SSJデータ･アーカイブ　第6回公開セミナー
社会調査の2次分析シリーズ 第2回
2002年11月11日 (月) 15：30-17：30

共分散構造分析の基礎と実際
----応用編----

狩野　裕（大阪大学 大学院人間学研究科）

SEMによる
因果分析について

因果の方向を探る
因果の大きさを正確に測定する

因果推論

- 因果の方向を如何に探るか
  - 適合度, 同値モデル, 操作変数法, 共分散選択
- 因果の大きさを如何に正確に (バイアスなく) 測るか
  - 第三変数の影響, 共変量選択 (バックドア規準)

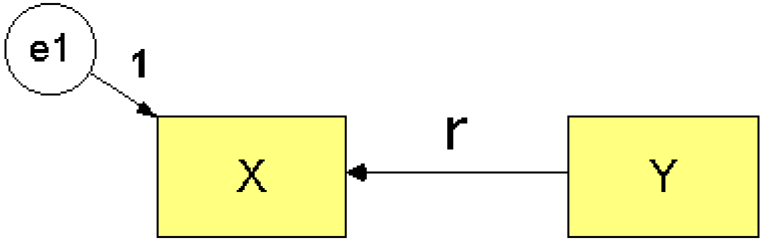
近年の発展

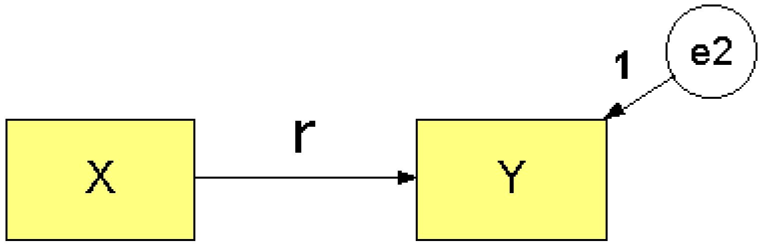
- 適合度による比較 (SEM)
  - モデルが大きいとき, 特に有益
  - 同値モデルの同定が瞬時
- 操作変数法
  - 受身的な変数選定ではない
  - 興味ある因果の同定に関して積極的に働きかける変数選定
- 双方向因果モデル (非逐次モデル) の推定

相関係数から因果の方向は決まらない ---同値モデルの問題---

相関構造

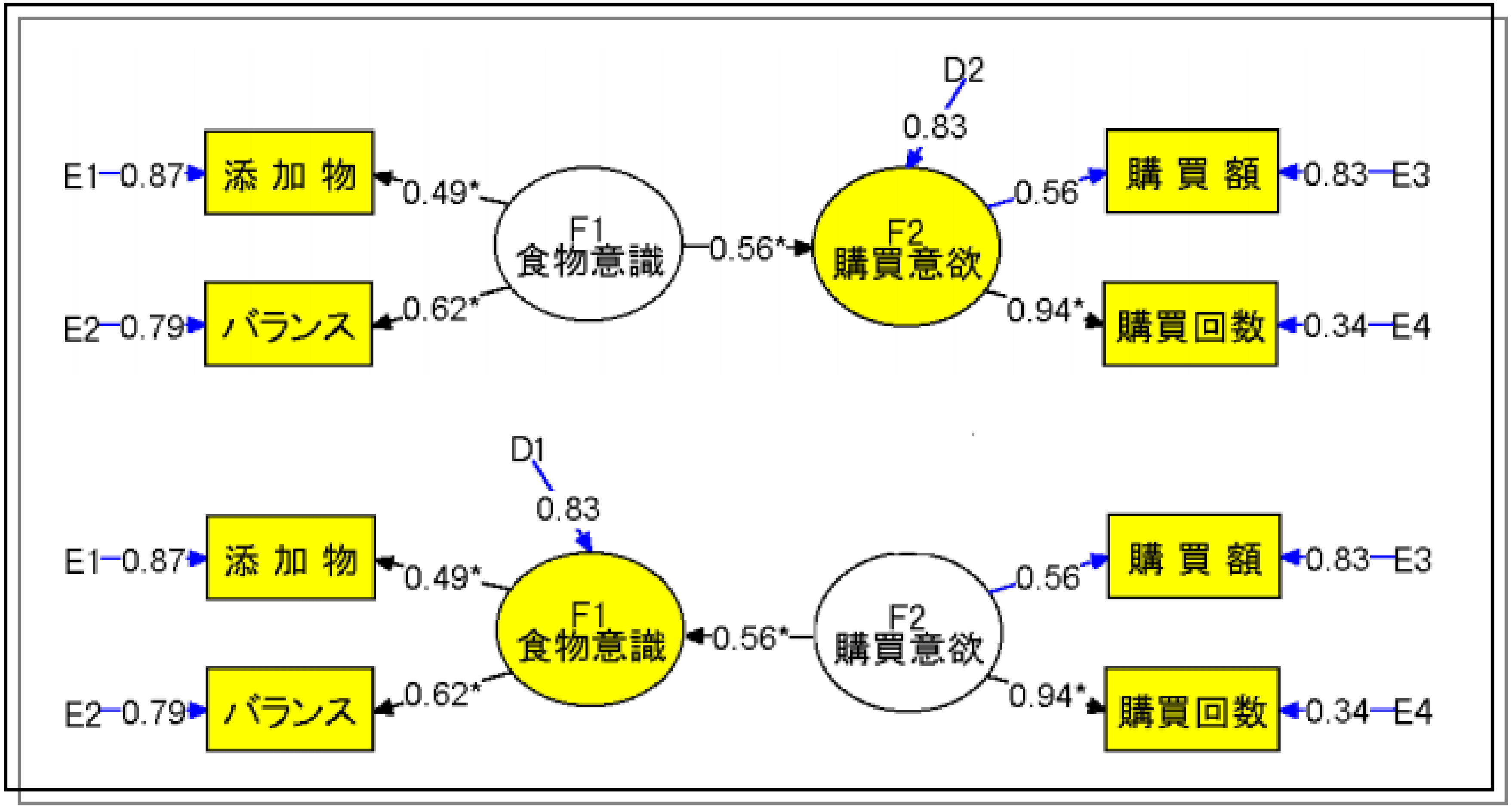
	X	Y
X	1	r
Y	r	1





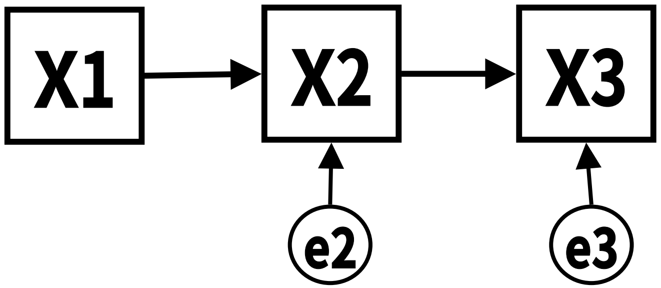
- データから区別できないモデルを同値モデルという
- 「区別できない」とは適合度が同一であることをいう

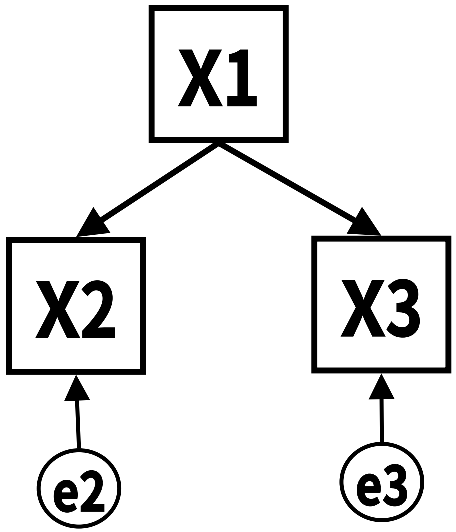
同値モデル例


D2
0.83
E1
0.87
添 加 物
0.49*
F1
食物意識
0.56*
F2
購買意欲
0.56
購 買 額
0.83
E3
E2
0.79
バランス
0.62*
0.94*
購買回数
0.34
E4
D1
0.83
E1
0.87
添 加 物
0.49*
F1
食物意識
0.56*
F2
購買意欲
0.56
購 買 額
0.83
E3
E2
0.79
バランス
0.62*
0.94*
購買回数
0.34
E4

Blalock-Simonによる古典的パス解析

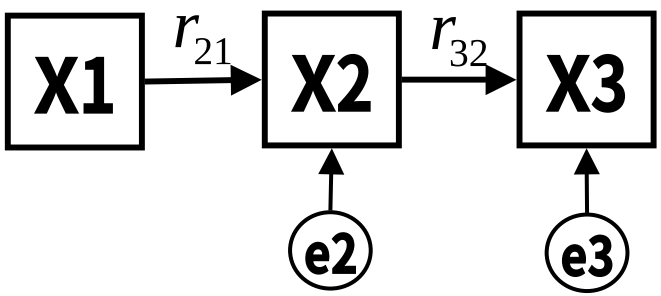
- 連鎖モデルと擬相関モデル

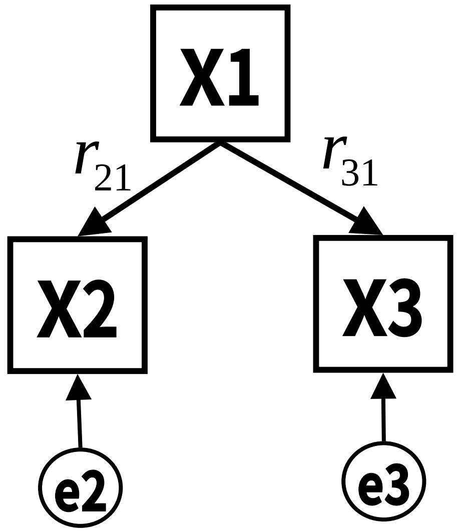




X1: 社会･経済的発展
X2: 政党間の競争
X3: 福祉政策の創出

モデルと相関係数の関係





$$Cor(X_1, X_3) = r_{32} r_{21}$$

or

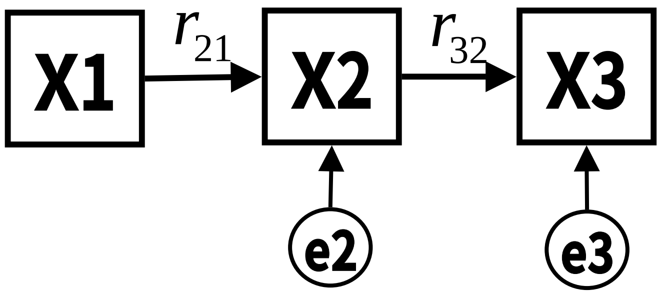
$$r_{13\bullet 2} = 0$$

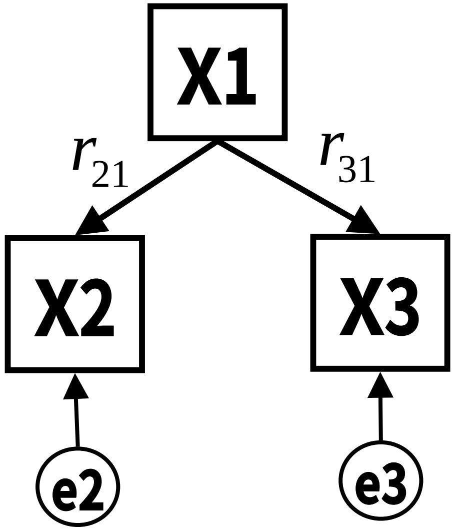
$$Cor(X_2, X_3) = r_{31} r_{21}$$

or

$$r_{23\bullet 1} = 0$$

モデルと相関係数の関係

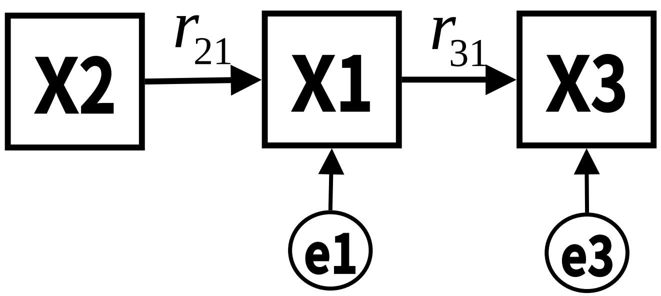


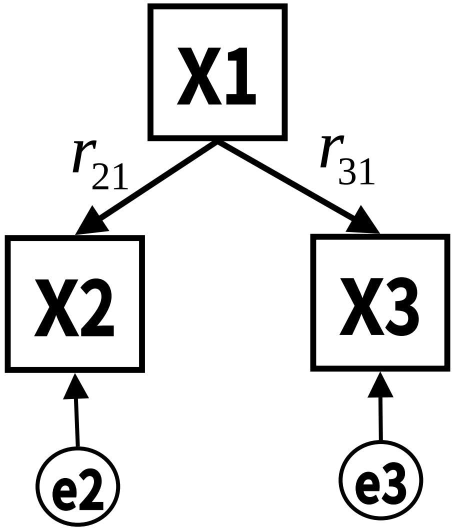


	X1	X2	X3
X1	1		
X2	r_{21}	1	
X3	$r_{21}r_{32}$	r_{32}	1

	X1	X2	X3
X1	1		
X2	r_{21}	1	
X3	r_{31}	$r_{21}r_{31}$	1

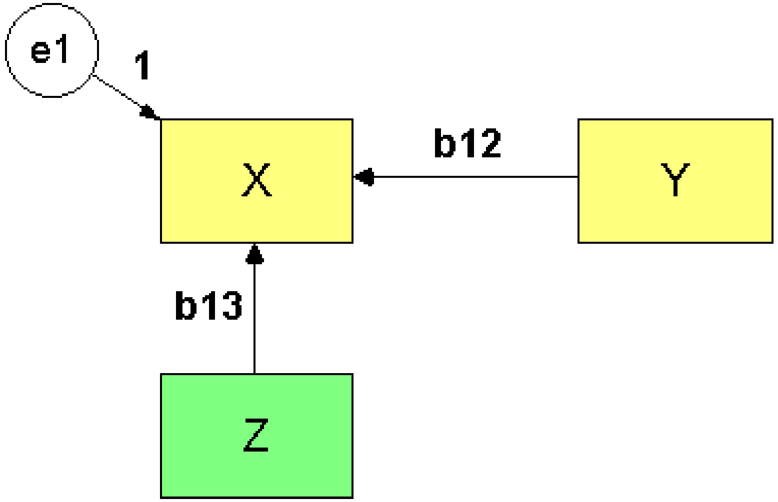
注：いつもうまくいくとは限らない

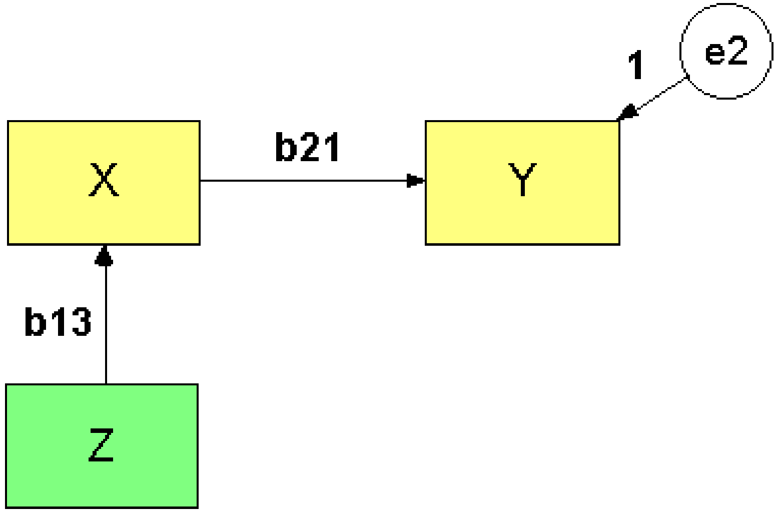




	X1	X2	X3
X1	1		
X2	r_{21}	1	
X3	r_{31}	$r_{21}r_{31}$	1

因果の方向を決める：
操作変数法 (Instrumental variable method)





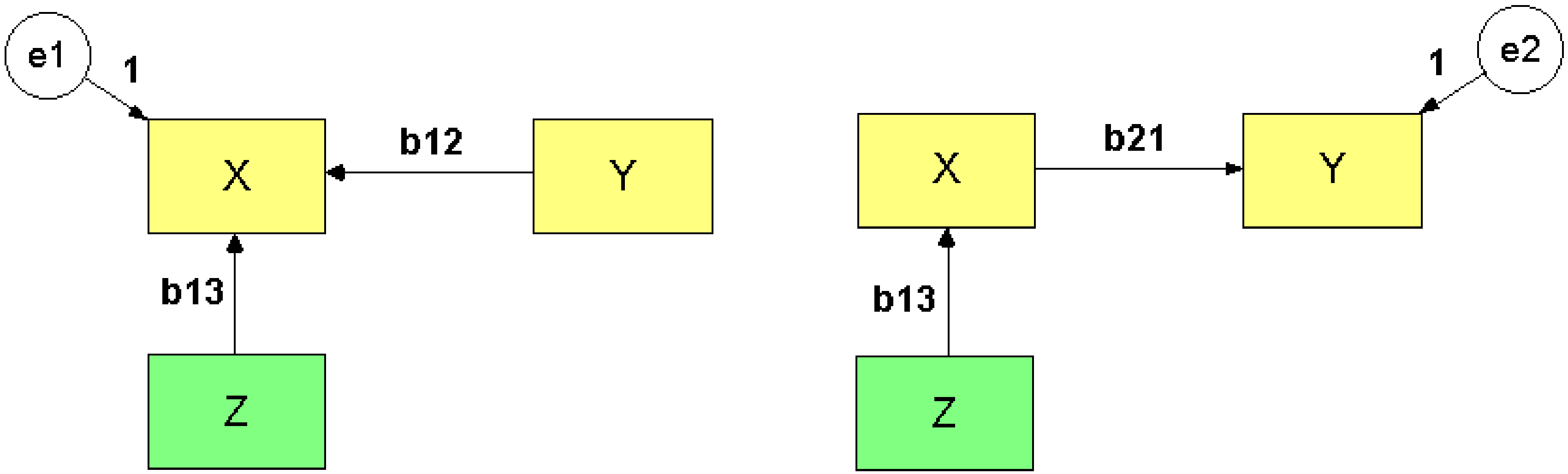
相関構造

	X	Y	Z
X	1	b12	b13
Y		1	0
Z			1

相関構造

	X	Y	Z
X	1	b21	b13
Y		1	b21b13
Z			1

因果の方向を決める：適合度との関係



適合度が低い

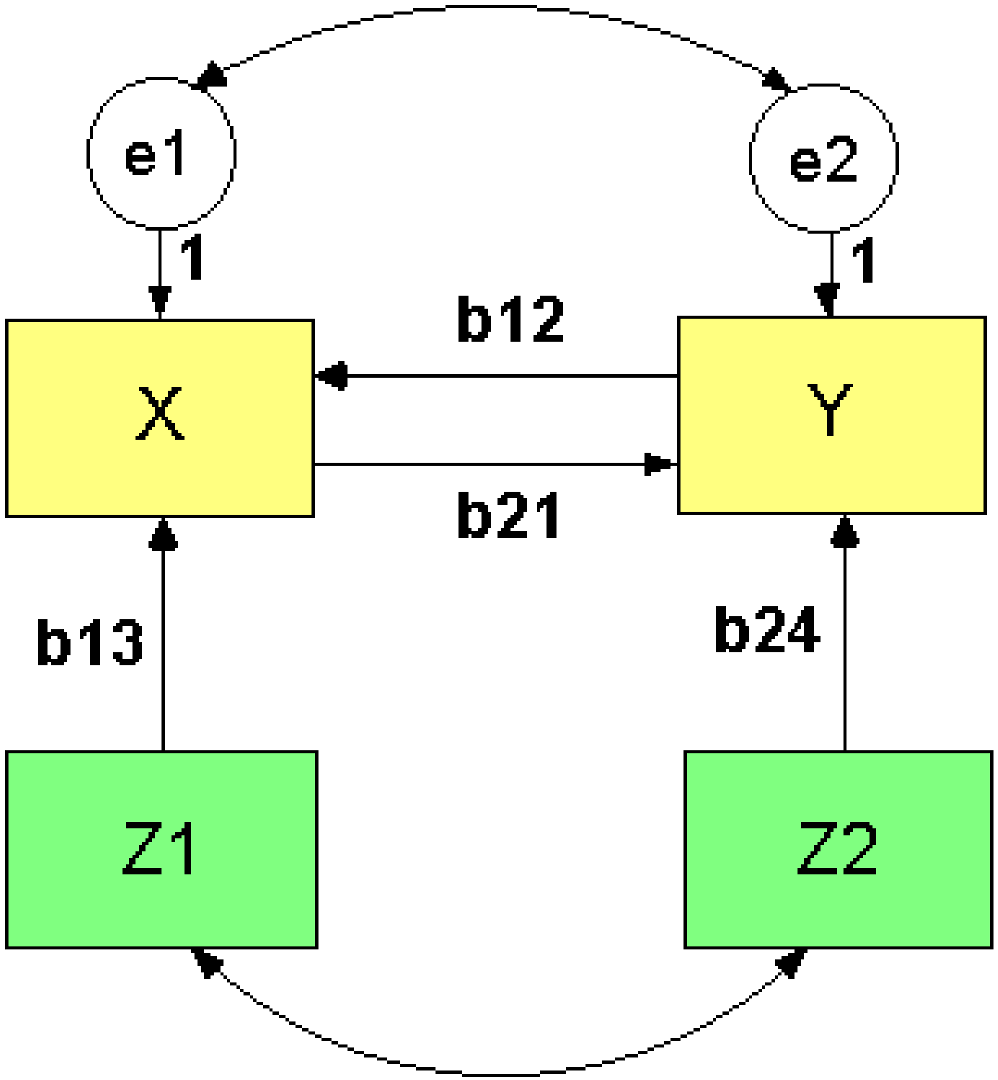
適合度が高い

X→Y の因果関係が示唆される

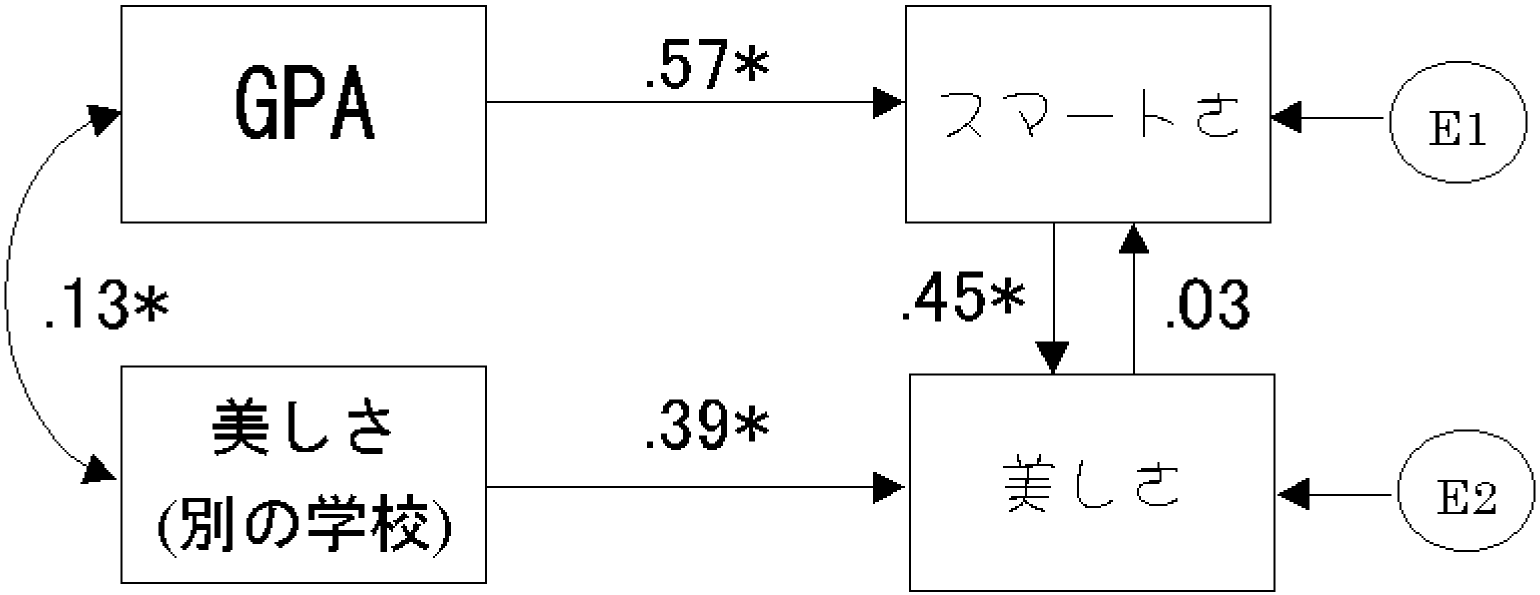
操作変数法とは

- X，Yのいずれかに影響を及ぼし，他方への直接効果をもたない変数Z（操作変数）を観測する
- X，Y，Zの相関構造から，X→Y or X←Yを判断する

双方向因果モデル (非逐次モデル)

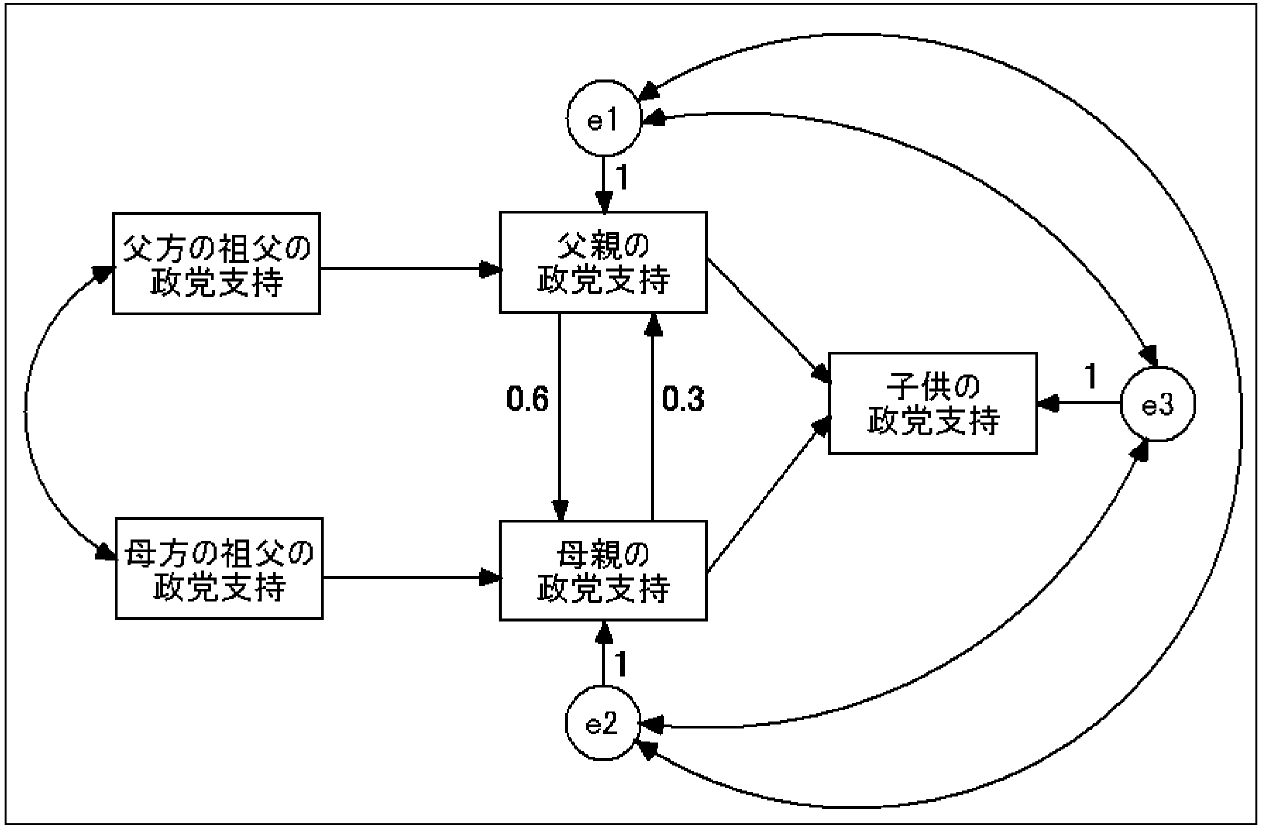


例1：Attractiveness implies perceived academic ability?



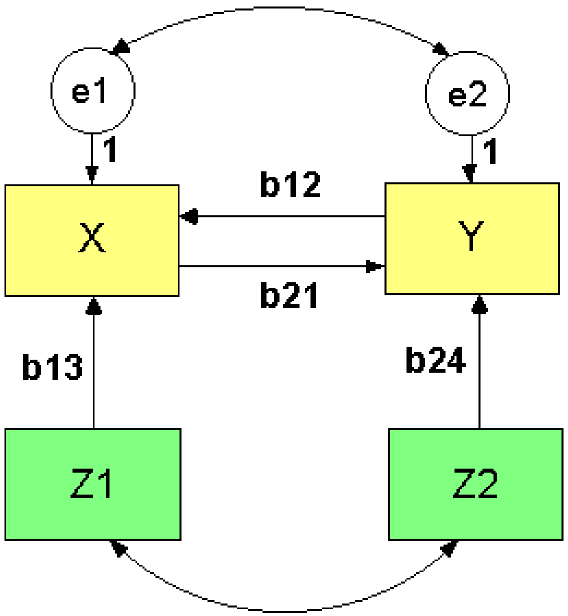
出展:
AMOSマニュアル

例2：政治的社会化モデル



出典：Asher(1976). Causal Modeling. Sage

双方向因果モデルの基礎仮定



$$\underbrace{\begin{bmatrix} X \\ Y \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} X \\ Y \end{bmatrix}}_{\mathbf{y}} + \underbrace{\begin{bmatrix} b_{13}Z_1 \\ b_{24}Z_2 \end{bmatrix}}_{\Gamma\mathbf{z}} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{\mathbf{e}}$$

$$\Rightarrow \mathbf{y} = B\mathbf{y} + \Gamma\mathbf{z} + \mathbf{e}$$

$$\mathbf{y}_{t+1} = B\mathbf{y}_t + \Gamma\mathbf{z} + \mathbf{e}_t \quad (t = 0,1,2,\cdots,\infty)$$

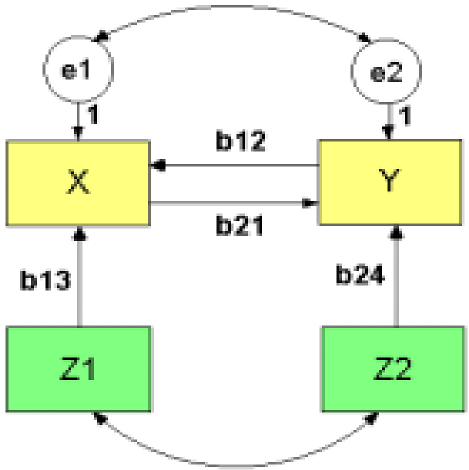
$$\Rightarrow$$

$$\mathbf{y} = B\mathbf{y} + \Gamma\mathbf{z} + \mathbf{e}$$

$$\Rightarrow$$

$$\mathbf{y} = (I - B)^{-1}\left(\Gamma\mathbf{z} + \mathbf{e}\right)$$

双方向因果モデルの解釈



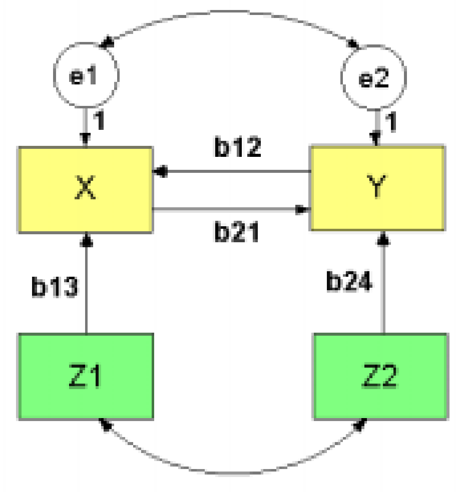
$$\mathbf{y} = (I - B)^{-1}(\Gamma\mathbf{z} + \mathbf{e})$$

$$\mathbf{y}_{t+1} = B\mathbf{y}_t + \Gamma\mathbf{z} + \mathbf{e}_t \quad (t = 0,1,2,\cdots,\infty)$$

$$\Rightarrow$$

$$\begin{aligned}
\mathbf{y}_{t+1} &= B\left(B\mathbf{y}_{t-1} + \Gamma\mathbf{z} + \mathbf{e}_{t-1}\right) + \Gamma\mathbf{z} + \mathbf{e}_t \\
&= B^2\mathbf{y}_{t-1} + B\Gamma\mathbf{z} + B\mathbf{e}_{t-1} + \Gamma\mathbf{z} + \mathbf{e}_t \\
&= B^2\mathbf{y}_{t-1} + \left(B + I\right)\Gamma\mathbf{z} + B\mathbf{e}_{t-1} + \mathbf{e}_t \\
&\cdots \\
&= B^{t+1}\mathbf{y}_0 + \left(\sum_{k=0}^{t} B^k\right)\Gamma\mathbf{z} + \left(\sum_{k=0}^{t} B^k\mathbf{e}_{t-k}\right) \\
&\overset{d}{=} B^{t+1}\mathbf{y}_0 + \left(\sum_{k=0}^{t} B^k\right)\Gamma\mathbf{z} + \left(\sum_{k=0}^{t} B^k\right)\mathbf{e} \quad (\mathbf{e}_t \sim \mathrm{Normal}) \\
&\to \left(I - B\right)^{-1}\left(\Gamma\mathbf{z} + \mathbf{e}\right) \qquad \left(\mathbf{y}_0 = \mathbf{0}\right)
\end{aligned}$$

双方向因果モデルの基礎仮定



$$\mathbf{y}_{t+1} = B\mathbf{y}_t + \Gamma\mathbf{z} + \mathbf{e}_t \quad (t = 0,1,2,\cdots,\infty)$$

$$\lim_{t\to\infty}\mathbf{y}_t = (I - B)^{-1}(\Gamma\mathbf{z} + \mathbf{e})$$

$$\mathbf{y} = (I - B)^{-1}(\Gamma\mathbf{z} + \mathbf{e})$$

- X，Yは，ある初期値(0)からスタートして，相互に無限回，影響し合った結果である
- 影響の大きさは変化せずBである
  - XとYの相互の影響関係が安定しているべき

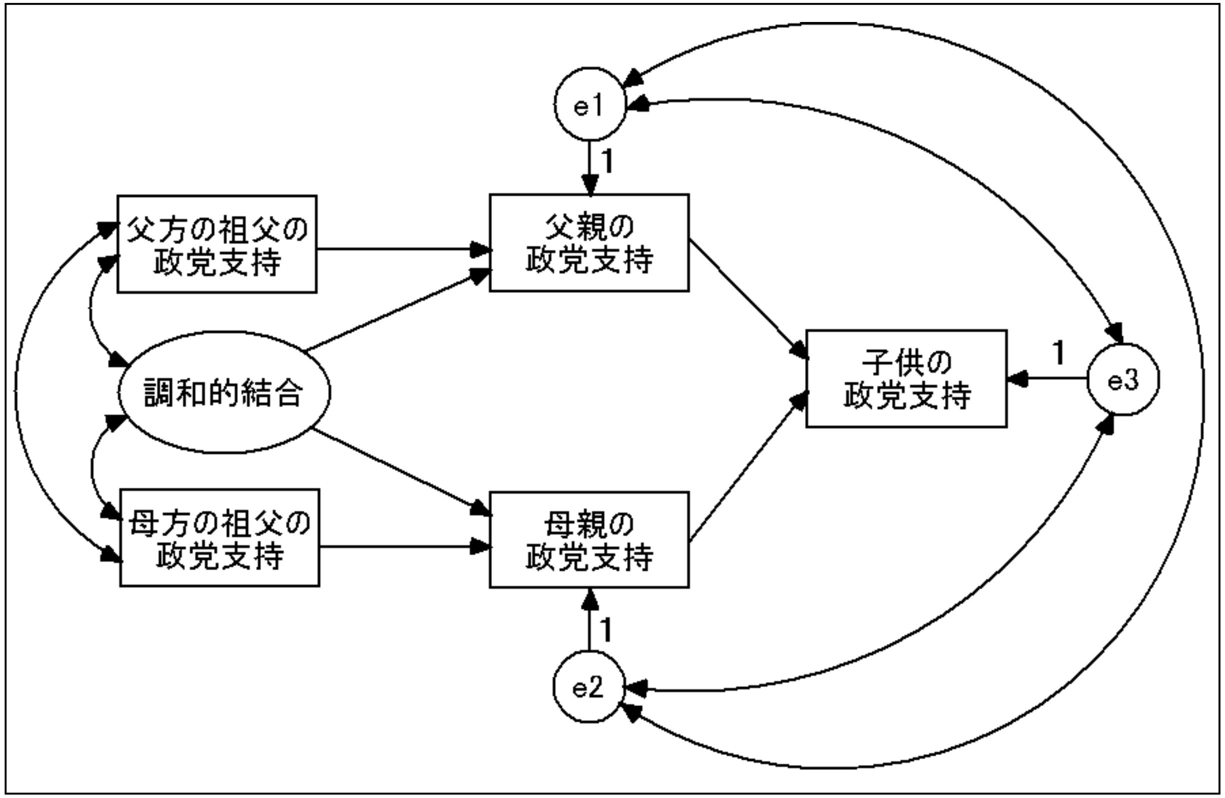
共分散構造分析による因果の決定

- 因果の方向に興味があるとき
  - 当該モデルが適合する
  - 対立モデルが適合しない
- 対立モデルが同値モデルにならないようなモデリングが必要
  - そのための方法が操作変数(道具的変数)の導入

有効性

- 因果を決定したか？
  - X→Y，X←Y，X←→Yのいずれかであるという仮定の下で，どれがデータに矛盾しないかを決定しているにすぎない．つまり，「因果関係にない」という可能性を排除していない
    - 交絡変数の問題
    - 横断的データに基づく因果推論には限界がある
  - 「X→Y」が分かっても，「XはYの“主要な”原因である」とは言いきれない．Ｒ２の大きさに注意すべきである．R2=0.1でもモデルは適合する（モデルの適合度とＲ２は別概念）．このような場合，XはYの原因のごく一部である

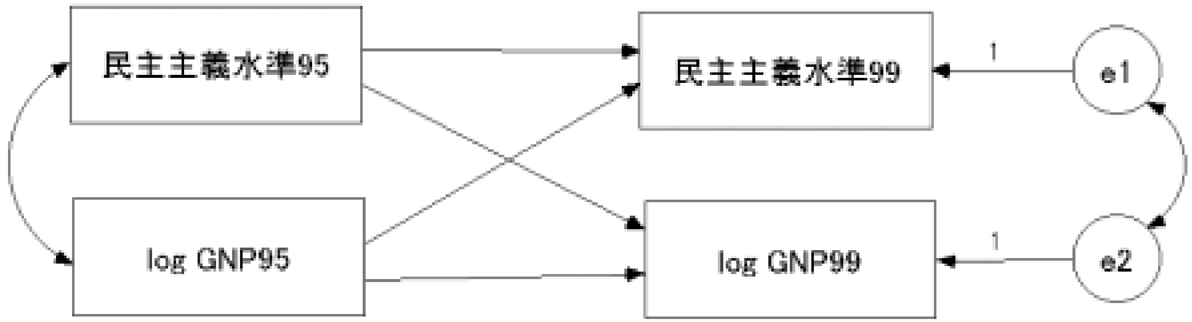
ウィークポイント：交絡変数

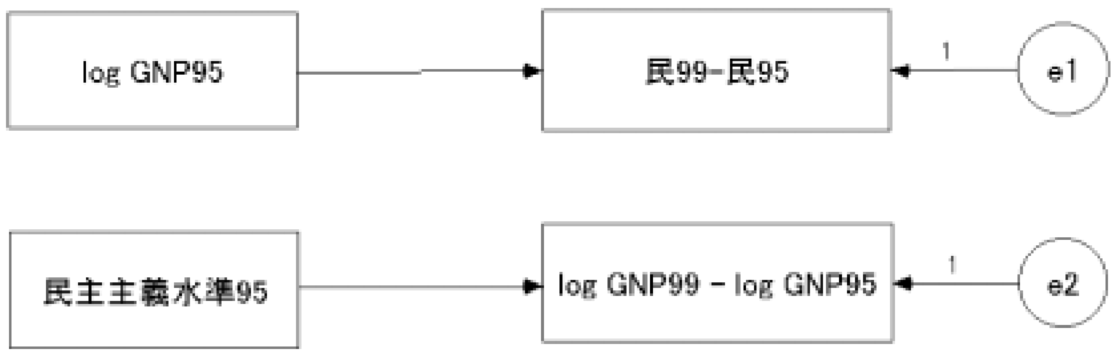


盛山(1986, 行動計量学)

縦断的データの利用

- 2時点でデータをとり，時間差を利用する
  - 民主主義⇒経済発展 or 経済発展⇒民主主義





因果の大きさを正確に測定する

第三変数（未分析変数）の影響

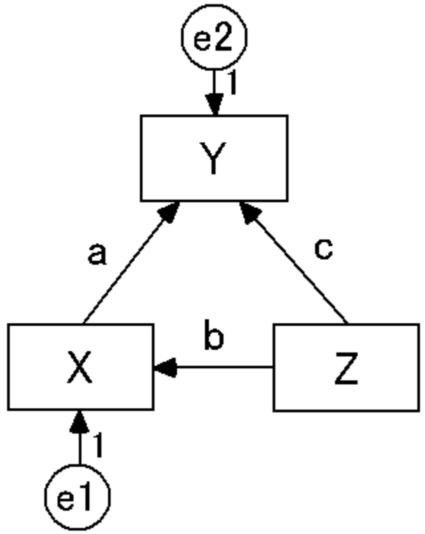
調査データ分析の最大の
ウィークポイント

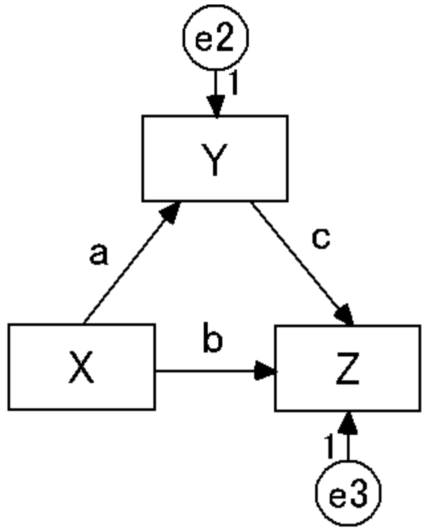
実験研究と調査研究

- 実験研究
  - 原因系の変数に水準を設定し，水準ごとに無作為化を行いデータを採取する
  - 水準の恣意性に問題
  - 無作為割付けができれば，第三変数 (or 剰余変数；未分析の交絡変数) の影響をシャットアウトできる
- 調査研究 (観察研究)
  (observational study; correlational study)
  - 無作為に被験者を抽出して全変数を観測する
  - 第三変数の影響が深刻

第三変数とは

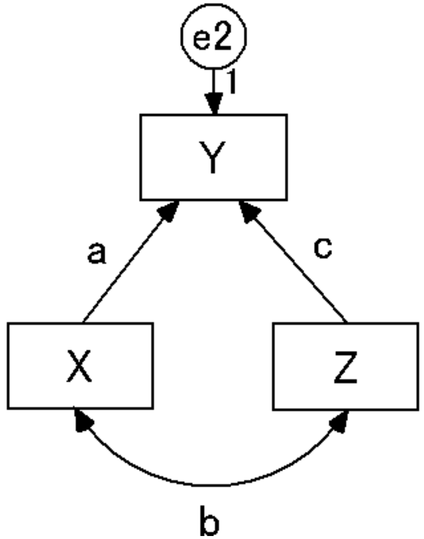
交絡変数

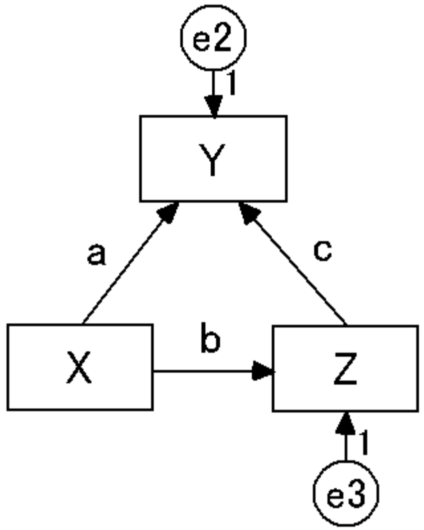




合流点

交絡変数



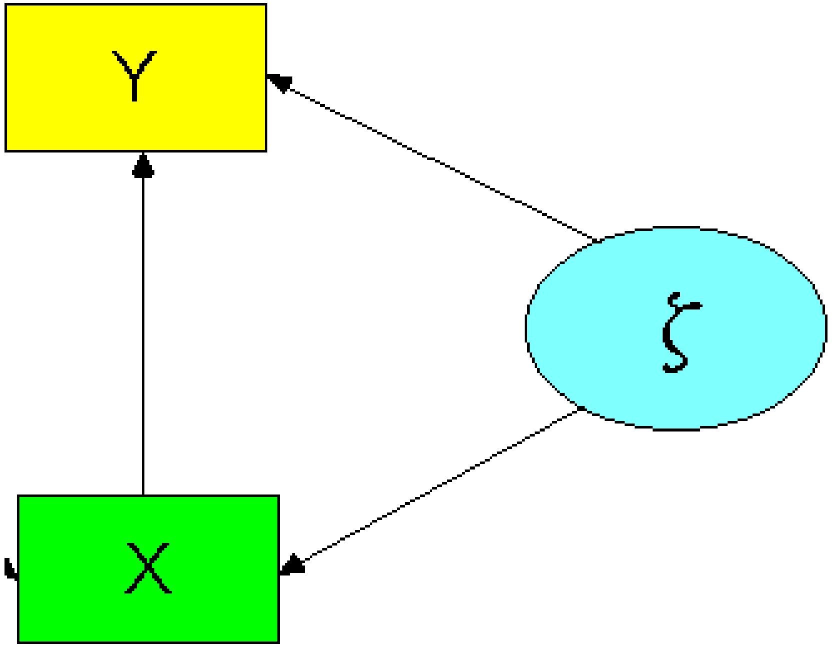


中間変数

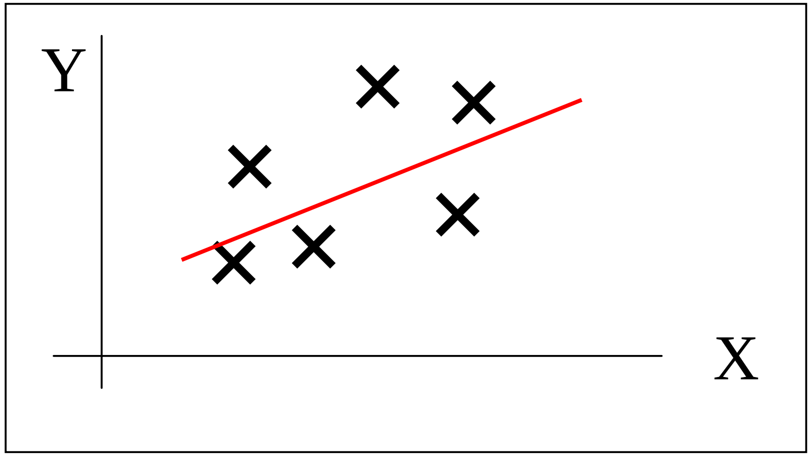
交絡変数いろいろ

- 第三変数
- 剰余変数
- 二次変数
- 媒介変数

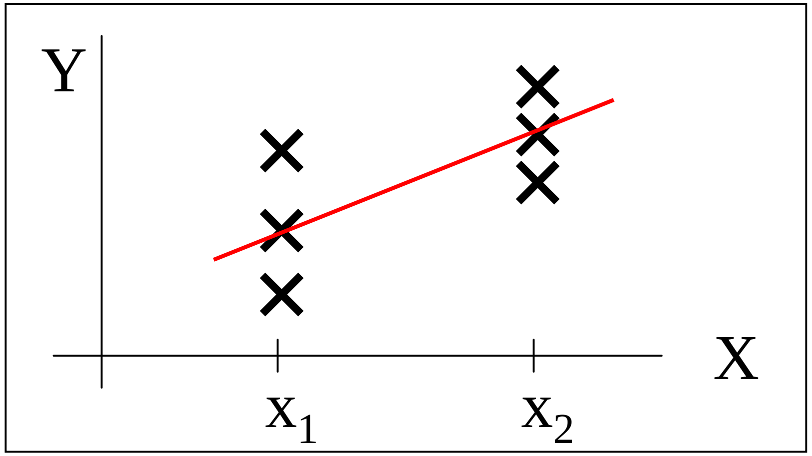
交絡変数の影響



調査研究にはバイアスが生じる

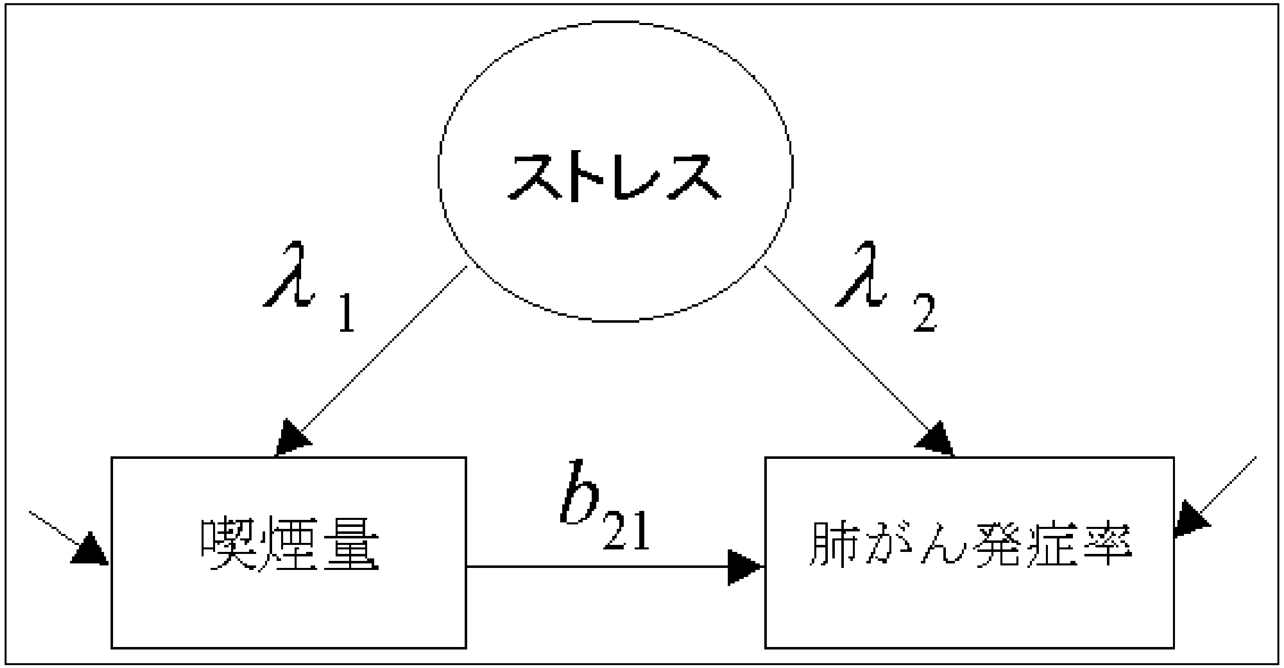


Xを無作為割付けすると
バイアスが生じない



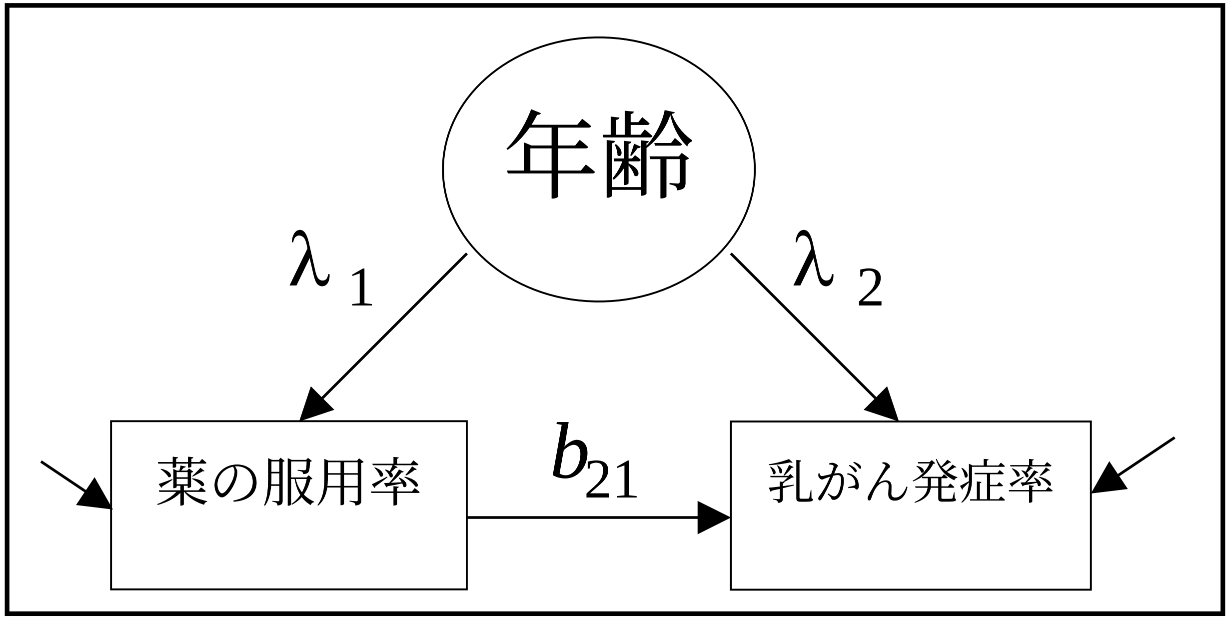
交絡変数の影響：例1

$$Cor(\text{喫煙量},\ \text{肺がん発症率}) = b_{21} + \lambda_1\lambda_2$$



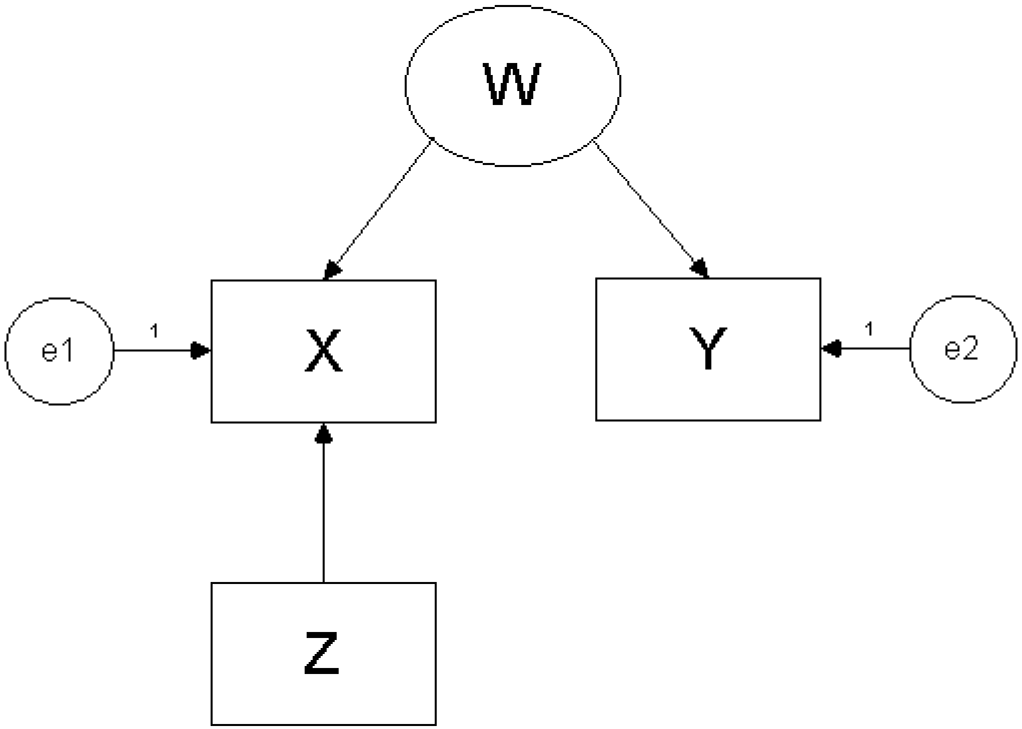
交絡変数の影響：例2

Cor(薬服用率, 乳がん発症率) ＝ $b_{21} + \lambda_1\lambda_2$



出典：Jick et al (1974). Reserpine and breast cancer *Lancet*, September 21.

因果方向決定にも影響



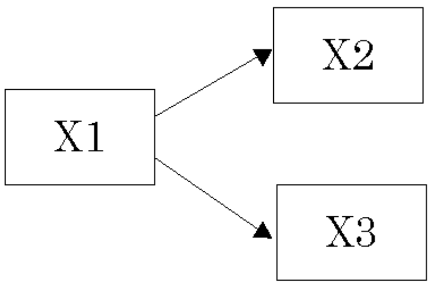
	X	Y	Z
X	1		
Y	*	1	
Z	*	0	1

誤ってY→Xと結論してしまう

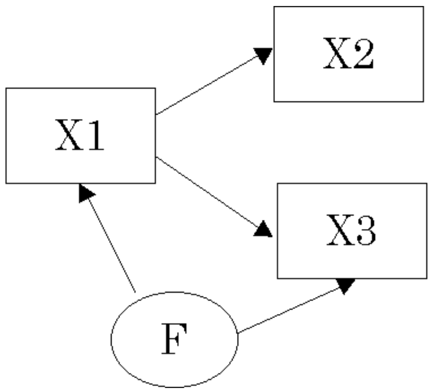
適合度による
交絡変数の影響の検出

検出可能な例

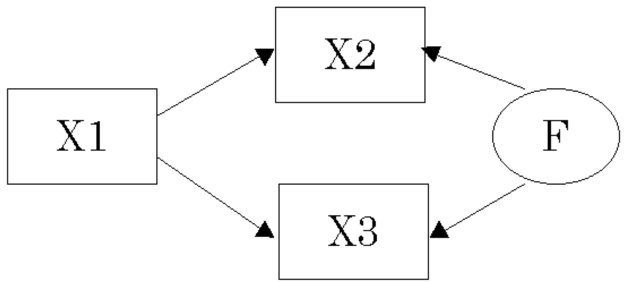
モデル



真の状況1



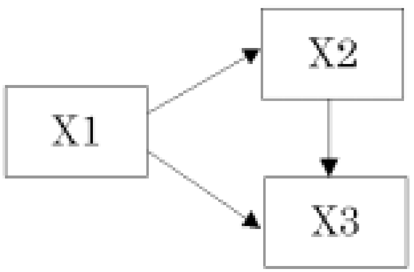
真の状況2



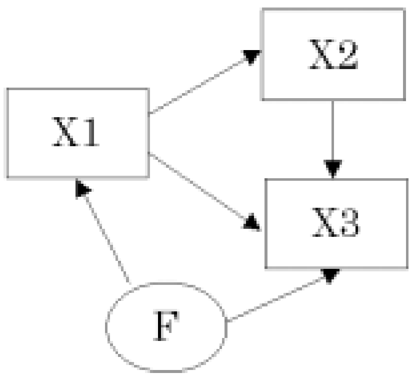
適合度による
交絡変数の影響の検出

検出できない例

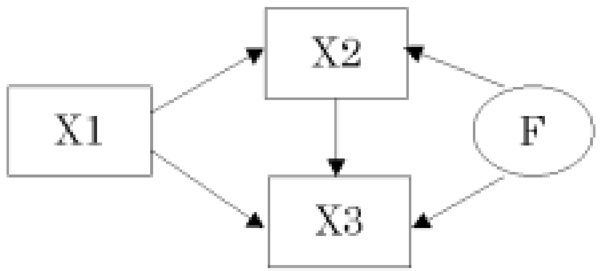
モデル



真の状況1



真の状況2

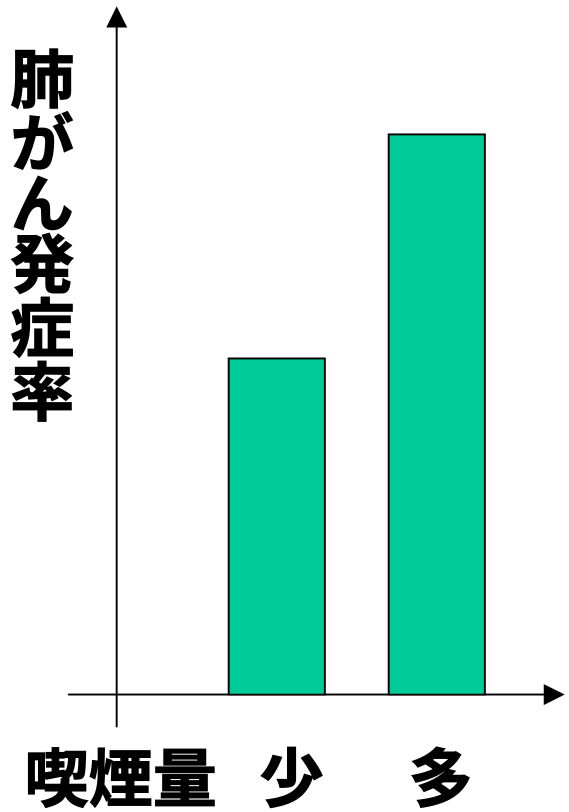


実験研究っぽい調査研究

1) ランダムに被験者を選び，喫煙量により群へ振分け

2) 喫煙量の多いｸﾞﾙｰﾌﾟと少ないｸﾞﾙｰﾌﾟのそれぞれからランダムに被験者を選ぶ

3) ランダムに被験者を選び，喫煙量を割付ける



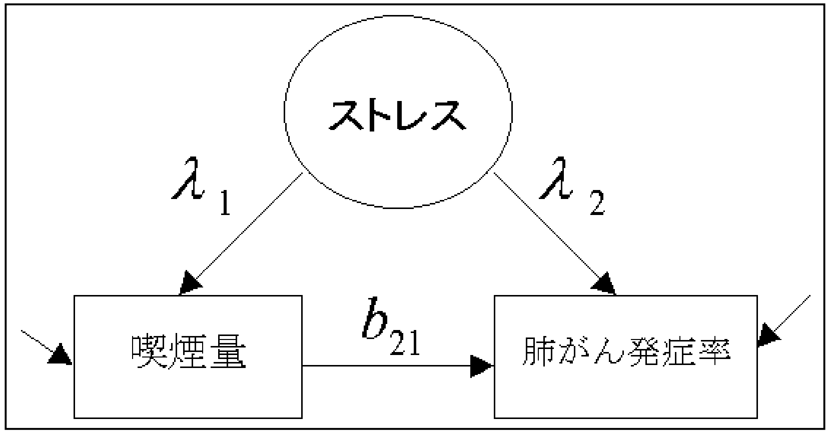
実験研究っぽい調査研究

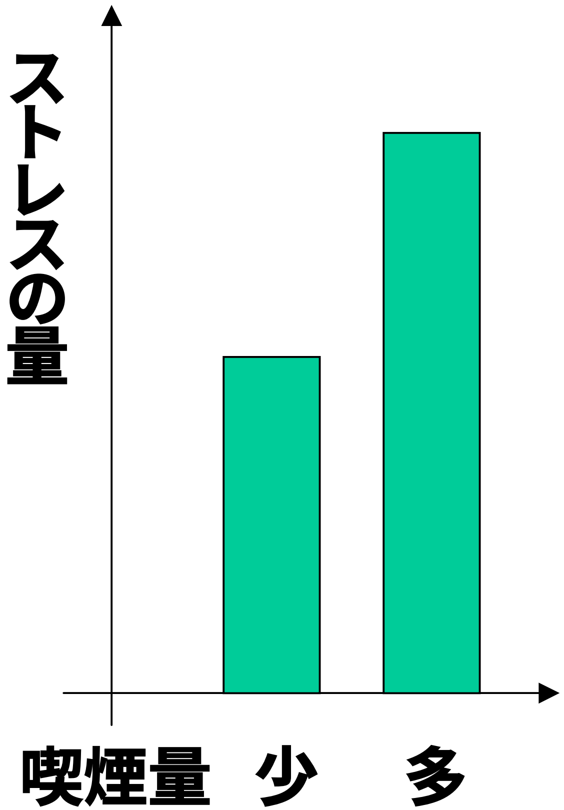
1) ランダムに被験者を選び，
   喫煙量により群へ振分け
2) 喫煙量の多いｸﾞﾙｰﾌﾟと少
   ないｸﾞﾙｰﾌﾟのそれぞれから
   ランダムに被験者を選ぶ
3) ランダムに被験者を選び，
   喫煙量を割付ける

交絡変数の
影響を受ける

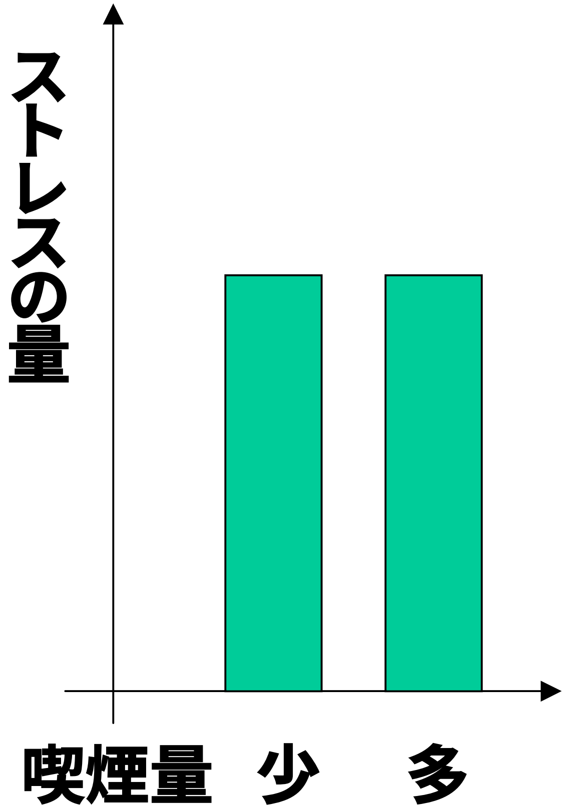
交絡変数の
影響を受けない

ストレスの量





1) と 2) の場合



3) の場合

因子の区別

- 標示因子
  - 水準の指定はできるが選択ができない
  - サンプルに割付できない
  - e.g., 男女, 性格, 知能
  - 交絡変数の影響を受ける
- 制御因子
  - 水準の指定と選択ができる
  - サンプルに割付できる
  - e.g., 学習時間, . . .
  - 無作為割付けした場合, 交絡変数の影響を避けることができる

ではどのようにするか

- 標示因子は無作為割付けができない
  - 無作為割付けによって交絡変数の影響を断つことはできない
  - この意味で調査研究に分類
- 交絡変数の影響をコントロールする
  - 調査研究では重要な交絡変数を必ず観測する
    - データ採取前の検討が重要
  - 交絡変数をモデル化する
    - パス解析
  - 交絡変数を統制する
    - ストレスをそろえる，IQをそろえる
    - 傾向スコアの利用

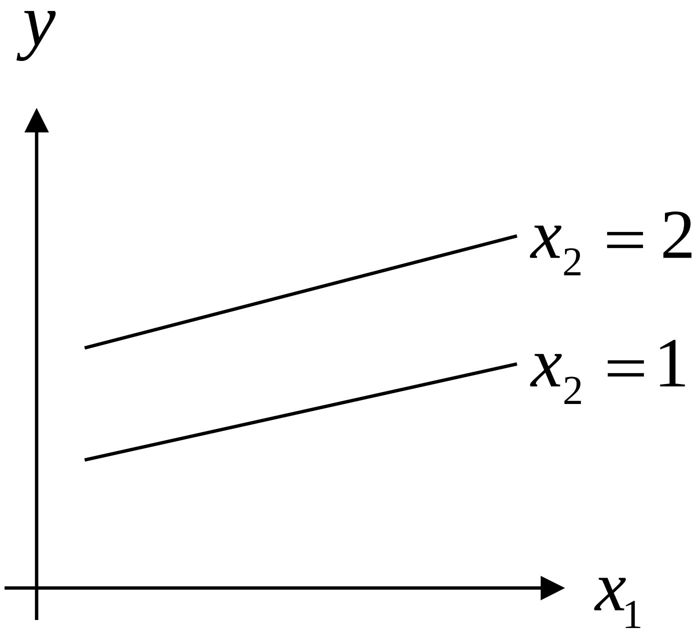
標示因子は原因変数？

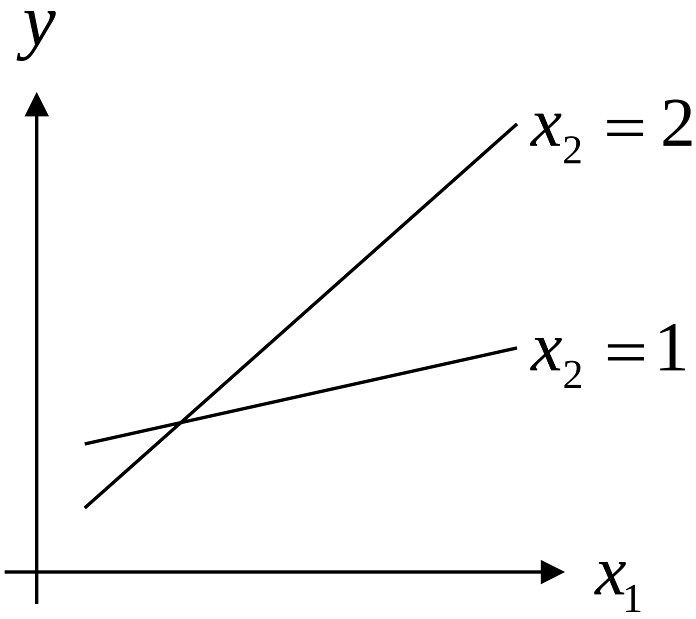
- 標示因子は原因変数と考えることができないという極論もある(Holland,1986)
  - 性別，国籍，年齢 etc は統計的に原因変数として扱うことができないという主張
  - 因子の各水準に曝露可能である必要
- 標示因子であっても予測には使える

交互作用

- ナイーブなパス解析は交互作用は無視
- 交互作用を検出したいとき
  - 連続量を二値化し分散分析
  - 積の項を入れた回帰分析
  - SEM
    - 潜在変数･観測変数の交互作用
    - 現在ソフトウェアが発展中

積の項の役割





$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \\ &= \beta_0 + (\beta_1 + \beta_{12} x_2) x_1 + \beta_2 x_2 \end{aligned}$$

注：中心化しておく　$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12}(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)$

「因果分析」のまとめ

- 共分散構造分析は，調査研究の弱点である交絡変数の影響を受ける
  - 状況によっては、適合度の悪化により交絡変数による影響を疑うことができるが．．．
- モデル構築の段階で，因果仮説を十分に吟味し，重要な影響を与える変数を分析から落とさない
  - データを採る前が大事
- 因果の方向に興味があるときは，操作変数法などを用いて，対立モデルを識別できるようにしたモデル構成を行う
- 調査研究は積み重ねることが重要