
Toegepaste Statistiek Inductieve technieken

Arie van Peet
Kirsten Namesnik
Joop Hox

Derde druk



Noordhoff Uitgevers

Toegepaste statistiek

Inductieve technieken

Arie van Peet

Kirsten Namesnik

Joop Hox

Derde druk

Noordhoff Uitgevers Groningen/Houten

Ontwerp omslag: G2K, Groningen/Amsterdam

Omslagillustratie: iStockphoto

Eventuele op- en aanmerkingen over deze of andere uitgaven kunt u richten aan:
Noordhoff Uitgevers bv, Afdeling Hoger Onderwijs, Antwoordnummer 13, 9700 VB
Groningen, e-mail: info@noordhoff.nl

Deze uitgave is gedrukt op FSC-papier.

0 / 12

© 2012 Noordhoff Uitgevers bv Groningen/Houten, The Netherlands.

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 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 16h Auteurswet 1912 dient men de daarvoor verschuldigde vergoedingen te voldoen aan 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 1912) kan men zich wenden tot Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, postbus 3060, 2130 KB Hoofddorp, www.stichting-pro.nl).

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

ISBN (ebook) 978-90-01-84705-0

ISBN 978-90-01-81327-7

NUR 916

Woord vooraf bij de derde druk

Dit boek is het deel *Inductieve Technieken* uit de serie *Toegepaste Statistiek*. Het sluit aan op het deel *Beschrijvende Technieken*.

De delen *Beschrijvende technieken* en *Inductieve technieken* zijn gebaseerd op de jarenlange ervaring van de drie auteurs in het geven van statistiekcolleges aan studenten in de sociale wetenschappen. Vergeleken met de eerdere edities is één auteur vervangen: K.T. Namesnik heeft de plaats ingenomen van G.L.H. van den Wittenboer. Ook zij geeft onderwijs in statistiek en methoden aan studenten in de sociale wetenschappen.

Studenten in de sociale wetenschappen zijn doorgaans meer geïnteresseerd in de toepassing van statistiek dan in de mathematische achtergronden van de statistische technieken. We hebben met deze voorkeur zo veel mogelijk rekening gehouden. Daarbij gaan we ervan uit dat bij de analyse van sociaal-wetenschappelijke gegevens een computerprogramma wordt gebruikt. De nadruk ligt daardoor vooral op het begrijpen van de gedachtegang die aan de gepresenteerde technieken ten grondslag ligt, de selectie van een geschikte analysetechniek gegeven de vraagstelling en de voorliggende gegevens, en de correcte interpretatie en presentatie van de resultaten.

Voor de analyse van sociaalwetenschappelijke gegevens zijn verschillende statistische programma's beschikbaar. Bij de eerste druk van *Inductieve technieken* in 1995 werd nog gebruikgemaakt van SPSS/PC+. In deze derde druk zijn de voorbeelden gebaseerd op SPSS onder Windows, omdat dit in Nederland het meest gebruikte statistische pakket is. De versie waarop wij ons hier baseren is versie 20.0.

In de tekst wordt regelmatig aan het eind van een hoofdstuk aangegeven hoe een en ander met behulp van SPSS is te berekenen. De voorbeelden in de delen *Beschrijvende technieken* en *Inductieve technieken* zijn echter algemeen gehouden; de delen kunnen dus ook gebruikt worden door docenten die voor de analyses andere programma's zoals SAS of STATA gebruiken.

Nieuw is dat aan het eind van elk hoofdstuk een aantal opgaven worden gegeven die gemaakt kunnen worden om na te gaan of de betreffende stof is begrepen. Op de website www.inductievetechnieken.noordhoff.nl staan de opgaven die voorheen in de werkboeken waren opgenomen. Op deze zelfde website staat ook een uitgebreide begrippenlijst met een korte beschrijving van elk begrip en met vermelding in welk hoofdstuk het betreffende begrip wordt behandeld. Ook een lijst met alle formules uit *Beschrijvende technieken* kan men op deze website vinden. Voorts is een aantal onduidelijke passages verbeterd in deze nieuwe druk en zijn voorbeelden geactualiseerd.



Vergeleken met de tweede druk zijn er drie hoofdstukken toegevoegd. Eén nieuw hoofdstuk behandelt multiële regressieanalyse. In het deel *Beschrijvende technieken* hebben we al kort naar regressieanalyse gekeken. Hier behandelen we de toetsende aspecten van deze techniek. Het tweede nieuwe hoofdstuk is gewijd aan het gebruik van computerintensieve analysemethoden. Het zijn zogenoemde permutatietoetsen en bootstrapmethoden die in recente versies van SPSS beschikbaar zijn. Deze methoden zijn een alternatief voor de nonparametrische technieken in die gevallen dat aan de verdelingsaanname van de standaard toetsen niet voldaan is. Daarnaast is een hoofdstuk opgenomen dat een beknopt overzicht biedt van de multivariate technieken, dat zijn technieken die gebruikt worden om een aantal variabelen tegelijk te analyseren.

Hoewel het onderscheid tussen *Beschrijvende technieken* en *Inductieve technieken* een bepaalde volgorde van behandelen van de stof suggereert, is het geenszins onmogelijk om daarvan af te wijken. Elke docent zal hier een eigen keuze kunnen maken. Voor u ligt thans het deel *Inductieve technieken*. Dit deel begint met een inleiding waarin de opzet van het boek beschreven wordt en waarin kort de inhoud van elk hoofdstuk wordt aangegeven. Het deel *Inductieve technieken* bevat de gebruikelijke technieken voor de beschrijving van cijfermatige gegevens.

Bij de herziening hebben wij veel profijt gehad van kritische commentaren van gebruikers. Ook bij deze herziene druk houden wij ons aanbevolen voor op- en aanmerkingen.

A.A.J. van Peet
K.T. Namesnik
J.J. Hox
Utrecht, voorjaar 2012

Inhoud

Inleiding 9

1 Sociaalwetenschappelijk onderzoek 13

- 1.1 Causale conclusie en generalisatie 15
- 1.2 Interne validiteit 15
- 1.3 Externe validiteit 18
- Samenvatting 20
- Opgaven 21

2 Statistische generalisatie: steekproeven en populaties 25

- 2.1 Kernprobleem van de inductieve statistiek 27
- 2.2 Populaties 28
- 2.3 Steekproeven 32
- 2.4 Manieren om steekproeven te trekken 37
- Samenvatting 43
- Opgaven 44

3 Inleiding toetsende statistiek 47

- 3.1 Schatten 49
- 3.2 De steekproevenverdeling van het gemiddelde 49
- 3.3 Het toetsen van hypothesen 57
- 3.4 Mogelijke fouten 62
- 3.5 Hypothesen met en zonder richting 64
- Samenvatting 70
- Opgaven 71

4 Betrouwbaarheidsintervallen en toetsende statistiek 75

- 4.1 Toetsen met kritieke waarden 77
- 4.2 Het verband tussen de p -waarde en een kritieke waarde 82
- 4.3 Betrouwbaarheidsintervallen 84
- 4.4 Het verband tussen betrouwbaarheidsintervallen en hypothesetoetsen 88
- 4.5 Afrondende opmerkingen 90
- Samenvatting 92
- Opgaven 93

5 Parametrische toetsen voor één populatie 97

- 5.1 De t -toets voor het gemiddelde 99
 - 5.2 Toets voor gepaarde waarnemingen 108
 - 5.3 Toets voor de correlatie 115
 - 5.4 Toets voor een proportie 118
 - 5.5 Afrondende opmerkingen 123
- [Samenvatting 124](#)
[Opgaven 125](#)

6 Parametrische toetsen voor twee populaties 129

- 6.1 Toets voor het verschil tussen twee gemiddelden 131
 - 6.2 Toets voor het verschil tussen twee proporties 140
 - 6.3 Toets voor twee varianties 145
 - 6.4 Afrondende opmerkingen 150
- [Samenvatting 152](#)
[Opgaven 153](#)

7 Het onderscheidingsvermogen van een statistische toets 159

- 7.1 Het onderscheidingsvermogen 161
 - 7.2 Door welke factoren wordt het onderscheidingsvermogen beïnvloed? 163
 - 7.3 Onderscheidingsvermogen van een aantal specifieke toetsen 165
 - 7.4 Afrondende opmerkingen 173
- [Samenvatting 174](#)
[Opgaven 175](#)

8 Nonparametrische toetsen 179

- 8.1 Inleiding 181
 - 8.2 Binomiaaltoets voor proporties 182
 - 8.3 Chikwadraattoets voor frequenties van één nominale variabele 185
 - 8.4 De chikwadraattoets voor het verband tussen twee nominale variabelen 194
 - 8.5 Rangcorrelatie van Spearman voor het verband tussen twee ordinale variabelen 202
- [Samenvatting 210](#)
[Opgaven 211](#)

9 Computerintensieve technieken 215

- 9.1 Permutatietoetsen 217
 - 9.2 Voorbeelden van permutatietoetsen 218
 - 9.3 Bootstrappen 224
 - 9.4 Voorbeelden van bootstrappen 226
 - 9.5 Afrondende opmerkingen 231
- [Samenvatting 233](#)
[Opgaven 234](#)

10 Variantieanalyse 237

- 10.1 Eenwegs-variantieanalyse 239
- 10.2 Tweewegs-variantieanalyse 252
- 10.3 Veronderstellingen 268
 - Samenvatting 271
 - Opgaven 272

11 Meerdere regressieanalyse 277

- 11.1 Meerdere regressieanalyse en de meerdere correlatie 279
- 11.2 Toevoeging van predictoren en selecteren van predictoren 284
- 11.3 Regressiegewichten en toetsing van regressiegewichten 287
- 11.4 Aannamen meerdere regressieanalyse 291
- 11.5 Voorbeeld regressieanalyse met behulp van SPSS 294
- 11.6 Afrondende opmerkingen 300
 - Samenvatting 301
 - Opgaven 302

12 Overzicht multivariate analysetechnieken 307

- 12.1 Variantieanalyse 309
- 12.2 Regressieanalyse 312
- 12.3 Verschillende toepassingen gebaseerd op regressieanalyse 313
- 12.4 Complexere modellen 316
- 12.5 Afrondende opmerkingen 324

Antwoorden opgaven 325

Literatuuroverzicht 333

Illustratieverantwoording 335

Bijlage 1: Rechteroverschrijdingskansen in de standaard normale verdeling 336

Bijlage 2: Rechter kritieke waarden van de t -verdeling 339

Bijlage 3: Rechter kritieke waarden van de F -verdeling 342

Bijlage 4: Rechter kritieke waarden van de χ^2 -verdeling 345

Bijlage 5: Kritieke waarden van de rangcorrelatie van Spearman 346

Register 347

Inleiding

Net als *Beschrijvende technieken* is dit boek geschreven om duidelijk het verband vast te houden met onderzoek waarin de computer een belangrijke rol vervult. Dat heeft consequenties voor de opzet. We houden in dit boek nadrukkelijk rekening met het feit dat berekeningen met de computer worden uitgevoerd. Er worden kleine voorbeelden gegeven met handmatige berekeningen, die vooral de bedoeling hebben duidelijk te maken wat de logica van een specifieke techniek is. Tot in specifieke details weten hoe iets moet worden uitgerekend, is minder belangrijk geworden.

Handmatige berekeningen aan de hand van tabellen vinden niet meer plaats. De computer rekent statistische grootheden niet uit aan de hand van tabellen, maar aan de hand van de ingevoerde ruwe gegevens. De computer geeft in zulke gevallen de correcte waarde. Bovendien rekent de computer doorgaans in meer decimalen nauwkeurig dan een onderzoeker met pen en papier. Dit wil natuurlijk niet zeggen dat we de computer dus maar blindelings moeten vertrouwen. Als we de computer geen deugdelijke dingen laten doen (en dat komt meer voor dan men denkt) krijgen we ook geen deugdelijke uitkomsten.

Moderne statistiekprogramma's bevatten meer mogelijkheden dan in de gebruikelijke leerboeken over statistiek wordt behandeld. Dit brengt het risico met zich mee dat onderzoekers de computer allerlei berekeningen kunnen laten uitvoeren zonder dat zij precies weten waar die voor dienen. Dat betekent dat het des te belangrijker wordt om een breed overzicht te krijgen van de beschikbare statistische technieken.

Hoewel we uitgaan van het gebruik van een computer, blijft dit boek niettemin gelden als een leerboek voor inductieve statistiek. Het is geen handleiding voor de computer, of voor het gebruik van SPSS. Het boek is opgezet als een leerboek. Aan de hand van de definitieformules voor de behandelde technieken voeren wij berekeningen uit op kleine voorbeelden, al is het soms alleen maar om te laten zien waar verschillen bestaan tussen het handmatige rekenwerk en de computer. Dat betekent ook dat alle passages, inclusief de voorbeelden, nauwkeurig gelezen moeten worden en bij voorkeur moeten worden nagerekend, want de gehele tekst bevat informatie die voor een beter begrip van belang kan zijn.

Een aantal keren beschrijven wij wat we in SPSS moeten doen om een bepaald resultaat te krijgen. We baseren ons in deze editie op SPSS 20.0 voor Windows. Deze passages zijn aan het eind van elk hoofdstuk geplaatst. Om na te gaan of de stof van een bepaald hoofdstuk begrepen is, wordt aan het eind



van elk hoofdstuk doorgaans een tiental opgaven gegeven. De antwoorden staan achter in het boek. Op de website www.inductievetechnieken.noordhoff.nl worden nog veel meer opgaven gegeven en zijn ook de uitwerkingen te vinden.

De opbouw van het boek is als volgt. Elk hoofdstuk begint met een onderzoeksvoorbeeld waarvan de onderzoeksgegevens gebruikt worden om de stof in dat hoofdstuk toe te lichten. Het boek begint met een hoofdstuk over *sociaalwetenschappelijk onderzoek* waarin wordt aangegeven wat het verschil tussen interne en externe validiteit is en wanneer wij spreken van statistische generalisatie. Daarna volgt een hoofdstuk over *steekproeven en populatie* (hoofdstuk 2). Hoofdstuk 3 gaat over de *logica* van statistische toetsing van hypothesen, en hoe we kunnen vaststellen of een uitspraak over de populatie (de hypothese) houdbaar is gezien de resultaten uit de steekproef. In hoofdstuk 4 wordt een ander basisprincipe van de inductieve statistiek behandeld, het berekenen van *betrouwbaarheidsintervallen*. De hoofdstukken 3 en 4 zijn de kern van het boek; voor een goed begrip van alle volgende hoofdstukken is het begrijpen van de stof in deze hoofdstukken absoluut noodzakelijk. De hoofdstukken 5 tot en met 11 behandelen een grote variëteit aan *statistische technieken*. In hoofdstuk 5 worden parametrische toetsen behandeld voor één populatie, zoals de toets voor één gemiddelde en de toets voor een correlatie. Hoofdstuk 6 behandelt de parametrische toetsen voor twee populaties, zoals de toets voor het verschil tussen twee gemiddelden. Hoofdstuk 7 gaat in op het onderscheidingsvermogen van de statistische toets; het gaat hier onder meer om de vraag hoe groot de steekproef moet zijn om een bestaand verschil ook aan te kunnen tonen. Hoofdstuk 8 bespreekt een aantal nonparametrische toetsen, toetsen die gebruikt kunnen worden wanneer niet aan de verdeelingsaanname van de parametrische toetsen is voldaan. Hoofdstuk 9 bespreekt een aantal moderne technieken die de rekenkracht van de computer gebruiken om parametrische toetsen toch te kunnen gebruiken in gevallen dat er niet aan de verdeelingsaanname is voldaan. Hoofdstuk 10 bespreekt variantieanalyse, en hoofdstuk 11 multiële regressieanalyse. Deze twee technieken behoren tot de multivariate technieken; technieken waarmee de invloed van twee of meer onafhankelijke variabelen tegelijk wordt geanalyseerd. Er zijn heel veel multivariate technieken, en hoofdstuk 12 geeft daarvan een beknopt overzicht.

De behandeling van deze statistische technieken geschiedt doorgaans volgens een vaste volgorde:

Toetsingssituatie In welk soort situaties wordt de statistische toets toegepast?

Voorwaarden Aan welke voorwaarden moet voldaan zijn (welke aannamen moeten we doen), om de toets te mogen toepassen?

Hypothesen Welke hypothesen kunnen we toetsen met behulp van de statistische techniek?

Toetsingsgrootheid Wat is de toetsingsgrootheid en hoe is deze verdeeld?

Beslissingsregel Hoe luidt de regel volgens welke we de nulhypothese al dan niet kunnen verwerpen? Deze regel kan worden geformuleerd met behulp van kritieke waarden of met behulp van overschrijdingskansen. Computerprogramma's berekenen doorgaans de exacte overschrijdingskansen.

Betrouwbaarheidsinterval Van een aantal statistische toetsen wordt de formule voor het betrouwbaarheidsinterval gegeven.

Voorbeeld Het gebruik van de statistische toets wordt geïllustreerd met behulp van een voorbeeld.

SPSS en het voorbeeld Hetzelfde voorbeeld wordt uitgevoerd met behulp van SPSS (indien de toets in SPSS is opgenomen) en de uitvoer wordt kort besproken.

Afrondende opmerkingen (alleen indien nodig).



1

Sociaalwetenschappelijk onderzoek

- 1.1 Causale conclusie en generalisatie
 - 1.2 Interne validiteit
 - 1.3 Externe validiteit
- Samenvatting
Opgaven

Het doel van veel onderzoek is om op basis van verzamelde gegevens van een steekproef uitspraken te doen over de gehele populatie. In dit hoofdstuk gaan we dieper in op twee belangrijke vragen. Ten eerste bekijken we de vraag of het gekozen onderzoeksontwerp ons in staat stelt om causale conclusies te trekken. Ten tweede overwegen we de vraag in hoeverre we de resultaten die in de steekproef gevonden zijn, terug kunnen vinden in de populatie.

Interne validiteit 15

Externe validiteit 15

Causaal verband 16

Voormeting 17

Storende variabelen 17

Verklarende variabele 17

Spurieuze correlatie 17

Generaliseerbaarheid over situaties 18

Generaliseerbaarheid over methoden 18

Generaliseerbaarheid over perioden 18

Statistische generalisatie 19

Theoretisch universum 19

Zelf lezen en voorgelezen worden



Samen met kinderen in boeken kijken en voorlezen betekent niet alleen plezier voor kinderen en volwassenen maar is ook nog eens een goede stimulans voor de taal- en leesontwikkeling van het kind. In een (fictief) onderzoek worden gegevens verzameld van 20 kinderen van 6 jaar die dagelijks door hun

ouders worden voorgelezen (= groep 1) en van 20 kinderen van eveneens 6 jaar die zelden of nooit door hun ouders worden voorgelezen (= groep 2). Na een halfjaar voorlezen wordt bij alle 40 kinderen een leestoets afgenomen waarvan de scores lopen van 1 tot en met 10. De scores van beide groepen staan in tabel 1.1.

TABEL 1.1 Scores op leestoets van twee groepen kinderen (groep 1: wel voorgelezen, groep 2: niet voorgelezen)

Groep	Score	Groep	Score	Groep	Score	Groep	Score
1	7	1	5	2	7	2	6
1	4	1	8	2	4	2	3
1	9	1	5	2	7	2	7
1	7	1	9	2	8	2	6
1	6	1	8	2	5	2	4
1	9	1	7	2	7	2	7
1	7	1	9	2	7	2	5
1	5	1	10	2	4	2	8
1	8	1	7	2	8	2	5
1	4	1	5	2	4	2	8

1.1 Causale conclusie en generalisatie

Statistische technieken kunnen globaal onderverdeeld worden in beschrijvende en generaliserende statistiek.

Bij *beschrijvende statistiek* gaat het om het beschrijven van de voorliggende gegevens. Bijvoorbeeld, stel dat we bij de uitgang van een bioscoop aan 20 bioscoopgangers hebben gevraagd wat ze van de film vonden, uitgedrukt in een schaal die loopt van 0 (vreselijk) tot 10 (fantastisch). Beschrijvende statistiek richt zich dan op vragen als: 'Wat is door de bank genomen het oordeel over de film?' of 'Lopen de oordelen erg uiteen of zijn ze vrijwel hetzelfde?'. Zulke vragen kunnen in principe beantwoord worden door van de 20 oordelen het gemiddelde en de standaardafwijking te berekenen.

De *inductieve statistiek*, ook wel de *generaliserende statistiek* genoemd, richt zich op de vraag of we zulke resultaten, berekend uit de gegevens van 20 bioscoopgangers, mogen generaliseren naar iedereen die de film gezien heeft. Voor de beantwoording van de vraag in hoeverre we gegevens die in een steekproef zijn verzameld, mogen generaliseren naar de algemene populatie zijn allerlei technieken beschikbaar. In de komende hoofdstukken van dit boek wordt de redenering achter het generaliseren van statistische resultaten behandeld, en wordt een aantal belangrijke technieken gedetailleerd uitgelegd.

Het doel van veel sociaalwetenschappelijk onderzoek is om op basis van verzamelde steekproefgegevens conclusies te trekken over een hypothese. Voor het voorbeeld in de casus luidt de hypothese als volgt: Kinderen die vaak door hun ouders worden voorgelezen, leren beter lezen dan kinderen die niet worden voorgelezen.

We berekenen de gemiddelden van de twee groepen kinderen in het voorbeeld. De kinderen die veel voorgelezen worden halen gemiddeld een 7, kinderen die niet worden voorgelezen een 6. Mogen we nu concluderen dat voorlezen helpt en dat dit geldt voor alle kinderen in de betreffende leeftijd?

We hebben hier te maken met twee verschillende vragen. De *eerste vraag* is of het verschil tussen de 7 en de 6 veroorzaakt wordt door het feit dat aan de ene groep kinderen veel werd voorgelezen en aan de andere groep niet of nauwelijks, en niet door iets anders. Dit is de vraag naar de *interne validiteit* van het onderzoeksontwerp. Die vraag beantwoorden we in paragraaf 1.2. De *tweede vraag* is of voor alle kinderen geldt dat kinderen die voorgelezen worden een hogere leesscore halen dan kinderen die niet worden voorgelezen. Dit is een vraag naar de *externe validiteit*, die in paragraaf 1.3 aan de orde wordt gesteld.

Interne
validiteit

Externe
validiteit

1.2 Interne validiteit

Interne validiteit kunnen we als volgt omschrijven:

Interne validiteit is de mate waarin een onderzoeksontwerp ons in staat stelt causale conclusies te trekken over het effect van een specifieke onafhankelijke variabele op een afhankelijke variabele.

Waarschijnlijk willen we graag de conclusie trekken dat voorlezen de kinderen helpt om zelf ook beter te leren lezen. We geven aan onze resultaten daarmee een causale verklaring: voorlezen wordt als oorzaak gezien voor het beter leren lezen van kinderen. Om te kunnen concluderen dat tussen voorlezen en zelf lezen een *causaal verband* bestaat, moet aan drie eisen worden voldaan:

- 1 dat wel-of-niet voorgelezen worden op de voorspelde wijze samengaat met zelf goed of slecht lezen;
- 2 dat het verschil tussen goed en slecht lezen in de tijd na het voorlezen ontstaat; en
- 3 dat er behalve het voorlezen geen andere invloeden op het lezen een rol hebben gespeeld.

Veronderstel dat we in ons onderzoek (1) een verschil hebben gevonden in de voorspelde richting. Dan kunnen (2) en (3) nog allerlei problemen geven. Punt (2) eist simpelweg dat de oorzaak (het voorlezen) in de tijd vooraf moet gaan aan het gevolg (het beter lezen). Daar komen we in onze onderzoeksopzet niets over te weten. We weten om te beginnen niet of het verschil is ontstaan na het voorlezen. Misschien waren de kinderen die veel voorgelezen werden al beter in lezen voordat de ouders begonnen met voorlezen. In dat geval bestonden de verschillen al voordat het voorlezen begon. Of misschien vroegen de kinderen die beter zijn in lezen wel vaker aan hun ouders of die hen wilden voorlezen. In dat geval is de verklaring zelfs precies omgekeerd: het kunnen lezen veroorzaakt het voorgelezen worden!

Punt (3) eist dat er geen andere verklaringen mogen zijn. Ook daar kunnen problemen bij opduiken. Misschien is het wel zo dat slimmere ouders ook slimmere kinderen krijgen én meer voorlezen. Dan is het verschil in lezen een gevolg van intelligentie. Of misschien hebben rijke ouders meer tijd om hun kinderen voor te lezen, en gaan kinderen van rijke ouders bovendien naar betere scholen. Dan is het verschil in lezen een gevolg van inkomensverschillen. Bovendien sluiten deze alternatieve verklaringen elkaar niet uit; misschien zijn ze allebei waar en zijn slimme ouders bovendien vaker rijk! Hoewel de voorgelezen kinderen hogere leesscores behaald hebben dan de kinderen die niet werden voorgelezen, kunnen we op grond van het voorbeeldonderzoek niet concluderen dat voorlezen helpt.

Het grootste probleem van dit onderzoeksontwerp is dat we achteraf proberen te onderzoeken of voorlezen geholpen heeft. Beter zouden we misschien een experiment kunnen opzetten waarbij we beginnen met 40 kinderen die nog niet kunnen lezen. Bij 20 kinderen spreken we met de ouders af dat die heel veel voorlezen. Bij 20 andere kinderen spreken we met de ouders af dat ze de kinderen niet voorlezen. We bepalen het wel-of-niet voorgelezen worden door puur loten. Na een half jaar evalueren we de leesprestaties van de kinderen. Als we dán een gemiddelde score van 7 vinden voor de ene groep en 6 voor de andere, staan we veel sterker in onze conclusie.

Het voorbeeld maakt eveneens duidelijk waarom desondanks niet altijd gekozen wordt voor een 'echt' experiment, hoewel dat sterke conclusies mogelijk maakt. Wanneer voorlezen inderdaad helpt, dan is ons experiment ethisch niet verantwoord. De 20 kinderen die niet worden voorgelezen, worden immers in hun ontwikkeling beperkt! Daarnaast is het niet

waarschijnlijk dat ouders zomaar bereid zijn om aan zo'n experiment mee te doen. In de praktijk worden daarom dikwijls onderzoeksopzetten gebruikt die niet volledig voldoen aan de eisen van een echt experiment, maar waarin door aanvullend onderzoek wordt geprobeerd alternatieve verklaringen voor de resultaten uit te sluiten.

We kunnen het onderzoek in veel gevallen beginnen met een voormeting. We kunnen in het leesonderzoek bijvoorbeeld beginnen met een groep kinderen waarvan we op basis van een voormeting weten dat ze niet of nauwelijks van elkaar verschillen wat betreft leesprestatie. Wanneer we nu later vinden dat kinderen die veel worden voorgelezen beter leren lezen dan de andere kinderen, dan kunnen we in ieder geval stellen dat het verschil niet al vooraf bestond. En als we denken dat onze conclusies 'verstoord' kunnen worden door verschillen in intelligentie of het inkomen van ouders, kunnen we deze twee *storende variabelen* proberen mee te nemen in ons onderzoek. Dat is niet ideaal, maar het is aanzienlijk beter dan niets.

Voormeting

In veel methodologieboeken worden onderzoeksopzetten besproken die ons in staat stellen om toch zo sterk mogelijke conclusies te trekken in situaties waarin experimenteren niet mogelijk is (zie: Boeije e.a., 2009; Hoyle e.a., 2002; Landsheer, 2003; Segers, 2002). Verder is het mogelijk om met speciale analysetechnieken (zie Pedhazur, 1997) na te gaan of mogelijke storende variabelen die we in ons onderzoek hebben meegenomen, van invloed zijn op de resultaten. We kunnen bijvoorbeeld nagaan of het statistische verband tussen voorlezen en de leesprestatie verdwijnt wanneer we rekening houden met de intelligentie van de ouders. Als dat zo is, dan is intelligentie de 'echte' verklarende variabele, en niet het 'al dan niet voorlezen'. De invloed van zo'n derde variabele wordt fraai geïllustreerd in het volgende voorbeeld. Het heeft betrekking op het verband tussen de schade door brand en het aantal betrokken brandweermannen. Hoe meer brandweermannen bij een brand aanwezig zijn, des te groter is doorgaans de brandschade. Wordt de brandschade dus door de brandweer zelf veroorzaakt? Het antwoord is dat in dit voorbeeld een belangrijke derde variabele is weggelaten, namelijk de grootte van de brand! Bij een grote brand zijn meer brandweermannen nodig, en ontstaat er desondanks een grote schade. Het verband tussen het aantal brandweermannen en de grootte van de schade is dus een schijnverband; het wijst niet op een oorzaak-gevolg (causale) relatie. In de methodologie wordt voor zo'n schijnverband wel de term *spurieuze correlatie* gebruikt.

Storende variabelen**Verklarende variabele****Spurieuze correlatie**

De conclusie van dit betoog is niet dat alleen echte experimenten 'goed' zijn, en evenmin dat 'goed' kwantitatief onderzoek in de sociale wetenschappen niet mogelijk is. De conclusie is wél dat geprobeerd moet worden het onderzoek zo op te zetten dat alternatieve verklaringen voor de verwachte resultaten zo veel mogelijk kunnen worden uitgesloten. Dat is in eerste instantie geen statistisch probleem. Het houdt in dat een zo goed mogelijk onderzoeksontwerp moet worden gekozen, en dat het zin heeft daarvoor methodologieboeken te raadplegen. Het houdt ook in dat mogelijke storende variabelen opgespoord moeten worden vóórdat het onderzoek van start gaat. Wederom is dat geen statistisch, maar een inhoudelijk probleem. Op basis van de theorie, aangevuld met resultaten van eerder onderzoek, moeten onderzoekers bij zichzelf nagaan welke storende variabelen er zouden kunnen zijn, en hoe ze daar in het onderzoek mee om zullen gaan.

1.3 Externe validiteit

Externe validiteit kunnen we als volgt omschrijven:

Externe validiteit is de mate waarin wij de resultaten van het onderzoek kunnen generaliseren over *situaties, methoden, tijd en populaties* (Hoyle e.a., 2002).

Vaak wordt met ‘generaliseren’ statistische generalisering bedoeld: in hoeverre zijn de steekproefresultaten geldig voor de populatie? Maar dat is dus niet de enige vorm van generaliseren zoals uit de omschrijving van externe validiteit naar voren komt. We behandelen nu eerst in het kort generaliseren over ‘situaties, methoden en tijd’ om daarna in te gaan op het generaliseren over populaties.

Generaliseerbaarheid over situaties

Generaliseerbaarheid over situaties heeft betrekking op de vraag in hoeverre een onderzoekssituatie en de dagelijkse situatie op elkaar lijken. Wanneer ouders hun kinderen in de onderzoekssituatie voorlezen in rustige kamertjes, terwijl het voorlezen in de dagelijkse werkelijkheid plaatsvindt wanneer andere kinderen eromheen spelen, dan is het de vraag in hoeverre de resultaten gegeneraliseerd kunnen worden naar zulke situaties. Dus ook het overdragen van conclusies naar mogelijke andere situaties is een vorm van generaliseren. Dit is echter geen *statistische generalisatie*.

Generaliseerbaarheid over methoden

Generaliseerbaarheid over methoden gaat over de vraag of dezelfde resultaten gevonden zouden worden wanneer het verschijnsel op een andere manier zou zijn onderzocht.

In het voorleesonderzoek is ‘beter lezen’ gemeten met behulp van een leestoets waarbij de kinderen hardop iets moesten voorlezen. Voor dat voorlezen kregen zij een score. Het ‘beter lezen’ zou echter ook gemeten kunnen worden door de kinderen een toetsje af te nemen met een aantal vragen over een zojuist gelezen passage. De ene manier om ‘beter lezen’ te meten is niet per definitie beter dan de andere, maar het is zeer wel denkbaar dat de twee verschillende methoden verschillende resultaten te zien zullen geven. Generaliseerbaarheid over methoden is wel degelijk generalisatie, maar het is niet in de eerste plaats wat we onder *statistische generalisatie* verstaan.

Generaliseerbaarheid over perioden

Generaliseerbaarheid over perioden heeft betrekking op de vraag in hoeverre de steekproefresultaten ook geldig zijn voor een andere periode. In het voorbeeld van het voorleesonderzoek maakt de leeftijd van de kinderen waarop het onderzoek wordt uitgevoerd nogal verschil uit. Zouden we dezelfde resultaten hebben gevonden als we in plaats van de kinderen uit groep 3 kinderen uit groep 1 in het onderzoek gebruikt hadden? De vraag of de conclusies ook voor andere perioden gelden, is wel degelijk een generaliseerbaarheidsvraag, maar het is niet een *statistische generaliseerbaarheidsvraag*.

Wanneer het gaat om de vraag of de resultaten in de steekproef van toepassing zijn op een *andere* populatie dan die waaruit onze steekproef is getrokken, dan is dat nog steeds geen statistische generalisatie. Wanneer onze conclusies op basis van steekproefresultaten betrekking hebben op kinderen en we vragen ons af of zij ook geldig zijn voor ouderen dan is ook dat geen statistische generalisatie, maar een kwestie van beargumenteren

en van verder empirisch onderzoek of die generalisatie naar andere populaties geoorloofd is.

De term *statistische generalisatie* kunnen we als volgt definiëren:

We zijn bezig met statistische generalisatie wanneer we ons afvragen in hoeverre de resultaten die wij in ons onderzoek vonden, ook gelden voor de populatie waaruit onze aselechte steekproef is getrokken.

**Statistische
generalisatie**

1

Bij statistische generalisatie gaat het om de volgende vragen:

- Als we in de *steekproef* een bepaald verschil tussen twee gemiddelden vinden, hoe groot is dan de kans dat dit verschil in de *populatie* ook echt bestaat?
- Als er in de *populatie* een bepaald verschil tussen twee gemiddelden is, hoe groot is dan de kans dat we dit verschil in de *steekproef* terugvinden?

Het is belangrijk om op te merken dat *alleen wanneer we over aselechte steekproeven uit een populatie beschikken, we de statistische generalisatie-technieken die in dit deel Inductieve technieken worden besproken, kunnen toepassen.*

Heel vaak echter weten we niet zeker of de steekproef wel aselechte is, of beter gezegd, heel vaak weten we zeker dat de steekproef niet aselechte is. Toch passen we in veel gevallen dan de generalisatietechnieken toe, voeren we toetsingen uit en berekenen we betrouwbaarheidsintervallen. Wat is dan de betekenis van dergelijke berekeningen en resultaten?

Soms doen we alsof de steekproef een aselechte steekproef is uit een *theoretisch universum* en dat de steekproefresultaten wél naar een dergelijke denkbeeldige populatie gegeneraliseerd kunnen worden. Wat in dat geval de betekenis is van zo'n generalisatie, is niet geheel duidelijk. De onderzoeker heeft dan niet alleen te beargumenteren waarom een dergelijk theoretisch universum zinvol is, maar tevens waarom de steekproef uit dat universum wél als een aselechte steekproef beschouwd mag worden.

**Theoretisch
universum**

Een verwante redenering is om de steekproef op te vatten als een aselechte steekproef uit een denkbeeldige populatie. We beschouwen dan bijvoorbeeld het verschil in de resultaten van de leestoets als een minimumvoorwaarde voor de vraag of een bepaalde grootte van het verschil de *moeite van het interpreteren waard* is. Ook hierbij dient heel wat te worden uitgelegd. De 'moeite van het interpreteren waard' betekent niets anders dan 'aanwezig zijn in de onderzochte populatie'!

Samenvatting

1

-
- ▶ De kwaliteit van een onderzoeksontwerp hangt af van het antwoord op twee belangrijke vragen, namelijk de vraag naar de interne validiteit en de vraag naar de externe validiteit.
 - ▶ *Interne validiteit* is de mate waarin het onderzoeksontwerp ons in staat stelt om causale conclusies te trekken over het effect van een onafhankelijke variabele op de afhankelijke variabele.
 - ▶ *Externe validiteit* is de mate waarin wij de resultaten van het onderzoek kunnen generaliseren over *situaties, methoden, tijd* en *populaties*.
 - ▶ Alleen wanneer we over een aselechte steekproef uit een populatie beschikken, kunnen we de statistische generalisatie-technieken die in dit boek worden besproken toepassen.
 - ▶ Het gaat dan steeds om de vraag in hoeverre de resultaten die wij in onze steekproef vonden, ook terug te vinden zijn in de populatie waaruit onze steekproef getrokken is.
-

Opgaven

- 1.1** Aan de ouders van kinderen in een peuterspeelzaal wordt een vragenlijst afgenomen waarin wordt gevraagd om op een schaal van 1 tot en met 10 aan te geven hoe tevreden zij zijn over de peuterspeelzaal. De gemiddelde tevredenheidsscore blijkt 6,38 te zijn. Dit is een voorbeeld van
- A** beschrijvende statistiek.
 - B** generaliserende statistiek.
 - C** toetsende statistiek.
- 1.2** Een onderzoekshypothese luidt dat kinderen die een peuterspeelzaal bezoeken sociaal vaardiger zullen zijn dan de kinderen die dat niet doen. Een pedagoge beoordeelt 20 aselekt gekozen driejarige kinderen die een peuterspeelzaal bezoeken op een sociale vaardigheidsschaal van 1 tot en met 7. Hetzelfde doet zij met 20 aselekt gekozen driejarige kinderen die géén peuterspeelzaal bezoeken. Na toetsing is de conclusie dat het bezoeken van een peuterspeelzaal de kinderen sociaal vaardiger maakt. Die conclusie is
- A** onjuist, omdat de beoordeling van de sociale vaardigheid slechts door één persoon geschiedde.
 - B** juist, omdat beide groepen aselekt gekozen zijn.
 - C** onjuist, omdat er vele alternatieve verklaringen mogelijk zijn voor het verschil in sociale vaardigheid.
- 1.3** Bepaal voor ieder van de volgende uitspraken of deze waar of niet waar is.
- a** Statistische generalisatie heeft vooral betrekking op de stap van begrip-zoals-bedoeld naar begrip-zoals-bepaald.
 - b** Statistische generalisatie heeft vooral betrekking op de stap van steekproef naar populatie.
- 1.4** Wat wordt bedoeld met 'storende' variabelen?
- 1.5** Om te kunnen concluderen dat er tussen het volgen van een niet-verplichte cursus statistiek en 'statistisch redeneren' een effect bestaat, moeten we de volgende twee punten kunnen aantonen:
- I** dat het al of niet volgen van de cursus op de voorspelde wijze samenhangt met goed en minder goed statistisch kunnen redeneren.
 - II** dat het verschil tussen goed en minder goed statistisch kunnen redeneren in de tijd is ontstaan na het volgen van de statistiekcursus.

Voor het aantonen van een causaal verband zijn deze twee punten:

- A** voldoende, want dan is het verband causaal.
- B** voldoende, maar het verband is slechts gedeeltelijk causaal.
- C** niet voldoende, want er is niet aan alle eisen voldaan.

- 1.6** Wanneer een onderzoeksontwerp een onderzoeker in staat stelt om duidelijke conclusies te trekken over het effect van de onafhankelijke variabele op de afhankelijke variabele, dan is er een hoge mate van
- A** begripsvaliditeit.
 - B** externe validiteit.
 - C** interne validiteit.
- 1.7** In een onderzoek blijkt dat er een positieve correlatie is tussen de variabelen *inkomen*, *intelligentie* en *opleiding*. De onderzoeker beargumenteert dat het verband tussen *inkomen* en *opleiding* volledig causaal is. Wat gebeurt er met de correlatie tussen *inkomen* en *opleiding* wanneer *intelligentie* constant gehouden wordt en zijn argumentatie juist is?
- 1.8** Bepaal voor ieder van de volgende uitspraken of deze waar of niet waar is.
- a** *Externe validiteit* is de mate waarin een test een gedrag buiten de testsituatie kan voorspellen.
 - b** *Interne validiteit* is de mate waarin twee verschillende methoden die gebruikt worden om eenzelfde onderwerp te onderzoeken, gelijke resultaten opleveren.
- 1.9** Wanneer we het hebben over ‘statistische generalisatie’ hebben we het vooral over de vraag in hoeverre de steekproefresultaten
- A** ook geldig zijn in een andere steekproef.
 - B** gelden voor de populatie waaruit de steekproef afkomstig is.
 - C** ook gelden voor een andere populatie dan die waaruit de steekproef getrokken is.
- 1.10** Wanneer gegeven is dat een steekproef een aselechte steekproef is uit een goed omschreven theoretisch universum, dan ...
- a** Waar of niet waar? ... kan de onderzoeker zonder meer de generalisatie-technieken gebruiken die in dit boek worden behandeld.
 - b** Waar of niet waar? ... moet de onderzoeker wel aangeven waarom dit theoretisch universum relevant is.