

Statistiek als hulpmiddel

© 2023, Uitgeverij 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](http://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](http://www.stichting-pro.nl)).

NUR 916

ISBN folioboek: 9789023259350

ISBN ebook: 9789023259367

1e druk 2008

2e druk 2011

3e herziene druk 2013

4e herziene druk 2023

Uitgave: Uitgeverij Koninklijke Van Gorcum, Assen

Grafische verzorging: Richard Bos

Omslagontwerp: Viesrood grafisch & interactief ontwerp, Zwolle

Redactie Lilian Eefting, Leef in tekst, Groningen

Druk: Drukkerij Van Gorcum, Meppel

 MET TROTS  
GEPRODUCEERD  
IN NEDERLAND

# STATISTIEK ALS HULPMIDDEL

Pieter van Groenestijn

Mark Visser

4e herziene druk

## INHOUD

|   |           |
|---|-----------|
| Voorwoord   | 7         |
| Statistiek als hulpmiddel   | 8         |
| <b>1 Statistische gegevens</b>                                      | <b>10</b> |
| 1.1 Inleiding   | 11        |
| 1.2 Vier soorten variabelen   | 11        |
| 1.3 Selectie van analyse-eenheden: de steekproeftrekking            | 14        |
| 1.4 Verzamelen van statistische gegevens                            | 16        |
| 1.5 Kwaliteit van statistische gegevens                             | 18        |
| 1.6 Van statistische gegevens naar uitspraken                       | 21        |
| <b>2 Excel</b>  | <b>22</b> |
| 2.1 Inleiding   | 23        |
| 2.2 De fictieve gegevens in de voorbeelden                          | 23        |
| 2.3 Exporteren van SPSS-databestanden                               | 24        |
| 2.4 Bewerkingen   | 25        |
| 2.5 Grafieken   | 29        |
| 2.5.1 Cirkeldiagram   | 29        |
| 2.5.2 Staafdiagram  | 31        |
| 2.5.3 Histogram   | 33        |
| 2.5.4 Spreidingsdiagram   | 33        |
| 2.5.5 Lijndiagram   | 35        |
| <b>3 Beschrijvende statistiek</b>                                   | <b>36</b> |
| 3.1 Inleiding   | 37        |
| 3.2 Numerieke beschrijvingen van een variabele                      | 37        |
| 3.2.1 Maten voor het centrum van een verdeling                      | 38        |
| 3.2.2 Maten voor de spreiding in een verdeling                      | 42        |
| 3.2.3 Maten voor de relatieve plaats in een verdeling               | 48        |
| 3.3 Numerieke beschrijvingen van verbanden                          | 52        |
| 3.4 Grafische beschrijvingen van een variabele                      | 52        |
| 3.4.1 Staafdiagram  | 52        |
| 3.4.2 Taartdiagram  | 53        |
| 3.4.3 Histogram   | 54        |
| 3.4.4 Stamdiagram   | 54        |
| 3.5 Grafische beschrijvingen van het verband tussen twee variabelen | 55        |
| 3.5.1 Doosdiagram   | 55        |
| 3.5.2 Spreidingsdiagram   | 56        |
| 3.5.3 Lijndiagram   | 57        |
| 3.6 Samenvatting  | 58        |

|          |  |           |
|----------|--|-----------|
| <b>4</b> | <b>Toetsende statistiek</b>                          | <b>60</b> |
| 4.1      | Inleiding toetsingstheorie                           | 61        |
| 4.2      | Univariate toetsing                                  | 70        |
| 4.2.1    | Toets op gemiddelde                                  | 70        |
| 4.2.2    | Toets op proportie                                   | 72        |
| 4.3      | Bivariate toetsing                                   | 75        |
| 4.3.1    | Toetsen op verschil in gemiddelden                   | 75        |
| 4.3.2    | Samenhangsmaten voor nominale en ordinale variabelen | 82        |
| 4.3.3    | Samenhangsmaten voor interval- en ratiovariabelen    | 97        |
| 4.4      | Multivariate toetsing                                | 102       |
| 4.4.1    | Gangbare modellen                                    | 105       |
| 4.4.2    | Meervoudige lineaire regressieanalyse                | 107       |
| 4.5      | Tot besluit  | 114       |
|          | Index  | 117       |

# VOORWOORD BIJ DE GEHEEL GEWIJZIGDE VIERDE DRUK

In de sociale wetenschappen is kwantitatief onderzoek wijdverspreid. Voor dergelijk onderzoek is kennis van statistiek onmisbaar. Die kennis is vastgelegd in een groot aantal boeken. Vaak zijn het lijvige, Engelstalige boekwerken waarin tal van statistische onderwerpen uitvoerig worden besproken. *Statistiek als hulpmiddel* vormt alleen al door zijn omvang een uitzondering. Afwijkend is ook dat de nadruk ligt op de praktijk en niet op de statistische theorie. Het gaat dus niet om allerlei ingewikkelde formules en complexe berekeningen, maar om hoe je het op een praktische manier kunt toepassen. Dit betekent dat men geen wiskundige kennis hoeft te bezitten om de inhoud van dit boek te kunnen begrijpen. De bedoeling is om lezers in betrekkelijk korte tijd vertrouwd te maken met de statistische toepassingen die gangbaar zijn binnen hun vakgebied.

De statistische toepassingen worden uitgelegd met voorbeelden uit de dagelijkse onderzoekspraktijk. Deze voorbeelden worden uitgewerkt met behulp van het statistische programma SPSS. Voor de grafieken hebben we ervoor gekozen gebruik te maken van Microsoft Excel, zodat het laagdrempelig blijft. Bovendien spelen wij zo in op de onderzoekspraktijk.

Pieter van Groenestijn en Mark Visser  
Nijmegen, 2023

# STATISTIEK ALS HULPMIDDEL

## INLEIDING

Statistiek is er in vele soorten en maten. Vaak associeert men statistiek echter met complexe vormen van wiskunde, met als gevolg dat men statistiek – op voorhand! – als lastig ervaart. In de sociale wetenschappen, zoals de sociologie, pedagogiek en psychologie, is kennis van statistiek geen doel maar een praktisch hulpmiddel. Het heeft dan ook niet veel zin om te leren hoe men complexe formules af kan leiden of om de grondslagen van de statistiek tot in de finesses te kennen. In het statistiekonderwijs is dat in het verleden wel gebeurd, met als resultaat dat studenten weliswaar deskundig waren op het gebied van de statistiek, maar minder goed wisten hoe zij deze kennis konden inzetten voor onderzoek.

In *Statistiek als hulpmiddel* staan de complexere onderdelen uit de statistiek niet centraal en ligt de nadruk op de praktische toepasbaarheid. Met behulp van diverse onderzoeksgegevens wordt geïllustreerd waarom de statistiek als hulpmiddel onontbeerlijk is. Het uiteindelijke leerdoel is dan ook niet om lezers te voorzien van uitputtende kennis van de statistiek. In plaats daarvan hopen we dat de kennis uit dit boek bijdraagt aan een verantwoord gebruik van de statistiek om onderzoeksvragen uit de dagelijkse onderzoekspraktijk te beantwoorden.

## OPBOUW

In het eerste hoofdstuk behandelen we het verzamelen van kwantitatieve gegevens. We gaan onder andere in op gegevens die zijn verzameld via aselecte steekproeven. Aangezien op het internet een schat aan kwantitatieve gegevens beschikbaar is, bieden we ook een beknopt overzicht van enkele belangrijke vindplaatsen.

Hoofdstuk 2 is gewijd aan Microsoft Excel. Dit spreadsheetprogramma is erg populair in de onderzoekspraktijk vanwege de eenvoudige en aantrekkelijke mogelijkheden om middels grafieken onderzoeksresultaten te presenteren. In de overige hoofdstukken maken we, waar mogelijk, gebruik van Excel bij het presenteren van afbeeldingen.

In het derde hoofdstuk komen belangrijke onderdelen van de beschrijvende statistiek aan de orde. Centraal staat de vraag hoe de vaak grote hoeveelheid kwantitatieve gegevens kan worden samengevat. Deze samenvatting kan getalsmatig (numeriek) zijn, bijvoorbeeld het gemiddelde, maar ook grafisch met grafieken, zoals het staafdiagram. Aan de hand van allerlei voorbeelden uit de dagelijkse onderzoekspraktijk zullen we illustreren welke mogelijkheden er zijn om een goede samenvatting te maken.

Hoofdstuk 4 behandelt de toetsende statistiek. Centraal staat de vraag hoe men uitspraken kan doen over een populatie terwijl men slechts de beschikking heeft over informatie afkomstig van een klein deel van deze populatie (steekproef). Het hoofdstuk bevat relatief eenvoudige toetsen op gemiddelde en proportie, maar ook ingewikkeldere toetsingen door middel van regressieanalyse. Steeds staan echter de praktische toepasbaarheid en juiste interpretatie voorop.

## SOFTWARE

Omdat de nadruk ligt op praktische toepassingen, kunnen we niet zonder een goede gereedschapskist in de vorm van een computerprogramma. Gekozen is voor SPSS (oorspronkelijk: **S**tatistical **P**ackage for the **S**ocial **S**ciences) aangezien dit programma veel wordt gebruikt in onderwijs en onderzoek, mede dankzij de gebruikersvriendelijkheid. Ons streven was een betaalbaar, niet te omvangrijk boek te schrijven; het bevat dan ook vrijwel geen uitleg over SPSS zelf. Hiervoor verwijzen we naar twee eerdere publicaties die door Koninklijke Van Gorcum zijn uitgegeven, namelijk *Basicursus SPSS* en *SPSS met Syntax*. SPSS hanteert, net als veel andere statistische softwarepakketten, de Engelse termen voor statistische toepassingen. Deze zullen waar nodig tussen haakjes worden vermeld achter de Nederlandse benamingen. Hoewel je ook met SPSS grafieken kunt maken die bij de statistische analyses horen, heeft het onze voorkeur dit in Excel uit te voeren. Dit komt overeen met de onderzoekspraktijk.

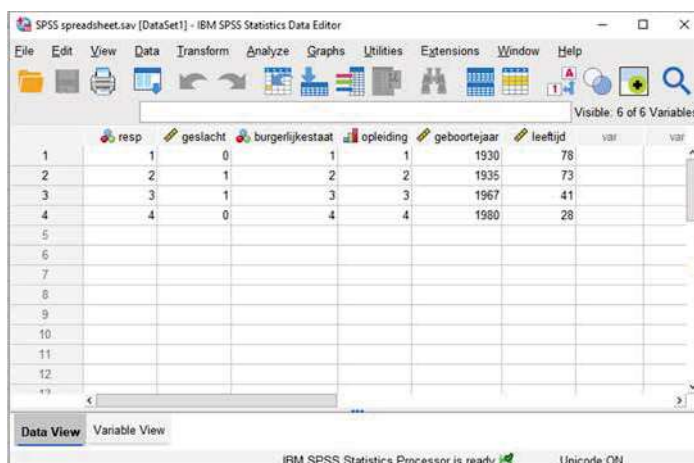


1

# STATISTISCHE GEGEVENS

## 1.1 INLEIDING

Wil men statistische analyses toepassen, dan zijn gegevens nodig. Deze gegevens moeten aan bepaalde eisen voldoen. Een belangrijke eis is dat de gegevens numeriek zijn, dat wil zeggen dat de informatie in de vorm van getallen is opgeslagen. Uiteraard zijn er ook gegevens die in woorden zijn uitgedrukt (alfanumeriek genoemd), maar die worden omgezet in getallen alvorens er statistische bewerkingen plaatsvinden. De numerieke gegevens zijn doorgaans opgeslagen in een zogenaamde spreadsheet. In de rijen van een spreadsheet staan normaal gesproken de analyse-eenheden, dat wil zeggen de eenheden waarover we uitspraken willen doen. In de sociale wetenschappen gaat het hier meestal om personen (in onderzoek ook wel respondenten of proefpersonen genoemd). In de kolommen staan de variabelen, met variërende informatie over de analyse-eenheden. Als de analyse-eenheden personen zijn, dan zijn bijvoorbeeld geslacht, geboortejaar, opleiding, inkomen en burgerlijke staat veelvoorkomende variabelen. In de cellen staan ten slotte de getallen. In afbeelding 1.1 staat een SPSS-spreadsheet afgebeeld met daarin vier respondenten en hun scores op zes variabelen.



|    | resp | geslacht | burgerlijke staat | opleiding | geboortejaar | leeftijd | var | var |
|----|------|----------|-------------------|-----------|--------------|----------|-----|-----|
| 1  | 1    | 0        | 1                 | 1         | 1930         | 78       |     |     |
| 2  | 2    | 1        | 2                 | 2         | 1935         | 73       |     |     |
| 3  | 3    | 1        | 3                 | 3         | 1967         | 41       |     |     |
| 4  | 4    | 0        | 4                 | 4         | 1980         | 28       |     |     |
| 5  |      |          |                   |           |              |          |     |     |
| 6  |      |          |                   |           |              |          |     |     |
| 7  |      |          |                   |           |              |          |     |     |
| 8  |      |          |                   |           |              |          |     |     |
| 9  |      |          |                   |           |              |          |     |     |
| 10 |      |          |                   |           |              |          |     |     |
| 11 |      |          |                   |           |              |          |     |     |
| 12 |      |          |                   |           |              |          |     |     |
| 13 |      |          |                   |           |              |          |     |     |

Afbeelding 1.1 SPSS-spreadsheet (datavenster) met zes variabelen

## 1.2 VIER SOORTEN VARIABELEN

Een variabele meet een bepaald kenmerk van de analyse-eenheid en kan verschillende waarden aannemen. Zo hebben alle respondenten een bepaalde leeftijd en een bepaald opleidingsniveau en in beide zit over het algemeen een behoorlijke variatie. Voor elke variabele moet duidelijk zijn welke situaties zich kunnen voordoen en welke getallen in de spreadsheet kunnen voorkomen. De variabelen hebben daarom een aantal vaste categorieën waaraan unieke waarden zijn verbonden. Zo heeft de variabele 'burgerlijke staat' uit afbeelding 1.1 vier

categorieën die respectievelijk zijn gecodeerd als 1, 2, 3 en 4. De inhoudelijke betekenis van deze waarden is te vinden in het codeboek dat hoort bij de verzamelde gegevens, maar is vaak ook vermeld in het variabelenvenster ('Variable View') van SPSS (zie afbeelding 1.2). Uit deze afbeelding valt af te lezen dat de waarden van 'burgerlijke staat' staan voor 'Ongehuwd' (code 1), 'Gehuwd' (2), 'Gescheiden' (3) en 'Weduwe/Weduwenaar' (4).



Afbeelding 1.2 Het variabelenvenster (boven) en dialoogvenster (onder) met categorienamen in SPSS

In de statistiek worden variabelen verdeeld in vier meetniveaus:

- Nominaal
- Ordinaal
- Interval
- Ratio

*Nominale variabelen* vormen het laagste meetniveau. De categorieën worden enkel onderscheiden door de naamgeving. De gebruikte codes voor de categorieën zijn willekeurig, zolang ze maar verschillend zijn. De codes zijn willekeurig, omdat ze alleen maar hoeven aan te geven dat de categorieën van elkaar verschillen. De variabele 'burgerlijke staat' is een voorbeeld van een nominale variabele, omdat er verschillende categorieën zijn zonder rangorde. Het is namelijk niet zo dat mensen die ongehuwd zijn in enig opzicht 'minder' zijn dan gehuwden. Dit betekent dat men de codering ook mag veranderen in bijvoorbeeld 6 (Ongehuwd), 1 (Gehuwd), 4 (Gescheiden) en -6 (Weduwe/Weduwenaar), zonder dat er inhoudelijk wat verandert.

*Ordinale variabelen* kunnen niet willekeurig worden gecodeerd, omdat er nu wel sprake is van een ordening of rangorde. Zo is de variabele 'opleiding' van ordinaal meetniveau als men ervan uitgaat dat respondenten met een havo-diploma een hoger kennisniveau hebben dan respondenten met een vmbo-diploma, vwo'ers op hun beurt meer kennis bezitten dan havisten, hbo'ers