Universiteit
Leiden
The Netherlands

**Voorbij het gemiddelde**
Dusseldorp, E.M.L.

**Citation**
Dusseldorp, E. M. L. (2025). *Voorbij het gemiddelde*. Retrieved from https://hdl.handle.net/1887/4248525

| | |
|---|---|
| Version: | Publisher's Version |
| License: | Leiden University Non-exclusive license |
| Downloaded from: | https://hdl.handle.net/1887/4248525 |

**Note:** To cite this publication please use the final published version (if applicable).

Prof. Dr. Elise M. L. Dusseldorp

# Voorbij het gemiddelde

Universiteit Leiden

Bij ons leer je de wereld kennen

# Voorbij het gemiddelde

Oratie uitgesproken door


## Prof. Dr. Elise M. L. Dusseldorp


bij de aanvaarding van het ambt van hoogleraar

Methodologie en Statistiek van Psychologisch Onderzoek

aan de Universiteit Leiden

op vrijdag 13 juni 2025


Universiteit
Leiden

"We need both exploratory and confirmatory"

John W. Tukey (1980)

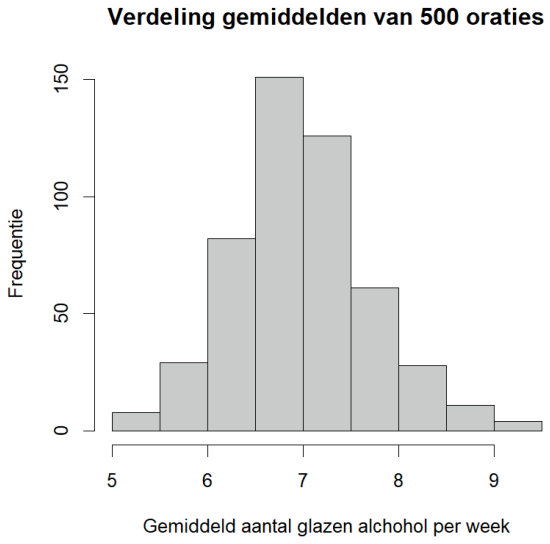
*Mevrouw de Rector Magnificus, mevrouw de Decaan, zeer gewaardeerde aanwezigen,*

Vanmiddag wil ik het graag met u hebben over het gebruik van statistiek bij vraagstukken in de psychologie en de ontwikkeling van nieuwe statistische methodes voor het beantwoorden van onderzoeksvragen. Het komt daarbij goed uit dat mijn leerstoel getiteld is "Methodologie en Statistiek van Psychologisch Onderzoek". Wie zich afvraagt hoe boeiend een verhandeling over statistiek kan zijn, hoop ik gerust te stellen door te melden dat ik zoveel mogelijk praktijkvoorbeelden uit psychologisch onderzoek zal gebruiken waar ik bij betrokken ben geweest. Deze voorbeelden gaan over het effect van behandelingen voor het verminderen van depressieve gevoelens tot behandelingen voor het bevorderen van een gezonde leefstijl. Dus ik verwacht dat ik daarmee uw aandacht wel kan trekken. Voor degenen onder u die vooral geïnteresseerd zijn in getallen en modellen heb ik een oefening bedacht, maar daarover later meer.

Op het gebied van het ontwikkelen van statistische methodes richt mijn onderzoek zich op het opsporen van subgroepen die *voorbij* het gemiddelde van de totale groep scoren, dus subgroepen die behoorlijk meer dan gemiddeld of minder dan gemiddeld scoren. Het uiteindelijke doel hiervan is om meer te weten te komen over de vraag "Wat werkt voor Wie?" en in de praktijk een behandeling op maat te kunnen bieden. Dat wil zeggen een behandeling die het beste past bij de persoon in kwestie in plaats van een behandeling die gemiddeld genomen passend is. Op dit gebied heb ik twee onderzoekslijnen. In de eerste onderzoekslijn gebruik ik gegevens van personen die toegewezen zijn aan verschillende behandelingen. In de tweede onderzoekslijn gebruik ik de resultaten van meerdere studies naar het effect van een en dezelfde behandeling. Voordat ik hierover ga uitwijden zou ik willen beginnen bij de basis: het gemiddelde zelf.

**Het gemiddelde als normaal en de normale verdeling**

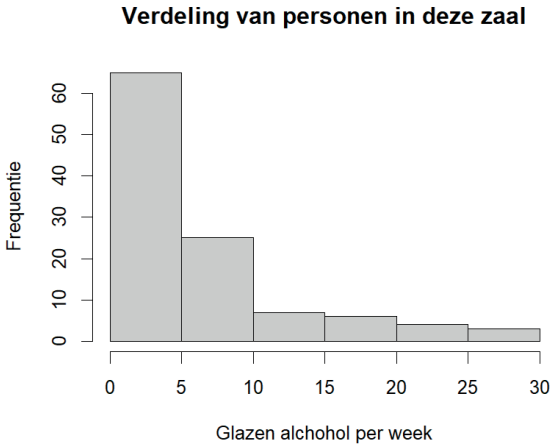
In de statistiek bekijken we hoe een eigenschap varieert tussen mensen en schatten we het gemiddelde van deze eigenschap. Bij het woord "eigenschap" in psychologisch onderzoek kunt u denken aan persoonlijkheidseigenschappen, maar ook aan gedragingen, emoties, lichamelijke verschijnselen en symptomen van bijvoorbeeld depressie. We kunnen het gemiddelde van een representatieve groep mensen gebruiken om een uitspraak te doen over een hele grote groep mensen. Ik zal dit illustreren met een voorbeeld. Stel, de eigenschap waar onze interesse naar uit gaat, is het alcoholgebruik van volwassenen in Nederland. Op de website StatLine van het Centraal Bureau voor de Statistiek [1] vind ik dat de volwassen Nederlander gemiddeld genomen 1 glas alcohol per dag drinkt, dus dit komt neer op 7 glazen (alcohol) per week. Ik wil de schatting van dit gemiddelde graag gaan controleren door zelf op onderzoek uit te gaan, bijvoorbeeld op de volgende manier. Ik selecteer een groep volwassenen in Nederland, dus ik zou kunnen beginnen bij ieder van u in deze zaal, en vraag hoeveel glazen alcohol u doorgaans drinkt per week. Ik noteer alle scores en bereken vervolgens het gemiddelde van deze groep. Stel dat dit gemiddelde ligt op 6,5 glas per week (iets onder het landelijk gemiddelde). Vervolgens vraag ik de pedel, degene die in dit gebouw de promoties en oraties organiseert, om bij de komende 500 oraties nogmaals de aanwezigen te vragen naar hun alcoholgebruik en om steeds het gemiddelde van de mensen in de zaal te noteren. Stel dat de pedel er in zou slagen om van 500 oraties steeds het gemiddelde alcoholgebruik te noteren, dan zal de verdeling er als volgt uit kunnen zien:






Op de verticale as van de figuur staat de frequentie, dus het aantal oraties dat een bepaalde gemiddelde waarde heeft gescoord. Hoe hoger de staaf hoe meer oraties deze waarde hebben gescoord. Zoals u ziet begint de horizontale as van deze figuur rond de 5. Het laagste gemiddelde dat de pedel vond was om precies te zijn 5,1 en het hoogste gemiddelde was 9,4. Omdat deze symmetrische, klokvormige verdeling erg vaak voorkwam in onderzoek, dus erg normaal was, noemden statistici aan het einde van de 18e eeuw al snel deze verdeling de "normaal verdeling" en zo noemen we die nog steeds. Over het algemeen kunnen we stellen dat als je meerdere steekproeven neemt van een willekeurige eigenschap, dat de verdeling van het gemiddelde van deze steekproeven er uitziet als een normaal verdeling. Het gemiddelde van deze verdeling is de beste schatting voor het gemiddelde van de hele grote groep waarover je een uitspraak wilt doen. Of de aanwezigen in deze zaal representatief zijn voor de volwassenen in Nederland laat ik even in het midden.

In plaats van naar de verdeling van de gemiddelden te kijken van veel steekproeven, kunnen we ook naar de verdeling van de scores kijken in 1 steekproef, zoals de verdeling van alcoholgebruik in deze zaal. Die verdeling ziet er bijvoorbeeld zo uit:




In tegenstelling tot de normaal verdeling die we daarnet hebben gezien, is deze verdeling erg scheef met veel mensen die 5 glazen of minder drinken en wat uitschieters aan de rechter kant. U ziet, er zit nogal wat variatie in jullie scores op alcoholgebruik (zie appendix voor code van de simulatie). De score varieert van 0 tot wel 30 glazen per week. Dergelijke scheve verdelingen komen vaak voor in psychologisch onderzoek bij de algemene bevolking. Veel mensen hebben bijvoorbeeld weinig depressieve klachten en een weinig mensen ervaren heel veel klachten.

Omdat scores op een eigenschap variëren, noemen we de eigenschap zelf in de statistiek "een variabele". In onderzoek is het prettig als de scores van personen op een variabele verschillend zijn, dus veel variëren. Een voorwaarde hierbij is wel dat de score betrouwbaar is gemeten, dus als we

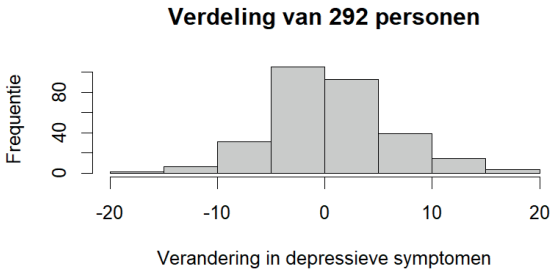
een persoon in korte tijd nogmaals zouden meten, zou de meting op dezelfde score moeten uitkomen. Dan kunnen we vervolgens proberen of we de variatie in de scores kunnen verklaren met andere variabelen. Bijvoorbeeld: zorgt meer bewegen voor minder alcoholgebruik? Of zelfs proberen of we op basis van andere variabelen de score op een eigenschap kunnen voorspellen. Hierover later meer. Ik heb u nu voldoende basisuitleg gegeven over het gemiddelde en neem u graag mee naar de introductie van mijn 1e onderzoekslijn.

**Het effect van een behandeling op basis van gemiddelden**

Naast het verklaren en voorspellen van een eigenschap van mensen, richt veel psychologisch onderzoek zich op het uitzoeken of een behandeling effectief is. Veel behandelingen in de psychologie hebben gedragsverandering voor ogen: bijvoorbeeld het bevorderen van gezonde leefgewoonten (zoals meer bewegen, betere voedingsgewoontes, en minder alcoholgebruik) of het verminderen van psychische klachten; dit veranderen van gedrag moet u dus ruim zien. Ik geef u graag een voorbeeld van een effectiviteitsonderzoek waarbij ik zelf als statisticus betrokken was.

*"Onderzoeker Marleen ter Avest is geïnteresseerd in de effectiviteit van een nieuwe vorm van cognitieve gedragstherapie, waarbij ook mindfulness technieken worden gebruikt [2]. Voor de kenners onder jullie is de volledige naam van deze therapie: Mindfulness based cognitieve therapie. Ik kort dit hier voor het gemak af tot mindfulness therapie. In deze therapie train je jouw aandacht om stil te staan bij wat er nu gebeurt, in het huidige moment, zonder enig oordeel. Je wordt jezelf bewust van de gedachten die door je heen gaan, emoties die je ervaart en lichamelijke sensaties die aanwezig zijn. Deze bewustwording kan helpen om gedrag dat je op de automatische piloot doet te veranderen. De doelgroep van het onderzoek van Marleen zijn volwassenen die depressieve symptomen ervaren. Ze heeft de gegevens van 292 volwassenen; 150 van hen hebben alleen de gebruikelijke behandeling gekregen, zoals antidepressieve medicatie of ondersteuning door een psychiatrisch verpleegkundige; de rest van de deelnemers, in totaal 142, hebben de nieuwe mindfulness therapie gevolgd, bovenop de gebruikelijke behandeling. De mate van depressie is gescoord op de Hamilton schaal, een meetinstrument voor depressieve symptomen (denk daarbij aan de mate van eetlust, vermoeidheid en prikkelbaarheid). De score wordt voor iedere deelnemer voorafgaand aan de behandeling bepaald en vlak na de behandeling. Marleen is geïnteresseerd in de verandering van deze scores. Ze wil weten of de verbetering in depressieve symptomen gemiddeld genomen groter is in de groep die de nieuwe therapie heeft gevolgd vergeleken met de groep die alleen de gebruikelijke behandeling heeft gekregen."*

De verandering in de scores kunnen we bepalen door de score na de behandeling af te trekken van de score voor de behandeling. Dit levert een veranderingsscore op, waarbij een positieve score betekent een verbetering van depressieve symptomen. De verdeling van de veranderingsscore voor de hele groep van Marleen ziet er als volgt uit:




Zoals u ziet zijn er behoorlijke verschillen tussen de deelnemers. De laagste score is -17, dit is een verslechtering van de depressieve symptomen en de hoogste score is 18, dus een verbetering ervan. Voor de groep die de nieuwe therapie heeft gevolgd blijkt de verandering gemiddeld genomen 2,1 te zijn, oftewel een lichte verbetering. De verandering voor de groep met alleen de gebruikelijke behandeling is 0,2, dus de

score blijft min of meer gelijk voor deze groep. Het verschil tussen de twee groepsgemiddelden is 1,9 (onthoud dit getal, want ik kom hier later nog op terug).

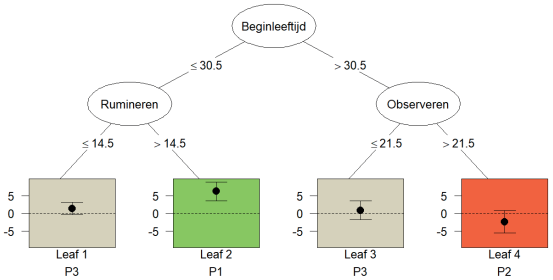
Dit verschil van 1,9 tussen de gemiddelden wordt het behandelingseffect genoemd en we toetsen of dit verschil statistisch gezien afwijkt van 0. Als dit zo is, en dit was hier inderdaad het geval, dan mogen we op basis van de steekproef de conclusie trekken dat in de gehele doelgroep van Marleen (volwassen met depressieve symptomen) het verschil afwijkt van 0. Dit noemen we statistische inferentie. Omdat het verschil hier positief is, concludeert Marleen dat de nieuwe mindfulness therapie effectief is. Belangrijk hierbij is dat de personen aselect (oftewel random) zijn toegewezen aan de behandelgroepen. Hierdoor verschillen de twee groepen aan de start van de studie alleen op de behandeling die ze volgen en zijn de gemiddelden van de groepen gelijk op allerlei andere variabelen, zoals bijvoorbeeld leeftijd.



Dit brengt me op de vervolgvraag waarmee ik deze oratie begon: Als de nieuwe behandeling gemiddeld genomen beter is betekent het dan ook dat deze voor iedere persoon beter is? Hiervoor moeten we voorbij het gemiddelde effect kijken. Vaak is het behandelingseffect heterogeen, dit betekent dat het varieert tussen personen. We zagen hiervan al een teken: de verandering in depressieve symptomen varieerde van min 17 tot wel plus 18. Een onderzoeksvraag waar ik in mijn 1e onderzoekslijn aan werk is:

**Voor welke personen werkt de behandeling meer of minder dan gemiddeld?**
**(kernvraag 1e onderzoekslijn)**
Voor het beantwoorden van deze vraag zijn we geïnteresseerd in variabelen die de sterkte van het behandelingseffect, oftewel de mate van verschil in gemiddelden tussen de twee behandelgroepen, beïnvloeden. Als we vooraf al weten welke variabele dit is (bijvoorbeeld leeftijd) kunnen we statistisch toetsen of het effect van type behandeling afhangt van het niveau van deze variabele. Echter in veel studies zijn er van te voren geen duidelijke hypotheses welke variabelen het effect beïnvloeden, simpelweg omdat er nog weinig onderzoek naar gedaan is. Dit was ook het geval bij de studie van Marleen. Ze had een lijst van 12 variabelen opgesteld die gemeten waren bij alle 292 deelnemers voorafgaand aan de behandeling en die mogelijk van invloed zouden kunnen zijn op het behandelingseffect (i.e., moderatoren).

Samen met andere onderzoekers heb ik voor deze situatie een analysemethode ontwikkeld [3-5]. Als we de methode toepassen op de gegevens van Marleen vinden we het volgende resultaat:



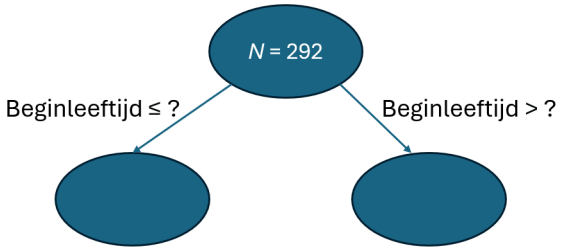

Dit resultaat wordt een boom genoemd met verschillende knopen. Voor het gemak heb ik een gedeelte van de boom weergegeven, de werkelijke boom was iets groter. De boom begint in de bovenste knoop met het opsplitsen van de totale groep van 292 personen in 2 subgroepen op basis van de variabele beginleeftijd van depressie. De personen met een beginleeftijd jonger of gelijk aan 30,5 jaar gaan naar links en de personen met een beginleeftijd van ouder dan 30,5 jaar gaan naar rechts. Vervolgens worden deze 2 subgroepen verder opgesplitst. De linker knoop wordt gesplitst op de variabele rumineren. Rumineren betekent het voortdurend malen ofwel herkauwen van negatieve gedachtes [6]. Deelnemers die dit vaak doen hebben een hogere score op deze variabele. De rechter knoop wordt gesplitst op de variabele observeren,

die aangeeft hoe goed de deelnemers in staat zijn om interne ervaringen waar te nemen, zoals gedachtes en emoties. Uiteindelijk heeft de boom vier verschillende eindknopen, de bladeren genoemd. De zwarte punten in de bladeren geven de grootte van het behandelingseffect weer; de stippellijn geeft de waarde 0 aan, dit betekent geen verschil tussen de gemiddelden van de twee behandelingen. Voor de grijs gekleurde bladeren is dit het geval: de zwarte punt ligt dicht bij 0. Voor het groen gekleurde blad heeft het behandelingseffect een hoge positieve waarde. Voor het rood gekleurde blad heeft het een negatieve waarde: de zwarte punt licht onder de 0.

Ik ga iets dieper in op de groene subgroep. Hierin zitten 66 deelnemers, 30 van hen hebben de nieuwe therapie gevolgd en de overige 36 deelnemers alleen de gebruikelijke behandeling. Het verschil in gemiddelde veranderingsscore tussen deze twee behandelgroepen is gelijk aan 6,2. Dit is ruim drie keer zoveel dan het verschil van 1,9 dat we vonden voor de totale groep. In deze groene subgroep zitten de deelnemers met een eerste depressieve periode voor de leeftijd van of gelijk aan 30,5 jaar en een score hoger dan 14,5 op rumineren; dit is relatief hoog omdat het gemiddelde ligt op 12,5. Marleen concludeert uit deze resultaten dat de nieuwe mindfulness therapie waarschijnlijk het beste werkt voor deze groene subgroep.

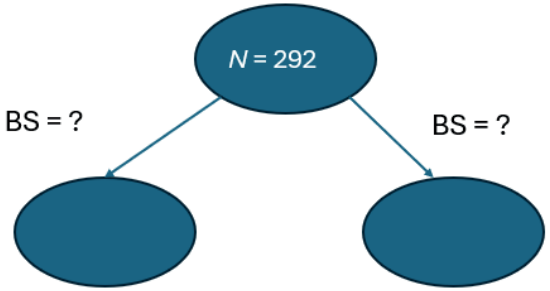
Graag leg ik u uit hoe de methode tot dit resultaat is gekomen. De methode wordt een algoritme genoemd en valt onder het populaire en brede onderzoeksgebied van Artificiële Intelligentie. Om de methode toe te passen op de studie van Marleen, is de veranderingsscore als uitkomstmaat gebruikt. Vervolgens zijn alle 12 variabelen als input meegegeven. Voor degenen onder u die geïnteresseerd zijn in getallen, maak ik graag een kleine uitstap naar de techniek.

Het algoritme begint met het zoeken onder alle input variabelen naar de beste 1e split. De input variabelen kunnen numeriek van aard zijn, zoals bijvoorbeeld de leeftijd van de eerste depressieve periode, maar ook categorisch, zoals bijvoorbeeld burgerlijke staat. Laten we stilstaan bij de beginleeftijd. Stel dat de waardes op deze variabele variëren van 27 tot en met 32 jaar. In deze oefening ga ik ervan uit dat de leeftijd in hele jaren is gemeten, dus de waargenomen waardes zijn 27, 28, 29, 30, 31, of 32 jaar. Op hoeveel verschillende manieren kunnen we dan een 1e split maken op deze variabele? Ik geef de liefhebbers van deze oefening onder u graag even de tijd om dit uit te zoeken.






Ik geef u drie mogelijke antwoorden: 5, 6 of 7 mogelijke splits. Wie schat in de zaal dat het 5 is? Graag uw hand omhoog. Wie denkt 6? En wie denkt 7? Het goede antwoord is 5. Hier ziet u alle mogelijke splits. Laten we dezelfde oefening doen voor Burgerlijke Staat (BS). Deze variabele heeft 4 categorieën. A: ongehuwd, B: gehuwd/partnerschap, C: gescheiden, D: verweduwd. Op hoeveel verschillende manieren kunnen we een 1e split maken op deze variabele?



Ik geef u eerst een tip en de tijd om het antwoord te vinden. Er zijn weer drie mogelijke antwoorden: 4, 6, of 7 mogelijkheden voor de 1e split. Wie schat in de zaal dat het 4 is? Wie 6? Wie 7? Het goede antwoord hier is 7.

Het bepalen van alle mogelijkheden voor de 1e split wordt herhaald voor alle input variabelen en vervolgens wordt de variabele en splitpunt gekozen die zorgt voor het grootste verschil in behandelingseffect tussen de subgroepen. Daarna gaat het algoritme verder met splits vinden voor de volgende knopen. Het algoritme maakt eerst een grote boom en vervolgens worden de takken die instabiel zijn weggesnoeid. Hoe dit precies gebeurt, laat ik voor nu achterwege. Belangrijk is om te onthouden dat het snoeien ons ervoor behoedt dat we toevalligheden vinden. Tot zover deze uitstap naar de techniek zelf.

U vraagt zich misschien af in hoeverre we dit resultaat kunnen generaliseren naar de doelgroep van mensen met depressieve symptomen en u ziet wellicht allerlei haken en ogen omdat de methode wel erg veel aan het zoeken is geweest in de gegevens. Daar heeft u helemaal gelijk in. Door dit uitvoerig zoeken kunnen we niet zo gemakkelijk een statistische toets uitvoeren. Maar wat hebben we dan aan dit resultaat?

De boom die we exploratief hebben gevonden kan gebruikt worden om hypotheses te genereren. In een vervolgonderzoek werden de resultaten van Marleen gebruikt om vooraf de hypothese op te stellen dat de mindfulness therapie beter zal werken voor personen met een verhoogde score op rumineren [7]. In dit onderzoek werd deze hypothese getoetst en het resultaat bevestigde de eerdere bevinding van Marleen. Op basis hiervan kunnen we het resultaat generaliseren naar de doelgroep.

U begrijpt dat de werkwijze om eerst hypotheses te genereren en vervolgens deze hypotheses te toetsen een arbeidsintensieve en tijdrovende aangelegenheid is. De studie van Marleen stamt bijvoorbeeld uit 2019 en de studie met het bevestigende resultaat werd pas 5 jaar later gepubliceerd. We weten dat wetenschappelijke vooruitgang langzaam gaat maar kan dit misschien wat sneller? Ik zal u hierover meer vertellen als ik de nieuwe perspectieven beschrijf, maar sluit voor nu deze 1e onderzoekslijn af.

*Introductie van mijn 2e onderzoekslijn: exploratieve methode voor meta-regressie*

Ter introductie van mijn 2e onderzoekslijn neem ik u mee naar een geheel andere onderzoeksstrategie. Deze strategie werkt als volgt: we zoeken in de internationale wetenschappelijke literatuur naar studies die het effect van een zelfde behandeling hebben onderzocht en hierover hebben gerapporteerd. Het analyseren van de resultaten van meerdere studies wordt meta-analyse genoemd. Het doel hiervan is om een uitspraak te doen over het gemiddelde effect van eenzelfde type behandeling.

**Het gemiddelde effect van een behandeling op basis van meerdere studies**
Ik deel graag met u een voorbeeld hiervan uit de gezondheidspsychologie.

*"Onderzoeker Susan Michie [8] is geïnteresseerd in het effect van behandelingen ter bevordering van een gezonde leefstijl. --Denkt u hierbij aan trainingen die gericht zijn op meer bewegen en/of het bevorderen van gezonde eetgewoontes. Deze trainingen zijn zorgvuldig ontwikkeld en worden op een standaard manier uitgevoerd. We noemen deze trainingen leefstijlinterventies--. Van de meer dan 100 studies die Susan en haar collega's hebben gevonden in de literatuur noteert ze de gemiddelde verandering in gezond gedrag van de groep die de leefstijlinterventie heeft gevolgd en de gemiddelde verandering in de groep die de interventie niet heeft gehad. Ook noteert ze de standaarddeviatie van de verandering, dit is een maat voor de spreiding in de scores, en ze noteert de groepsgroottes. Met deze gegevens voert Susan een meta-analyse uit en ze vindt dat de leefstijlinterventies gemiddeld genomen een effect hebben van 0,27. "*

Hoe moeten we dit effect duiden? Het betekent dat gemiddeld genomen de groep die de leefstijlinterventie heeft gevolgd 0,27 standaarddeviatie hoger scoort op gezondheidsgedrag dan de groep die dit niet heeft gehad. Is dit nu een groot effect? In de psychologie gebruiken we vaak de vuistregel van Cohen [9] die aangeeft dat 0,2 een klein effect is, 0,5 een middelgroot en 0,8 een groot effect. Dus gemiddeld genomen hebben de leefstijlinterventies een klein tot middelgroot effect.

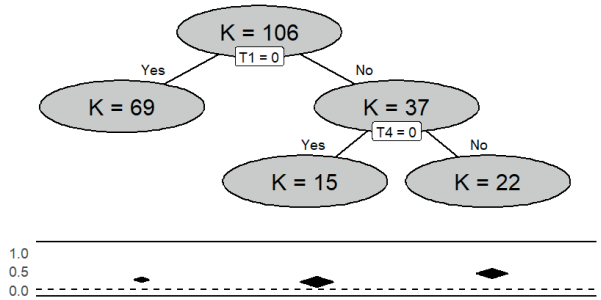
Susan constateert verder dat dat er veel variatie zit in de afzonderlijke effectgroottes van de studies. Sommige studies laten een heel groot effect zien en andere geen effect of een negatief effect. Ze vermoedt dat dit komt doordat de leefstijlinterventies verschillend zijn in de technieken die ze gebruiken om het gezonde gedrag te bevorderen. Een voorbeeld van zo'n techniek is "het stimuleren tot het opstellen van een intentie". Door deze techniek leren de deelnemers om een intentie te formuleren die passend is voor hen. Voor iemand die bijvoorbeeld niet aan sport doet, kan zo'n intentie zijn "ik wil tenminste 2 keer per week naar de sportschool gaan". Susan besluit om de gedragsveranderingstechnieken te coderen voor alle leefstijlinterventies in haar meta-analyse. In het totaal codeert ze 26 technieken: met een 1 geeft ze aan dat de techniek wordt gebruikt en met een 0 als deze niet is gebruikt. Ik laat u ter illustratie haar gegevens zien van 11 studies en 5 technieken:

| study | g | vi | T1 | T2 | T3 | T4 | T25 |
|---|---|---|---|---|---|---|---|
| ALDANA 2005 (PA) | 0.61 | 0.02 | 1 | 1 | 0 | 1 | 0 |
| ANDERSON 2006 | 0.75 | 0.05 | 0 | 0 | 0 | 1 | 0 |
| ARAO 2007 (PA) | 0.73 | 0.07 | 0 | 0 | 0 | 1 | 0 |
| BABAZONO 2007 (PA) | 0.88 | 0.10 | 0 | 0 | 0 | 1 | 0 |
| BAKER 2008 | 0.74 | 0.05 | 0 | 1 | 0 | 0 | 0 |
| BENNETT 2008 | 0.16 | 0.05 | 0 | 0 | 0 | 1 | 1 |
| BLISSMER 2002 | 0.40 | 0.05 | 0 | 1 | 0 | 0 | 0 |
| BOLOGNESI 2006 | 0.52 | 0.04 | 1 | 1 | 0 | 0 | 0 |
| BULL 1999 | 0.18 | 0.01 | 0 | 1 | 0 | 0 | 0 |
| CALFAS 1996 | 0.19 | 0.02 | 0 | 1 | 0 | 1 | 0 |
| CALFAS 2000 (F) | 0.00 | 0.02 | 0 | 1 | 0 | 0 | 0 |

De eerste studie van Aldana heeft een gemiddeld behandelingseffect gevonden van 0,61 (dit is de kolom met de g) en de interventie gebruikt de technieken 1, 2, en 4. De studie helemaal onderaan van Calfas uit 2000 vindt een behandelingseffect van 0,00, dus geen effect. Deze studie gebruikt alleen techniek 2. Susan wil graag weten of een bepaalde combinatie van technieken zorgt voor een hoger effect. Zij vraagt zich af:

**Kunnen we groepen studies identificeren met een hoger effect dan het gemiddelde?**

Dit is de kernvraag van mijn 2e onderzoekslijn. Tijdens mijn werk bij TNO maakte ik samen met anderen een eerste versie van een methode om deze vraag te beantwoorden [10]. Op het mathematisch instituut heeft promovendus Xinru Li de methode verder ontwikkeld [11]. Sinds vorige maand hebben we versie 3 beschikbaar [12, 13]. De methode werkt als volgt: Op basis van kenmerken worden de studies verdeeld in subgroepen die een lager of juist hoger gemiddeld effect laten zien. Als we de methode toepassen op de gegevens van Susan levert dit het volgende resultaat op:




De boom begint met 106 studies die worden gesplitst op basis van techniek 1. Deze techniek houdt in "informatie verschaffen over hoe meer bewegen en gezonder eten bijdraagt aan je gezondheid". In de subgroep van 69 studies in de meest linker knoop is de score op deze techniek gelijk aan 0, dus hij wordt niet gebruikt. De studies die deze techniek wel gebruiken worden vervolgens gesplitst op techniek 4 "het stimuleren tot het opstellen van een intentie". De rechtse subgroep van 22 studies die beide technieken gebruiken laat gemiddeld genomen het sterkste effect zien, namelijk een effect grootte van 0,43, een middelgroot effect. In de andere 2 subgroepen is het gemiddelde effect ongeveer de helft hiervan (i.e., 0,26 en 0,21).

Deze boommethode is ook een exploratieve methode en heeft veel weg van het vorige algoritme. Onder water gebeuren er hele andere berekeningen maar in grote lijnen komen de algoritmes overeen. Wat nieuw is aan deze zogenaamde meta-bomen, is dat we niet alleen hypotheses genereren maar ook een manier hebben gevonden om te testen of de boomstructuur de spreiding in de effect groottes kan verklaren. Dit doen we door de gegevens te husselen (permuteren)[1]. Voor de boom op de data van Susan, resulteerde de toets in een *p*-waarde van 0,02. Dit betekent dat de opdeling in 3 subgroepen op basis van de 2 technieken de spreiding in de effect groottes significant verklaard. We hebben ook een manier gevonden om de onzekerheid in de effectgroottes in de bladeren van de boom te schatten (bootstrapping). Als u hier nieuwsgierig naar bent, kunt dit zelf uitvoeren op de gegevens van Susan, met de code in de appendix van het boekje van mijn oratie.

Deze vernieuwingen maken statistische inferentie mogelijk bij een exploratieve methode. Door de data op een slimme manier te hergebruiken kunnen we de resultaten voorzichtig generaliseren naar de grotere groep. Ik spreek hier van voorzichtig omdat er nog meer toekomstig onderzoek voor nodig is. Voor nu sluit ik mijn 2e onderzoekslijn hier af. Dit brengt mij op:

1 We fixeren de kolommen met de effectgrootte (g) en de variantie (vi) en gaan de rijen met de scores op de technieken op een willekeurige manier verwisselen en dit herhalen we een heleboel keer. Iedere keer berekenen we hoeveel spreiding we kunnen verklaren met een boom op de gehusselde gegevens en we noteren of dit meer is dan de boom op de originele gegevens. Zo kunnen we een *p*-waarde bepalen.

**Nieuwe perspectieven en aandachtspunten**

*Trainen en testen*

Er zijn nog meer manieren waarop we een dataset intensiever kunnen gebruiken om te leren van de gegevens. We kunnen bijvoorbeeld de dataset splitsen in tweeën: het ene deel om te exploreren en een model, zoals onze boom, te "trainen" (ontwikkelen) en vervolgens het andere deel om het model te testen. Deze werkwijze wordt vaak gebruikt in "statistical learning" [14]. Een andere benaming hiervoor is "machine learning".

Het splitsen van gegevens is echter niet zo vaak mogelijk in de praktijk. Voor de boommethode die ik heb toegepast op de gegevens van Marleen hebben we al minimaal 300 personen nodig om te trainen. Dit brengt me op de spagaat waar we vaak in terecht komen, omdat de datasets niet al te groot zijn. Dus "voorbij het gemiddelde" kijken is helaas vaak buiten onze middelen ("beyond the mean is often beyond our means"). Ik vond in de literatuur meerdere voorbeelden van onderzoekers die in hun studieprotocol melden het boomalgoritme te zullen gaan gebruiken, maar het uiteindelijk toch niet gebruikten omdat de dataverzameling tegenviel [15].

*Open Science*

Recente ontwikkelingen in Open Science bieden mogelijk een oplossing hiervoor. Door deze beweging stellen steeds meer onderzoekers hun data beschikbaar voor anderen. Op deze manier kunnen we persoonsgegevens van meerdere onderzoeken samennemen en analyseren. Om dit goed te kunnen doen, zijn er wel wat voorwaarden; idealiter worden voor tenminste een deel dezelfde meetinstrumenten en hetzelfde type behandeling gebruikt. Dit zou wat mij betreft wat vaker mogen gebeuren in de sociale en gedragswetenschappen.

*Andere manieren van meten en modelleren van het behandelingseffect*

Er zijn ook vernieuwingen gaande op het gebied van de methodologie. Door nieuwe technieken zoals tablets en smartwatches kan het aantal metingen fors worden uitgebreid. Zo scoren bijvoorbeeld de deelnemers meerdere keren op een dag hoe ze zich voelen of worden via de smart watch fysiologische gegevens (zoals hartslag) van hen verzameld. Mijn naaste collega Mark de Rooij onderzoekt samen met promovendi hoe verschillende type databestanden het beste kunnen worden gecombineerd om goede voorspellingen te doen [16].

Ook het gebruik van slechts één samenvattende score voor een psychologische eigenschap, zoals de Hamilton schaal voor depressieve symptomen, staat ter discussie. De studie van mijn promovendus Bunga Citra Pratiwi wees uit dat het gebruik van een samenvattende maat in combinatie met een statistical learning methode in de meeste gevallen beter was in het voorspellen dan het gebruik van de losse items van een psychologische test [17].

*Voorspellen op individueel niveau*

Er bestaan algoritmes die niet 1 boom maken maar wel 1000 bomen, dus kortom een heel bos. Deze methodes hebben hippe namen als Random Forest en Gradient Boosting en zijn recentelijk zo aangepast dat ze ook het effect van een behandeling kunnen voorspellen [18][2].

2 Omdat deelnemers aan onderzoek meestal maar één behandeling hebben ontvangen en we niet weten wat de score op de uitkomstmaat zou zijn als de persoon de andere behandeling zou hebben gehad is het ontwikkelen van een goede analysemethode hiervoor een hele uitdaging. Door gebruik te maken van technieken waarmee missende waardes worden geschat wordt de potentiële waarde op de uitkomstvariabele geschat voor de behandeling die een persoon niet heeft gehad. Dus uiteindelijk hebben we voor 1 persoon 2 scores: een waargenomen score en een potentiële score. Het verschil tussen deze twee scores is het behandelingseffect per persoon. Deze verschilscore gebruiken we vervolgens als uitkomst maat om bijvoorbeeld een Random Forest te maken, zie [18] voor een overzichtsartikel.

Het combineren van de resultaten van meerdere bomen heeft meerdere voordelen:

1) De voorspelling blijkt in de meeste gevallen beter te zijn dan de voorspelling door 1 boom.

2) Het behandelingseffect kan voorspeld worden op individueel niveau in plaats van op subgroep niveau. Het stelt ons dus instaat om *voorbij het gemiddelde* effect van de totale groep en ook *voorbij gemiddelde* effecten van subgroepen te kijken.

Het combineren heeft ook nadelen. We begrijpen niet meer zo goed hoe de voorspelling tot stand is gekomen. We zien als het ware door het bos de afzonderlijke bomen niet meer. Dat maakt de acceptatie voor het gebruik van deze methodes in de praktijk lastig. Stel dat u naar de huisarts gaat en de huisarts geeft u een behandeling omdat de computer heeft voorspeld dat deze het beste is voor u, zonder dat de huisarts kan uitleggen waarom. Nemen we dit dan aan? En voelt de huisarts zich daar comfortabel bij? En wat is de onzekerheidsmarge bij de voorspelling?



Hiervoor is nog meer onderzoek nodig. Daarom deel ik graag de resterende tijd met u:

**Waar ik in toekomstig onderzoek voor wil gaan en waar ik voor sta**

*Mensgerichte Artificiële Intelligentie*

Ik verwacht dat het in de toekomst mogelijk zal zijn om met een handzame, begrijpelijke set van regels een **mensgerichte AI-tool** voor bijvoorbeeld de zorgpraktijk te scheppen. Mijn naaste collega Marjolein Fokkema maakt hierin veelbelovende stappen door kleine bomen te combineren voor een goede voorspelling, waardoor een betere interpretatie van de resultaten mogelijk is [19]. Een andere aanpak is om door experts opgestelde beslisregels te verbeteren met machine learning, zoals master student Jurgens Fourie onlangs aantoonde [20]. Er is nog veel te onderzoeken op dit gebied van "Explainable AI (XAI)" en ik wil hierin samen optrekken met statistici binnen de interfacultaire master Statistics & Data Science, GGZ instellingen, het instituut voor chemische neurowetenschappen (iCNS) en onderzoekers in het thema "Gezonde Maatschappij" van het samenwerkingsverband tussen Leiden, Delft en de Erasmus Universiteit.

*Postmodelsectie inferentie*

Een recent onderzoeksgebied richt zich specifiek op het toepassen van statistische inferentie op de resultaten van exploratieve methodes. Dit heet met een chic woord: post modelselectie inferentie. Onlangs lieten Anna Neufeld en collega's [21] zien hoe je verschillen in gemiddelden kunt toetsen in de subgroepen die gevonden zijn door een boomalgoritme (regressiebomen). Ik wil dit graag in toekomstig onderzoek verder gaan uitwerken voor algoritmes gericht op behandelingseffecten.

*Belang van toepassingen en statistische consultatie*

Naast ontwikkelen van algoritmes houd ik mij graag bezig met toepassingen. Ik denk dat u dat wel hebt gemerkt. Nadat een nieuw ontwikkelde methode uitvoerig is getest (met artificiële data), komt de geloofwaardigheid en bruikbaarheid ervan het beste tot zijn recht door goede toepassingen. Andersom kunnen toepassingen de statisticus inspireren tot het ontwikkelen van een nieuwe methode. Ik haal hier graag het standpunt aan van Stella Cunliffe, die als eerste vrouw werd benoemd tot President van de Britse Royal Statistical Society. Zij pleitte ervoor dat statistici uit hun ivoren toren klimmen, hun waarde gedurende het gehele onderzoekstraject bewijzen en de taal van de onderzoeker leren spreken [22]. Ik ben het hier helemaal mee eens en zou in aanvulling hierop willen pleiten dat onderzoekers al vanaf het begin van hun

onderzoekstraject een statisticus bij de hand nemen en niet pas aan het eind als bijvoorbeeld een reviewer een probleem met de gehanteerde analysemethode opwerpt. Er zijn geluiden dat statistici sinds de komst van ChatGPT niet meer nodig zijn. Uiteraard ben ik het hier helemaal niet mee eens. Samen met mijn naaste collega Anikó Lovik heb ik dit verder uitgezocht. We vonden dat ChatGPT af en toe opmerkelijk goede code kan genereren, maar dat de statisticus hard nodig is om te beoordelen of deze code past bij de onderzoeksvraag en om een goede conclusie uit de resultaten te trekken [23].

*Belang van onderwijs als voorbereiding op de praktijk*

Aan de universiteit houden we ons naast onderzoek bezig met het geven van onderwijs. Ik vind het belangrijk dat we onze studenten opleiden tot kritisch en zelfstandig denkende professionals en hen voorbereiden op de praktijk. Ik heb samen met mijn collega's van Methodologie en Statistiek ervoor gezorgd dat we in de bachelor Psychologie zijn overgegaan op de open software R. Zo leren de studenten stapsgewijs met oefeningen wat een statistische toets inhoudt en hoe ze onderzoeksgegevens kunnen analyseren. Dit in tegenstelling tot klik klik klik in een gebruikersinterface die kant en klare toetsen voorschotelt. In het masteronderwijs geef ik het vak statistische consultatie, waarin de studenten in paren werken aan het oplossen van echte analyseproblemen van onderzoekers. Ze leren al doende om de statistische kennis die ze hebben te vertalen naar de taal van de onderzoeker. De interactie met de masterstudenten in dit vak en het zien groeien van hun vaardigheden, geven mij enorm veel voldoening.

**Dankwoord**

Tot slot wil ik graag allen bedanken die mij hebben ondersteund direct of indirect op mijn wetenschappelijke pad. Ik wil iedereen in de zaal en via de livestream bedanken voor uw luisterend oor. Degenen die aan de totstandkoming van mijn benoeming hebben bijgedragen wil ik bedanken voor hun vertrouwen in mij: het college van bestuur, de vorige decaan en de vorige wetenschappelijke directeur van ons instituut, Paul Wouters en Andrea Evers, en hun directe opvolgers Sarah de Rijcke en Hanneke Hulst.

Tijdens mijn master Methoden en Technieken aan de universiteit van Leiden voelde ik mij als student-assistent bij de sectie Gezondheidspsychologie als een vis in het water. Ik wil mijn collega's bij die sectie bedanken voor de fijne samenwerkingen en Stan Maes in het bijzonder die mij stimuleerde om me verder te ontwikkelen in de wetenschap en te gaan promoveren bij hem en Jacqueline Meulman.

Jacqueline, ik ben jou erg dankbaar voor jouw stimulerende aanpak tijdens mijn promotietraject. Ik was nog geen maand promovendus of je stuurde me naar een cursus Modern Regression & Classification in Boston die werd gegeven door Trevor Hastie en Robert Tibshirani. Tijdens deze cursus kwam ik op het idee om bomen te gaan gebruiken om differentiële behandelingseffecten aan te tonen. Dit was bepalend voor mijn verdere ontwikkeling. Jouw ervaring met exploratieve data analyse is erg inspirerend geweest en je gaf me veel positieve feedback waardoor mijn zelfvertrouwen groeide. Ik wil mijn collega's van datatheorie bedanken voor de leuke en inspirerende congresreizen die we samen ondernamen.

Ik wil Paul Eilers bedanken voor zijn stimulans om te gaan solliciteren buiten de universiteit. Bij TNO kwam ik terecht bij de statistiek groep binnen de thema's Jeugd en Gezondheid. Ik wil Stef, Paula, Hedwig en alle andere TNO collega's bedanken voor de enorm leerzame tijd die ik daar heb gehad om wetenschappelijk kennis naar de praktijk toe te vertalen.

Mark de Rooij en Serge Rombouts, dank jullie wel voor het aannemen van mij bij de sectie Methodologie en Statistiek. Dit is voor mij een droomplek waar alles samenkomt: lesgeven, consultatie en fundamenteel onderzoek. Sinds drie jaar ben

ik voorzitter van de sectie en ik vind het heel fijn dat ik vaak bij jullie terecht kan voor sparren en advies. Het besturen van de sectie doe ik samen met Hemmo, Tom W. en Wouter, met ondersteuning door onze onmisbare management assistent Ingrid. Ik wil jullie enorm bedanken voor de vruchtbare samenwerking. Alle collega's van de sectie Methodologie en Statistiek: dank je wel voor de fijne werksfeer, de inhoudelijke discussies en de lachsalvo's tijdens de lunches. Door jullie fiets ik met veel plezier naar mijn werk.

Ook mijn collega's bij de andere secties van ons instituut en bij Pedagogiek; dank jullie wel voor de fijne samenwerkingen en sommigen van jullie voor het meedoen als onderzoeker aan het vak Statistische Consultatie.

Mijn zus, aanwezige familie en mijn vriendinnen wil ik bedanken voor de leuke momenten van ontspanning en de fijne gesprekken. De band die ik met jullie heb is me heel dierbaar.



Mijn ouders hebben een grote rol gespeeld in mijn ontwikkeling en hebben ieder op hun eigen manier mij gevormd. Ze zijn er vandaag niet bij omdat mijn vader is overleden en mijn moeder in een verpleeghuis zit, maar ze hebben allebei gelukkig wel meegekregen dat ik hoogleraar werd. Pap, door jou heb ik mijn fascinatie voor getallen. Ik heb ervoor gekozen om vrijdag de 13e te kiezen voor mijn oratie als knipoog naar jou, omdat je ervan genoot als jouw verjaardag op een vrijdag de 13e viel en hoopte dat alles op die dag goed zou gaan. Mam, door jou zie ik vrede en rijkdom niet als vanzelfsprekend en dop ik mijn eigen boontjes. Je hebt je hele leven gewerkt als docent en was zeer zelfstandig. Je bent een voorbeeld voor me in hoe je werk en sociale en maatschappelijke betrokkenheid kon combineren.

Graag wil ik tenslotte mij richten tot Teun en Pieter. Lieve Teun, het is een voorrecht om jou te zien opgroeien en jouw eigen weg te zien vinden. Je hebt interesse in getallen en proces mining en ik vind het heel leuk om het met jou daarover te hebben. Je doet de dingen op jouw eigen manier, daar ben ik enorm trots op. Dank je wel voor jouw aandacht voor waar ik mee bezig ben en voor jouw steun. Lieve Pieter, zonder jou ben ik niet wie ik nu ben. Jouw liefde en steun door dik en dun heeft het mogelijk gemaakt dat ik hier nu sta. Dank je wel!

'Ik heb gezegd'.

**Referenties**


1. StatLine - Leefstijl; persoonskenmerken, geraadpleegd op 1 mei 2025. Geselecteerd op jaar 2024 en onderwerp alcoholgebruik.
2. Ter Avest, M. J., Dusseldorp, E., Huijbers, M. J., van Aalderen, J. R., Cladder-Micus, M. B., Spinhoven, P., Greven, C. U., & Speckens, A. E. (2019). Added value of mindfulness-based cognitive therapy for depression: A Tree-based qualitative interaction analysis. *Behavior Research and Therapy, 122*, 103467. doi: 10.1016/j.brat.2019.103467.
3. Dusseldorp, E., & Meulman, J. J. (2004). The regression trunk approach to discover treatment covariate interaction. *Psychometrika*, *69*(3), 355-374.
4. Dusseldorp, E., & Van Mechelen, I. (2014). Qualitative interaction trees: a tool to identify qualitative treatment–subgroup interactions. *Statistics in Medicine*, *33*(2), 219-237.
5. Dusseldorp, E., Doove, L., & Mechelen, I. V. (2016). Quint: An R package for the identification of subgroups of clients who differ in which treatment alternative is best for them. *Behavior Research Methods*, *48*, 650-663.
6. Venhuizen, G. (2021). Aandacht voor depressie. NRC, 19-09-2021.
7. Lubbers, J., Geurts, D. E., Spinhoven, P., Cladder-Micus, M. B., Ennen, D., Speckens, A. E., & Spijker, J. (2024). Rumination and self-compassion moderate mindfulness-based cognitive therapy for patients with recurrent and persistent major depressive disorder: A controlled trial. *Depression and Anxiety, 2024*(1), 3511703.
8. Michie, S., Abraham, C., Whittington, C., McAteer, J., & Gupta, S. (2009). Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychology, 28(6)*, 690.
9. Cohen (1988). *Statistical power analysis for the behavioral scientist* (2nd Ed.). Lawrence Erlbaum Associates.
10. Dusseldorp, E., Van Genugten, L., van Buuren, S., Verheijden, M. W., & van Empelen, P. (2014). Combinations of techniques that effectively change health behavior: Evidence from Meta-CART analysis. *Health Psychology, 33(12)*, 1530.
11. Li, X., Dusseldorp, E., & Meulman, J. J. (2019). A flexible approach to identify interaction effects between moderators in meta-analysis. *Research Synthesis Methods*, *10*, 134–152.
12. Li, X., Dusseldorp, E., Claramunt González, J. (2025). metacart: Meta-CART: A Flexible Approach to Identify Moderators in Meta-Analysis. R package version 3.0.0, <https://CRAN.R-project.org/package=metacart>.
13. Li, X., Dusseldorp, E., Claramunt-González, J., Su, X., Van Megen, J., & Meulman, J. J. (submitted) Enhanced tree-based subgroup identification in meta-analysis.
14. Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction.* Springer Publishing Company.
15. van Santen, J., Dröes, R. M., Twisk, J. W., Henkemans, O. A. B., van Straten, A., & Meiland, F. J. (2020). Effects of exergaming on cognitive and social functioning of people with dementia: a randomized controlled trial. *Journal of the American Medical Directors Association*, *21*(12), 1958-1967.
16. van Loon, W., Fokkema, M., Szabo, B., & de Rooij, M. (2024). View selection in multi-view stacking: choosing the meta-learner. *Advances in Data Analysis and Classification*, 1-39.
17. Pratiwi, B. C., Dusseldorp, E., Karch, J. D., & de Rooij, M. (2023). Predictive performance of psychological tests: Is it better to use items than subscales?. *Computational Statistics & Data Analysis*, *185*, 107767.

18. Lipkovich, I., Svensson, D., Ratitch, B., & Dmitrienko, A. (2024). Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data. *Statistics in Medicine*, *43*(22), 4388-4436.
19. Fokkema, M., & Strobl, C. (2020). Fitting prediction rule ensembles to psychological research data: An introduction and tutorial. *Psychological Methods*, *25*(5), 636.
20. Fourie, J. J. (2025). Rules for transport mode determination in smart travel surveys. *Master thesis Statistics and Data Science,* Leiden University.
21. Neufeld, A. C., Gao, L. L., & Witten, D. M. (2022). Tree-values: selective inference for regression trees. *Journal of Machine Learning Research*, *23*(305), 1-43.
22. Starmans, R (2025). Statistische prijzen en de erfenis van Stella Cunliffe. *STAtOR: periodiek van de VVSOR, 26*(1), 30-34
23. Lovik, A., & Dusseldorp, E. (2025). Statistical consulting guidelines for new researchers in psychiatry and mental health–beyond ChatGPT. *BJPsych Advances*, *31*(2), 91-102.

**Appendix met R-code**

```
##generate alcohol data

#positively skewed variable alcohol use in one sample of 110
persons

n_pers <- 110; set.seed(130625)

var_alc <- rnbinom(n_pers, size=1, mu = 7.0) #negative
binomial; count data

hist(var_alc, xlab = "Glazen alchohol per week", ylab =
"Frequentie", main = "Verdeling van personen in deze zaal")

summary(var_alc)

#repeat 500 times and save the mean

K <- 500; mean_alc <- numeric(K)

set.seed(130625)

for (i in 1 : K) {

var_alc <- rnbinom(n_pers, size=1, mu = 7.0)

mean_alc[i] <- mean(var_alc)

}

hist(mean_alc, xlab = "Glazen alchohol per week", ylab =
"Frequentie", main = paste( "Verdeling gemiddelden van", K,
"oraties"))

summary(mean_alc)

## meta analyses with part of the data from Susan Michie

library(metacart)

data("dat.BCT2009")

#create formula with effect size g as outcome and BCTs as
possible moderators

#T3 is omitted due to no variance in the scores

formBCT<- as.formula(g ~ T1 + T2 + T4 + T25)

set.seed(111111)

modBCT.replus_perm <-  REmrt(formBCT, data=dat.
BCT2009, vi=vi, c.pruning = 0,

                lookahead = TRUE, sss = FALSE, perm = 100)

summary(modBCT.replus_perm)  #results with 100
permutations

#bootstrap confidence intervals:

set.seed(111111)

Results_boot <- BootCI(modBCT.replus_perm, nboot = 50)

summary(Results_boot); plot(Results_boot)
```

## Prof. Dr. Elise M. L. Dusseldorp

Na haar conservatoriumopleiding te Rotterdam (richting Schoolmuziek) studeerde Elise Dusseldorp (1966) psychologie aan de Universiteit Leiden. Ze promoveerde in 2001 bij Stan Maes en Jacqueline Meulman op het proefschrift “Discovering Treatment Covariate Interaction: An Integration of Regression Trees and Multiple Regression”.

Sinds 2015 werkt ze bij de sectie Methodologie en Statistiek waar ze in 2023 hoogleraar “Methodologie en Statistiek van Psychologisch Onderzoek” werd. In haar onderzoek richt ze zich op effecten van combinaties van factoren (interacties) en is ze geïnteresseerd in meta-analyse en het statistisch beantwoorden van de vraag: Welke behandeling werkt voor wie? Ze heeft een passie voor statistische consultatie en geniet ervan als ze samen met onderzoekers in de sociale en gedragswetenschappen een brug kan slaan tussen inhoudelijke en statistische expertise.