

# Handleiding bij SPSS

bij

## Praktijkgericht onderzoek in de (para)medische zorg

**Eveline Wouters**  
**Yvonne van Zaalen**  
**Janna Bruijning**

Eveline Wouters, oktober 2011  
Fontys Paramedische Hogeschool  
Lectoraat Health Innovations and Technology (HIT)  
© Lectoraat HIT

uitgeverij | C  
coutinho

bussum 2015

Deze handleiding bij SPSS hoort bij [Praktijkgericht onderzoek in de \(para\)medische zorg](#) van Eveline Wouters, Yvonne van Zaalen en Janna Bruijning.

© 2012 Uitgeverij Coutinho bv

Alle rechten voorbehouden.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van reprografische verveelvoudigingen uit deze uitgave is toegestaan op grond van artikel 16 h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, [www.reprorecht.nl](http://www.reprorecht.nl)). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, [www.stichting-pro.nl](http://www.stichting-pro.nl)).

Eerste druk 2012

Tweede, herziene druk 2015

Uitgeverij Coutinho

Postbus 333

1400 AH Bussum

[info@coutinho.nl](mailto:info@coutinho.nl)

[www.coutinho.nl](http://www.coutinho.nl)

Noot van de uitgever

Wij hebben alle moeite gedaan om rechthebbenden van copyright te achterhalen. Personen of instanties die aanspraak maken op bepaalde rechten, wordt vriendelijk verzocht contact op te nemen met de uitgever.

ISBN 978 90 469 0427 5

NUR 890

## Voorwoord

Voor het analyseren van kwantitatieve data is het programma SPSS een erg handig hulpmiddel. Er zijn diverse boeken die je met deze software op weg kunnen helpen. Deze handleiding is bedoeld voor een allereerste kennismaking tijdens een korte introductie-cursus van ongeveer anderhalf uur. Omdat alle instructies stap voor stap zijn beschreven, kun je echter ook zonder begeleiding aan de slag gaan.

Bij het samenstellen van deze handleiding is gebruikgemaakt van de boeken waarnaar gerefereerd wordt in de bronnenlijst en van de handleiding SPSS die gebruikt wordt voor de opleiding Physician Assistant, Hogeschool Rotterdam.

In de paragrafen 6 tot en met 8 maken we gebruik van een dataset die je kunt downloaden van [www.coutinho.nl/wouters](http://www.coutinho.nl/wouters).

Eveline Wouters  
Najaar 2014

# Inhoudsopgave

## Handleiding bij SPSS | 5

Inleiding | 5

**1** Starten van SPSS | 5

**2** Algemene informatie | 6

**3** Invoeren van gegevens | 8

**4** Analyseren van data | 10

**5** Opslaan van gegevens | 11

**6** Statistische basistechnieken | 12

**7** Grafische uitvoer | 16

**8** Toetsen | 17

Bronnen | 17

# Handleiding bij SPSS

## Inleiding

Er zijn verschillende versies van SPSS (Statistical Package for the Social Sciences). In 2009 veranderde de naam SPSS voor enige tijd in PASW (Predictive Analytics SoftWare). Dit betreft de versies 17.0.3 tot en met 18.0.3. Met het uitkomen van versie 19 keerde 'SPSS' weer terug in de officiële naam en sindsdien heeft men het over IBM SPSS Statistics. De meest recente versie is IBM SPSS Statistics 20.0.

De versies zijn allemaal hetzelfde in het gebruik, maar hoe nieuwer de versie, hoe uitgebreider de mogelijkheden. Verschillen tussen de versies zul je daarom alleen tegenkomen bij complexere analyses. Dus ook als je een oudere versie van SPSS gebruikt, zul je met deze handleiding vooruit kunnen.

Studenten van de meeste hogescholen krijgen korting op het programma als ze het bestellen via Surfspot. De aanschaf bedraagt dan ongeveer 16 euro en een licentie is een jaar geldig. Voor een klein bedrag kun je de licentie daarna verlengen ([www.surfspot.nl](http://www.surfspot.nl)).

Dit is een introductie in SPSS. Na het doorlopen van de instructies kun je je eigen data in een database zetten en enkele eenvoudige analyses uitvoeren. We beperken ons wat betreft de statistische toetsing tot één voorbeeld. Afhankelijk van je eigen onderzoeksvraag kun je dan volgens hetzelfde principe andere toetsen toepassen.

Je leert het snelst met het programma te werken door er direct mee aan de slag te gaan. Deze handleiding bevat dan ook opdrachten die je zelfstandig uit moet voeren. Probeer ook verschillende opties uit, zodat je erachter komt wat je zelf het handigst vindt.

## 1 Starten van SPSS

Start SPSS (PASW) op. Meestal verschijnt er een klein venster met de vraag: 'What would you like to do?' Hiermee kun je onder andere recent geopende bestanden openen. Als je met een nieuw bestand wilt beginnen, druk je op **<Cancel>** waardoor je vanzelf in het beginvenster van SPSS komt: **IBM SPSS Statistics Data Editor** (of: **PASW Statistics Data Editor**).

De bovenste balk is de *titelbalk*. Hierin staat de naam die je het bestand hebt gegeven nadat je het een keer hebt opgeslagen. Begin je met een nieuwe dataset, dan is er nog geen naam en zie je links in de titelbalk **Untitled1** staan.

Onder de titelbalk staat de *menubalk* met het hoofdmenu: **File, Edit, View, Data, Transform, Analyze**, etc. Deze items kun je aanklikken waarna het submenu verschijnt (vergelijkbaar met Word).

De derde balk is de *toolbar*. Als je met de muispijl op een knop van de toolbar gaat staan, zie je in de *statusbalk* onder in het venster wat de betekenis van die knop is.

## 2 Algemene informatie

SPSS werkt met vensters die je allemaal afzonderlijk moet openen en ook afzonderlijk moet opslaan. De belangrijkste zijn:

- de *data-editor*: hierin staan de data en de variabelen.
- het *syntaxvenster*: hierin sla je de syntax van de SPSS-opdrachten op. Je kunt hiermee een opdracht die al eerder uitgevoerd is nogmaals runnen. De syntax is de formule waarmee je bijvoorbeeld een statistische toets hebt uitgevoerd. Hoe je het syntaxvenster gebruikt, valt buiten de doelstellingen van deze handleiding.
- het *outputvenster*: hierin komen de resultaten van de analyses. Het outputvenster kun je alleen maar openen nadat je een analyse gedaan hebt.

Je opent de vensters door middel van **<File> <Open> <Data...>**, **<File> <Open> <Syntax...>** of **<File> <Open> <Output...>**.

### De data-editor

In de data-editor werk je met gegevens (data) en gegevensbestanden (datafiles). De gegevens zijn de basis van je statistische analyse. De data-editor bestaat uit twee tabbladen, die je afwisselend kunt oproepen door linksonder in het venster op de betreffende tabs te klikken:

- 1 **Data View**: bevat de gegevens van je onderzoek die geanalyseerd moeten worden. Boven aan de kolommen (verticaal) staan de variabelen (bijvoorbeeld *IDnummer*, *geslacht*, *BMI*, *leeftijd*). De waarden van de variabelen staan in de cellen. In de rijen (horizontaal) staan de onderzoeksgevallen (cases, deelnemers). Bij een enquête is bijvoorbeeld elke respondent een case.
- 2 **Variable View**: geeft een overzicht van de definities van de verschillende variabelen (zie hieronder).

### Het tabblad Variable View

In het tabblad **Variable View** van de data-editor worden de eigenschappen van de variabelen gedefinieerd. Van links naar rechts zijn dat:

**Name** De naam van de variabele. Het is handig om een naam te kiezen die goed past bij de inhoud. SPSS stelt een aantal voorwaarden aan de naamgeving, maar welke dat precies zijn, is afhankelijk van de versie die je gebruikt. In versie 20 zijn bijvoorbeeld geen spaties en bepaalde andere tekens toegestaan. Je merkt vanzelf wanneer een naam niet aan de voorwaarden voldoet, want SPSS geeft dan de melding 'Variable name contains an illegal character' of iets dergelijks. Voorbeeld: voor het identificatienummer van een deelnemer zou je de naam *IDnummer* kunnen kiezen.

**Type** Het type variabele. De meest gebruikte typen zijn:

- *numeriek*: de variabele bestaat uitsluitend uit cijfers (ook decimalen) en je kunt er dus mee rekenen.

- *alfanumeriek* (Engels: *strings*): de variabele bestaat uit letters of andere tekens, bijvoorbeeld de naam van de kliniek waar het onderzoek heeft plaatsgevonden.

**Width** Het aantal tekens (karakters) dat de ingevoerde waarde mag hebben. Het getal 45,456 heeft bijvoorbeeld 6 tekens en 'Catharinaziekenhuis' heeft 19 tekens. Standaard staat het veld **Width** op 8 wat voor getallen doorgaans meer dan voldoende is, maar voor woorden niet.

**Decimals** Het aantal decimalen van de ingevoerde waarde. Houd alles zo duidelijk mogelijk door het aantal decimalen op 0 te zetten als er geen decimalen nodig zijn, bijvoorbeeld als de variabele de leeftijd in hele jaren weergeeft. Standaard staat deze waarde op 2 (decimalen).

**Label** Een uitgebreidere omschrijving voor de variabele. Hier kun je ook de hele vraag of stelling van een enquête opschrijven. De naam van de variabele is dan bijvoorbeeld *vraag-1* en het label: 'trap lopen kost mij veel moeite'.

**Values** Codes die aan verschillende waarden van een variabele zijn toegekend. Bij de variabele *geslacht* kun je bijvoorbeeld de code 1 voor 'man' gebruiken en de code 2 voor 'vrouw'. Dit werkt sneller bij het invoeren van de data. Welke cijfers je gebruikt, bepaal je zelf. (Vaak worden 0 en 1, of 1 en 2 gebruikt.) Een in eerste instantie alfanumerieke variabele wordt door het toekennen van codes dus numeriek. Tip: probeer dit zo veel mogelijk te doen voor variabelen die een antwoord weergeven. Voorbeeld: 'eens', 'oneens' en 'geen mening' kun je coderen met 1, 2 en 3 (1 = 'eens', 2 = 'oneens', 3 = 'geen mening').

**Missing** Waarden die in het bestand ontbreken, bijvoorbeeld als een respondent niet alle vragen ingevuld heeft. Hierbij vul je ofwel niets in (no missing values) ofwel je geeft er een onmogelijke waarde aan (bijvoorbeeld 999 voor gewicht). Dit wordt discrete missing value genoemd.

De volgende twee eigenschappen hebben geen invloed op de statistische analyse en zijn dus minder van belang:

**Columns** De kolombreedte van de variabele op het tabblad **Data View**.

**Align** De uitlijning van het tabblad **Variable View** (alfanumerieke variabelen links en numerieke variabelen rechts uitgelijnd).

De laatste eigenschap is wel weer van belang:

**Measure** De meetschaal. Je kunt kiezen uit **<Nominal>**, **<Ordinal>** en **<Scale>**. Een variabele waarvan de waarden in klassen zijn verdeeld, zoals de variabelen *geslacht* en *overgewicht*, noemen we *nominale variabelen*. Aan de verschillende klassen kunnen getallen worden toegekend (bijvoorbeeld: 0 = 'geen overgewicht', 1 = 'overgewicht'), maar met deze getallen kun je niet rekenen.

Bij *ordinale variabelen* hebben de klassen bovendien een bepaalde rangorde (bijvoorbeeld meer of minder opleiding: 1 = mbo, 2 = hbo, 3 = wo). Heb je te maken met variabelen waar je wel mee kunt rekenen, zoals leeftijd en gewicht, dan kies je voor **<Scale>** (Nederlands: intervallschaal/ratioschaal). Vermeld de rekeneenheid, bijvoorbeeld jaren of aantal kg, in het label van de variabele.

### 3 Invoeren van gegevens

Aan de hand van een voorbeeld ga je nu een minidatabase bouwen. Voordat je de waarden van de variabelen in kunt voeren, moet je eerst de database *opzetten*. Hieronder word je met logische stappen door dit proces geloodst.

Stel dat je voor je onderzoek vier medische dossiers hebt verzameld waarvan je de volgende gegevens wilt analyseren:

- gewicht in kg
- medicatiegebruik
- leeftijd in jaren
- wel of geen operatie in de voorgeschiedenis

De gegevens zijn als volgt in een tabel samengevat:

Casenummer	Lichaamsgewicht	Medicatiegebruik?	Leeftijd	Operatie in voorgeschiedenis?
1	65,4	nee	25	ja
2	70,0	ja	35	ja
3	100,5	ja	32	ja
4	80,9	nee	28	nee

Dit zijn dus 4 cases met 5 variabelen. Het invoeren gaat in twee stappen: (1) het invoeren van de namen van de variabelen en (2) het invoeren van de getallen.

#### Stap 1 ■ Het invoeren van de namen van de variabelen

Bedenk hoe je de variabelen wilt noemen, rekening houdend met de voorwaarden die SPSS stelt aan de naamgeving. De hierboven genoemde variabelen kunnen bijvoorbeeld worden ingevoerd als *casenr*, *gewicht*, *medicatie*, *leeftijd* en *operatie*.

Start SPSS op en ga naar de data-editor, tabblad **Variable View**. Je kunt tussen de tabbladen wisselen door linksonder in het venster op de tabs te klikken. De eerste regel is geel gemarkeerd en is bestemd voor de eerste variabele uit je database – meestal het casenummer van het dossier of de respondent. Typ *casenr* in de cel onder **Name** en druk op **<Enter>**. In de overige cellen van de rij zet SPSS nu automatisch een standaardwaarde. Klik vervolgens met de muis in de tweede cel onder **Name**, typ hier *gewicht* en druk weer op **<Enter>**. Vul de overige drie variabelen ook op deze manier in.



Alle vijf de variabelen maken we van het type numeriek. Je ziet in de kolom onder **Type** dat SPSS dit standaard al heeft ingevoerd. Als je een ander type variabele wilt, klik je op de cel in de kolom **Type** en vervolgens op het grijze vierkantje met de drie puntjes, waarna een keuzeschermbijbeeld verschijnt. Klik bijvoorbeeld op het rondje voor **String** als je een alfanumerieke variabele wilt hebben. Daarnaast kun je in dit keuzeschermbijbeeld het maximale aantal cijfers en het aantal decimalen wijzigen, maar deze eigenschappen kun je uiteraard ook rechtstreeks in de volgende twee kolommen wijzigen (**Width** en **Decimals**). Tip: zet **Decimals** waar dit kan op 0 om alles zo overzichtelijk mogelijk te houden. Dit is dus het geval voor *casenr*, *medicatie*, *leeftijd* en *operatie*. Stel *gewicht* in op één decimaal.

De kolom **Labels** hoeft niet ingevuld te worden om een analyse te kunnen uitvoeren, maar dit mag wel. Een label geeft vaak duidelijkheid, bijvoorbeeld als de variabelenaam een afkorting is. Door een uitgebreidere omschrijving als label toe te voegen kun je de variabelen tijdens de analyse gemakkelijker herkennen. Ook als je met meerdere onderzoekers aan een dataset werkt, voorkomt het gebruik van labels foute interpretaties.

In dit voorbeeld vul je de volgende labels in door te klikken op de betreffende cel in de kolom **Label**:

- bij variabele 1 (*casenr*): *casenummer*
- bij variabele 2 (*gewicht*): *gewicht in kg*
- bij variabele 3 (*medicatie*): *wel of geen medicatiegebruik*
- bij variabele 4 (*leeftijd*): *leeftijd in jaren*
- bij variabele 5 (*operatie*): *wel of geen operatie in de voorgeschiedenis*

Vervolgens ken je valuelabels toe aan de (nominale) variabelen *medicatie* en *operatie* om vast te leggen met behulp van welke nummers je straks de waarden van deze variabelen gaat invoeren. Klik daarvoor in de kolom **Values** op de cel van de variabele *medicatie* en vervolgens op het grijze vierkantje met de drie puntjes. Hiermee roep je het venster **Value Labels** op. Typ achter **Value** de code 0 in en achter **Label** de omschrijving *geen medicatie*. Klik vervolgens op **<Add>**. Vervolgens herhaal je de handeling en koppel je de omschrijving *wel medicatie* aan de code 1. Klik hierna op **<OK>** en doe hetzelfde voor de variabele *operatie*: *geen operatie* krijgt code 0, *wel operatie* code 1. Tip: kies steeds dezelfde code voor 'ja', 'nee', 'wel', 'niet', etc. Dit maakt het invoeren van de data veel gemakkelijker en voorkomt vergissingen, vooral als je met meerdere personen aan een onderzoek werkt.

Je kunt de valuelabels van alle variabelen controleren door in de menubalk te klikken op **<Utilities>** en **<Variables>**, waarna een venstertje verschijnt met een overzicht van de variabelen. Door op een variabele te klikken zie je in het rechterschermbijbeeld informatie over die variabele, waaronder de toegekende valuelabels. Hier klik je aan wat van toepassing is.

## Stap 2 ■ Het invoeren van de getallen

Het geraamte voor je database (dit wordt ook wel het *codebook* genoemd) is nu klaar en je kunt beginnen met het invoeren van de data die je hebt verzameld. Ga hiervoor naar het tabblad **Data View** en vul vervolgens de data uit de voorbeeldtabel op bladzijde 8 in. Let erop dat je bij de variabelen *medicatie* en *operatie* de codes gebruikt die je bij de valuelabels hebt ingevoerd. Je kunt deze codes natuurlijk terugvinden via het tabblad **Variable**

**View**, maar je kunt ook op het icoontje **<Value Labels>** in de toolbar klikken (tweede van rechts). De ingevoerde code (0 of 1) verandert dan meteen in het bijbehorende valuelabel.

De variabelen *casenr*, *gewicht* en *leeftijd* kun je gewoon als getallen invullen. Let op: gebruik bij decimalen een komma en geen punt. De punt wordt door de nieuwere versies van SPSS namelijk niet herkend: 34.5 wordt dan geïnterpreteerd als 345 en afhankelijk van het ingestelde aantal decimalen bijvoorbeeld weergegeven als 345,00. In de oudere versies van SPSS (versie 17 en daarvoor) wordt wel met punten gewerkt. Het omgekeerde kan overigens ook voorkomen, namelijk dat de komma niet wordt herkend, maar wel de punt. In dat geval heeft je computer waarschijnlijk Engelse instellingen (Windows: zie de landinstellingen in het configuratiescherm). Overigens hoeft je niet de landinstellingen van Windows te wijzigen als je wilt switchen tussen punten en komma's. Het zou ook binnen SPSS moeten kunnen met een van de commando's SET LOCALE of SET DECIMAL in het syntaxvenster.

### Missing values

Als er gegevens ontbreken, dan spreken we van *missing values*. Er zijn twee soorten missing values:

- **User missing values:** dit zijn waarden die de gebruiker symbolisch heeft ingevoerd als er geen 'echte' waarde voorhanden is, bijvoorbeeld als een respondent is vergeten een vraag in te vullen. De gebruiker kan dan een fictief getal invullen, dat buiten de range van de variabele ligt (bijvoorbeeld 999).
- **System missing values:** de waarden zijn door SPSS zelf gecreëerd wanneer een case of een cel niet is ingevuld en worden weergegeven door een puntje.

## 4 Analyseren van data

De volgende stap is het analyseren en wegschrijven van gegevens uit de data-editor naar het outputvenster. Stel je wilt de variabelen *gewicht* en *leeftijd* met elkaar vergelijken. Ga in de menubalk naar **<Analyze> <Reports> <Case Summaries...>**. Er verschijnt nu een venster met twee kolommen. De linker kolom is een overzicht van de variabelen die je hebt ingevoerd, de rechter kolom is nog leeg.

Selecteer nu met de muis in de linker kolom de variabele *gewicht* en verplaats deze met behulp van de pijlknop tussen de kolommen (of door te slepen) naar rechts. Doe dit ook met de variabele *leeftijd*. Je kunt ook meerdere variabelen in één keer overbrengen door de Ctrl-toets ingedrukt te houden tijdens het selecteren van de variabelen. Druk vervolgens op **<OK>**. SPSS gaat aan het rekenen en het outputvenster verschijnt. Hierin zie je een overzicht van de opgegeven variabelen met hun waarden.

Ga nu terug naar het venster **Summarize Cases (<Analyze> <Reports> <Case Summaries...>)**. Rechts zit de knop **<Statistics...>**. Door deze aan te klikken kun je voor de geselecteerde variabelen een groot aantal statistische maten laten berekenen. Verplaats bijvoorbeeld **Mean** van de linker naar de rechter kolom als je de gemiddelde waarden van de variabelen wilt weten. Druk vervolgens op **<Continue>** en **<OK>**.

Kijk of je in het outputvenster de gemiddelde waarden kunt weergeven (79,2 kg en 30 jaar) en de laagst en hoogst voorkomende waarden (65,4 resp. 100,5 kg en 25 resp. 35 jaar). Let op: zolang je het outputvenster niet afsluit, worden de uitkomsten van al je opdrachten onder elkaar in het openstaande outputvenster geplakt. Om het overzichtelijk te houden is het soms handig om per vraag uit je enquête of per onderwerp uit je cases een nieuwe outputfile te openen (via **<File> <New> <Output>**). Zo kun je de gegevens later gemakkelijker terugvinden.

Je kunt zelf kiezen of je in het outputvenster de naam van de variabele krijgt te zien, het label van de variabele of allebei. Dit kun je instellen door in de data-editor te klikken op **<Edit> <Options...>** en dan te kiezen voor het tabblad **Output Labels**. In dit scherm staan vier te wijzigen regels. De derde (**Pivot Tables Labeling – Variables in labels shown as:**) kun je op drie manieren instellen:

- **Labels** (in de output komt: *wel of geen medicatiegebruik*)
- **Names** (in de output komt: *medicatie*)
- **Names and Labels** (in de output komt: *medicatie wel of geen medicatiegebruik*)

## 5 Opslaan van gegevens

Om je werk te bewaren moet je de ingevoerde gegevens én de output opslaan. Ga voor het opslaan van de gegevens naar de data-editor. Het maakt niet uit of je in het tabblad **Data View** of **Variable View** staat. Klik op **<File> <Save As...>** en kies waar je het bestand op wilt slaan. SPSS vult automatisch het juiste bestandstype in. Klik vervolgens op **<Save>**. De tabbladen **Data View** en **Variable View** zijn nu beide opgeslagen in één bestand met de extensie **.sav**.

Voor het opslaan van de output herhaal je de bovenstaande procedure in het outputvenster. De outputfile wordt altijd opgeslagen met de extensie **.spv** (oudere versies: **.spo**).

Let op: wanneer je **<Ctrl+s>** gebruikt om op te slaan, let er dan op dat het juiste venster actief is. Om vergissingen te voorkomen kun je het beste eerst met de muis klikken in het venster dat je wilt opslaan.

## 6 Statistische basistechnieken

Bij het werken met onderzoeksgegevens komen twee soorten statistiek kijken:

- 1 *beschrijvende statistiek*: deze vorm van statistiek wordt gebruikt voor het ordenen en presenteren van onderzoeksgegevens in bijvoorbeeld frequentietabellen of grafieken. Beschrijvende statistiek wordt ook gebruikt om gegevens te reduceren, bijvoorbeeld door het gemiddelde en de standaarddeviatie aan te geven.
- 2 *toetsende statistiek*: deze vorm van statistiek wordt gebruikt om conclusies te trekken over een populatie op basis van een representatieve steekproef. Hieronder vallen verschillende toetsen zoals de chi-kwadraattoets, de t-toets etc.

Met beschrijvende statistiek verkrijgt je inzicht in de verdeling van een variabele. De meeste procedures hiervoor in SPSS vind je door in de menubalk op **<Analyze>** te klikken en vervolgens naar **<Descriptive Statistics>** te gaan. De procedure die je vervolgens gebruikt (**<Frequencies...>** of **<Descriptives...>**), is onder andere afhankelijk van het type variabele (nominaal, ordinaal, interval) waarover je iets wilt weten. Het heeft bijvoorbeeld geen zin om het gemiddelde van een nominale variabele (zoals geslacht) uit te rekenen. In plaats daarvan zul je een frequentietabel willen genereren. Voor leeftijd en gewicht is het wel zinvol om een gemiddelde te laten berekenen.

### Frequentietabellen

Stel je wilt weten hoeveel mensen uit je databestand geopereerd zijn. Klik hiervoor in de menubalk op **<Analyze>** **<Descriptive Statistics>** **<Frequencies...>**. Het venster **Frequencies** verschijnt. Selecteer in de linker kolom de variabele *operatie* en verplaats deze met behulp van de pijltjesknop naar de rechter kolom. Druk vervolgens op **<OK>**. Het outputvenster verschijnt met een frequentietabel waarin zichtbaar is dat 1 persoon (25%) niet geopereerd is en 3 personen (75%) wel geopereerd zijn.

Met het venster **Frequencies** is meer mogelijk dan alleen het genereren van een frequentietabel. Als je in dit venster op **<Statistics...>** klikt, kun je bijvoorbeeld ook percentielen, het gemiddelde en de standaarddeviatie opvragen. Via de knop **<Charts...>** kun je een staaf- of cirkeldiagram laten maken, waarbij je bovendien kunt kiezen voor een weergave in aantallen of een weergave in percentages. Met **<Format...>** wijzig je de opmaak van de frequentietabel. Ten slotte kun je er nog voor kiezen om de frequentietabel niet weer te geven in de output, bijvoorbeeld omdat je alleen geïnteresseerd bent in het bijbehorende staafdiagram. In dat geval haal je het vinkje voor **Display frequency tables** weg.

Maak nu via de knop **<Charts...>** in het venster **Frequencies** een staafdiagram (*bar charts*) van de variabele *operatie*. Zorg ervoor dat de frequentietabel niet nog een keer in het outputvenster verschijnt.

### Statistische kengetallen

Om statistische kengetallen (gemiddelde, standaarddeviatie, variantie, minimum, maximum, etc.) te berekenen, ga je naar **<Analyze>** **<Descriptive Statistics>** **<Descriptives...>**. Selecteer in het venster **Descriptives** de variabele *gewicht* en breng deze met de pijlknop naar de rechterkolom. Standaard worden nu het gemiddelde, de standaarddevi-

atie, het minimum en het maximum berekend. Onder de knop <Options...> kunnen nog meer kengetallen gekozen worden.

### Voorbeeld: de gemiddelde leeftijd van een groep respondenten

Download van [www.coutinho.nl/wouters](http://www.coutinho.nl/wouters) het voorbeeldbestand **Dataset bij SPSS handleiding.sav** en open dit in SPSS. In het kader hieronder lees je hoe de dataset tot stand is gekomen.

#### De dataset

Er wordt een enquête gehouden onder studenten van verschillende opleidingen. Het doel van de enquête is na te gaan of de studentenfolder *Gezonde voeding*, die verspreid is onder alle eerstejaars, effect heeft gehad. Er wordt nagegaan of de folder ontvangen, gelezen en/of besproken is. Ook worden leeftijd, woonsituatie en studierichting gevraagd. Ten slotte wordt een aantal inhoudsvragen gesteld, waarvan er een in de dataset is opgenomen: een flesje Bacardi breezer bevat meer calorieën dan een flesje (gewoon) bier. Waar of niet waar?

We gaan de gemiddelde leeftijd van de respondenten berekenen, inclusief de standaarddeviatie en de mediaan. Probeer eerst zelf de oplossing te vinden, voordat je verder leest.

Ga in de menubalk naar <Analyze> <Descriptive Statistics> <Frequencies...>. Je zou verwachten dat je <Descriptives...> moet gebruiken omdat het om statistische kengetallen gaat, maar de mediaan is alleen via <Frequencies...> op te vragen.

Je zit nu in het venster **Frequencies**. Selecteer in de linker kolom de variabele *leeftijd* en verplaats deze met de pijl naar de rechter kolom. Klik vervolgens op <Statistics...> rechts in het venster. Hier kun je verschillende uitkomsten selecteren. Voor dit voorbeeld selecteer je **Mean**, **Median** en **St.Deviation**. Druk vervolgens op <Continue> en <OK>. In het outputvenster verschijnen twee tabellen. De eerste tabel bevat het gemiddelde (18,89), de standaarddeviatie (0,896) en de mediaan (19) voor de variabele *leeftijd*. De tweede tabel is de frequentietabel en bevat alle leeftijden met bijbehorende frequenties.

#### Kruistabellen (Crosstabs)

Een kruistabel is een handige manier om inzicht te krijgen in categorische data (veelal nominale variabelen). Een kruistabel is een frequentietabel waarin de waarden van de ene variabele (kolommen) worden gekruist tegen de waarden van de andere variabele (rijen). Het kruispunt van een rij en een kolom is een cel en alle cellen samen vormen de kruistabel. In de cellen staan de waargenomen aantallen (de geobserveerde frequenties).

Een kruistabel wordt gemaakt met de procedure **Crosstabs**. We leggen deze procedure weer uit aan de hand van het bestand **Dataset bij SPSS handleiding.sav**. Open dit bestand en zorg (als oefening) ervoor dat je in het tabblad **Data View** woorden ziet en geen numerieke codes.

Klik achtereenvolgens op **<Analyze>** **<Descriptive Statistics>** **<Crosstabs...>**. Vervolgens kun je aangeven welke variabele in de rijen en welke variabele in de kolommen moet komen te staan.

Selecteer uit het rijtje links in het venster de variabele *Fol\_GEZ\_gelezen* en verplaats deze met de pijlknop naar het vak **Row(s)**. Selecteer vervolgens de variabele *Woonsit* en verplaats deze naar het vak **Column(s)**. Druk op **<OK>**.

Als het goed is, zie je in het outputvenster de volgende tabel (boven deze tabel staat nog een andere tabel met een overzicht van het totale aantal cases die gebruikt zijn en de missing values):

Crosstabulation

Count

		Thuiswonend of op kamers wonend		Total
		thuiswonend	op kamers wonend	
<b>Studentenfolder gezonde voeding gelezen?</b>	nee	3	11	14
	ja	15	32	47
Total		18	43	61

Je ziet dat 47 studenten de folder hebben gelezen. Daarvan wonen er 15 thuis en 32 op kamers.

Als je onderzoek doet, werk je behalve met absolute getallen ook vaak met percentages. SPSS kan hiervoor een uitgebreidere kruistabel maken dan bovenstaande. Herhaal bovenstaande opdracht, maar klik rechts in het venster **Crosstabs** ook op de knop **<Cells...>**, waardoor je in het venster **Crosstabs: Cell Display** komt. Hier heb je de keuzemogelijkheid voor het weergeven van percentages. Klik in het subvenster **Percentages** alle drie de opties aan (**Row**, **Column**, **Total**) en druk vervolgens op **<Continue>** en **<OK>**. In het outputvenster verschijnt de volgende, uitgebreide tabel:

Crosstabulation

			Thuiswonend of op kamers wonend		Total
			thuiswonend	op kamers wonend	
<b>Studentenfolder gezonde voeding gelezen?</b>	nee	Count	3	11	14
		% within Stdentenfolder gezonde voeding gelezen?	21,4%	78,6%	100,0%
		% within Thuiswonend of op kamers wonend	16,7%	25,6%	23,0%
		% of Total	4,9%	18,0%	23,0%
	ja	Count	15	32	47
		% within Stdentenfolder gezonde voeding gelezen?	31,9%	68,1%	100,0%
		% within Thuiswonend of op kamers wonend	83,3%	74,4%	77,0%
		% of Total	24,6%	52,5%	77,0%
Total	Count	18	43	61	
	% within Stdentenfolder gezonde voeding gelezen?	29,5%	70,5%	100,0%	
	% within Thuiswonend of op kamers wonend	100,0%	100,0%	100,0%	
	% of Total	29,5%	70,5%	100,0%	

Hieruit kun je aflezen dat 77% ( $n = 47$ ) de folder gelezen heeft. Hoeveel procent van de thuiswonende en hoeveel procent van de op kamers wonende studenten heeft de folder niet gelezen?

Je zult merken dat als je je gaat verdiepen in de tabel, deze nog best moeilijk te lezen is. Maak het jezelf gemakkelijk door alleen weer te geven waarin je geïnteresseerd bent. Als je bijvoorbeeld alleen de rijpercentages wilt zien, kruis dan in het venster **Crosstabs: Cell Display** alleen dat hokje aan.



## 7 Grafische uitvoer

In de vorige paragraaf heb je gezien dat je via het venster **Frequencies** staaf- en cirkel-diagrammen kunt laten maken (<Analyze> <Descriptive Statistics> <Frequencies...> <Charts...>). Dit kan echter ook via het menu <Graphs>, waar je bovendien nog veel meer grafische mogelijkheden hebt. Open weer het bestand **Dataset bij SPSS handleiding.sav**.

### Histogram

*Versie 20:*

Kies voor <Graphs> in de menubalk en vervolgens voor <Chart Builder...>. Klik op <OK> mocht er een venster met een waarschuwing verschijnen. Je hebt nu het venster **Chart Builder** voor je. Op het tabblad **Gallery**, onder in het venster, kun je aangeven welk soort grafiek je wilt. Kies het eerste icoontje in de rubriek **Histogram**. Sleep het icoontje naar het vak rechtsboven in het venster (er komt een nieuw venster bij: **Element Properties**) en sleep de variabele *leeftijd* naar de *x*-as. In het venster **Element Properties** kun je het histogram verder aanpassen. Vink bijvoorbeeld **Display normal curve** aan – en druk op <Apply> – als je tevens een normaalkromme in het histogram wilt afbeelden. Klik ten slotte op <OK>.

*Versie 18:*

Kies voor <Graphs> in de menubalk en vervolgens voor <Histogram...>. Kies uit het onderste venster de soort grafiek die je wilt (in dit geval de eerste: **simple histogram**) en sleep die naar het bovenste venster. Verplaats de variabele *leeftijd* naar de *x*-as en druk op <OK>.

*Opmerking:*

Bij een histogram deelt SPSS de gegevens in klassen in. De klassengrenzen lopen altijd tot de bovengrens.

### Bar chart (staafgrafiek)

We maken nu een *bar chart* van de variabele *leeftijd*. Kies <Graphs> <Bar> <Simple> <Define>. Klik bij <Bars represent> **N of cases** aan. Definieer de leeftijd van de respondent als **Category Axis**. Klik nu op <OK>.



## 8 Toetsen

Als je in je onderzoek verschillende groepen met elkaar vergelijkt en wilt weten of gevonden verschillen significant zijn, kun je de chi-kwadraattoets (of *Fisher's exact test*) gebruiken. Bij kleine steekproeven (< 25) en een 2x2-tabel kun je beter een *Fisher's exact test* doen.

In deze paragraaf maken we weer gebruik van het voorbeeldbestand **Dataset bij SPSS handleiding.sav**. We willen weten of thuiswoners de stelling 'Een flesje Bacardi breezer bevat meer calorieën dan een flesje gewoon bier' anders beoordelen dan kamerbewoners.

Ga naar **<Analyze> <Descriptive Statistics> <Crosstabs...>**. In het nu geopende venster verplaats je de variabele *woonsituatie* naar het venster **Row(s)** en *inhoudvraag* naar het venster **Column(s)**. Klik vervolgens op **<Statistics...>** en vink **<Chi-square>** aan. Klik dan op **<Continue>** en in het volgende venster op **<OK>**. Je krijgt de volgende output:

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	10,302a	1	,001		
Continuity Correctionb	8,552	1	,003		
Likelihood Ratio	10,381	1	,001		
Fisher's Exact Test				,002	,002
Linear-by-Linear Association	10,133	1	,001		
N of Valid Cases	61				

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 7,38.

b Computed only for a 2x2 table

Is het verschil tussen respondenten met verschillende woonsituaties significant? Kijk hierbij naar de kolom **Asymp. Sig. (2-sided)** en de rij **Pearson Chi-Square**.

Wat betekent dit resultaat in woorden?

### Bronnen

Baarda, D.B., Goede, M.P.M. de & Dijkum, C.J. van (2011). *Basisboek Statistiek met SPSS*. Groningen: Noordhoff Uitgevers BV.

Vocht, A. de (2011). *Basishandboek SPSS 19*. Utrecht: Bijleveld Press.