Vince Penders

PIRATEN, PERZIKEN EN P-WAARDEN

Productie en distributie: Mosae Verbo, Maastricht
www.boekenplan.nl

Omslag en illustraties: Chelsy Penders

ISBN 978 90 8666 389 7
NUR 123

INHOUDSOPGAVE

ZIJ DIE GEMAAKT HEBBEN, GROETEN U

Welkom, beste lezer! Voor je neus ligt een splinternieuw handboek, en wat zijn we trots op het resultaat. Als makers willen wij ons kort aan je voorstellen. Mag dat?

Ik ben **VINCE PENDERS**… noem me maar **Vincenzo**. Het brein achter deze lesmethode – dat ben ik. Creatief tot op het bot, een liefhebber van koken, bakken, muziek en Japanse videogames. Zo ongeveer alles interesseert me, vooral als het met mensen en maatschappijen te maken heeft. Voor de kost geef ik les aan studenten, met ontzettend veel plezier, en ik schrijf boeken. Ook boeken die niet over statistiek gaan: spannende verhalen met een vleugje sciencefiction! Mijn debuut *Zwaluwhart* is in mei 2015 uitgekomen bij Uitgeverij Macc. Leuk voor als je je statistiekvak gehaald hebt en eens even heerlijk wilt ontspannen. Mits je van mij nog geen tabak gekregen hebt dan.

Mijn zus **CHELSY PENDERS** is minstens even creatief, maar dan op haar manier: ze zingt, danst, ambieert een carrière als interieurstylist en kan verdraaid goed tekenen. Alle illustraties in dit boek komen van haar hand, evenals het grafisch ontwerp, met de fruitige kleuren en gelikte tabellen. Zonder haar inzicht waren deze pagina's een stuk saaier, amateuristischer en soms een beetje pijnlijk voor het oog geweest. Voor de afwerking riep ze de Photoshop-*skillz* in van Luca Britti.

En dit is **WARD SCHOONBROOD**, mijn goede vriend. Zelfstandig ondernemer en adviseur met een focus op de creatieve industrie in de regio. Een familiemens met hart voor zijn geboortestad Maastricht. Min of meer eigenhandig sloot Ward zich aan toen ik mijn idee voor een eigen lesmethode aan hem voorlegde: hij tekende onze multimediale aanpak grotendeels uit, en kwam gelijk even met de projectnaam – *Piraten, perziken en p-waarden*. Hij onderhoudt onze website, www.pppwaarden.nl, en is verantwoordelijk voor al ons videomateriaal. "Eigenlijk heeft hij dus niks met het handboek te maken?" Stomme vraag, lezer. Ward hoort er gewoon bij. ☺

Dan nog een paar eervolle vermeldingen als dankbetuigingen. De eerste is voor René Bouman, eigenaar van Boekenplan, de uitgever met wie we dit handboek hebben gerealiseerd. De tweede gaat naar de vakgroep Methodologie en Statistiek van de Universiteit Maastricht. *Piraten, perziken en p-waarden* is een volledig onafhankelijk project, maar zonder de docenten van de vakgroep met hun kundigheid en geduld had ik nooit alles geleerd wat ik jou, beste lezer, nu kan doorgeven. Ten slotte een dikke omhelzing voor alle studenten die ik de afgelopen jaren heb mogen helpen, die me hebben geïnspireerd, gepassioneerd en overgehaald om dit project te lanceren. Jullie zullen hopelijk niet de laatsten zijn.

WAAROM ZOU JE DIT HANDBOEK GEBRUIKEN?

BESTE STUDENT,

Waarschijnlijk ben je afgekomen op de geruchten: geruchten over een handboek dat het anders doet. Een lesmethode die jou begrijpt, die je serieus neemt, en die je inzicht écht kan verbeteren. Statistiek is een vak waar je nooit om hebt gevraagd; je studeert psychologie, gezondheid of een sociale wetenschap, en bent geen expert in de wiskunde. Maar omdat statistiek zo belangrijk is, ook voor jou, kun je er niet omheen. Zal het ooit spannender en concreter worden dan een spelletje van *X*'en en *Y*'en?
Mijn antwoord: ja. Na twee jaar experimenteren, ervaring en creativiteit weet ik dat het mogelijk is: iedereen – en ik overdrijf niet – kan statistiek begrijpen. Voor je ligt mijn aanpak van kraakheldere en complete uitleg met een knipoog, vertaald naar papier, vergezeld door levendige opdrachten die je nieuwe vaardigheden meteen op de proef stellen. Je zult achtbanen testen, bananen tellen, snorren meten en nog veel meer. Het zal je slimmer maken dan je voor mogelijk had gehouden.
Kortom: probeer het, en laat je overtuigen. Bouw mee aan de revolutie. Zie je jezelf als een perzik in statistiek? Nog even en je wordt een piraat!

Groeten,
Vincenzo

BESTE DOCENT,

Misschien slaat u dit handboek open met enige scepsis. Wat moet zo'n jonge snotneus met een eigen lesmethode? Had hij die niet beter kunnen overlaten aan ervaren docenten, die zich door jaren van lesgeven en onderzoek doen hebben ontwikkeld tot experts? Wat bezielt hem eigenlijk, om studenten ertoe te verleiden de aanbevolen literatuur terzijde te leggen en over te stappen op zijn ongevraagde alternatief?
Over de stelling dat ik minder kennis van zaken heb dan een gepromoveerd wetenschapper in de statistiek, kunnen we kort zijn: u hebt ongetwijfeld gelijk. Ik heb veel te bieden, maar ik ben begrensd. Wellicht schuilt daarin een deel van mijn kracht. Ik kan me inleven in de student, die soms bij nul moet beginnen, en vaak geen gevoel voor wiskunde heeft of dit nog niet heeft ontwikkeld. U kunt wel raden dat ik om deze reden heb gekozen voor onorthodoxe casussen en een lichtvoetige toon. De kans bestaat dat u beide niet veel meer dan opsmuk vindt, of zelfs – in het ongunstigste geval – iets wat de student afleidt van de dingen die hij of zij zou moeten leren. Mijn visie is anders. Staat u me toe om het een en ander uit de doeken te doen.
Ten eerste vermoed ik dat *structuur* en *overzicht* een van de belangrijkste pijlers zijn die een student nodig heeft. Los van mijn voorbeelden en schrijfstijl heb ik mijn best gedaan deze te bieden. Het organiseren van nieuwe kennis ervaren de meesten als buitengewoon uitdagend, met name als het vakgebied hen aanvankelijk niet ligt. Door enig voorwerk te verrichten, voorkom ik een hoop frustratie. Elk hoofdstuk begint met scherpe samenvattingen van de besproken theorie in tabellen, die door het handboek heen een duidelijke continuïteit laten zien: onderzoeksdesigns, wiskundige formules,

assumpties van statistische toetsen en stappenplannen voor een analyse staan puntsgewijs vermeld. Het algemene stappenplan voor toetsanalyse, tussen hoofdstuk 6 en 7, biedt verdere ondersteuning bij de ontwikkeling van een vogelvluchtperspectief en kan ook in latere stadia door studenten en jonge onderzoekers gebruikt worden om de statistiek van hun onderzoek in goede banen te leiden. Mogelijk bent u van mening dat ik de studenten te veel werk uit handen neem. In mijn eigen ervaring leren ze niettemin het effectiefst als ze deze structuur ten minste één keer aangereikt krijgen. In contexten waar interactie mogelijk is, zoals een college of les, kan een middenweg bewandeld worden die wat mij betreft het optimum is: de studenten vullen dan samen met de docent de samenvattingen zelf in. Ik adviseer lezers van dit handboek ook om op eigen aantekeningen hetzelfde te doen (zie de volgende bladzijde).

Ten tweede ben ik ervan overtuigd dat een lesmethode statistiek zowel toegankelijk moet zijn als volledig. Toegankelijk in de zin dat iedereen hem moet begrijpen; volledig in de zin dat hij zo min mogelijk stappen overslaat. Grofweg kent het statistiekonderwijs twee uitersten. Het ene uiterste is de aanpak 'neem deze formule en vul maar in'; je bespaart de student de details, maar hij of zij heeft nog steeds geen idee waar het eigenlijk over gaat en krijgt geen kans om echte inzichten te ontwikkelen. Aan het andere uiterste vinden we formele wiskundige definities, waar met abstract taalgebruik, matrices, subscripten en bewijzen een zwaarbewaakt kasteel wordt gebouwd zonder bezoekersingang – en vergeet niet: de student is een bezoeker in het statistische universum, een immigrant die zijn weg nog moet vinden. Beide uitersten bevallen mij niet. Dit handboek is een routekaart die zoveel mogelijk locaties wil aandoen, via paden zonder valkuilen, en met tips voor de fijnproever. Als zich een formule aandient, vertel ik hoe deze in elkaar is gezet; als voor een statistische toets een assumptie geldt, leg ik uit waarom. Voert een probleem echt te ver voor een normale cursus statistiek, dan bied ik in voetnoten en bonusparagrafen de oplossing voor de lezer die tóch tot het gaatje wil gaan. Heel wat studenten (met name universitaire) vinden het namelijk vreselijk om dingen maar te moeten aannemen; ze willen liever door het vuur gaan en hun kennis wortelen in stevige grond – waarna ze die kennis gegarandeerd beter onthouden.

Ten derde stel ook ik altijd de kritische vraag wanneer ik een nieuwe casus verzin: helpt dit voorbeeld of leidt het af? Is het bizarre interessantdoenerij of kan een student er echt iets van leren? Natuurlijk zullen sommigen het voorhoofd fronsen als hun gevraagd wordt wasmachines te testen, cupcakes te proeven en orks te observeren. Waarom komt deze lesmethode niet ter zake? Maar al snel ontdekken diezelfde studenten hoe levendig ze zich de data kunnen voorstellen. Verbanden en effecten ontvouwen zich in hun hoofd, en op dat moment geef ik die verbanden en effecten een statistisch gezicht. De student begint te zien waar dat wiskundige gewauwel voor bedoeld is, wat significantie en *confounding* en interactie betekenen, en slaagt er op het eind in de zaak om te draaien en de moeilijkste vertaalslag te maken van allemaal: van het analyseresultaat naar de conclusie van het onderzoek. De ironie? De student leert niet ondanks, maar *dankzij* de vrolijke voorbeelden, die voor hem of haar oneindig veel tastbaarder zijn dan een variabele X en een variabele Y. En als hij of zij zich later probeert te herinneren hoe multipele regressie ook alweer ging, welke herinnering zou dan eerder bovenkomen? 'O, dat met die X'en!'… of toch: 'O, dat met die piraten!'?

Kortom: *Piraten, perziken en p-waarden* wil bruggen slaan. Een brug tussen totaliteit en toegankelijkheid, een tussen werkelijkheid en wiskunde, en een tussen komedie en kennis. En niet te vergeten: een brug tussen docenten. Zoals ik al ruiterlijk toegaf, is uw expertise in de statistiek groter dan de mijne. Dit handboek is dan ook vanaf de eerste alinea bedoeld geweest om tot samenwerking te komen. Ik denk dat u en ik elkaar kunnen complementeren vanuit onze eigen specialiteit, om zo het statistiekvak eindelijk te beroven van die hardnekkige status van 'moeilijk' en 'saai'. Mocht u voorwaarden zien verrijzen waaronder deze lesmethode een waardevolle aanvulling zou vormen op uw curriculum, aarzelt u dan niet en neem contact op via www.pppwaarden.nl. Ik kijk uit naar ons eerste gesprek.

Hartelijke groeten,
Vince Penders

HOE GEBRUIK JE DIT HANDBOEK?

Piraten, perziken en p-waarden kan in principe fungeren als een op zichzelf staande cursus statistiek. In de praktijk zullen de meeste lezers het waarschijnlijk gebruiken ter ondersteuning van een vak op de hogeschool of universiteit. Daarom heb ik me in deze eerste versie geconcentreerd op het uitleggen van de theorie. Opdrachten om te oefenen vormen een kleiner onderdeel van het geheel, want daarvan krijg je er waarschijnlijk al een groot aantal aangeboden. De bedoeling is echter wel om het aanbod in de toekomst uit te breiden; goede opdrachten zijn leerzaam en voegen daarom veel toe.
Kan ik je alles leren wat je weten moet? Doe je een universitaire bachelor in Psychologie, dan is die kans groot. Ook voor medici bespreek ik een arsenaal aan nuttige technieken. Mis je iets in dit handboek, laat het ons dan weten via www.pppwaarden.nl! We staan te trappelen om de volgende versie nog volledi ger te maken.

BEN JE ZELFSTANDIG BEZIG MET STATISTIEK?

Begin dan gewoon bij hoofdstuk 1 of het onderwerp waarover je op dit moment graag iets wilt leren. Zie voor de rest het volgende kopje.

VOLG JE EEN STATISTIEKVAK OP DE HOGESCHOOL, UNIVERSITEIT OF AVONDSCHOOL?

Ga naar de inhoudsopgave en zoek het onderwerp op dat je moet leren. Kun je het niet vinden, raadpleeg dan eens de lijst *Spraakverwarring* (op de volgende bladzijde) en daarna het deel *Overzichten en kernbegrippen*. Levert dit allemaal geen resultaat op, dan kun je een e-mail sturen naar info@pppwaarden.nl en vragen of ik jouw leerstof ergens in dit handboek bespreek.
Onderwerp gevonden? De voorkennis die je nodig hebt, staat telkens aan het begin van het hoofdstuk vermeld. Lees eerst met een goeie kop koffie of thee de theorie en maak aantekeningen. Schrijf de overzichtsparagraaf over, maar laat de meeste vakjes leeg en probeer deze in te vullen onder het lezen. Maak daarna de opdrachten om je nieuwgeboren kennis op de proef te stellen. De uitwerkingen vind je gratis op de website, www.pppwaarden.nl.
Raak vooral niet ontmoedigd als de opdrachten je niet meteen goed afgaan. Statistiek heeft aandacht en oefening nodig. Soms gaat het een dag later al veel beter, als de nieuwe stof een beetje is bezonken. Veel succes! ☺

HEB JE STATISTIEK NODIG VOOR JE EIGEN ONDERZOEK?

Waarschijnlijk ben jij al bekend met de onderwerpen die ik in dit handboek bespreek. Je bent echter een aantal dingen vergeten, of wilt een paar details controleren. In dat geval is het belangrijk dat je gemakkelijk kunt vinden wat je zoekt. Raadpleeg in elk geval het **Stappenplan toetsanalyse**. Zodra je de statistische analyse hebt gekozen die bij jouw onderzoek past, kun je naar het desbetreffende hoofdstuk gaan. Tussen de regels door vind je ook instructies om SPSS aan te sturen. Moet je daarnaast je kennis opfrissen over een specifiek begrip, dan biedt het deel *Overzichten en kernbegrippen* uitkomst. Ik hoop dat je begeleider onder de indruk zal zijn van jouw staaltje statistiek.

SPRAAKVERWARRING

Het is een irritant, maar onvermijdelijk probleem in een grote wetenschap als die van de statistiek: onderzoekers uit diverse windstreken komen met verschillende namen voor exact hetzelfde ding. Met deze lijst probeer ik te voorkomen dat je docent een begrip of toets bespreekt die ik behandel, zonder dat jij als lezer dat doorhebt. Komt er iets niet voor in de hoofdstuktitels en ook niet bij *Overzichten en kernbegrippen*? Kijk dan eens of je het hier vinden kunt.

Aanname: assumptie
ANACOVA: ANCOVA
Dependent t-test: *paired samples t-test*
Effectmodificatie: interactie
Gebalanceerd design: orthogonaal design
Gepaarde t-toets: *paired samples t-test*
Herhaalde-metingen-ANOVA: *within-subjects*-ANOVA
Least squares regression: lineaire regressie
Matched-pairs t-test: *paired samples t-test*
Mixed ANOVA: *split-plot*-ANOVA
Moderatie: interactie
Multicollineariteit: collineariteit
Onderscheidend vermogen: *power*
Ongepaarde t-toets: *independent samples t-test*
Ordinary least squares (OLS): lineaire regressie
Repeated measures ANOVA: *within-subjects*-ANOVA
Two samples t-test: *independent samples t-test*
Uitbijter: uitschieter
χ^2-toets voor de gelijkheid van k verdelingen: χ^2-kruistabeltoets
χ^2-toets voor onafhankelijkheid: χ^2-kruistabeltoets

OVERZICHTEN EN KERNBEGRIPPEN

De statistiek die in dit handboek aan bod komt, kent vele centrale concepten. Deze komen telkens terug. Dit overzicht bespreekt ze een voor een, en wil dan ook een basis zijn waarop je altijd kunt terugvallen. Gebruik dit deel niet om nieuwe dingen te leren, maar keer er steeds naar terug als je iets bekends vergeten bent. Wanneer in de hoofdstukken een begrip groen is gemarkeerd, word je verwezen naar deze proloog.

ASSUMPTIES

Geïntroduceerd in hoofdstuk 7

Elke statistische toets heeft een aantal zogenaamde assumpties. Wat zijn dat? Abstract gezegd: iets waarvan de toets uitgaat bij zijn berekening van het toetsresultaat, zodanig dat de uitkomst van de toets alleen ergens op slaat als dat uitgangspunt correct was.
¿Que? zullen de meeste lezers nu denken. Gelukkig kan dit concreter, want niet alleen statistische toetsen hebben assumpties. Ook een gewone keukenweegschaal heeft er een: zwaartekracht. Meer specifiek: de aanname dat de weegschaal zich bevindt op de planeet Aarde, waar een bepaalde zwaartekracht geldt. We zouden met deze weegschaal ook prima een krop sla kunnen wegen als we ons op de maan bevonden – er verschijnt heus wel een aantal kilogrammen op het display, alleen klopt dit getal voor geen meter. Ander voorbeeld: de klok heeft eveneens de assumptie dat de gebruiker zich op de planeet Aarde bevindt. Meer specifiek: dat er een 24-uursritme geldt. Als we op de planeet Mars een horloge gebruiken, is het de ene keer om 16:00 uur klaarlichte dag en de andere keer pikdonker. Oftewel, de waarde die we aflezen slaat nergens op.
Kortom: een assumptie is een situatie waarvan de toets uitgaat bij zijn berekeningen. Als een assumptie geschonden wordt, is de uitkomst van de toets meestal onjuist.
Sommige schendingen van assumpties maken de toets echter niet meteen onbruikbaar; als er aan een speciale voorwaarde voldaan is, wordt de toets **robuust** tegen schending. Let in de hoofdstukken op de kleurcodes: een rood gedrukte assumptie kan niet worden gepasseerd (of niet zonder aanpassingen van de data, het statistische model en/of de toets); een blauw gedrukte assumptie wel.

BETROUWBAARHEIDSINTERVAL

Geïntroduceerd in hoofdstuk 6

Om een beeld te krijgen van een populatieparameter, bijvoorbeeld een gemiddelde, kunnen we onder meer een betrouwbaarheidsinterval opstellen. Ik beperk me hier tot een snelle herhaling. Een 95%-betrouwbaarheidsinterval is zo opgesteld dat 95% van alle intervallen die ik maak (als ik heel vaak dezelfde steekproef opnieuw zou trekken) de parameter van de populatie bevat. Als bijvoorbeeld de gemiddelde snorlengte van Nederlandse mannen in de populatie gelijk is aan 5 millimeter, zal 95% van alle betrouwbaarheidsintervallen de waarde 5 bevatten. We kunnen dan ook zeggen (bijvoorbeeld): 'In mijn steekproef heb ik een gemiddelde gevonden van 5,27 millimeter; ik weet dat dat niet precies het

populatiegemiddelde is, want het is maar een steekproef, maar het populatiegemiddelde ligt waarschijnlijk tussen de 2,96 en 7,58 (mijn betrouwbaarheidsinterval).' Vervolgens mogen we eventueel ook zeggen: '3 is dus een aannemelijk populatiegemiddelde, dus ik zou de nulhypothese $\mu = 3$ niet verwerpen. Maar 8 is géén aannemelijk populatiegemiddelde, dus ik zou de nulhypothese $\mu = 8$ wel verwerpen.' Zie voor een uitgebreidere uitleg hoofdstuk 6.
Opmerking: het is naar mijn mening al heel wat als je weet dat de populatieparameter waarschijnlijk tussen de grenswaarden van het interval ligt. Ook gaan we in latere hoofdstukken geen intervallen meer met de hand construeren.

BONFERRONI-CORRECTIE

Geïntroduceerd in hoofdstuk 10

Soms bekijken we met meerdere statistische toetsen achter elkaar of proefpersonen verschillen op een afhankelijke variabele Y. Neem bijvoorbeeld het aantal krakelingen dat klanten bij een banketbakker kopen, uit hoofdstuk 10: schaffen ze even veel krakelingen aan bij drie verschillende reclamebordjes? We zouden de drie condities allemaal een op een kunnen vergelijken, met drie aparte t-toetsen:

- 'Krakelingen' versus 'Krrrakelingen';
- 'Krakelingen' versus 'Ambachtelijke roomboterkrakelingen';
- 'Krrrakelingen' versus 'Ambachtelijke roomboterkrakelingen'.

Echter, aan die aanpak kleeft een probleem. Elke keer dat we een statistische toets uitvoeren, hebben we kans een Type I-fout te maken. We voeren nu drie unieke toetsen uit. De kans op minstens één Type I-fout is daardoor veel groter dan 5%, namelijk bijna 15%. We noemen deze kans de ***familywise error rate***: de mate waarin er een fout kan worden gemaakt over de 'familie' van alle t-toetsen.
De Bonferroni-correctie is een basale methode om de *familywise error rate* terug te brengen naar 5%. Deze correctie bestaat eruit dat we het significantieniveau delen door het aantal toetsen: in dit geval $\alpha = \frac{0{,}05}{3} = 0{,}017$ (afgerond). Het significantieniveau wordt dus verlaagd, waardoor het moeilijker wordt de nulhypothese te verwerpen, en de kans afneemt dat dit onterecht gebeurt (Type I-fout). P-waarden moeten nu worden vergeleken met het aangepaste significantieniveau. Is een p-waarde nog steeds lager dan dit nieuwe significantieniveau, dan hebben we een significant resultaat.
Ook mogelijk is de p-waarde te vermenigvuldigen met het aantal toetsen, opdat we die weer kunnen vergelijken met het oorspronkelijke significantieniveau. Dit levert exact dezelfde resultaten op. Waarom?

$$p \quad vs. \quad \frac{\alpha}{aantal\ toetsen}$$

Vermenigvuldig links en rechts met het aantal vergelijkingen, en je hebt:

$$p * aantal\ toetsen \quad vs. \quad \alpha$$

Effectief – en ook in SPSS – is de Bonferroni-correctie dus $\boldsymbol{p * aantal\ toetsen}$: in dit geval $p * 3$.

CENTREREN
>>> Datatransformatie

COHENS *d*

Geïntroduceerd in hoofdstuk 6

Een indicator van effectgrootte. Neem nog eens de modecampagne voor langere snorren uit hoofdstuk 6. In dit hoofdstuk bleek dat de gemiddelde snorlengte van Nederlandse mannen was toegenomen met 2 millimeter (twee maanden na de start van de campagne). Is dit een beetje indrukwekkend? Als individuen al heel erg verschillen in snorlengte (kijk eens naar de standaardafwijking), stelt een gemiddelde stijging van 2 millimeter niet veel voor.

Cohens d kan steeds op een andere manier berekend worden, afhankelijk van het onderzoeksdesign en de betrokken variabelen; zie steeds het betreffende hoofdstuk. Het principe is echter altijd hetzelfde: we delen het gevonden effect door de standaardafwijking van de individuen. Enige richtlijnen voor de uitkomst (niet heilig) zijn:

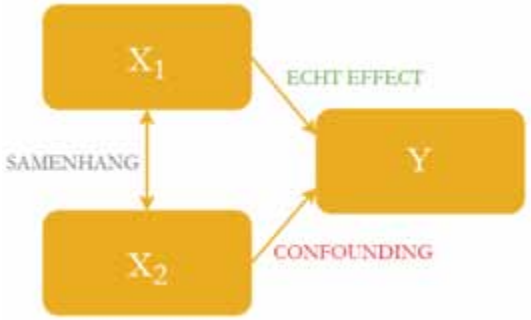
- 0,2 beschouwen we als een klein effect;
- 0,5 vinden we een medium effect;
- 0,8 noemen we een groot effect.

Ook al is een bepaald effect significant (waarmee we aantonen dat het zich ook voordoet in de populatie), dat maakt het nog niet tot een indrukwekkend effect. Is het effect vrij groot of juist erg klein? Deze vraag is minstens even belangrijk.

CONFOUNDING

Geïntroduceerd in hoofdstuk 11 en 20

Interactie en *confounding* worden vaak door elkaar gehaald. Toch zijn het twee compleet verschillende fenomenen; kijk goed naar de gele illustraties. In hoofdstuk 8 presenteerde fabrikant Rodent vol trots zijn ActiBleach-formule als effectief: wie hiermee poetste, kreeg gemiddeld wittere tanden dan de personen die poetsten met een anonieme concurrent. Maar stel dat wij nu de data van Rodent nader inspecteren, en zien dat de ActiBleach-gebruikers relatief vaak lang hun tanden poetsten, en de concurrent-gebruikers juist relatief vaak kort. Dat de ActiBleach-gebruikers de witste tanden hebben, kan dus óók zijn gekomen doordat zij langer poetsten. Op dat moment is poetstijd een ***confounder***.



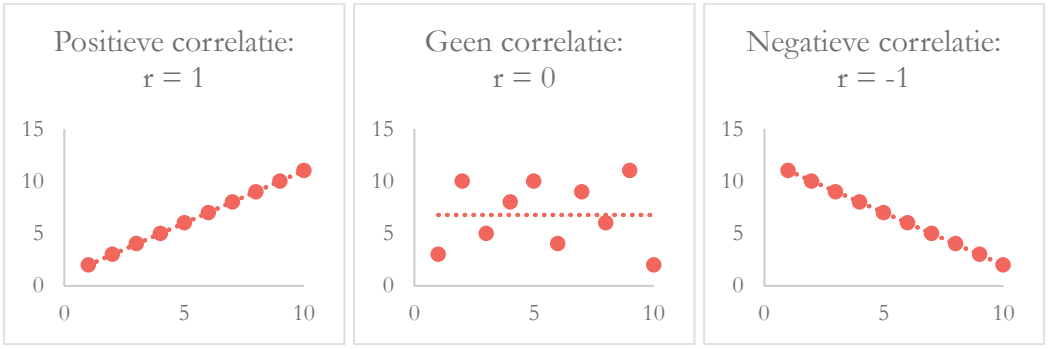
We spreken van *confounding* als de effectmeting van X_1 op Y verstoord wordt door een derde variabele, X_2, die met X_1 en Y samenhangt. (Tegelijk kan X_1 trouwens ook een *confounder* zijn voor het effect van X_2 op Y.) We moeten voor deze *confounding* corrigeren, en proberen het 'pure' effect van de variabelen te meten. SPSS heeft daar bepaalde methoden voor.

Kortom: *confounding* is een meetprobleem. Interactie (zie verderop) is dat niet! Door *confounders* kunnen we niet goed meten hoe het effect van een variabele eruitziet in de populatie. Als zich echter een significante interactie voordoet, is het effect van een variabele – ook in de populatie – simpelweg niet altijd hetzelfde.

CORRELATIE

Geïntroduceerd in hoofdstuk 3

Als we het lineaire verband willen beschrijven tussen twee kwantitatieve variabelen, kunnen we gebruikmaken van een correlatiecoëfficiënt. Dit is een gestandaardiseerde maat en heeft daarom altijd een waarde tussen -1 en 1. Zie ook de volgende pagina. Als r_{XY} ongeveer 0 is, is het verband tussen de twee variabelen erg zwak of zelfs afwezig. Hoe verder r_{XY} afwijkt van 0, des te sterker is het verband. Hij kan zoals gezegd minimaal -1 zijn; in dat geval spreken we van **perfecte negatieve samenhang**. Als de correlatiecoëfficiënt 1 is (maximaal), is er **perfecte positieve samenhang**.



Je zou de lineairecorrelatiecoëfficiënt zelf kunnen berekenen, maar dan kom je er wel achter waarom we tegenwoordig een computer het werk laten doen. 😐 De meest gebruikelijke formule is:

$$r_{XY} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})/(N-1)}{s_X s_Y}$$

Deze geldt voor steekproeven. Voor een uitwerking kun je kijken in hoofdstuk 3.

DATATRANSFORMATIE

Geïntroduceerd in hoofdstuk 1

Wie bijvoorbeeld de lengte van een aantal bananen heeft gemeten in meters, kan ervoor kiezen om alle scores om te zetten in centimeters (dat lijkt mij ook wat handiger in dit geval). Hij of zij voert dan een datatransformatie uit: alle data worden omgezet naar andere waarden, maar hun betekenis blijft hetzelfde. De lengte van de bananen verandert tenslotte niet als ik van 0,2 meter naar 20 centimer ga. Het simpelste soort datatransformaties zijn **lineaire transformaties**: centreren, standaardiseren (naar z-scores) en multipliceren. Zie hiervoor de laatste paragraaf van hoofdstuk 1.

EFFICIËNTE SCHATTER
>>> *Schatter*

HOOFDEFFECT
>>> *Interactie, hoofdeffecten en simpele effecten*

HYPOTHESETOETSING

Geïntroduceerd in hoofdstuk 6

Een cruciaal onderdeel van de statistiek in dit handboek is het toetsen van hypothesen. Zorgt ActiBleach-tandpasta voor wittere tanden dan een concurrerend merk? Dat kunnen we uittesten op een beperkte groep personen (een **steekproef**), maar als de ActiBleach-formule in deze steekproef tot wittere tanden lijkt te leiden, kan dit ook door toeval komen: doordat de groep van ActiBleach-gebruikers toevallig wat meer mensen met een beter gebit bevat. Sterker nog: door toeval zullen de twee steekproefgemiddelden vrijwel nooit precies hetzelfde zijn. Dat betekent dat zich in de steekproef bijna altijd wel een effect van de tandpasta voordoet. We moeten kunnen uitsluiten dat dit puur door toeval komt.
Voor dit doel zijn alle statistische toetsen ontwikkeld. Ze voltrekken zich in drie stappen.

I. HYPOTHESEN OPSTELLEN

Elke toets heeft een **nulhypothese**: zijn uitgangspunt. Dit uitgangspunt is telkens dat de onderzoeksfactor geen effect heeft, en dat alle effecten die we in steekproeven kunnen tegenkomen,

dus veroorzaakt worden door toeval. In ons voorbeeld luidt de nulhypothese dan ook: 'het soort tandpasta heeft geen effect op de witheid van je tanden'. In een formule:

$$H_0: \mu_{ActiBleach} = \mu_{concurrent}$$

Oftewel, in de populatie hebben mensen die ActiBleach gebruiken gemiddeld even witte tanden als gebruikers van de concurrerende tandpasta.
Ons streven is dat we deze nulhypothese kunnen verwerpen. Voor het geval dat dit lukt, hebben we ook een **alternatieve hypothese**. Die kan bijvoorbeeld zijn:

$$H_A: \mu_{ActiBleach} \neq \mu_{concurrent}$$

Oftewel, als we de nulhypothese verwerpen, aanvaarden we de alternatieve. We concluderen dan dat de ActiBleach-tandpasta niet tot even witte tanden leidt als de concurrerende formule; er is wél een effect van de tandpasta.[1]
Maar zover zijn we nu nog niet. We vertrekken vanuit de nulhypothese.

II. KANS BEREKENEN

We trekken de steekproef, en we krijgen twee steekproefgemiddelden: één gemiddelde witheid van de ActiBleach-groep, en één van de concurrentgroep. Het verschil tussen de gemiddelden blijkt gelijk aan 1 punt (op een schaal van 0 tot 20). Hoe vaak zouden we zo'n groot verschil meten, als er in werkelijkheid – dus in de populatie – helemaal geen verschil is? We kunnen dit uitrekenen; in de introductiehoofdstukken doen we dat nog veel met de hand, maar op een gegeven moment zullen we het meeste aan SPSS overlaten. Stel, de uitkomst is 0,42. Dat wil zeggen: als de twee groepen op populatieniveau niet verschillen qua gemiddelde witheid van hun tanden, zullen we 42% van de keren dat we een steekproef trekken, tóch een verschil vinden van 1 punt of groter.
Die 0,42 noemen we de **p-waarde: de kans op het gevonden steekproefeffect – of een nog groter effect – <u>indien de nulhypothese waar is</u>**.
Nu, een steekproefeffect dat 42% van de keren zou voorkomen is niet bijzonder, of wel soms? Dit effect past prima bij het plaatje van de nulhypothese, en dus kunnen we niet aantonen dat ons gevonden resultaat meer is dan puur toeval.
Stel dat de situatie ietsje anders is. We meten een verschil van 2,5 punt. De p-waarde blijkt gelijk aan 0,11. Opvallend. Een steekproefeffect van deze omvang (of een nog grotere) zou slechts 11% van de keren voorkomen, als ActiBleach- en concurrent-gebruikers op populatieniveau gemiddeld even witte tanden hadden. Het is dus erg toevallig dat wij zo'n zeldzaam verschil in onze steekproef meten. Maar goed, we moeten het maar geloven.
Stel, ten slotte, dat de ActiBleach-groep in onze steekproef gemiddeld 4 punten wittere tanden heeft dan de concurrent-groep. De bijbehorende p-waarde blijkt 0,003. Ho even… dus slechts 0,3% van de keren zouden we zo'n verschil tussen deze twee groepjes proefpersonen meten, als alle ActiBleach-gebruikers samen gemiddeld niet verschillen van alle concurrent-gebruikers? Dan hebben wij een enorm zeldzame steekproef getrokken. Te zeldzaam. We kunnen letterlijk zeggen: <u>"Dit is geen toeval meer!"</u> Nee, veel waarschijnlijker is het dat we helemaal niet zo'n zeldzame steekproef hebben getrokken – en dat de tandpasta wel degelijk een effect heeft op de witheid van je tanden. Op dit moment is het toetsresultaat **significant**. Het geeft aan *('signifies')* dat de nulhypothese niet juist is.
Er is een grenswaarde waarbij we de p-waarde zo klein vinden dat we hem significant noemen. Deze grenswaarde noemen we het **significantieniveau**, aangegeven met $\boldsymbol{\alpha}$. Een onderzoeker kiest het significantieniveau in principe zelf, maar in de psychologie gaan we standaard voor 0,05 (5%).

III. CONCLUSIE TREKKEN

Het beslismodel ziet er dus uiteindelijk zo uit als op de volgende bladzijde.
Vergeet niet de conclusie van het beslismodel altijd naar de inhoud van het onderzoek te vertalen. In dit geval: het is aangetoond dat mensen die ActiBleach gebruiken van Rodent, gemiddeld <u>niet</u> even witte tanden hebben als mensen die poetsen met een anonieme concurrent.

[1] In de praktijk werken we bijna altijd met tweezijdige toetsen. Wil je het onderscheid weten tussen eenzijdig en tweezijdig toetsen? Zie dan hoofdstuk 6.