

Methoden en Statistiek 1

true

2 juli 2020

Contents

Voorwoord	5
 Deel I: Methodologie	 9
1 Inleiding	9
1.1 Wetenschappelijk onderzoek	9
1.2 Paradigmata	11
1.3 Instrumentatie-onderzoek	12
1.4 Beschrijvend onderzoek	13
1.5 Experimenteel onderzoek	14
1.6 Vooruitblik	17

Voorwoord

In dit boek hebben we gepoogd om de belangrijkste kwantitatieve methoden en statistische technieken die relevant zijn voor de Geesteswetenschappen uit te leggen. De tekst is waar mogelijk gevrijwaard van wiskundige afleidingen en formules, omdat we die voor studenten Geesteswetenschappen minder bruikbaar achten. Indien nodig geven we de belangrijkste formules voor statistische beschrijvingen en analyses in een aparte paragraaf.

Het tekstboek bevat ook aanwijzingen over hoe de besproken statistische analyses en visualisaties uitgevoerd kunnen worden in twee veelgebruikte programma's, nl. SPSS (versie 22 en later) en R (versie 3.0 en later). Ook deze aanwijzingen staan los van de hoofdtekst, in afzonderlijke paragrafen.

In aansluiting op het internationale gebruik en op de conventies van Engelstalige tijdschriften gebruiken we de punt als decimaalteken; we schrijven dus $\frac{3}{2} = 1.5$. Hierbij is een waarschuwing op zijn plaats: het decimale symbool kan verschillen tussen computers, en zelfs tussen programma's op dezelfde computer. Controleer dus welk decimaal symbool gebruikt wordt door (elk programma op) jouw computer.

Graag willen we onze mede-docenten danken voor de vele discussies en voorbeelden die op enige wijze verwerkt zijn in dit tekstboek. Onze studenten danken we voor hun nieuwsgierigheid en nauwkeurigheid die geleid heeft tot deze versie van dit tekstboek.

Ook betonen wij grote dank aan Willemijn Heeren, Gerrit Bloothoofd, Marijn Struiksma, Margot van den Berg, Els Rose, Tobias Quené en Kirsten Schutter voor hun adviezen, data, en/of commentaar bij het manuscript van dit digitale tekstboek.

Utrecht, december 2016 - juli 2020

Hugo Quené, <https://www.hugoquene.nl>

Huub van den Bergh

This document is licensed under the *GNU GPL 3* license (for details see <https://www.gnu.org/licenses/gpl-3.0.en.html>). It was created with the **bookdown** package (Xie, 2020) in Rstudio.

Deel I: Methodologie

Chapter 1

Inleiding

In dit tekstboek worden de grondbeginselen, methoden en technieken van empirisch wetenschappelijk onderzoek besproken, zowel in algemene zin als toegespitst op het brede domein van taal en communicatie. We zullen ons bezighouden met vragen als: Wat is een goede onderzoeksvraag? Welke methode is de beste om de onderzoeksvraag te beantwoorden? Hoe kunnen onderzoekers zinnige en valide conclusies trekken uit (statistische analyses van) hun gegevens? In dit tekstboek beperken we ons tot de belangrijkste grondbeginselen, en tot de belangrijkste methoden en technieken. In dit eerste hoofdstuk zullen we een overzicht geven van verschillende typen en vormen van wetenschappelijk onderzoek. In het vervolg van dit tekstboek geven we de meeste aandacht aan methoden van wetenschappelijk onderzoek waarbij empirische observaties uitgedrukt worden in de vorm van getallen (kwantitatief), die geanalyseerd worden met behulp van statistische technieken.

1.1 Wetenschappelijk onderzoek

Om te beginnen moeten we een vraag stellen die terugslaat op de allereerste zin hierboven: wat is eigenlijk wetenschappelijk onderzoek? Wat is het verschil tussen wetenschappelijk en niet-wetenschappelijk onderzoek (bijv. door onderzoeksjournalisten)? Onderzoek dat een wetenschapper uitvoert, hoeft nog geen wetenschappelijk onderzoek te zijn. Evenmin is journalistiek onderzoek per definitie onwetenschappelijk omdat het door een journalist wordt uitgevoerd. In dit tekstboek hanteren we de volgende definitie (Kerlinger and Lee, 2000, p.14):

“Scientific research is systematic, controlled, empirical, amoral, public, and critical investigation of natural phenomena. It is guided by theory and hypotheses about the presumed relations among such phenomena.”

Wetenschappelijk onderzoek is systematisch en gecontroleerd. Wetenschappelijk onderzoek is zodanig ontworpen dat we geloof kunnen hechten aan de conclusies, omdat die conclusies goed onderbouwd zijn. Het onderzoek kan door anderen herhaald worden, met (hopelijk) dezelfde resultaten. Deze eis van replicerbaarheid maakt ook dat wetenschappelijk onderzoek zeer nauwgezet wordt ontworpen en uitgevoerd (zie Hoofdstukken ?? en ??). De sterkste vorm van controle is die van een wetenschappelijk experiment; we besteden daarom in dit tekstboek veel aandacht aan experimenteel onderzoek (§1.5). Mogelijke alternatieve verklaringen voor het onderzochte verschijnsel worden één voor één onderzocht en zo mogelijk uitgesloten, zodat tenslotte slechts één verklaring overblijft (Kerlinger and Lee, 2000). Die verklaring vormt dan onze wetenschappelijk onderbouwde conclusie of theorie over het onderzochte verschijnsel.

Ook wordt in de definitie gesteld dat wetenschappelijk onderzoek *empirisch* van aard is. De conclusies die de onderzoeker trekt moeten uiteindelijk gebaseerd zijn op (systematische en gecontroleerde) waarnemingen of observaties van een verschijnsel in de werkelijkheid — bijvoorbeeld op de waargenomen inhoud van een tekst, of op het waargenomen gedrag van een proefpersoon. Als die waarneming ontbreekt, dan kunnen de eventuele conclusies niet logisch verbonden worden met de werkelijkheid, waardoor ze geen wetenschappelijke waarde hebben. Vertrouwelijke gegevens uit een onbekende bron, of inzichten verkregen in een droom of in een mystieke beleving, zijn niet empirisch onderbouwd, en kunnen dus niet de basis vormen van een wetenschappelijke theorie.

1.1.1 Theorie

Het doel van wetenschappelijk onderzoek is te komen tot een theorie over een deel van de werkelijkheid. Die theorie is te zien als een coherente en consistente verzameling van “justified true beliefs” (Morton, 2003). In deze overtuigingen, en in de theorie, wordt geabstraheerd van de complexe werkelijkheid van de natuurlijke verschijnselen, naar een abstract mentaal *construct*, dat uit zijn aard niet rechtstreeks waarneembaar is. Voorbeelden van dergelijke constructen zijn: leesvaardigheid, intelligentie, activatie-niveau, verstaanbaarheid, omvang van iemands actieve woordenschat, schoenmaat, woon-werk-afstand, introvertheid, etc.

Een onderzoeker definieert in een theorie niet alleen verschillende constructen, maar ook specificeert hij de *verbanden* of relaties tussen deze constructen. Pas wanneer zowel de constructen gedefinieerd zijn als de relaties tussen de constructen gespecificeerd zijn, kan een onderzoeker komen tot een systematische verklaring van het onderzochte verschijnsel. Deze verklaring of theorie kan weer de basis zijn van een *voorspelling* over het onderzochte verschijnsel: het aantal gesproken talen op de wereld zal verminderen in de 21e eeuw; teksten zonder voegwoorden zullen moeilijker te begrijpen zijn dan teksten met voegwoorden; kinderen die tweetalig opgroeien zullen niet slechter presteren op school dan eentalige kinderen.

Wetenschappelijk onderzoek is er in vele verschillende typen en vormen, die op verschillende manieren ingedeeld kunnen worden. In de volgende sectie 1.2 bespreken we een indeling op basis van paradigma, de manier waarop de onderzoeker tegen de werkelijkheid aankijkt. Onderzoek kan ook ingedeeld worden op een continuüm van ‘zuiver theoretisch’ naar ‘toegepast’. Een derde manier om onderzoek in te delen is gericht op het type onderzoek, bijvoorbeeld instrumentatieonderzoek (§1.3), beschrijvend onderzoek (§1.4), en experimenteel onderzoek (§1.5).

1.2 Paradigmata

Eén criterium om typen onderzoek te onderscheiden is op basis van het gebruikte paradigma, de manier waarop de onderzoeker tegen de werkelijkheid aankijkt. In dit tekstboek besteden we nagenoeg alleen aandacht aan het empirisch-analytisch paradigma, omdat dit het meest uitgewerkte en meest invloedrijke paradigma is. Heden ten dage kan deze benadering opgevat worden als ‘de’ standaardopvatting, waar andere paradigma’s zich min of meer tegen afzetten.

Binnen het *empirisch-analytische paradigma* onderscheiden we twee varianten: het positivisme en het kritisch-rationalisme. Beide stromingen hebben gemeen dat er aangenomen wordt dat er wetmatigheden zijn die ‘ontdekt’ kunnen worden: verschijnselen kunnen beschreven en verklaard worden in abstracte termen (constructen). Het verschil tussen beide stromingen binnen de empirisch-analytische traditie is gelegen in de pretentie van de uitspraken die gedaan worden. Volgens de positivisten is het mogelijk om uitspraken te doen vanuit feitelijke waarnemingen naar een theorie. Op basis van de observaties kunnen we generaliseren naar een algemeen geldende regel, door middel van inductie. (De vogels die ik zie, die hoor ik ook fluiten, dus alle vogels fluiten.)

De tweede stroming is het kritisch-rationalisme. De aanhangers van deze stroming keren zich tegen bovengenoemde inducties: al hoor ik talloze vogels ook fluiten, dan nog kan ik geen zekerheid verkrijgen over de veronderstelde algemene regel. Maar we kunnen het wel omkeren, en proberen aan te tonen dat de veronderstelde algemene regel of hypothese *niet* juist is. Hoe werkt dat? Op basis van de algemeen geldende regel kunnen we voorspellingen afleiden voor specifieke observaties, door middel van deductie. (Als alle vogels fluiten, dan moet het zo zijn dat alle vogels in mijn steekproef fluiten.) Als niet alle vogels in mijn steekproef fluiten, dan is de algemene regel blijkbaar onjuist. Dit wordt het falsificatie-principe genoemd; we bespreken dat uitgebreider in sectie ??.

Ook aan het kritisch-rationalisme kleven echter tenminste twee bezwaren. Met het falsificatieprincipe kunnen waarnemingen (empirische feiten, observaties, onderzoeksresultaten) gebruikt worden om theoretische uitspraken te doen (met betrekking tot hypothesen). Strikt genomen moet een veronderstelde algemene regel meteen verworpen worden na één geslaagde falsificatie (een van de vogels in mijn steekproef fluit niet): als theorie en observatie niet overeenstemmen,

dan faalt de theorie, volgens de kritisch-rationalisten. Maar om te komen tot een observatie moet een onderzoeker vele keuzes maken (bijv.: hoe maak ik een goede steekproef, wat is een vogel, hoe bepaal ik of een vogel fluit?), die de geldigheid van de observaties onzeker kunnen maken. Er kan dus ook iets mis zijn met de waarnemingen zelf (horen), of met de operationalisaties van de gebruikte constructen (vogels, fluiten).

Een tweede probleem is dat er in de praktijk eigenlijk zeer weinig theorieën zijn die werkelijk iets uitsluiten. Wanneer er discrepanties waargenomen worden tussen theorie en observaties, dan wordt de theorie bijgeschaafd, zodat de nieuwe observaties toch weer binnen de theorie passen. Theorieën worden dan ook zelden volledig verworpen.

Een tweede paradigma is de kritische benadering. Het *kritische paradigma* onderscheidt zich van andere paradigmata in de nadruk op maatschappelijke bepaaldheden; ‘de’ werkelijkheid bestaat niet, ons beeld ervan is een voorlopige, door maatschappelijke oorzaken bepaalde werkelijkheid. Inzicht in de maatschappelijke verhoudingen heeft zelf dus ook invloed op die werkelijkheid. Onze wetenschapsopvatting zoals verwoord in bovengenoemde definities van onderzoek en theorie wordt in het kritische paradigma dan ook afgewezen. Kritische onderzoekers menen dat onderzoeksprocessen niet los gezien kunnen worden van de maatschappelijke context waarin het onderzoek is verricht. Deze laatste visie wordt overigens overgenomen door steeds meer onderzoekers, ook door hen die andere paradigmata aanhangen.

1.3 Instrumentatie-onderzoek

Onderzoek is, zoals gezegd, een gesystematiseerde en gecontroleerde wijze om empirische gegevens te verzamelen en te interpreteren. Onderzoekers streven naar inzicht in natuurlijke verschijnselen, en in de wijze waarop (de constructen van) die verschijnselen met elkaar samenhangen. Een voorwaarde hiervoor is dat de onderzoeker deze verschijnselen daadwerkelijk kan meten, d.i. uitdrukken in een observatie (bij voorkeur in de vorm van een getal). Instrumentatieonderzoek is voornamelijk gericht op de constructie van instrumenten of methoden om verschijnselen, gedrag, vaardigheden, attitudes, etc. meetbaar te maken. De ontwikkeling van goede meetinstrumenten is bepaald geen sinecure: het is ambachtelijk handwerk, waarbij de constructeur vele valkuilen moet zien te vermijden. Het meetbaar maken van verschijnselen, van gedrag of van constructen noemen we de *operationalisatie*. Een concrete leestoets is bijvoorbeeld op te vatten als een operationalisatie van het abstracte construct ‘leesvaardigheid’.

We kunnen een nuttig onderscheid maken tussen het abstracte theoretische construct en het gemeten construct, ofwel een onderscheid tussen: het begrip-zoals-bedoeld en het begrip-zoals-bepaald. Het is uiteraard de bedoeling dat het begrip-zoals-bepaald (de toets, de vragenlijst, de observatie) het begrip-zoals-bedoeld (het theoretische construct) zo goed mogelijk benadert. Indien het

theoretische construct goed wordt benaderd, dan spreken we van een adequate of valide meting.

Bij operationalisatie van een begrip-zoals-bedoeld moeten talloze keuzen gemaakt worden. Zo moet het CITO (Centraal instituut voor toetsontwikkeling) elk jaar tekstbegripstoetsen construeren om de leesvaardigheid van eindexamenkandidaten te meten. Daarvoor moet allereerst een tekst gekozen of geredigeerd worden. Deze tekst mag niet te moeilijk, maar ook niet te makkelijk zijn voor de doelgroep. Voorts mag het onderwerp van de tekst niet al te bekend zijn, omdat anders de bij sommige leerlingen aanwezige algemene kennis kan interfereren met de meningen en standpunten die in de tekst naar voren gebracht worden. Vervolgens moeten de vragen zó ontworpen worden dat de verschillende passages in de tekst aan bod komen. Ook moeten de vragen zó samengesteld zijn dat het theoretische construct ‘leesvaardigheid’ adequaat geoperationaliseerd wordt. Tot slot moet ook nog rekening gehouden worden met de examens uit voorafgaande jaren; het nieuwe examen mag immers niet al te veel afwijken van oude examens.

Een construct moet dus op de juiste wijze geoperationaliseerd zijn, om observaties te verkrijgen die niet alleen valide zijn (een goede benadering van het abstracte construct, zie Hoofdstuk ??) maar die ook betrouwbaar zijn (ongeveer gelijke observaties bij herhaalde meting, zie Hoofdstuk ??). In ieder onderzoek zijn de validiteit en de betrouwbaarheid van een meting cruciaal; we besteden dan ook twee hoofdstukken aan deze begrippen. Maar in instrumentatieonderzoek zijn deze begrippen zelfs essentieel, omdat dit type onderzoek juist beoogt om valide en betrouwbare instrumenten te leveren, die een goede operationalisatie zijn van het abstracte construct-zoals-bedoeld.

1.4 Beschrijvend onderzoek

Met beschrijvend onderzoek bedoelen we onderzoek dat voornamelijk gericht is op de beschrijving van een bepaald natuurlijk verschijnsel in de werkelijkheid. De onderzoeker richt zich dus vooral op een *beschrijving* van het verschijnsel: het huidige vaardigheidsniveau, het verloop van een proces of een discussie, de wijze waarop de lessen Nederlands in het voortgezet onderwijs vorm worden gegeven, de politieke voorkeur van stemmers vlak voor verkiezingen, de samenhang tussen het aantal uren zelfstudie en het eindcijfer dat een student behaalt, etc. Kortom, ook de onderwerpen voor beschrijvend onderzoek kunnen zeer divers zijn.

Voorbeeld 1.1: (Dingemanse et al., 2013) hebben opnames van conversaties gekozen of gemaakt in 10 talen. Uit die opgenomen conversaties zijn woorden genomen waarmee een luisteraar om “open verduidelijking” vraagt: woordjes als *hè* (Nederlands), *huh* (Engels), *ã?*

(Siwu). Van deze woorden is de klankvorm en het toonhoogteverloop vastgesteld, met akoestische metingen en met fonetische transcriptions door experts. Een conclusie van dit beschrijvende onderzoek luidt dat deze tussenvoegsels in de verschillende talen veel meer op elkaar lijken (in klankvorm en toonhoogteverloop) dan op grond van toeval te verwachten is.

Dit voorbeeld illustreert dat beschrijvend onderzoek niet ophoudt als de gegevens (klankvormen, toonhoogteverloop) beschreven zijn. Vaak zijn *verbanden* tussen de verzamelde gegevens ook zeer interessant (zie §1.1). Zo wordt in opiniepeilingen naar het stemgedrag bij verkiezingen vaak een verband gelegd tussen het gepeilde stemgedrag enerzijds, en leeftijd, geslacht en opleidingsniveau van de respondent anderzijds. En evenzo wordt in onderwijskundig onderzoek een verband gelegd tussen aantal uren studietijd enerzijds, en studiesucces van de respondent anderzijds. Dit type van beschrijvend onderzoek, waarbij een correlatie wordt vastgesteld tussen mogelijke oorzaken en mogelijke gevolgen, wordt ook aangeduid als *correlationeel onderzoek*.

Het essentiële verschil tussen beschrijvend en experimenteel onderzoek is gelegen in de vraag naar oorzaak en gevolg. Op basis van beschrijvend onderzoek kan een causaal verband tussen oorzaak en gevolg *niet* goed vastgesteld worden. Uit beschrijvend onderzoek zou kunnen blijken dat er een samenhang is tussen een bepaald soort voeding en een langere levensduur. Is het voedingspatroon dan ook de oorzaak van de langere levensduur? Dat hoeft bepaald niet het geval te zijn: het is ook mogelijk dat dat soort voeding vooral genuttigd wordt door mensen die relatief hoog opgeleid en welvarend zijn, en door deze *andere* factoren ook relatief langer in leven zijn¹. Om vast te kunnen stellen of er een causaal verband is, moeten we experimenteel onderzoek opzetten en uitvoeren.

1.5 Experimenteel onderzoek

Experimenteel onderzoek wordt gekenmerkt doordat de onderzoeker een bepaald aspect van de onderzoeksomstandigheden systematisch varieert (Shadish et al., 2002). Het effect van deze manipulatie staat dan centraal in het onderzoek. Een onderzoeker vermoedt bijvoorbeeld dat een bepaalde nieuwe lesmethode zal resulteren in betere prestatie van de leerlingen dan de huidige lesmethode. De onderzoeker wil deze hypothese toetsen door middel van een experimenteel onderzoek. Hij manipuleert het type onderwijs: sommige klassen of groepen krijgen les volgens de nieuwe experimentele lesmethode en andere klassen of

¹Het is zelfs mogelijk dat het onderzochte voedingspatroon de oorzaak is van een relatief *kortere* levensduur, maar dat dit negatieve effect gemaskeerd wordt door de sterkere, positieve effecten van opleidingsniveau en welvaartsniveau op de levensduur.

groepen krijgen les op de traditionele wijze. Het effect van de nieuwe lesmethode wordt geëvalueerd door de prestaties van de twee soorten schoolklassen te vergelijken, na de ‘behandeling’ met de oude vs. nieuwe lesmethode.

Experimenteel onderzoek heeft als voordeel dat we de onderzoeksresultaten doorgaans mogen interpreteren als het gevolg van de experimentele manipulatie. Omdat de onderzoeker het onderzoek systematisch controleert en slechts één aspect (i.c. de lesmethode) varieert, kunnen eventuele verschillen tussen de prestaties van de twee categorieën alleen toegeschreven worden aan het veranderde kenmerk, i.c. aan de lesmethode. Dit veranderde kenmerk moet dan logischerwijs wel de oorzaak zijn van de geobserveerde verschillen. Experimenteel onderzoek is dus gericht op de evaluatie van causale verbanden.

Deze redenering vereist wel dat proefpersonen (of schoolklassen, in bovenstaand voorbeeld) volgens het toeval, aselekt (Eng. ‘at random’), worden toegewezen aan de experimentele condities (i.c. de oude of nieuwe lesmethode). Deze aselechte toewijzing (Eng. ‘random assignment’) is de beste methode om eventuele niet-relevante verschillen tussen de behandelcondities uit te sluiten. Een dergelijk experiment met aselechte toewijzing van proefpersonen aan condities wordt een *gerandomiseerd experiment* genoemd (Eng. ‘randomized experiment’, ‘true experiment’,). Om bij ons voorbeeld te blijven: als de onderzoeker de oude lesmethode zou inzetten bij jongens, en de nieuwe lesmethode bij meisjes, dan is een eventueel verschil in prestaties niet meer uitsluitend toe te schrijven aan het gemanipuleerde kenmerk (de lesmethode), maar ook aan een niet-gemanipuleerd maar wel relevant kenmerk, hier het geslacht van de leerlingen. Zo’n mogelijk verstrend kenmerk wordt een *storende variabele* (Eng. ‘confound’) genoemd. In Hoofdstuk ?? bespreken we hoe we deze storende variabelen kunnen neutraliseren, door random toewijzing van proefpersonen (of schoolklassen) aan de experimentele condities, in combinatie met andere maatregelen.

Er is ook experimenteel onderzoek waarbij een bepaald aspect (zoals lesmethode) wel systematisch varieert, maar waarbij de proefpersonen of schoolklassen *niet aselekt* zijn toegewezen aan de experimentele condities; dit wordt *quasi-experimenteel* onderzoek genoemd (Shadish et al., 2002). In het bovenstaande voorbeeld is daarvan sprake als de gebruikte lesmethode onderzocht wordt, met gegevens van schoolklassen waarvan niet de onderzoeker maar de docent bepaald heeft of de oude of nieuwe lesmethode gebruikt wordt. Als de nieuwe lesmethode betere prestaties zou opleveren, dan weten we *niet* met zekerheid dat het verschil in prestaties toe te schrijven is aan de lesmethode. Ook het enthousiasme of de werkhouding van de docent kan een storende variabele zijn geweest in dit quasi-experiment. In dit tekstboek zullen we verschillende voorbeelden van quasi-experimenteel onderzoek tegenkomen.

Binnen het type van experimenteel onderzoek kunnen we een verdere verdeling aanbrengen, tussen laboratoriumonderzoek en veldonderzoek. In beide typen experimenteel onderzoek wordt een aspect van de werkelijkheid gemanipuleerd. Het verschil tussen beide typen onderzoek is gelegen in de mate waarin de

onderzoeker in staat is om allerlei storende aspecten van de werkelijkheid onder controle te houden. In laboratoriumonderzoek kan de onderzoeker zeer exact bepalen onder welke omgevingscondities de observaties worden gedaan, en kan de onderzoeker dus ook vele mogelijke storende variabelen onder controle houden (denk aan verlichting, temperatuur, omgevingslawaaï, etc.). In veldonderzoek is dit niet het geval. De onderzoeker is ‘in het vrije veld’ niet in staat om alle (mogelijk relevante) aspecten van de werkelijkheid volledig onder controle te houden.

Voorbeeld 1.2: Margot van den Berg onderzocht samen met collega’s van de Universiteit van Ghana en de Universiteit van Lomé hoe meertalige sprekers hun talen gebruiken als zij eigenschappen zoals kleur, grootte en waarde moeten benoemen door middel van een zogenaamde *Director-Matcher task* (Van den Berg et al., 2017). In deze taak gaf de ene onderzoeksdeelnemer (de directeur) aanwijzingen aan een ander (de uitvoerder) om een reeks voorwerpen in een bepaalde volgorde neer te zetten. Zo konden in een kort tijdsbestek veel voorkomens van eigenschapswoorden worden verzameld (‘Zet de gele auto naast de rode auto maar boven de kleine slipper’). De gesprekken werden opgenomen, uitgeschreven en vervolgens onderzocht op taalkeuze, moment van taalwisseling en type grammaticale constructie. Bij dergelijk veldwerk kunnen echter allerlei niet-gecontroleerde aspecten in de omgeving van invloed zijn op de geluidsopnames, en daarmee op de gegevens: “kakelende kippen, een buurman die z’n motor aan het repareren is en ’m om de haverklap moet starten terwijl je een gesprek aan het opnemen bent, keiharde regen op het aluminium dak van het gebouw waar de interviews plaats vinden.” (Margot van den Berg, pers.comm.)

Voorbeeld 1.3: Bij het luisteren naar gesproken zinnen kunnen we uit de oogbewegingen van een proefpersoon afleiden, hoe die gesproken zinnen worden verwerkt. In een zgn. ‘visual world’-taak krijgen luisteraars een zin te horen (bijv. “Bert zegt dat het konijn is gegroeid”), terwijl ze kijken naar meerdere afbeeldingen op het scherm (meestal 4, bijv. een schelp, pauw, zaag, en wortel). Luisteraars blijken vooral te kijken naar de afbeelding die geassocieerd is aan het woord dat ze op dat moment mentaal verwerken: als ze het woord *konijn* verwerken, dan kijken ze naar de wortel (exacter gezegd: ze kijken vaker en langer naar de wortel dan naar de andere afbeeldingen). Met een zgn. ‘eye tracker’ kan worden vastgesteld naar welke positie van het scherm de proefpersoon kijkt (door observatie van de pupillen).

De onderzoeker kan zo dus observeren welk woord op welk moment mentaal verwerkt wordt (Koring et al., 2012). Dergelijk onderzoek kan het beste uitgevoerd worden in een laboratorium, met controle over achtergrondgeluiden, verlichting, en positie van de ogen t.o.v. computerscherm.

Laboratoriumonderzoek en veldonderzoek hebben beide voordelen en nadelen. Het grote voordeel van laboratoriumonderzoek is natuurlijk de mate waarin de onderzoeker allerlei externe zaken onder controle kan houden. In een laboratorium zal het experiment niet vaak verstoord worden door een startende motor of door een regenbui. Dit voordeel van laboratoriumonderzoek is echter ook een belangrijk nadeel, nl. dat het onderzoek plaatsvindt in een min of meer kunstmatige omgeving. Het is dan nog maar de vraag in hoeverre resultaten die onder kunstmatige omstandigheden verkregen zijn, ook zullen gelden in het leven van alledag buiten het laboratorium. Dit laatste is dan ook een punt in het voordeel van veldonderzoek: het onderzoek wordt verricht onder natuurlijke omstandigheden. Het nadeel van veldonderzoek is dan weer dat er in het veld kan van alles gebeuren wat de onderzoeksresultaten beïnvloedt, maar waar de onderzoeker geen controle over kan houden (zie het bovenstaande voorbeeld). De keuze die een onderzoeker maakt tussen beide typen experimenteel onderzoek wordt uiteraard sterk bepaald door de vraagstelling van het onderzoek. Sommige vraagstellingen laten zich beter in laboratoriumsituaties onderzoeken, terwijl andere beter in veldsituaties onderzocht kunnen worden (zoals bovenstaande voorbeelden illustreren).

1.6 Vooruitblik

Dit tekstboek bestaat uit drie delen. Deel I (hoofdstukken 1 tot en met 7) van dit tekstboek behandelt methoden van onderzoek, en geeft een toelichting bij allerlei termen en begrippen die van belang zijn bij het ontwerpen en opzetten van goed wetenschappelijk onderzoek.

In deel II (hoofdstukken 8 tot en met 12) van het tekstboek behandelen we de beschrijvende statistiek (Eng. ‘descriptive statistics’) en in deel III (hoofdstukken 13 tot en met 17) behandelen we de elementaire technieken uit de toetsende statistiek (Eng. ‘inferential statistics’). Met deze laatste twee delen streven we drie doelen na.

Allereerst willen we dat je in staat bent om artikelen en andere verslagen waarin statistische verwerkings- en toetsingstechnieken zijn gebruikt, kritisch te beoordelen. Ten tweede willen we dat je de noodzakelijke kennis en inzicht hebt in de belangrijkste statistische procedures. Ten derde willen we met deze

statistische delen bereiken dat je in staat bent om zelfstandig statistische bewerkingen uit te voeren voor je eigen onderzoek, bijvoorbeeld voor je stage of eindwerkstuk.

Deze drie doelen zijn geordend in volgorde van belangrijkheid. Wij menen dat een adequate en kritische interpretatie van statistische resultaten en de conclusies die daaraan verbonden kunnen worden van groot belang is voor alle studenten. Om die reden besteden we in dit tekstboek dan ook relatief veel aandacht (in deel I) aan de ‘filosofie’ of methodologie achter de besproken statistische technieken en analyses. Ook geven we aan hoe je de besproken statistische analyses zelf kunt uitvoeren in SPSS (een populair pakket voor statistische analyses) en in R (een wat moeilijker, maar ook krachtiger en veelzijdiger pakket, met stijgende populariteit). Beide statistische pakketten zijn geïnstalleerd in de computerleerzalen van de Faculteit Geesteswetenschappen. SPSS is beschikbaar via SurfSpot.nl voor een sterk gereduceerde prijs. R is vrijelijk beschikbaar via www.R-project.org. Meer achtergrond over het gebruik van R is te vinden via <https://hugoquene.github.io/emlar2020/>.

Bibliography

- Dingemanse, M., Torreira, F., and Enfield, N. (2013). Is “huh?” a universal word? conversational infrastructure and the convergent evolution of linguistic items. *PLOS One*, 8(11):e78273.
- Kerlinger, F. N. and Lee, H. B. (2000). *Foundations of Behavioral Research*. Harcourt College Publishers, Fort Worth, 4th edition.
- Koring, L., Mak, P., and Reuland, E. (2012). The time course of argument re-activation revealed: Using the visual world paradigm. *Cognition*, 123(3):361–379.
- Morton, A. (2003). *A Guide through the Theory of Knowledge*. Blackwell, Malden, MA, 3e edition.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth, Belmont, CA.
- Van den Berg, M., Amuzu, E. K., Essizewa, K., Yevudey, E., and Tagba, K. (2017). Crosslinguistic effects in adjectivization strategies in Suriname, Ghana and Togo. In Cutler, C., Vrzić, Z., and Angermeyer, P., editors, *Language Contact in Africa and the African Diaspora in the Americas: in honor of John V. Singler*, pages 343–362. Benjamins, s.l.
- Xie, Y. (2020). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.18.