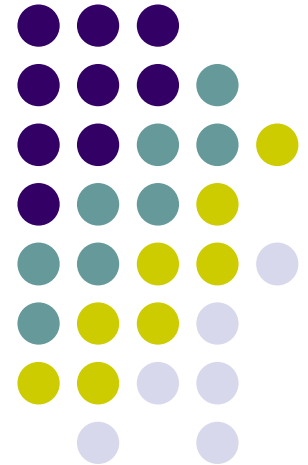


Kausaalisuus ja kausaalipäätely

Pertti Töttö





27.8.2019



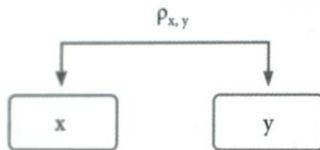
?????

- "Tilastollinen riippuvuus ei ole yhtä kuin kausaalisuus"
- "Ei korrelaatiota ilman kausaatiota"
- "Näennäiskorrelaatio"
- "Suunnattu korrelaatio"

LIITE B. TILASTOLLISTEN MALLIEN GRAAFINEN ESITYSTAPA

Tilastolliset mallit yksinkertaisesta korrelaatiosta monimutkaisiin latenteja muuttujia sisältäviin välittäjähypoteeseihin voidaan kuvata rakenneyhtälömallin graafisella esitystavalla. Tähän liitteeseen on koottu tälle teokselle keskeisimpiä tilastollisia käsitteitä ja tekniikoita rakenneyhtälömallimuodossa. Samaa graafista esitystapaa on käytetty teoksen kaikissa luvuissa.

KORRELAATIO (CORRELATION)



Yksinkertainen korrelaatio ($\rho_{x,y}$) kuvaa yhteyttä kahden muuttujan välillä ottamatta mitenkään kantaa siihen, aiheuttaako jompikumpi varianssia toiseen (tästä syystä kaksisuuntainen nuoli). Ellei toisin ole ilmoitettu, tässä teoksessa korrelaatio tarkoittaa muuttujien välistä lineaarista kovarianssia. Jos tämä korrelaatio standardoidaan muuttujien variansseilla, päädytään Pearsonin korrelaatioon.

SUUNNATTU KORRELAATIO (DIRECTED CORRELATION)

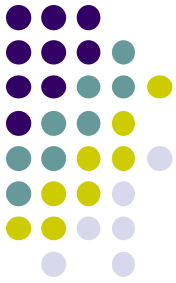


Suunnattu korrelaatio (β_{yx}) tarkoittaa, että yhden (eksogeenisen) muuttujan (x) varianssi heijastuu (jollain mekanismilla) toisen (endogeenisen) muuttujan (y) varianssiin: muutos x:n arvossa implikoi muutoksen y:n arvossa. Koska y-muuttujan varianssi ei ole kuitenkaan koskaan täysin x:n määrittämä, y-muuttujalla on lisäksi

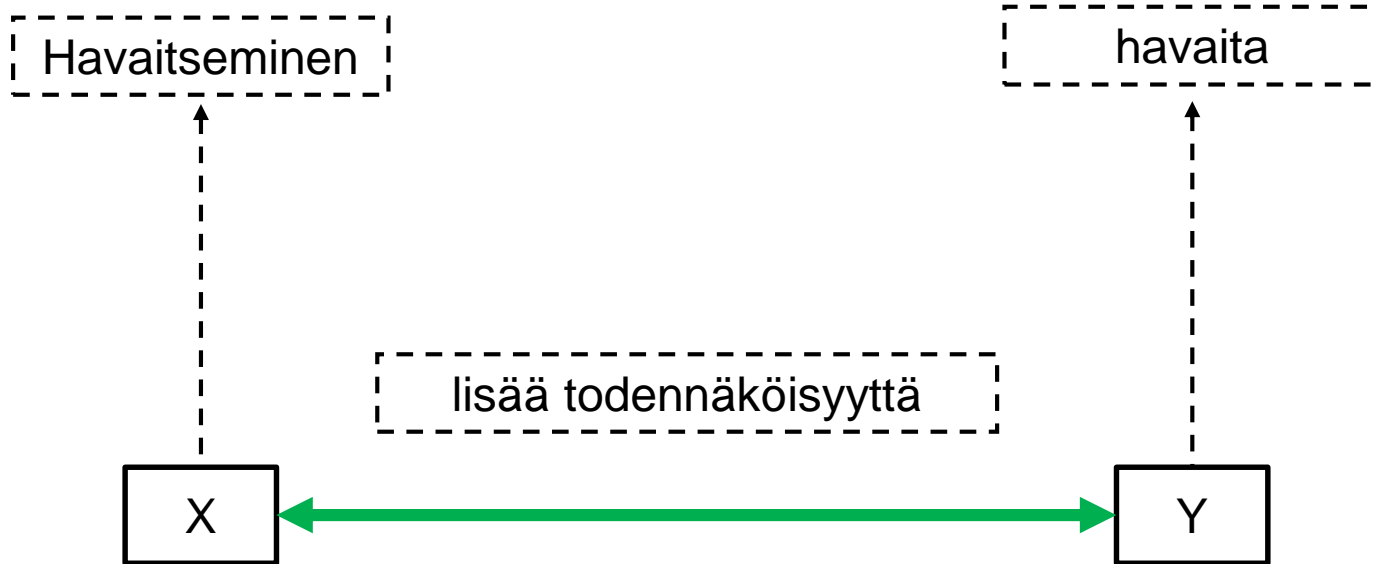
virhetermi ϵ . Virhetermi voidaan ajatella siksi osaksi y-muuttujan varianssia, joka ei selity x-muuttujan varianssilla. Tarkemmin muotoiltuna muutos x:n arvossa implikoi muutoksen y:n odotusarvossa.

Korrelaation $\rho_{x,y}$ laskeminen aineistosta on suoraviivaista. Suunnatun korrelaation β_{yx} osoittaminen on (potentiaalisen endogeenisuuden ongelman takia) paljon monimutkaisempaa.

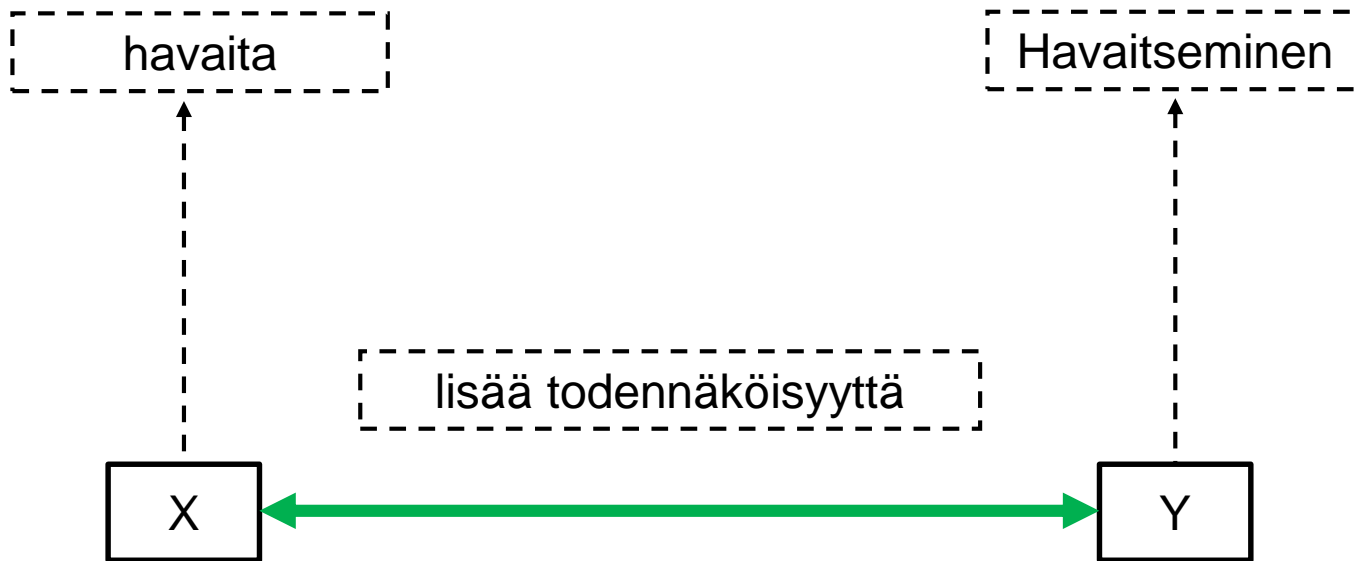
27.8.2019



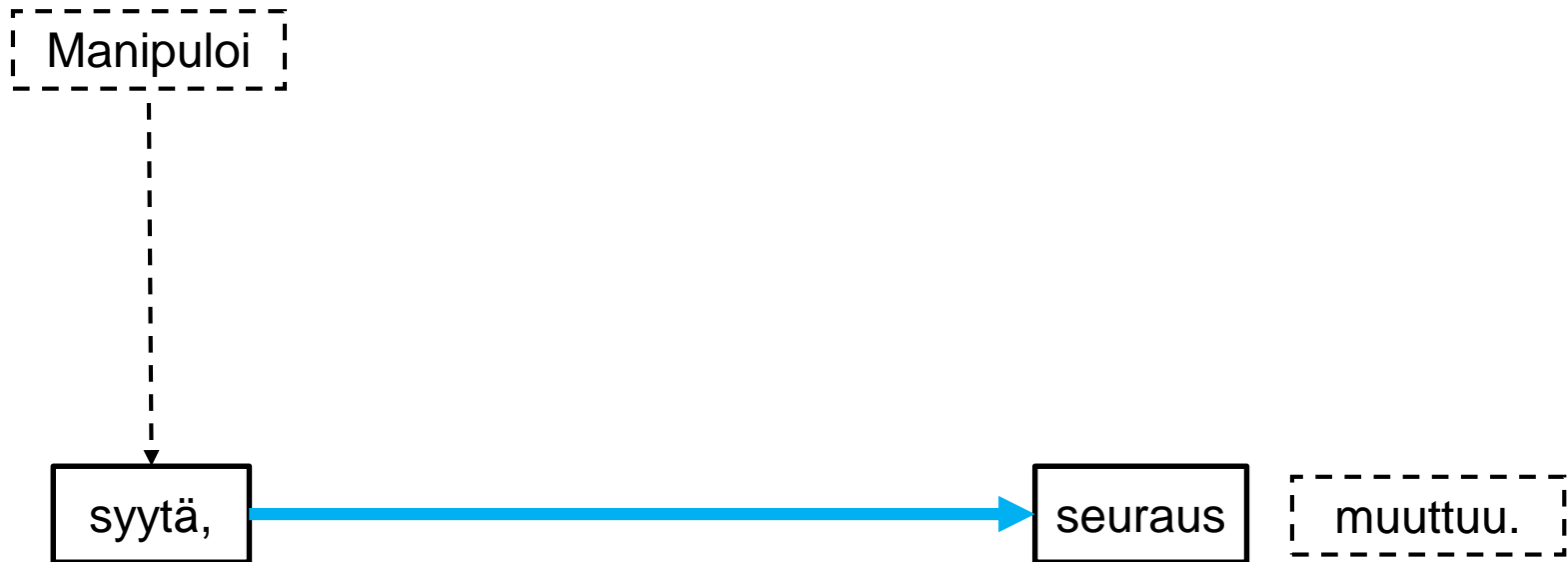
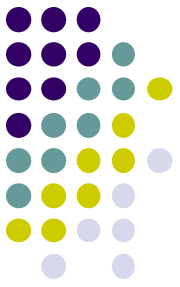
Korrelaatio: symmetrinen



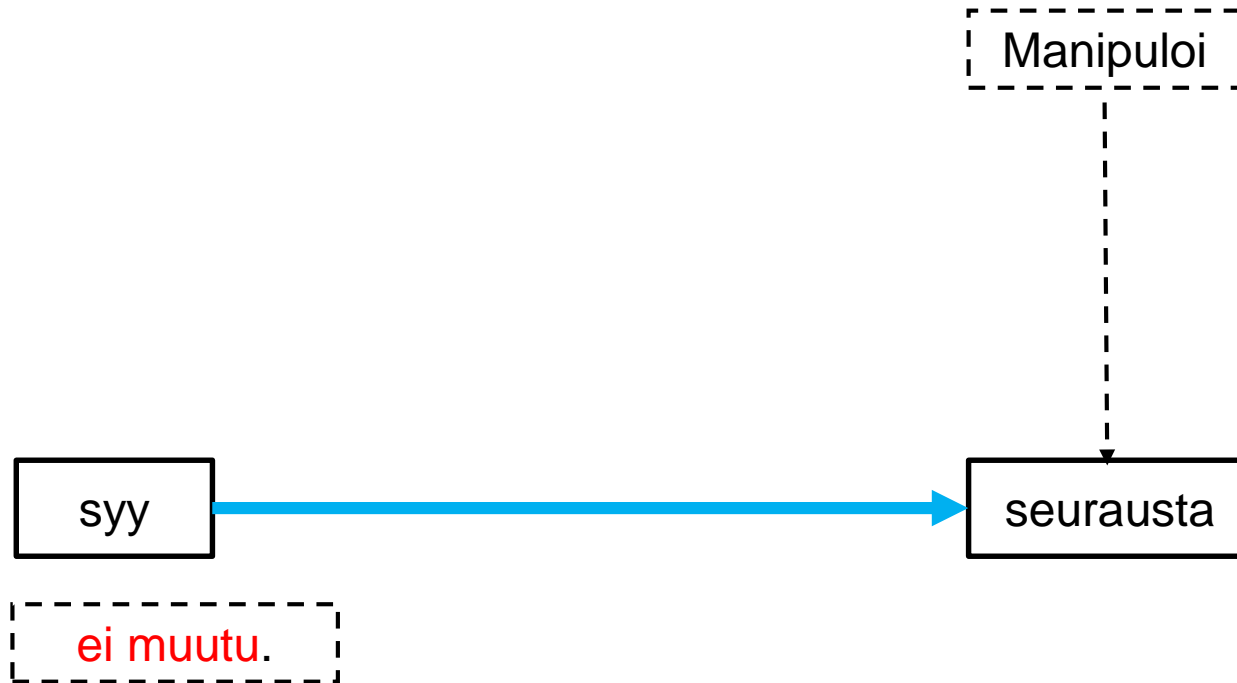
Korrelaatio: symmetrinen



Kausaatio - epäsymmetrinen



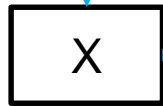
Kausaatio - epäsymmetrinen





Kausaalisuuden määritelmä

Muutos



muuttaa

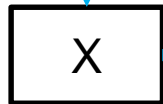


Jos X on Y:n syy, muuttamalla X:ää voi muuttaa Y:tä, mutta Y:n muuttaminen ei muuta X:ää.



Kausaalisuuden määritelmä

Muutos



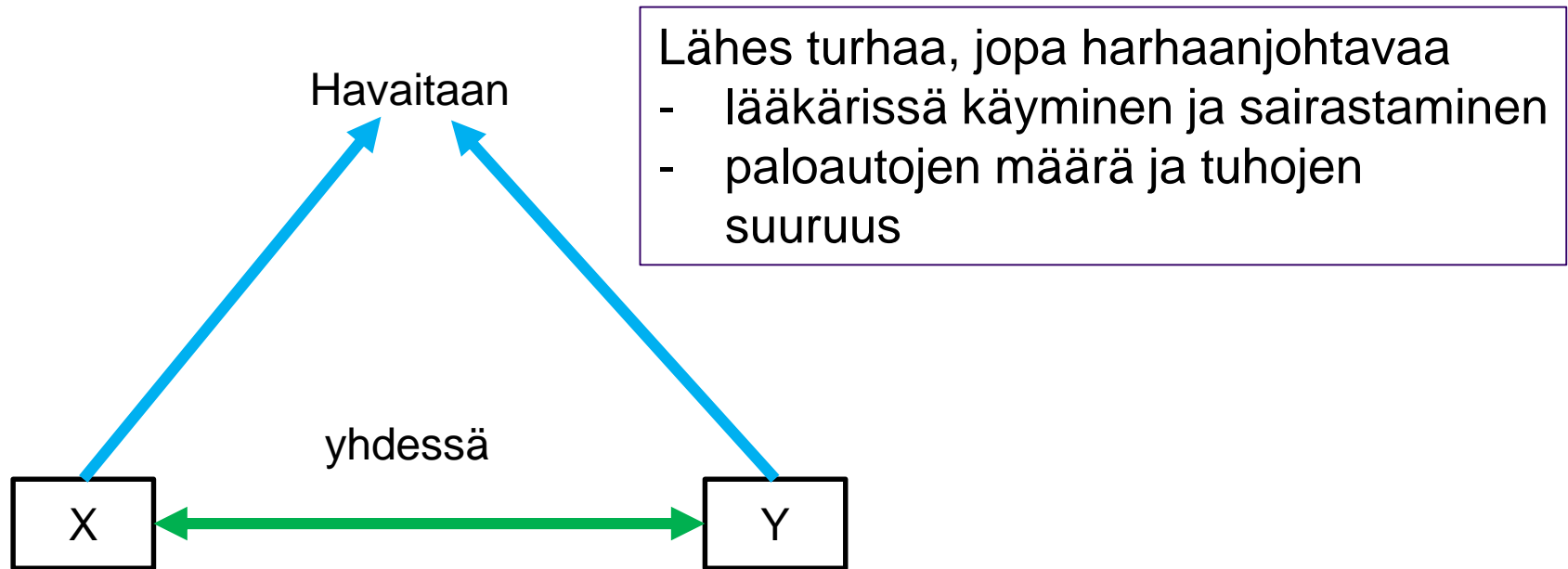
muuttaa



Määritelmä

- ei edellytä tietoista toimijaa
- on kehämäinen (ekalla nuolella määritellään toinen)
- käsite on primitiivi (Kant ja rotat)

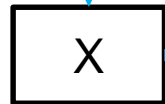
Korrelatiivinen tieto



Kausaalinen tieto...



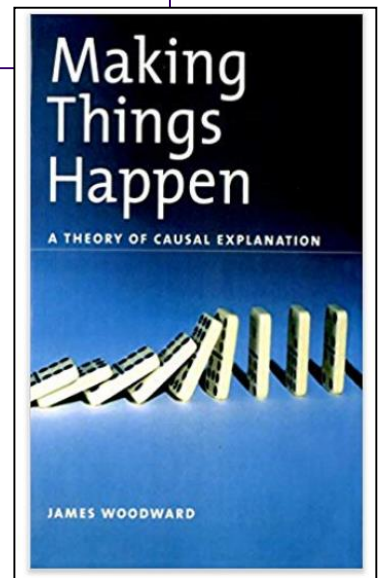
Muutos



muuttaa



Kaiken toiminnan (pidättymisen, välttämisen, vaikuttamisen, hallinnon, politiikan jne.) välttämätön edellytys



... Ja miten sitä saadaan?





... kausaalipäätelyllä!

Jos X:n muuttaminen muuttaa Y:tä,
silloin X on Y:n syy.

Muutos





Kausaalipäätelyllä???

- Mitataan opiskelijaryhmän Y (itsetunto) ennen X :ää (Jari Sarasvuon luento) ja sen jälkeen
- Onko $\Delta Y = Y_{\text{jälkeen}} - Y_{\text{ennen}}$ yhtä kuin X :n kausaalinen vaikutus Y :hyn?



Kausaalipäätelyllä???

- Mitataan opiskelijaryhmän Y (itsetunto) ennen X :ää (Jari Sarasvuon luento) ja sen jälkeen
- Onko $\Delta Y = Y_{\text{jälkeen}} - Y_{\text{ennen}}$ yhtä kuin X :n kausaalinen vaikutus Y :hyn?
- Ei ole, vaan mikä on?



Kausaalipäätelyllä???

- Mitataan opiskelijaryhmän Y (itsetunto) ennen X :ää (Jari Sarasvuon luento) ja sen jälkeen
- Onko $\Delta Y = Y_{\text{jälkeen}} - Y_{\text{ennen}}$ yhtä kuin X :n kausaalinen vaikutus Y :hyn?
- Ei ole, vaan tämä:
- $Y_{\text{jälkeen}} - Y^*_{\text{jälkeen}}$



Kausaalipäätelyllä???

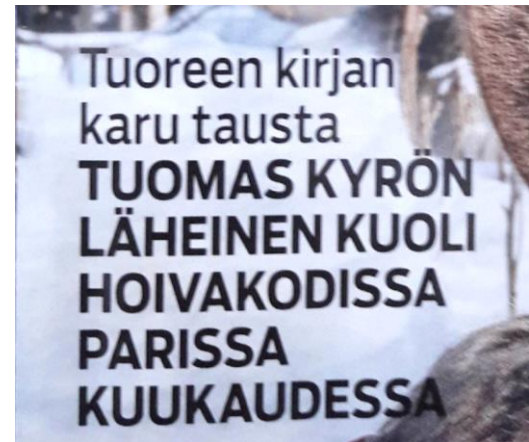
- Mitataan opiskelijaryhmän Y (itsetunto) ennen X :ää (Jari Sarasvuon luento) ja sen jälkeen
- Onko $\Delta Y = Y_{\text{jälkeen}} - Y_{\text{ennen}}$ yhtä kuin X :n kausaalinen vaikutus Y :hyn?
- Ei ole, vaan tämä:
- $Y_{\text{jälkeen}} - Y^*_{\text{jälkeen}}$
- $Y^* =$ mitä Y olisi ollut, jos X :ää ei olisi ollut



Kausaalipäätely on ...

- Y^* = mitä Y olisi ollut, jos X :ää ei olisi ollut

... kontrafaktuaalista päätelyä





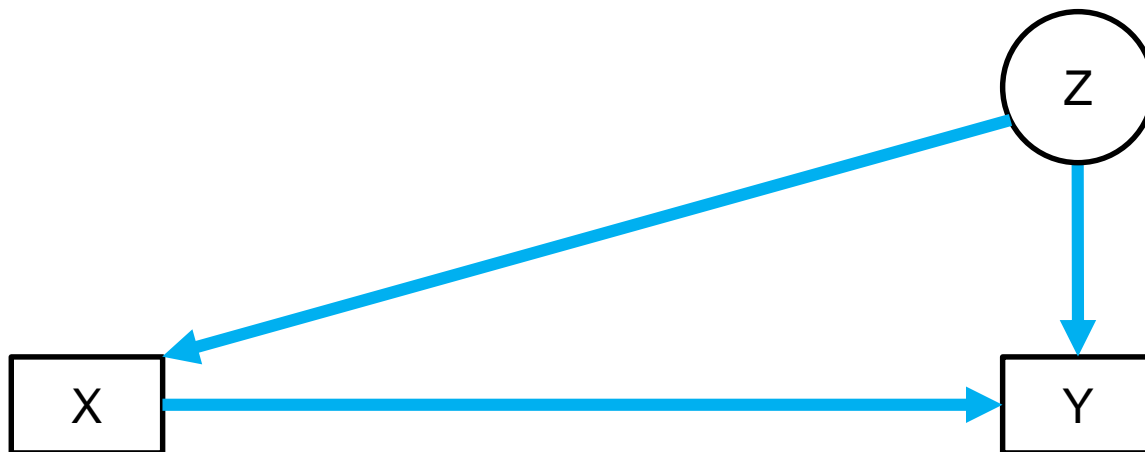
RCT-asetelma

- Y^* = mitä Y olisi ollut, jos X :ää ei olisi ollut
- \approx kontrolliryhmän Y jälkeen



RCT-asetelma

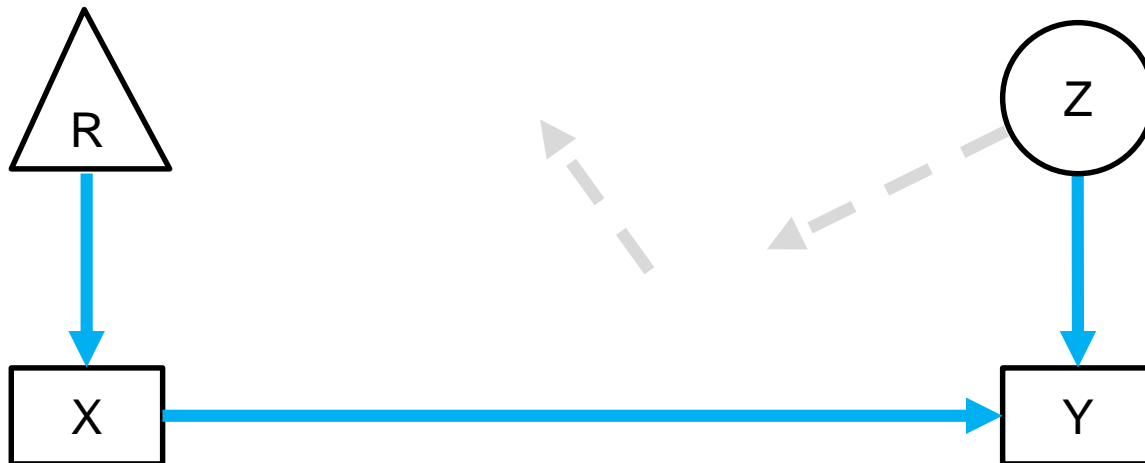
- Y^* = mitä Y olisi ollut, jos X :ää ei olisi ollut
- \approx kontrolliryhmän Y jälkeen





RCT-asetelma

- Y^* = mitä Y olisi ollut, jos X :ää ei olisi ollut
- \approx kontrolliryhmän Y jälkeen



Satunnaistaminen



$$E(Z|X=0) = E(Z|X=1)$$





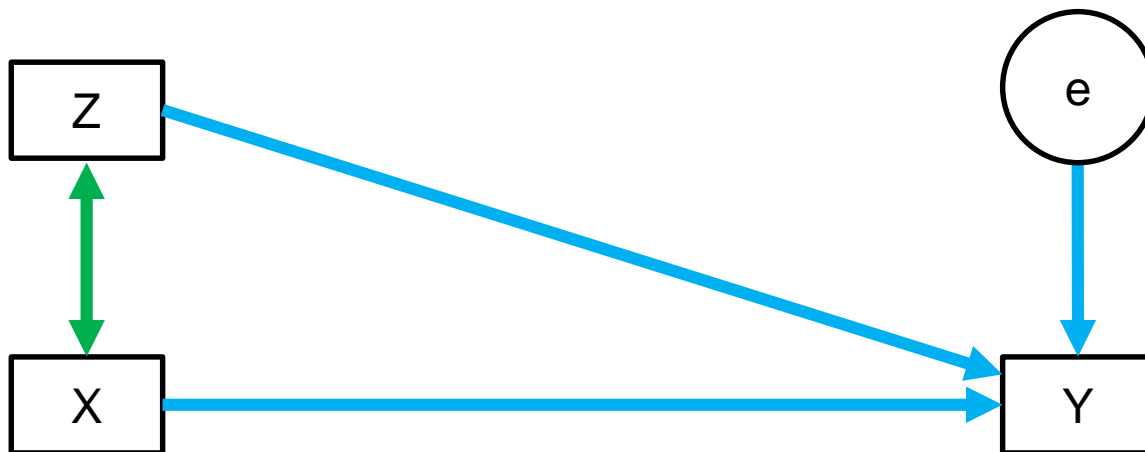
Tilastollinen vakiointi

- adjusting, conditioning, controlling, stratifying
- X:n ja Y:n riippuvuus Z:n luokissa, tasoilla, ositteissa
- $Y = a + b_1X + b_2Z + e$ (lineaarinen erikoistapaus)

Satunnaistaminen



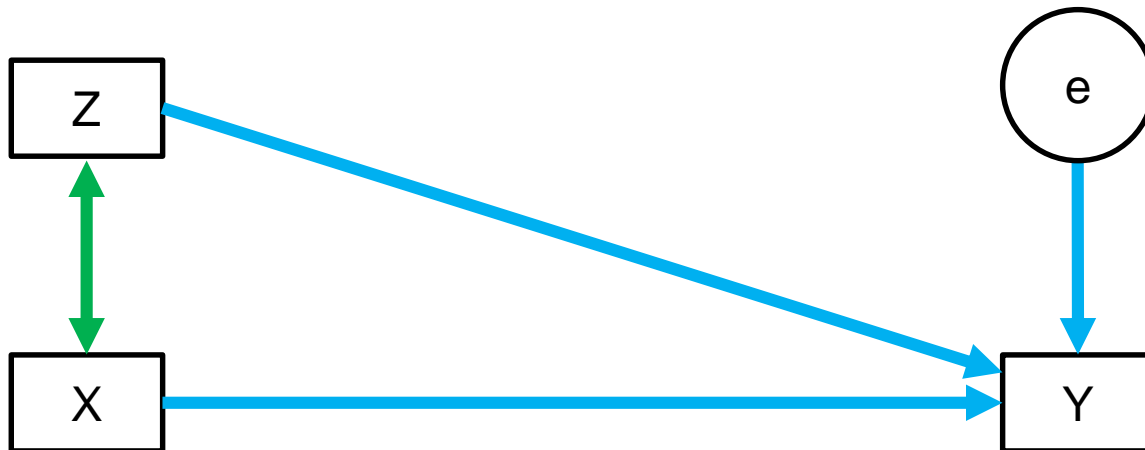
Vakiointi





Vakioinnin ongelma

- Mitkä Z-muuttujat pitää (kannattaa, on tapana) vakioida?



Johdon sukupuolen ja yrityksen kannattavuuden välinen yhteys

Annu Kotiranta
Tutkija

Helsingin Yliopisto,
Elinkeinoelämän Tutkimuslaitos ETLA

Petri Rouvinen
Tutkimusjohtaja

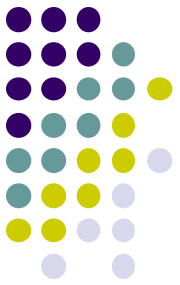
Elinkeinoelämän Tutkimuslaitos ETLA,
Etlatieto Oy

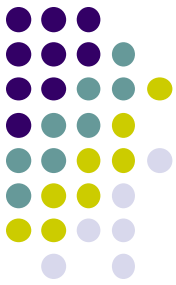
Anne Kovalainen
Professori

Turun kauppakorkeakoulu

1. Johdanto

Viime vuosina Suomi on saanut monta sukuta johtotehtävistä (Eurostat), mikä on alle





Taulukko 2. Regressiokertoimet naisjohtajuusmuuttujien osalta

	Selitetty: Koko pääoman tuottoaste (ROA)	Selitetty: Sijoitetun pääoman tuottoaste (ROI)	Selitetty: Liikevoitto-prosentti (ROS)
(1.) Pienimmän neliösumman menetelmä (selitettävä muuttuja sellaisenaan, koko otos, 46 selittävää muuttujaa)			
Naispuolinen toimitusjohtaja	1,31	4,63*	-0,92
Naispuolinen hall. puheenjoht.	-0,03	-0,99	7,14***
Naisten osuus hallituksessa	2,20***	3,94*	9,39***
(2.) Pienimmän neliösumman menetelmä (selitettävä muuttuja kohtuullistettuna, koko otos, 46 selittävää muuttujaa)			
Naispuolinen toimitusjohtaja	1,18**	2,975***	0,07
Naispuolinen hall. puheenjoht.	-0,50	-1,122	-0,08
Naisten osuus hallituksessa	1,53***	1,278	1,41***
(3.) Pienimmän neliösumman menetelmä (selitettävä muuttuja sellaisenaan, ei äärihavaintoja, 46 selittävää muuttujaa)			
Naispuolinen toimitusjohtaja	1,27**	3,73***	-0,01
Naispuolinen hall. puheenjoht.	-0,64	-1,06	-0,46**
Naisten osuus hallituksessa	1,40**	1,65	0,00

45
Z-muuttujaa!!!



3. Aineisto

Koska mielenkiintomme kohteena on (palkka)johtajan sukupuolen ja yrityksen kannattavuuden välinen yhteys, rajaamme tarkastelun osakeyhtiöihin, jotka työllistivät vuonna 2003 vähintään 10 henkeä. Tulkintamme mukaan näissä yksiköissä omistajuuden ja johtamisen erottaminen on mahdollista ja yleistäkin sekä nämä yritykset ovat pääasiallisesti voittoa tavoittelevia. Näissä yrityksissä on usein myös toimiva hallitus sekä mahdollisesti muutoinkin selkeä hierarkkinen rakenne ja yritysjohton kesken määritellyt vastualueet.

Tilastokeskuksen yritys- ja toimipaikkarekisteri määritteli tämän tutkimuksen kohdejoukon: vuonna 2003 Suomessa toimi 14 020 vähintään 10 henkeä työllistävää osakeyhtiötä. Täydentävät tiedot haettiin Asiakastieto Oy:n tietokannoista. Käyttökelpoinen 12 738 yrityksen otos kattaa 91 % kohdejoukosta.

Toimitusjohtajasta kontrolloidaan hänen sukupuolensa ohella ikä (keskimäärin 48 vuotta) sekä hallituksesta sen koko (keskimäärin 4 henkeä), keski-ikä (49 vuotta) sekä vanhimman ja nuorimman jäsenen ikä-ero (keskimäärin 16 vuotta). Mukana on indikaattori myös siitä, toimiiko toimitusjohtaja hallituksen puheenjohtajana (näin on 6 %:ssa otoksen yrityksistä).

Mukana ovat myös seuraavat yritystason kontrollit: vientitoiminnan harjoittaminen (28 % otoksen yrityksistä); pääomaintensiteetti (taseen loppusumman ja myynnin suhde keskimäärin 1,9); ulkomaalaisomisteisuus (9 % otoksen yrityksistä); kuuluminen useamman yrityksen muodostamaan konserniin (27 % otoksen yrityksistä); luottoluokitus (keskimäärin 78 asteikolla nolasta (huonoin) sataan (paras)); indikaattori sille, että tilintarkastajalla ei ole ollut huomautettavaa yrityksen tilinpäätöksestä (totta yhdeksässä kymmenestä



Vakioi kaikki mahdolliset?

- Tilastotieteen näkökulmasta ei ole oikeita eikä vääriä kontrollimuuttujia, joten kaikki käy. Mutta miksi kontrolloida ainuttakaan? Miksi jotkut korrelaatiot olisivat näennäisiä?
- RCT kontrolloi kaikki mahdolliset Z-muuttujat, myös tuntemattomat ja ei-mitatut
- Kausaalipäättelyn näkökulmasta jotkin muuttujat voivat olla vääriä kontrolleja



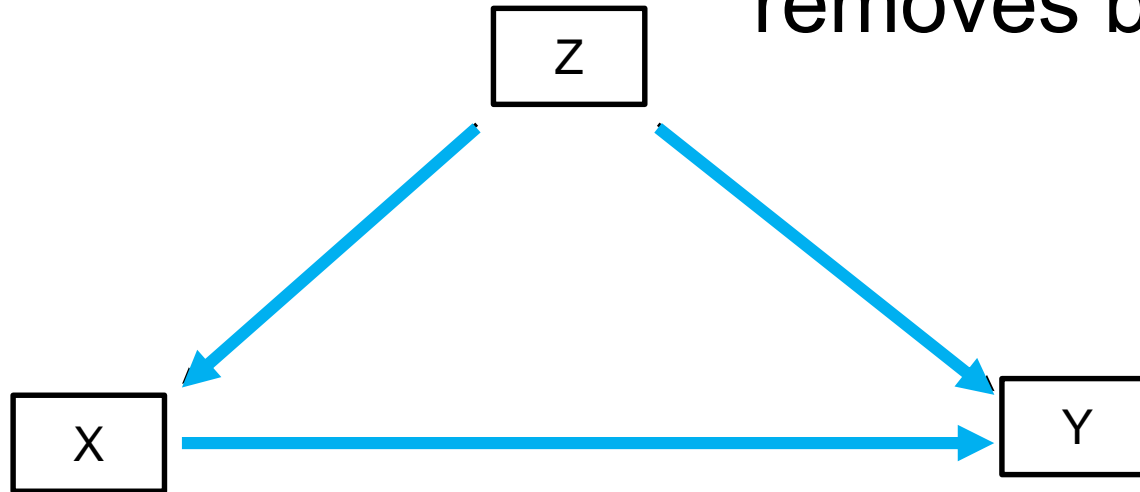
Vakioi...

- ... Z-muuttujat, jotka korreloivat sekä Y:n että X:n kanssa!
- Miksi?
- Kausaalipäättelyn kannalta riittämätön tai harhaanjohtava ohje

Z korreloi sekä X:n että Y:n kanssa



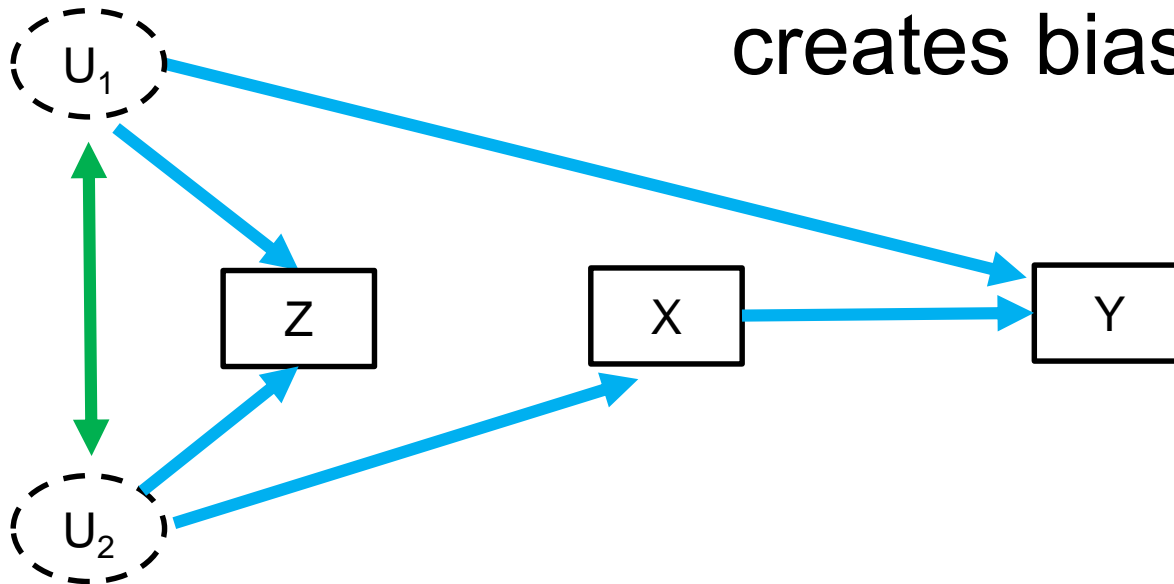
”Conditioning on Z
removes bias”



Z korreloi sekä X:n että Y:n kanssa



”Conditioning on Z
creates bias”



Ergo



- Pitää olla eksplisiittinen kausaalimalli, jonka perusteella voi ratkaista, mitä pitää vakioida halutun efektin identifioimiseksi
- Pearl: etuovi-kriteeri, takaovi-kriteeri, tee-laskenta
- Wright: polkuanalyysi (→ rakenneyhtälömallit)

Polkuanalyysin peruskäsitteet



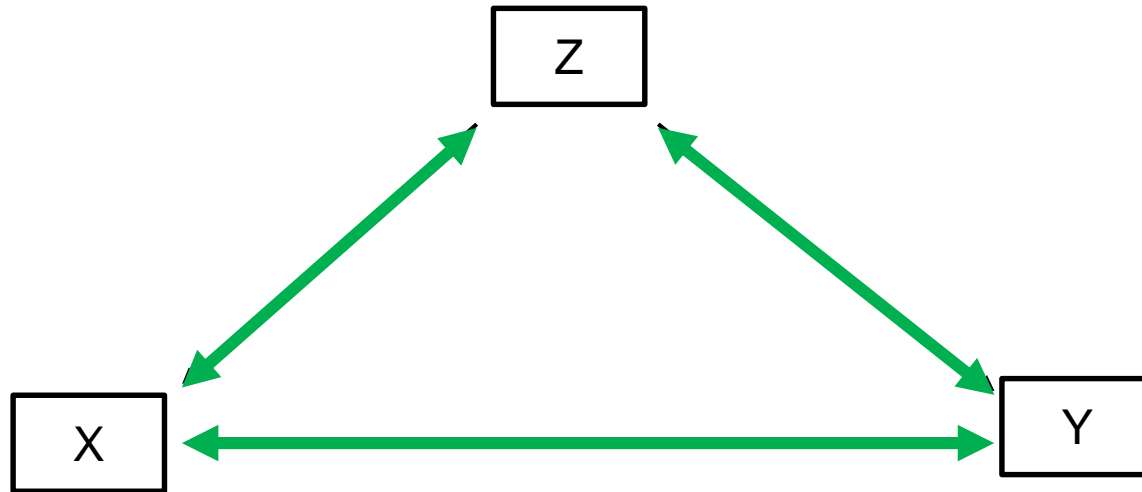
- Suora vaikutus
- Epäsuora vaikutus
- Kokonaisvaikutus

Polkuanalyysin peruskäsitteet



- Suora vaikutus
- Epäsuora vaikutus
- Kokonaisvaikutus
- Näennäisvaikutus

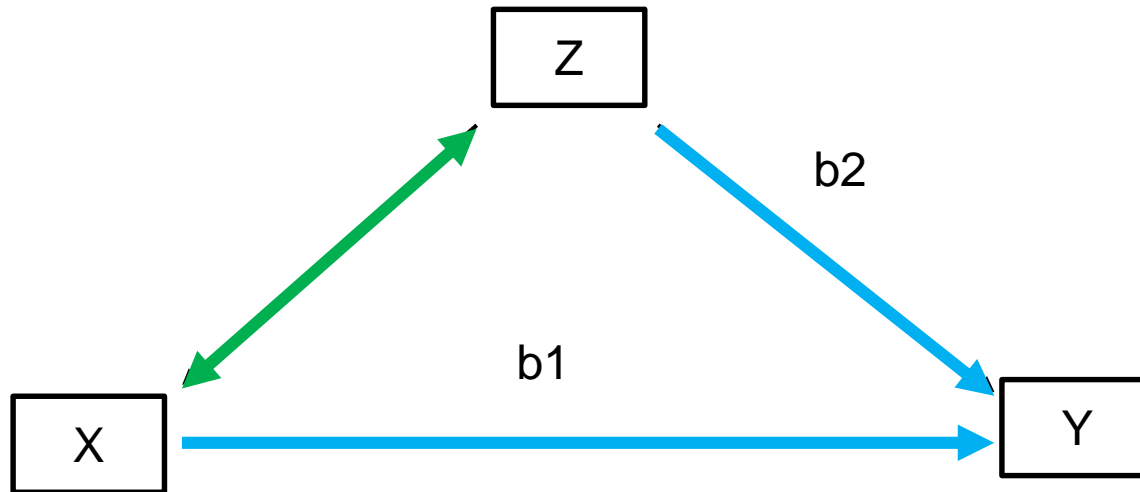
Polkuanalyysin peruskäsitteet



Suora vaikutus



X:n suora vaikutus Y:hyn
= b_1

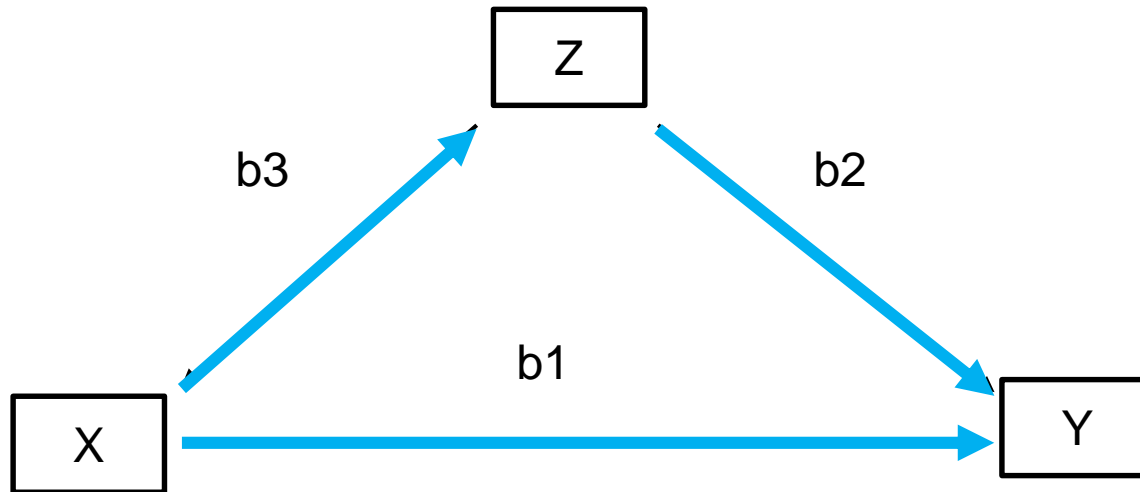




Epäsuora vaikutus

Lineaarisessa tapauksessa

X:n epäsuora vaikutus Y:hyn
 $= b_3 \cdot b_2$

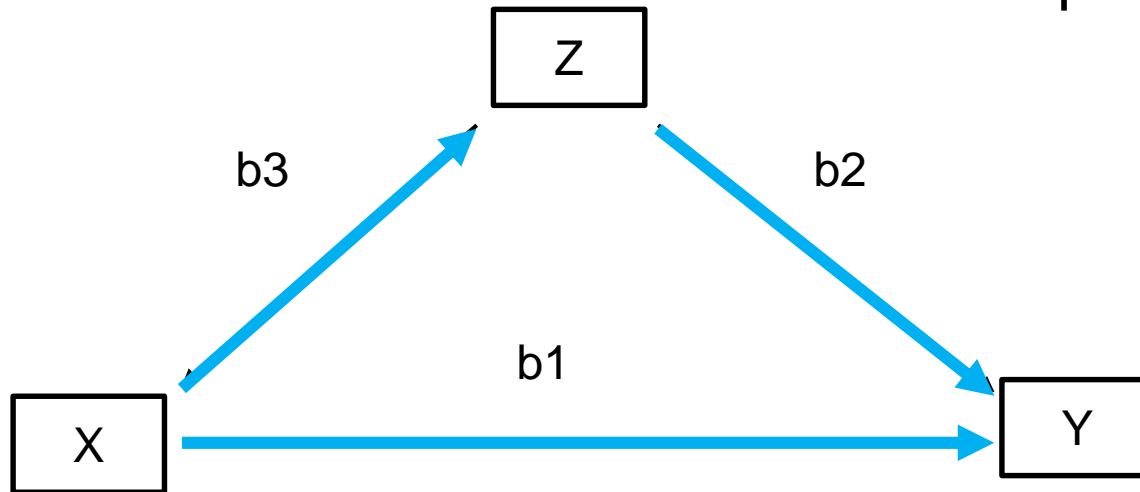




Kokonaisvaikutus

Lineaarisessa tapauksessa

X:n kokonaisvaikutus Y:hyn
= $b1 + b3 \cdot b2$
= suora + epäsuora vaikutus

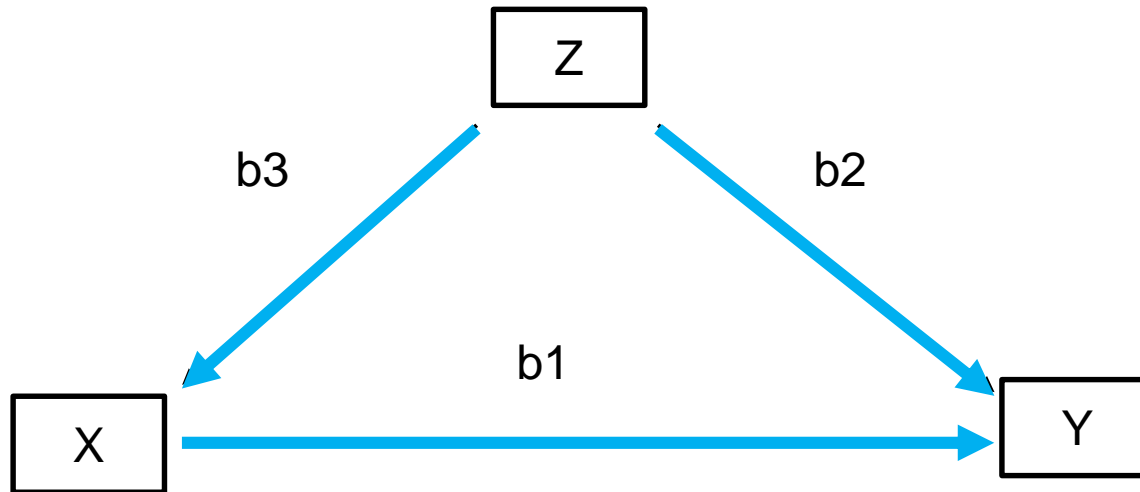




Kokonaisvaikutus

Lineaarisessa tapauksessa

X:n kokonaisvaikutus Y:hyn
= b_1
= suora vaikutus

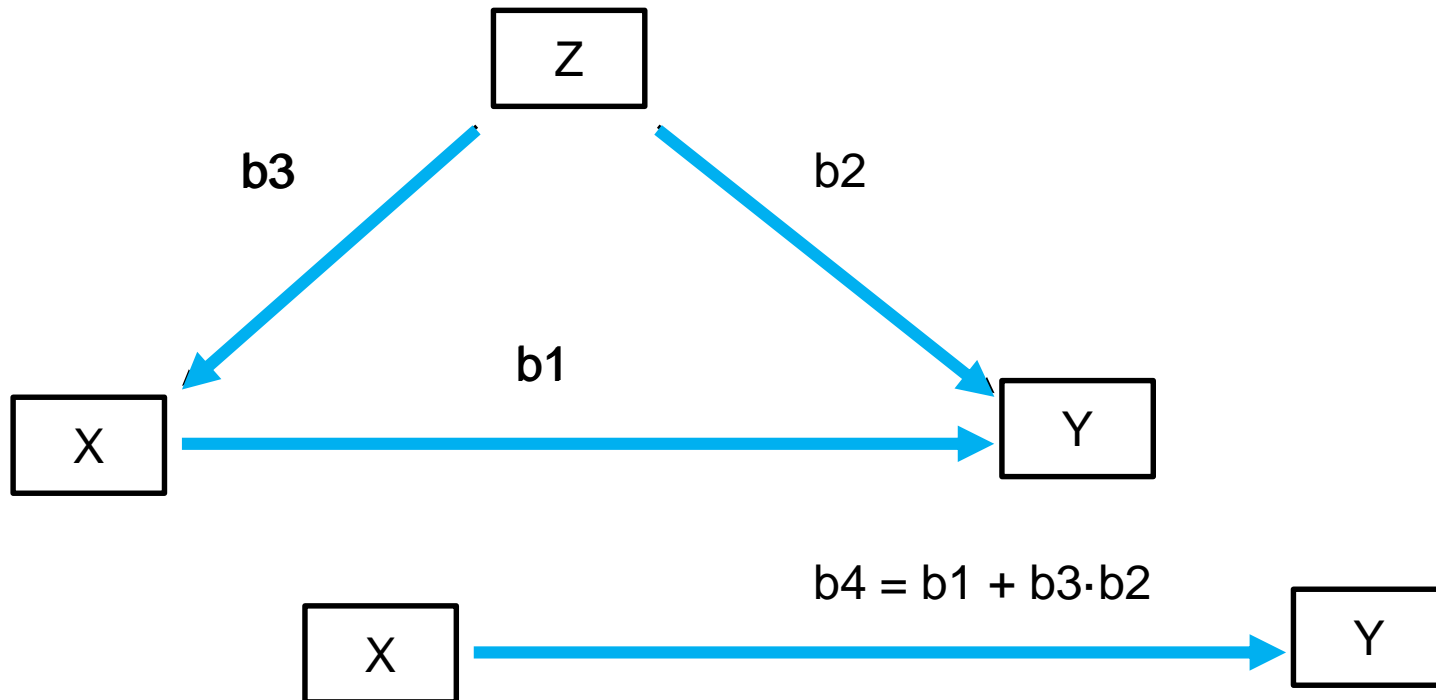




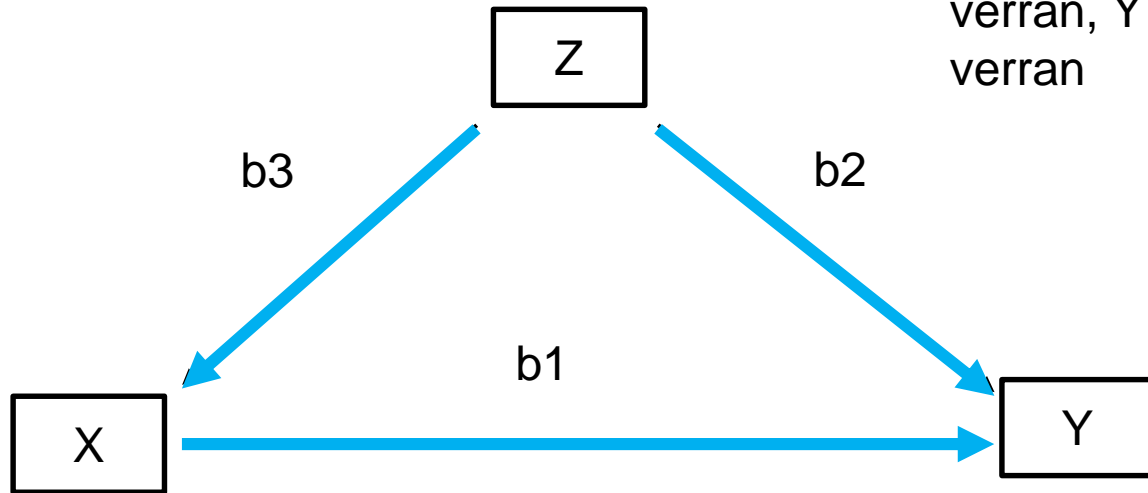
Näennäisvaikutus

Lineaarisessa tapauksessa

X:n näennäisvaikutus Y:hyn
 $= b4 - b1 = b3 \cdot b2$

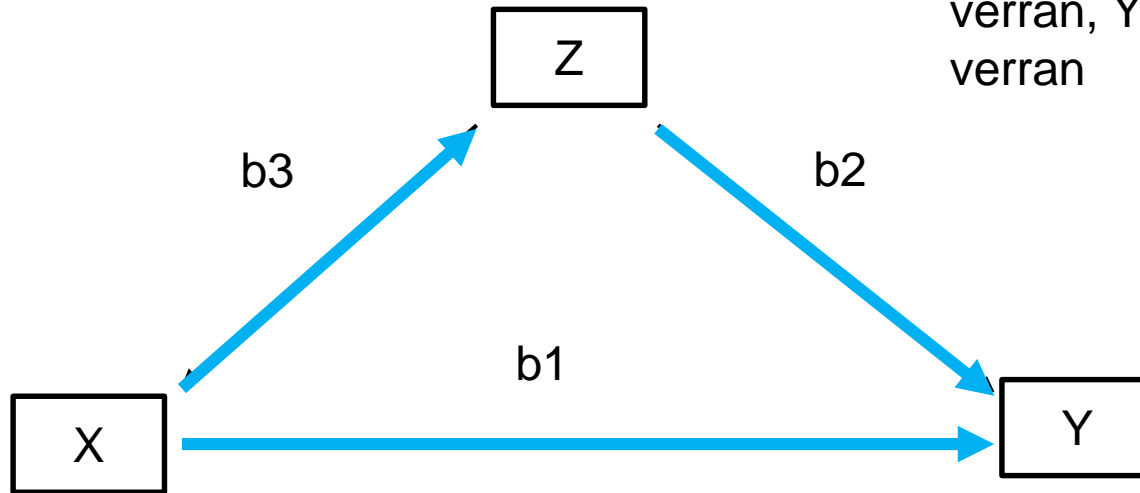


Suoran ja kokonaisvaikutuksen kausaalinen ero



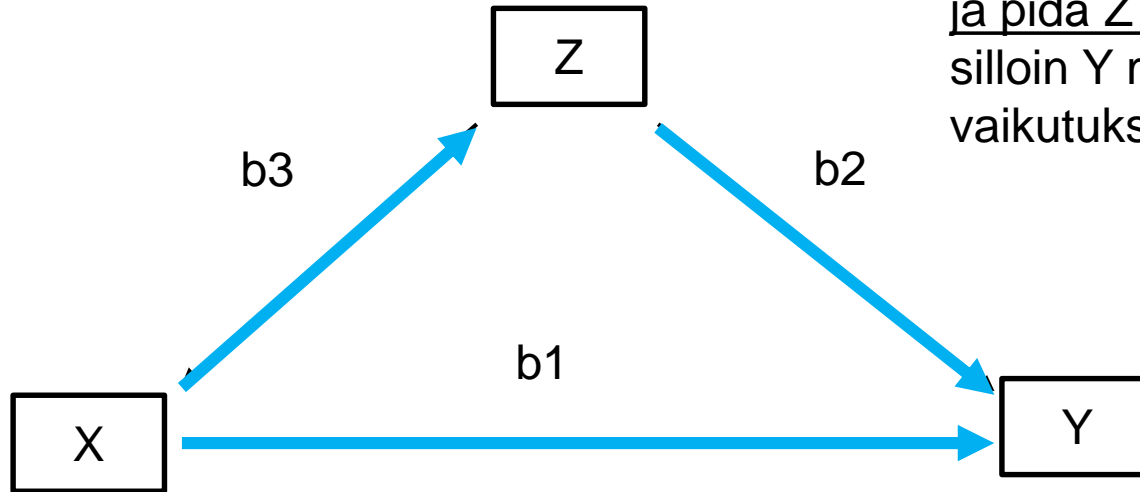
Muuta X:ää yhden yksikön verran, Y muuttuu b_1 :n verran

Suoran ja kokonaisvaikutuksen kausaalinen ero



Muuta X:ää yhden yksikön verran, Y muuttuu $(b1 + b3 \cdot b2)$:n verran

Suoran ja kokonaisvaikutuksen kausaalinen ero



Muuta X:ää yhden yksikön verran ja pidä Z muuttumattomana, silloin Y muuttuu suoran vaikutuksen (b1) verran

Miksi miehet saavat parempia gradu-arvosanoja kuin naiset?





Keskiarvot

	Naiset	Miehet	p
Tampere	3,97 (n = 3223)	3,95 (n = 1526)	.601
Turku	4,12 (n = 1966)	4,09 (n = 1522)	.476
Jyväskylä	4,08 (n = 7955)	4.22 (n = 4498)	.000000000002



Tieteenala	Sukupuoli	Keskiarvo	N
Arts	nainen		7000
	mies		2000
Sciences	nainen		2000
	mies		2000
A+S	nainen		9000
	mies		4000



Tieteenala	Sukupuoli	Keskiarvo	N
Arts	nainen	4,0	7000
	mies	4,0	2000
Sciences	nainen	4,5	2000
	mies	4,5	2000
A+S	nainen		9000
	mies		4000

Simpsonin paradoksi:



Tieteenala	Sukupuoli	Keskiarvo	N
Arts	nainen	4,0	7000
	mies	4,0	2000
Sciences	nainen	4,5	2000
	mies	4,5	2000
A+S	nainen	4,11	9000
	mies	4,25	4000



$$Y = 4,0 + 0,5 \text{ tieteenala} + 0 \text{ sukupuoli}$$

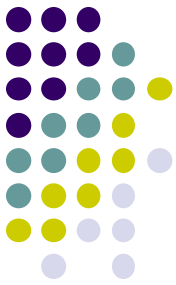
Tieteenala	Sukupuoli	Keskiarvo	N
Arts	nainen	4,0	2000
	mies	4,0	2000
Sciences	nainen	4,5	2000
	mies	4,5	2000
A+S	nainen	4,25	4000
	mies	4,25	4000



Report

gradun arvolause

Aine	sukup	Mean	N	Std. Deviation
ei ole Tampereella	nainen	4,45	1196	1,011
	mies	4,63	1240	1,063
	Total	4,54	2436	1,042
on Tampereella	nainen	4,02	6759	1,017
	mies	4,07	3258	1,175
	Total	4,04	10017	1,071
Total	nainen	4,08	7955	1,028
	mies	4,22	4498	1,172
	Total	4,14	12453	1,084



Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.246 ^a	.061	.059	1.052

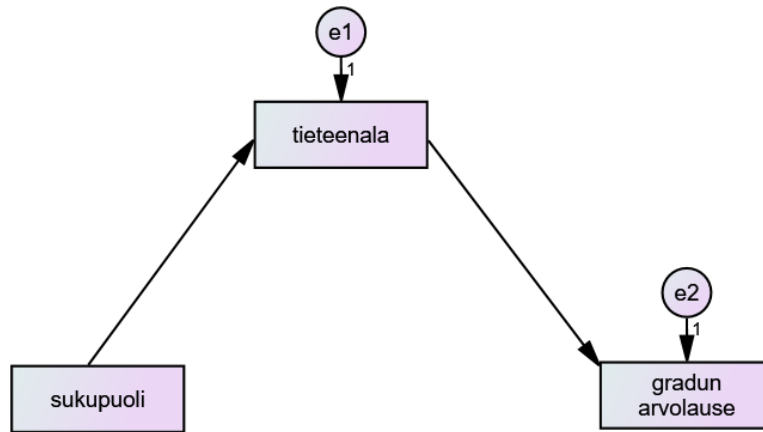
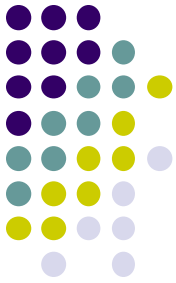
a. Predictors: (Constant), tieteenala, sukupuoli

Otos N = 1221

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	3.974	.039	100.860	.000	
	sukupuoli	.080	.064	.035	1.253	.211
	tieteenala	.637	.075	.238	8.492	.000

a. Dependent Variable: gradun arvolause



Otos N = 1221

```
Mplus - [jyvaskyla.out]
File Edit View Mplus Plot Diagram Window Help
Chi-Square Test of Model Fit
Value 1.413*
Degrees of Freedom 1
P-Value 0.2346
Scaling Correction Factor 1.1126
for MLR
* The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
for chi-square difference testing in the regular way. MLM, MLR and WLSM
chi-square difference testing is described on the Mplus website. MLMV, WLSMV,
and ULSMV difference testing is done using the DIFFTEST option.
RMSEA (Root Mean Square Error Of Approximation)
Estimate 0.018
90 Percent C.I. 0.000 0.081
Probability RMSEA <= .05 0.713
CFI/TLI
CFI 0.996
TLI 0.987
Ready Ln 1, Col 1
```