

## 第2講 因果関係分析

### 2.1 はじめに

経済分析において因果関係を明らかにすることはきわめて重要な研究テーマである。とりわけ、政策の効果を分析する際には因果関係の分析が中心になるといっても過言ではない。ある政策が目標を達成できたかどうかは、政策手段から政策目標への確実な因果関係が成り立つことが確認されることによってはじめて明らかになる。本講では、因果関係分析の考え方と手法について解説する。

### 2.2 因果関係分析の考え方

医薬品の効果を分析する場合を考えてみよう。ある薬がある病気に効果があるかどうかはその薬を処方することによって、病気が治るということで判断される。もちろん、時間の経過とともに自然治癒することもあるし、他人の効果を比較すると各個人の治癒力の違いもあるので薬の純粋な効果を測るためには、同じ患者に対して薬を投与した場合としなかった場合について調べることが理想的である。しかし、同一個人に対して、同時期に薬を投与することとしないことを実現することは不可能である。そこで患者に薬を投与したとすると、あたかもその患者と同じような病状にありかつ薬を投与していない患者を見つけて(対照サンプル)、事後的な健康状態の違い(差)をみることでその薬の効果を測るという手法を用いる。患者があたかも薬が投与されなかった場合を考えるというのは counterfactuals (反事実) を考えることになる。薬の効果が適切に測定されるためには、この counterfactuals の選択が重要になってくる。

医療の場合もそうだが、経済や政治に関わる政策の効果を測定することは、論理学や物理現象における因果関係とは違って不確実性が入ってくる(確率論的因果関係)。例えば、タバコを吸うと肺ガンになりやすいという因果関係は議論できても、すべての喫煙者がガンになるわけではないことも事実である。ガンになるということと喫煙の間にさらにいくつかの因果関係があり、それが掌握されていなければタバコとガンの間の因果関係は確率論的なものにとどまらざるを得ない。

また政策の因果関係を分析するときに、ある政策を導入することが他の政策の削減に結びついて、それが2次的な効果をもたらすような場合には、政策の一般均衡的效果を考えなければならない。

同一個人が繰り返し同じような政策を経験した場合には、過去の政策の経験が記憶に残っている場合と残っていない場合で対応が異なってくることがある。個人の学習によって政策の効果が違ってくるといことは、ケインズ政策に対する批判としてよく論じられた点ではあるが、政策評価を行う上で

も重要な点である。

## 2.3 因果関係の哲学的分析

### 2.4 個人処置効果の推定

個人  $i$  がある (政策の処置を受けた場合 (1 と表示) と受けなかった場合 (0 と表示) の効果は原理的には次のように推定できる。

$$\text{個人処置効果} : y_{1i} - y_{0i} \quad (1)$$

すでに論じたように、(1) を同一個人について推定することは不可能である。そこで処置を受けたグループと受けなかった (対照) グループの平均をとって、その差をみるのが考えられる。

$$\text{平均処置効果} : E(y_1 - y_0) = E(y_1) - E(y_0) \quad (2)$$

もし処置グループと対照グループの分布が対称分布でないような場合には、中位値を調べるほうが適切である。

$$\begin{array}{l} \text{中位処置効果} \\ (\text{median effect}) \end{array} : \text{Med}(y_1 - y_0) \neq \text{Med}(y_1) - \text{Med}(y_0) \quad (3)$$

(3) から明らかなように、 $y_1 - y_0$  の中位値と  $y_1$  と  $y_0$  の中位値の差は等しくない。中位処理効果は実際の計算上、困難が多いので、平均処置効果を用いることが多い。

処置効果を同一条件下 (例えば  $x = x_0$ ) で推定することができれば限界平均処置効果がわかる。

$$\begin{array}{l} \text{限界平均処置効果} \\ (\text{marginal mean effect}) \end{array} : E(y_1 - y_0) = E(y_1 - y_0 | x = x_0) \quad (4)$$

ここで  $x$  は連続変数でもいいし、あるカテゴリーに属していることを表す指標であつてもいい。例えば  $x_0$  は男性であれば 1、女性であれば 0 となるような性別を表す指標であり、平均処理効果を男女別に推定したと考えればいいだろう。

個人処置効果が測れないことは問題ではあるが、ある政策の効果を評価するということは個人にとっての効果というより、社会全体に対して平均的にどれぐらいの効果を与えたかということであろう。とすれば平均処置効果および限界平均処置効果が適切に推定できればある意味では十分である。

処置効果が適切に推定できるための条件を以下に列挙しておく。

#### (1) 異時点間での関係安定性

処置効果を推定するときに過去の処置に対する経験の有無が現在の処置の効果には影響を与えない。

(2) 効果の大きさに個人差はない

処置効果が個人間で異なるとすれば、先に述べた counterfactuals の手法は用いることができなくなる。同様に処置を受けなかった場合の結果にも個人差はないと考える。

(3) 独立性

処置プログラムの実施は他の説明変数とは独立に選ばれている。すなわちプログラムの参加者はランダムに選ばれていると考え、特定の属性によってプログラム参加者が選択されているということは考えない。

(4) 説明変数の単位の均一性

処置効果を推定するときに処置を受けた個人  $i$  が影響を受ける変数  $u_i$  と処置を受けなかった個人  $j$  が影響を受ける変数  $u_j$  の単位が均一でなければ処置効果を厳密に識別することができない。

例えば、 $u$  をカロリーの摂取量だとする。カロリー計算上は共通の単位で計算できるかもしれないが、個人によって体格が違ったあり、年齢、性別によって必要摂取量が違ってくるとすると、そのような違いを調整したうえでの比較可能な単位に計算しなおす必要がある。

処置群 (グループ) と対照群 (グループ) は処置の有無をのぞいては統計的性質に差がないことが重要になる。そのためにいかに対照群をランダムに選ぶかが問題になる。

社会実験としてある政策効果を推定したいということで政策を実施するのであれば、まず、全サンプルからランダムに標本を抽出し、さらにその上でランダムに処置群と対照群に分ければ無作為抽出による標本を 2 つ作ることができ、一方に処置を施し、他方には処置を施さないという実験を行うことができる。

しかし、社会政策は一般に管理実験ではなく、現実の諸々の問題に対処するために実施するものであって、処置群と対照群を事前に決めて政策を実施するという事は難しい。

The Causal effect defined by an experiment where individuals with given  $W$  characteristics are randomly assigned to treatment and control groups, and the causal effect is the expected value of the outcome.

The Differences-in-Differences Estimator

With panel-data, the causal effect can be estimated using the differences-in-differences estimator, which is the average change in  $y$  in the treatment

group over the course of the experiment, minus the average change in  $y$  in the control group over the same time.

#### Causality Analysis (1)

$\left\{ \begin{array}{l} \text{Counterfactual causality} \\ \text{Probabilistic causality} \end{array} \right. \quad \text{Probability theory}$   
 e.g. タバコを吸うとガンになる  
 $\left\{ \begin{array}{l} \text{Partial equilibrium} \\ \text{General equilibrium} \end{array} \right.$

Temporal causality  $\left\{ \begin{array}{l} \text{without memory} \\ \text{with memory} \end{array} \right.$

individual treatment effect:  $y_{1i} - y_{0i}$

mean (average) effect:  $E(y_1 - y_0) = E(y_1) - E(y_0)$

median effect:  $Med(y_1 - y_0) \neq Med(y_1) - Med(y_0)$

mean effect is easier to handle.

marginal mean effect:  $E(y_1 - y_0) = E(y_1 - y_0 | x = x_0)$

the strongest version of no effect:  $y_{1i} - y_{0i} \forall i$

$y_1$  and  $y_0$  are exchangeable:  $P(y_0 \leq t_0, -y_1 \leq t_1) = P(y_1 \leq t_0, -y_0 \leq t_1) \forall t_0, t_1$

#### Causality Analysis (2)

Rubin's Model

$y_t(u) - y_c(u)$

$\Rightarrow t$  causes the effect  $y_t(u) - y_c(u)$

Fundamental Problem of Causal Inference: It is impossible to observe the value of  $y_t(u)$  and  $y_c(u)$  on the same unit and, therefore, it is impossible to observe the effect of  $t$  on  $u$ .

The statistical solution: The average causal effect  $T$  of  $t$  over  $U$  is the expected value of the difference

$$E(y_t - y_c) = T \Rightarrow T = E(y_t) - E(y_c)$$

The statistical solution replaces the impossible-to-observe causal effect of  $t$  on a specific unit with the possible-to-estimate average causal effect of  $t$  over a population of units.

#### Causality Analysis (4)

(2) unit homogeneity 効果に個人差なし。

$y_t(u_1) = y_t(u_2)$  and  $y_c(u_1) = y_c(u_2)$  for two units  $u_1$  and  $u_2$ . the causal effect of  $t$  is taken to be the value of  $y_t(u_1) = y_c(u_2)$ .

## (3) Independence

Cause ( $t$  or  $c$ )  $u$  is exposed to be regarded as statistically independent of all other variables, including  $y_t$  and  $y_c$ . This means if the physical randomization is carried out correctly, then it is plausible that  $s$  is independent of  $y_t$  and  $y_c$  and other variables over  $u$ .

$$\begin{aligned} E(y_t) &= E(y_t | s = t) \\ E(y_c) &= E(y_c | s = c) \\ \Rightarrow T &= E(y_s | s = t) - E(y_s | s = c) \end{aligned}$$

The data  $(s, y_s)$  is used to estimate  $T$  by taking the difference between the average value of the observed response  $y_s$  for the units with  $s = t$  and  $s = c$ .

## (4) Constant effect

The assumption of constant effect is that the effect of  $t$  on every unit is the same.

$$T = y_t(u) - y_c(u) \quad \forall u \text{ is } U.$$

This assumption is also called additivity in statistical models for experiments because the treatment  $t$  adds a constant amount  $T$  to be control response for each unit.

Philosopher's idea on causality.

- Aristotle  $\left\{ \begin{array}{l} \text{material} \\ \text{formal} \\ \text{efficient} \\ \text{final} \end{array} \right\}$  cause

He is concerned with the causes of a thing rather than the effects of causes.

- Hume emphasized that causation is a relation between experiences rather than one between facts three criteria (a) spatial/temporal contiguity, (b) succession, (c) constant conjunction.

- Mill

- (1) the method of concomitant variation
- (2) the method of difference: caused effect
- (3) the method of residues:  $y_{ab}(u) - y_a(u)$

- (4) the method of agreement: rule out possible causes 起こったことのうち唯一の違いを見つける

● Suppes Probabilistic theory of causality

- $\left\{ \begin{array}{l} \text{a prima facie cause} \\ \text{spurious cause} \\ \text{a genuine cause (a prima facie not spurious)} \end{array} \right.$

notion of a genuine cause is simply a correlation between a cause and effect that will not go away by partially and legitimate competing causes.

What can be a cause?

- (A) She did well on the exam because she is a woman.  
 (B) She did well on the exam because she studied for it.  
 (C) She did well on the exam because she was coached by her teacher.

(A) cause is an attributed she possesses (属性)

(B) 彼女の内政的努力、(C) 外生的に与えられた行事による。

(A) 試験の成績は女性のほうが男性より良いことを意味しているが、これは現実的ではない。

(C) コーチがつかなければ成績は悪かっただろう。

(B) は内政的努力ということであるが、努力しても成績の悪かった人もいる。内生的問題は因果関係を考えるうえで難しい。

Fisher's attach on those who used the association between smoking and lung cancer as evidence of causal link between them as an example of the difficulty in deciding whether or not smoking is an attribute or a cause.

⇒ Fisher は smoking と cancer の間に遺伝子的特徴 (certain genetic differences) を考慮する必要を説いた。

Granger

a "cause" ought to improve our ability to predict an effect in a probabilistic system.

⇒ temporal  $\left\{ \begin{array}{l} \text{cause} \\ \text{succession} \end{array} \right.$

$X$  is not a granger cause of  $Y$  (relative to the information  $Z$ ) if  $X$  and  $Y$  are conditionally independent given  $Z$ .

⇒ Granger noncausality is similar to Suppes's notion of a spurious cause.

Controlled group と treated group がそれ以外の要因では差がないということが重要。

図 1

いくつかの考えられる問題点

non-randomness job training program の参加者を名前のアルファベット順の上半分と下半分で分けたとする。苗字のアルファベットそのものが人種的

な傾向を反映しており、必ずしもランダムに分布しているわけではない。労働経験、学歴等も人種的な差がある。つまり、Treatment  $X_i$  と誤差項に相関がある。この場合、効果の推定にはバイアスが入ってくる。

Failure of treatment protocol = partial compliance with the treatment protocol

選択された人が参加しなかったり、選択されなかった人が他所で処置を受けることがあれば選択事態はランダムでも処置を受けたかどうかはランダムでなくなることがある。

Attrition = drop out after being randomly assigned to the treatment or control group

脱落の理由がランダムではなく処置そのものに関係しているとバイアスが生じる。

Experimental effects: Hawthorne effect, General Electric Company's productivity study.

実験化において処置群グループか Control group (対照群) に入っているかは自明で、それが処置以上の効果をもたらす可能性がある。薬学では二重盲検 (double-blind experiment) を行って薬の効果を測るが、社会科学では盲検不可能であり、その結果、処置以外の効果も含まれてくる。(ワシントンの国立病院ではガンの処置についても二重盲検を行っている) (placebo (偽薬) の使用の倫理性)

Small samples: サンプルの大きさは予算に制約される。small sample bias はあるかもしれないし、ないかもしれない。因果関係は不適切に推定される。

実験的な意味での因果分析はすべての要因が管理できるという前提で行われている。社会生活における因果分析は管理できずにあいまいな部分が残る。そのあいまいさが議論の対象となる。

## 図 2

事件がおきたときの判断 公正な裁判における事実認定とその原因の公正な確定などには不可欠な議論

a, b は e を起こそうとしたが、a が結果的に e を起こした。では b は e の原因ではないといえるのか?

— e が起こるための必要十分条件を確認する Third explanation many need

— a が起こらなければ e は起こらない、すなわち a を阻止すれば e は決して起こらないといえるか。 Causation by omission

## Causation under probabilistic laws

確率的因果性

c は e の原因であるとする。

c が起こらなければ e が起こる確率は c が起こったときよりはるかに低い。

## 図 3

a が起こったが e に影響 (届く) を与える前に e が反応して起こる。この場合 a は e の原因なのか?

trumping preemption

図 4

causation by omission

不在がある結果をもたらす。空気がなければ死ぬとか

Transitivity

c causes d, d causes e, then c causes e?

原因と結果はそれぞれのペアで独立している。風が吹けば桶屋が儲かるという落語があるが、中間にたくさんの原因結果を入れることで transitivity の無意味さを表現しているが、原因と結果の間に隔たりがあればそのような suprisous な原因性が入ってくる。

電車事故の原因は何か?

①運転手のミス、②電車の故障、③スケジュールの過密、④会社の体質、⑤  
予期せぬ対応の遅れ

⇒ 帰結 107 名の死者、400 名を超える障害

1. 運転手の行動 (何によって行動が引き起こされたか) が事故を通して乗客の死をもたらした。
2. 運転手が乗客を殺した。
3. 会社が乗客を殺した。
4. 病院の対応が追いつかず乗客が死んだ。

政策効果といった場合に政策をどう捉えるのかということが重要。原因 (責任者、参加者) は多数に上る。政策の実施には多くの人が動員されすべてが有機的に動く必要がある。

図 5

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (5)$$

$x_i$  is the treatment level

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_{1i} + \dots + \beta_{1+r} w_{ri} + u_i \quad i=1, \dots, n \quad (6)$$

$w_{ri}$  is additional regressors to control  $y_i$



(2) is consistent under the conditional mean independence (i.e.  $u_i$  depends on  $w_{ri}$ , but not on  $x_i$ )

$$(i) E(u_i | x_i, w_{1i}, \dots, w_{ri}) = 0$$

(ii)  $x_i$  is randomly assigned,

(iii) treatment  $x_i$  is assigned randomly, conditional on  $w_{ri}$

If  $w_i$  is a set of indicator variables, conditional mean independence means that  $x_i$  is randomly assigned within each group or "block" (block randomization)

Under the conditional mean assumption,  $\beta_1$  is the treatment effect.

$$E(y_i | x_i, w_{ri}, \dots, w_{ri}) = \beta_0 + \beta_1 x_i + \beta_2 w_{1i} + \dots + \beta_{1+r} w_{ri} + E(u_i | x_i, w_{1i}, \dots, w_{ri}) \quad (7)$$

Evaluating the conditional expectation (3) at  $x_i = 1$  and at  $x_i = 0$ , i.e.

$$E(y_i | x_i = 0, w_{1i}, \dots, w_{ri}) = \beta_1 \quad (8)$$

Checking out covariate balance

ランダム度の検定 Testing for Randomization

(0) 処置と対照グループの平均標準偏差が等しいかどうかを検定する。

(1) 処置をランダムに受けているとすれば、処置  $x_i$  は他の  $w_{ri}$  属性に依存しない。これを検定するには

$$x_i = \alpha + \beta w_{1i} + \gamma w_{2i} + \dots + \phi w_{ri}$$

$$H_0 : \beta = \alpha = \phi = \dots = 0$$

F-test for  $H_0$  をおこなう  $z_i$ : assignment

(2) 処置がランダムに施されて (assigned) いるとすれば、 $z_i$  は  $w_i$  とは相関していない。

$$z_i = \alpha + \beta w_{1i} + \gamma w_{2i} + \dots + \phi w_{ri} + u_i$$

$$H_0 : \beta = \gamma = \phi = \dots = 0$$

F-test for  $H_0$

—社会科学における因果性分析の難しさ

1. 管理実験が出来ない
2. 理論検証のための合意がいられていない
3. 複数の経済モデルの競合状況にあり、それぞれのモデルを厳密に反証するにはデータが不十分 (反証可能性が獲得されにくい)
4. 社会観の違いが社会経済政策に反映されることを良しとする伝統がある
5. 歴史における "if what" の厳密な意味での検証が行われていない (comter-factual の徹底的探求)

图1

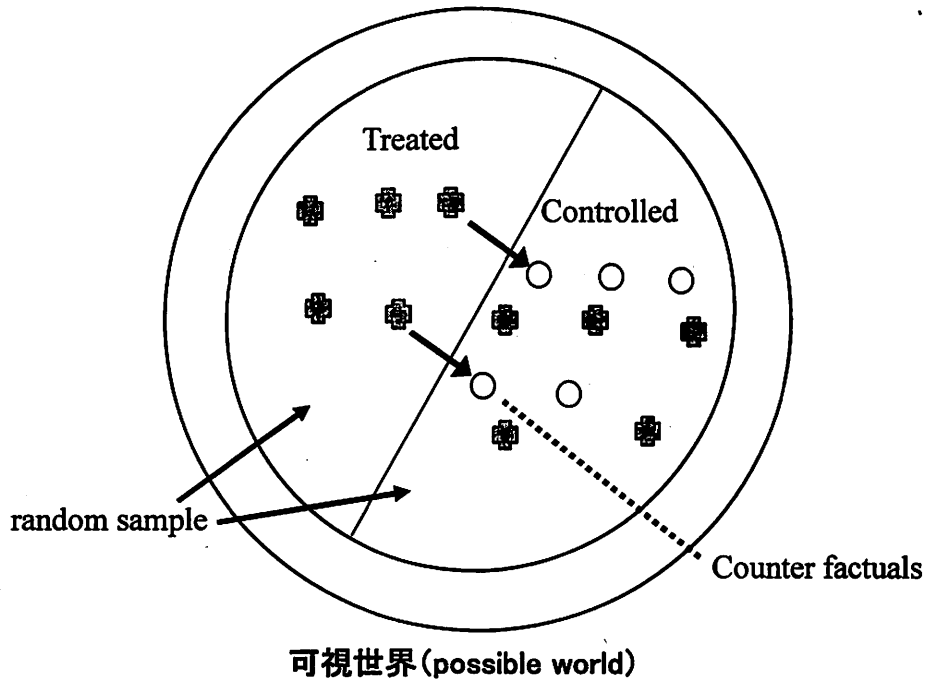


图2

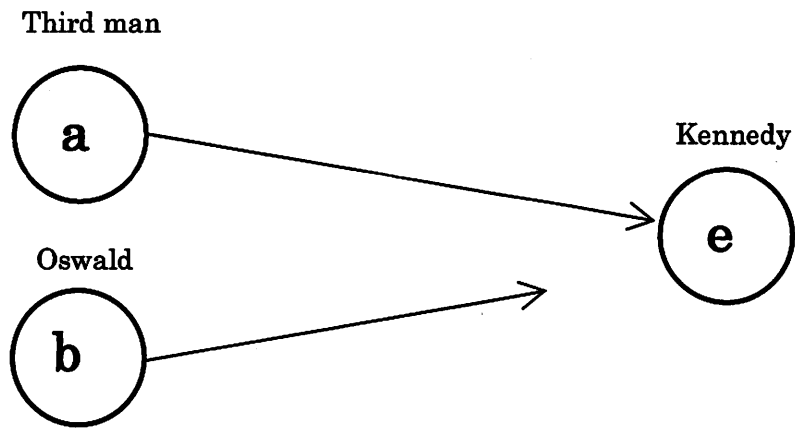


图3

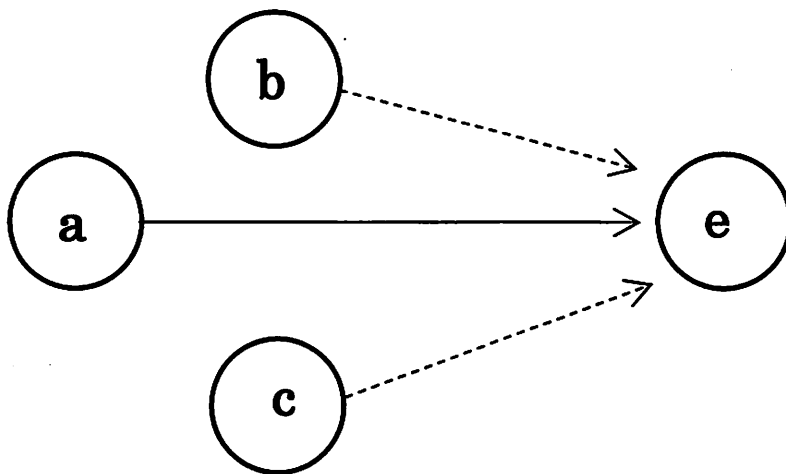
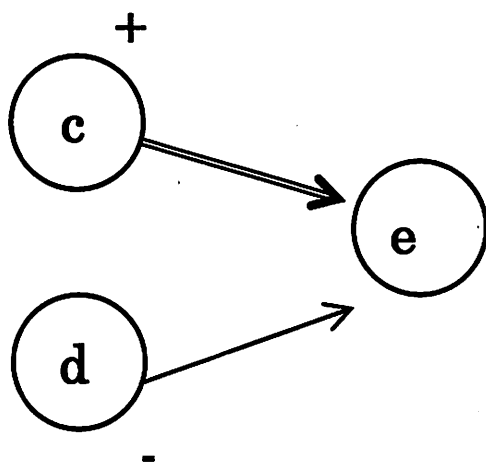
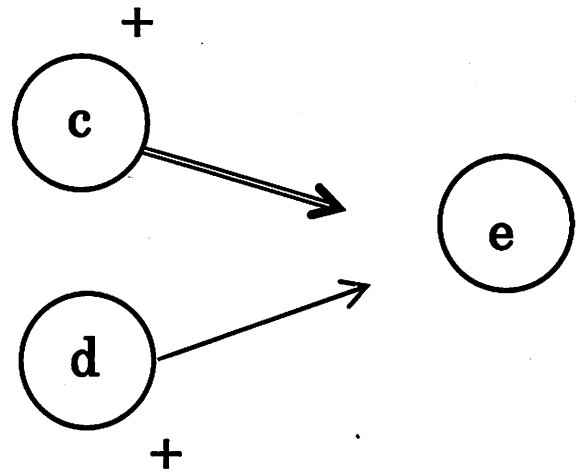


図4

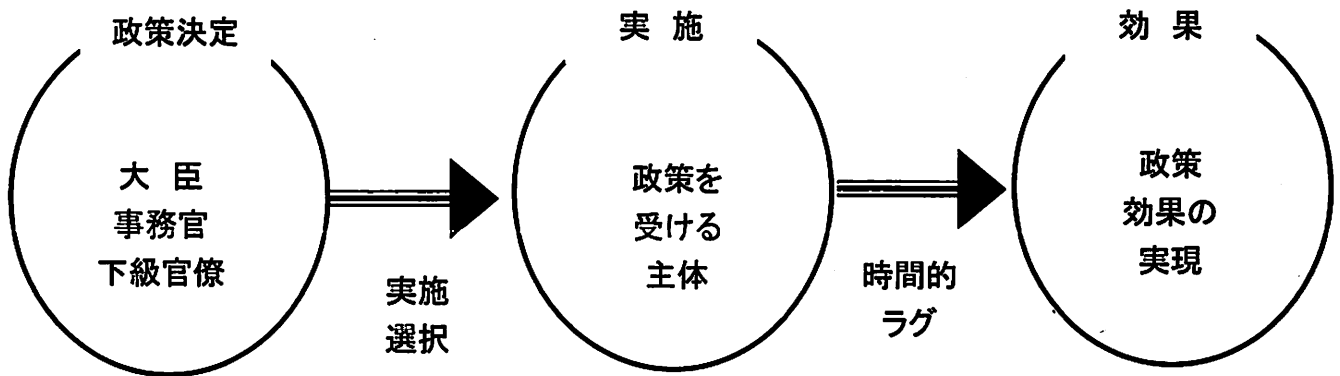


dは負の効果があるにもかかわらずeが起こったとすればcがeを起こしたといえる。



cとdはともに正の効果があるがcの方がより強い(intense)効果がeにあるということを示す必要がある。

図5



政策が実施されるためにはかなり多くの人参加が必要

主体の属性やプログラムの質の違いによって効果が変わってくる