

Martien G. Schriemer

# STATISTIEK

voor de beroepspraktijk

$$\mu_B = \sum_{j=1}^{10} \frac{x_j}{10} = 52,0$$

B

67 57 47 77  
37 87 97  
27  
17

3

7

1 5 10  
6 8 9  
4


A

2

$$\mu_A = \sum_{i=1}^{10} \frac{x_i}{10} = 5,5$$

Statistiek leren lezen, daarna begrijpen en berekenen met SPSS  
Voor HBO en WO

## Ondersteuning en hulp bij leren

- Studenten kunnen via [www.HetHikkendeHeksje.nl](http://www.HetHikkendeHeksje.nl) (zonder in te loggen) de datasets downloaden die benodigd zijn voor het maken van de opgaven.
- Docenten kunnen via de site tentamenmateriaal en ander ondersteunend collegemateriaal toegesonden krijgen, zoals SPSS-databestanden, Excel-bestanden en dergelijke.
- Alle clips staan op het YouTube-kanaal van Martien Schriemer: <https://www.youtube.com/channel/UCVgKcfUJw3kCkZnrMSoK1w>. In de tekst zijn de verwijzingen naar de clips te herkennen aan 

Eventuele op- en aanmerkingen over deze uitgave (of eraan gelieerde uitgaven) kunt u richten aan de auteur via [info@HetHikkendeHeksje.nl](mailto:info@HetHikkendeHeksje.nl).

Eerste druk 2016

De APA-literatuurverwijzing voor deze uitgave:

Schriemer, M.G. (2016). *Statistiek voor de beroepspraktijk. Statistiek leren lezen, daarna begrijpen en berekenen met SPSS. Voor HBO en WO*. Haarlem: SVW.

Uitgegeven door SVW te Haarlem.

ISBN: 978-90-826322-0-0

Ontwerp omslag: De Zagerij ontwerp bureau, Den Haag

Illustratieverantwoording: Figuur 1.1, 3.1: Martien G. Schriemer

Druk: Pumbo

©2016, Martien G. Schriemer

Alle rechten voorbehouden.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden veeleenvoudigd, 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 auteur.

Voor zover het maken van reprografische veeleenvoudingen uit deze uitgave is toegestaan op grond van artikel 16 h Auteurswet 1912 dient men 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.cedar.nl/pro](http://www.cedar.nl/pro)).

*All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, in any form or by any means, mechanical, photocopying, recording, or otherwise, without prior written permission of the publisher.*

- ▶ **Continue variabelen** zijn gegevens die alle waarden kunnen aannemen. De waarden kunnen dus naast gehele getallen ook breuken bevatten zoals

$$\frac{a}{b} = \frac{9}{11} \text{ en ook irreele getallen zoals } \pi = 3,1428\dots$$

### 2.2.2 Meetniveaus van variabelen

Er zijn vier soorten meetniveaus van variabelen. De soortkenmerken van de niveaus zijn te herkennen aan een viertal eigenschappen: onderscheidend vermogen, ordeningsmogelijkheden, onderlinge verschillen en de onderlinge verhoudingen. Om te kunnen bepalen welk meetniveau een variabele heeft dient er dus gekeken te worden naar de eigenschappen van die variabele. Aan de hand van voorbeelden wordt de eigenschap toegelicht.

Tabel 2.2. Meetniveau en de soortkenmerken van variabelen.

Meetniveau	Nominaal	Ordinaal	Interval	Ratio
Voorbeeld	Sekse	Opleiding	Intelligentiequotiënt	Saldo bankrekening
Eigenschappen	Onderscheid	Onderscheid	Onderscheid	Onderscheid
		Ordening	Ordening	Ordening
			Verschillen	Verschillen
				Verhoudingen

- ▶ Een variabele bezit de eigenschap **onderscheid** als de elementen uit een verzameling van elkaar verschillen op dat punt. Er wordt gesproken van een **nominaal** meetniveau. Zo kunnen mensen van elkaar worden onderscheiden door middel van hun geslacht: vrouw of man. Een nominaal meetniveau betekent dat die variabele alleen maar onderscheid maakt. Oftewel een voorbeeld in wiskundige notatie  $Nom = \{a, b, c, \dots, z\}$ .
- ▶ Een variabele bezit de eigenschap **ordening** als de elementen van de variabele gerangschikt kunnen worden van bijvoorbeeld laag naar hoog of van weinig naar veel. Met andere woorden, als alle elementen uit de verzameling op volgorde zijn gelegd, eerst  $x_i$  dan  $x_{i+1}$ , dan geldt dat elke volgende element groter is dan zijn voorganger. Een **ordinaal** meetniveau houdt in dat die variabele als eigenschap heeft de kenmerken van personen te onderscheiden (bijvoorbeeld 'middelbaar' versus 'laag'), maar ook een ordening aan te brengen. Namelijk iemand die een lage opleiding heeft, heeft minder opleidingsjaren genoten dan iemand met 'middelbaar' en die heeft weer minder opleidingsjaren genoten dan iemand met opleidingsniveau 'hoog'. Dus laag < middelbaar < hoog. Oftewel in wiskundige notatie:  $x_i < x_{i+1}$ .
- ▶ Een variabele bezit de eigenschap **verschil**, als de verschillen tussen de elementen gelijk zijn en dus eenduidig te interpreteren zijn. Met andere woorden van elke drie opeenvolgende elementen geldt dat het verschil tussen de eerste twee en de laatste twee gelijk moet zijn. Dit meetniveau maakt dus naast onderscheid en ordening ook nog een gebruik van verschillen. Een

$$f_7 = \frac{\sum_{i=1}^7 \text{koffie}_i}{\sum_{i=1}^7 \text{bezoekers}_i} = \frac{122+134+121+101+122+122+167}{322+333+326+280+277+256+277} = \frac{889}{2071} = 0,43 = 43\%.$$

Dus naarmate het steekproefkader groter wordt, blijkt dat de relatieve frequentie  $f_i$  tendeert naar 42%. De relatieve kans op de verkoop van koffie aan een bezoeker is dus in dit voorbeeld 42%. Dit komt overeen met het (rekenkundig) gemiddelde.

### 3.7 Dichotomieën

Wanneer in een vragenlijst gevraagd wordt naar het geslacht van de respondent, dan zijn er twee keuzes: man of vrouw<sup>24</sup>. Zo ook als de respondent gevraagd wordt te reageren op een stelling met eens of oneens. Zie kader 3.2 In die gevallen wordt een nominale variabele verkregen waarbij de antwoordopties uit slechts twee mogelijkheden bestaan. Die nominale variabele wordt een **dichotomie** genoemd.

*Kader 3.2. Voorbeelden van een dichotome vraagstelling.*

Bent u:

- Vrouw
- Man
- Zeg ik liever niet

Wat vindt u van de stelling: 'het huidige kabinet regeert het land adequaat'?

- Oneens
- Eens
- Weet ik niet, geen mening

In bovenstaande voorbeelden zijn steeds de eerste twee antwoordopties valide antwoorden (bruikbare respons). Als de respondent geen antwoord wil geven, of het niet weet, is dat de optie die hoort bij een onbruikbare respons. Met behulp van SPSS kan die waarde als 'missing value' gedefinieerd worden.

► Een **dichotomie** is een variabele met uitsluitend twee mogelijke uitkomsten.

👁 Zie **Dichotomie**. Uitleg over de kans op 'ja' en 'nee' met de bijbehorende formules voor proporties. <https://youtu.be/KxPD7knTHIU>

Het is gebruik om dichotomieën met de waarden 0 en 1 te coderen. Bijvoorbeeld 1 = 'eens' en 0 = 'oneens' en -vooruit- voor de volledigheid de missing value 9 = 'zeg ik liever niet'.

<sup>24</sup> In Duitsland is sinds 2013 wettelijk vastgelegd dat er een derde keuze is: 'onbepaald' oftewel blanco (Transgenderinfo, 2013).

## 4.9 Samenvatting: begrippen en wiskundige notaties

Onderstaand wordt in vogelvlucht de begrippen (soms in wiskundige notatie) uit de toetstheorie samengevat en kort toegelicht.

- ▶ Symmetrische relatie      Een relatie tussen twee grootheden A en B is symmetrisch als er (op voorhand) geen richting aan de relatie kan worden gegeven:  $A \leftrightarrow B$ , de samenhang is onbepaald in richting.
- ▶ Asymmetrische relatie      Een relatie tussen twee grootheden A en B is asymmetrisch als er (op voorhand) wel een richting aan de relatie kan worden gegeven:  $A \rightarrow B$ , of  $B \rightarrow A$  de samenhang kent een bepaalde richting.
- ▶ Bias      Een bias in de steekproef is een niet-toevallige fout in de steekproeftrekking.
- ▶ Afhankelijke relatie      Afhankelijkheid van de relatie houdt in dat de uitkomst tussen die twee grootheden een bepaalde mate van voorspelbaarheid bezit, ze beïnvloeden elkaar.
- ▶ Onafhankelijke relatie      De relatie is onafhankelijk te noemen als er geen voorspelbaarheid in zit. De afhankelijkheid van variabelen wordt ook wel een gepaarde relatie genoemd en onafhankelijkheid een ongepaarde relatie. In het Engels (en in SPSS) wordt de terminologie 'paired' en 'unpaired' gehanteerd.
- ▶ Standaard-normalisatie      Dit is een transformatie van  $x$  naar  $z_i$  als volgt  

$$z_i = \frac{x_i - \mu}{\sigma}$$
 Spreek uit als 'z i is gelijk aan x i min mu gedeeld door sigma'. Dit houdt in dat de oorspronkelijke variabele  $x_i$  wordt gecorrigeerd voor zijn gemiddelde  $\mu$  en vervolgens wordt gedeeld door zijn eigen standaarddeviatie  $\sigma$ .
- ▶ Hypothese      Een hypothese is voorafgaand aan het feitelijke onderzoek een vastgelegde en geformuleerde te bewijzen stelling met behulp van onderzoek.
- ▶ Nulhypothese,  $H_0$       In de nulhypothese wordt de verwachting van het onderzoek gepostuleerd, ofwel daarin wordt de hoop uitgesproken wat er uit het onderzoek zal komen.

### 8.3.2 Evaluatie

Sommige vragenlijsten hanteren aan het eind van de vragenlijst nog een paar (korte) vragen die geheel gericht zijn op de evaluatie van het onderzoek. Zo wordt een beeld verkregen in hoeverre de respons serieus genomen moet worden. Een andere analyse in dit kader is dat de **gemiddelde invultijd** geanalyseerd wordt. De 'snelle invullers' dienen geïdentificeerd te worden en vergeleken te worden met de overige respondenten.

- **Snelle invullers** zijn respondenten die in vergelijking met andere respondenten de vragenlijst zeer snel ingevuld hebben en waarschijnlijk de vragenlijst niet serieus hebben genomen. Sommige onderzoeksbureaus hebben kengetallen voor een gemiddelde invultijd, naar gelang de lengte van de vragenlijst.

Zie kader 8.1 voor de bepaling van snelle invullers. Deze analyse heeft mogelijk tot gevolg dat de hoeveelheid bruikbare respons verkleind wordt.

#### *Kader 8.1. Snelle invullers.*

Het vaststellen wie de groep snelle invullers zijn binnen een bepaalde responsgroep kan met behulp van een hercodering. Er dienen twee groepen gemaakt te worden:  $\{R_1, R_2\}$ .

$R_1$ : deze groep respondenten hebben een invultijd  $R$  (responsietijd, vandaar de letter  $R$ ) dat minder is dan  $R < \bar{x} - 1,96 \cdot s$ , waarbij  $\bar{x}$  gelijk is aan de **gemiddelde invultijd** van de enquête en  $s$  de bijbehorende steekproefstandaarddeviatie. De groep  $R_2$  verkregen door  $R \geq \bar{x} - 1,96 \cdot s$  zijn respondenten die tot de 2,5% snelste invullers behoren. De antwoorden op de vragen uit het onderzoek van deze groep mogen op dit punt niet significant afwijken, anders is er immers grond om de respondenten van groep  $R_2$  uit te sluiten.

Respondenten die (heel) lang doen over hun vragenlijst zouden ook een groep respondenten kunnen zijn die afwijkende meningen hebben, maar ervaring leert dat de responsietijd vaak door externe oorzaken lang duurt doordat men tussendoor koffie gaat zetten, men de deurbel beantwoordt, telefoneert, etc.

Als laatste onderdeel dient de **kwaliteit van de respons** nader geanalyseerd te worden. Er zijn nogal wat vragenlijsten die (misschien onterecht) veel gebruik maken van meervoudige stellingvragen. Deze vragen worden vaak in een één keer uitgevraagd. Respondenten die geen zin meer in de vragenlijst hebben, of de vragen te ingewikkeld vinden, of bijvoorbeeld de vragen niet zo van toepassing vinden, geven soms ongenueanceerde en eenzijdige antwoorden. Bijvoorbeeld bij een stellingvraag 'wat vindt u van de winkel' met een vijfpuntsschaal als antwoordcategorie wordt dan overall maar uit het oogpunt van gemak 'tevreden' ingevuld. Dergelijke respons valt niet onder 'kwalitatieve goede respons'. Deze analyse leidt ook weer tot een hercodering van twee groepen aan de hand van de spreiding van de antwoorden (zie kader 8.1). De respondenten die dan juist weinig spreiding vertonen in hun antwoorden kunnen eventueel uitgesloten worden voor de vervolganalyses. Ook deze analyse heeft mogelijk tot gevolg dat de hoeveelheid bruikbare respons verkleind wordt.



## Martien Schriemer over **STATISTIEK** voor de beroepspraktijk

**“Kun je lezen? Kun je tellen?  
Dan kun je ook statistiek, iedereen kan het leren!”**

**“Statistiek is een doe-vak”**

**“Alles is toeval”**

**“Statistiek kan niet zonder verzamelingen”**

**“Spreading is de basis van statistiek”**



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

ISBN 978-90-826322-0-0  
Uitgever SVW te Haarlem



9 789082 632200 >