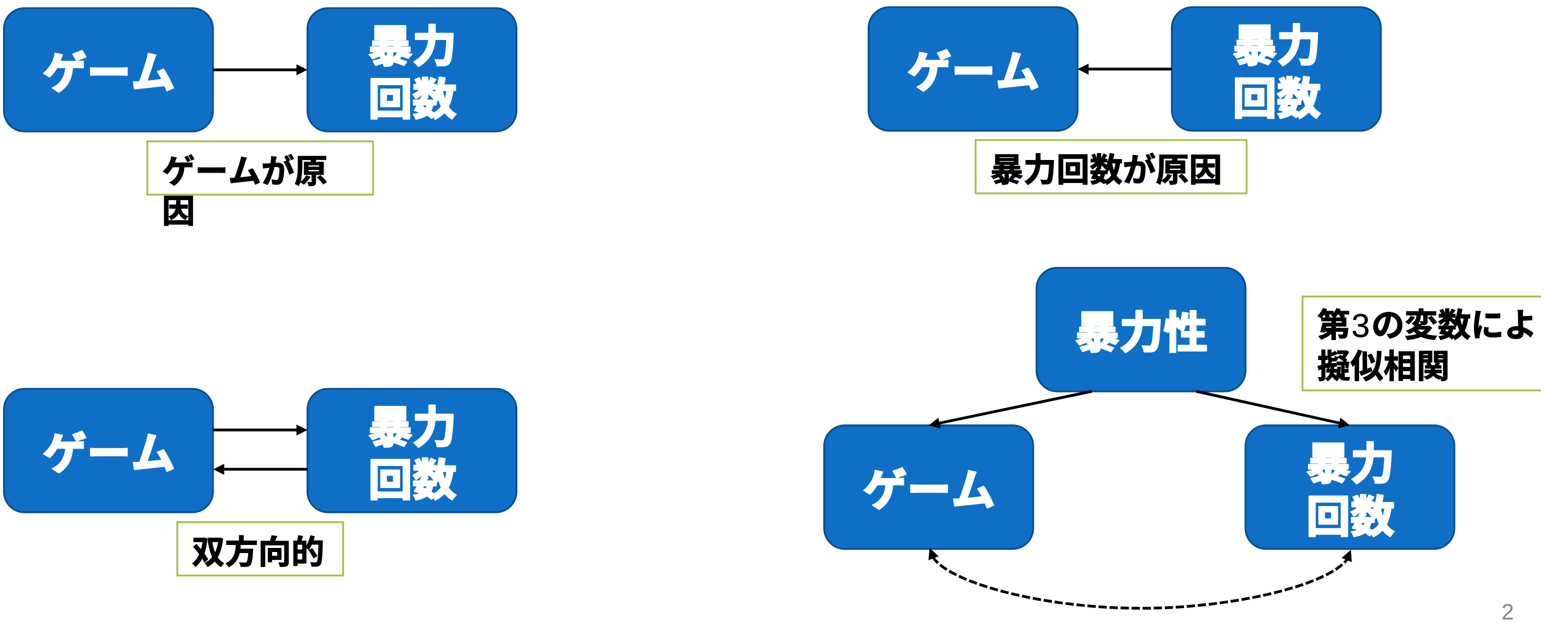
第9章　縦断データ解析による因果関係の探索

教育学研究科　修士1回
石黒翔

相関研究と因果

「小学生の暴力的ゲームで遊ぶ時間」と「周囲に暴力を振るう回数の間に正の相関が見られたとしても、因果関係はわからない。



因果関係を示すための条件

X→Yという因果関係を示すための条件

①XはYよりも時間的に先行していること
②XとYの間に関連があること
③他の因果的説明が排除されていること
（ミルの3原則）

無作為化配置実験と因果関係

①XはYよりも時間的に先行していること
Xを実験的に操作することができるため条件を満たす

②XとYの間に関連があること
検定が有意ならば条件を満たす（と判断する）

③他の因果的説明が排除されていること
無作為化により統制されていると考えられるため条件を満たす

無作為化配置実験が行えない状況

無作為化配置実験は因果関係を検討するための強力な方法である
しかし、無作為化配置実験を行うことができない状況もある

例：暴力的なゲームで遊ぶことの長期的な影響を検討する研究
一ヶ月間、毎日2時間ゲームで遊ぶ条件に強制的に参加者を割り当てることなどはできない（倫理的問題が生じるため）。

無作為化配置実験に代わる方法として縦断調査により、因果関係に迫ることができる。

縦断調査と因果関係

①XはYよりも時間的に先行していること
同一の参加者に複数時点でデータ収集を行うため条件を満たす

②XとYの間に関連があること
検定が有意ならば条件を満たす（と判断する）

③他の因果的説明が排除されていること
他の変数を投入し、当該の関係についての偏回帰係数を検討することで、<u>他の因果的説明をある程度排除することはできる</u>

今回のデータと分析

データ

架空のデータであり、子ども200名から得られたものとする
時点1および時点2の2回データ測定が行われていると考える

今回用いる変数

game1, game2: **時点1または2における暴力的ゲームの遊び時間**
shool1, schoo2: **時点1または2における学校内での暴力回数**
violence: **暴力性**
family1, family2: **時点1または2における家庭内での暴力回数**
manga1, manga2: **時点1または2における暴力内容の漫画の読み時間**

分析方法

構造方程式モデリング（SEM）

構造方程式モデリングとは？
Structural Equation Modeling, SEM

構造方程式モデリングとは？

構成概念や観測変数の性質を調べるために集めた多くの観測変数を同時に分析するための統計的手法である　（豊田、1998）

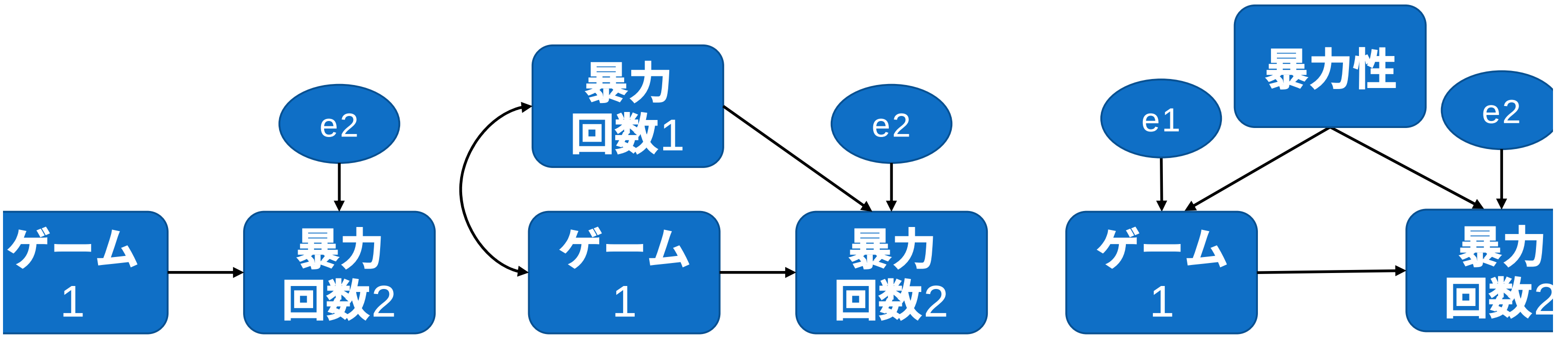
※共分散構造分析という呼び方もある。

モデルの特徴

柔軟なモデル構成を行うことができる

回帰分析や因子分析などの分析を下位モデルとして含み、これらを統合するようなモデルも構成可能である

SEMと他の分析の関係（パス図による表現



※単回帰分析　※重回帰分析　※パス解析

全てSEMの枠組みの下で分析可能

Rによる構造方程式モデリング

*lavaan*パッケージを使用

lavaanパッケージを用いたSEMの具体的な流れ

1.**ローデータあるいは、共分散行列を用意**

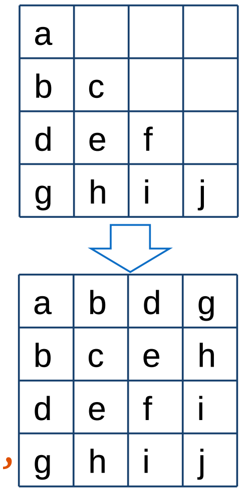
2.**研究仮説のモデルをパス図で表現**

3.**パス図より、**lavaan**の書式にしたがって、モデル式を記述**

4.lavaan**の関数により推定値の算出**

lavaanの読込みと共分散行列の読込み

```
install.packages("lavaan") #lavaanパッケージのインストール
library(lavaan) #lavaanパッケージの読み込み
lower <- '
2.45,
1.88, 2.45,
0.86, 0.75, 1.96,
0.87, 0.82, 1.51, 2.05,
1.51, 1.41, 1.26, 1.32, 2.49,
1.65, 1.51, 1.32, 1.46, 1.92, 2.59,
0.72, 0.73, 1.30, 1.29, 1.17, 1.32, 2.10,
0.79, 0.58, 1.13, 1.30, 1.23, 1.24, 1.55, 2.07
1.08, 1.12, 1.32, 1.09, 1.12, 1.06, 1.24, 1.18, 3.61’
```

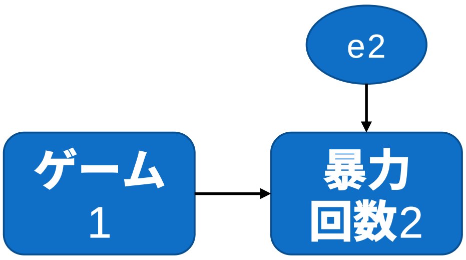



```
#共分散行列の作成
game.cov <- getCov(lower, names=c("game1","manga1","school1","family1", "game
"manga2", "school2","family2","violence"))
```


lavaanパッケージにおける変数間の関係の表現

記号	意味	例	例の意味
~	**構造方程式**	y ~ x1	x1→y **の回帰分析**
=~	**測定方程式**	f =~ x1 + x2 + x3	**因子**f**は観測変数**x1,x2,x3**に** **より測定される**
~~	**共分散**	x1 ~~ x2	x1**と**x2**の共分散**

単回帰モデル



変数

game1：**時点1における暴力的ゲームの遊び時**

shool2：**時点2における学校内での暴力回数**

```
#単回帰モデル
model1 <- '
school2 ~ game1
school2 ~~ school2
'
```

school2 ~ game1　**は**game1→schoo1**を表している**

school2 ~~ school2 **は**school2**の誤差分散を表している**(**図中**e2**に相当**)

※lavaan**では単方向矢印を受けている変数**y**について**’y~~y’**を用いると**y**の誤差分散として扱**

単回帰モデルのデータへの当てはめ

```
#単回帰モデル(model1)を当てはめる
fit1 <- lavaan(model1, sample.cov=game.cov, sample.nobs=200)
summary(fit1, standardized=T, rsquare=T)
#standardizedは標準化推定値の出力の有無
#rsquareは決定係数の出力の有無

#出力
lavaan (0.5-18) converged normally after 11 iterations

Number of observations 200

Estimator ML
~略~
```

単回帰分析の結果

Number of observations:	**サンプルサイズ**
Estimator:	**推定方法（**ML＝**最尤法）**
Minimum Function Test Statistic:	**モデルのカイ二乗値**
Degrees of freedom:	**自由度**
P-value:	**カイ二乗検定の***p***値**

※単回帰モデルでは上記3**つの値が全て**0**になるため参照する意味は無い**

Regressions:	**回帰分析**
Estimate:	**推定値**
Std.err:	**標準誤差**
Z-value:	**検定統計量***z*
	帰無仮説「母集団において回帰係数=0

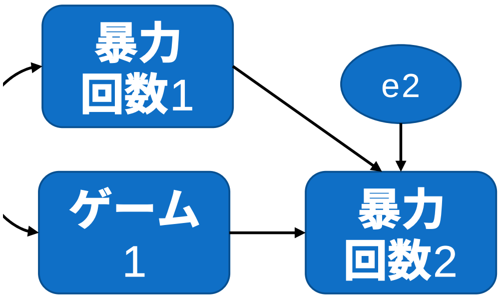
先の分析で重要な点

school2 ~ game1**の回帰係数の推定値は**0.294**で有意**

この結果からは、「ある一ヶ月間に暴力的ゲームで1**時間長く遊ぶと、次の一ヶ月間に学校で平均的に**0.294**回暴力を多く働くようになる」と解釈される。**

他の可能性も探るために重回帰分析およびパス解析を行う

重回帰モデル



変数

game1：**時点1における暴力的ゲームの遊び時間**
shool1：**時点1における学校内での暴力回数**
shool2：**時点2における学校内での暴力回数**

重回帰モデル

```
del2 <- '
hool2 ~ game1 + school1
hool2 ~~ school2
```

`ıool2 ~ game1 + schoo1` **は独立変数が**game1**と**school1**、従属変数が**shool2**であると示す**
ゲーム1**と暴力回数**1**の間の共分散は**lavaan**が自動的に指定するため、明示する必要はない**

単回帰モデルのデータへの当てはめ

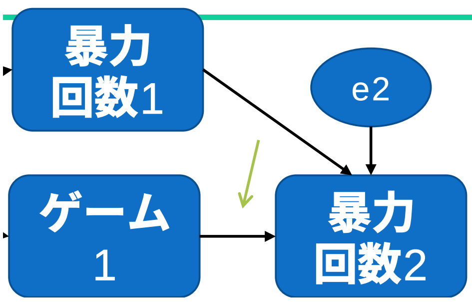
#重回帰モデル(model2)を当てはめる

```
fit2 <- lavaan(model2, sample.cov=game.cov, sample.nobs=200)
summary(fit2, standardized=T, rsquare=T)
```

#出力 回帰係数部

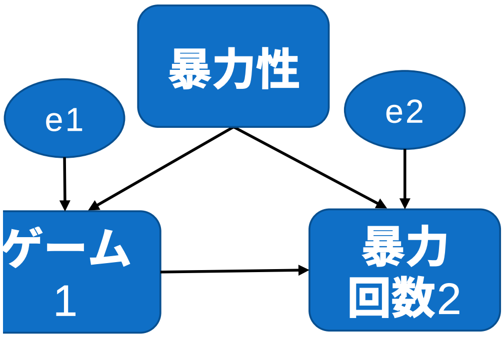
```
      Estimate   Std.err    Z-value    P(>|z|)    Std.lv
      Std.all
Regressions:
school2 ~
game1   0.072     0.054      1.327      0.185      0.072      0.07
school1 0.632     0.061      10.384     0.000      0.632      0.61
```

有意ではない



暴力的ゲームで遊ぶこと自体が
暴力行為を増加させているわけではないこ
が示唆される

パス解析モデル



変数

game1：**時点1における暴力的ゲームの遊び時間**
shool1：**時点1における学校内での暴力回数**
violence：**暴力性**

パス解析モデル

```
del3 <- '
hool2 ~ game1 + violence
me1 ~ violence
hool2 ~~ school2; game1 ~~ game1
```

；’により一行に複数の式を記述することもできる

パス解析モデルのデータへの当てはめ

#パス解析モデル(model3)**を当てはめる**

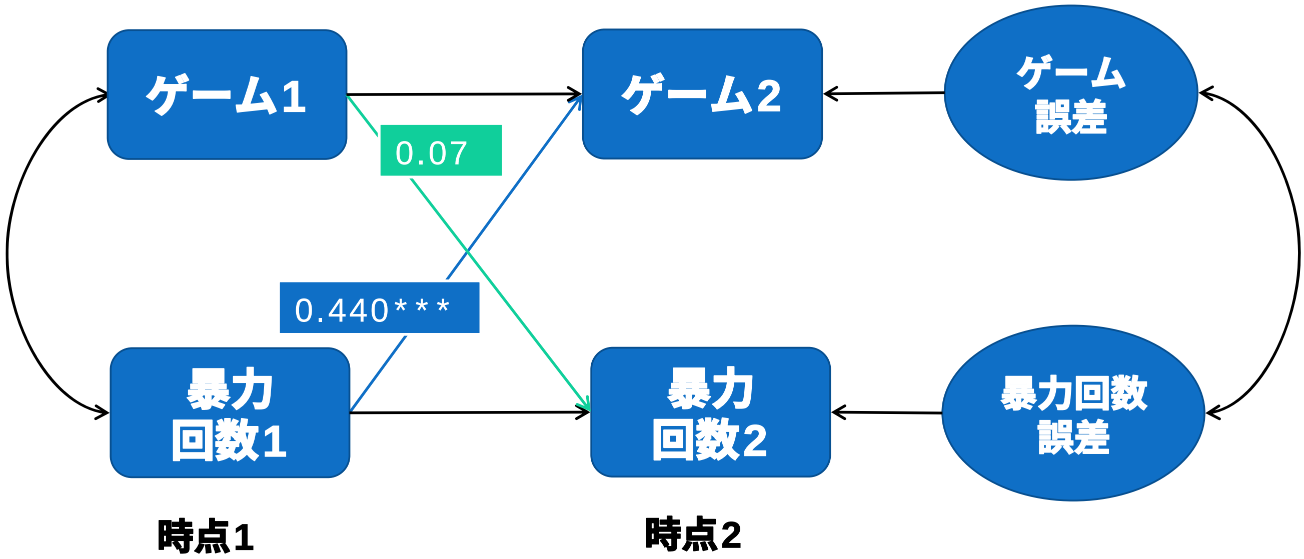
```
fit3 <- lavaan(model3, sample.cov=game.cov, sample.nobs=200)
summary(fit2, standardized=T, rsquare=T)
```

#出力 回帰係数部

```
     Estimate   Std.err    Z-value    P(>|z|)    Std.lv
     Std.all
Regressions:
school2 ~
game1    0.164   0.062      2.662      0.008      0.164      0.17
violence 0.294   0.051      5.796      0.000      0.294      0.38
game1 ~
violence 0.299   0.054      5.512      0.000      0.299      0.36
```

shool2 ~ game1**が有意であり、暴力性が第**3**の変数として機能せず、**
暴力性がゲームと暴力行動の関係を説明できるわけではないことがわかる。

交差遅延モデル(cross-lagged model)



暴力回数1→ゲーム2のパス係数と
ゲーム1→暴力回数2のパス係数を同時に見ることができる
2時点での測定を行っているため、ある程度因果関係の推測が行える

交差遅延モデルのデータへの当てはめ

```
#交差遅延モデルその1
model4 <- '
school2 ~ game1 + school1
game2 ~ game1 + school1
school2 ~~ game2; school2 ~~ school2; game2 ~~ game2
'
fit4 <- lavaan(model4, sample.cov=game.cov, sample.nobs=200)
summary(fit4, standardized=T, rsquare=T, fit.measure=T)
```

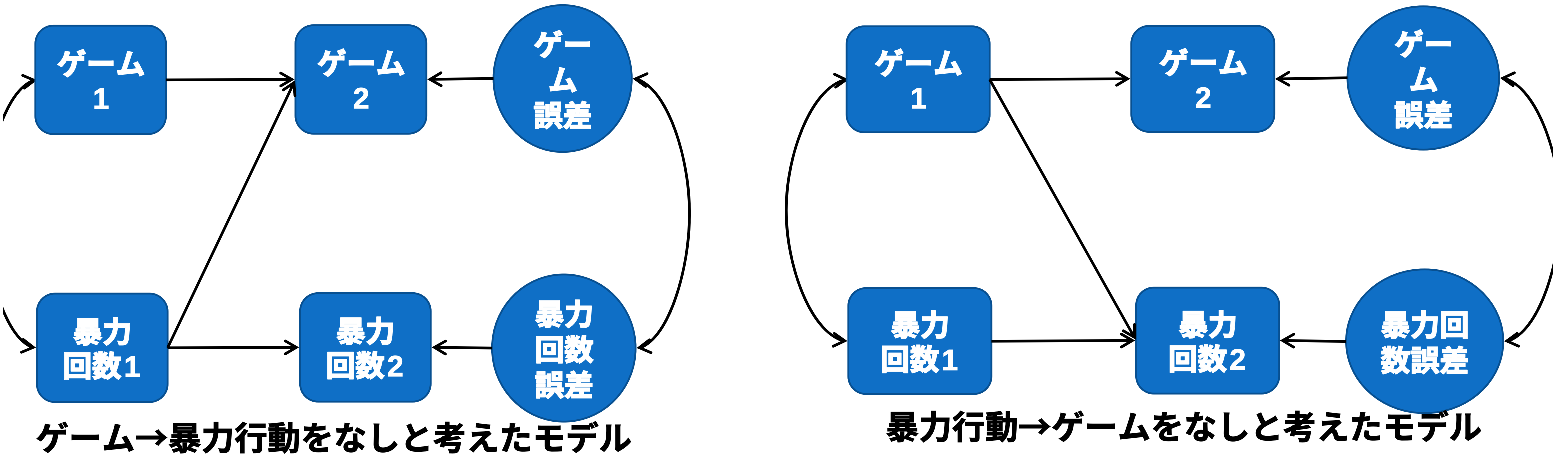
出力から、
game1→school2**のパス係数は**0.072。p**値は**0.184**で有意ではない**
schoo1→game2**のパス係数は**0.440。p**値は**0.000**で有意である。**

暴力回数→ゲームの因果関係の方がもっともらしいと判断できる。
（厳密には直接の比較についての検定が必要か？）

参考　モデル間の比較と適合度

同一データに対して複数のモデルを考えることもできる。

例　先までのデータに対して



モデルの適合度(AIC**や**CFI**など)から、**
どのモデルがデータに対し当てはまりが良いか検討する必要があ

潜在変数

潜在変数：
複数の観測変数に影響を与えていると仮定される潜在的な変数
（豊田、1998**）**

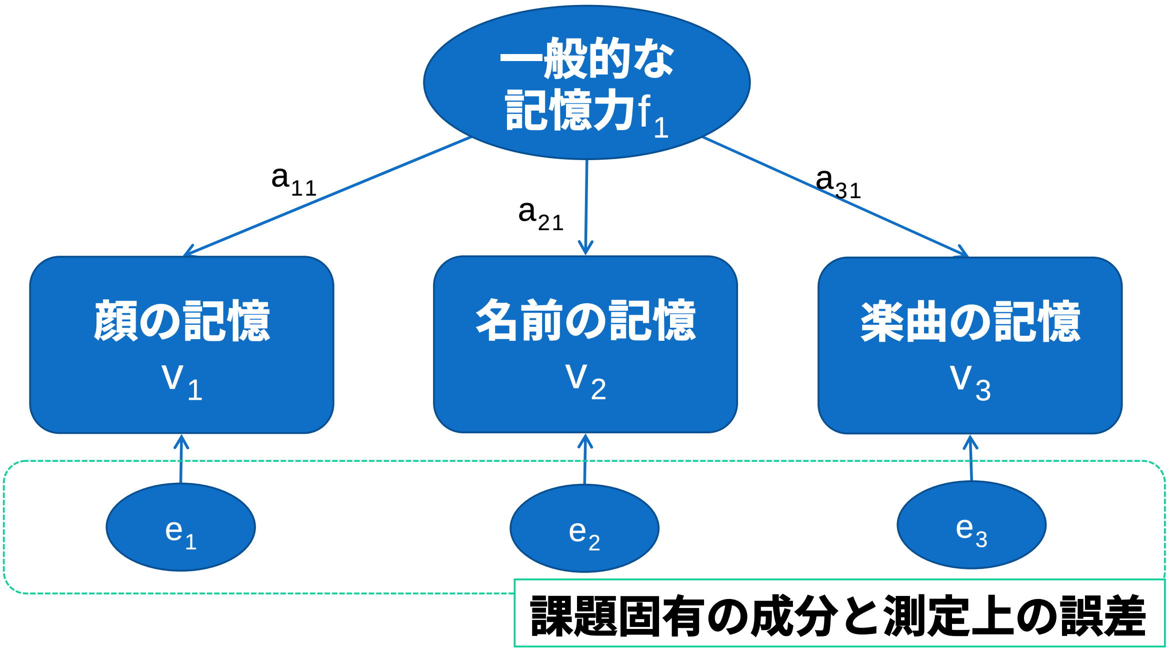
心理学においては、構成概念が潜在変数として表現される。

e.g. **実際に課題の成績を検討するのではなく、課題の成績に影響を与えている変数を心理学は対象としている。**

潜在変数と構成概念

概念

潜在変
（因子

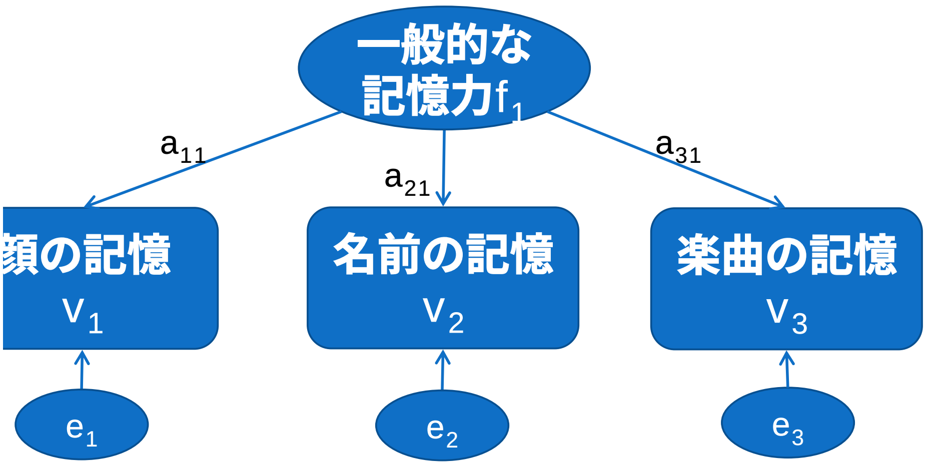


観測変

エラー

題固有の成分と測定上の誤差を取り除いた成分（e.g. f_1）を
ことができる

考　潜在変数と観測変数の関係



※vは偏差ベクトル、$f_1 \sim N(0$
v_1, v_2, v_3はf_1とそれぞれのエ
で説明されると考える

このため

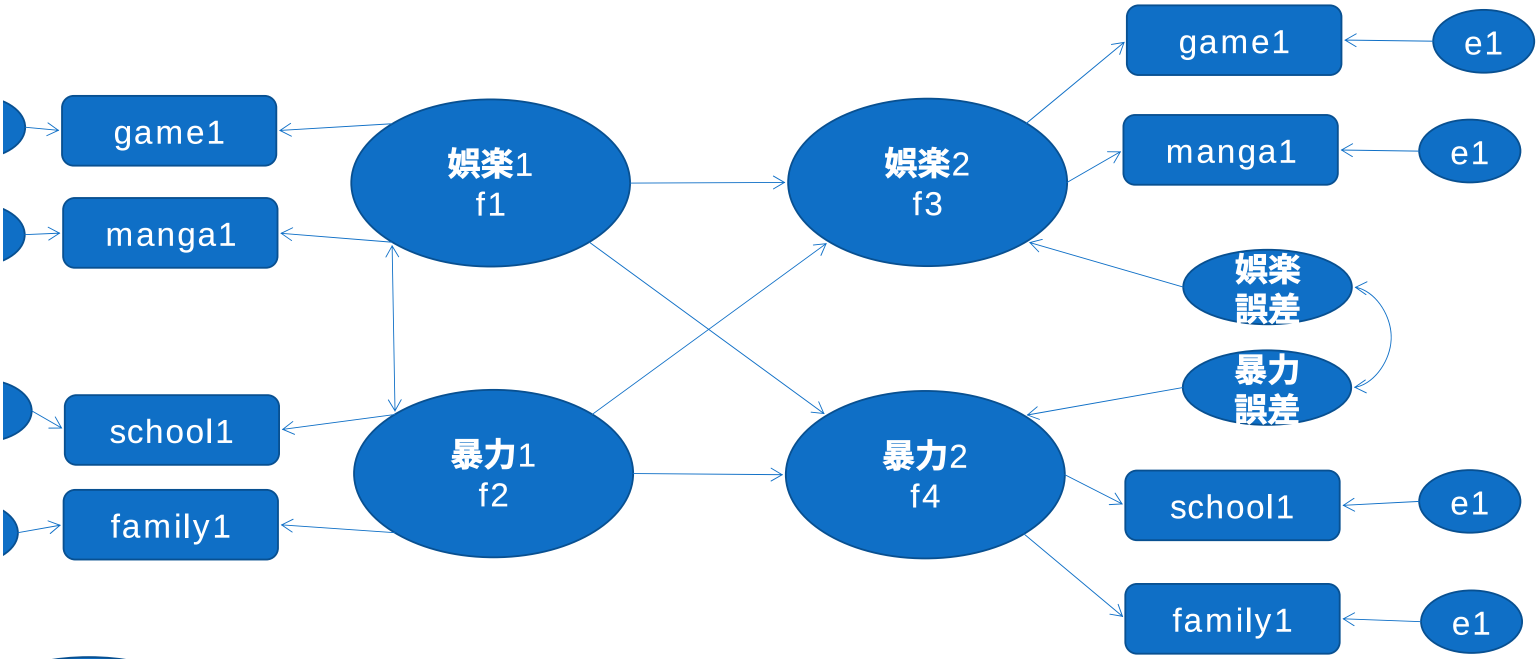
$$v_1 = a_{11}f_1 + e_1$$
$$v_2 = a_{21}f_1 + e_2$$
$$v_3 = a_{31}f_1 + e_3$$

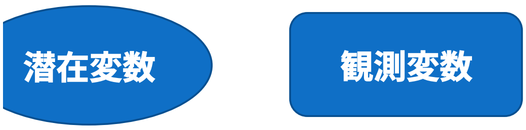
の3本の方程式が立つ。

お、これらの方程式の母数を用いて、観測変数の分散と共分散が表現できる。
e.g. v_1の分散は $a^2_{11} + \sigma^2_{e1}$ となる。v_1とv_2の共分散は$a_{21}a_{11}$となる。

らに、実際に得られた標本分散・標本共分散を利用することで
$_{11}, a_{21}, a_{31}, e_1, e_2, e_3$を推定することができる。

潜在変数を導入した交差遅延モデル（因子分析的手法の導入）





潜在変数は楕円で、観測変数は長方形で表されてい

潜在変数を導入した交差遅延モデル

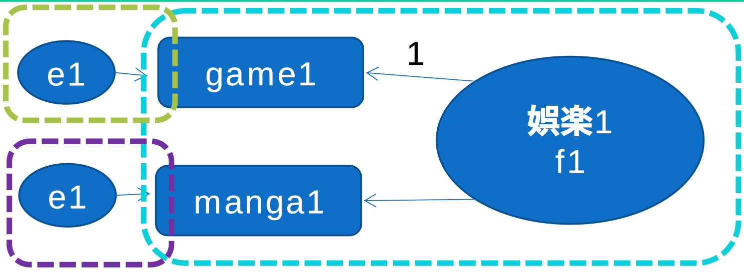
潜在変数間の交差遅延モデルその1

```
del7 <- '
 =~ 1*game1 + manga1
 =~ 1*school1 + family1
 =~ 1*game2 + manga2
 =~ 1*school2 + family2
me1 ~~ game1; manga1 ~~ manga1; school1 ~~ school1; family1 ~~
mily1
me2 ~~ game2; manga2 ~~ manga2; school2 ~~ school2; family2 ~~
mily2
 ~ f1 + f2
 ~ f1 + f2
 ~~ f2 ; f3 ~~ f4; f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4

t7 <- lavaan(model7, sample.cov=game.cov, sample.nobs=200)
mmary(fit7, standardized=T, rsquare=T, fit.measure=T)
```

因子（潜在変数）の測定

```
 =~ 1*game1 + manga1
me1 ~~ game1
nga1 ~~ manga1
```



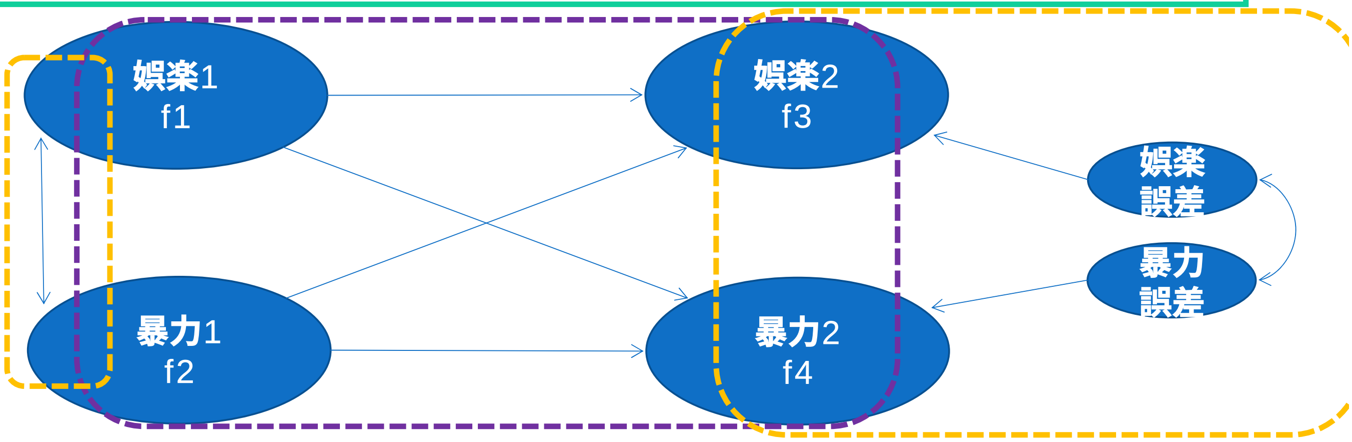
=~ 1*game1 + manga1 は観測変数game1とmanga1の背後に因子f1
Eすることを意味している。

game1 はgame1に対する因子負荷量を1に固定することを意味する。

ne1 ~~ game1 は誤差e1の分散を表す（manga1についても同様）。

因子間のパス解析

```
 ~ f1 + f2
 ~ f1 + f2
 ~~ f2 ; f3 ~~ f4; f1 ~~ f1; f2 ~~ f2; f3 ~~ f3; f4 ~~ f4
```



3 ~ f1 + f2とf4 ~ f1 + f2 は因子間のパス解析
1 ~~ f2 ; f3 ~~ f4 はそれぞれの因子間の共分散
1 ~~ f1; f2 ~~ f2 はそれぞれの因子の分散
3 ~~ f3; f4 ~~ f4 はそれぞれの因子に関する誤差の分

モデルの適合度

```
summary(fit7, standardized=T, rsquare=T, fit.measure=T)
#fit.measureオプションにより、モデルの適合度を表示できる。
```

モデル	X二乗値	自由度	*p*値	AIC	CFI	RMSEA	SRMR
交差遅延その1	23.072	14	0.059	4791.353	0.992	0.057	0.014

絶対的指標　CFI, RMSEA, SRMR

CFI**は**0.95**以上、**RMSEA**と**SRMR**は**0.05**以下のときによい当てはまりであると言われている。**

相対的指標　AIC　**（モデルの比較に用いる）**

AIC**は他のモデルと比較したときに、小さい値を示すモデルを当てはまりのよいモデルとみなす。**

参考文献

豊田秀樹(1998). **共分散構造分析 入門編―構造方程式モデリング 朝倉書店**

豊田秀樹(2014). **共分散構造分析**[**R編**]　--**構造方程式モデリング 東京図書出版**

山田剛史(編著) (2015). R**による心理学研究法入門　北大路書房**