultaten in mijn onderzoeksrapport?
- Hoe schrijf ik mijn conclusie(s) op?

Voor elke besproken techniek laten we een voorbeeld van de SPSS-uitvoer zien. Aan de hand van dit voorbeeld leggen we uit hoe je het moet lezen en wat het betekent. Om dit laatste te illustreren, laten we telkens zien hoe je over het resultaat kunt rapporteren.

Er is veel aandacht besteed aan de didactische opbouw van het boek. Elk hoofdstuk kent de volgende indeling:

- *voorkennis*: de kennis die nodig is om het betreffende hoofdstuk te lezen
- *vragen*: de vragen die in het hoofdstuk beantwoord worden
- *begrippen*: de belangrijkste begrippen die in het hoofdstuk behandeld worden
- *inhoudelijke informatie*: de specifieke stof voor het in het hoofdstuk behandelde onderwerp
- *samenvatting*

#### *Het didactische concept is veranderd*

De vorige versies van dit boek kenden veel opsommingen en samenvattingen. Deze zijn grotendeels vervangen door schermafdrucken, waarin instructies zijn opgenomen. Telkens worden de schermafdrucken getoond die de student te zien krijgt als hij de opdrachten uitvoert. In de schermafdrucken hebben wij informatie gezet, waardoor de student weet waar hij welke informatie moet plaatsen en op welke knoppen hij moet klikken.

Bij het schrijven van dit boek hebben wij gebruikgemaakt van SPSS versie 25.0. Dit programma is op de computers van de meeste hogescholen en universiteiten geïnstalleerd. Je kunt het als student met korting bestellen op [www.surfspot.nl](http://www.surfspot.nl).

De databestanden waarop we de analyses uitvoeren, zijn te vinden op internet via <http://basisboekstatistiekmetsspss.noordhoff.nl>



# For I don't care too much for money, money can't buy me love

— John Lennon & Paul McCartney

## **Voorkennis**

Er wordt geen specifieke voorkennis vereist. Wel wordt aanbevolen om naast dit boek het *Basisboek Methoden en Technieken* (6e herziene druk, 2017) of *Dit is onderzoek!* (3e druk, 2019) en het *Statistiekkwartet* (1e druk, 2010) te gebruiken.

## 1

# Hoe bereid ik mij voor op statistiek met SPSS?

In dit hoofdstuk worden de volgende vragen en begrippen behandeld:

- Welke onderwerpen worden in dit boek behandeld (1.1)?
- Waarover gaat de casus 'Geld, geluk en gezondheid' (1.2)?
- Voor de analyse is het van belang om te weten (1.3):
  - Gaat het in de onderzoeksvraag om een frequentie, verschil, of om een samenhang (1.3.1)?
  - Wat is het meetniveau (1.3.2)?
  - Gaat het om een steekproef of een populatie (1.3.3)?
- Wat houden de statistische begrippen normaalverdeling, standaardfout, betrouwbaarheid, zekerheid, significantie, een- of tweezijdige toetsing, relevantie, effectgrootte en vrijheidsgraden in (1.4)?
- Hoe werkt SPSS onder Windows (1.5)?

---

Frequentie 14

Verskil 14

Samenhang 14

Nominaal meetniveau 14

Ordinaal meetniveau 14

Intervalmeetniveau 15

Ratiomeetniveau 15

Nulpunt 15

Continue variabele 15

Discrete variabele 15

Beschrijvende statistiek 16

Inductieve/inferentiële statistiek 16

Normaalverdeling 17

Standaardfout 17

Significantie 18

Een- of tweezijdig toetsen 18

Effectgrootte 18

Vrijheidsgraden 18

Data View-scherf 20

Variabelenscherf 21

## 1.1 Inleiding

Als je onderzoek doet, is het kiezen van de juiste statistische techniek om de verzamelde gegevens te analyseren een belangrijke schakel in de lange keten van beslissingen die je moet nemen. Het uiteindelijke doel is het beantwoorden van de onderzoeksvraag of -vragen.

Om de plaats van de data-analyse in de onderzoekscyclus als geheel duidelijk te maken, zetten we de gebruikelijke fasen van een onderzoek hierna op een rij. Elke fase van de onderzoekscyclus is in de vorm van een vraag opgenomen:

- 1 Wat is de onderzoeksvraag of -vragen en wat is de doelstelling van mijn onderzoek?
- 2 Hoe zoek ik informatie (onder meer literatuurstudie)?
- 3 Wat voor type onderzoek ga ik doen?
- 4 Hoe ziet mijn onderzoeksontwerp eruit?
- 5 Betrek ik de populatie in mijn onderzoek, of trek ik een steekproef?
- 6 Welke dataverzamelmethode ga ik gebruiken?
- 7 Hoe prepareer ik mijn data voor de analyse?
- 8 Hoe analyseer ik mijn data?
- 9 Hoe rapporteer en evalueer ik mijn onderzoek?

In dit boek staan de fasen 7 en 8 en een deel van fase 9 centraal: preparatie, analyse en beschrijving van de onderzoeksgegevens die met SPSS geanalyseerd worden. Aan de hand van een onderzoek naar de relatie tussen geld, geluk en gezondheid (zie figuur 1.1), behandelen we stap voor stap:

- Hoe prepareer je onderzoeksgegevens om deze te kunnen invoeren (hoofdstuk 2)?
- Hoe kun je gegevens met SPSS aanpassen en veranderen? Voordat je aan de analyse begint moet je vaak eerst van bepaalde gegevens de waarden ompolen (hercoderen), of de waarden van gegevens combineren tot een nieuwe score (hoofdstuk 3) en vervolgens controleren op hun betrouwbaarheid (paragraaf 3.4 en bijlage I).
- Hoe kies je de juiste analysetechniek? Om de juiste analysetechniek te kiezen stel je eerst vast wat het karakter is van je onderzoeksvraag (subparagraaf 1.3.1). Gaat het om frequenties (hoofdstuk 4), verschillen (hoofdstuk 5) of om samenhangen (hoofdstuk 6). Vervolgens ga je na wat het meetniveau van je gegevens is (subparagraaf 1.3.2). Tot slot stel je vast of het in je onderzoek om een steekproef of een populatie gaat (subparagraaf 1.3.3). Je kunt dan aan de hand van het blokschema 'Hoe analyseer ik mijn data?' vaststellen welke analysetechniek het best past bij jouw onderzoeksvraag. Je vindt het schema op de binnenkant van de kaft in dit boek.
- Hoe voer je de analyse met SPSS uit en hoe interpreteer je de resultaten? Dit geven we telkens aan bij elke statistische techniek die we behandelen. We bespreken hoe je verslag kunt doen van de resultaten.

In dit hoofdstuk bespreken we verder een aantal belangrijke statistische begrippen zoals normaalverdeling, significantie, kans en standaardfout (paragraaf 1.4). In de laatste paragraaf van dit hoofdstuk (paragraaf 1.5) leggen we uit hoe je SPSS kunt starten.

## 1.2 Geld, geluk en gezondheid; toelichting op de casus

Voordat we ingaan op het gebruik van SPSS, introduceren we eerst het voorbeeldonderzoek 'Geld, geluk en gezondheid' (zie figuur 1.1). De vragen uit figuur 1.1 zijn fictief voorgelegd aan een voor de Nederlandse volwassen bevolking representatieve steekproef van 500 Nederlandse mannen en 490 Nederlandse vrouwen van 25 tot en met 60 jaar. Er is bewust voor deze leeftijdsgrenzen gekozen. Veel jonge mensen studeren nog en hebben daardoor geen vast inkomen. Bij mensen ouder dan 60 jaar is er vaak al sprake van (gedeeltelijke) uittreding uit het arbeidsproces waardoor zij in een andere financiële situatie verkeren. De data van dit onderzoek vind je op de website <http://basisboekstatistiekmetspss.noordhoff.nl> onder de naam 'data20181'.

FIGUUR 1.1 Voorbeeldonderzoek 'Geld, geluk en gezondheid'

### Maakt geld gelukkig?

Een onderzoeker wil weten of er een verband bestaat tussen geld, geluk en gezondheid. Hij vraagt zich af of geld gelukkig maakt en ook of het een positief effect op de gezondheid heeft. Zijn centrale onderzoeksvraag luidt dan ook: 'Is er een positief verband tussen de mate waarin iemand over geld beschikt en de mate waarin hij/zij zich gelukkig en gezond voelt?' Het begrip 'geld' heeft de onderzoeker als volgt gedefinieerd: geld is de hoeveelheid financiële middelen waar iemand over kan beschikken. Deze definitie is met opzet ruim gekozen. Hierdoor worden niet alleen het inkomen en het vermogen, maar ook andere financiële bronnen waar iemand over kan beschikken in het onderzoek betrokken. Geluk wordt als volgt gedefinieerd: geluk is de mate waarin iemand tevreden is met het leven dat hij/zij leidt. Verder is gevraagd hoe gezond men zich voelt. De onderzoeker heeft deze drie begrippen gemeten met een vragenlijst. Zowel voor het meten van het begrip geld, als voor het begrip geluk heeft hij vijf uitspraken of items gemaakt, voor de gezondheid is één vraag gesteld.

Geld is gemeten met de items:

- 1 Ik ben in het bezit van een auto *ja/nee*
- 2 Ik heb een koopwoning/flat *ja/nee*
- 3 Ik bezit een tablet *ja/nee*
- 4 Ik krijg zorgtoeslag *ja/nee*
- 5 Ik krijg huurtoeslag *ja/nee*

Geluk is gemeten met de volgende vragen:

- 1 Als ik mijn leven over zou mogen doen zou ik het ... op dezelfde manier doen.  
 *absoluut niet*    *niet*    *ten dele*    *wel*    *absoluut wel*
- 2 De meeste anderen hebben het beter dan ik.  
 *absoluut niet*    *niet*    *ten dele*    *wel*    *absoluut wel*
- 3 Ik heb het ... naar mijn zin.  
 *absoluut niet*    *niet*    *ten dele*    *wel*    *absoluut wel*
- 4 Het leven is zwaar.  
 *absoluut niet*    *niet*    *ten dele*    *wel*    *absoluut wel*

5 Ik voel mij eenzaam.

- absoluut niet*    *ten dele*    *wel*    *absoluut wel*

Gezondheid is gemeten met de vraag:  
Hoe is uw gezondheid?

- zeer goed*    *goed*    *voldoende*    *slecht*    *zeer slecht*

Omdat geluk en gezondheid niet alleen van financiële middelen afhangen, maar ook van andere zaken, vraagt de onderzoeker tevens naar een aantal gemakkelijk te meten kenmerken zoals geslacht, leeftijd, opleidingsniveau en leefsituatie van de personen die hij enquêteert.

Geslacht  *man*    *vrouw*

Leeftijd in jaren ...

Leefsituatie  *alleen*

*met partner*

*met partner en kinderen*

Hoogste afgeronde opleiding

*lager (beroeps)onderwijs (lagere school, lager technisch onderwijs, lager voortgezet onderwijs)*

*middelbaar (beroeps)onderwijs (middelbaar technisch onderwijs, middelbaar voortgezet onderwijs, mavo, mulo, vmbo, enz.)*

*hoger (beroeps)onderwijs (universiteit, hbo, vwo, havo)*

## 1.3 Hoe analyseer ik mijn data? Een gebruikswijzer!

Voor de keuze van een statistische analysetechniek zijn de antwoorden op de volgende vragen van belang:

- 1 Gaat het in de vraagstelling om frequenties (hoe vaak/in welke mate), om een verschil of om een samenhang? Of gaat het om een combinatie daarvan?
- 2 Wat is het meetniveau (nominaal, ordinaal of interval-/rationiveau) van de gegevens die je hebt verzameld?
- 3 Gaat het om een steekproef of om een populatie?

Het blokschema 'Hoe analyseer ik mijn data' (te vinden aan de binnenzijde van de kaft van dit boek) is ontworpen aan de hand van deze vragen. In de volgende subparagrafen gaan we daar nader op in.

### 1.3.1 Om wat voor specifieke onderzoeksvragen gaat het in mijn onderzoek?

Bij het beantwoorden van de vraag welke statistische techniek je kunt gebruiken, vormt de *onderzoeksvraag of onderzoeksvragen het uitgangspunt*. Het onderzoek omvat altijd één of meer specifieke onderzoeksvragen waarop een antwoord moet worden gegeven. Globaal zijn er drie soorten onderzoeksvragen te onderscheiden:

- vragen waarbij het erom gaat hoe vaak of in welke mate (**frequentie**) iets voorkomt; bijvoorbeeld: 'In welke mate zijn Nederlanders gelukkig?' of 'Hoeveel procent van de Nederlanders is in het bezit van een auto?'
- vragen waarbij het gaat om een **verschil**; bijvoorbeeld: 'Zijn mannen gezonder dan vrouwen?'
- vragen waarbij het gaat om een **samenhang**; bijvoorbeeld: 'Is er een samenhang tussen geld en geluk?'

### 1.3.2 Wat is het meetniveau van mijn gegevens?

Als je hebt bepaald om wat voor type onderzoeksvragen het in je onderzoek gaat (zie de eerste kolom in het blokschema 'Hoe analyseer ik mijn data'), dan ga je vervolgens na op welk meetniveau de variabele(n) is/zijn gemeten. Zie daarvoor in het genoemde blokschema de cellen onder frequentie, verschil of samenhang. Per onderzoeksvraag geef je aan wat het meetniveau is van de betreffende variabelen. Bij de onderzoeksvraag of er verschil is tussen mannen en vrouwen in de mate waarin zij zich als gezond beoordelen, is het meetniveau van de betreffende variabelen namelijk anders (en lager) dan het meetniveau van de variabelen in bijvoorbeeld de onderzoeksvraag of er een samenhang is tussen geld en geluk. Bij de variabele geslacht zijn er maar twee categorieën of waarden, namelijk 'man' en 'vrouw'. Hierbij is alleen sprake van een verschil, maar niet van meer of minder. Een man is anders dan een vrouw, maar niet meer of minder. Hetzelfde geldt voor leefsituatie; iemand is gehuwd, samenwonend of alleenstaand. Bij dit type antwoordmogelijkheden gaat het om een *nominaal meetniveau*. Je kunt alleen zeggen hoeveel mannen of vrouwen een auto hebben, maar niet dat iemand meer 'mans' of meer 'vrouws' is. Je bent in deze test man of vrouw.

Dat kan wel bij gegevens die zijn gemeten op respectievelijk ordinaal, interval- of rationiveau. Bij gegevens op *ordinaal meetniveau* is er wél sprake van meer of minder, maar het verschil tussen de categorieën is niet in een getal uit te drukken. Bij opleidingsniveau bijvoorbeeld is er duidelijk sprake

Nominaal  
meetniveau

Ordinaal  
meetniveau



van meer en minder. De havo is hoger dan het vmbo, maar er is niet aan te geven hoeveel hoger. Dat geldt ook voor de medailleverdeling op een kampioenschap. Het is duidelijk dat bijvoorbeeld een 100-meter loper die goud wint sneller heeft gelopen dan een loper die zilver heeft gewonnen. Het feit dat hij goud heeft gewonnen geeft aan dat hij sneller heeft gelopen, maar niet hoeveel sneller.

Bij interval- en ratiomeetniveau is dat verschil tussen categorieën in termen van meer of minder wel in een getal uit te drukken. Temperatuur is daar een goed voorbeeld van. Het verschil tussen 5 en 10 graden Celsius is even groot als het verschil tussen 45 en 50 graden. Bij het *intervalmeetniveau* is er alleen geen sprake van een natuurlijk nulpunt, zoals wel het geval is bij gegevens die op *ratiomeetniveau* zijn gemeten, bijvoorbeeld gewicht en lengte. Nul graden Celsius is geen natuurlijk nulpunt. Het natuurlijke *nulpunt* voor de temperatuur is  $-273$  graden Celsius, wat overeenkomt met nul graden Kelvin. Wanneer je de temperatuur weergeeft in graden Kelvin is er wel sprake van een ratiometing, want hier is er een natuurlijk nulpunt.

Dit heeft gevolgen voor de rekenkundige bewerkingen die je mag uitvoeren. Bij temperatuur gemeten in graden Celsius mag je niet zeggen dat 20 graden tweemaal zo veel is als 10 graden. Bij een meting in graden Kelvin mag je wel zeggen dat 20 graden tweemaal zo veel is als 10 graden en bij een gewichtsmeting dat 20 kilo tweemaal zo zwaar is als 10 kilo. Overigens vat SPSS het interval- en ratiomeetniveau samen onder het begrip *Scale*. De beide andere meetniveaus worden *Nominal* en *Ordinal* genoemd.

We onderscheiden verder continue en discrete variabelen. Bij *continue variabelen* kun je je een lijn voorstellen waarop waarden een aaneengesloten rij punten vormen: een continuüm. Tussen twee punten liggen altijd nog (oneindig veel) andere mogelijke waarden. Voorbeelden van continue variabelen zijn de lengte van een persoon, leeftijd en intelligentie. Variabelen die alleen hele waarden kunnen aannemen, noemen we *discrete variabelen*, zoals het aantal auto's dat iemand bezit of het aantal kinderen in een gezin.

Interval-  
meetniveau

Ratiomeetniveau

Nulpunt

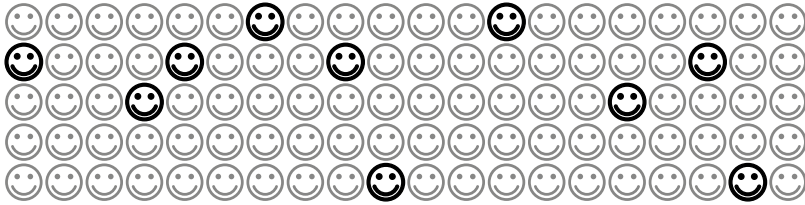
Continue  
variabelenDiscrete  
variabelen

**TABEL 1.1** Overzicht van meetniveaus, hun rekenkundige consequenties en voorbeelden

Meetniveau	Rekenkundige consequentie	Voorbeeld
Nominaal	Tellen, percentages (alleen onderscheid)	Geslacht
Ordinaal	Tellen, percentages en hoger/lager (onderscheid en ordening)	Opleidingsniveau
Interval	Tellen, hoger/lager, waarbij verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding (onderscheid en ordening)	Intelligentie
Ratio	Tellen, hoger/lager, waarbij verschillen in eenheden zijn uit te drukken, gemiddelde, spreiding en het berekenen van verhoudingen (onderscheid en ordening)	Leeftijd

### 1.3.3 Gaat het om een populatie of om een steekproef?

FIGUUR 1.2 Voorbeeld van een populatie en een aselechte steekproef



N=100 hele populatie van werknemers

n=10 aselechte getrokken steekproef uit populatie van werknemers

#### Beschrijvende statistiek

Er zijn twee vormen van statistiek: de beschrijvende en de inductieve of inferentiële statistiek. *Beschrijvende statistiek* gebruik je wanneer je onderzoek doet bij een populatie. Er is sprake van een populatie wanneer alle eenheden waarover je uitspraken wilt doen in je onderzoek worden betrokken. Dus wanneer je bijvoorbeeld alle werknemers van een bedrijf enquêteert om de mate van tevredenheid over het werk vast te stellen. Om kosten te besparen kun je ook een deel van de werknemers enquêteren (steekproef), die je aselechte uit het totale bestand van werknemers trekt (zie figuur 1.2). Bij een steekproef is de opzet dat je uitspraken doet over de totale populatie van werknemers. In dit geval zul je gebruik moeten maken van de *inductieve of inferentiële statistiek*: je wilt op grond van een speciaal geval (een steekproef) algemene uitspraken doen (over de populatie). Raadpleeg een methoden- of statistiekboek voor meer informatie. Ook op Wikipedia vind je uitgebreide informatie over het trekken van steekproeven [http://en.wikipedia.org/wiki/Sampling\\_\(statistics\)](http://en.wikipedia.org/wiki/Sampling_(statistics)).

#### Inductieve/inferentiële statistiek

Voordat je aan de analyse van je gegevens begint, stel je jezelf de vraag over welke eenheden (wie of wat) je uitspraken wilt doen. Wanneer dat alleen de personen of zaken zijn die in je onderzoek zijn betrokken, dan is er sprake van een populatieonderzoek. Wil je ook uitspraken doen over de personen of zaken die niet betrokken zijn bij je onderzoek, maar als het ware gerepresenteerd worden door de onderzoekseenheden die je geselecteerd hebt, dan gaat het om een steekproefonderzoek. In de volgende paragraaf behandelen we in het kort enkele statistische termen die je steeds weer tegenkomt bij het statistisch toetsen. Het gaat er dan om wat de kans is dat de resultaten die je in een steekproef vindt op toeval berusten, of met een bepaalde marge aan onzekerheid kunnen worden gegeneraliseerd naar de populatie waaruit de steekproef is getrokken.

## 1.4 Enkele algemene statistische begrippen

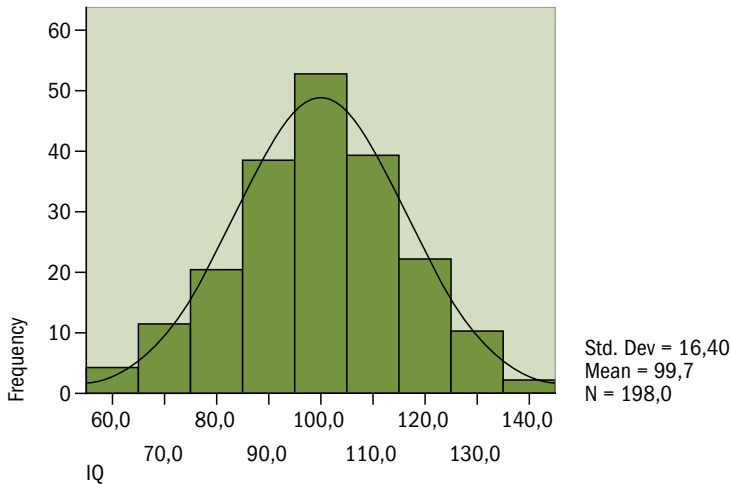
Zoals in de voorgaande paragraaf is aangegeven, zijn er twee vormen van statistiek: de beschrijvende en de inductieve, of inferentiële statistiek. Het doel van de beschrijvende statistiek is op een inzichtelijke en overzichtelijke manier je gegevens te presenteren. Als je van bijna duizend werknemers de mate van tevredenheid met het werk hebt vastgesteld, heeft het weinig zin al die gegevens los te presenteren. Meestal vat je ze samen in bijvoorbeeld een histogram (subparagraaf 4.4.1) of in de vorm van percen-

tages of een gemiddelde (paragraaf 4.1). Je beschrijft je gegevens in een gereduceerde en daardoor overzichtelijke vorm.

Als je gegevens grafisch weergeeft, is het resultaat nogal eens de zogenoemde *normaalverdeling*. In figuur 1.3 zie je daar een (fictief) voorbeeld van.

Normaal-  
verdeling

**FIGUUR 1.3** Het intelligentieniveau van de 198 werknemers van de firma Arbeid



De figuur geeft de verdeling van scores op een intelligentietest weer die is afgenomen bij 198 werknemers van de firma Arbeid. Die verdeling neemt nagenoeg de vorm van een normaalverdeling aan. Ter vergelijking is de normaalverdeling ingetekend. Deze verdeling wordt ook wel de verdeling van *Gauss* of *Gausskromme* genoemd. Het karakteristieke kenmerk van de normaalverdeling is de vorm van een klok en de symmetrie van de linker- en de rechterhelft.

Met SPSS kun je controleren of de verdeling van jouw gegevens de normaalverdeling benadert (paragraaf 4.4).

Wanneer de 198 werknemers een aselechte steekproef vormen uit het totale bestand van werknemers van de firma Arbeid ( $N = 2213$ ), moeten we gebruikmaken van de inductieve of inferentiële statistiek. De vraag is dan in welke mate het gevonden gemiddelde IQ van 99,7 representatief is voor de totale populatie van werknemers. Met andere woorden, wat is de kans op een gemiddeld IQ van 99,7 in de populatie wanneer je inderdaad de gehele populatie in je onderzoek zou kunnen betrekken? Die kans is uiteraard niet zo groot. Want het gemiddelde dat je in de steekproef hebt gevonden is afhankelijk van de toevallige samenstelling van deze steekproef. Wanneer we opnieuw een aselechte steekproef trekken en nog een keer et cetera, dan zal het gemiddelde IQ waarschijnlijk steeds iets hoger of iets lager zijn. De gevonden waarden zullen wel iets van elkaar afwijken, maar waarschijnlijk ook weer niet erg veel. In SPSS kun je de standaardfout laten berekenen (paragraaf 4.2). De *standaardfout* geeft aan in hoeverre het gevonden steekproefgemiddelde een *betrouwbare* schatting is van het populatiegemiddelde. De standaardfout is groter naarmate het verschil in IQ binnen de groep groter en de steekproef kleiner is. De standaardfout

Standaardfout

wordt dus bepaald door de steekproefgrootte en de homogeniteit van de steekproef. Op basis van de standaardfout kun je bijvoorbeeld met minstens 95% zekerheid aangeven dat het populatiegemiddelde ligt tussen het gevonden steekproefgemiddelde minus tweemaal de standaardfout en het gemiddelde plus tweemaal de standaardfout.

Het begrip *zekerheid* of *kans* speelt een belangrijke rol in de inductieve statistiek. Ook wanneer je de gemiddelden van twee steekproeven vergelijkt, is het de vraag wat de kans is dat je een gevonden verschil in gemiddelden terugvindt in de populatie. Stel dat de steekproef van werknemers van de firma Arbeid uit 99 vrouwen en 99 mannen bestaat. Je vindt dat het gemiddelde IQ van de vrouwen 102,2 is en dat van de mannen 98,2. Kun je dan stellen dat de vrouwelijke werknemers van de firma Arbeid gemiddeld intelligenter zijn dan de mannelijke werknemers? Of dat verschil 'significant' is, kun je toetsen. In hoofdstuk 5 leggen we uit hoe je dat doet voor verschillen en in hoofdstuk 6 hoe je dat doet bij samenhangen. Wanneer spreek je nu van *significantie*? Men houdt over het algemeen de regel aan dat er van significantie sprake is als de overschrijdingskans kleiner is dan 5% of bij grotere steekproeven (> 1000) kleiner is dan 1%. Vaak staat er in de SPSS-uitdraai ook bij of er *een- of tweezijdig* getoetst is. Je toetst eenzijdig (*one-tailed*) wanneer je een hypothese of verwachting hebt geformuleerd. Als je een theorie hebt op grond waarvan je kunt verwachten dat de vrouwelijke werknemers intelligenter zijn, dan kun je eenzijdig toetsen. Heb je echter geen idee of er sprake is van een verschil en zeker niet van de richting van dat verschil, dan toets je tweezijdig (*two-tailed*).

Het bepalen van de significantie is gebaseerd op enkele kenmerken van de steekproef. Dat zijn vaak de omvang en de homogeniteit van de steekproef. Naarmate de steekproef groter is, is de kans op toeval uiteraard kleiner. En naarmate de verschillen op een variabele binnen groepen kleiner zijn (homogene groepen), is de kans dat de verschillen tussen groepen op toeval berusten eveneens kleiner.

Als een verschil significant is, wil dat niet automatisch zeggen dat het ook *relevant* is. Je ziet in het voorbeeld dat het verschil in gemiddeld IQ vier IQ-punten in het voordeel van de vrouwen is. Dat verschil is inderdaad significant, de kans dat het op toeval berust, is 2%, dus kleiner dan 5%. Het verschil van vier IQ-punten is dan wel significant, maar niet erg relevant. Het zegt erg weinig over de verschillen tussen mannelijke en vrouwelijke werknemers. Je ziet tegenwoordig in onderzoeksverslagen dat naast de significantie steeds vaker een *effectgrootte* wordt vermeld. De gebruikelijke maat voor effectgrootte is *Cohen's d*. De Cohen's *d* in het voorbeeld is 0,17, wat een te verwaarlozen effect is, pas als *d* groter is dan 0,20 wordt er van een klein effect gesproken. Het geslacht verklaart hier namelijk nog geen 1% van de verschillen in IQ. In hoofdstuk 5 leggen we uit hoe je de effectgrootte kunt berekenen.

In de SPSS-uitvoer kom je ook vaak de term *vrijheidsgraden* (*df = degrees of freedom*) tegen. Het aantal vrijheidsgraden geeft de mate aan waarin scores kunnen variëren. Als je van twee getallen er maar één kent (namelijk 36) en je weet dat het gemiddelde 40 is, dan moet het andere getal 44 zijn. Je hebt hier één vrijheidsgraad. Als je namelijk het ene getal weet, weet je het andere ook. Bij veel toetsen, zoals de t-toets (subparagraaf 5.4.1), is het aantal vrijheidsgraden het aantal steekproefelementen minus 1. Bij een kruistabel (zie paragraaf 5.2) is het aantal vrijheidsgraden het aantal rijen minus 1, vermenigvuldigd met het aantal kolommen minus 1. Voor een 2x2-tabel is het aantal vrijheidsgraden dus 1. Als de randtotalen van een 2x2-kruistabel bekend zijn en je weet ook een van de celfrequen-

### Significantie

### Een- of tweezijdig toetsen

### Effectgrootte

### Vrijheidsgraden

ties, dan kun je de andere celfrequenties berekenen. Vrijheidsgraden zijn van belang als je op basis van een steekproef een schatting wilt maken van bijvoorbeeld het gemiddelde van de populatie. Of een waargenomen verschil of samenhang in een steekproef significant is, hangt dus mede af van het aantal vrijheidsgraden. Dit aantal vrijheidsgraden is, met uitzondering van kruistabellen, vaak afhankelijk van de grootte van de steekproef.

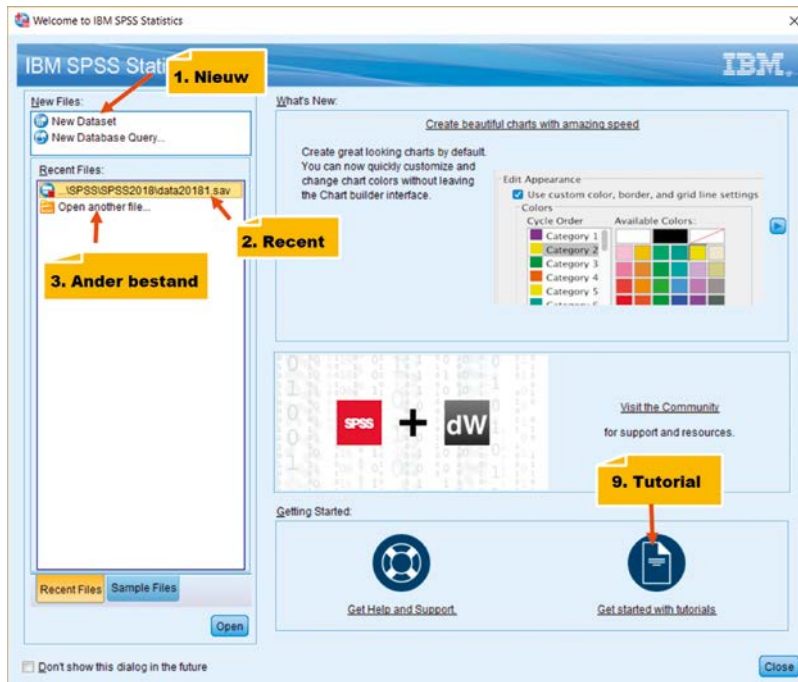
## 1.5 Hoe werkt SPSS onder Windows?

Zoals in de inleiding al is aangegeven, is SPSS een uitgebreid statistisch softwarepakket dat veel gebruikt wordt. Als SPSS is geïnstalleerd, vind je, als het goed is, het icoon op je scherm zoals afgebeeld in figuur 1.4.

FIGUUR 1.4 Het IBM-SPSS-icoon



FIGUUR 1.5 Het openingscherm van SPSS



Je ziet dan het openingsscherm zoals dat in figuur 1.5 is weergegeven.

In het openingsscherm zijn verschillende mogelijkheden aangegeven:

1. *New Files*

Als eerste biedt SPSS je de mogelijkheid om te beginnen met het maken van een nieuw databestand. Als je op 'New Dataset' dubbelklikt, opent SPSS een lege datasheet, die je kunt vullen met de gegevens van bijvoorbeeld je enquête. Hoe je dat doet leggen we uit in hoofdstuk 2.

2. *Recent files*

Als je verder wilt gaan met een bestand dat je al eerder gebruikt hebt, vind je het waarschijnlijk onder het kopje 'Recent files'. Je dubbelklikt dan op het betreffende bestand en SPSS opent het (figuur 1.5). SPSS heeft het meest gebruikte recente bestand gemarkeerd.

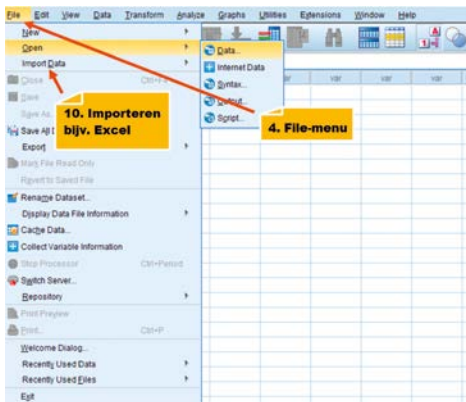
3. *Open another file*

Als je op 'Open another file' dubbelklikt, kun je andere SPSS-databestanden die bijvoorbeeld op je harde schijf staan openen. Je geeft dan aan waar een bestand staat dat je wilt gebruiken en SPSS opent het.

4. *File-menu*

Je kunt een bestand ook op de meer gebruikelijke manier openen met de file-knop die je in de Data Editor vindt (zie figuur 1.6). Je ziet daar dat je kunt kiezen uit verschillende soorten files. We concentreren ons hier op de data-file. De andere soorten files komen later aan bod.

**FIGUUR 1.6** Het openen van een datafile via het file-menu



5. *Data View*

Het eerste wat je ziet als je een bestand opent is het *Data View*-scherm; daar staan alle ingevoerde gegevens (figuur 1.7), zoals de scores op de geld- en geluksvragen van de respondenten

FIGUUR 1.7 Het Data View-scherm als je Data2018 opent

	idcnt	gld1	gld2	gld3	gld4	gld5	glik1	glik2	glik3	glik4
1	1	1,00	2,00	2,00	2,00	2,00	4,00	1,00	4,00	1,00
2	2			1,00	1,00	2,00	2,00	3,00	2,00	3,00
3	3	2,00	2,00	2,00	1,00	1,00	3,00	4,00	2,00	3,00
4	4	2,00	1,00	2,00	2,00	2,00	4,00	2,00	4,00	2,00
5	5	1,00	1,00	1,00	2,00	2,00	5,00	5,00	4,00	2,00
6	6	2,00	2,00	1,00	2,00	2,00	3,00	1,00	4,00	3,00
7	7	1,00	2,00	2,00	1,00	1,00	4,00	4,00	3,00	2,00
8	8	1,00	1,00	1,00	1,00	2,00	4,00	4,00	3,00	2,00
9	9	1,00	2,00	1,00	2,00	2,00	4,00	2,00	4,00	1,00
10	10	2,00	2,00	1,00	1,00	1,00	1,00	4,00	3,00	4,00
11	11	1,00	1,00	1,00	2,00	2,00	4,00	1,00	4,00	1,00
12	12	1,00	1,00	2,00	1,00	2,00	4,00	2,00	4,00	2,00
13	13	1,00	1,00	1,00	2,00	2,00	5,00	5,00	4,00	2,00
14	14	1,00	2,00	2,00	1,00	1,00	4,00	4,00	3,00	2,00
15	15	1,00	1,00	1,00	1,00	2,00	4,00	4,00	3,00	2,00
16	16	1,00	2,00	1,00	2,00	2,00	4,00	2,00	4,00	1,00
17	17	2,00	2,00	1,00	1,00	1,00	1,00	4,00	3,00	4,00
18	18	1,00	1,00	1,00	2,00	2,00	4,00	1,00	4,00	1,00
19	19	2,00	2,00	1,00	1,00	2,00	2,00	3,00	2,00	3,00
20	20	2,00	2,00	2,00	1,00	1,00	3,00	4,00	2,00	3,00
21	21	1,00	1,00	1,00	2,00	2,00	1,00	5,00	4,00	2,00
22	22	2,00	2,00	1,00	2,00	2,00	3,00	1,00	4,00	3,00
23	23	1,00	2,00	2,00	1,00	1,00	4,00	4,00	3,00	2,00

6. Variable View

Als je links onderin op 'Variabele View' klikt, kom je in het variabelenscherm. Daarin geeft SPSS aan welke variabelen in het bestand voorkomen en wat de eigenschappen van die variabelen zijn (figuur 1.8), zoals het feit dat de variabele gld1 gaat over het bezit van een auto. In hoofdstuk 2 leggen we uit wat al die eigenschappen uit het Variable View-scherm inhouden en hoe je ze kunt veranderen.

Variabelenscherm

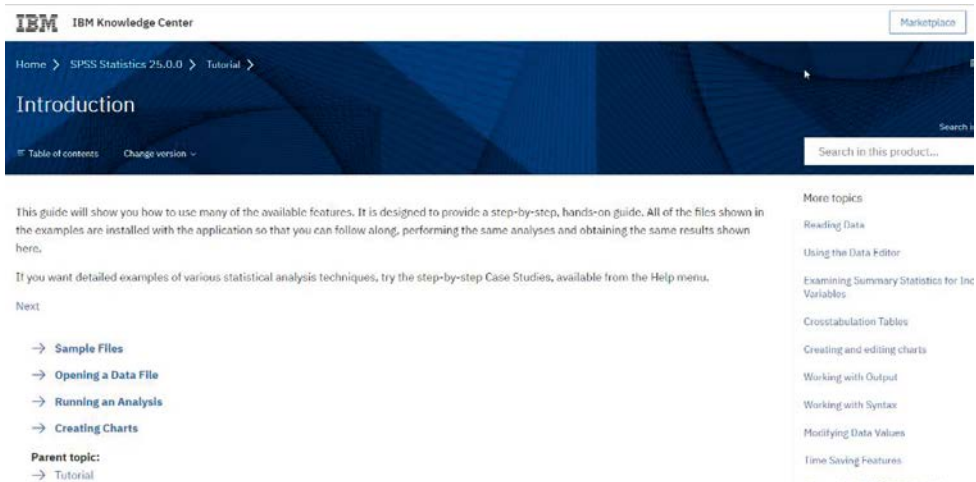
FIGUUR 1.8 Het Variable View-scherm als je Data2018 opent

Je kunt nu een aantal dingen doen. Je kunt de data gaan bewerken. Je kunt er ook analyses op uitvoeren via 'Analyze' (7), of grafieken maken met 'Graphs' (8). In de volgende hoofdstukken werken we deze mogelijkheden uit aan de hand van concrete voorbeelden.

## 9. Tutorials

Get started with tutorials biedt je de mogelijkheid om over allerlei onderwerpen (zie figuur 1.9) nadere uitleg te krijgen.

FIGUUR 1.9 De onderwerpen in het Tutorial-onderdeel



## 10. Import Data

In SPSS kun je ook werken met gegevens uit andere dataverwerkingsprogramma's zoals Excel (figuur 1.6). Alleen moet je dan wel aangeven hoe SPSS de Excel-gegevens moet lezen. SPSS helpt je bij het inlezen van de voor SPSS vreemde data. Het gaat voor dit Basisboek te ver, om uit te leggen hoe dit precies in zijn werk gaat. Wil je meer weten over het importeren van Excel-gegevens in SPSS, kijk dan eens naar de volgende video:  
[www.youtube.com/watch?v=6E9IhVKRS4A](http://www.youtube.com/watch?v=6E9IhVKRS4A)

Bij afsluiten van het programma wordt je gevraagd of de (eventuele) wijzigingen die je in het bestand hebt aangebracht moeten worden opgeslagen. Het is verstandig een bestand waarin je wijzigingen hebt aangebracht op te slaan onder een nieuwe naam, mocht je een fout hebben gemaakt dan heb je altijd nog de laatste versie van het bestand.



# Samenvatting

- 
- ▶ In paragraaf 1.1 wordt aangegeven wat in dit boek besproken wordt aan de hand van de voorbeeldcasus 'Geld, geluk en gezondheid' (paragraaf 1.2).
  - ▶ Het is het verstandig om wanneer je zelf een onderzoek doet voor je begint jezelf een aantal vragen te stellen:
    - Gaat het in mijn onderzoek om een frequentie, verschil of samenhang? (subparagraaf 1.3.1)
    - Wat is het meetniveau van mijn gegevens: nominaal, ordinaal, interval of ratio? (subparagraaf 1.3.2)
    - Betreft het een populatie of een steekproef, want in het geval van een steekproef moet je gebruikmaken van inductieve of inferentiële statistiek
  - ▶ In paragraaf 1.4 worden kort de volgende statistische begrippen besproken: normaalverdeling, standaardfout, betrouwbaarheid, zekerheid, significantie, één- of tweezijdige toetsing, relevantie, effectgrootte, vrijheidsgraden.
  - ▶ In paragraaf 1.5 wordt uitgelegd hoe je SPSS start, uit welke schermen (de hoofdschermen: 'Variable View' en 'Data View') het is opgebouwd en waar je binnen SPSS hulp kunt vinden (de 'Tutorials').
-