Hoofdstuk 5

5.1

Causaliteit is een groot probleem in statistiek en filosofie. Een reden hiervoor is dat er niet een exacte definitie is van het woord causaliteit. Het word meestal geloofd als dat causaliteit kan bepaald wanneer er aan deze drie criteria is voldaan: associatie, richting van invloed en isolatie. We zullen deze drie in meer detail bespreken:

Associatie

Het eerste wat we leren is dat correlatie niet betekend dat er ook causaliteit is. Maar je moet ook onthouden dat wanneer er causaliteit is dit betekend er ook correlatie is. Wanneer twee variabelen een causaal verband hebben, moet een verandering in de een, een verandering in de andere veroorzaken. Daarom een statistische associatie (correlatie) is nodig om te kunnen zeggen dat er causaliteit is.

Richting van causaliteit

Wanneer twee variabelen (A en B) geassocieerd zijn met elkaar kunnen er drie reden zijn voor dit:

- Het kan dat A is de oorzaak van B
- Het kan zijn dat B is de oorzaak van A
- Een andere variabele, C is de oorzaak van de beide A en B.

Dus wanneer er een correlatie is tussen twee variabelen, is het niet duidelijk welke causale mechanisme hiervoor heeft gezorgd. Dus we weten niet de richting van de causaliteit. Maar hoe kunnen we zeggen of A, B veroorzaakt of dat B, A veroorzaakt? Het antwoord is dat we altijd ervan uit gaan dat oorzaak eerst kom en daarna pas het effect. Wanneer A veroorzaakt B, zal een verandering in A een verandering in B zullen veroorzaken na een bepaalde tijd. Dus we moeten dit kunnen laten zien, welke eerste is. Dut we verandering van de afhankelijke variabele moet worden geobserveerd na een verandering in de onafhankelijke variabele. Dit tijd interval tussen de oorzaak en het effect kan erg verschillen. Dit idee van temporale priority is ook aanwezig in het design van een experiment, omdat de manipulatie van de onafhankelijke variabele altijd eerst moet gebeuren voordat de afhankelijke variabele word gemeten. Dit kan helaas niet worden gedaan in niet experimenteel of cross-sectie onderzoek, waarbij alle gegevens meestal worden verzameld op 1 punt. Dan moeten we dus gebruik maken van zogenaamde mental experimenten. Dit zijn beslissingen over de richting van de causaliteit gebaseerd op theorie, voorafgaand onderzoek, en gezond verstand.

Isolatie

Om zeker te weten dat een onafhankelijke variabele, A de oorzaak is van de afhankelijke variabele B, moeten we afhankelijke variabele (B) isoleren om te zorgen dat deze niet door andere dingen word beïnvloed dan A. Meestal kan deze isolatie niet helemaal worden bereikt, maar wel een zogenaamde pseudo-isolatie. Experimentele controle kan worden gebruikt om de onafhankelijke variabelen te isoleren, in niet experimentele onderzoeken kan dit niet. Dit moet worden gedaan op een andere manier. We weten al dat de regressie hellingen het effect van de onafhankelijke variabele op de afhankelijke variabele laten zien, terwijl ze de andere effect van de andere onafhankelijke variabelen gelijk houden.

Dus de regressie modellen kunnen worden gebruikt om dit te bekijken.

Om deze verschillende criteria bij elkaar te voegen, moeten we niet alleen naar de statistische analyse kijken, maar ook naar theorie. Wanneer psychologen gegevens verzamelen willen ze zeker weten dat deze nauwkeurig zijn. Theorie heeft een belangrijke functie in een succesvol gebruik van de regressie. Een serie van verschillende bevindingen, die allemaal correleren kunnen bewijs zijn voor een causale relatie. Deze serie van verschillende bevindingen word de signature van dat proces genoemd.

5.2

We zullen nu gaan kijken naar het effect van de steekproefgrootte op de regressie analyse. Het idee is vooral dat hoe groter de steekproefgrootte des te beter. De standaardfout van het gemiddelde is gelijk aan:

$$se(\bar{x}) = \sqrt{\frac{sd^2}{n}} \qquad (50)$$

Uit deze formule kan je opmaken dat wanneer de steekproefgrootte groter wordt, de noemer ook groter wordt, en dus zal de standaardfout kleiner worden. Dit betekent dat de geschatte waarden voor de parameter preciezer zullen worden. Daarnaast zal een kleinere standaardfout de kans op het vinden van een significante waarde vergroten. Maar er zijn ook problemen met steekproeven die te groot zijn. Want het zoeken van meer gegevens kost veel tijd. Daarnaast zijn ook ethische besturen meer bezig met het onderzoeken van de steekproef groottes. Zij zeggen dat deelnemers hun tijd opgeven in de hoop dat er iets goeds mee wordt gedaan.Er zijn twee manieren om een goede steekproefgrootte te bepalen. De eerste zijn vuistregels. Dit zijn makkelijke regels. De tweede is de power analyse. We zullen beide methoden bespreken.

Vuistregels zijn meestal erg makkelijk. Green heeft een methode van vuistregels gemaakt om een minimale steekproefgrootte te bepalen. Hij zegt dat een minimale steekproefgrootte groter moet zijn dan 50 +8k, waarin k het aantal onafhankelijke variabelen is. Ook heeft hij gezegd wanneer je een significantie toets wil uitvoeren op de regressie hellingen, de steekproefgrootte groter moet zijn dan 104 + k. Het nadeel van deze regels is dat ze de verwachte effectgrootte of gewilde power van de toets niet meenemen. Dus deze regels missen generaliteit. Om een power analyse te gebruiken hebben we de volgende informatie nodig:

- De waarde van alpha. Deze waarde is meestal 0.05 (5%). Wanneer alpha groter is wordt de kans dat we een significant effect vinden groter, maar tegelijkertijd wordt ook de kans op het vinden van een onecht resultaat groter. De kans op een type I fout is gelijk aan de waarde van alpha.
- De effectgrootte van de populatie waar we in geïnteresseerd zijn. De effectgrootte in een meervoudige regressie (multiple regression) is gelijk aan R2. Hoe groter de effectgrootte, des te groter de kans om het te vinden. Maar wanneer de effectgrootte is heel klein, dan zal het vinden van het ook niet handig zijn. De effectgrootte kan bepaald worden op drie manieren:
    - Effect gebaseerd op werkelijke kennis
    - Baseer de schatting van de effectgrootte op voorgaand onderzoek
    - Gebruikt bepaalde regels on de verwachte effectgrootte te bepalen. Cohen heeft waarden van R2 bepaald die de hoeveelheid van de effectgrootte aangeven. Wanneer R2 =0.02 is er een klein effect, wanneer R2=0.13 is er een gemiddeld effect en wanneer R2 = 0.26 is er een groot effect.

Een geschikt level van de power moet worden bepaald. De power is de kans op het vinden van een resultaat gegeven dat het effect bestaat in de populatie. Een regel is dat de power wordt gezet op 0.80 (80%). Dat betekent dat er een 80% kans is op het vinden van een significant resultaat wanneer er een effect is in de populatie. De kans op een type II fout is gelijk aan 1-power, dus 1-0.80=0.20 (20%).

Uit deze informatie kunnen we de benodigde aantal deelnemers bepalen. Programma's zoals G*Power kunnen worden gebruikt om dit te bereken. In G*Power worden grafieken gemaakt, welke verschillende aantallen van deelnemers laten zien. Uit deze grafieken kan je conclusies trekken, maar je kan niet een nauwgezette power berekening mee doen (zie de grafieken op pagina 122-125.)

5.3

We zullen nu gaan kijken naar collineariteit, ook wel multicollineariteit genoemd. Collineariteit refereert naar de grootte van de correlaties tussen de onafhankelijke variabelen in een regressie. Het komt voor omdat twee (of meer) onafhankelijke variabelen correleren. Dit betekend dat het moeilijke is voor de regressie berekening om te bepalen welke van de variabelen echt belangrijk is, eentje van de twee, of misschien beiden. Dus de onzekerheid (standaardfouten) zal vergroten en ook de onnauwkeurigheid (helling coëfficiënten) zullen vergroten. Dus we kunnen niet bepalen welke variabele belangrijk is voor het resultaat. De regressie berekeningen houden rekening met deze onzekerheid en hebben grotere standaardfouten. Dus meer formeel, wanneer er correlatie is tussen twee onafhankelijke variabelen gelijk is aan 1 (of dichtbij 1) of wanneer de multiple correlatie tussen elke onafhankelijke variabele 1 (of dichtbij 1) is, is er perfecte of complete Collineariteit.

Er is een aanname van het regressie model, dat er geen perfecte collineariteit mag optreden, en wanneer dit wel zo is, zullen de meeste statistische computerprogramma's stoppen en een fout laten zien. Maar perfecte collineariteit komt maar weinig voor in echte gegevens. Wanneer dit wel gebeurd, is het hoogstwaarschijnlijk dat er twee of meer onafhankelijke variabelen bij elkaar zijn gevoegd om zo een extra variabele te gebruiken.

Meestal wanneer er collineariteit is is het hoog genoeg om problemen te veroorzaken, maar niet hoog genoeg om de aanname niet waar te maken.

Wanneer je een regressie analyse hebt waarbij de regressie coëfficiënten niet significant zijn, maar de totale regressie wel significant is, moet je onthouden dat collineariteit hier een rol kan hebben gespeeld. Het kan namelijk gebeuren dat de regressie analyse weet dat een groot deel van de variatie kan worden uitgelegd door de onafhankelijke variabelen, maar wee niet wat de grootte van de geschatte parameters is bij welke onafhankelijke variabele. Wanneer je denkt dat er problemen met collineariteit zijn, zijn er verschillende mogelijkheden om te bepalen hoe erg dit probleem is.

Je kan je matrix van de correlaties bekijken. Wanneer een variabele erg correleert met twee andere variabelen, kan die laten zien dat er collineariteit is. Maar een lage correlatie betekend niet meteen dat er geen probleem is. De correlatie laat ons zien hoe veel variatie de twee variabelen delen. De waarde die we willen weten is het deel van de variatie van elke onafhankelijke variabele welke is gedeeld met alle andere onafhankelijke variabelen. We kunnen dit vinden door een meervoudige regressie analyse uit te voeren. We moeten een aparte regressie analyse doen voor elke onafhankelijke variabele. Dit kost veel werk. Maar, de meeste statistische software geeft je ook andere waarden die we kunnen gebruiken.

Dit zijn tolerance en de variance inflation factor (VIF). Tolerance is een uitbreiding van de R2. De tolerance van een onafhankelijke variabele is de hoeveelheid van de onafhankelijke variabele die niet kan worden voorspeld door de andere onafhankelijke variabele in de regressie analyse.

De waarde van R (in meervoudige correlatie) is gelijk aan de waarde van r (bivariate correlatie), en de tolerance kan daarom via r worden berekend. De waarde van tolerance ligt tussen 0-1. Een tolerance van 0 voor een variabele betekend dat deze variabele is helemaal voorspeld door de andere onafhankelijke variabele. Dit is perfecte collineariteit. Wanneer een variabele een tolerance heeft van 1, betekend dit dat de variabele helemaal niet correleert met de andere onafhankelijke variabelen.

De variance inflation factor (VIF) ligt dichtbij tolerance. Wanneer er meer dan twee onafhankelijke variabelen zijn berekenden we de VIF als volgt:

$$VIF = 1/tolerance \qquad (110)$$

Deze waarde laat zien met hoeveel de standaardfout van de variabele is toegenomen door collineariteit. De toename in de standaardfout is gelijk aan de wortel van de VIF.

Maar wat moet je doen wanneer je collineariteit in je gegevens vind? Het beste is, wanneer collineariteit een groot probleem is, om je oude gegevens weg te doen, en nieuwe gegevens te verzamelen. Dit is veel werk, maar er zijn ook andere opties:

- Meer gegevens verzamelen. Het is niet een grote verbetering. Collineariteit zorgt voor een toename in de standaard fouten. Een toename in de steekproefgrootte heeft kleinere standaard fouten, dus een meer gegevens, zal een beetje helpen tegen de effecten van collineariteit. Maar het helpt niet wanneer je perfecte collineariteit hebt.
- Verwijder of combineer variabelen. Wanneer variabelen een hoge correlatie hebben betekend dit dat ze dezelfde dingen meten, en dat de informatie van deze variabelen (deels) overbodig is. Wanneer je veel onafhankelijke variabelen hebt, kan je dit aantal verkleinen door principal components analysis (PCA). Dit lijkt op een factor analyse, omdat ze beiden groepen maken van de originele variabelen. Dit zorgt voor minder gecorreleerde factoren.
- Stepwise entry. Dit is een vorm van hierarchical regressie. Je kan het gebruiken wanneer de collineariteit een probleem geeft in het selecteren van de variabelen voor de analyse.
- Ridge regressie. Wanneer de collineariteit zo groot is dat de regressie procedure niet kan door gaan, is een mogelijke oplossing een ridge regressie. Dit is een complexe methode, het is ook moeilijk om te interpreteren dus het word haast nooit gebruikt.

5.4

Deze paragraaf gaat over het “one-factor repeated measures ANOVA (“ORMA”)” model (ORMA is geen officiële afkorting, maar hier handig om te gebruiken).

Laten we ten eerste de karakteristieken van dit model bespreken: Het ORMA model maakt het mogelijk om twee of meer metingen te analyseren, en vormt zo een uitbreiding van de “dependent t-test”. Het “repeated” deel van ORMA betekent dat elk subject reageert op elk level van de factor A (dit wordt ook wel “within-subjects design” genoemd). Op deze manier functioneren de subjecten als hun eiegen controle, omdat er zo rekening gehouden worgt met de individuele verschillen. De consequentie hiervan is, echter, dat de scores vand e subjecten niet onafhankelijk zijn over de verschillende levels van de factor A.

Vanwege de subjecten en de variaties veroorzaakt door de interactie tussen A en de subjecten in ORMA, wordt de "residual variation" verder ontbonden in gewone variatie ("variation"). Dit betekent een vermindering in de residuensom ("residual sum of squares"), een sterker model, en meer precisie in de schatting van de effecten op A (what dus betekent dat er minder subjecten nodig zijn).

De ORMA is een gemixt model. Dit is vanwege het feit dat de subject-factor een "random" effect is, terwijl de A-factor over het algemeen een vast ("fixed") effect is. De ORMA kan ook gezien worden als een speciaal geval van de "two-factor mixed-effects design", maar dan met maar één subject per cel (n=1).

De meer negatieve aspectenn van de ORMA zijn onderanderen dat er gevaar is can "caryover effects" van één level van A naar een ander, vanwege het feit dat elk subject op elk level van A reageert. Deze effecten kunnen geminimaliseerd worden door (1) de volgorde van toedoenen van de levels van A te compenseren op zo'n manier dat elk subject niet delfde volgorde van de levels van A ontvangt; (2) tijd te laten passeren tussen het toedoenen van de levels; of (3) soortgelijke subjecten te koppelen/blokkeren, met de assumptie dat de subjecten binnen een block willekeurig aan een level van A zijn toegewezen (ook wel: "randomized block design)".

De lay-out van het ORNA model is als volgt:

	Level of Factor A (Repeated Factor)				
Level of Factor S	1	2	…	J	Row mean
1	Y11	Y12	…	Y1J	$\bar{Y}$1.
2	Y21	Y22	…	Y2J	$\bar{Y}$2.
….	…	…	…	…	…
n	Yn1	Yn2	…	YnJ	$\bar{Y}$n.
Column mean	$\bar{Y}$.1	$\bar{Y}$.2	…	$\bar{Y}$.J	$\bar{Y}$..

Het kan hier ook gezien worden dat het ORMA model een vorm van het "two-factor model" is, maar dan met maar één observatie per cel.

Kollommen laten de levels van A zien, en de rijen de subjecten (factor S). Kolommen representeren hier dus de verschillende metingen. De subject gemiddelde worden hier afgebeeld, maar worden zelden gebruikt.

De formule voor dit model is as volgt (geschreven in termen van "population parameters"):

$Yij = \mu + \alpha j + si + (s\alpha)ij + \varepsilon ij$

Yij is de geobserveerde score van de responsvariable voor individu i die reageerd op level j van factor A. μ staat voor het algemene populatie gemiddelde. α is het vaste effect for level j van factor A. si is het "random effect" voor subject i van de subject factor. $(s\alpha)ij$ is de interactie tussen subject i en level j. εij is de "random residual error" voor individu i in level j.

Als het aankomt op assumpties, is de ORMA weer erg gelijkend op het "two-factor mixed-effects model". Net zoals in dit model gaan de assumpties van ORMA vooral over de distributie van de "random effects" en de scores van de responsvariabele. De ORMA heeft wel twee nieuwe assumpties die het "two-factor mixed-effects" model niet heeft:

- “Compound symmetry”: Deze assumptie stelt dat de co-variantie tussen de subject scores constant blijft over de levels van de herhaalde factor A. Deze assumptie wordt vaak geschonden in ANOVA, zeker wanneer de factor A tijd is (sinds er dan continue verandering is kunnen de co-varianties niet constant zijn). Als deze assumptie geschonden wordt zijn er drie opties: (1) de levels van A limiteren tot (a) degene die de assumptie bevestigen, of (b) twee herhaalde metingen; (2) gebruik aangepaste (“adjusted”) F-tests; of (3) gebruik MANOVA (“multiple analysis of variance”), die wel minder krachtig kan zijn, maar geen last heeft van deze assumptie. De eerste maatregel tegen “carryover effects” kan ook problemen met deze assumptie minimaliseren.
- “Sphericity”: Deze asusmptie stelt dat voor elk paar factor levels de variantie van de verschillende scores hetzelfde is. Dit is de noodzakelijke en voldoende conditie voor de validiteit van de F-test (“compoun symmetry” is wel voldoence maar niet noodzakelijk).

De ANOVA tabel voor dit model is als volgt:

Source	SS	df	MS	F
A	SSA	J – 1	MSA	MSA/MSSA
S	SSS	n – 1	MSS	
SA	SSSA	(J – 1)(n – 1)	MSSA	
Total	SStotal	N - 1		

Wat betreft variatie bronnen (“sources of variation”) is de ORMA weer erg gelijkend op het twee-factor model, met de uitzondering dat ORMA geen binnen-celse variatie heeft. Zoals de tabel laat zien zijn de variatie bronnen hier: A (de herhaalde meting), S (de subjecten), SA (de interactie tussen A en S), en het totaal. Hoewel dit laat zien dat we drie “main squares”-termen kunnen berekenen, is er alleen een R-ratio resultaat voor factor A. Dit laat zien dat er geen passende fout-term (“error term”) is foor het subject-effect, en dat dit niet getest kan worden.

De “sum of squares” (SS) moet overwogen worden. Het ontbinden van de SS wordt als volgt gedaan:

SStotaal = SSA + SSS + SSSA

De verwachte “mean squares” (MS) zijn belangrijk voor de formatie van de juiste F-ratio. Maar, de verwachte MS is afhankelijke van of the null hypothese (gemiddelde is hetzelfde voor elke van de metingen) waar is of niet. Als H0 waar is dan zijn de verwachte MS:

- $E(MSA) = \sigma\varepsilon 2$
- $E(MSS) = \sigma\varepsilon 2$
- $E (MSSA) = \sigma\varepsilon 2$

Hier is $\sigma\varepsilon 2$ de populatie variantie van de fout residuen (“residual errors”).

Als H0 niet waar is dan zijn de verwachte MS:

$$E\left(MS_A\right) = \sigma_\varepsilon^2 + \sigma_{s\alpha}^2 + n\left(\frac{\sum_{j=1}^{J}\alpha_j^2}{J-1}\right)$$

E(MSS) = σε2 + Jσ s2

E(MSSA) = σε2 + σsα2

Hier is σs2 de variatie vanwege de subjecten, en σsα2 is de interactie van factor A en subjecten.

De juiste F-ratio wordt gevormd door gebruik van deze formule:

$$F = \frac{systematic\ variability + error\ variability}{error\ variability}$$

Vanwege de eerder besproken "compoun symmetry" assumptie, is de volgende volgorde van procedure aangeraden voor de test van factor A: (1) Doe de normale F-test, ondanks het feit dat deze regelmatig de H0 teveel afwijst. (2a) Als H0 niet afgewezen wordt, stop. (2b) Als H0 afgewezen worde, gebruik de Geisser & Greenhouse (1958) conservatieve F-test. De graden vrijheid ("degrees of freedom") voor de kritische F-waarde ("F-critical-value") worden in dit model aangepast tot "1" en "n – 1". (3a) Als H0 afgewezen wordt, stop, omdat dit aangeeft dat beide tests dezelfde conclusie hebben bereikt. (3b) Als H0 niet afgewezen wordt, moet er een verdere test gebruikt worden, om het gelijkspel te beeindigen. Dit is een aangepaste ("adjusted") F-test, waarbij de aanpassing ook wel bekend is als Box's (1954b) correctie (ook wel: de Huyn & Feldt procedure). De graden vrijheid van de teller zijn "(J – 1)ε", en de graden vrijheid van de noemer zijn "(J -1)(n – 1)ε". De "ε" hier is de correctie-factor ("correction factor"; dus niet dezelfde "ε" dat naar residuen refereert).

Als er meer dan twee levels van de herhaalde factor A zijn, en de H0 (voor de herhaalde factor) afgewezen wordt, dan kan het belangrijk zijn om te weten welke gemiddelden anders zijn van elkaar. Dit kan geëvalueerd worden door gebruik te maken van "multiple comparison procedures" (MCP). Verschillende MCP's worden besproken in hoofdstuk 2, en de meeste hiervan kunnen ook gebruikt worden voor een ORMA.

Echter, een geschonden "compound symmetry" assumptie heeft ook effect op MCP's. Er zijn dus twee alternatieven mocht dit het geval zijn. de eerste is om een reserve fout-term ("spare error term") te gebruiken voor elke geteste vergelijking (i.p.v. dezelfde fout-term te gebruiken, makelijk MSSA). het tweede alternatief is om multipele afhankelijke t-tests te gebruiken ("multiple dependent t-tests"), waarin het α-level aangapast is om een soortgelijke manier als in de Bonferonni-procedure.

Er zijn verschillende alternatieven voor het ORMA model. Mees aanzienswaardig is de friedman test, wat een non-parametrische procedure is gebaseerd op rangen (zoals de Kruskal-Wallis test), maar die ook gebruikt kan worden in een herhaalde metingen-model. De Friedman tests wordt als volgt uitgevoerd:

- Scores worden gerangsschikt binnen het subject (als er 6 levels van factor A zijn, dan zijn de scores voor elk subject gerangschikt van 1 tot 6).
- Van deze rangen wordt een gemiddelde rang voor elk level van factor A berekent.
- Dan wordt H0 een test van of the gemiddelde rangen voor de levels van A gelijk zijn.
- De test statistiek (χ2) wordt vergeleken met de kritische waarde van αχJ-12. Als de test statistiek groter is dan de kritische waarde, wordt H0 afgewezen.

Deze test heeft helaas wel het probleem dat de test-statistiek niet precies verspreid wordt zoals χ2 als de n of klein is (specifiek als een van beide onder de 6 is). Hierom is het belangrijk om the tabel van kritische waarden ("critical values") te raadplegen in Marascuilo & McSweeney (1997, Table A-22, p. 521).

Net zoals de kruskal-Wallis test gaat de friedman test ervan uit dat de vorm en variatie van de populatie distributies gelijk is, en dat de responsvariabele continue is.

Voor de friedman test kunnen er meerder MCP's gebruikt worden. Een "ultiple-matched-pair Wilcoxon tests" in Bonferonnie vorm zijn het best in het geval ven een geplande paarsgewijze vergelijking.