

# Random en systematische meetfouten

Harry B.G. Ganzeboom

ADEK UvS – College 3

5 maart 2011

# Recap

- Je hebt random en systematische meetfouten: onbetrouwbaarheid en invaliditeit.
- Random meetfouten spoor je op door de meting te herhalen.
- Systematische meetfouten spoor je op door de meetfout te herhalen.
- Wanneer je random meetfouten probeert op te sporen via alternatieve vraagformuleringen, ligt het voor de hand dat je systematische meetfouten introduceert.

# Opleiding en Beroep

- Hermeting van opleidingsniveau in hetzelfde interview via alternatieve formuleringen:
  - Hoogste kwalificatie
  - Totale opleidingsduur
- Hermeting van beroep via alternatieve formuleringen:
  - Open vraag, gecodeerd in standaardclassificatie
  - Gesloten vraag
  - (Beide worden uitgedrukt in statusschaal)

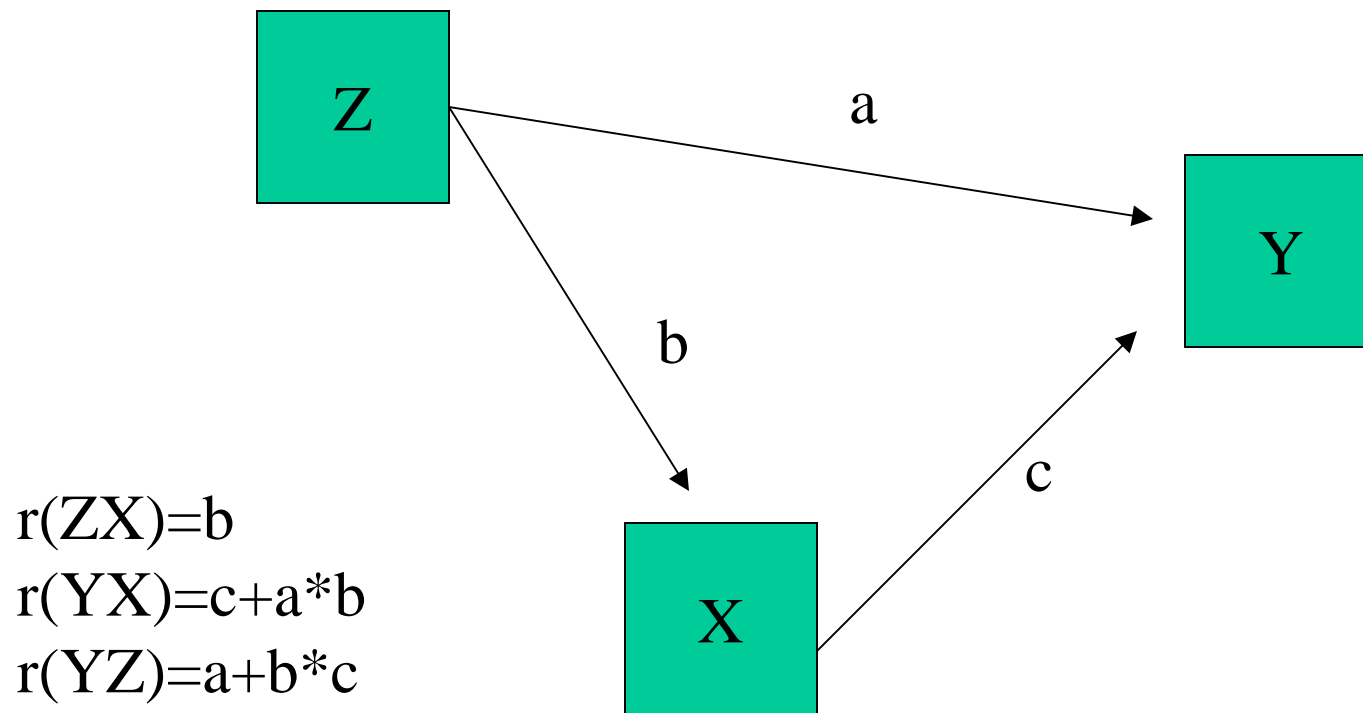
# Dubbele metingen opleiding

Correlations				
	EDLEV	EDDUR	PEDLEV	PEDDUR
EDLEV	1	,764	,645	,529
		,000	,000	,000
	2777	2636	1974	1876
EDDUR	,764	1	,536	,598
	,000		,000	,000
	2636	2657	1900	1872
PEDLEV	,645	,536	1	,782
	,000	,000		,000
	1974	1900	1982	1835
PEDDUR	,529	,598	,782	1
	,000	,000	,000	
	1876	1872	1835	1888

# Dubbele meting beroep

Correlations				
	ISEI1	ISEI2	PISEI1	PISEI2
ISEI1	1	,654	,409	,275
		,000	,000	,000
	2551	2458	1817	1779
ISEI2	,654	1	,312	,273
	,000		,000	,000
	2458	2559	1802	1788
PISEI1	,409	,312	1	,695
	,000	,000		,000
	1817	1802	1949	1838
PISEI2	,275	,273	,695	1
	,000	,000	,000	
	1779	1788	1838	1918

# Elementaire algebra van causale modellen

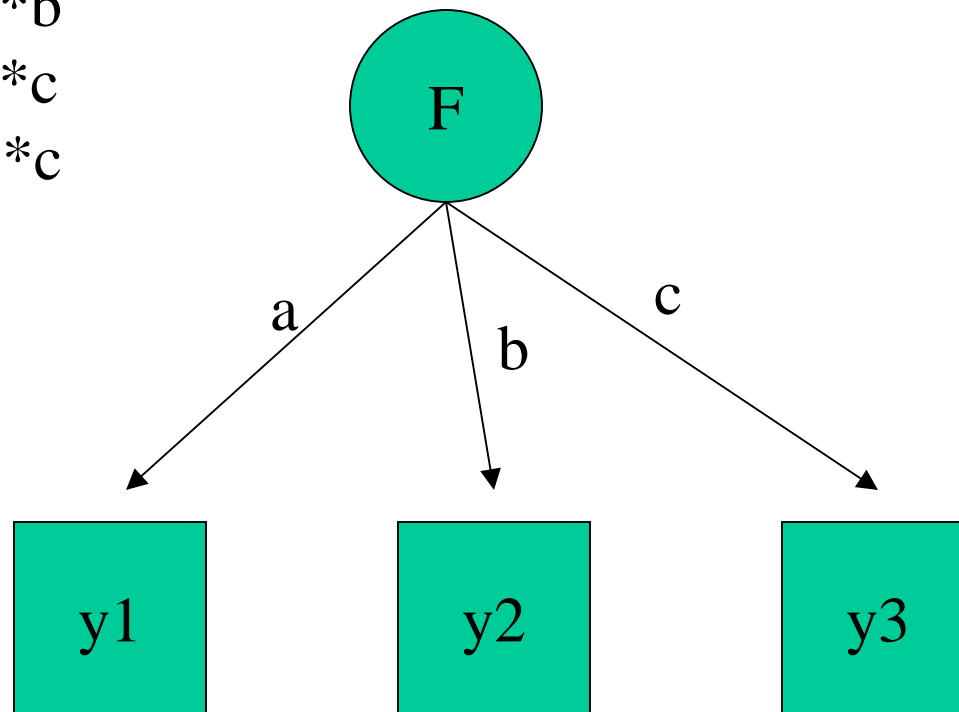


# Algebra van het elementaire causale model

- Y: afhankelijke variabele
- X: onafhankelijke (oorzaak) variabele
- Z: confounder
- Totale correlatie =
  - Direct effect +
  - + indirect effect
  - + schijnbaar causaal effect
- Uit drie correlaties kun je de drie partiële effecten berekenen (drie vergelijkingen met drie onbekenden).

# Meten als causaal model

$$r(y1,y2)=a*b$$
$$r(y1,y3)=a*c$$
$$r(y2,y3)=b*c$$





# Metten als causaal model

- Het enkelvoudige factormodel met drie indicatoren is precies geïdentificeerd: drie correlaties (vergelijkingen) met drie onbekenden.
- Als je twee indicatoren zou hebben, kun je wel een gemiddelde  $a$  en  $b$  bepalen ( $a=b$ ), maar niet hun verhouding.
- Daarom: goed meten begint bij drie indicatoren!!
- (Maar met twee indicatoren lukt het ook in een uitgebreider causaal model.)

# Dubbele metingen, twee concepten, alleen random meetfouten

$$r(x1,x2)=a*b$$

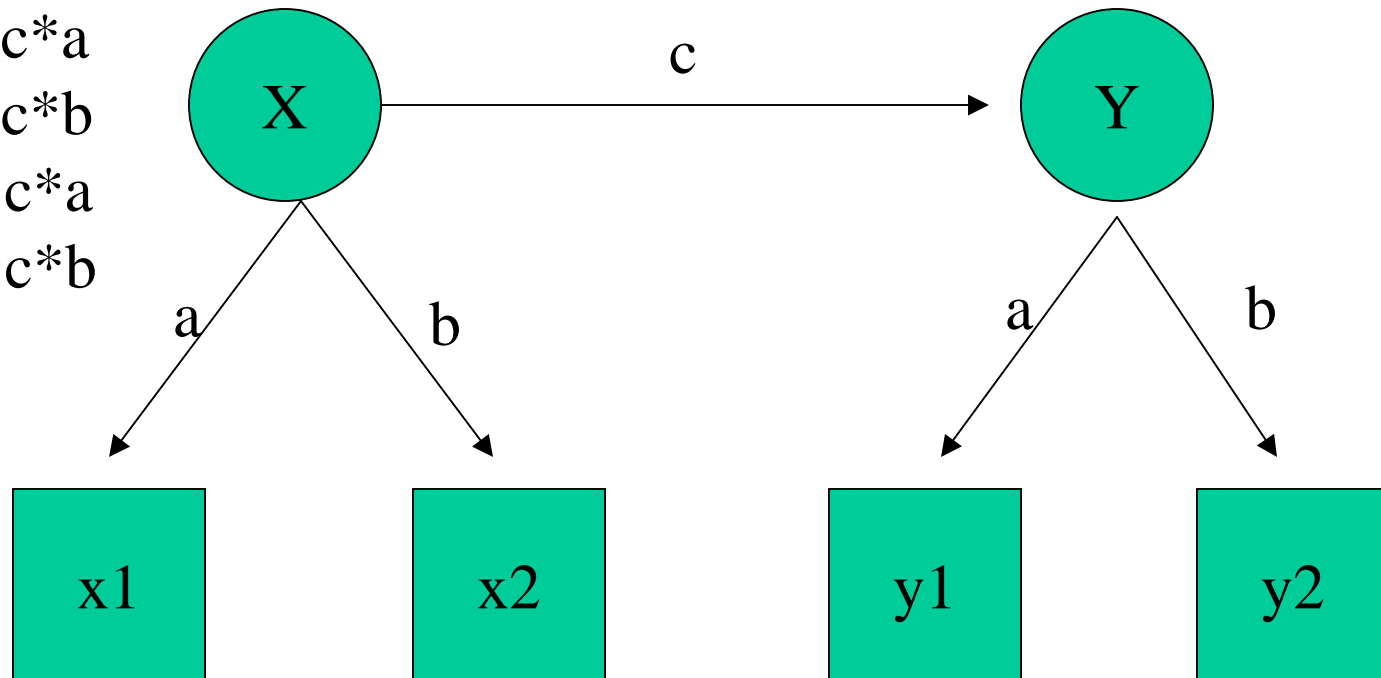
$$r(y1,y2)=a*b$$

$$r(x1,y1)=a*c*a$$

$$r(x1,y2)=a*c*b$$

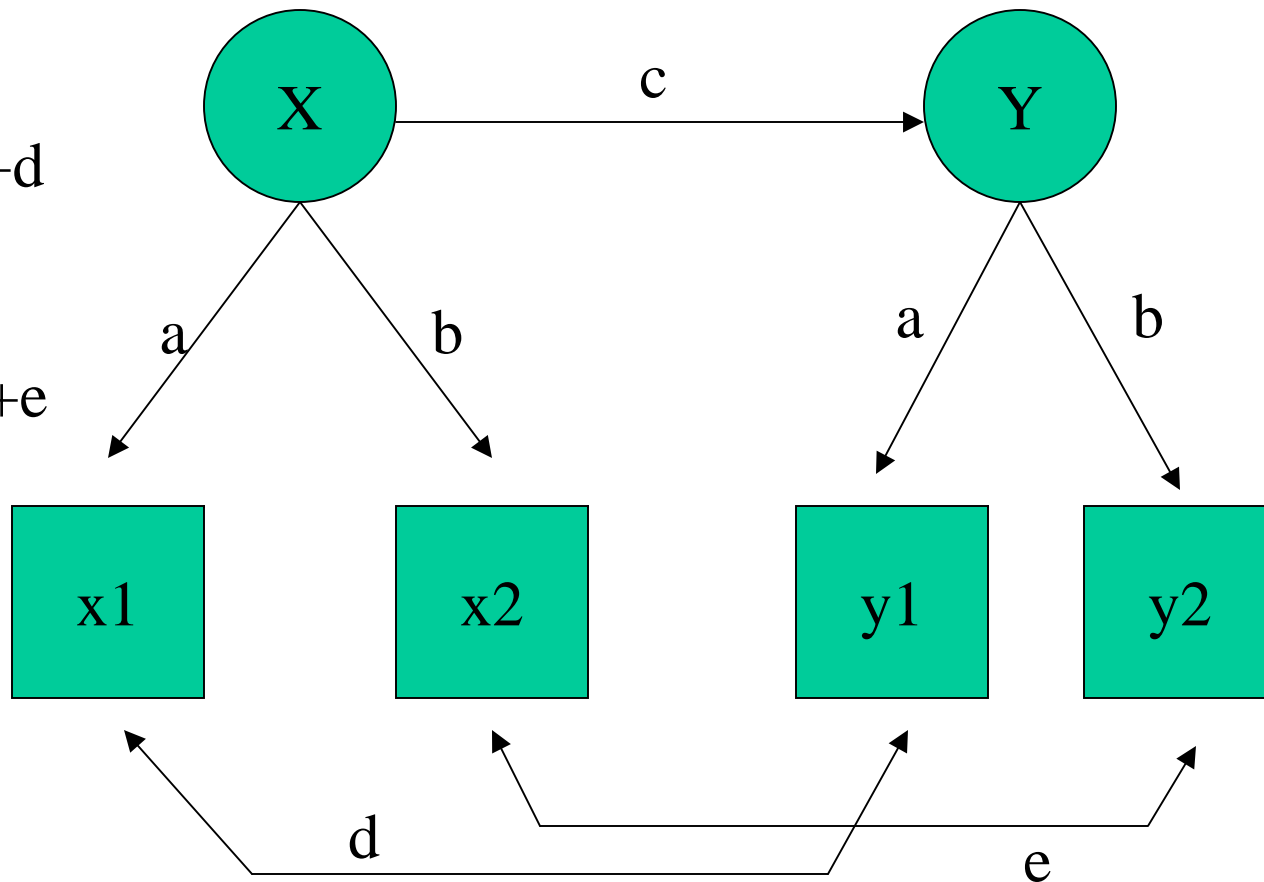
$$r(x2,y1)=b*c*a$$

$$r(x2,y2)=b*c*b$$



# Dubbele metingen, twee concepten, random en systematische meetfouten

$$\begin{aligned}r(x1,x2) &= a*b \\ r(y1,y2) &= a*b \\ r(x1,y1) &= a*c*a+d \\ r(x1,y2) &= a*c*b \\ r(x2,y1) &= b*c*a \\ r(x2,y2) &= b*c*b+e\end{aligned}$$



# Dubbele metingen, twee concepten, random en systematische meetfouten

- Het model is het meest elementaire voorbeeld van een MTMM (multi-trait multi-method) model.
- Het lijkt (over)geïdentificeerd (6 vergelijkingen met 5 onbekenden), maar het is het niet.
- Identificatie is er wel als we een groter model (met meer variabelen) schatten.

# Opleidingen en beroepen

	EDLEV	EDDUR	PEDLEV	PEDDUR	ISE1	ISE2	PISE1	PISE2
EDLEV	1,000	0,764	0,645	0,529	0,578	0,521	0,381	0,325
EDDUR	0,764	1,000	0,536	0,598	0,495	0,438	0,321	0,271
PEDLEV	0,645	0,536	1,000	0,782	0,396	0,353	0,640	0,567
PEDDUR	0,529	0,598	0,782	1,000	0,331	0,300	0,571	0,509
ISE1	0,578	0,495	0,396	0,331	1,000	0,654	0,409	0,275
ISE2	0,521	0,438	0,353	0,300	0,654	1,000	0,312	0,273
PISE1	0,381	0,321	0,640	0,571	0,409	0,312	1,000	0,695
PISE2	0,325	0,271	0,567	0,509	0,275	0,273	0,695	1,000

# Ook zonder algebra kun je zien..

- Dat EDLEV aanzienlijk minder random meetfouten heeft dan EDDUR.
- Dat ISEI1 (gesloten) aanzienlijk minder random meetfouten heeft dan ISEI2 (open).
- Dat er ook systematische (gecorrleerde) meetfouten zijn in EDDUR en wellicht ISEI1.

# De Vries & Ganzeboom, 2008

- Vergelijken en modelleren de meetkwaliteit van gesloten en open vraagstelling naa beroepsstatus.
- Data: vier afleveringen ISSP-NL.
- Variabelen: beroepen van respondent, vader en moeder (dubbel gemeten) + opleidingen en inkomen (enkelvoudig gemeten).

# Beroepsmeting -- open

- Naar beroep wordt doorgaans gevraagd met een open vraagstelling. Er zit veel codeerwerk in.
- Het coderen gebeurt aan de hand van een standaard beroepenclassificatie, doorgaans de International Standard Classification of Occupations (ISCO).
- Dit is veel werk en er worden vast ook veel fouten mee gemaakt.
- In analyses worden de beroepscode omgezet in een statusschaal, zoals de International Socio-Economic Index [ISEI].



# Beroepsmeting -- gesloten

- In de ISSP 1987 is een gesloten vraagstelling gehanteerd. Deze bestaat in wezen uit een zeer vergroefde ISCO (1-digit) + een aantal voorbeeld beroepen.
- In vijf ISSP landen zijn naast de gesloten vragen ook open vragen gesteld. Deze gegevens zijn lang blijven liggen, maar Ganzeboom (2005) heeft ze geanalyseerd.

# Ganzeboom, 2005

- Gesloten vraag is betrouwbaarder dan open vraag!
- Systematische meetfouten niet geanalyseerd, maar deze zijn bij nader inzien tamelijk gering.
- De gesloten vraag wordt nu ook gebruikt in ESS en in ISSP 2009.

# De Vries & Ganzeboom, 2008

- Gesloten vraag is iets betrouwbaarder dan open vraag.
- Er is nauwelijks sprake van systematische meetfouten.
- Toepassing van beide metingen leidt tot inhoudelijk andere conclusies, met name waar het inkomensverwerving aangaat.