研究仮設のモデル化とその定量的分析

メタデータ	言語: Japanese
	出版者:
	公開日: 2007-12-26
	キーワード (Ja):
	キーワード (En):
	作成者: 高木, 廣文
	メールアドレス:
	所属:
URL	http://hdl.handle.net/10285/170

研究仮説のモデル化とその定量的分析

高木廣文

I. はじめに

医学，看護学の分野ではもちろんのこと，心理学や社会学など他の多くの研究領域において，研究開始時にあらかじめ仮説を設定することの重要性が強調されてきた。調査や分析は，その研究仮説にそって行なわれ，仮説を検証することに主眼が置かれてきた。そのような研究方法は，いわゆる仮説検証的方法と呼ばれるものであり，実験のように各種の条件設定が比較的容易な場合には，極めて有効なものであるが，調査という人間集団を対象とする場合，必ずしも優れて機能するわけではない。

実験と異なり，人間集団では単純な仮説を検証する場合であっても，そのための条件設定が困難であることが大部分である。現在では，人間集団を扱う研究では，データ解析 data analysis の立場から研究を行うのが一般的となってきた[1,2]。この立場をより明確にしたのは，J. W. Tukey らによる探索的データ解析 exploratory data analysis という考え方であるといえよう[2~4]。特定の研究において，あらかじめ研究上の仮説を枠組として固定するかわりに，先入観なしにデータを分析し，データの中からその意味する内容を探索し，その結果を踏まえて，次のデータ解析を行っていくという立場が，上記の探索的データ解析であると考えられる。そのようなデータ解析の過程を通して，研究上重要な仮説が構築され，現象の真実の姿が理解できるようになるものと期待される。このような考え方からすると，新たな課題について研究を始める場合，研究仮説を頑強に構築するのは，自由で創造的な活動の妨げとなる恐れすらあるといえよう。この点からいえば，仮説検証的方法は古典的であり，研究当初において避けるべき方法といえないこともない。ただし，仮説検証的方法にも優れた点がある。その一つとして，統計学の手法を用いることにより，仮説の正当性を判断するために，確率という客観的な基準を用いることができる。また，探索的データ解析の立場といえども，研究の進展に伴い，特定の仮説が提示され，それを検証する必要が生じることがあるだろう。ここでは，そのような場合において，仮説をモデルとして図示し，定量的に分析するための方法として，パス解析の手法を紹介する。

II. 研究仮説のモデル化の意義

一般に，研究上の仮説は文章により記述されている。例えば，「塩分の過剰摂取は高血圧の原因である」といった形式をとることが多い。そのような研究仮説は，統計的仮説検定の結果を踏まえて考察されることになるが，必ずしも直接的に研究仮説を検証している訳でない点に注意がいる。例えば，2変数間の関係を調べるために，相関係数を求め，その検定を行う場合を考えてみよう。通常，相関係数の検定といえば「無相関の検定」を意味しており，検定結果が有意であれば，それは単に母相関が0でないことを示すにすぎない。すなわち，その結果から母相関が0.6とか0.7とかをいえる訳ではない。また逆に，検定が有意でないからといって，母相関が0であるということを，積極的に支持できる訳でもない。この辺りの事情については，一般の統計学の成書に詳しく述べられているので，これ以上説明はしない。ただ，統計的仮説検定による結果の解釈が，不当に拡大されているという場合が多くみられる点に注意を向けるべきだろう。その理由の一つは，研究者の統計学に対する勉強不足であり，さらに自己の立てた研究仮説に対する愛着が，より大きな理由の一つとなっていると考えられる。すなわち，研究者が分析結果に対して，自己の仮説に対して有利なように結果を解釈しがちである。これは，一つには研究仮説が未熟なため，定式化されておらず，言語により表現されていることにもよる。言葉はいくらその概念を規定しても，その規定自体が言葉でなされるのであるから，本質的にどこかに曖昧さが残される。そのような曖昧さが，分析結果の解釈にも影響していると考えられよう。そのような曖昧さを除くための一手法として，実測可能な変数のみを用いて，研究仮説をモデル化することが考えられる。当然，そのようなモデル化には既存の文献等が十分活用される必要があるが，ここでは特にその点については触れない。ただし，苦

労して作成したモデルが，単に定性的にしか評価できなければ，それは徒労に終ることになろう。モデルの定量的評価のための手法として，ここではパス解析の方法を中心に紹介し，さらにその限界などにも若干の考察を加えることにする。

III. パス解析 path analysis

パス解析は，生物学者 Wright[5]により考案され，主に社会学の分野で発展した。とくに，変数間の相関・偏相関から因果推論を行うための Blalock の因果律は有名である[6]。パス解析は，形式的には変数間の相関関係をもとに，因果推論を行うため，因果分析法 causal analysis として知られているが，この呼称はしばしば誤解を招く。すなわち，因果推論はあくまで推論であって，因果関係を正しく表わしているのではないことを見落すべきではない。以後，特にこの点について断わることはしないが，常に応用にあたって最も注意すべき点の一つである。また，実際の場では，人間集団を対象とした標本調査により，正しくデータが収集される必要がある。この問題についても，ここでは扱わない。しかし，それはデータ収集が容易だからではなく，分析よりも多くの問題を内包しているため，とてもここでは説明しきれないという理由による。この点に関しては，林[7]を参照すればよいだろう。

1．研究仮説のモデル化

最も単純な研究上の仮説として，「変数Xは変数Yに影響する」というものを考えてみよう。上記の仮説をより大胆にいえば，「……に影響する」を「……の原因である」となるが，ここでは因果関係を示唆する用語は，差し控えることにする。

上記の仮説は，言葉により表わされているので，このままでは統計学の手法に馴染まない。一般的に考えると，上記の仮説は変数YがXのある関数 f(x)により表わされると仮定しているものと考えられる。すなわち

$$y = f(x) \qquad (1)$$

とかけることを示すものである。

関数 f(x)は，Yの変動を説明するのに最適なものである必要がある。例えば，Yとして集団中の特定の疾患の死亡率を，Xとしてある物質の摂取量が考えられているとしよう。そのような場合，Xの増加につれ，Yは直線的に増加するのではなく，S字形の曲線を描いて上昇するのが一般的である。したがって，f(x)としては良く知られているように，

$$f(x) = 1/[1+exp\{-(ax+b)\}] \qquad (2)$$

のような，ロジスティック関数を用いることが多い。ただし，ここでaは変数Xにかかる重みであり，bは定数である。

上記のような特別な場合を除いて，一般にはより簡単なモデルを考えるのが普通である。すなわち，変数YはXの一次式で表わされるとするものである。

$$y = ax+b \qquad (3)$$

である。上式は，変数Yを基準変数，Xを説明変数とした場合の単回帰式である。

パス解析でも，考え方はほぼ同様であるが，数式よりもまず図によってモデルを表わすことから分析が始まるといえる。図1は，前述の仮説を図示したものであり，パスモデル path model と呼ばれる。図中のXからYに到る矢印はパス path と呼ばれ，XがYに影響を及ぼしていることを示している。また，図中のUは攪乱因子 disturbance と呼ばれるものであり，X以外でYに影響を及ぼしているものの総体を示している。パス上のa，bは，変数X，UのYへの各影響の程度を示すもので，パス係数 path coefficient と呼ばれる。パス解析は，各変数間の相関係数をもとに，各パス係数を推定することに主眼がある。そのためには，他の変数の影響を受ける変数（内生変数）について，(3)式のように関係式をかきだす必要がある。図1については，

$$Y = aX+bU \qquad (4)$$

となる。(4)式は(3)式と異なり，攪乱因子についても考慮している。上記の式は，構造式 structure equation と呼ばれることがある。上式をもとに，パス係数を推定するわけであるが，そのための基本的な仮定と手順を次に示すことにする。

2．パス係数の推定

パス解析では，基本的な仮定として，すべての変数（攪乱因子も含めて）は平均0，分散1に標準化されているものとする。この仮定により，変数間の共分散 Cov(X，Y)は相関係数 ρ_{xy} と一致する。したがって，パス係数と各相関係数の関係式を求めるのが，極めて容易となる。例えば（4）式より，

$$\begin{aligned}\rho_{xy} &= C_{ov}(X,\ Y)\\ &= C_{ov}(X,\ aX+bU)\\ &= C_{ov}(X,\ aX)+C_{ov}(X,\ bU)\\ &= a\cdot C_{ov}(X,\ X)+b\cdot C_{ov}(X,\ U)\\ &= a+b\rho_{xu}\end{aligned} \qquad (5)$$

となる。同様にして，

$$\begin{aligned}1 &= \rho_{yy} = C_{ov}(Y,\ Y)\\ &= a\cdot C_{ov}(X,\ Y)+b\cdot C_{ov}(Y,\ U)\\ &= a\rho_{xy}+b\rho_{yu} \qquad (6)\end{aligned}$$

となり，

$$\rho_{yu} = a\rho_{xy}+b \qquad (7)$$

となる。ここで，ρ_{xy}，ρ_{xu}，ρ_{yu}などは各変数間の相関係数を示すものである。上記の関係式から，パス係数を求めればよいが，未知数がa，b，ρ_{xu}，ρ_{yu}の4つあり，方程式は3つしかないので，このままでは解が求まらない。解を求めるためには，もう一つ条件式があればよい。図1のパスモデルにおいて，変数Xと攪乱因子Uの間には，相互に到達可能なパスによる経路が存在していないので，相関を0と仮定する。すなわち，$\rho_{xu}=0$とする。なお，この仮定はパス解析において一般的なものである。詳しく述べると，2変数間の経路として，他の変数を経由する場合もありえるが，そのような場合には，経由する変数に向うパスと，他の変数に向うパスの両方が存在している必要がある。変数間に到達可能なパスによる経路があれば，相関の存在が仮定され，なければ無相関とされる。

図1の場合，$\rho_{xu}=0$であるから，(5)式より，

$$a = \rho_{xy} \qquad (8)$$

となる。また，(6)，(7)式より，

$$b^2 = 1-a^2 = 1-\rho_{xy}^2 \qquad (9)$$

と求められる。したがって，図1のモデルでは，XのYに対する影響の大きさは，相関係数ρ_{xy}となる。また，攪乱因子の影響は，決定係数ρ_{xy}^2を1から引いたものの平方根となっている。

3．因果連鎖モデル

ここでは，前述のパス解析の手法をより明確に理解するために，若干複雑なモデルを扱うことにする。

研究仮説として，「変数Xは変数Yに影響し，さらにYは変数Zに影響する」とする。上記の仮説をパスモデルにより図示したものが図2である。このようなモデルは，変数間の連鎖的影響を示すものであるため，因果連鎖モデル causal chain model と呼ばれる。図中のU，Vは変数Y，Zの各攪乱因子を示している。図2から，YとZについての構造式を求めると，

$$Y = aX+cU \qquad (10)$$
$$Z = bY+dV \qquad (11)$$

となる。すでに述べた方法により，変数間の相関係数

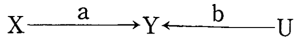

図1　簡単なパスモデル

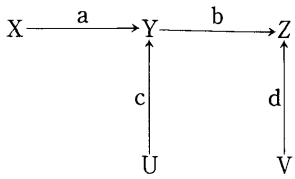

図2　因果連鎖モデル

とパス係数との関係式をすべて求める。まず，(10)式より，

$$\left.\begin{aligned}\rho_{xy} &= a \quad\ \ +c\rho_{xu}\\ 1 &= a\rho_{xy}+c\rho_{yu}\\ \rho_{zy} &= a\rho_{zx}+c\rho_{zu}\end{aligned}\right\} \qquad (12)$$

となり，(11)式から，

$$\left.\begin{aligned}\rho_{zx} &= b\rho_{xy}+d\rho_{xv}\\ \rho_{yz} &= b \quad\ \ +d\rho_{yv}\\ 1 &= b\rho_{yz}+d\rho_{zv}\end{aligned}\right\} \qquad (13)$$

が得られる。また，(10)，(11)式から，

$$\rho_{yz} = ab\rho_{xy}+bc\rho_{yu}+ad\rho_{xv}+cd\rho_{uv} \qquad (14)$$

となる。ここで，図から，

$$\rho_{xu} = \rho_{xv} = \rho_{yv} = \rho_{uv} = 0$$

が仮定されていることに注意すると，

$$\left.\begin{aligned}a &= \rho_{xy},\text{または}\ a = \rho_{zx}/\rho_{yz}\\ b &= \rho_{yz},\text{または}\ b = \rho_{zx}/\rho_{xy}\end{aligned}\right\} \qquad (15)$$

となり，

$$c^2 = 1-a^2,\ d^2 = 1-b^2 \qquad (16)$$

となる。(15)式で示したように，パス係数a，bの解は2通り存在することになる。2通りの解が，唯一の数値をとるためには，

$$ab = \rho_{xy}\ \rho_{yz} = \rho_{xz} \qquad (17)$$

が成り立てばよい。この条件は，制約条件 restriction と呼ばれることがある。

上記のように，解が一意でない理由は，変数X，Y，Z間の異なる変数間の相関係数の個数が3であるのに，未知のパス係数はa，bの2つしかないためであ

る。ただし，攪乱因子のパス係数は，他の数値により定められるので考慮しなくてよい。このように，未知のパス係数の個数が，異なる変数間の相関係数の個数よりも少ない場合，モデルは識別性過度 overidentification であるといわれ，解は複数個存在する。逆に，パス係数の個数が異なる変数間の相関係数の個数より多い場合，モデルは識別性不足 underidentification であるといわれ，解を求めることができない。そのような場合には，いくつかのパス係数間に特定の関係式を仮定するなどして，自由度を減らし，推定すべきパス係数の個数と相関係数の個数が一致するように調整する必要がある。未知のパス係数の個数と異なる変数間の相関係数の個数が等しい場合，すなわち識別性適合 justidentifieation の場合のみ，解を一意に定めることができる。

ところで，(17)式の制約条件は，XとZからYの影響を除いた偏相関係数 $\rho_{x/y\cdot z/y}$,

$$\rho_{x/y\cdot z/y} = \frac{\rho_{xz} - \rho_{xy}\ \rho_{yz}}{\sqrt{(1-\rho_{xy}{}^2)(1-\rho_{yz}{}^2)}} \tag{18}$$

が0という条件に等しい。したがって，図2のモデルが妥当か否かは，標本偏相関係数 $r_{x/y\cdot z/y}$ をデータから計算し，無相関の検定を行えばよい。検定結果が有意でなければ，仮定したモデルは正しいものと考えられるし，有意ならばモデルは適当な修正が必要となるだろう。実際にパス係数を算出する場合，制約条件が完全に満たされることは稀であるから，パス係数の2つ以上の推定値について，その算術平均か幾何平均を求めればよいだろう。

ところで，(17)式は因果連鎖をなす変数間の相関について，重要な関係を示している。すなわち，変数間の相関係数が全て非負ならば，相関係数は1以下であるから，

$$\rho_{xz} \leqq \rho_{xy} \text{かつ } \rho_{xz} \leqq \rho_{yz} \tag{19}$$

の関係が成り立つ。このことは，因果連鎖をなす変数間では，相互に隣り合う変数間の相関は，離れた位置にある変数間の相関に比べ，その値が大きいことを示すものである。なお，上記の関係は，4変数以上の因果連鎖モデルについても，容易に拡張できる[6]。したがって，既存の知識からパスモデルを設定する場合，変数間の相関係数が他の組合せのものより大きい変数は，それぞれ相互に隣接させ，相関係数が小さい場合には，相互に遠い位置になるように，各変数をモデル上で配置すべきことがわかる。Takagi は，上記の考えから，標本相関係数をもとに，客観的にパスモデルを作成するためのアルゴリズムを提唱している[8,9]。

4．フィードバックループのあるモデル

ここでは，変数間にフィードバックループが存在する場合について考えてみよう。研究仮説は，「変数Xは変数Yに影響を及ぼすが，YもXに影響を及ぼす」となる。この仮説をパスモデルにかくと，図3のように

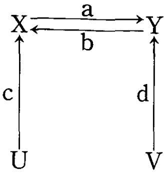

図3　2変数間のフィードバック

なる。XとYについて，その構造式をかくと，

$$Y = aX + dV \tag{20}$$
$$X = bY + cU \tag{21}$$

となる。$\rho_{uv} = 0$ に注意して，まず(20)式から相関係数とパス係数の関係式を求めると，

$$\left.\begin{array}{ll} \rho_{xy} = a + d\rho_{xv}, & 1 = a\rho_{xy} + d\rho_{yv} \\ \rho_{yu} = a\rho_{xu}, & \rho_{yv} = a\rho_{xv} + d \end{array}\right\} \tag{22}$$

となり，(21)式からは，

$$\left.\begin{array}{ll} \rho_{xy} = b + c\rho_{yu}, & 1 = b\rho_{xy} + c\rho_{xu} \\ \rho_{xu} = b\rho_{yu} + c, & \rho_{xv} = b\rho_{yv} \end{array}\right\} \tag{23}$$

となる。これを整理することで，

$$\left.\begin{array}{ll} \rho_{xu} = c/(1-ab), & \rho_{yu} = ac/(1-ab) \\ \rho_{xv} = bd/(1-ab), & \rho_{yv} = d/(1-ab) \end{array}\right\} \tag{24}$$

となることがわかるので，

$$\rho_{xy} = (a+b)/(1+ab) \tag{25}$$

および，

$$\left.\begin{array}{l} c^2 = (1-b\rho_{xy})(1-ab) \\ d^2 = (1-a\rho_{xy})(1-ab) \end{array}\right\} \tag{26}$$

と求められる。しかし，すでに述べたように，2つのパス係数a，bを，1つの相関係数 ρ_{xy} のみから推定することは，(25)式で明らかなように不可能である。このような場合，aとbの間に特定の関係を仮定する必要がある。XとYの影響の及ぼし方が等しいと仮定できるのならば，a＝bとすればよいし，Xに比べYはその半分の影響しか及ぼさないとすれば，b＝0.5aとす

ればよい。いま，b＝0.5a と仮定すると，(25)式より，

$$a = \frac{3 \pm \sqrt{9 - 8\rho_{xy}^2}}{2\rho_{xy}} \qquad (27)$$

となる。$\rho_{xy}=0.8$とすれば，

$$\left.\begin{array}{ll} a \fallingdotseq 3.106, & b \fallingdotseq 1.553 \\ c^2 \fallingdotseq 0.927, & d^2 \fallingdotseq 5.667 \end{array}\right\} \qquad (28)$$

または，

$$\left.\begin{array}{ll} a \fallingdotseq 0.644, & b \fallingdotseq 0.322 \\ c^2 \fallingdotseq 0.588, & d^2 \fallingdotseq 0.384 \end{array}\right\} \qquad (29)$$

と求められる。上記のように，2通りの推定値が得られたが，c^2とd^2は重回帰分析において，分散を1とした場合の残差分散に対応するものと考えられる。したがって，その数値が1以上となる(28)式による解は不適と考えられるので，(29)式の方を採用すべきである。

図4に，より複雑なフィードバックループの例を示した。すなわち，「変数Xは変数Yに影響し，変数Yは変数Zに影響し，そしてZはXに影響する」というモデルである。

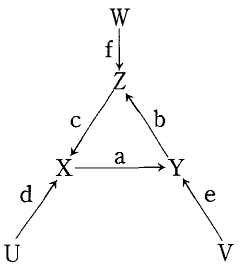

図4　3変数のフィードバック

$$\left.\begin{array}{l} Y = aX + eV \\ Z = bY + fW \\ X = cZ + dU \end{array}\right\} \qquad (30)$$

である。上式から，各相関係数とパス係数の関係式を求めればよいが，その計算はかなり複雑である。結果を整理して示すと，

$$\left.\begin{array}{l} \rho_{xy} = (a+bc)/(1+abc),\ \rho_{yz} = (b+ca)/(1+abc) \\ \rho_{zy} = (c+ab)/(1+abc),\ d^2 = (1-c\rho_{zx})(1-abc) \\ e^2 = (1-a\rho_{xy})(1-abc),\ f^2 = (1-b\rho_{yz})(1-abc) \end{array}\right\} \qquad (31)$$

となる。上式から，a，b，cが推定できるように考えられるが，実際には代数的に簡単な形で各パス係数を表わすことができない。このような場合，(31)式から，

$$\left.\begin{array}{l} a = (\rho_{xy} - bc)/(1 - bc\rho_{xy}) \\ b = (\rho_{yz} - ca)/(1 - ca\rho_{yz}) \\ c = (\rho_{zx} - ab)/(1 - ab\rho_{zx}) \end{array}\right\} \qquad (32)$$

のように変形し，a，b，cに適当な初期値（－1から1の間の数値）を代入し，各値が一定になるまで反復計算を行い，数値的に解を求めればよい。

ここで，$\rho_{xy}=0.8$, $\rho_{yz}=0.7$, $\rho_{zx}=0.6$としよう。(32)式を用いて反復計算を行うと，60～70回程度の反復により，

$$\left.\begin{array}{lll} a \fallingdotseq 0.756, & b \fallingdotseq 0.624, & c \fallingdotseq 0.180 \\ d^2 \fallingdotseq 0.817, & e^2 \fallingdotseq 0.362, & f^2 \fallingdotseq 0.516 \end{array}\right\} \qquad (33)$$

と求められる。なお，(33)式による推定値が妥当なものかは，(31)式に各値を代入し，相関係数の値が再現されるかを調べればよい。

IV. 統計的手法とモデルについて

研究仮説のモデル化による図示，およびそれを用いての定量的分析という点から，パス解析についての説明を行った。しかし，特定のモデルを前提にして分析を行うのは，何もパス解析に限るわけではない。統計学の手法の多くは，特定の変数が他の変数の線形結合（ある値を変数にかけたり加えたりしたもの）により表わされるとする線形モデル linear model を仮定し，さらに誤差の分布として正規分布を仮定することがしばしばである。その代表的なものとして，重回帰モデルがあるし，その特別な場合が単回帰モデルである。したがって，2変数の相関を検討する場合，かりに2変数間に曲線関係が存在していれば，通常用いるようなピアソンの積率相関係数は無効となる。なぜならば，相関係数は単回帰モデルから導入されたものであり，変数間の線形関係を示すための統計量であるからである。この辺りの事情が了解されれば，無定見な相関係数の使用はずっと減少するはずである。

同様に，ここで紹介したパス解析も，パスモデルの考案，作成から，変数間の構造式という線形モデルを導き，それを用いて分析しているという点を忘れるべきではないだろう。したがって，変数間の関係が線形でなく，曲線関係などの非線形なものであることが認められる場合，前述したロジスティック・モデルなどが有効な場合もあるし，他のより適当なモデルを考えた方がよい場合も多いはずである。ただし，現実には変数間に線形関係を仮定しても，それが重大な誤りとなることは少い。したがって，ここで説明したようなパス解析の手法を用いて分析を行うことによる誤りよ

りも，有用な情報が入手できる可能性の方が大きいものと考えてよいと思われる。

最後に，パス解析も含めて統計的手法を用いる場合の問題点について整理しよう。

第一に,因果推論について考える必要があるだろう。どのような方法であれ，統計的手法のみから「因果関係を立証することはできない」ということは事実である。調査に基づくデータをどのように加工しても，因果関係を推論することはできても，直接的にその存在を証明することはできない。したがって，パス解析といえども，分析して得た情報は，たんにモデルについてのものにすぎない。統計的仮説検定が有意である場合，その結果のみから研究仮説が立証されたかのように結論付けている例が少なからずみられるが，そのような論理は誤りといえる。因果関係の立証のためには，調査ではなく，厳密に条件設定を管理された実験によらねばならない。

第二に，分析結果の数値に絶対の信頼を置くべきではない。調査データは，必ず何らかの誤差を含むものであり，その分析結果から得られた数値が，母集団のそれと定全に一致することは稀である。したがって，ある程度の誤差があるものと考えて，解釈する方が無難である。できれば，信頼区間を求めることが望ましい。

第三に，モデルとして採用したものが，唯一のものではないという点である。実際の研究において，どのような仮説的モデルを用いるかは，研究者のそれまでの知識と知恵により決定されるのであるから，視点を少しでも変えれば他のモデルも考えられるはずである。少くとも，自己の仮定したモデルに対して客観視ができないようでは，分析した結果に対する解釈もできないように思われる。

以上,注意すべき点について簡単な説明を加えたが,実際にはこの他にもいくつかあるかもしれない。しかし，結局のところデータ解析は理論のみでなく，データを実際にどのくらい扱ったかにより，力量の差が大きく異なるものである。モデルの設定や分析の仕方，解釈などは，実際にデータにまみれることで，上達もより早くなるようである。

なお，パス解析の方法をより詳しく知りたいのならば，Kenny[10]，Heise[11]，豊川[12]，柳井ら[13]を参照すればよいだろう。

V. まとめ

研究上の仮説をモデルとして図示し，変数間の相互の影響力を定量的に評価するための一手法として，主にパス解析の方法を紹介した。実際に，いくつかの研究仮説をモデル化し，変数間の相関係数を用いて，各変数のパス係数の推定方法について詳述した。

また，実際にデータ解析を行う場合の問題点を指摘した。すなわち，統計的分析方法における線形モデルの仮定の問題，調査データの誤差に付随する推定値の解釈上の問題，そして研究上のモデルの選択に関する問題について指摘し，あわせてデータ解析における基本的考え方についても若干の説明を加えた。

参考文献

1) 林知己夫：データ解析の考え方，東洋経済新報社（東京），1977
2) Tukey, J. W.: The future of data analysis, the Annals of Mathematical Statistics, 33, 1-67, 1962
3) Tukey, J. W.: Exploratory Data Analysis, Addison-Wesely, 1977
4) Hartwig, F. and Dearing, B. E.: Exploratory Data Analysis, Sage Pub., 1979（柳井晴夫，高木廣文訳：探索的データ解析の方法，朝倉書店，1981）
5) Wright, S.: Correlation and causation, Journal of Agricultural Research, 20, 557-585, 1921
6) Blalock, H. M. Jr.: Causal Inferences in Nonexperimental Research, the Univ. of North Carolina, Chapel Hill, 1964
7) 林知己夫：調査の科学ー社会調査の考え方と方法，講談社，1984
8) Takagi, H.: An algorithm for causal modeling in epidemiological studies, Behaviormetrika, 8, 41-55, 1980
9) 高木廣文：疫学研究における因果モデル作成上の新しい手法の開発とその応用について，聖路加看護大学紀要，8，11-20，1982
10) Kenny, D. A.: Correlation and Causality, John Wiley & Sons, 1979
11) Heise, D. R.: Causal Analysis, John Wiley & Sons, 1975
12) 豊川裕之編：疫学，メヂカルフレンド社，1984
13) 柳井晴夫，高木廣文編著：多変量解析ハンドブック，現代数学社，1986

（昭和60年12月20日受付）

聖路加看護大学紀要 Vol. 11. 1986. 5.

—欧文抄録—

MODELING THE HYPOTHESIS ON A NONEXPERIMENTAL STUDY AND ITS QUANTITATIVE ANALYSIS

Hirofumi Takagi

Path analysis was introduced and described as a method of modeling the working hypothesis on a nonexperimental research and as a technique of evaluating the internal effects among variables used in the model shown as a path diagram.

By the use of some simple examples, explained in detail were the method of building the concrete path model from the conceptual hypothesis and also the estimation method of path coefficients which indicate the degree of influence from a variable to another one.

Furthermore, it was indicated that some serious problems exist in data analysis on our real studies; i.e., the problem of the statistical hypothesis which includes both of assumptions of the linear model and the normal distribution in the statistical methods, the problem of the causal inference of estimated values computed from the data with the inevitable sample error, and the problem of the model selecton where the alternative model always exists if we try to consider it.

In addtion, the fundamental concept and attitude toward data analysis were described to some extent.