心理データ解析演習
～ 回帰分析・交互作用項

2009/06/10

経済学研究科

D2　中本龍市

発表内容

- (社会科学における)交互作用項の流行
- 最も基本的な注意
- 関数の形の特定
- 交互作用項の係数の解釈

交互作用項の流行

- 研究上の要請
;非線形の関数は、理論的にも、実証的にも面白い
→　複雑な仮説の検証が可能　例)「ある条件Aの下での、変数Aと変数Bの間にある効果」

- 理論的な問題点
;関数の形を正当化できる理論的な説得力とデータセットの統計的特徴を結びつける回帰分析が必要
→　後付けの結果が多いのは、仕方がないことだが･･･

想定されうる因果関係

- 因果関係の種類
  ;(1)直接的な因果関係
  ;(2)擬似的な関係
  ;(3)分析されていない関係(分析できない関係)
  ;(4)間接的な因果関係
  ;(5)相互的な因果関係
  ;(6)モデレートされた因果関係

交互作用項で取り扱う因果関係

•例) 回帰分析

•例) 構造方程式モデル

•例) 回帰分析、偏相関

•例) 構造方程式モデル

•例) 事例分析？

•例) 交互作用項

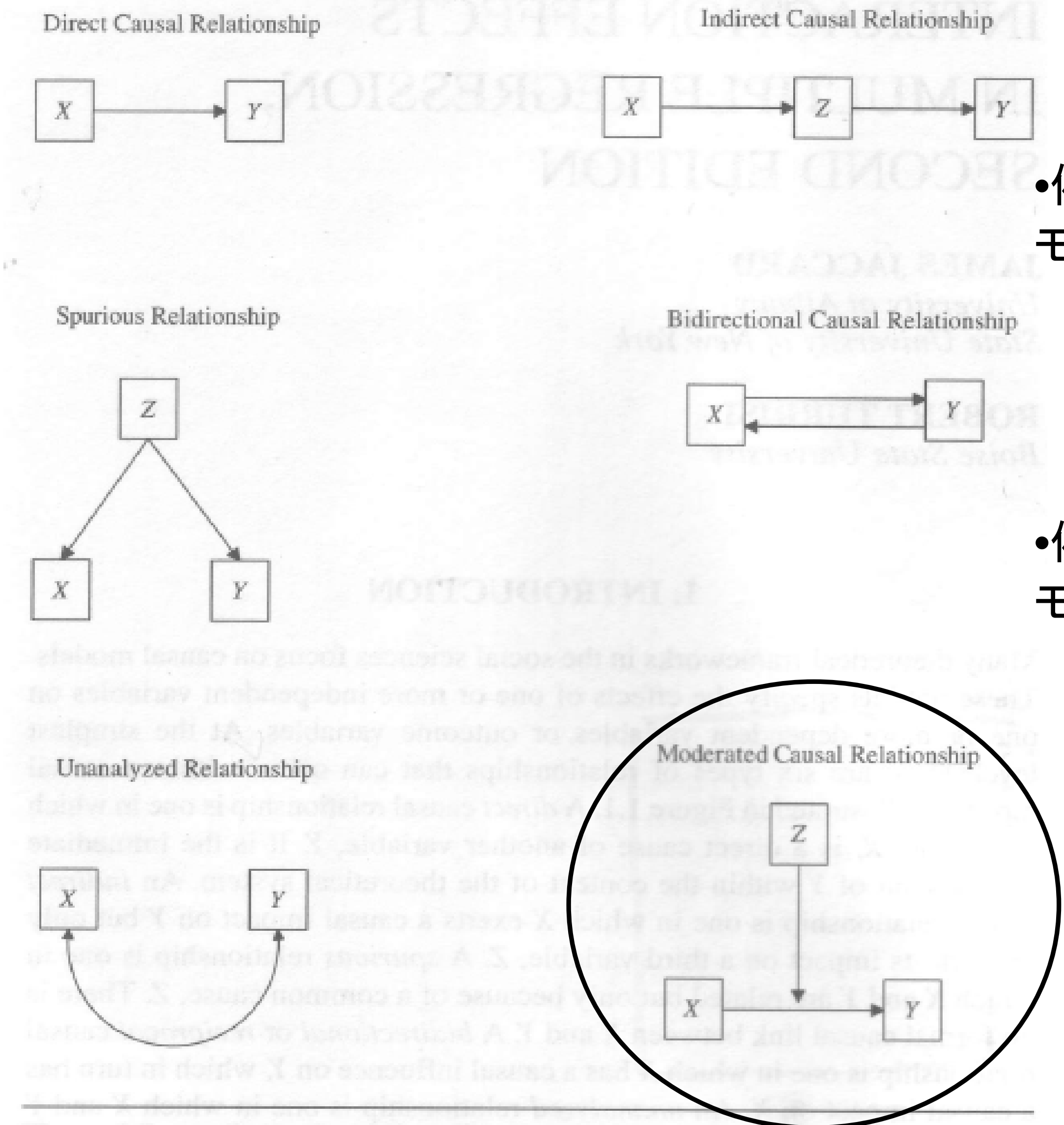


Figure 1.1 Examples of Causal Relationships

交互作用項

- Interaction term
  ;交互作用項、交差項など分野によって定訳が異なる

- 変数Xと変数Yの間の関係に、変数Zがどういう影響を与えるのか、影響があるのかを確かめる

交互作用項への期待

- 交互作用項に関する注意
  ;「エディターとしての実務経験上、非線形関数を扱う際には注意をすべきである。というのは、多くの著者が気付いていないことがあるからである」(Meyer2009)
- 具体的に
  ;説明変数の範囲では、関係は常に安定しており、線形の回帰分析で検証できる場合もある
  ;だが、非線形の場合もあるし、範囲に制限をかけているから線形であることもある

最も基本的な注意点(1)

- 「非線形の研究を受け取ると、普段、私はまず最初に、封筒の裏に簡単な計算をする。驚くべきことに、そうすると、発見事実の妥当性が否定される。このテストは次のようなものである。微分して、最小値か最大値をとる値を計算する。そして、この値を、説明変数の定義域と比較する。驚くことに、この値は、平均値から±2SD以上離れていたり、場合によっては、説明変数の範囲から外れている。」(Meyer2009)
- 非線形の解釈
 ;プロットしてみる、か、変曲点を計算してみる。

最も基本的な注意点(2)

- プロットしてみて・・・
  ;もし、最大値が定義域の外側になってしまった場合、
  ;逆U字型ではないが、「正で、増加率は低減している」ということがわかる。となると、ログ関数が適切であろう。
- 変曲点が、定義域の範囲でも、平均値から遠く離れている場合、
  ;一つかいくつかの外れ値が結果に影響していることがある。
  ;この場合、外れ値を外して回帰分析をやり直すか、プロットしてみる
- 曲線が見た目では、高い水準で相対的に平らである場合、低いレベルでは正の効果があるが、高いレベルでは、影響が少ないことをしめしている。この場合には、ログ関数が良い。

曲線関係と定義域

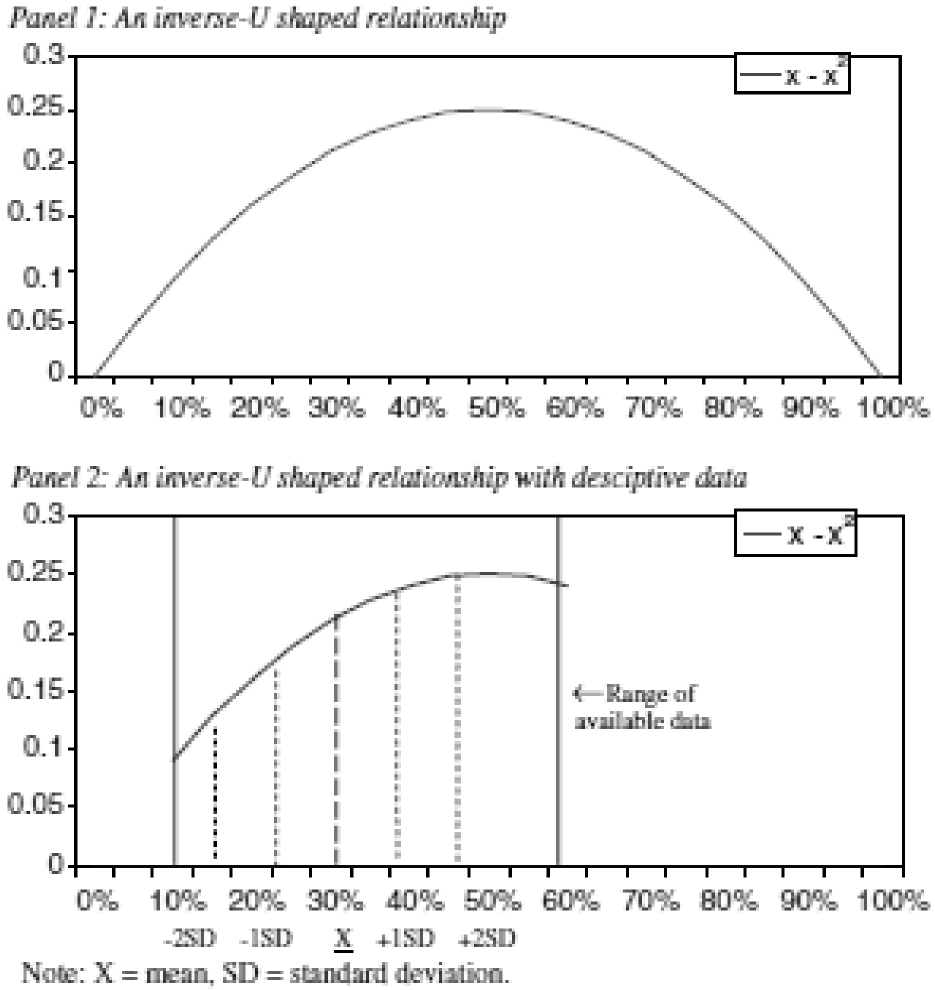


Figure 1 Interpreting a quadratic relationship.

- 記述統計は、なぜ掲載しなければならないのか？
  → 図のように、定義域と曲線の関係を検証できる。
- 下の図の場合、曲線関係があると、言えるのか？

Meyer(2009)

理論からの関数特定

- 表.1
  ;表1は、Aが正の効果、Bが負の効果をもって、相互作用する場合を示す

- 関数の特定
  ;非線形の関係は、一つの可能性に過ぎない。一つに特定できない。相互作用の本質が分からない場合には、理論によって、関数の形を決定することはできない。理論を考えると、2変数の関係を示唆する仮説に行き着くが、関数の正確な形まではわからないことが多い
  ;既存研究から推測することはある(例　Takeuchi, et al 2007)

- 交互作用に意味があるなら、そこまで理論からの導出にこだわる必要はない(Meyerの立場)

関数の形(1)

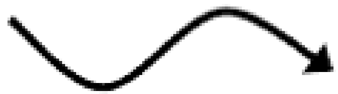
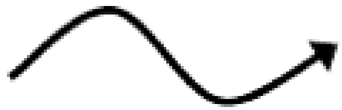
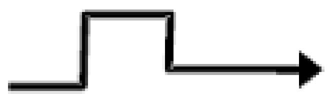
Table 1 Alternative functional specifications.

Case	Primary effects (A = positive, B = negative)	Aggregate effect	Functional form
1	A and B cancel each other out	No aggregate effect	No significant effect
2	A dominates over B		Linear: $\beta_1 x$, with $\beta_1 > 0$
3	B dominates over A		Linear: $\beta_1 x$, with $\beta_1 < 0$
4	A dominates at low level, B dominates at high levels		Quadratic (concave or inverse-U): $\beta_1 x + \beta_2 x^2$, with $\beta_1 > 0$, and $\beta_2 < 0$
5	B dominates at low level, A dominates at high levels		Quadratic (convex or U-shaped): $\beta_1 x + \beta_2 x^2$, with $\beta_1 < 0$, and $\beta_2 > 0$

•ケース4と5では、U字型。しかし、なぜ、ある値を超えたら、AがBをしのぐのか、を説得的に説明できるか？

続き

関数の形(2)

Case	Primary effects (A = positive, B = negative)	Aggregate effect	Functional form
6	A dominates at high and low levels, B dominates at intermediate levels		Cubic (S-shape): $\beta_1 x+ \beta_2 x^2+ \beta_2 x^2$, with $\beta_1>0$, $\beta_2<0$ and $\beta_3>0$
7	B dominates at high and low levels, A dominates at intermediate levels		Cubic (inverse-S): $\beta_1 x+ \beta_2 x^2+ \beta_2 x^2$, with $\beta_1>0$, $\beta_2<0$ and $\beta_3>0$
8	A dominates, but weakens for higher values		Logarithmic: $\beta_1 \log(x)$, with $\beta_1>0$
9	B dominates, but weaken for higher values		Logarithmic: $\beta_1 \log(x)$, with $\beta_1<0$
10	Qualitative shifts at certain benchmark values		Step-function

•質的に別のステージになる場合(ケース10)；例えば、株の所有率49.9%と51.1%の質的な違い。ステップ関数。

回帰分析の基本モデル

- $Y=\alpha+\beta_1X+\beta_2Z+\varepsilon$

- 係数の解釈

交互作用項を含むモデル

- $Y=\alpha+\beta_1X+\beta_2Z+\varepsilon$
- $\beta_1=\alpha'+\beta_3Z$
- $Y=\alpha+(\alpha'+\beta_3Z)X+\beta_2Z+\varepsilon$
- $Y=\alpha+\alpha'X+\beta_3XZ+\beta_2Z+\varepsilon$
- $Y=\alpha+\beta_1X+\beta_2Z+\beta_3XZ+\varepsilon$

数値例

TABLE 2.2
Cell Means as a Function of *X* and *Z*

	Peer Pressure (Z)				
Attitude (X)	*1*	*2*	*3*	*4*	*5*
5	25	21	17	13	9
4	20	17	14	11	8
3	15	13	11	9	7
2	10	9	8	7	6
1	5	5	5	5	5

b_1 at a score of 1 on $Z = 5.0$
b_1 at a score of 2 on $Z = 4.0$
b_1 at a score of 3 on $Z = 3.0$
b_1 at a score of 4 on $Z = 2.0$
b_1 at a score of 5 on $Z = 1.0$

交互作用項と係数の問題

- 積の項が入ったモデルの場合、XとZの係数は、それぞれ、片方が、0の場合の値
→主効果とは異なる

高次の関数

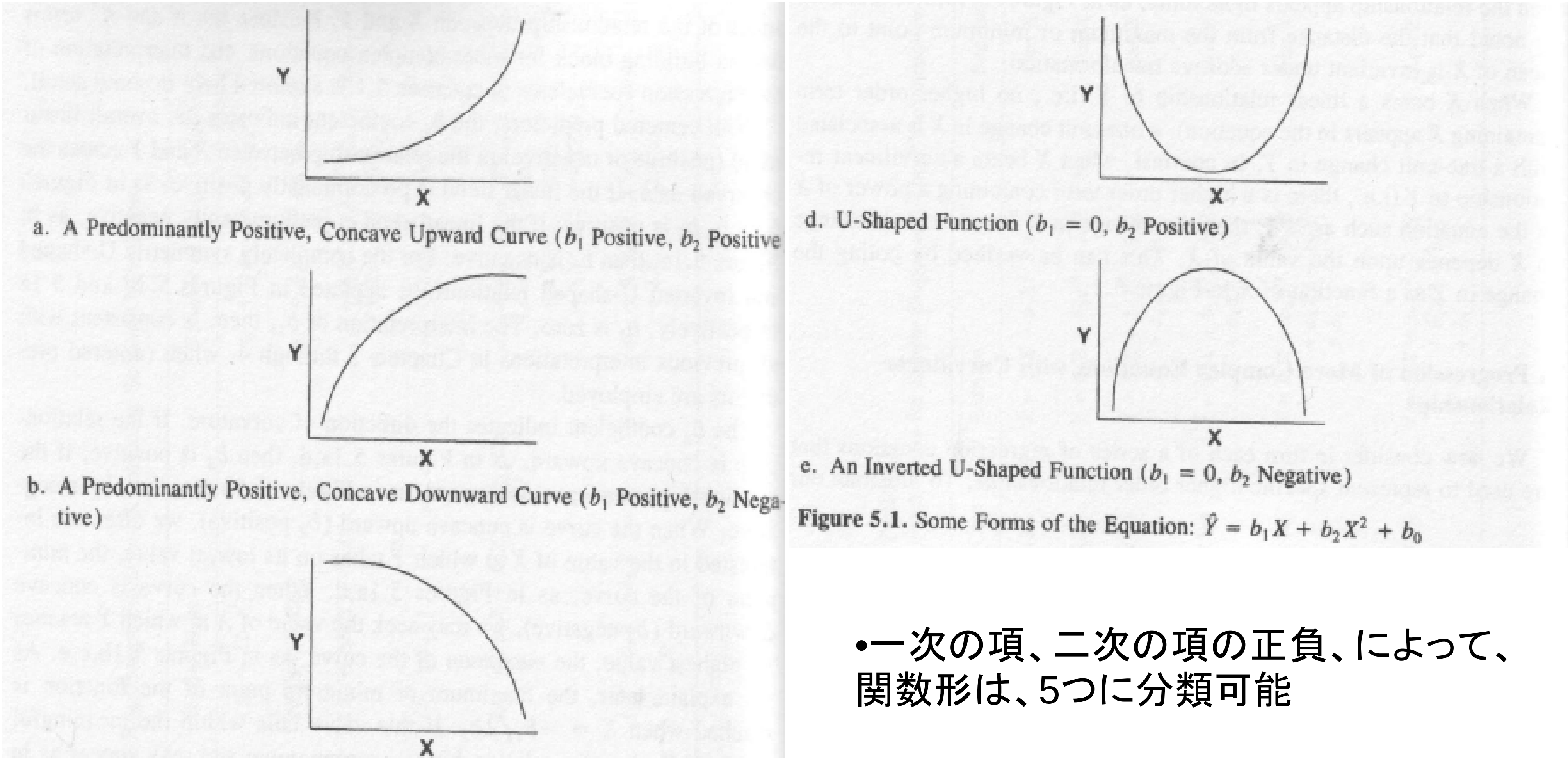
- 係数の正負の基本的な問題
  ;微分した結果によって関数形は異なる

高次の関数

Table 5.1
Expressions for Simple Slopes and Their Variances for Various Regression Equations Containing Second-Order Terms

Case	Equation	Regression	Simple Slope	Variance of Simple Slope (s_b^2)
(1)	$\hat{Y} = b_1X + b_2X^2 + b_0$	Y on X	$(b_1 + 2b_2X)$	$s_{11} + 4Xs_{12} + 4X^2s_{22}$
(2)	$\hat{Y} = b_1X + b_2X^2 + b_3Z + b_0$	Y on X	$(b_1 + 2b_2X)$	$s_{11} + 4Xs_{12} + 4X^2s_{22}$
(3a)	$\hat{Y} = b_1X + b_2X^2 + b_3Z + b_4XZ + b_0$	Y on X	$(b_1 + 2b_2X + b_4Z)$	$s_{11} + 4X^2s_{22} + Z^2s_{44} + 4Xs_{12} + 2Zs_{14} + 4XZs_{24}$
(3b)	$\hat{Y} = b_1X + b_2X^2 + b_3Z + b_4XZ + b_0$	Y on Z	$(b_3 + b_4X)$	$s_{33} + 2Xs_{34} + X^2s_{44}$
(4a)	$\hat{Y} = b_1X + b_2X^2 + b_3Z + b_4XZ + b_5X^2Z + b_0$	Y on X	$(b_1 + 2b_2X + b_4Z + 2b_5XZ)$	$s_{11} + 4X^2s_{22} + Z^2s_{44} + 4X^2Z^2s_{55} + 4Xs_{12} + 2Zs_{14} + 4XZs_{24} + 4XZs_{15} + 8X^2Zs_{25} + 4XZ^2s_{45}$
(4b)	$\hat{Y} = b_1X + b_2X^2 + b_3Z + b_4XZ + b_5X^2Z + b_0$	Y on Z	$(b_3 + b_4X + b_5X^2)$	$s_{33} + X^2s_{44} + X^4s_{55} + 2Xs_{34} + 2X^2s_{35} + 2X^3s_{45}$

高次の関数



a. A Predominantly Positive, Concave Upward Curve (b_1 Positive, b_2 Positive)

b. A Predominantly Positive, Concave Downward Curve (b_1 Positive, b_2 Negative)

d. U-Shaped Function ($b_1 = 0$, b_2 Positive)

e. An Inverted U-Shaped Function ($b_1 = 0$, b_2 Negative)

Figure 5.1. Some Forms of the Equation: $\hat{Y} = b_1 X + b_2 X^2 + b_0$

- 一次の項、二次の項の正負、によって、関数形は、5つに分類可能

交互作用の有無

$$F = \frac{(R_2^2 - R_1^2)/(k_2 - k_1)}{(1 - R_2^2)/(N - k_2 - 1)}$$

参考文献

- Meyer K.E.(2009) “Motivating, testing, and publishing curvilinear effects in management research”, Asia Pacific Journal of Management, 26, pp187-193
- 佐和隆光(1979)『回帰分析』朝倉書店