Mediatie-analyse
College 4+
Cursus PMC Statistiek Plus

Harry Ganzeboom

1 maart 2019

Hoofdpunten

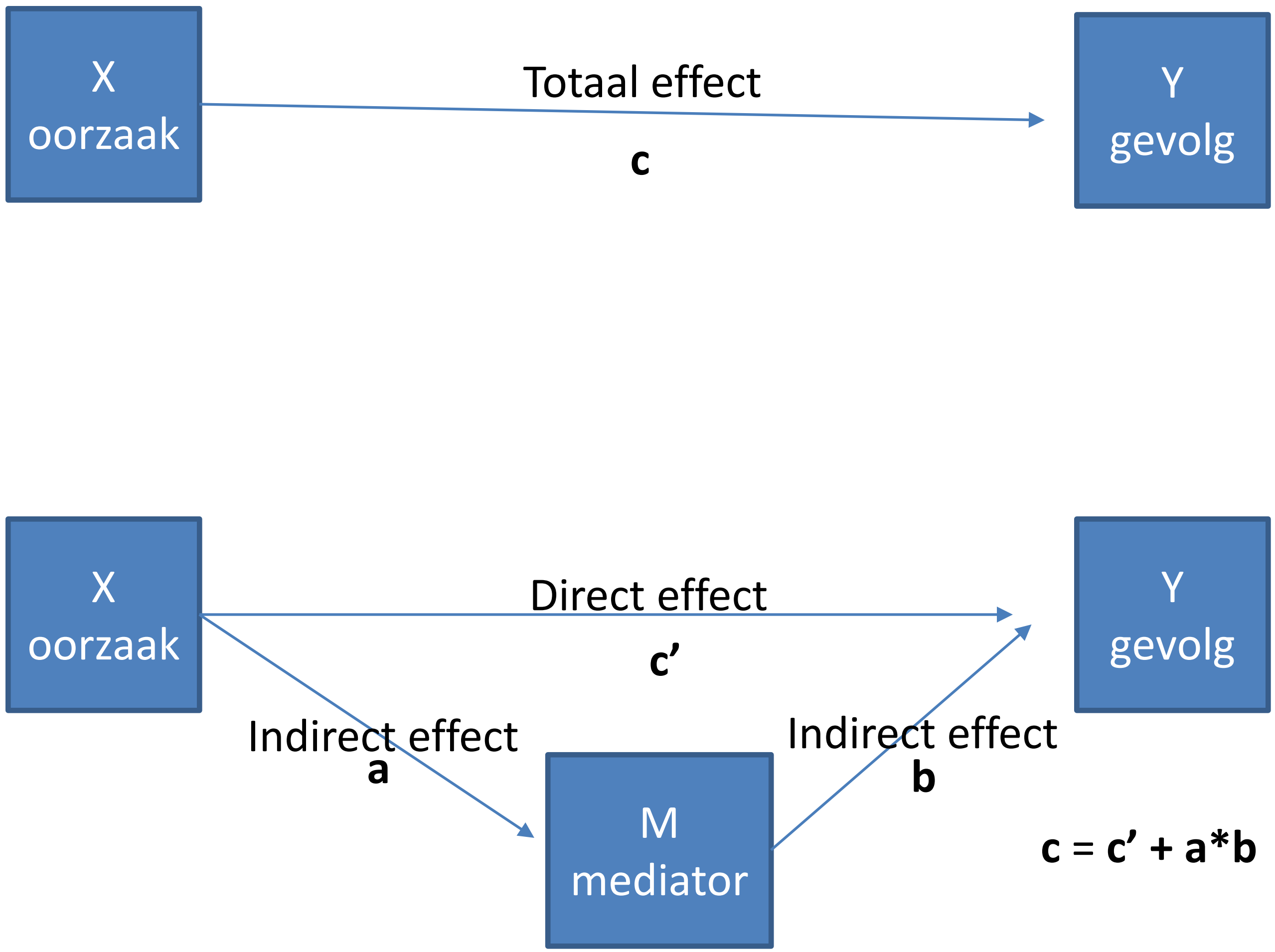
- Wat is mediatie? Wat is confounding?
- Waarschuwing: Mediatie is geen moderatie!!
- Multipele regressie: hoe werkt het?
- Mediatie-analyse via multiple regressie
  - Stapsgewijze regressie
  - Berekening samenstelling indirect effecten
  - Toetsing van significant indirect effect
- Stappenplan

MEDIATIE

Mediatie-analyse

- Mediatie-analyse: in hoeverre verloopt de invloed van een oorzaakvariabele X op een gevolgvariabele Y via een tussenliggende variabele M?
- Voorbeeld: in hoeverre hebben vrouwen minder inkomen dan mannen omdat vrouwen minder uren werken?
  - X: sekse
  - Y: inkomen
  - M: arbeidsaanbod (aantal uren betaalde arbeid)
- X: oorzaak, onafhankelijke *(independent)* variabele
- Y: gevolg, uitkomst, afhankelijke *(dependent)* variabele
- M: mediator, mechanisme, tussenliggende variabele

Causaal model



MEDIATIE IS GEEN MODERATIE!!

- In wetenschappelijke literatuur wordt **mediatie** vaak besproken in samenhang met **moderatie**: Baron, Reuben M. and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51(6):1173–82.
- Aan Baron & Kenny hebben we ook de labeling van de effecten **c, c', a, b** te danken.
- Wij bespreken moderatie in een volgend college. Hier alvast de waarschuwingen:
  - Studenten (maar niet alleen studenten…) zijn vaak in de war over het onderscheid tussen mediatie en moderatie.
  - Mediatie en moderatie zijn twee heel verschillende probleemstellingen (die je beide met multiple regressie te lijf gaat). Bij mediatie gaat het om **indirecte effecten**, bij moderatie om **interactie-effecten.**
  - Mediatie is wel verwant met confounding (de invloed van achterliggende variabelen)
- **Mediatie:** in welke mate verloopt de invloed van X op Y via een tussenliggende variabele M?
- **Confounding**: in welke mate wordt de correlatie tussen X en Y geproduceerd door een achterliggende variabele Z?
- **Moderatie:** in welke mate verandert een (tussenliggende of achterliggende) variabele de invloed van X op Y?

Totaal, direct en indirect effect

- ***Totaal effect***: wat is de invloed van X op Y als je niet rekening houdt met de mediator M?
- ***Indirect effect***: welke deel van deze invloed verloopt via M: X → M → Y?
- ***Direct effect***: welk deel van deze invloed verloopt NIET via M: X → Y || M. [||||*: wanneer we M constant houden*]
- Mediatie-analyse stelt je in staat deze effecten keurig te kwantificeren:

Totaal effect = indirect effect + direct effect

- Direct effect is eigenlijk geen gelukkige benaming. Beter zou zijn: residueel effect, onverklaard effect.
- Als het directe effect 0 is, kun je spreken van ***volledige*** mediatie, anders van ***partiële*** mediatie.

Correlatie en totaal effect

- Het totaal effect X → Y is onderdeel van de correlatie:
  - Correlatie = totaal effect + schijneffect
  - Schijneffecten = correlatie – totaal effect
- Het is gebruikelijk om in mediatie in gestandaardiseerde termen te spreken (correlatie en beta), maar het kan ook ongestandaardiseerd (covariante en B).

Constanthouden

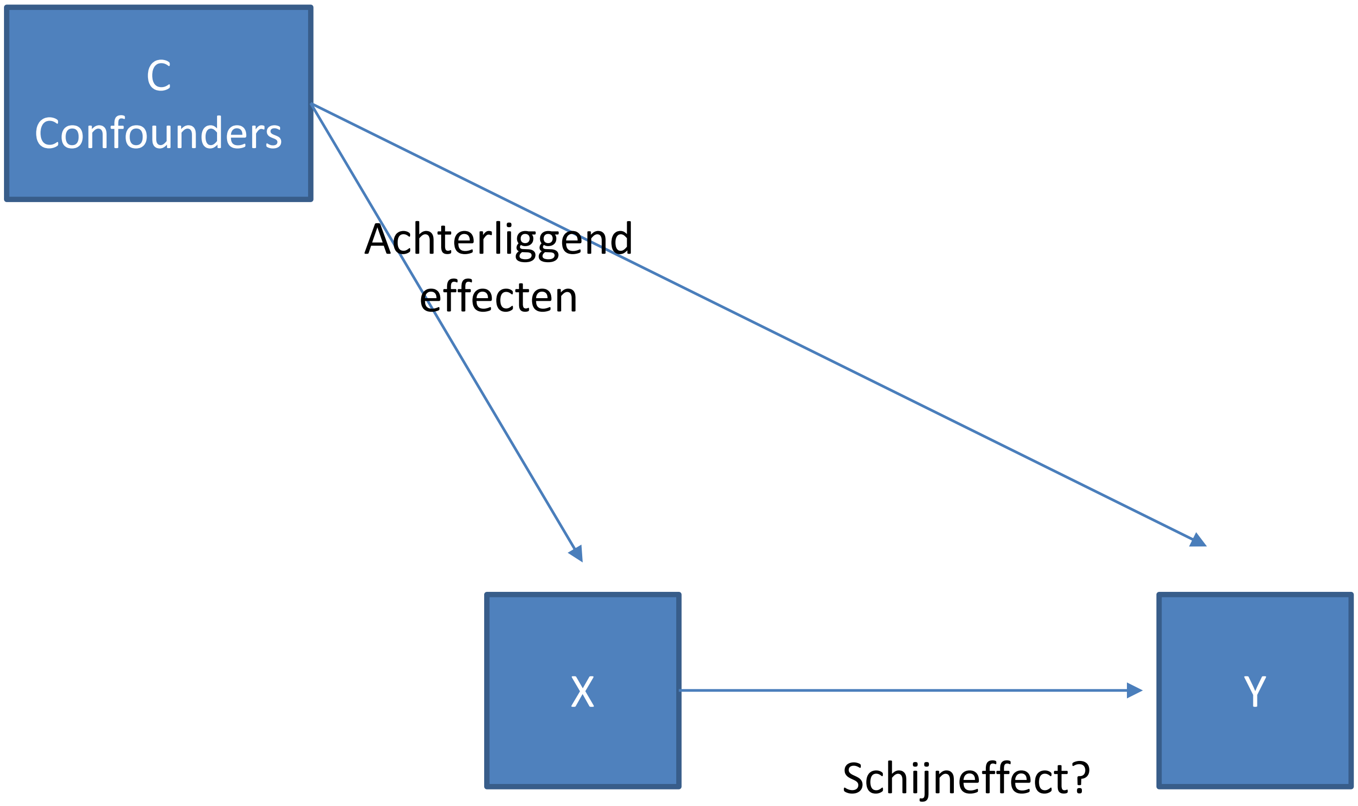
- Om mediatie en confounding te kwantificeren, moeten we de invloed X → Y berekenen bij constant houden van de andere betrokken variabelen: M (bij mediatie), C (bij confounding).
- Het constant houden van een andere variabele kunnen we uitvoeren met tabellen met conditionele gemiddelden (tabelsplitsing, tabelelaboratie), maar in de praktijk doen we het altijd met Multiple Regression, oftewel het *(OLS) linear model*.
- Analyse van mediatie (en ook confounding) komt uiteindelijk neer op:
  - het berekenen van twee regressiemodellen;
  - nabewerking van de resultaten via:
    - berekening direct en indirect effect
    - toetsen van significantie van het indirect effect via de Sobel-test

Confounding

CONFOUNDING

- Confounding == gemeenschappelijk (achterliggende) oorzaak.
- Bekende (afgezaagde) voorbeelden:
  - Hoe meer ooievaars in een gebied, des te hoger het geboortecijfer.
  - Hoe groter de schoenmaat, des des te hoger het IQ.
  - Hoe meer brandweerlieden bij een brand, des te groter de schade.
  - Kinderen die met een nachtlampje gaan slapen, worden later bijziend.
- We spreken over *spurious effects*: schijnbare invloed.
- NB: niet schijn<u>correlatie</u>, het is een schijn<u>effect</u>. De correlatie bestaat wel, het effect bestaat niet.

Het causaal model bij confounding



Confounding

- Bij confounding analyseer je de samenhang tussen X en Y met uitschakeling van de achterliggende oorzaken van X en Y.
- Bij volledige confounding: de relatie van X → Y verdwijnt bij constant houden van C.
- Gedeeltelijke confounding: de relatie van X → Y verzwakt bij constant houden van C.
- Suppressie: de relatie van X → Y versterkt bij constant houden van C.

Confounding en mediatie

- Confounding en mediatie lijken sterk op elkaar en zijn eigenlijk ***twee perspectieven op hetzelfde causale model***.
- De statistische behandeling van confounding en mediatie is identiek: wat gebeurt er met X → Y als je C / M (controlevariabelen) constant houdt?
- Maar de interpretatie is radicaal verschillend:
  - Confounding: hoe sterk is het zuivere effect X →Y. Is er wel een oorzaak-gevolg relatie?
  - Mediation: hoe verloopt het effect X → Y? Hoe zit de oorzaak-gevolg relatie in elkaar?
- Het verschil wordt gemaakt door wat je als primaire oorzaak-gevolg relatie bekijkt. Het zit in je hoofd.

Causale volgorde

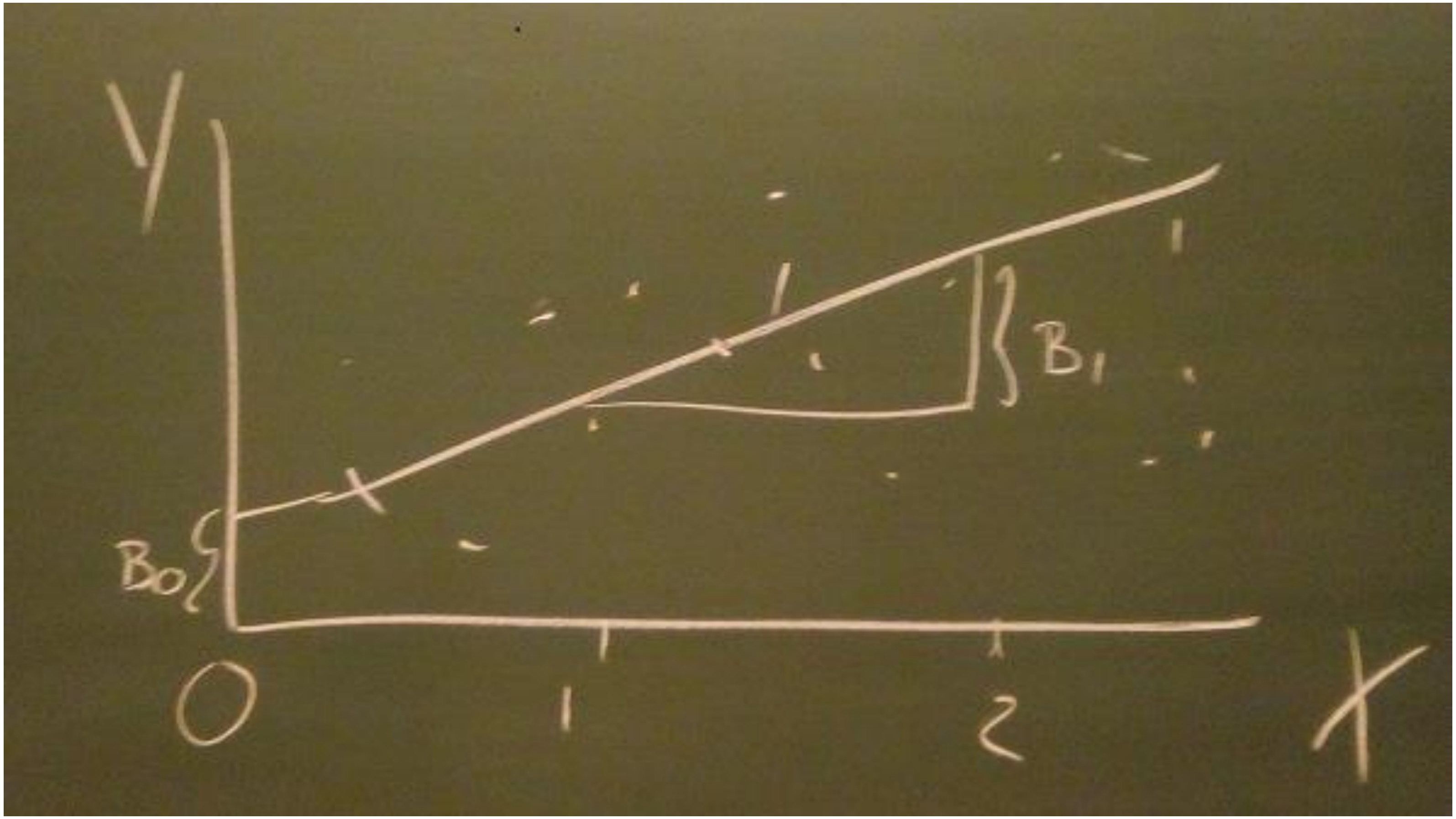
- Welke variabele als X, Y, M of C optreedt, is een aanname (assumptie) van de onderzoeker. Je kunt dit niet zien aan de statistische output.
- De onderzoeker dient te argumenten wat de causale volgorde van de variabelen is: wat kan oorzaak zijn en wat gevolg?
- Meest gebruikte argumenten voor het kiezen van een volgorde:
  - Optreden in de tijd, levensloop.
  - Bv. eerst geslacht, geboortejaar, dan opleiding, dan arbeidsaanbod, en ten slotte inkomen.

MULTIPELE REGRESSIE

Het enkelvoudig lineair model *(simple regression)*

- Het enkelvoudig lineair model luidt:
  - Y = B0 + B1*X1 (ook wel: Y = a + b*X)
- Het model bestaat uit de formule van een best passende rechte lijn. Hierin zijn:
  - B0: de constante / *intercept* = de verwachting voor Y als X1 = 0
  - B1: de hellingshoek / *slope* = hoeveel Y naar verwachting toeneemt / afneemt als X1 een eenheid groter wordt
- De best passende lijn wordt (door SPSS Regression) berekend via de methode van *Ordinary Least Squares* (OLS = kleinste kwadratenmethode).

OLS simple regression



Het meervoudig lineair model *(multiple regression)*

- Model: Y = B0 + B1*X1 + B2*X2 + …
- (het voorbeeld is voor twee X-variabelen, maar uitbreidbaar naar –tig X-en)
- Berekening van het best passende model vindt opnieuw plaats (door SPSS Regression) met de OLS (kleinste kwadraten) methode.
- B0: de verwachting voor Y, wanneer X1 = X2 = 0.
- B1: de toe-/afname van Y wanneer X1 met een eenheid toeneemt, waarbij X2 constant wordt gehouden.
- B2: de toe-/afname van Y wanneer X2 met een eenheid toeneemt, waarbij X1 constant wordt gehouden.
- De interpretatie van B1 en B2 is ***partieel***: we houden de andere X-variabele constant.

Constant houden

- Het bijzondere van het MR model is de **partiële interpretatie** van de B-coëfficiënten. Het levert je een beeld op van de unieke invloed van een X-variabele op Y, terwijl je de invloed van de andere X-variabelen uitschakelt / neutraliseert door deze constant te houden.
- Het MR model is niet de enige methode van constant houden, maar wel een heel algemeen toepasbare, in het bijzonder als het om veelwaardige (bv. continue) variabele gaat.
- Een alternatieve, inzichtelijke maar onbruikbare methode is constant houden via tabelsplitsing (‘table-elaboration’).

Een voorbeeld met twee dichotome X-variabelen

- Wanneer we twee dichotome X-variabelen hebben, kunnen we ook een (tabel)analyse met conditionele gemiddelden gebruiken
- Voorbeeld:
  - Y: inkomen (continu)
  - X1: vrouw (0/1)
  - X2: full-time werk (0/1) meer of minder dan 20 uur werken.
- Dit kan zowel met conditionele gemiddelden als met een regressie-model.

Voorbeeld in SPSS

Tabel 1: Conditionele gemiddelden van Inkomen en arbeidsaanbod (al dan niet FULLTIME) voor mannen en vrouwen

FEMALE		FULLTIME	PINC
0	Mean	0.98	2617
	N	384	384
1	Mean	0.84	1820
	N	366	366
Total	Mean	0.91	2228
	N	750	750

Fulltime: > 20 uur per week

Conclusies:

- Mannen werken 98% fulltime, vrouwen 84%
- Mannen verdienen 2617-1820=797 euro (per maand) meer dan vrouwen

Conditionele gemiddelden

Tabel 2: Conditionele gemiddelden van Inkomen en arbeidsaanbod voor mannen en vrouwen als ze fulltime of parttime werken

FULLTIME	FEMALE		Mean	N
0	0		1933	6
	1		1192	60
		Total	1259	66
1	0		2628	378
	1		1943	306
		Total	2321	684
Total	0		2617	384
	1		1820	366
		Total	2228	750

Binnen elke rij van FULLTIME is het arbeidsaanbod constant (0 of 1) !!

Conclusies over inkomensverschil tussen vrouwen en mannen :

- Voor parttimers is het verschil 1192-1933 = -741
- Voor fulltimers is het verschil 1943-2628-= -685
- Gemiddeld (gewogen) is het (66/750)*(-741) + (684/750)*(-685) = -687

Enkelvoudige regressie

Tabel 3: Inkomensverschil tussen mannen en vrouwen als lineair model

	B	SE	Beta	t	p
1					
(Constant)	2617	49		53.4	0.000
FEMALE	-797	70	-0.383	-11.4	0.000

Conclusies:

- Vrouwen verdienen 797 euro minder dan mannen
- Het inkomensverschil is dik significant (t=-11.4).

Multiple regressie

Tabel 4: Inkomensverschil tussen mannen en vrouwen en tussen parttimers en fulltimers als multiple (meervoudig) lineair model

Model		B	SE	Beta	t	P
1	(Constant)	1884	133		14.2	0.000
	FEMALE	-687	71	-0.330	-9.6	0.000
	FULLTIME	745	126	0.203	5.9	0.000

Conclusies:

- Mannen die part-time werken hebben een verwacht inkomen van 1884.
- Full-timers verdienen gemiddeld 745 euro meer dan parttimers, als je rekening houdt met de (gecontroleerde) inkomenschillen tussen mannen en vrouwen
- Vrouwen verdienen gemiddeld 687 euro minder dan mannen, als je rekening houdt met het gecontroleerde verschil in inkomen tussen fulltime en parttime werken.
- Het inkomensverschil tussen mannen en vrouwen is verminderd van -797 naar -687 door het constant houden van full-time werken.

Een voorbeeld met continue M

- Het is natuurlijk een vergroving om het verschil in arbeidsaanbod tussen mannen en vrouwen te vereenvoudigen tot het al dan niet full-time (= meer of minder dan 20 (!) uur) werken.
- We willen liever arbeidsaanbod als continue variabele (het preciese aantal uren werk per week) constant houden.
- Met tabelanalyse gaat dit moeilijk (je zou heel veel tabellen moeten maken en de resultaten worden erg onregelmatig door kleine aantallen), maar het gaat heel goed met MR.

Voorbeeld in SPSS

Tabel 5: Verwacht inkomensverschil tussen mannen en vrouwen zonder en met constant houden van arbeidsaanbod (=aantal uren betaalde arbeid)

Model		B	SE	Beta	t	Sig.
1	(Constant)	2617	49		53.4	0.000
	FEMALE	-797	70	-0.383	-11.4	0.000
2	(Constant)	810	200		4.0	0.000
	FEMALE	-417	78	-0.200	-5.3	0.000
	HOURS	47.7	5.1	0.349	9.3	0.000

Conclusies:

- Per uur (per week) verdient men gemiddeld bijna 48 euro (per maand!)
- Als je rekening houdt met het aantal gewerkte uren, is bijna de helft van het inkomensverschil tussen mannen en vrouwen verklaard

STATISTISCH UITSTAPJE

B en Beta

- Het geschatte regressiemodel staat in de kolom B, die achtereenvolgens de coefficienten B0, B1, B2 etc bevat.

  Y = B0 + B1*X1 + B2*X2 + etc.

- Bij een X is dit de formule van een rechte lijn, die je gemakkelijk kunt tekenen. Bij meer dan een X is dit de formule van een hypervlak en is een grafiek niet meer te tekenen.
- Beta is hetzelfde regressiemodel, maar nu in gestandaardiseerde termen zX en zY: zX = (X – Mean(X))/SD(X) en zY = (Y – Mean(Y))/SD(Y).
- Beta-coefficienten variëren tussen -1 en +1 (zoals correlaties) en kunnen daarom gemakkelijk met elkaar vergeleken worden.

SE

- SE is het centrale ding in de inferentiële statistiek.
- Je leert wat een SE is en hoe je hem uit moet rekenen doorgaans in verband met een gemiddelde [SEMEAN = SD/√N] en een proportie [SE(P) = √((1-P)*(P)/N)], maar het is belangrijk te weten dat SE's over alle statistische grootheden bestaan, waaronder de regressiecoefficienten B.
- Daarom: Wat is een SE en wat kun je ermee doen?

Wat is een SE?

- Een steekproefgrootheid (zoals een correlatie, B, Beta, gemiddelde, proportie, SD) is niet hetzelfde als de overeenkomstige populatiegrootheid.
- We zijn in de populatiegrootheid geïntesseerd, de steekproefgrootheid is daarvan slechts een schatting, die ook anders had kunnen uitvallen.
- Maar hoeveel anders had het kunnen uitvallen? Daarvoor doen we een gedachtenexperiment: de steekproevenverdeling (sampling distribution).

Steekproevenverdeling

- Stel: we trekken een groot aantal steekproeven (van grootte N) en noteren de uitkomst van een bepaalde statistische grootheid (bv. B1).
- Over de verdeling van de grootheid bij een groot aantal (denkbeeldige) steekproeven, weten we intuïtief:
  - Deze verdeling is gecentreerd rondom de populatiewaarde: gemiddeld komt er de populatiewaarde uit.
  - De afwijkingen van de populatiewaarde zijn normaal verdeeld: kleine afwijkingen tussen steekproefwaarden en populatiewaarde komen veel vaker voor dan grote afwijkingen.
  - Als N groter is, zullen de afwijkingen tussen steekproefwaarden en populatiewaarde gemiddelde kleiner zijn: de SD van de steekproevenverdeling is dan kleiner.

SE: SD van de steekproevenverdeling

- De SD van een steekproevenverdeling heeft een bijzonder naam: de Standard Error (ook: Sampling Error of Standaardfout).
- Omdat (en voorzover) de steekproevenverdeling normaal is, zou het mooi zijn als we de SE zouden weten: dan weten we immers gelijk met welke waarschijnlijkheid een bepaalde afwijking van de populatiewaarde optreedt.
- Gelukkig hebben knappe statistici formules ontwikkeld waarmee SE's geschat kunnen worden. Die hebben ze doorgegeven aan SPSS, en voor ons gewone stervelingen rekent SPSS het uit en drukt het af.
- Voor ons is een SE dus weinig anders dan een getal in de output; niettemin moeten we wel begrijpen wat er achter zit en wat je ermee kunt doen.
- Een paar formules van SE's staan al hierboven. Een completere collectie vind je hier: http://www.harryganzeboom.nl/Teaching/formules_HG.pdf.
- Formules voor een SE(B) zijn veel ingewikkelder dan die voor een gemiddelde.

Toepassing 1: Schatten, Betrouwbaarheidsintervallen

- Als je eenmaal een SE weet, kun je daarmee een betrouwbaarheidsinterval (CI: Confidence Interval) bereken.
- Doorgaans: CI_{95}: steekproefwaarde ± 1.96*SE.
- Interpretatie: als je heel vaak een betrouwbaarheidsintervallen berekent, bevat die in 19/20 gevallen de ware populatiewaarden – in 5% van je schattingen niet.
- NB: Confidence: steekproefbetrouwbaarheid. Dit is iets anders dan Reliability: meetbetrouwbaarheid. In het Nederlands hebben we helaas daarvoor maar een woord.

Toepassing 2: Toetsing van (statistische) significantie.

- Bij significatie toetsing stellen we een H0 op: een veronderstelling dat de populatiewaarde van de bestudeerde grootheid 0 is.
- Rondom deze denkbeeldige grootheid kunnen we weer een CI opstellen: 0 ± 1.96*SE.
- We berekenen hoeveel SE onze steekproefwaarde van de H0 af ligt: t = steekproefwaarde/SE.
- Als t > 1.96 (of t < -1.96) dan kunnen we concluderen dat de veronderstelde H0 met minder dan 5% kans waar is – we verwerpen dan de H0.
- Kort door de bocht: toetsen is hetzelfde als kijken of de steekproefwaarde in het CI van de H0 ligt.

Enige nuanceringen

- Strikt genomen zijn steekproevenverdeling niet normaal, maar t-verdeeld. Normale en t-verdelingen gelijken sterk bij N > 35 en zijn praktisch niet te onderscheiden bij N > 100.
- SPSS past de juiste verdeling (Normaal of T toe) en geeft je de bijbehorende overschrijdingskans (‘Sig’): de kans dat het steekproefresultaat zou verschijnen in een random sample als de H0 opgaat in de populatie.

H0 bij regressie-analyse

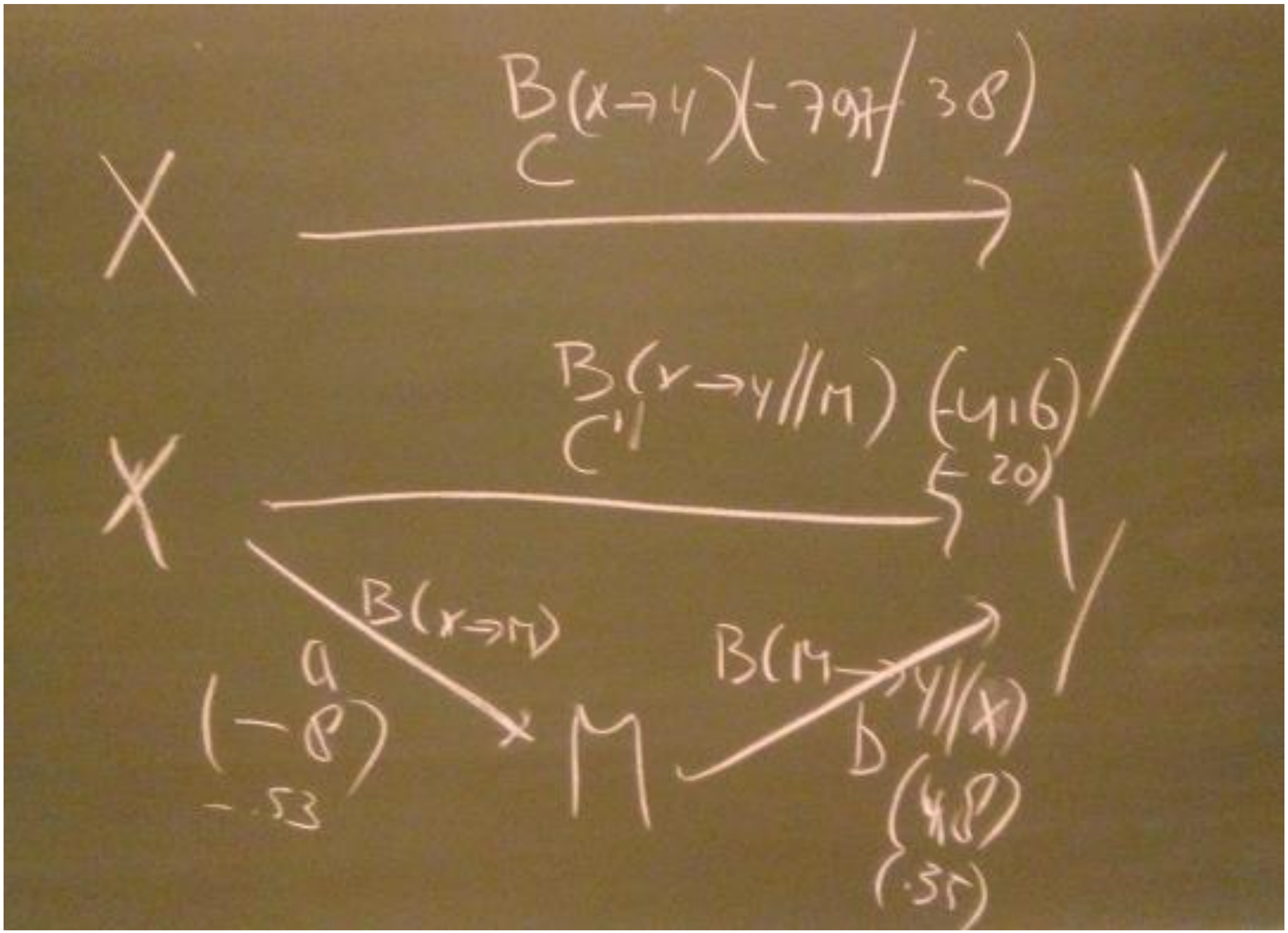
- SPSS veronderstelt bij elke toetsing H0 = 0.
- Voor regressie-coefficienten B1, B2 etc. is dit een mooie veronderstelling, die betekent: geen effect.
- Voor de constante B0 is ook de veronderstelde H0: B0 = 0. Dat is heel vaak een zinloze veronderstelling en de betreffende toetsing is niet interessant.

MEDIATIE-ANALYSE VIA MR

Mediatie-analyse via MR

- Bij multiple regressie maken we nog geen veronderstelling over de causale volgorde van X1 en X2.
- Pas door te beslissen welke causale volgorde er bestaat tussen X1 en X2 wordt het een mediatie-analyse, met een direct en een indirect effect.
- Welke van de twee is de oorzaak variabele en welke de tussenliggende variabele?
- In ons voorbeeld is het antwoord niet moeilijk: sekse (X) gaat vooraf aan arbeidsaanbod (M).

Mediatie-model



Conclusies over mediatie

- Het totale effect Sekse → Inkomen: -797 euro
- Het directe (partiële, overblijvende) effect: -417 euro
- Het indirecte (verklaarde) effect = totaal effect – direct effect = -380 euro.

NB: dit directe effect is niet per se 'discriminatie van vrouwen op de arbeidsmarkt'. Andere mogelijke mediatoren zijn: opleiding, beroep, arbeidservaring.

Over het indirecte effect komen nu twee vragen op:

- Hoe is het indirecte effect samengesteld?
- Is het indirecte effect statistisch significant?

Samenstelling indirect effect

- We kunnen de omvang van het indirecte effect X → M → Y berekenen via vermenigvuldiging van de effecten **a** en **b**.
- Hiervoor moeten we **twee modellen** berekenen:
  - B(X → M) (M is de *dependent variable*)
  - B(M → Y || X) (Y is de *dependent variable*)

  Dit kan in SPSS alleen via twee afzonderlijke regressie-modellen.
- De berekening werkt zowel met ongestandaardiseerde (B) als gestandaardiseerde (Beta) coëfficiënten. We geven doorgaans de voorkeur aan de Beta coefficienten.

Berekening indirect effect

- We kunnen het indirect effect berekenen via vermenigvuldiging van de beide effecten (**a*b**)
- Bij FULLTIME als mediator:

  a*b = -.15 * 745 = -112 euro

- Bij HOURS als mediator:

  a*b = -8 * 47.6 = -381 euro

Samenstelling totaal effect

- Met Fulltime als mediator:
  - -112 – 687 = -797
- Met Hours als mediator
  - -381 – 417 = -797

Berekening gestandaardiseerde direct en indirect effect

- Het is in mediatie-analyse gebruikelijk om indirect en direct effect (ook) uit te drukken in gestandaar-diseerde coëfficiënten (Beta).
- Gestandaardiseerde effecten zijn allemaal in sterkte met elkaar vergelijkbaar.
- Dat is hier belangrijk omdat bij verschillende afhankelijke variabelen (M en Y) de meeteenheid van de B effecten niet vergelijkbaar is: we kunnen van de ongestandaardiseerde effecten niet de relatieve sterkte zien.

Gestandaardiseerde berekening

- Met FULLTIME als mediator:
  - Totaal effect = c = -0.383
  - Direct effect = c’ = -0.330
  - Indirect effect = a*b = -0.262*0.203 = -0.053
- Met HOURS als mediator:
  - Totaal effect = c = -0.383
  - Direct effect = c’ = -0.200
  - Indirect effect = a*b = -0.525*0.349 = -0.183
- Uit deze analyse leren we drie dingen:
  - Arbeidsaanbod (als je het goed operationaliseert) is voor bijna 50% verantwoordelijk voor de inkomensverschillen tussen mannen en vrouwen.
  - Het sterke verschil in arbeidsaanbod (0.523) tussen mannen en vrouwen is belangrijker voor het indirecte effect dan het uurloon-effect (0.349).
  - Het maakt veel uit of arbeidsaanbod operationaliseert als al dan niet FULL-TIME werken of als continu aantal HOURS → bij mediatie-analyse is de kwaliteit van meting van de mediator van cruciaal belang.

Significantie indirect effect

- Aan de SPSS output kunnen we niet zien of het indirecte effect statistisch significant is.
- Toch is dat een heel belangrijke vraag.
- Aan de berekening van de SE van indirecte effecten en de daarop berustende significantie-testen is een belangrijke bijdrage geleverd door de Amerikaanse socioloog (!) Michael Sobel.
- Zijn werk is gepopulariseerd in de website over de Sobeltest: http://quantpsy.org/sobel/sobel.htm

Sobel-test

- Je kunt de sobel-test op twee manieren invoeren:
- Via de t-waarden van B(X→M) en B(M→Y||X).
- Via B(X→M) en B(M→Y||X) zelf, en hun bijbehoren SE.
- Methode 1 geeft alleen een significantietest, methode 2 geeft ook de SE van het indirecte effect.

Sobel-test

- Met HOURS als mediator:
  - t(a) = -9.3 t(b) = 17.9 → Sobel t: 8.0
- Met FULLTIME als mediator:
  - t(a) = -7.4 t(b) = 5.9 → Sobel t: 4.6
- Het indirecte effect is dus in beide gevallen statistisch significant ($t > 1.96$).
- Dat hoeft niet zo af te lopen:
  - Als een van beide samenstellende effecten (a of b) niet significant is, is a*b nooit significant.
  - Als a of b (of allebei) marginaal significant is, is a*b vaak niet significant. Exacte uitslag geeft de sobel-test.

Mediatie in de praktijk

- In de praktijk worden indirecte effecten en de sobel-test vaak niet berekend.
- Maar wordt volstaan met een stapsgewijze regressietabel, waarin naast elkaar worden gezet:
  - Model 1: Correlatie tussen X en Y als enkelvoudige regressive.
  - Model 2: Totaal effect van X op Y (multipele regressie met constant houden van de confounders)
  - Model 3: Partiële effecten van X en M op Y (multipele regressie met constanthouden van confounders en mediatoren)
- Het commentaar gaat er dan over welke veranderingen optreden in totale en partiële effect van X op Y.
- Het effect van X op M wordt veelal niet bediscussieerd.

Mediatie-analyse via stapsgewijze regressie

Tabel 7a: Stapsgewijze regressie ter verklaring van inkomensverschillen tussen mannen en vrouwen via al dan niet full-time werken

		B	SE	Beta	t
1	(Constant)	2617	49		53.4
	FEMALE	-797	70	-.383	-11.4
2	(Constant)	1918	100		19.2
	FEMALE	-526	76	-.253	-7.0
	FULLTIME	725	91	.289	7.9

Mediatie-analyse via stapsgewijze regressie

Tabel 7b: Stapsgewijze regressie ter verklaring van loonverschillen tussen mannen en vrouwen via aantal uren arbeidsaanbod

	Model 1				Model 2			
	B	SE	t	Beta	B	SE	t	Beta
Constante	2617	49.0	53.4		810	200.3	4.0	
FEMALE	-797	70.2	-11.4	-0.383	-417	78.2	-5.3	-0.200
HOURS					47.7	5.14	9.3	0.349
adj R2				14.6%				23.3%

Y: Maandinkomen in euro's; FEMALE: 0/1; HOURS: 1-40. Sample: N=750 mannen en vrouwen, ISSP-NL 2013-2014

De tweede regressie vergelijking

Tabel 8a: Arbeidsaanbod van mannen en vrouwen (uren betaalde arbeid)

Model		B	SE	Beta	t	Sig.
1	(Constant)	37.91	0.33		114.6	0.000
	FEMALE	-7.98	0.47	-0.525	-16.8	0.000

Tabel 8b: Arbeidsaanbod van mannen en vrouwen (al dan niet full-time)

Model		B	SE	Beta	t	Sig.
1	(Constant)	0.984	0.014		70.4	0.000
	FEMALE	-0.148	0.020	-0.262	-7.4	0.000

Conclusies:

- Vrouwen werken minder uren (-8) en minder vaak fulltime (-15%) dan mannen.
- Het verschil in arbeidsaanbod is veel sterker wanneer je het arbeidsaanbod afmeet aan het aantal uren dan aan al dan niet full-time.

Stappenplan mediatie

- Stap 1: bepaal de causale volgorde van je variabelen: wat is X, Y en M? Deze beslissing komt voort uit je design en theoretische overwegingen.
- Stap 2: Bereken regressiemodellen
  a) Bereken model X → Y
  b) Bereken model X, M → Y
  c) Bereken model X → M

  Stap 2a en 2b kun je in SPSS handig combineren.
- Stap 3: Maak stapsgewijze regressietabel
- Stap 4: Bereken de sobel-test

Appendix

- SPSS syntax staat afzonderlijk op Canvas