

Basiscursus SPSS

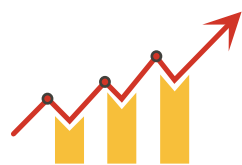
versie 20-23

Het statistische programma SPSS is al geruime tijd op de Nederlandse markt. In die periode zijn er veel leerboeken verschenen waarin het programma uitvoerig werd beschreven, vaak in combinatie met de bijbehorende statistische theorie. Dit boek wijkt af van deze aanpak doordat de auteurs zich beperken tot het SPSS-programma zelf en daarvan de onderwerpen behandelen die een gemiddelde gebruiker vaak zal tegenkomen.

De didactische opzet is dusdanig dat de lezer vanaf hoofdstuk 2 zelfstandig opdrachten uitvoert aan de hand van een praktijkvoorbeeld. De handelingen zijn duidelijk beschreven en de belangrijkste uitkomsten zijn weergegeven ter controle. Er is geen wiskundige achtergrond vereist. De basiscursus vormt daarmee een laagdrempelige kennismaking met het analyseren van kwantitatieve gegevens met behulp van SPSS.

Manfred te Grotenhuis is universitair hoofddocent kwantitatieve analysetechnieken bij de sectie Sociologie van de Radboud Universiteit. Anneke Matthijssen is onderwijskundige (specialisatie: zelfstandig leren) en werkzaam als beleidsmedewerker / Institutional Researcher bij de Radboud Universiteit.

*In dit boek wordt gebruikgemaakt van diverse bestanden.
Deze zijn beschikbaar via: webwinkel.vangorcum.nl/spss*



 koninklijke
van gorcum



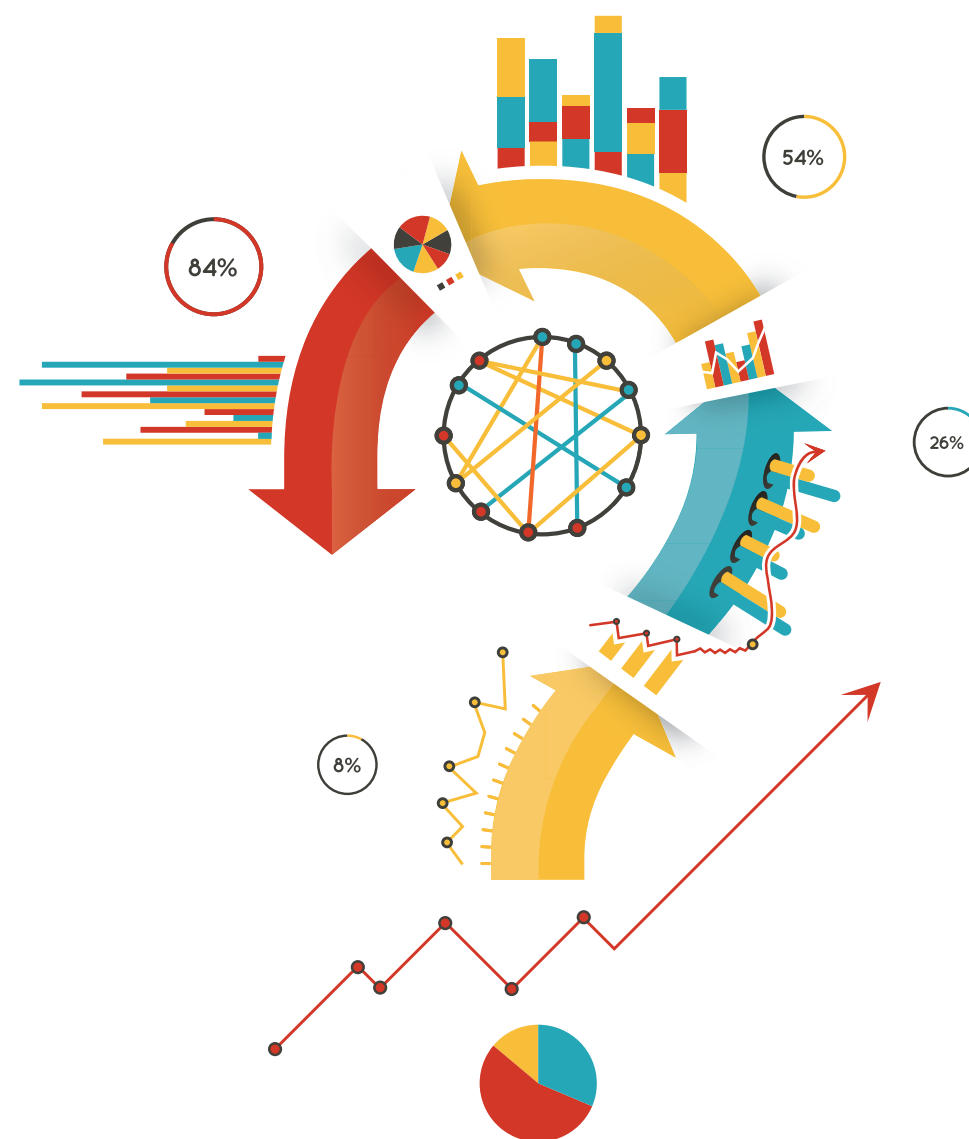
Basiscursus SPSS

versie 20-23

Te Grotenhuis & Matthijssen



Manfred te Grotenhuis &
Anneke Matthijssen



Basiscursus SPSS

versie 20-23

Basiscursus SPSS

Versie 20-23

Manfred te Grotenhuis
Anneke Matthijsen

6e druk

© 2016, Koninklijke Van Gorcum BV, Postbus 43, 9400 AA Assen.

Behoudens de in of krachtens de Auteurswet 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 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3060, 2130 KB Hoofddorp www.reprorecht.nl). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.stichting-pro.nl).

ISBN folio 978 90 232 5500 0

ISBN e-book 978 90 232 5501 7

NUR 916

Uitgave: Uitgeverij Koninklijke Van Gorcum, Assen
Omslagontwerp: Kim Boeren, Viesrood grafisch & interactief ontwerp
Grafische verzorging: Imago Mediabuilders, Amersfoort
Druk: Drukkerij Van Gorcum, Assen

VOORWOORD

V

In 1968 startte in Californië op Stanford University een project dat zou uitgroeien tot één van de meest populaire statistische programma's: SPSS (oorspronkelijk: **S**tatistical **P**ackage for the **S**ocial **S**ciences). Het programma is in een groot aantal leerboeken beschreven, vaak in combinatie met statistische theorie. Aan de Radboud Universiteit bleek er echter behoefte aan een relatief eenvoudige cursus SPSS om in korte tijd de basis-beginselen over te dragen. Daarom werd de uitgebreide uitleg van de statistiek geschrapt en kwam de nadruk te liggen op het zelfstandig uitvoeren van de juiste computerhandelingen. Voor dit doel werd een bescheiden dictaat geschreven, dat na elk cursusjaar werd aangepast en verbeterd. Zo is uiteindelijk dit boekje ontstaan, waarmee voldoende kennis, maar vooral vaardigheid wordt ontwikkeld om gangbare statistische bewerkingen en analyses zelfstandig uit te voeren.

Nijmegen, 2004

VOORWOORD BIJ DE ZESDE DRUK

De Basiscursus SPSS wordt inmiddels meer dan 10 jaar gebruikt op hogescholen en universiteiten in Nederland en België. Daarnaast is er in 2015 een internationale editie verschenen. Dit is voor ons het beste bewijs dat het nuttig is om de basishandelingen in het programma SPSS al doende uit te leggen zonder uitgebreid in te gaan op de statistische achtergronden.

De afbeeldingen en de instructies zijn afgestemd op de versies 20-23 van SPSS. Gebruikers van versie 18 en 19 kunnen echter ook probleemloos van deze druk gebruikmaken. De onderwerpen zijn grotendeels hetzelfde gebleven in vergelijking met de vorige druk. De tekst is op sommige plaatsen herzien en het gebruik van het populaire programma Excel is uitvoeriger beschreven in hoofdstuk 2. Verder is de meervoudige regressie-analyse toegevoegd aan hoofdstuk 5. Om de leesbaarheid te verhogen is de opmaak verbeterd en is de indeling licht gewijzigd.

De uitleg van de 'multiple response analyse' blijft gratis beschikbaar via de bijbehorende internetpagina. Wij bedanken alle studenten en de vele referenten voor de enthousiaste reacties en de opbouwende commentaren uit de afgelopen jaren.

De auteurs
Nijmegen, 2016

INHOUD

<i>Voorwoord</i>	V
<i>Voorwoord bij de zesde druk</i>	V
Hoofdstuk 1 Het programma SPSS	1
1.1 Wat is SPSS?.....	1
1.2 Waarvoor dient SPSS?.....	1
1.3 Structuur van het boek.....	2
1.4 De fictieve gegevens in de voorbeelden.....	2
1.5 Vensters, balken en menu's.....	4
1.5.1 Het data-venster.....	4
1.5.2 De titelbalk.....	5
1.5.3 De menubalk.....	5
1.5.4 De knoppenbalk.....	6
1.5.5 De statusbalk.....	7
1.6 Werken met vensters.....	7
1.6.1 Het datavenster.....	7
1.6.2 Het outputvenster.....	8
1.6.3 Het syntaxvenster.....	9
1.7 Bestandstypen.....	10
Hoofdstuk 2 SPSS-bestanden	11
2.1 Inleiding.....	11
2.2 Het openen van bestanden.....	11
2.3 Het aanmaken van bestanden.....	13
2.3.1 Het definiëren van variabelen in SPSS.....	13
2.3.2 Variabelen voorzien van labels.....	16
2.3.3 Variabelen categorienamen geven (value labels).....	17
2.3.4 Het definiëren van ongeldige scores (missing values).....	19
2.3.5 Het invoeren van gegevens.....	20
2.4 Het openen van Excel-bestanden.....	24
2.5 Het opslaan van SPSS-bestanden.....	29
2.6 Opdrachten.....	31
Hoofdstuk 3 Bewerken van variabelen	35
3.1 Inleiding.....	35
3.2 Het hercoderen van variabelen.....	35
3.3 Het berekenen van nieuwe variabelen.....	39
3.4 Het selecteren van analyse-eenheden (cases).....	42
3.5 Het opsplitsen van bestanden.....	46
3.6 Opdrachten.....	49

<i>Hoofdstuk 4 Beschrijvende statistiek</i>	51
4.1 Inleiding.....	51
4.2 Het maken van frequentieverdelingen.....	52
4.3 Het maken en bewerken van grafieken (taartdiagram).....	57
4.4 Het maken van doosdiagrammen (boxplots).....	61
4.5 Grafieken met 'Chart Builder'.....	63
4.6 Het maken van kruistabellen.....	66
4.6.1 <i>Bivariate kruistabel</i>	66
4.6.2 <i>Andere bivariate tabellen en trivariate tabellen</i>	71
4.7 Uitvoer naar tekstverwerkingsprogramma's.....	71
4.8 Opdrachten.....	73
<i>Hoofdstuk 5 Toetsende statistiek</i>	75
5.1 Inleiding.....	75
5.2 Samenhang in een kruistabel.....	75
5.3 Binomiale toets op één proportie.....	79
5.4 T-toets op één gemiddelde.....	81
5.5 T-toets op verschil tussen twee gemiddelden.....	83
5.5.1 <i>Onafhankelijke groepen</i>	83
5.5.2 <i>Afhankelijke (gekoppelde) groepen</i>	86
5.6 Variantie-analyse.....	88
5.7 Correlatie.....	92
5.8 Regressie-analyse.....	94
5.8.1 <i>Enkelvoudige regressie-analyse</i>	94
5.8.2 <i>Meervoudige regressie-analyse</i>	97
5.9 Opdrachten.....	100
<i>Tot besluit</i>	102
<i>IBM SPSS</i>	102
<i>Index</i>	103


2.3.3 VARIABELN CATEGORIENAMEN GEVEN (VALUE LABELS)

Het is ook mogelijk om omschrijvingen toe te voegen aan de codes of categorieën van een variabele.

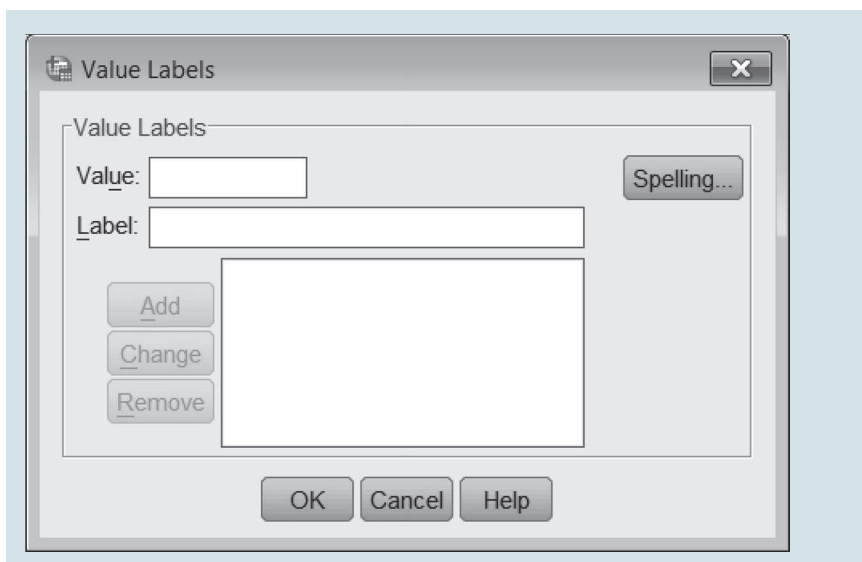
VOORBEELD

In het datavenster staat de variabele 'sexe'. Deze variabele bestaat uit twee codes/categorieën: 0 = man en 1 = vrouw. Het is handig om aan te geven (te labelen) waar die 0 en 1 voor staan, zodat dit later in tabellen en dergelijke wordt weergegeven. Bij het bekijken van bijvoorbeeld de frequentieverdeling, wordt automatisch de categorienaam (value label) door SPSS vermeld, zodat dit niet opgezocht hoeft te worden.



- Klik op **None** in de tweede cel onder **Values** (dit is een cel uit de rij waarin 'sexe' staat).
- Klik op het icoontje  rechts naast **None**.

Nu verschijnt het dialoogvenster 'Value Labels' (zie afbeelding 2.5).



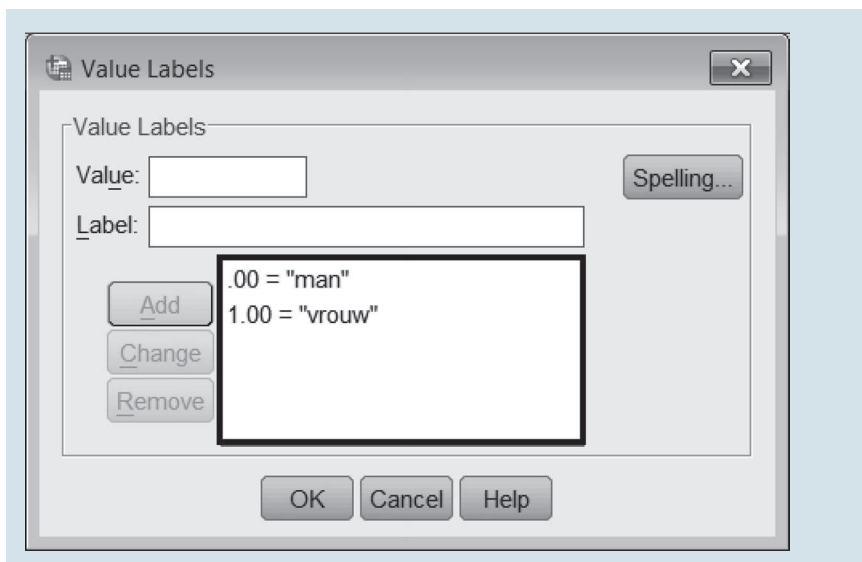
Afbeelding 2.5 Het dialoogvenster 'Value Labels'

18



- Voer in het lege veld achter **Value** het cijfer 0 in en klik met de muis in het veld achter **Label** en vul het woord *man* in.
- Klik op **Add** om de opdracht uit te voeren.
- Voer nu achter **Value** het cijfer 1 in en voer achter **Label** *vrouw* in.
- Klik wederom op **Add** om de opdracht uit te voeren.

Controleer de handelingen aan de hand van afbeelding 2.6.



Afbeelding 2.6 Het dialoogvenster 'Value Labels' ingevuld



Klik afsluitend op **OK**
(*'sexe'* heeft nu de juiste categorienamen).

2.3.4 HET DEFINIËREN VAN ONGELDIGE SCORES (MISSING VALUES)

Het komt vaak voor dat men één of meerdere categorieën wil uitsluiten van analyse. Bijvoorbeeld omdat het gaat om categorieën zoals 'niet ingevuld', 'weet niet' of om extreem lage of hoge scores (zogenaamde 'uitschieters' ('outliers')). Dit uitsluiten gebeurt door in SPSS deze categorieën als ongeldige scores op te geven. De scores worden dan niet meegenomen in de statistische bewerkingen. Het opgeven van ongeldige scores kan via het dialoogvenster 'Missing Values' (zie afbeelding 2.7).

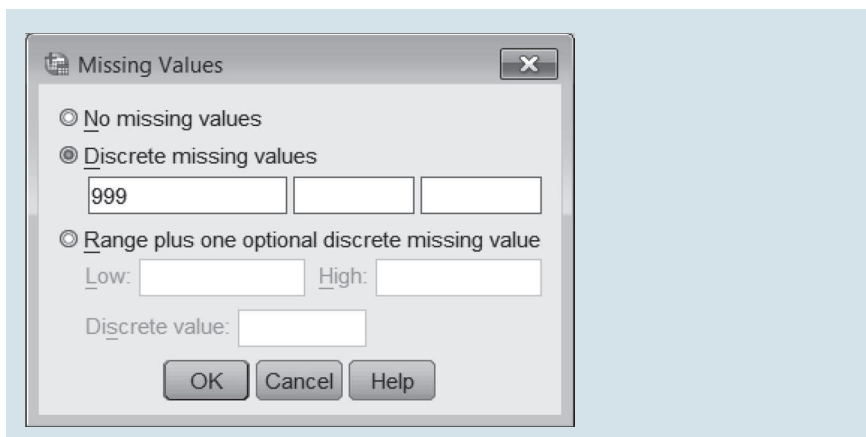
VOORBEELD

Bij de variabele 'huistijd' waarbij de respondenten (studenten) de tijd invullen die zij thuis hebben besteed aan de statistiekopgaven, heeft een aantal van hen niets ingevuld. Waarschijnlijk hebben ze niet bijgehouden hoeveel tijd ze thuis hebben besteed. Voor deze respondenten is de score 999 gebruikt, omdat dit een waarde is die niet voor zal komen bij de gegeven geldige antwoorden.



- Klik op **None** in de eerste cel onder **Missing** in het variabelenvenster (dit is de rij met de variabele 'huistijd').
- Klik op rechts naast **None**.

Nu verschijnt het dialoogvenster 'Missing Values':

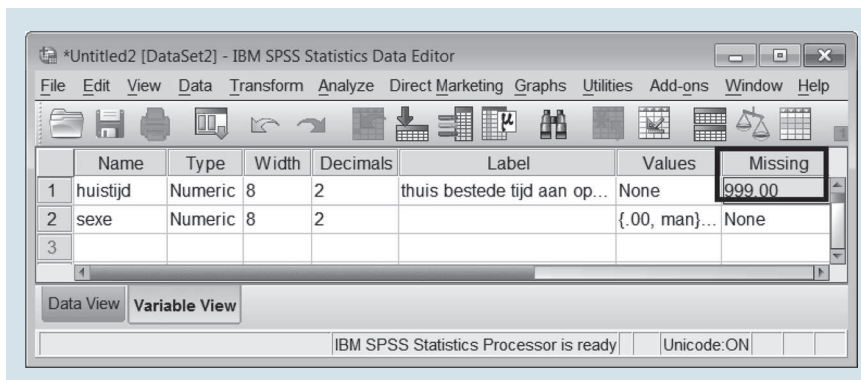


Afbeelding 2.7 Het dialoogvenster 'Missing Values'



- Vink **Discrete missing values** aan, klik in het eerste lege veld daaronder en typ de cijfers 999 in (zie afbeelding 2.7).
- Klik op **OK** voor bevestiging.

Als alles goed is uitgevoerd, ziet het variabelenvenster er als volgt uit (zie laatste kolom (Missing) in onderstaande afbeelding 2.8):



Afbeelding 2.8 Het variabelenvenster met 'huistijd' en 'sexe' (999 is aangegeven als ongeldige score ('missing') bij de variabele 'huistijd')

De volledige gegevens van de twee variabelen zijn nu ingevoerd. Normaal gesproken moeten er echter nog meer variabelen worden aangeemaakt. Dit is een vrij tijdrovende bezigheid, daarom wordt het hier achterwege gelaten. Het bestand met alle variabelen dat is gebruikt om de gegevens uit de evaluatie van de cursus statistiek in te voeren, wordt in de volgende paragraaf opgevraagd.

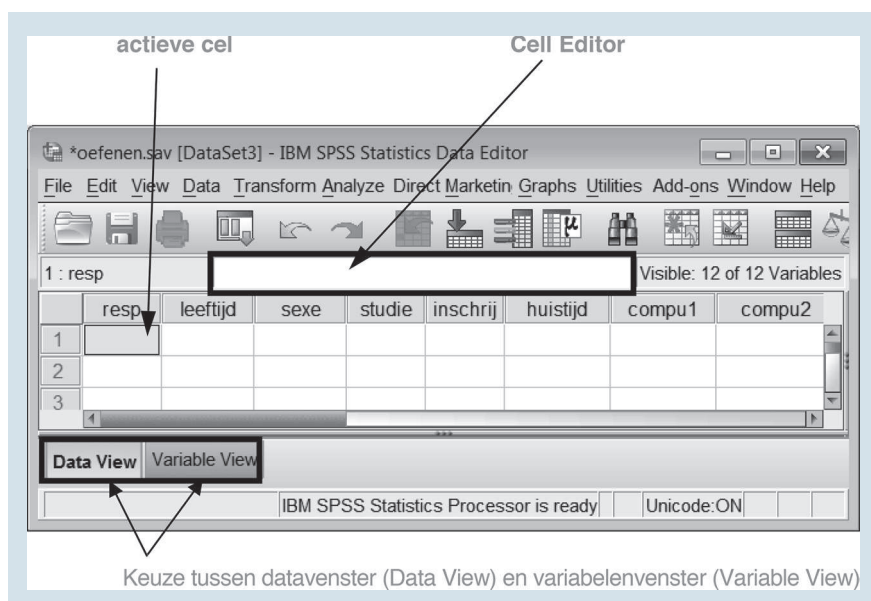
2.3.5 HET INVOEREN VAN GEGEVENS



- Open het databestand *oefenen.sav*. Zie voor handelingen paragraaf 2.2.
- Als het variabelenvenster (afbeelding 2.8) verschijnt, klik dan op **Data View** linksonder in het variabelenvenster.

Het datavenster van SPSS is nu zichtbaar (zie afbeelding 2.9). Dit venster is een eenvoudige vorm van een zogeheten *spreadsheet*. Een spreadsheet bestaat uit rijen en kolommen. De analyse-eenheden (hier: de studenten) staan in de rijen, terwijl de variabelen in de kolommen worden weergegeven. Elk kruispunt van een rij en een kolom wordt een *cel* genoemd. In deze cellen staat één enkele observatie, vaak uitgedrukt in een getal. De cel waarin de cursor staat wordt de *actieve cel* genoemd. Een cel kan actief worden gemaakt door deze aan te wijzen met de muis en vervolgens te klikken met de linker muisknop.

De actieve cel is de enige cel waarin gegevens kunnen worden ingevoerd of gewijzigd. De inhoud van de actieve cel wordt altijd afgebeeld in de *Cell Editor*. Dit is de regel boven de spreadsheet. Als je gegevens invoert, worden ze eerst altijd in de Cell Editor geplaatst. Om de gegevens daadwerkelijk in de actieve cel te zetten, moet je op de [Enter] of [Tab] toets drukken of met de muis een andere cel activeren.



Afbeelding 2.9 Het 'Data View'-venster

De meest gebruikelijke volgorde is om per analyse-eenheid ('case' in het Engels) de waarden/codes in te voeren. De gegevens van één analyse-eenheid (student) staan doorgaans op één regel.

Het invoeren van de gegevens gaat als volgt:



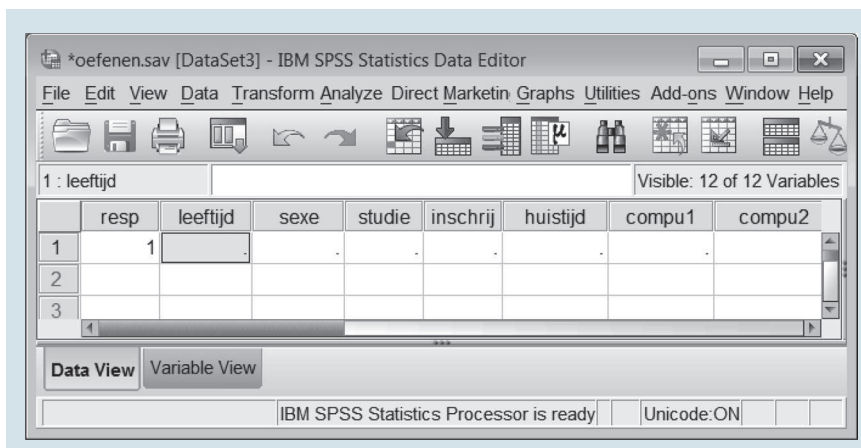
Klik op de eerste lege cel onder de eerste variabele ('resp').

De cel is geel en actief nu (zie afbeelding 2.9) De score op de variabele voor deze student kan nu worden ingetypt. In dit geval gaat het om de variabele 'resp' (= respondentnummer) die dient om de verschillende studenten te kunnen onderscheiden van elkaar. De waarde voor de eerste student is 1.



- Typ de waarde 1 in de lege cel onder resp in.
- Druk op de [Tab] toets op het toetsenbord.

De ingevoerde waarde 1 wordt nu in de actieve cel geplaatst en de cel onder 'leeftijd' wordt actief, zie afbeelding 2.10.

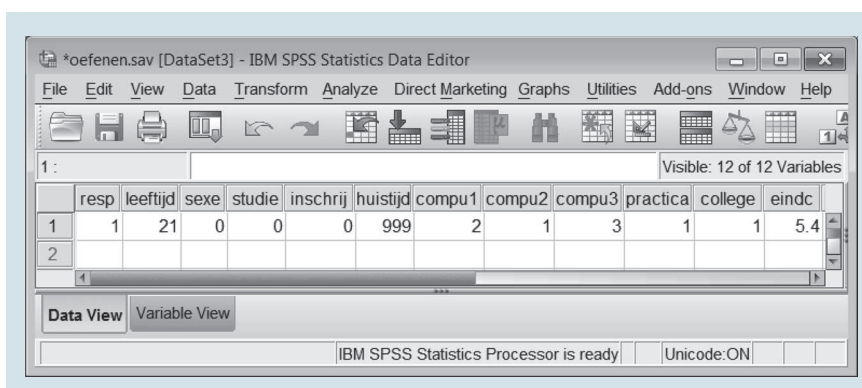


Afbeelding 2.10 De invoer van het eerste getal (respondentnummer 1)



- De eerste student is 21 jaar, man, volgt de studie Pedagogiek, is voltijdstudent, vulde *niet* in hoeveel tijd hij thuis besteedde aan statistiek, heeft weinig ervaring met tekstverwerken (Word), zeer weinig ervaring met spreadsheets (Excel), gemiddelde ervaring met Windows, volgde 1 practicum en 1 college en haalde het cijfer 5,4. Kijk in paragraaf 1.4 welke codes/waarden horen bij bovenstaande gegevens en vul deze in de juiste cellen in.
Let op: computers met Nederlandse notatie accepteren alleen 5,4 en niet 5.4 (Engelse instelling).

Na invullen van de waarden ziet het datavenster er als volgt uit:



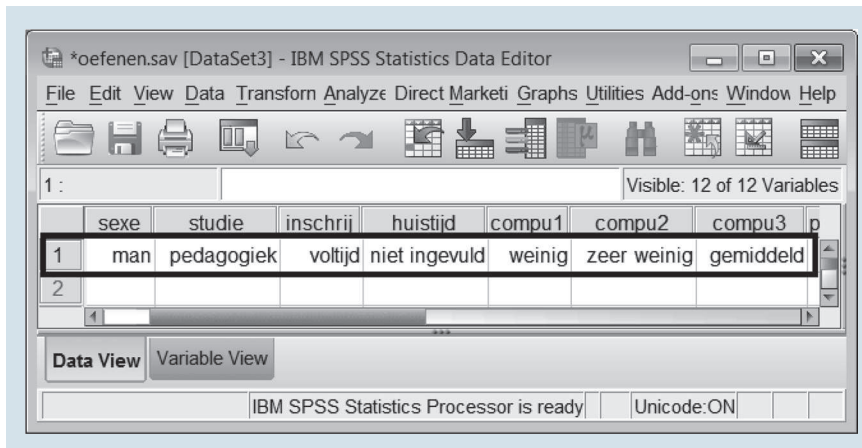
Afbeelding 2.11 Het datavenster met ingevulde waarden van één student



- Kies in het datavenster in de menubalk voor **View**.
- Klik op **Value Labels**.

Vergelijk de inhoud van het eigen datavenster met afbeelding 2.12 en controleer of student (respondent 1) inderdaad man is, Pedagogiek studeert, voltijdstudent is, niet invulde hoeveel tijd hij thuis besteedde aan de opgaven, weinig ervaring heeft met tekstverwerken, zeer weinig ervaring heeft met spreadsheets en gemiddelde ervaring heeft met Windows.

Indien dat niet het geval is, verander dan de bewuste cel waarde (zie uitleg over 'veranderen van cel waarden' verderop) tot dat alles correct is. Het bestand hoeft niet te worden opgeslagen, het was alleen bedoeld om te oefenen met het invoeren van gegevens.



Abbeelding 2.12 Het datavenster met 'value labels' aangevinkt

Ingevoerde cel waarden zijn achteraf op verschillende manieren te bewerken:

- **Veranderen van cel waarden**
Waarden kunnen worden veranderd door op de betreffende cel te klikken en de foutieve waarde te verbeteren.
- **Verwijderen van een variabele of analyse-eenheid**
Een variabele (kolom) of analyse-eenheid (rij) kan worden verwijderd door te klikken op de variabele-naam boven aan de kolom respectievelijk door te klikken op het rijnummer aan de linkerkant en dan te drukken op [**Delete**] op het toetsenbord.

2.4 HET OPENEN VAN EXCEL-BESTANDEN

Het komt in de praktijk regelmatig voor dat de antwoorden als rijtjes getallen in de computer zijn ingevoerd, vaak met andere programma's dan SPSS. Ook het fictieve databestand in deze basiscursus bestond in het begin uit zulke rijtjes (zie het onderstaande voorbeeld). Er bestaan veel zogenaamde 'spreadsheet' programma's die dergelijke gegevens kunnen