大型データの統計処理

清田徹*，内野愛†，中村剛‡，北村右一§

1　はじめに

当研究室においては、過去７年間，大阪国際大学経済学部教授・山本勇次氏がネパールで調査収集してきたセンサス・データの統計処理を、卒業研究として行っている。このデータは，ネパールの都市ポカラにおける全数調査で、約9千世帯，５万人に及ぶ大型のものである。

原データは，山本氏が現地の大学生を組織して収集したものである。それを OCR（光学読取り装置）用紙に転写し，大型計算機で読み取ってフロッピーディスクに保存したものが，コンピュータで利用可能になった最初のデータである。固定長のテキスト形式データで，約7メガバイトと巨大なものであった。

これをパーソナルコンピュータで利用するために，媒体（８インチから５インチへ）とフォーマット（IBMフォーマットから MS-DOS フォーマットへ）を変換した。さらに，フィールド区切りとしてタブを，レコード区切りとして改行を用いることによって，データサイズが約３メガバイトに減少した。これを自由形式データと呼ぶ。

本研究を開始した初年度は，統計処理の十分なソフトウェア環境が完備していなかったこともあり，プログラミング言語(C言語を用いた)に頼っていた。そのため，学生の研究目標も，統計処理というよりは，むしろ，多数に分割されたテキストファイルから，必要なデータを取り出すアルゴリズムの確立が主であった。

しかし，C言語に頼った処理では，

・実用的なプログラムを組めるようになるための，言語の習得に時間がかかる

・プログラムで実行される集計処理そのものは高速であるが，その開発にかなりの時間と努力を要する

などの欠点があった。そこで，翌年度からはデータベースソフトを導入し，これで統計処理を行うことにした。それでも，自由形式データの巨大さゆえ，必要な情報のみを取り出し，データベースソフトで読み込み可能な大きさに縮小するためには，プログラミング言語の利用が欠かせなかった。

＊長崎大学教育学部数学専攻４年
†長崎大学教育学部数学選修４年
‡長崎大学医療技術短期大学部・教授
§長崎大学教育学部数学・助教授

このような環境の下で，２変量間のクロス集計を中心に行ってきたが，分析対象項目が広がるにつれ，データベースソフトで直接利用可能な形式のデータが蓄積された。現在では，情報を取り出すために，学生がプログラミング言語を利用する必要はない。

これまでの，学生に対する教育目標は以下のものであった。

・基本アプリケーションの習熟

データベース操作・表計算・グラフ作成などのアプリケーションは，ワードプロセッサに次いで利用頻度が高い。これらの利用目的と，異なるアプリケーション間でのデータ共有などに習熟する。

・アルゴリズムの確立

データベースソフトなどでは，手作業による処理も可能である。しかし，大型のデータを誤りなく確実に処理するには，再現性のあるプログラムやマクロなどを利用する必要がある。その基本技術を学ぶ。

・アルゴリズムの検証

上記のプログラム・マクロの正当性を検証するために，小さくかつ適切なデータで完全な検査をすることの必要性を知る。

・処理の効率化

大型のデータでは，アルゴリズムの違いで，処理時間が数十倍にもなることがある。そこで，最適なアルゴリズムを選択出来る力を養う。

・最適な情報提示能力

同じ情報でも，提示方法により，見る者に訴える力が異なる。その方法や技術を学ぶ。

つまり，コンピュータによる情報処理能力の育成が主な目標で，結果として得られる統計分析は，その副産物であった。数学科の卒業研究として，このような題材を選ぶことの利点は，現実のデータを扱えることである。コンピュータ処理にしても，統計処理にしても，現実から離れた架空のデータでは，興味が削がれてしまいがちである。また，山本氏の助言を受けながら学生達が解析した結果は，彼の論文の中で利用され，公表さてきた。これも，学習意欲を増す動機となるであろう。

本年度は，「多変量間の分析によって，ポカラの社会構造を把握したい」という，山本氏からの強い要望が出され，大型計算機を利用した多変量解析を行うことになった。しかし，統計学に重点をおいた指導には，筆者(北村)の力の及ばぬところがあり，統計学の専門家である中村の協力を仰ぐことにした。

本稿では，実際にコンピュータ処理を行った清田と内野が，それぞれ，データ処理とその解析結果について述べ，中村が彼らの教育指導に関する部分を記述する。　　(北村右一)

2　データ処理の概要

北村ゼミ・ネパール班は，大阪国際大学の山本勇次氏がネパール・ポカラ市で調査を行ったデータについてのデータ処理に卒論研究として取り組んできた。その成果として，世帯に関する情報（約9,000世帯）と個人に関する情報(約50,000人分)が分類整理され，データベースで利用可能な形のデータとして蓄積されてきた。このデータには「カースト」,「年齢」,「教育程度」等の項目が納められている。

これまでの研究では各項目の度数分布図，並びに2項目の分割表でまとめた分析が行われていた。今年は，多変量解析による統計処理を行うことにより，職業決定に影響していると推測される要因（以下，背景因子という）から，職業を判別することを試みた。

今回の解析は，大型計算機を用いて行った。そのため既存のデータベースから必要な情報を抽出し，大型計算機に受け渡す必要があった。大型計算機では，データの特徴を見るために職業と各背景因子との関連を調べたのち，判別分析によりそれらの背景因子の値のみから職業の判別を行うことを試みた。

以下で，データ，多変量解析および，大型計算機を用いた解析過程の概要について述べる。

データについて

これまで使用してきたデータベースから，男性の世帯主に限ったもの（約7,000件）を用いた。計算機の容量制限もあり，全数（約50,000件）はあまりにも大きすぎるためである。また，女性を省いたのは女性が世帯主全体の約1割と少なかったからである。

各項目は，データの特徴をつかむためにおおまかに分類している。例えば，カーストは20余り存在し，教育程度も40余りある。このように細かい分類のままでは解析を行っても結果に対する解釈は困難だと考えられたからである。また年齢，居住期間のように，値のレンジの広い連続変量については，数値をそのまま用いるのではなく，いくつかのカテゴリーに分類した方が，一般により信頼できる結果が得られる。数値をそのまま用いることは，それらの変量と目的変数との関数関係がその広いレンジで近似的に一次式で表されることを仮定している。しかし，この仮定は一般には成り立たないので，本データのように関連が未知の時にはカテゴリー化は通常行われる工夫である。この操作により，カテゴリーと目的変数との関連を求めることが可能となる。

以下に，各項目の分類を示す。

情報一覧

◎ 職業　農業(AG)，経営者(BI)，日雇い労働者(LA)，恩給生活者(PE)，俸給生活者(SE)の5種
◎ 背景因子
(1) カースト　RANK1から4の4種
(2) 年齢　21－30歳(YOUNG)，31－50歳(MIDDLE)，51歳以上(OLD)の3種
(3) 教育程度　無し(NOTHING)，初等学力程度(ELE)，中等学力程度(JUNIOR)，教養学力程度(SENIOR)，大学卒業程度(COLLEGE)，大学院卒業程度(GRADUATE)の6種
(4) 居住期間　0－10，11－50，51－100，101年以上の4種
(5) 経済階級　富裕，中産，貧困の3種
(6) 宗教　ヒンズー教(Hindu)，仏教(Buddhist)の2種

多変量解析

多変量解析を用いて職業のように多元的側面を有していると見られる事象と，その事象の背後にあるとみられる要因との相関関係を分析すると，事象と個々の背景因子との関連の度合を推定することができる。さらにそれらの間の近似的な関数関係を求めて，予測や判別等を行うこともできる。

本研究では，判別分析（ロジスティック判別分析）を用いて，上の表に示した6つの背景因子による職業（目的変数）の判別を行った。判別分析とは，2つまたはそれ以上の群が存在するとき，所属が未知の個体から得られた多変量データをもとに，その個体がどの群に所属するかを判別するための手法である。そのため本研究に適していると考えられた。

また，ロジスティック分析は，その事象を規定する様々な要因から，その事象の起こる確率を予測しようとするものである。雨の降る確率を求める場合を例にあげる。Yが天気を示し（$Y=1$ が雨，$Y=0$ がその他とする），気圧・風力・湿度・雲量を因子とするとき，それらに適当な重み $b_1,\cdots,b_4$ をかけて加えた値をSとする。

$$S=b_0+b_1\times(\text{気圧})+b_2\times(\text{風力})+b_3\times(\text{湿度})+b_4\times(\text{雲量})$$

また，$Y=1$ の確率は次のように表される。

$$P(Y=1 \mid S)=\frac{e^S}{1+e^S}$$

この式を計算することで，気圧・風力・湿度・雲量から雨の降る確率が予測できることになる。

6つの背景因子から職業を判別する場合も同様に，各背景因子に重みをかけて加えた値から，その職業に就いているかどうかを確率的に予想できる。但し，ロジスティック判別分析は，2群間の判別を行うための多変量解析モデルであるため，職業毎にその職業に就いているかどうかの判別を行った。

解析の過程

大型計算機の統計パッケージ(BMDP)を使うために，これまでのデータを加工し，大型計算機へ受け渡した。その加工にはデータベースソフトを使用し，パソコンで行った。内容は，解析に必要な情報の既存データベースからの抽出と，文字による情報の数値化である。

大型計算機では，解析結果からデータの社会背景に即した解釈ができるように，各項目の度数分布図，職業と各背景因子との分割表を作り，データの特徴を調べた。以前にも同様の解析は行われていたが，世帯主に限ったものではなかったため，この解析を行った。職業と各背景因子との分割表は，1つのプログラムを実行させるだけで出力される。また各々の背景因子を組み合わせた分割表も簡単に出力することができる。

続いて，6つの背景因子による職業の判別を行った。最初は5つの職業を同時に判別する方法で計算を試みたが，計算時間が長くかかりすぎて打ち切られたため，分析できなかった。そのため，その職業に就いているか否かを職業別に判別する方法で，分析を行うことになった。この計算を実行することによって，縦軸が度数，横軸が確率のヒストグラム，および，ある確率で判別したときに正しく判別できる確率の表が出力される。ヒストグラムは，その職業に就いている人のものと，そうでない人のものの2つが出力される。それらに基づいて以後の分析を行った。この分析結果については次節

で述べる。

最後に，計算実行上の工夫点をあげておく。当初，プログラムの実行部とデータ部を一つのデータセットに入れていたが，データ量が大きすぎたためプログラムの書き換えや実行のたびに時間がかかっていた。通常，データ部は書き換える必要がないため，実行部とは別のデータセットに分けたところ，それらがスムーズに行えるようになった。（清田徹）

3　データ処理の分析と解釈

本卒業研究では，前節（情報一覧）であげた職業と6つの背景因子（変数）との関連を分析し，職業の特徴付けを行うことを当初の目的とした。多変量解析で得られた解釈を正しく把握するためには，職業と個別の背景因子との関連を知っておく必要があった。

(1)　職業とカースト

AGとSEにおけるランク4の割合は6.8%，3.7%と低く，BIにおけるランク3の割合は56.1%と高い。また，LAにおけるランク3とランク4の割合は43.4%，39.5%と高く，PEはランク3の割合が91.8%と極端に高い。（表1）

(2)　職業と年齢

AGにおけるMIDDLEとOLDの割合が44.0%，41.9%と高く，比較的高年齢層に偏っていることが分かる。また，BIとSEにおけるYOUNGとMIDDLEの割合は近似的に等しく，PEはYOUNGの割合が0.6%と極端に低い。（表2）

表1：職業におけるカースト（ランク別）の割合

	RANK1	RANK2	RANK3	RANK4	TOTAL
AG	36.1	30.9	26.1	6.8	100%
BI	13.8	13.7	56.1	16.5	100%
LA	7.2	9.9	43.4	39.5	100%
PE	1.9	6.0	91.8	0.2	100%
SE	31.0	26.0	39.4	3.7	100%
TOTAL	24.0	21.5	43.2	11.3	100%

表2：職業における年齢層の割合

	YOUNG	MIDDLE	OLD	TOTAL
AG	14.0	44.0	41.9	100%
BI	22.3	54.2	23.5	100%
LA	25.8	56.5	17.7	100%
PE	0.6	59.8	39.5	100%
SE	28.3	62.0	9.7	100%
TOTAL	20.2	53.8	26.0	100%

表３：職業における教育程度の割合

	NOTHING	ELE	JUNIOR	SENIOR	COLLEGE	GRADUATE	TOTAL
AG	33.4	51.1	9.4	4.8	1.0	0.3	100%
BI	17.7	59.1	9.3	10.2	0.3	0.5	100%
LA	56.4	40.2	2.3	0.7	0.4	0.0	100%
PE	10.2	82.9	3.9	2.8	0.2	0.0	100%
SE	7.0	42.6	17.2	17.7	11.3	4.3	100%
TOTAL	23.4	51.7	10.3	9.0	4.2	1.4	100%

(3) 職業と教育程度

職業における教育程度では，全体的に低学歴の割合が高いことが分かるが，BIとSEにおけるSENIOR以上の教育程度の割合が他の職業に比べ高い。（表３）

SENIOR以上の教育程度における職業の割合では，BIとSEの割合が他の職業に比べ高いことが，さらに明白である。（表４）

ここでは，BIには資本が必要な職種があり，この職種につける人は経済的にも教育を受けることが可能であったことも推測される。また，SEの殆どが公務員で，採用試験があるという社会背景から，SEが高学歴至高であることが理解できる。

(4) 職業と居住期間

AGは職業柄その土地に根付くようで，101年以上の割合が56.4%と高くなっている。一方，PEは，流動性の高い職業であるせいか０－10年の割合が71.7%と高い。（表５）

(5) 職業と経済階級

職業における経済階級では，全体的に富裕階級が3.3%と極端に低く，中産・貧困階級の割合はほぼ等しい。また，PEとSEにおける経済階級の分布が近似的に等しいことが分かる。（表６）

表４：教育程度における職業の割合

	NOTHING	ELE	JUNIOR	SENIOR	COLLEGE	GRADUATE	TOTAL
AG	45.0	31.1	28.6	16.7	7.3	7.7	31.5
BI	17.5	26.5	20.8	26.3	17.2	8.8	0.2
LA	26.2	8.4	2.4	0.9	1.1	0.0	10.9
PE	3.1	11.4	2.7	2.2	0.4	0.0	0.1
SE	8.1	22.6	45.5	53.9	74.0	83.5	27.4
TOTAL	100%	100%	100%	100%	100%	100%	100%

表５：職業における居住期間の割合

	0—10年	11—50年	51—100年	101年以上	TOTAL
AG	11.5	14.3	17.8	56.4	100%
BI	28.8	26.5	17.5	27.2	100%
LA	39.5	25.4	14.7	20.3	100%
PE	71.7	20.3	2.8	5.2	100%
SE	34.3	14.7	12.7	38.3	100%
TOTAL	29.2	18.9	14.9	37.0	100%

表６：職業における経済階級の割合

	富裕	中産	貧困	TOTAL
AG	6.1	62.8	31.1	100%
BI	2.7	34.4	62.9	100%
LA	0.0	22.8	77.2	100%
PE	1.7	46.6	51.7	100%
SE	2.4	45.5	52.0	100%
TOTAL	3.3	46.0	50.7	100%

表７：経済階級における職業の割合

	富裕	中産	貧困	TOTAL
AG	57.1	42.7	19.1	31.2
BI	19.2	17.5	28.9	23.3
LA	0.0	5.4	16.4	10.8
PE	3.7	7.2	7.2	7.1
SE	20.1	27.4	28.3	27.6
TOTAL	100%	100%	100%	100%

さらに，富裕階級におけるLAの割合は０％である。世帯資産とみなすことのできるデータが土地しかなく，農地と敷地を用いて経済階級を決定しているため，貧困階級におけるBIとSEの割合が28.9%，28.3%と他の職業に比べ比較的高くなっていることが分かる。(表７)

(6) 職業と宗教

ヒンズー教が８割以上を占めている。その中で，PEにおける仏教の割合が46.6%と高い。これは，PEはランク３の割合が高く，ランク３の人はもともと仏教徒だという社会背景によるものであろう。(表８)

以上見たように，各変数毎に，職業とその変数との関連を考察しても，特定の職業がある変数と深く関わり合っているというような単純な解釈は困難なことが分かる。そこで，職業と全ての変数との関連を同時に分析するための統計的方法である多変量解析を試みた。前節で述べたようにロジスティックモデルに基く判別分析法により，AGとAG以外(NON-AG)のような，一つの職業とそれ以外の職業とを６つの変数の値のみで判別することを試みた。これにより各職業の特徴付けを行えることが期待された。まず，AGとNON-AGとを６つの変数の値から判別するための判別式(詳細は省くが，基本的には６つの変数のうちから最良の変数の組み合わせを選択し，一次結合を構成したもの)を得，その判別式に各世帯主の６つの変数の値を代入して個人毎に，その人がAGである確率(p)を求めた。求めたpの分布をAGの人達(AG群)とそれ以外の人達(NON-AG群)とについてプロットしたのが図１である。二つの図を見較べてみると，分布が左右に分離されており，高い精度で判別可能なことが分かる。同じ解析をBI，LA，PE，SEにも行ったところ，LA，PEにおいても，やはり，AG同様に高い精度で判別可能である。

しかし，BIにおいては，図２を得，分布があまり左右に分離されておらず，判別することは困難であることが分かる。SEについても同様である。

表８：職業における宗教の割合

	HINDU	BUDDHIST	TOTAL
AG	94.8	5.2	100%
BI	82.4	17.6	100%
LA	87.6	12.4	100%
PE	53.4	46.6	100%
SE	85.9	14.1	100%
TOTAL	85.8	14.2	100%

表９：最良判別確率

AG or NON-AG	75%
BI or NON-BI	64%
LA or NON-LA	76%
PE or NON-PE	82%
SE or NON-SE	70%
BI or SE	69%

例えば，AGかNON-AGかを判別する場合，職業が未知である人の6つの変数の値をその判別式に代入して得られた値が，ある指定された値より大きい時にはAG，そうでないときにはNON-AGと推量することができる。同様の推量を職業が既知の人について行えば、その推量が正しいかどうか確認できることになる。そこでAGの人全てについて推量を行うことで、AGの人のうち何％を正しく判別できるかが分かり，この割合を「正しく判別できる確率」と考えることができる。ここで，その指定された値を判別点とよび，その時の「正しく判別できる確率」を判別確率ということにする。判別確率は判別点の値に依存するので，判別確率を最大にする判別点の値を最良判別点と呼び，そのときの判別確率を最良判別確率と呼ぶ。最良判別確率は，常に50％以上である。最良判別確率を，AGとNON-AG，BIとNON-BI，LAとNON-LA，PEとNON-PE，SEとNON-SEについて求めて表９に示した。これによると，BIとSEの判別確率は他に比べて低いため，今用いている６つの背景因子のみで他の職業と判別するのは困難であることが分かった。実際，個別の解析を見てもBIとSEはともに高学歴者の殆どを占めている等，他の職業と比べ特異な点が目につく。そこでこの２つの職業に強い関心を抱き，BIとSEにしぼって解析を進めることにした。

BIとSEに関するデータのみを取り出し，前述と同じ様に判別分析を行うことで，図３を得た。二つの図を見較べてみると，分布が左右に分離されていないことや，最良判別確率が69％と低いことから，BIであるかSEであるかを，今用いている６つの背景因子のみで判別することの困難性が分かった。

言い換えると、BIの６つの背景因子の分布（６次元の多変量分布）とSEの６つの背景因子の分布とは近似的に等しいことが示唆される。両者の判別を試みるには，新たな背景因子を加味する必要があろう。

（内野愛）

4　大規模データを用いた統計解析の教育

本研究は典型的な大規模探索的研究（Exploratory Study）である。小標本に基く検証的研究（Confirmatory Study）では，仮説の検定並びに母数の推測が興味ある主題となるが，本研究のデータは既に述べたように，約9,000世帯50,000人という膨大な情報からなり，特に検証すべき特定の仮説はない。探索的研究においては，調査項目間の関連，特異な部分集団の検出等を行い，最終的にはデータの全体像を明確にする統計指標の構成あるいは新たな仮説の提示が研究主題となる。探索的データは，医学における臨床データベース（例えば大学病院に訪れる患者さんのカルテや検査記録を電算機ファイル化し蓄積したもの），大学あるいは地域の健康診断記録，さらには政府統計，経済統計等幅広く見られる。それらに共通する点は，日常生活の営みの一部を記録した資料を大量に収集することにより，何らかの客観的かつ普遍的特徴を見い出そうとするものである。１ないし２変量の記述統計的方法（度数分布や相関係数等）はかなり発展しているが，３変量以上の記述統計的方法は，1970年代から関連図書が現れはじめ，因子分析，クラスター分析，パス解析等幾つかの有力な手法はあるものの，実は未発達である。このため，多変量での探索的データ解析法は試行錯誤の繰り返しが余儀なくされることが多い。

大規模探索的研究においては，データの前処理を十分行うことにより，実際の計算時間が長くなり

過ぎるのを防ぐのが肝要である。長崎大学総合情報処理センター汎用電算機で，もし全数約50,000件を用いたならば，記憶容量の制限から，計算の実行そのものが困難と思われた。このため，今年度は初めての試みでもあるので，安全のために世帯主約9,000件のみの解析に限った。それでもなお複雑な分析のときには，計算時間が長すぎて途中で打ち切られたこともあった。今後もし全数解析をする必要が生じたさいには，なんらかの特別な対策を考案する必要がある。大規模探索的研究にこのような問題はつきものであり，データの内容，統計的方法，情報機器の全てに精通しているか，あるいはそれらに精通した者が共同作業を行うことにより，初めて実行可能となる。

ここで用いられた統計パッケージBMDPは，多変量解析には定評があり，本研究には適していたといえる。しかしマニュアルも出力も英語なので，初めて使う時にはかなりの忍耐を必要とする。さらに統計的知識と経験が十分に無い学生にとっては英語による統計専門用語の理解も要求されるので，一層の困難であったと推察される。しかしながら，両学生が遂に使いこなしえたのは，努力の賜物であると同時に，利用目的が明確であったことも幸いしたと考えられる。もしBMDPの利用法と統計的方法の解説を一般論も含めて解説すれば，膨大な時間のかかることを思えば，初めから目的を定めて必要な知識のみを修得していったのは効率よい方法であった。統計的方法の修得は理論からよりも，実際のデータから入ったほうが好ましいことを実証したともいえる。

情報処理センターの汎用計算機(MSP)の利用方法も実は，きちんと解説すると数十時間かかり，なお十分な修得は困難なことを，医療技術短大での学生実習で経験していた。従って，ここでも，解説なしにまず最低限必要な操作を実演し，後は問題が生じた時に一緒にマニュアルをみたり，実験したりして解決していった。数週間後には，一人で使いこなし，新しい結果も出力できていた。情報機器の修得も，一般論の解説からではなくして，実際のデータの提供と必要な出力の指示から入るのが，効率良いことを実証したといえる。

しかしながら，まずデータと必要な出力を提示し，問題を解決することを通じて，統計的方法並びに情報機器の修得を促すことが困難な環境もあろう。特に，多人数を対象とした学生実習では，個々の学生の抱える問題へのマンツーマンでの対応が困難である。しかし，この場合も問題解決型の修得法を如何に組み込んでいくかは，今後解決すべき問題である。

ネパールデータは全数調査という点でもユニークである。全数であるからして，未知母数の推定という問題はありえない。それでもなお，検定すべき問題はありうる。例えば「学歴と職業との相関は有意か」という問題である。これも観察された相関係数が母数であるからして，その値が0でない以上母相関が0でないのだから，検定は無意味ということになる。しかしそのような「母集団－標本」という解釈ではなくして，「観察データにより条件付けされた確率化テスト」という考え方により上の検定は正当化される。観察された学歴と職業の個々の分布は与えられたものとし，さらに仮に学歴と職業とが互いに独立に無作為に割付けられたとしたときに，観察された相関係数をうる確率は小さすぎないか，という論理である。

本研究では多変量解析的方法により，職業と職業を規定する要因との関連を分析し，職業の特徴付けを行うことを当初の目的（Working Hypothesis）とした。職業は多くの要因に依存して決定されると思われるので，この解析法は適していたと考えられる。ここでは多変量解析法の一つであるロジ

スティック分析を記述統計的に用いているので，表面上検定の問題は出て来ない。しかし実は，ロジスティック判別分析における有効な変数の選択は，各変数の有意水準を基準にしているが，その論理は上のパラグラフで述べた確率化テストに基くものと解釈される。このような細かい理論上の問題は後から解説しても十分であり，初めに解説しても何の興味も湧かないであろう。

以上，本研究を，情報機器並びに統計学修得の為の新しい方法の実験とみなして，考察を述べたが，結果として予想以上の成果を観察できた。これは学生の自主的にして真摯な努力によるものである。

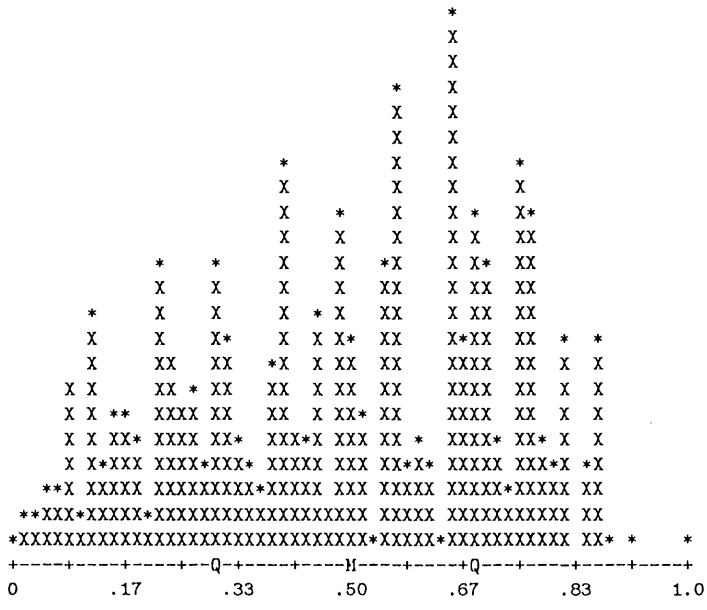
（中村剛）

5　資料

図1：AG or NON-AG

AGグループのヒストグラム

'X'は7人につき1つ，'*'は7人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。



NON-AGグループのヒストグラム

'X'は22人につき1つ，'*'は22人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。

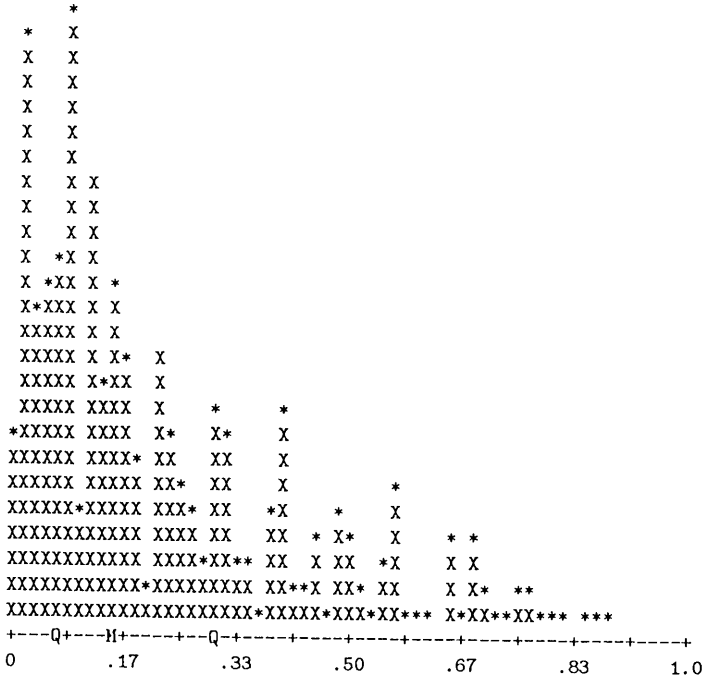
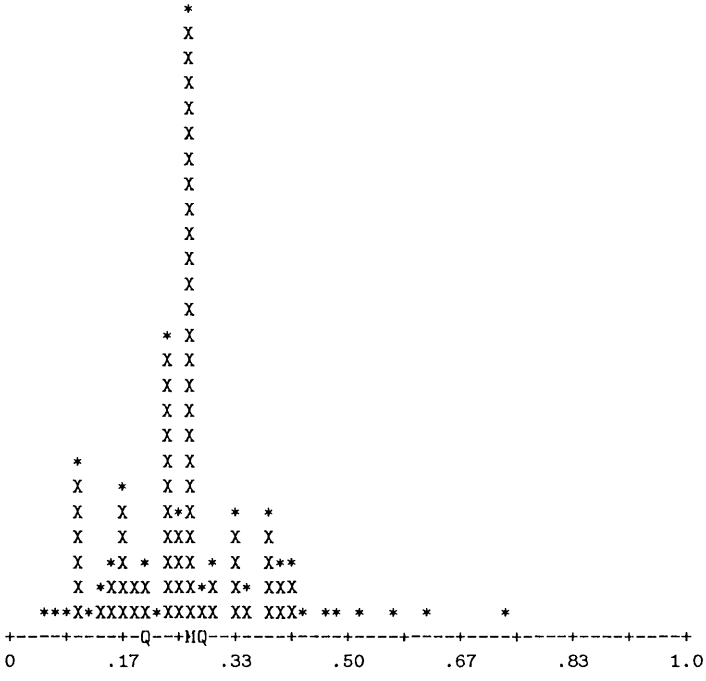


図2：ＢＩ or ＮＯＮ-ＢＩ

ＢＩグループのヒストグラム

'X'は21人につき1つ，'*'は21人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。



ＮＯＮ-ＢＩグループのヒストグラム

'X'は53人につき1つ，'*'は53人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。

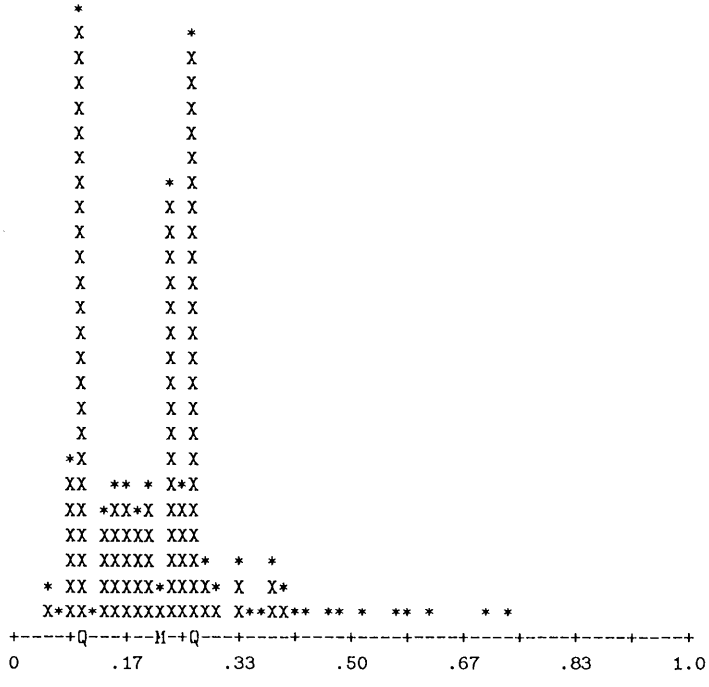
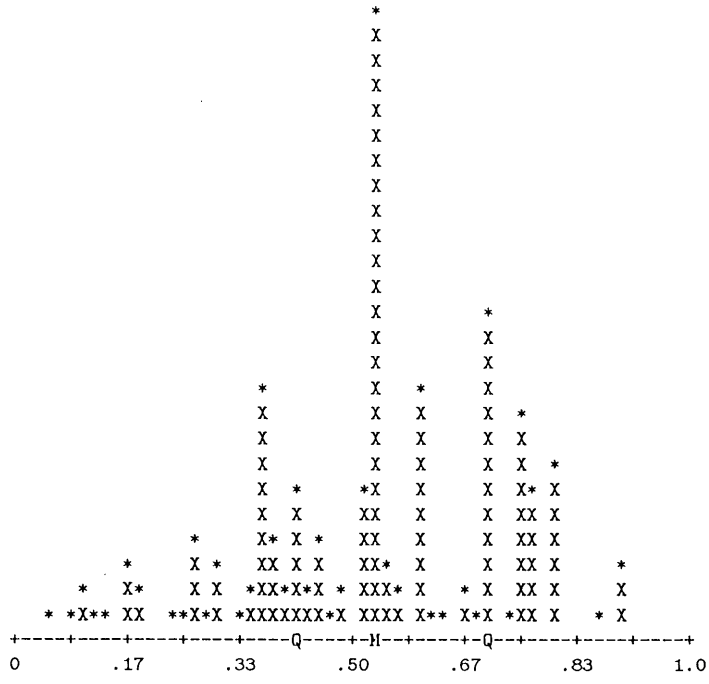


図３：ＢＩ or ＳＥ

ＢＩグループのヒストグラム

'X'は14人につき１つ，'*'は14人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。



ＳＥグループのヒストグラム

'X'は13人につき１つ，'*'は13人未満の場合に用いている。
'M'は度数の二等分点，'Q'は度数の四等分点を表す。