共分散分析とファジイ重回帰分析

吉田光雄

共分散分析とファジイ重回帰分析

1. 共分散分析

1. 1. 共分散分析におけるモデル

共分散分析（ANCOVA : Analysis of Covariance）とは分散分析の拡張であって，ある変数（目的変数，応答変数，response variable）の変動の分析をその変数に関連すると思われる他の変数（補助変数，随伴変数 concomitant variables または共変量 covariates 等と呼ばれる）の影響を取り去った後の残差から分析する手法である[2,7,8,12]。他変数の影響を取り去る方法として回帰が用いられ，その意味で回帰分析の応用と理解することもできる。すなわち，分散分析の要因（水準）または処理をクラスないしはグループとし，観測値が異なるグループに属する場合の回帰分析として，まず (1) 回帰がグループ毎に異なるかどうかに関心を持ち，(2) その結果をふまえて補助変数の影響を調整する方法，として理解することができる。成書に紹介されているのは，多くは補助変数が 1 変数の単回帰分析の場合であるが，ここでは多変数に拡張して整理する[7]。

共分散分析を数学的に整理すると，基本線型式

$$Y_{ij} = \mu + \alpha_j + \sum_l \beta_{jl} (X_{ijl} - \bar{X}_l) + \varepsilon_{ij}$$

ただし，i=1, …, n_j（オブザベーション），j=1, …, K（処理），l=1, …, p（変量）を仮定し，

仮説　$H : \beta_{1l} = \cdots = \beta_{Kl} \neq 0$

を採択し，共通の回帰勾配を確認した後，補助変数の影響を除去して

仮説　$H : \alpha_1 = \cdots = \alpha_K$

の採否を検討する方法といえる。回帰を導入することにより，通常の分散分析による分解

$$Y_{ij} = \mu + \alpha_j + e_{ij}$$

における残差 e_{ij} よりも小さい残差 ε_{ij} での分析が可能である。線型式は簡単のために一元配置固定モデルにおける式を示したが，容易に二元配置，多元配置に拡張しうる。

そして共分散分析が用いられるのは，Snedecor and Cochran[12] によれば，(1) 無作為化実験における精度の増大，(2) 観察による研究における偏り源の修正，(3) 無作為化実験において処理効果の性質を明らかにする，(4) 多元分類における回帰の研究，等であるが，本稿では回帰に関する研究の方法として取り上げ，さらに，回帰係数のあり方によって生じ得る

場合を，次のごときいくつかのモデルとして分類して一般的に論ずることとする。ただし，モデル名はDunn and Clark[2]に由来するが，本稿における仮称である。

[H_0：モデル0] Reduced Model　$E(Y \mid X) = \alpha_0$

回帰が有効ではなく，従って補助変数の導入が意味をなさず，かつグループ間の差も認められない場合。

[H_1：モデル1] ANOVA Model　$E(Y \mid X) = \alpha_j$

回帰が有効ではなく，従って補助変数の導入が意味をなさないが，目的変数によるグループ毎の分析は有効な場合。

[H_2：モデル2] Total Regression Model　$E(Y \mid X) = X\beta_0$

回帰は有効であるが，係数はグループ毎にすべて同じとみなし，重回帰分析を全データをこみにして行なう場合。

[H_3：モデル3] Within Regression Model　$E(Y \mid X) = X\beta_{0j}$

回帰は有効であるが，係数は変数によってグループ毎に同じものと異なるものが混在するとみなされる場合。すなわち，q個の補助変数のうち，r個を同じ（β_0），q−r個を異なる（β_j）とみなすものとする。共分散分析の場合，関心が持たれるのは要因の効果の有無であるため(fixed model)，補助変数を一括して同じとみなし，定数項が異なるかどうか，ということであるが，ここではその場合をも含めて，さらに一般化し任意の補助変数を分割して検討できるものとする。

[H_4：モデル4] Individual Regression Model　$E(Y \mid X) = X\beta_j$

回帰は有効であり，係数は変数によってグループ毎にすべて異なるものとみなし，重回帰分析を個別に行なう場合。

ただし，H_0，H_1におけるαは定数，H_2〜H_4におけるβはすべての回帰係数（このとき，補助変数$X_1 = 1$として定数項も含む）とする。回帰係数の選択は検定により行なってもよいし，先験的に与えてもよい。[モデル0]は全データが均等，[モデル1]は目的変数Yのみを有効とみなして，グループ毎の分析を行うものであり，実質的にはダミー・モデルといえる。

1. 2. モデルの階層構造

回帰係数の推移から明らかなように，これらのモデルは階層構造をなしている。次にこのことを，平方和の分解で示そう。

ただし，n_j：第j群のサンプル数，$i=1, \cdots, n_j$，$n=\sum n_j$：全サンプル数，p：補助変数の数$l=1, \cdots, q$，$q=p+1$：$X_l=1$をダミー変数として加えた全補助変数の数，K：グループ数，$j=1, \cdots, K$，$Y=X\beta_o$ or $X\beta_j$，添字o：グループ間で共通，添字j：グループ間で異なる値とし，また，モデルを示す添字の意味するところは以下の通りである。A：Anova，G：Total Regression，W：Within Regression，I：Individual Regression，R：Ressidual，例えばAR：Residual of ANOVA，WA：difference between Within Regression and Anova 等。

階層モデル

1. $H_0 \subset H_1$

H_0：$SS_T=\sum_{ij}(Y_{ij}-\bar{Y})^2 \quad df_T=n-1$

H_1：$SS_T=SS_{AR}+SS_A$ ····· (1)

$\sum_{ij}(Y_{ij}-\bar{Y})^2=\sum_{ij}(Y_{ij}-\bar{Y}_j)^2+\sum_{ij}(\bar{Y}_j-\bar{Y})^2$

$df_T=n-1 \qquad df_{AR}=n-K \qquad df_A=K-1$

2. $H_0 \subset H_2$

H_2：$SS_T=SS_{GR}+SS_G$ ····· (5)

$\sum_{ij}(Y_{ij}-\bar{Y})^2=\sum(Y_{ij}-\hat{Y}^G{}_{ij})^2+\sum(\hat{Y}^G{}_{ij}-\bar{Y})^2$

$df_T=n-1 \qquad df_{GR}=n-q \qquad df_G=q-1$

3. $H_1 \subset H_3$

H_3：$SS_T=SS_{WR}+SS_W$ ····· (3)

$\sum_{ij}(Y_{ij}-\bar{Y})^2=\sum(Y_{ij}-\hat{Y}^W{}_{ij})^2+\sum(\hat{Y}^W{}_{ij}-\bar{Y})^2$

$df_T=n-1 \quad df_{WR}=n-(q-r-1)-K \quad df_W=(q-r-1)+K-1$

1：$SS_W=SS_{WA}+SS_A$

$\sum_{ij}(\hat{Y}^W{}_{ij}-\bar{Y})^2=\sum(\hat{Y}^W{}_{ij}-\hat{Y}^A{}_{ij})^2+\sum(\hat{Y}^A{}_{ij}-\bar{Y})^2$

$df_W=(q-r-1)+K-1 \quad df_{WA}=q-r-1 \qquad df_A=K-1$

2：$SS_{AR}=SS_{WR}+SS_{WA}$ ····· (2)

$\sum_{ij}(Y_{ij}-\hat{Y}^A{}_{ij})^2=\sum(Y_{ij}-\hat{Y}^W{}_{ij})^2+\sum(\hat{Y}^W{}_{ij}-\hat{Y}^A{}_{ij})^2$

$df_{AR}=n-K \quad df_{WR}=n-(q-r-1)-K \quad df_{WA}=q-r-1$

4. $H_2 \subset H_3$

1：$SS_W=SS_{WG}+SS_G$ ····· (7)

$$\sum_{ij}(\hat{Y}^W_{ij}-\bar{Y})^2=\sum(\hat{Y}^W_{ij}-\hat{Y}^C_{ij})^2+\sum(\hat{Y}^C_{ij}-\bar{Y})^2$$

$$df_W=(q-r-1)+K-1 \quad df_{WC}=K-r-1 \quad df_C=q-1$$

2 : $SS_{GR}=SS_{WR}+SS_{WG}$ ····· (6)

$$\sum_{ij}(Y_{ij}-\hat{Y}^C_{ij})^2=\sum(Y_{ij}-\hat{Y}^W_{ij})^2+\sum(\hat{Y}^W_{ij}-\hat{Y}^C_{ij})^2$$

$$df_{GR}=n-q \quad df_{WR}=n-(q-r-1)-K \quad df_{WG}=K-r-1$$

5. $H_3 \subset H_4$

H_4 : $SS_T=SS_{IR}+SS_I$ ····· (4)

$$\sum_{ij}(Y_{ij}-\bar{Y})^2=\sum(Y^{ij}-\hat{Y}^I_{ij})^2+\sum(\hat{Y}^I_{ij}-\bar{Y})^2$$

$$df_T=n-1 \quad df_{IR}=n-Kq \quad df_I=Kq-1$$

1 : $SS_I=SS_{IW}+SS_W$ ····· (9)

$$\sum_{ij}(\hat{Y}^I_{ij}-\bar{Y})^2=\sum(\hat{Y}^I_{ij}-\hat{Y}^W_{ij})^2+\sum(\hat{Y}^W_{ij}-\bar{Y})^2$$

$$df_I=Kq-1 \quad df_{IW}=(q-1)(K-1)+r \quad df_W=(q-r-1)+K-1$$

2 : $SS_{WR}=SS_{IR}+SS_{IW}$ ····· (8)

$$\sum_{ij}(Y_{ij}-\hat{Y}^W_{ij})^2=\sum(Y_{ij}-\hat{Y}^I_{ij})^2+\sum(\hat{Y}^I_{ij}-\hat{Y}^W_{ij})^2$$

$$df_{WR}=n-(q-r-1)-K \quad df_{IR}=n-Kq \quad df_{IW}=(q-1)(K-1)+r$$

これらの関係を図示したものが，Fig.1，Fig.2 であり，(1) ～ (9) は図中に対応している。

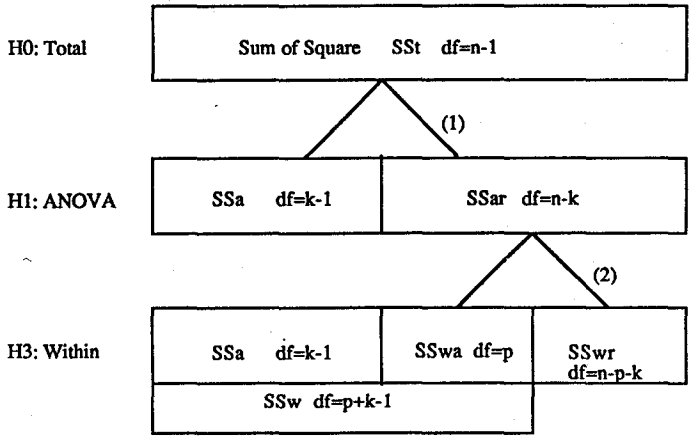

Fig. 1 Hierarchy of ANCOVA Models (1)

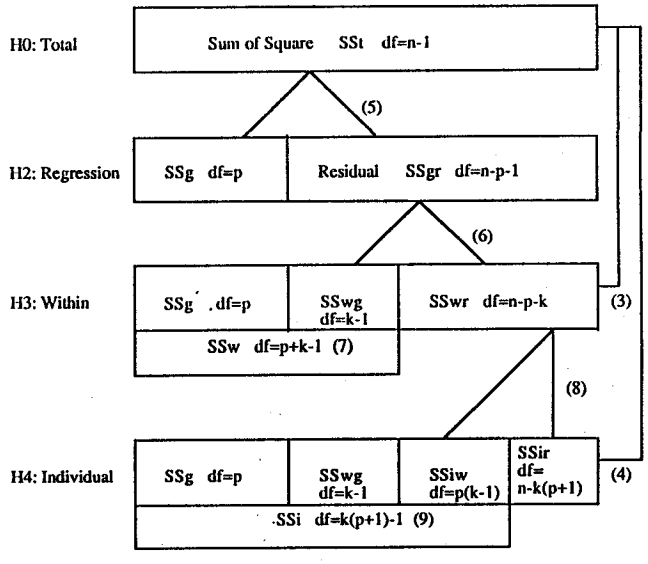

Fig. 2 Hierarchy of ANCOVA Models (2)

1. 3. 解法

各モデルのもとでの解は次のごとく与えられる。

[モデル0] $H_0: \hat{\alpha}_0 = \bar{Y}$

補助変数－目的変数間の相関が見られないので，回帰を用いず，目的変数のみで分析を行なう。また，オブザベーション・グループ間の差も見られないので，全データを1グループとみなし，平均値により推定を行なう。Yについての全平均を求めればよい。

[モデル1] $H_1: \hat{\alpha}_j = \bar{Y}_j$

補助変数－目的変数間の相関が見られないので，回帰を用いず，目的変数のみで分析を行なうが，オブザベーション・グループ間には差があり，グループ毎に平均値により推定を行なう。Yを用いる一元配置の分散分析モデルとして解けばよい。

[モデル2] $H_2: \hat{\boldsymbol{\beta}}_0 = (\boldsymbol{X}_T'\boldsymbol{X}_T)^{-1}\boldsymbol{X}_T'\boldsymbol{Y}$

補助変数－目的変数間の相関を利用して重回帰を用いて分析を行なう。ただし，オブザベーション・グループ間の差異を仮定せず，全データをプールして分析を行なう。通常の回帰分析と同様であり，共通の回帰係数を得る（Total Regression)。計画行列 $\boldsymbol{X}_T$ の次数は(n×q）である。ただし，第1列はすべて $X_1 = 1$ とし定数項を与えるダミー変数とする。

[モデル3]　$H_3 : \hat{\beta}_{0J} = (X'_W X_W)^{-1} X'_W Y$

補助変数－目的変数間の相関を利用して重回帰を用いて分析を行なう。ただし，オブザベーション・グループ間の差異を取り上げq個の補助変数中r個をグループ間で共通（β_0 = const），q－r個をグループ間で相違する（β_1），と仮定して，グループ毎の分析を行なう（Within Group Regression）。ダミー変数（定数項）は$X_1 = 1$としてもよいし，$X_{r+1} = 1$としてもよい。通常の共分散分析の場合は前者であるが，後者の場合に拡張することも可能である。

βはr個が共通，q－r個がグループによって異なるので，共通部分を先に並べ

$$\beta' = (\beta_{01}, \cdots, \beta_{0r}, \beta_{1,r+1}, \cdots, \beta_{1q}, \beta_{2,r+1}, \cdots, \beta_{2q}, \cdots, \beta_{K,r+1}, \cdots, \beta_{Kq})$$

と$r+k$ $(q-r)$次のベクトルとなる。このとき，計画行列X_Wは

$$1, \cdots, r, r+1, \cdots q, q+1, \cdots 2q-r, \cdots \quad \cdots$$

$$X_W = \begin{matrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ X0 & & & \\ & \cdots\cdots & & \\ 0 & 0 & & X_K \end{matrix}$$

とすればよく，ここでX_0(n×r)は各群に共通のr個の補助変数，X_Jは相互に異なる第J群のq－r個の補助変数（$n_j \times (q-r)$），0はゼロ行列である。

[モデル4]　$H_4 : \beta_J = (X_J'X_J)^{-1} X_J'Y_J \qquad j=1, \cdots, K$

補助変数－目的変数間の相関を利用して重回帰を用いて分析を行なう。ただし，全補助変数に関してオブザベーション・グループ間の差異を仮定し，従ってグループ毎に重回帰分析を行なうことに帰着する。データをグループ毎に分割し，個別に重回帰分析を反復して行なえばよい（Individual Regression）。X_Jは第J群の補助変数，Y_jは目的変数である。

所与のデータを上記各モデルのもとで，計画行列Xを選ぶことにより一般線型モデル（GLM）を用いて一般的に解くことができる。そして，どのモデルが最もよく適合するかの検討は次の2つの方法が可能である。

(1) 偏回帰係数の検定による方法：

$H_0 : \beta_l = 0$　　第l変量の係数が0かどうか

$H_0 : \beta_j = \beta_{jj}$　　2群 J, JJ の係数が同じかどうか

の検定結果を総合して，最適モデルを探索する。

(2) AIC（赤池）による方法：モデルの適合性を検討する方法として

$$AIC = -2log\,(\text{最大尤度}) + 2\,(\text{パラメータ数})$$

を計算し，$AIC = min$ とするモデルを採用する方法は広く用いられている。

選択されたモデルの下での，補助変数の影響を取り去った後の残差は

$$Y_{residual} = Y - \hat{Y} = Y - X\hat{\beta}$$

であり，これを用いて分散分析を行なえばよい。

2. ファジイ重回帰分析

2. 1. ファジィ理論

Zadeh, L. A. の Fuzzy Set に関する論文 (1965)[20] 以来，ファジイ理論はあいまいな概念を取り扱う理論として，今日まで着実な発展をみてきた。当初は純粋の数学的集合理論として展開されたが，現在では数学のみならず科学の様々な領域に浸透し，発展・体系化を見るに至っている。中でも工学の分野における応用はめざましく，新しい制御理論として実用化され，またいくつかの商品としても開発されている。同理論の特徴は，集合をメンバーシップ関数

$$\mu_A : X \rightarrow [0, 1]$$

によって特性づけられた集合とし，要素 $x \in X$ に対する値 $\mu_A(x)$ によって，集合の要素 x がファジイ集合 A に属する度合を表わし，ファジイ集合

$$A = \sum_i \mu_A(x_i) / x_i$$

に関する演算を行なうことである。

そして，区間 $[0, 1]$ を $\{0, 1\}$ とすれば，通常の集合（属するか，属さないかの2値しかとらない）に帰するため，ファジイ集合は通常の集合の拡張として展開されている。本稿で以下に取り上げるファジイ統計解析も，通常の統計解析の拡張として位置づけることができる。

メンバーシップ関数は任意の意味における集合への帰属を表わす測度であり，確率の公理をすべて満たす必要はない。確率は事象の生起のあいまいさを表わすのに対し，ファジイ理

論はメンバーシップ関数を用いて事象ないしは概念の「あいまいさ」を定量的に操作するところに，その特徴があり，両者は本質的に異なっている。

通常の集合に見られる基本性質，例えば反対称律，推移律，交換律，結合律，分配律等はファジイ集合においても成立するし，通常の集合における主要演算，例えば相等，包含，補集合，和集合，共通集合，代数積，代数和，等はファジィ集合でも成立する。また，本稿で取り扱うメンバーシップ関数を行列表示した，ファジイ行列についても，通常の行列に見られる演算や基本性質，例えば，相等，包含，和，積，行列積，行列式，余因子行列，逆行列等は一般の行列と同様に成立し，同様の演算が可能であることも確かめられている[5,6,11,18]。

ファジイ理論は「あいまいさ」を扱うものの，決して理論そのものは曖昧なものではなく，数学的に厳密に展開された一大理論体系である。

2. 2. ファジイ・グループ

そうしたファジィ理論を用いて共分散分析をさらに拡張し，ファジイ重回帰分析（FRA：Fuzzy Regression Analysis）として解くことを試みる。

重回帰式：$Y=X\beta$ の β にファジイを導入し，区間回帰分析として解いた研究[10,14,15,16,17]は多いが，ここではファジイの概念をサンプルのグループに導入し，ファジイ・グループを想定した回帰分析[9]として解くこととする。すなわち，共分散分析を多群の重回帰分析として解くとき，サンプルのグループを明白（crisp）な属性によるグループとしてではなく，ファジイ・グループとして用いる。

ファジイ・グループとは，Jajuga, K.[9]によれば，『等質ではあるが，しかしそれらは相互に必ずしも明瞭に分離できない，いくつかの下位集合からなる観測値の集合』と定義されている。例えば，社会調査における人口学的属性として，「男」・「女」は明瞭なグループであるが，調査項目に対する回答のパターンでみるとき，男性が女性に多く見られるパターンで回答する場合もあるであろうし，その逆に男性に特徴的な回答を多く示す女性のサンプルもあろう。こうした中間のグループを回答パターンからみるファジイ・グループとして取り出すことができれば，男，女に2分割するよりも，中間に〈男女〉のグループを導入し，ファジイ・グループとして処理することにより，データの特徴をくみ取った分析が可能となる。状態を相互に排反な2つのグループに分割するのではなく，中間の移行型を設けることにより，属性の変化を連続的に捉えることができる。そうした操作を必要としない場合には，メンバーシップ関数を0または1として選択すればよい。このとき通常の分析と同じ結果が得られる。

こうしたファジイ・グループの判別は，データにクラスター・アナリシスを施し，樹形図により，クラスター（グループ）形成の過程から行なうこともできるし，ファジイ判別分

析[19]を適用してファジイ・グループを抽出することも可能である。

2. 3. ファジイ統計量

ファジイ理論を統計学に導入するとき，必要となるのがメンバーシップ関数をどう導入して演算を行なうか，すなわち，ファジイ統計量をどう定義するか，である。すでに報告されているいくつかの例[1, 11, 18]をもとに，ここでは次のごとく導入する。

観測値 O_i のファジイ・グループ G_j へのメンバーシップ関数を f_{ij} とし，G_j における各種標本統計量をメンバーシップ関数をウェイトとして，次のごとく定義する。ただし，メンバーシップ関数に関しては確率や一般のウェイトのように

$$\sum_i f_{ij} = 1, \quad \sum_j f_{ij} = 1$$

等の束縛はない。また，メンバーシップ関数を0または1と設定すれば，通常の統計量に一致する。ファジイ統計量の記号として～を付す場合もあるが，以下本稿では統計処理としてはファジイ統計処理を指すので，特に誤解のない限り煩雑さを避け，一般の統計量と同じ記号を用いることとする。

ファジイ・サイズ　$n_j = \sum_i f_{ij}$

ファジイ平均　$\bar{X}_j = \sum_i X_i f_{ij} / n_j$

ファジイ分散　$S^2_j = \sum_i (X_i - \bar{X}_j)^2 f_{ij} / n_j$

K－次のファジイ積率　$m^k_j = \sum_i (X_i - \bar{X}_j)^k f_{ij} / n_j$

ファジイ平方和分解　$SS_T = SS_W + SS_B$

$$\sum_{ij} (X_{ij} - \bar{X})^2 f_{ij} = \sum (X_{ij} - \bar{X}_j)^2 f_{ij} + \sum (\bar{X}_j - \bar{X})^2 f_{ij}$$

2. 4. ファジイ重回帰分析

共分散分析におけるグループは明瞭なインデックスによって区別されたcrisp setであるが，ファジイ・グループのもとでの重回帰分析はファジイ統計量を用いて，次のごとくに解くことができる。

オブザベーション O_i のファジイ・グループ G_j へのメンバーシップ関数を対角行列

$$\boldsymbol{F}_j (n \times n) = \mathrm{diag}\ \{f_{ij}\}$$

で表わし，これを補助変数に対するウェイトとして用いれば，yに対する推定値 $\hat{y}$ は

$$y_j = \sum_L x_{iL} f_{ij} \beta_{jL} = \boldsymbol{F}_j \boldsymbol{X} \boldsymbol{\beta}_j$$

で求められる。このとき，データyとの残差平方和は

$$Q(\beta_j) = \sum_i (y_i - \hat{y}_i)^2$$

$$= \sum_i (y_i - \sum_L x_{iL} f_{ij} \beta_{jL})^2 = \sum_i (y_i - f_j x \beta_j)^2$$
$$= (\boldsymbol{y} - \boldsymbol{F_j X \beta_j})' (\boldsymbol{y} - \boldsymbol{F_j X \beta_j})$$

となり，$Q(\beta_j) = min$ とする β を解として，求めればよい。解は

$$\partial Q(\beta_j) / \partial \beta_j = 0$$

より，正規方程式

$$(\boldsymbol{X'F_j'F_j X})\beta_j = \boldsymbol{XF_j y}$$

を解き，β_j の最小自乗解として

$$\beta_j = (\boldsymbol{X'F_j'F_j X})^{-1} \boldsymbol{X'F_j Y}$$

を得る。

そして，これをファジイ・グループ $j=1, \cdots, K$ について解いた解を記述の便より

$$\hat{B} = (\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_K)$$

としておく。また，いま

$$\boldsymbol{Q_j} = \boldsymbol{F_j'X} (\boldsymbol{X'F_j'F_j X})^{-1} \boldsymbol{X'F_j}$$

とおくと，容易に

$$\boldsymbol{Q'Q = Q}, \quad \boldsymbol{Q^2 = QQ = Q}$$

で示され，$\boldsymbol{Q}$ はベキ等行列である。これを用いれば，$\hat{\boldsymbol{y}}_j$ は

$$\boldsymbol{y_j} = \boldsymbol{F_j X \hat{\beta}_j}$$
$$= \boldsymbol{F_j X} (\boldsymbol{X'F_j'F_j X})^{-1} \boldsymbol{X'F_j y} = \boldsymbol{Q_j y}$$

として求められ，またこのとき，残差平方和を R_{0j}^2 とすれば

$$R_{0j}^2 = \sum (y - \hat{y}_j)^2$$
$$= (\boldsymbol{y} - \boldsymbol{Q_j y})'(\boldsymbol{y} - \boldsymbol{Q_j y}) = \boldsymbol{y'}(\boldsymbol{I} - \boldsymbol{Q_j})\boldsymbol{y}$$

となり，これもQを用いて容易に求めることができる。

残差に関して

$$\boldsymbol{y} - \hat{\boldsymbol{y}}_j = \boldsymbol{d_j} \sim N_n(0, \ \sigma^2 I_n)$$

と，正規性の仮定を導入すると（正規モデル），

$$\boldsymbol{y} = \hat{\boldsymbol{y}}_j + \boldsymbol{d_j} \sim N_n(y_j, \ \sigma^2 I_n)$$

となり，従って，yの線型結合

$$\hat{\beta}_j = (\boldsymbol{X'F_j'F_j X})^{-1} \boldsymbol{X'F_j'y} = \boldsymbol{C_j y}$$

もまた正規変数である。このとき，β_j の平均，分散は

$$\boldsymbol{E}(\hat{\beta}_j) = \boldsymbol{C_j E}(\boldsymbol{y}) = \boldsymbol{C_j \hat{y}_j}$$

$$=C_j F_j X\beta_j = \beta_j$$

$$V(\hat{\beta}_j) = C_j' V(y) C_j = C_j' C_j \sigma^2 I$$

$$= (X'F_j'F_j X)^{-1} \sigma^2 I$$

である。

以上のごとく，ファジイ重回帰分析の解法は通常の重回帰分析と同様に展開することができた。さらに，β_jの区間推定，検定に関しても，同様にして，ベキ等行列Qを用いて，次のごとく導出することができる。

β_jにおける，任意のウェイトをw（第L要素が1，他は0）

$$w' = (0, 0, \cdots, 1, \cdots)$$

とすれば，上述のごとく，$\hat{\beta}_j$が正規変数であったので$w'\hat{\beta}_j$もまた正規変数であり，平均，分散は次のごとくである。

$$E(w'\hat{\beta}_j) = w'E(\hat{\beta}_j) = w'\beta_j$$

$$V(w'\hat{\beta}_j) = w'V(\hat{\beta}_j)\, w = w'\,(X'F_j'F_j X)^{-1} w \sigma^2$$

従って，これを標準化した

$$Z = (w'\hat{\beta}_j - E(w'\hat{\beta}_j)) \Big/ \sqrt{(V(w'\hat{\beta}_j))}$$

は標準正規分布に従い，これとは独立に残差平方和による統計量R_0^2 / σ^2は自由度$df = \nu$のχ^2－分布に従うので，両者の比は自由度$df = \nu$のt－分布に従う。νの値はそれぞれのモデルにより異なる。すなわち，これを用いて，β_jにおける第l要素の偏回帰係数の検定（$H_0 : \beta_{jl} = \beta_0$）と，区間推定を

$$t = (\beta_{jl} - \beta_0) \Big/ \sqrt{(R_0^2 \times S^{-1}{}_{ll} / \nu)}$$

を用いて行なうことができる。ただし，$S^{-1}{}_{ll}$は逆行列$(X'F'FX)^{-1}$における第ll要素，dfは回帰分析のモデルにより異なる。

さらに残差R_0^2を用いて，AICは次のごとく与えられる。

$$AIC = n\log(R_0^2/n) + 2t = n\log(S_y^2(1 - R_m^2)) + 2t$$

ただし，R_m：重相関係数，t：有効な補助変数の数。tの値はそれぞれのモデルにより異なる。

こうした重回帰分析の特徴は以下の通りである。

オブザベーションの属するグループの属性に関して，共分散分析の各モデル（H_0～H_4）下では各モデルは相互に独立に処理されるが，本ファジィ重回帰分析の場合には，メンバーシップ関数を用いてグループへの帰属を連続的に導入し，全体として処理すると同時に，個

別化からの視点も無視することなく処理する，ということである。全体，群内，個別の各回帰分析を総合した，モデル間の中間に位置する解を得ることができる。もちろん，個別の処理が要求される場合にはメンバーシップ関数を0または1とすればよい。

一般的にいえば，集合の要素を集合に（帰属，非帰属）の二分法として取り扱う従来の集合論に比し，メンバーシップ関数を用いて中間を連続的に処理するファジィ理論は，統計学に導入されたとき，データの本性に忠実な統計学的処理法として特徴づけることができよう。多くのデータは連続的に変容する属性をある基準のもとにカテゴリに分類することで処理されるが，ファジィ理論はこうした強制から解放するものである。

そして，中間移行型を導入することにより，統計処理の精度は落ちるが，それと引き替えに，より忠実にデータの本性を抽出することが可能となり，いずれを選択するかは研究の目的に応じて決定すればよい。精度の低下は必ずしも本法の欠点とはならない。

3. コンピュータ・プログラム

以上の過程をTHINK Pascal（Macintosh版）により，コンピュータ・プログラミングを行った。プログラムは大きく（1）データの入力，（2）共分散分析，（3）ファジィ重回帰分析の処理部分に分けられている。データの入力はKB（キーボード）からでも，ファイルからでも可能である。変量数，オブザベーション数に次いで，データを入力し，あわせてオブザベーションの属するグループを与えれば，5つのモデル下での解が求められる。偏回帰係数の検定結果，AICの値をもとに最適モデルを検討すればよい。

ファジィ重回帰分析に際しては，メンバーシップ関数を事前に設定してもよいし，データから算出してもよい。以下の数値例では，メンバーシップ関数としてオブザベーション点からグループの重心までの距離を求め，それに逆比例する値を標準化して用いたが，他のアルゴリズムへの変更は容易である。

プログラムの妥当性のチェックとして，PC版SASによる出力[3,9,13]と比較し，偏回帰係数，平方和，残差ともに同様の結果が得られることを確認した。ファジィ重回帰については，メンバーシップ関数を0または1として与え，通常の重回帰分析の解と一致することを確認した。変数は単精度計算のため，反復計算による誤差の蓄積が心配されるが，倍精度への切り替え，誤差の検討，および結果のグラフィックス出力は今後の課題である。

4. 数値例

4. 1. 数値例1（共分散分析）

Snedecor & Cochran[12] の例がSASマニュアル[9]に用いられているので，プログラムおよび分析法のチェックデータとして取り上げる。2種類の薬を患者に投与し，投薬前後のスコアを比較して投薬の効果を調べたデータである。2種類の投薬群（A, D）に加えて他の1群（F）を統制群とするのでグループ数はあわせて3群。各群10名で総オブザベーション数は30。Xを投薬前のスコア，Yを投薬後のスコアとする。Fig.3に散布図を示す。各群別に見ると，A（●），D（■）群が線型の散布を示し回帰が有効のように思える。相関係数を算出してみると，r_A=0.7642，r_D=0.9114 となり，高い相関が示されている。これに比し統制群F（○）はややばらつきが大いようであるが，相関係数はr_F=0.6610で単調増大の傾向を示しており，各群ともにXを補助変数としてその効果を除去しての分析が可能のようである。

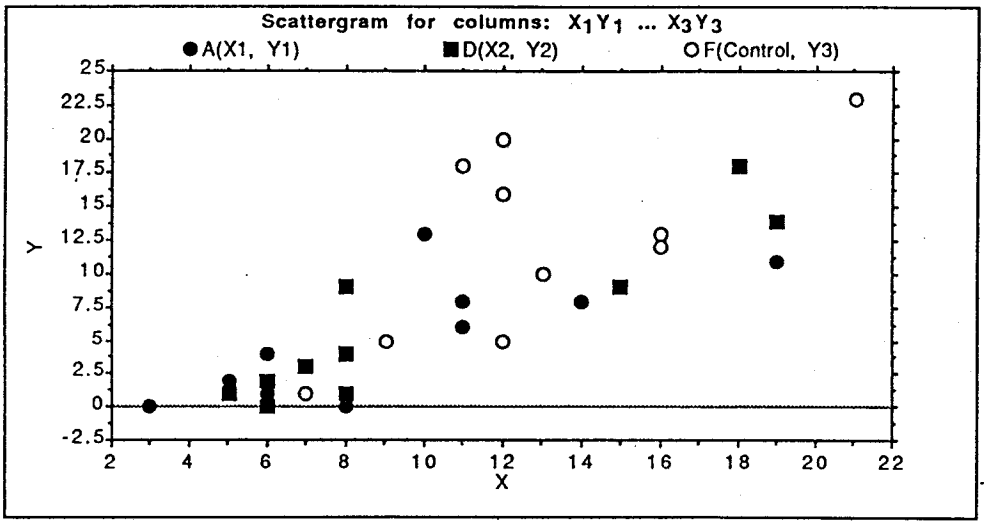

Fig. 3 Input Data of Example 1

Tab. 1 数値例1・ANCOVA Table（PC／SASによる処理）

Source		*SS* (*Typle* I)	*DF*	*MS*	*F*−*Value*	*Pr*
Model (H_3)	SS_W	871.497	3	290.499	18.10	Pr<0.0001
処理 (Drug)	SS_A	293.600	2	146.800	9.15	0.0010
回帰 (X)	SS_{WA}	577.897	1	577.897	36.01	0.0001
Residual	SS_{WR}	417.203	26	16.046		
Total	SS_T	1288.700	29			

SAS による分析では Tab. 1 のごとく，回帰によるモデルは有効で（$F_{Model}=18.10$, $Pr<0.001$），平方和を薬と X に分解しても共に有意であった（$F_{Drug}=9.15$, $Pr=0.01$, $F_X=36.01$, $Pr=0.0001$）。X を除去する以前から *3* 群は有意であるが，補助変数を用いて処理することによりさらに誤差を減じて，

$$Y^*_{ij}=Y_{ij}-X_{ij}\beta$$

による修正後の平方和，平均平方を用いることにより，より詳細な分析を行なうことができる。

SAS で共分散分析を行う場合には，GLM（一般線型モデル）プロシジアで

MODEL Y=A X

と指定すればよく，オプションで分析方法を選択することはできるが，モデルは H_3 の 1 通りの分析しかなく，前述の H_0～H_4 のモデルの選択肢はない。しかし，解かれた偏回帰係数から，

H_2 :	$Y=0.9872X-0.4347$	*All*
H_3 :	$Y=0.9872X-3.4461$	*A*
	$Y=0.9872X-3.3371$	*D*
	$Y=0.9872X-0$	*F*

が推定され，結果的に H_2, H_3 の 2 モデルについて解かれていることがわかる。予測による各群毎の Y の平均値は

$\bar{Y}_A=5.300$, $\bar{Y}_D=6.100$, $\bar{Y}_F=12.300$ 修正前
$\hat{\bar{Y}}_A=6.715$, $\hat{\bar{Y}}_D=6.824$, $\hat{\bar{Y}}_F=10.161$ 修正後

となり，2 種類の薬服用群と統制群との差が縮まるが，服用前の X の如何に拘らず，なお薬効が顕著のようである。

同データを H_1～H_4 の 4 種類のモデルについて解いた結果が Tab. 2 である。Y だけで分類を行う H_1 も有意であり，他の 3 種の回帰モデルもともに有意である。階層構造に従って残差平方和を分解したものを各モデルの下に追加したが，(1) は分散分析との関連，(2) において Total Regression から Within Regression にすることによる平方和の増大が $SS_{WG}=68.554$，さらに (3) において Within Regression を Individual Regression にすることにより，$SS_{IW}=19.645$ の平方和を取り出せること，等がわかる。ただし，本データの場合はいずれも有意差を示すほど大きくはない。これらをまとめたものが Tab. 3 である。

各モデルにおける偏回帰係数および重相関係数はTab. 4である。H_3における第3群（F）の切片，H_4における第1群（A）の勾配，切片が他と異なっている。AICからみるとH_3の場合が最小で最適モデルであることを示している。H_4はパラメタ数が多いため計算の負担が大きく，モデルとしては望ましくない。

Tab. 2 数値例1・共分散分析結果（モデル別）

H_1 : ANOVAMODEL $Y=\alpha_j$					
Sum of Square		*DF*	*MS*	*F-Value*	
SS_A	293.600	2	146.800	3.983	*
SS_{AR}	995.100	27	36.856	—	
SS_T	1288.700	29	—	—	
(1) SS_{WA}	577.897	1	577.897	36.015	**
SS_{WR}	417.203	26	16.046	—	
SS_{AR}	995.100	27	—	—	
H_2 : Total Regression $Y=\alpha_0+\beta_0$					
SS_G	802.99	1	802.944	46.283	**
SS_{GR}	485.756	28	17.348	—	
(2) SS_{WG}	68.554	2	34.277	2.136	
SS_{WR}	417.203	26	16.046	—	
SS_{GR}	485.756	28	—	—	
H_3 : Within Regression $Y=\alpha_j+X\beta_0$					
SS_W	871.497	3	290.499	18.104	**
SS_{WR}	417.203	26	16.046	—	
(3) SS_{IW}	19.645	2	9.822	0.593	
SS_{IR}	397.558	24	16.565	—	
SS_{WR}	417.203	26	—	—	
H_4 : Individual Regression $Y=\alpha_j+X\beta_0$					
SS_I	891.142	5	178.228	10.759	**
SS_{IR}	397.558	24	16.565	—	

Tab. 3 数値例1・各種回帰による平方和の追加

Source	*SS*	*DF*	*MS*	*F-Value*
全回帰による SS_G	802.944	1	802.944	48.472
群内回帰による追加 SS_{WG}	68.554	2	34.277	2.136
個別回帰による追加 SS_{IW}	19.645	2	9.822	0.593
Residual SS_{IR}	397.558	24	16.565	—
Total SS_T	1288.700	29	—	—

Tab. 4 数値例1・重回帰分析結果

	H_2	H_3	H_4			Fuzzy Regression		
β_1	1.098	0.987	0.745	1.069	1.194	1.133	1.122	1.025
α	−3.886	−3.881	−1.631	−4.585	−3.109	−4.634	−3.705	−2.978
		−3.772						
		−0.435						
R_m	0.7893	0.8224	0.7642	0.9114	0.6610	0.7893	0.7893	0.7893
AIC	87.54	86.97	127.975			88.327		

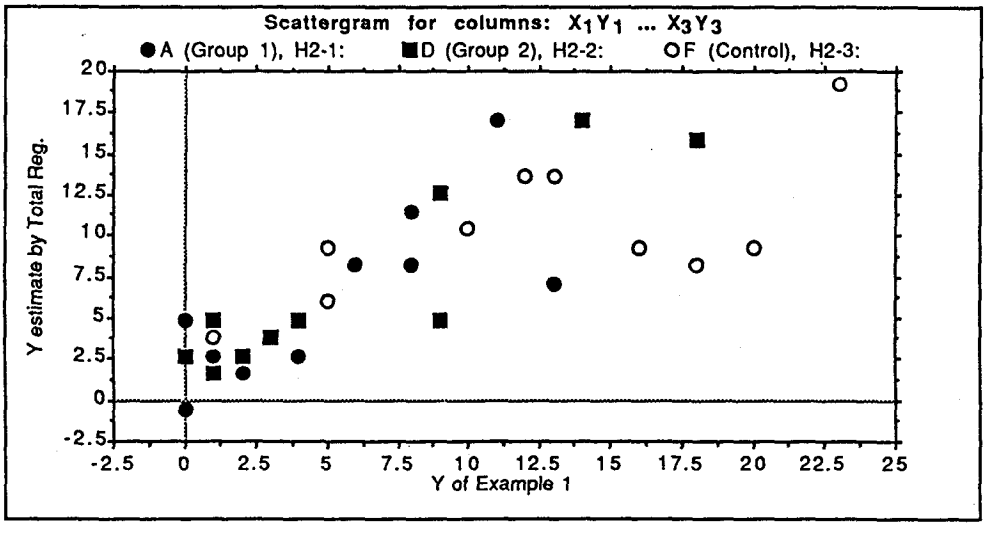

Fig. 4 Resuts of H2 Model (Total Regression)

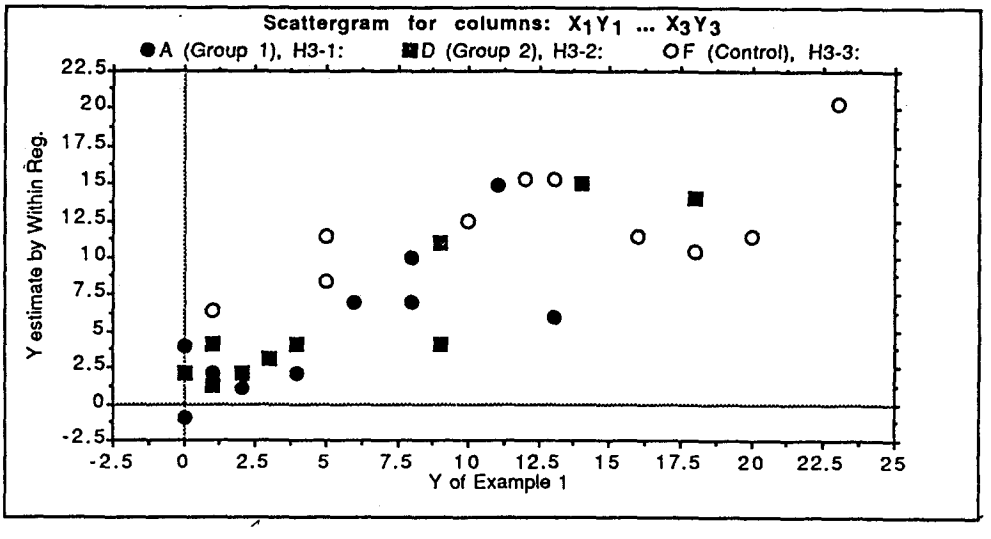

Fig. 5 Resuts of H3 Model (Withinl Regression)

H_2，H_3 モデルのもとでの Y の推定値をもとめ，Y との散布図を描いたものが Fig. 4, Fig. 5 である。$Y_{adj}=Y$ の 45° の直線からの偏差が残差であり，H_3 モデルが残差が少なく，補助変数により，より多く部分が説明され除去されていることがわかる。そして，統制群の Y の値の大きい部分における 3 サンプルのずれが大きく，これらが回帰を攪乱している。

4. 2. 数値例 2（ファジイ重回帰分析）

人工的データであるが，ファジイ重回帰分析の一つの適用例として以下のデータの分析を行なう。

説明変量数=2，グループ数=3 の Raw Data を用いる。先ず，共分散分析を行なった結果は Tab. 5 であり，回帰分析の結果は Tab. 6 である。各種モデル下での回帰による平方和は H_2，H_3 が有意で，H_1，H_4 は有意ではなかった。この意味するところは，データは全体とし

Tab. 5　数値例 2・共分散分析結果（モデル別）

H_1 : ANOVAMODEL　$Y=\alpha_j$					
Sum of Square		*DF*	*MS*	*F-Value*	
SS_A	11.738	2	5.869	1.773	
SS_{AR}	56.262	17	3.310	—	
SS_T	68.000	19	—	—	
H_2 : Total　Regression　$Y=\alpha_0+X\beta_0$					
SS_G	27.368	2	13.684	5.726	*
SS_{GR}	40.632	17	2.390	—	
H_3 : Within　Regression　$Y=\alpha_j+X\beta_0$					
SS_W	37.780	4	9.445	4.687	*
SS_{WR}	30.220	15	2.015	—	
H_4 : Individual　Regression　$Y=\alpha_j+X\beta_j$					
SS_I	44.114	8	5.514	2.540	
SS_{IR}	23.886	11	2.171	—	

Tab. 6　数値例 2・重回帰分析結果

	H_2	H_3	H_4			Fuzzy Regression		
β_1	0.311	0.650	0.862	−0.815	0.892	0.308	0.307	0.391
β_2	0.524	0.819	1.034	0.193	0.895	0.469	0.491	0.664
α	2.627	1.740	0.741	8.561	−2.445	2.945	2.700	1.589
		−.999						
		−.545						
R_m	0.6344	0.7454	0.6298	0.5680	0.9259	0.6341	0.6343	0.6344
AIC	20.18	18.26	36.095			17.954		

て一群をなしているとみなすこともできるし，回帰の勾配を同一にして解いてもよい，ということである。各群でデータがもう少し異なっていれば個別回帰の解が有意となったのであろう。本データは各群相互に類似しており，個別回帰で解かねばならないほど違ってはいない，ということを示している。このことは*AIC*（Tab. 6）で見ても，H_2，H_3で小さく，H_4で急激に増大していることにも反映されている。

同様のことは，Tab. 6の各モデル下での回帰係数からもいえる。H_2，H_3下におけるβ_1，β_2は類似しているが，H_4下では第2群の値が他とやや離れている。そして，データの変動が大で重相関係数も小さく，他の2群とは異なった値の回帰係数である。

ファジイ重回帰分析として解いた結果は同表の右欄である。切片，回帰係数ともに類似の解が得られ，H_2による値に近い。全体を見渡しながらも個別群の特徴をも考慮しつつ解いており，全体群と個別群の中間に位置する，ファジイなデータ処理が行なわれていることがわかる。

このことはファジイ重回帰分析の特徴であって，共分散分析のモデルH_2，H_4は相互に関連はなく，それぞれ独立に解かれるが，ファジイ重回帰の場合には，両者の中間にあって，全体解と個別解の両者の特徴を保持しつつ解くこととなる。すなわち，各群をファジイ・グループとみなし，相互の連関のうちに，全データを使用して解く解である。そして，メンバーシップ関数の選び方により，全体群にも個別群にも近づけた解を得ることができる。

本データの場合，共分散分析の結果からでは個別回帰における第2群の回帰係数が異なっていたが，ファジイ重回帰の結果では，第2群よりも第3群が他と離れているという結果が得られた。このことはデータの散布図からもみてもその傾向を見ることができる。H_4個別回帰では，少ないサンプル数で解くため，例えば特異値（outlier）等の影響を受けやすいが，ファジイ重回帰では常に全データを用いるので，そうした問題点からは免れている。

ファジイ重回帰分析の重相関係数はR_m＝0.6341～0.6344で，必ずしも共分散分析に比し向上するものではない。H_3で0.7454，H_4の第3群で0.9259を達成しているが，このような高い値ではなく，H_2の0.6344に近い値である。これはファジイ重回帰が全群のデータを同時に使用して解くためであり，むしろ分析の妥当性を示す数字と思われる。グループの境界をあいまいとして解くファジイ重回帰分析の特徴がここにも反映されている。

2つの数値例について，データ（Y）とファジイ回帰による推定値（$Y_{estimate}$）の散布図がFig. 6（数値例1），Fig. 7（数値例2）である。45°の直線からの偏差が予測のずれであり，Fig. 6ではFのコントロール群において，Fig. 7では第1，2群においてずれが大きいが，これは用いられた数値例に基づくものであって，分析法に根ざすものではない。

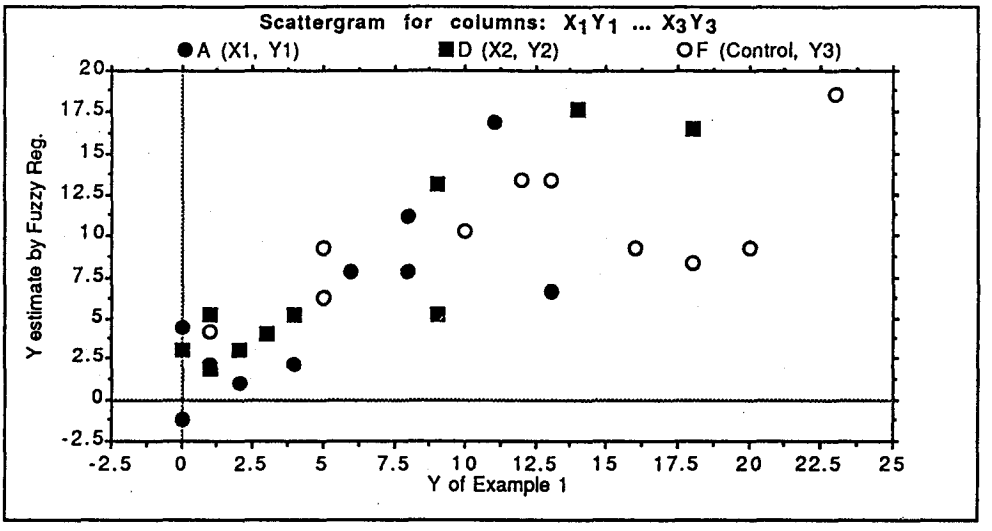

Fig. 6 Results of Fuzzy Regression Analyasis (Example 1)

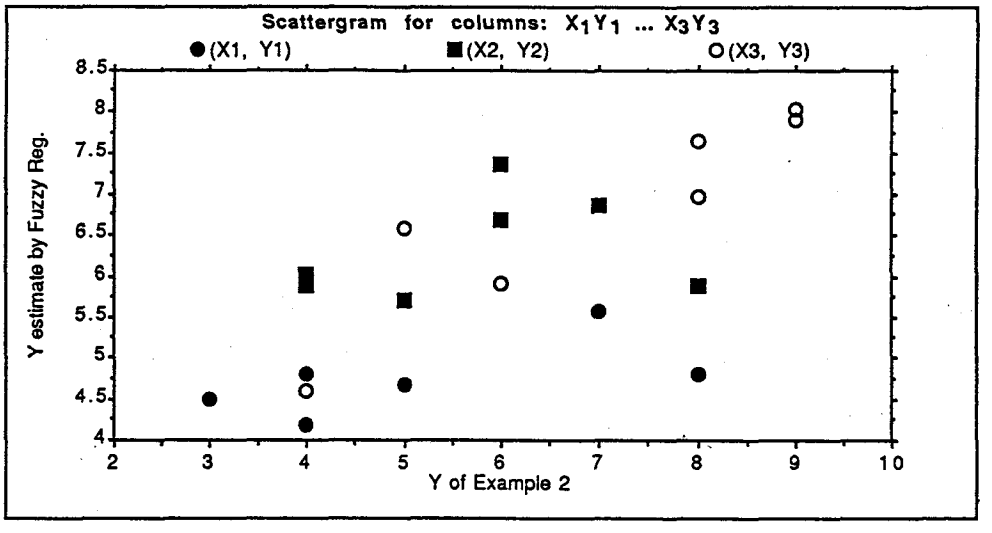

Fig. 7 Results of Fuzzy Regression Analyasis (Example 2)

5. まとめ

共分散分析は，随伴変数の影響を除去した後の目的変数の分析を行う統計学的手法であるが，オブザベーションをいくつかのグループに分類した場合の重回帰分析として理解することもできる。

グループ別における回帰係数のあり方から，いくつかのモデルに分類し，そのモデルのもとでの解法を一般線型モデルにより統一的に整理した。そして，計画行列をモデル別に修正して解くアルゴリズムを工夫し，コンピュータ・プログラミングを行った。

さらに，ファジィ理論を重回帰分析に適用し，ファジィ重回帰分析として展開した。ファジィの概念をオブザベーションのグループに適用してファジィ・グループとし，メンバーシップ関数を導入して，重回帰式を解く方法であり，メンバーシップ関数を0または1とすると，通常の重回帰分析となるため，本ファジィ重回帰分析は通常の重回帰分析の拡張といえる。

オブザベーションの属するグループの属性に関して，共分散分析のモデル下では各モデルは相互に独立に処理されるが，本ファジィ重回帰分析の特徴は全体として処理すると同時に，個別からの視点も無視することなく処理する，ということであり，モデル間の中間に位置する解を得ることができる。

集合要素を集合に帰属，非帰属として二分法的に取り扱う従来の集合論に比し，中間を連続的に処理するファジィ理論は，統計学に導入されたとき，データの本性に忠実な統計学的処理法として特徴づけることができよう。

文　献

1) Bardossy, A. Note on fuzzy regression, *Fuzzy Sets and Systems,* **37**, 65-75, 1990.
2) Dunn, O. J. and Clark, V. A. *Applied Statistics : Analysis of Variance and Regression,* John Wiley, 1974. 中村慶一訳, 応用統計学：分散分析と回帰分析, 森北出版, 1975.
3) Freund, R. J., Little, R. C. and Spector, P. C. *SAS System for Linear Model,* SAS Institute, 1986.
4) Jajuga, K. Linear Fuzzy Regression, *Fuzzy Sets and Systems,* **20**, 343-353, 1986.
5) Kaufmann, *A. Introduction to the Theory of Fuzzy Subsets,* Vol.I, Academic Press, 1975.
6) 水本雅晴, ファジイ理論とその応用, サイエンス社, 1988.
7) Morrison, D. F. *Applied Linear Statistical Model,* Prentice-Hall, 1983.
8) Pearce, S. C. Analysis of Covariance, in Ferber R. ed Encyclopedia of *Statistical Sciences,* Volume **1**, 61-69, John Wiley, 1982.
9) *SAS User's Guide : Statistics,* Version 5 Edition, SAS Institute, 1985.
同[日本語版], SAS ソフトウエア株式会社, 1990.
10) 坂和正敏, 矢野均, ファジィ入出力データに対する多目的ファジィ線形回帰分析, 日本ファジィ学会誌, **1**, 107-115, 1989.
11) 坂和正敏, ファジィ理論の基礎と応用, 森北出版, 1989.
12) Snedecor, G. W. & Cochran, W. G. Statistical method, 6th Edition, The Iowa State University Press, 1967. 畑村又好他訳, 統計的方法, 岩波書店, 1972.
13) 竹内啓監修, SAS による実験データの解析, 東京大学出版会, 1989.
14) 田中英夫, 林勲, 和多田淳三, 区間回帰分析, ファジィシステムシンポジウム論文集, **3**, 9-12, 1987.
15) Tanaka, H., Uejima, S. and Asai, K. Linear regression analysis with fuzzy model, *IEEE Trans. System Man Cybernet.* **12**, 903-907, 1982.
16) 田中英夫, 林勲, 和多田淳三, 区間回帰分析, ファジィシステムシンポジウム論文集, **3**, 9-13, 1987.
17) Tanaka, H., Hayashi, I. and Watada, J., Possibilistic linear regression analysis, *European Journal of Operation Research,* **40**, 389-396, 1989.
18) 寺野寿郎, 浅居喜代治, 菅野道夫共編, ファジィシステム入門, オーム社, 1987.
19) Yoshida, M. Fuzzy discriminant analysis on fuzzy groups, *Proceeding of 3rd IFSA Congress,* 759-762, 1989.
20) Zadeh, L. A. Fuzzy sets, *Information and Control,* **8**, 3, 338-353, 1965.

Analysis of Covariance and Fuzzy Regression Analysis

Mitsuo Yoshida

Analysis of Covariance (ANCOVA) is a statistical method for adjusting for any effect of other variables, named concomitant variables or covariates, using regression analysis. The components that affect on the response variable such as a general factor, effects among treatments or observation groups, amounts of concomitant variables and a random error which is normally distributed, are combined as a linear function and the third term of the concomitant variables is adjusted under the hypothesis that regression coefficients are equal among introduced treatments.

As an expansion of ANCOVA, five hierarchical models, Reduced, ANOVA, Total Regression, Within Regression and finally Individual Regression models, are proposed here according to varieties of regression coefficients and solved by design matrices of the general linear regression model (GLM) . Our interests are not the fixed effects as investigated by an intercept of the regression line, but the results of the multiple regression analysis asking which model shows the best fit to the given data.

Fuzzy Regression Analysis (FRA) is proposed as a further expansion of multiple regression analysis, which is an application of fuzzy set theory to regression analysis, introducing fuzziness to sample groups as fuzzy groups. In most statistical analysis, every observation always belongs to any crisp sample groups which are defined by demographic items, for example, male or female. Fuzzy group is defined as the set of observations that may be heterogeneous and may consist of several homogeneous subsets, but these subsets are not sufficiently separable.

By applying such fuzzy groups to regression analysis, we can obtain a fuzzy- and intermediate solution that would be obtained by not only total regression but individual regression in the above models. Membership functions are applicable either to any scores in advance or computed by the data. Numerical examples were processed by SAS, a statistical package, and THINK Pascal programs constructed by the author.

The advantage of the FRA is to adopt a soft algorithm that matches the state of nature of the data and to obtain the solution that comes from the whole as well as from the individuals.