# 統計表構造に沿った統計データベースの設計*

椿　　康和

## 1. はしがき

データベース・システムで管理しようとする統計データにはさまざまなものがある．筆者らは横山＝椿[7]において，地方統計のための DBS が持つべき機能について提案したが，その中で管理対象となるデータは，

a）国や地方自治体の調査や日常的業務によって収集された原資料である個票データ，

b）それらを集計し統計表の形式で公表された個別統計データ，

c）さらにある利用上の観点からそれらを再編成した総合統計データ，

に大別されている．これらの間には構造と利用形態において大きな相違があり，その共用化と統合的管理のためのデータベース・システムもそれに合わせて設計・構築しなければならない．

a）については筆者らは，その構造を記述しデータ処理に結びつける一つの方法を提案した（横山＝椿[8]）．b），c）については，佐藤＝穂高[6]など多くのシステムが設計されているが，それらはb）の集約統計データの構造を一意に表現する厳密なモデル化に基づいた標準形を想定してデータ構造を記述し，既存の統計データを根本的に再編成したうえでc）の総合的な統計 DB を構築しようとする立場からのものである．

本稿ではこれとは逆に，既存の個々の統計表について，その表頭・表側の構造定義に沿ってデータを管理することを目的とするシステムの設計を

* 本システムの設計にあたり，表頭定義行列による表頭構造の記述について御教示いただいた，本学部横山和典教授に深く感謝いたします．

行った.

このシステムの特徴は,

a）統計表の表頭・表側構造を分類変数と解析変数の表現式で端的に記述する.

b）分類変数の値の重複がなく,格納領域を節約する.

c）地域や産業等の範囲に関連した分類変数については,利用者が独自の分類体系を既存のものにつけ加えることができる.

d）入力時のデータの変換作業が少ない.

などである.

以下,2章では本システムを構築するうえで必要な,統計表とその DB 化に関する基礎概念について論じ,3章ではシステムの基本構想と検索を中心とした設計について述べる.

## 2. 統計データと統計表

### 2.1 統計データの性質

#### 2.1.1 分類変数と解析変数[2.1)]

統計データは,場所,時間によって規定され,ある実質的意味において共通性を持つ個体の集まり(統計集団)に関する属性を表すデータである.属性は分類・集計上の観点から,一般に分類変数と解析変数に大別され,次のように定義される[2.2)].

a）分類変数:個体の定性的な性質で分類カテゴリを定義するもの.

b）解析変数:個体の定量的な性質で集計・解析の対象となるもの.

分類変数は,単独またはその組み合わせにより部分集団を識別し,それらの分類・合併の基準となる.他方解析変数はその測定値を表す.例えば,広島市の金属製品製造業の従業者数という値では,広島市および金属製品製造業が,それぞれ分類変数,市町村および産業の1つの実現値で,従業

2.1）分類変数と解析変数は,それぞれ分類項目と統計項目,カテゴリ属性とサマリ属性などともよばれているが本稿ではこの表現を用いる.

2.2）Chan=Shoshani[1]

者数が解析変数になる[2.3]．

ところで，分類変数の中には相互に階層的意味関係が成立するものと，そうでないものがある．前者では市町村→県→地方のように変数値の間でｎ：１の対応関係が成立する．それらの組み合わせはクラスタ型抽象化(cluster abstraction) といわれ，統計集団の包含関係を表す．他方，後者では互いに独立な概念として，市町村×性別のように各属性値の可能なすべての組み合わせに対して分類が意味を持つ．これらは直積型抽象化(cross product abstraction) といわれ，統計集団の細分化を表す[2.4]．

2.1.2　データの集約

統計データに関する計算処理の対象は解析変数の測定値である．処理には変数間の演算による新しい変数の定義，平均や分散といった統計量の導出，多変量解析等の高度な分析手法による統計集団の分析など一般にデータ解析とよばれるものと，分類変数にしたがってデータを集計し，より上位の統計集団の変数値を求めるデータ集約とよばれるものがある．前者の処理は一般にデータベースからの検索結果を引き継ぐ各種の統計パッケージの機能であり，他方後者は統計データベースに付随する機能と考えられる．

データの集約は，より具体的には対象範囲と水準を定め，クラスタ型抽象化ではその下位レベルから上位レベルについて，直積型抽象化ではそれを構成する１つの分類変数について，集計操作を行うことと定義されるが，以下の諸問題がある．

(1)　集計方法

集計においていかなる方法をとるかは解析変数の性質に依存し，次のように大別される[2.5]．

ａ）加算性を持つならば単純集計あるいは加重集計が可能である．

ｂ）指数のように加算性を持たないものや，平均所得のように他の解析

---

2.3）この両者の概念上の相違は必ずしも絶対的なものではない．横山＝椿[7]

2.4）Chan＝Shoshani[1]

2.5）小菊ほか［5］

変数間の演算で導出されたものは，定義にしたがって再計算する必要がある．

c）物価指数や資本ストックのように，累積性を持たない時系列データの集計では，期間平均や期間内特定期の値をとる．

(2) 既集約データ

クラスタ型抽象化に沿った集約では，あらかじめ上位レベルのカテゴリに該当する値を格納しておく方法と，検索指示に応じてその都度集計処理を行う方法がある．前者では冗長性が，後者では効率性がそれぞれ問題となる．

(3) 分類体系の問題

a）既存の分類体系の中に存在しないカテゴリでの集約をどう処理するか．

b）市町村合併によって新しい行政区域が発生または消滅するように，分類体系が時間とともに変化する場合をどう扱うか．

(4) 秘匿値

データに統計法上の秘匿値が含まれる場合，集計操作では正しい結果を得ることができない．

## 2.2 統計表

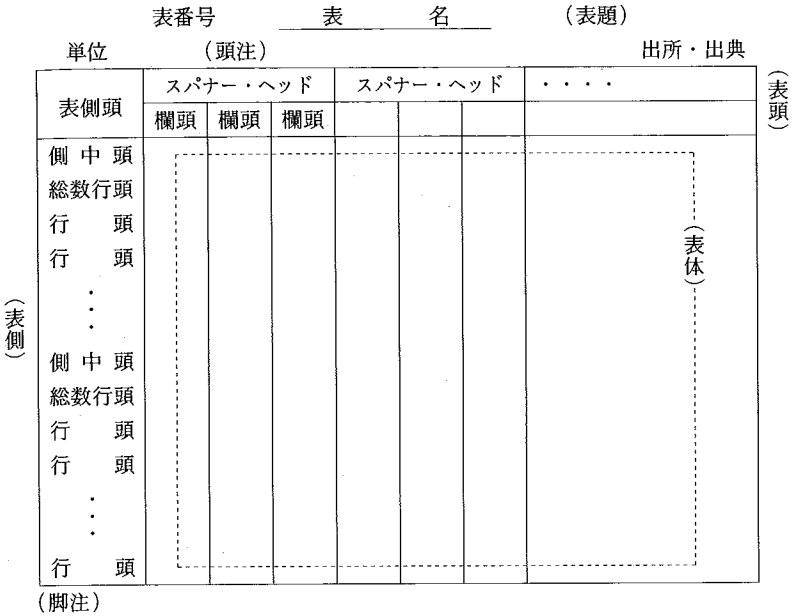
### 2.2.1 統計表の種類と構造

統計表は，場所，時間，数量，質の4つの分類基準によりその種類が定まり，それぞれを①場所的系列表（場所的分布表），②時系列表，③数量的構造表，④質的構造表という．③，④ではさらに分類基準の配置が一元であるか二元であるかにしたがって，③は度数分布表と相関表とに，④は質別構造表と連関表とに分けられる．

これらの表の中でなんらの記述的解説を伴わず，表単独で含まれるデータの持つ意味を表現できるものを正式統計表という．正式統計表は，表題，表側，表頭，表体の4つの要素から構成され，このほか頭注，脚注を含む[2.6]（図2.1）．

a）表題は表番号と表名から成る．

**図2.1** 正式統計表の構造




b）表側は表側頭，側中頭，行頭などから成る．表側頭は，表側に記入されている事項を総称するもので，市町村，商品，産業などの分類名や測定項目名が入る[2.7]．側中頭は，表側頭を細分化した事項を表し，分類変数のカテゴリ値を示す行頭の見出しとなる．行頭には総数を示す総数行頭とそれ以外の一般行頭がある[2.8]．さらに細分化が必要な場合には副次側中頭などを入れる．

---

2.6）石国［4］

2.7）表側頭がクラスタ型抽象化の全体（分類体系名）を示すなら，そこには各分類階層のカテゴリ値が現れる．またその体系に含まれないカテゴリ値が現れることもある（例：国－都道府県－人口５万人以上の市の計）．表側頭が直積型抽象化を表していてもそれで定義される全カテゴリが出現するわけではなく，一方のカテゴリ値により他方の定義域が異なることもある（例：市町村×産業で，市については産業大分類，町村については全産業）．

c）表頭には表側の分類項目に対する見出し項目を置き，項目が1個の時は一次元，複数個の時は二次元の表となる．二次元の表では，表頭は欄頭，スパナー・ヘッドなどから階層的に定義され，その全体で一つの意味を定義する．表頭は表側に比較して数値間の関係を強調する．

d）表体は統計数字を書き込む場所である．

統計データを表章する際に個々の分類変数と解析変数をa）～c）のいずれに配置するかについては一般的基準がなく，統計作成者の裁量にまかされている．これが種々の形式の統計表が作り出される原因である．

### 2.2.2　関係 DB による統計表の処理

統計表に収められたデータをDB化しようとする際，データが長方形の資料行列の形で存在していることから，その理論的モデルとして，表形式の関連するファイル（関係）の集まりによって DB を構成する関係モデル (Codd [2]) が用いられることが多い．

この場合，統計表の多様性を考慮せず両者の間の形式的類似性のみに注目して，単純に資料行列の列を属性，行をタプルに対応させて DB に格納するだけであるならば，作業はきわめて容易である．しかし，2.1節に述べたような統計データの性質から，そのままでは検索や集約等の処理に容易に対応できないのは明らかである．例えば，関係モデルではキーとなる属性とその他の属性との間には関数従属性以外に何の関連性も許容されていないため，資料行列の列が複数の分類変数の組み合わせにより定義されていても，それは属性を定義する記述的情報としてしか扱うことができない．したがって，DB 化において統計表の構造を管理し，それを検索や集約に結びつけるなんらかの手法が必要とされるのである．

統計表形式のデータを関係モデルで処理するにあたり，多様な形式の統計表を1つの標準的な形に変換して管理する方法がある．筆者らは，資料行列に対して適切な変換を施し，部分集団を一意的に識別する分類変数の組をタプルのキーに，その他の分類変数と解析変数をキーに関数従属する

2.8）合計はそれに含まれている数値の計，総数は全体での集計結果であり，丸めが行われた場合，両者は必ずしも一致しない．

**図2.2　基本レコード変換**

a）射影－結合変換

（県勢）

| 市町村 | 人口 | 面積 |
|---|---|---|
| 広島市 | 90 | 736 |
| 呉　市 | 23 | 145 |
| 竹原市 | 4 | 117 |
| 三原市 | 8 | 204 |
| 尾道市 | 10 | 111 |

射影 →　← 結合

（県勢Ⅰ）

| 市町村 | 人口 |
|---|---|
| 広島市 | 90 |
| 呉　市 | 23 |
| 竹原市 | 4 |
| 三原市 | 8 |
| 尾道市 | 10 |

＋

（県勢Ⅱ）

| 市町村 | 面積 |
|---|---|
| 広島市 | 736 |
| 呉　市 | 145 |
| 竹原市 | 117 |
| 三原市 | 204 |
| 尾道市 | 111 |

b）分割－合併変換

（階級別人口）

| 市町村 | 年齢階級 | 人口 |
|---|---|---|
| 広島市 | 0－15 | 22 |
| 広島市 | 15－65 | 62 |
| 広島市 | 65－ | 6 |
| 呉　市 | 0－15 | 5 |
| 呉　市 | 15－65 | 16 |
| 呉　市 | 65－ | 2 |

分割 →　← 合併

（階級別人口Ⅰ）

| 市町村 | 年齢階級 | 人口 |
|---|---|---|
| 広島市 | 0－15 | 22 |
| 広島市 | 15－65 | 62 |
| 広島市 | 65－ | 6 |

＋

（階級別人口Ⅱ）

| 市町村 | 年齢階級 | 人口 |
|---|---|---|
| 呉　市 | 0－15 | 5 |
| 呉　市 | 15－65 | 16 |
| 呉　市 | 65－ | 2 |

c）レコード－値変換

（産業別就業者）

| 市町村 | 産業 | 就業者 |
|---|---|---|
| 広島市 | 一次 | 12 |
| 広島市 | 二次 | 126 |
| 広島市 | 三次 | 286 |
| 呉　市 | 一次 | 2 |
| 呉　市 | 二次 | 39 |
| 呉　市 | 三次 | 63 |

→　←

（広島市）

| 産業 | 就業者 |
|---|---|
| 一次 | 12 |
| 二次 | 126 |
| 三次 | 286 |

（呉市）

| 産業 | 就業者 |
|---|---|
| 一次 | 2 |
| 二次 | 39 |
| 三次 | 63 |

d）フィールドーレコード変換

（工業統計）

| 市町村 | 製造品出荷額等 |
|---|---|
| 広島市 | 181,238 |
| 呉　市 | 60,329 |
| 竹原市 | 12,616 |
| 三原市 | 24,046 |
| 尾道市 | 11,906 |

⟶
⟵

（製造品出荷額等）

| 市町村 | 金額 |
|---|---|
| 広島市 | 181,238 |
| 呉　市 | 60,329 |
| 竹原市 | 12,616 |
| 三原市 | 24,046 |
| 尾道市 | 11,906 |

e）真ー偽変換

（人口10万人以上の市）

| 市 |
|---|
| 広島市 |
| 呉　市 |
| 尾道市 |

⟶
⟵

（市）

| 市 | TRUTH(人口) |
|---|---|
| 広島市 | 1 |
| 呉　市 | 1 |
| 竹原市 | 0 |
| 三原市 | 0 |
| 尾道市 | 1 |

f）空値変換

（人口）

| 市町村 | 人口 |
|---|---|
| 広島市 | 90 |
| 呉　市 | 23 |
| 竹原市 | 4 |
| 三原市 | 8 |
| 尾道市 | 10 |

⟶
⟵

（人口）

| 市町村 | 人口 | 世帯数 |
|---|---|---|
| 広島市 | 90 | — |
| 呉　市 | 23 | — |
| 竹原市 | 4 | — |
| 三原市 | 8 | — |
| 尾道市 | 10 | — |

属性に対応させることにより，第3正規形の関係を導出し，それをベースとした統計 DB の構築とそれに対する関係演算によるビューの持つ意味を検討している[2.9]．

各種の統計表からこの標準形を導出する過程は，Kent[3] の6種の基本レコード変換（elementaly record transformation，図2.2）とそれら

2.9）横山＝椿[7]

**図2.3** 事業所統計表の変換

a）原表：産業（大分類），経営組織（2区分），従業者規模（5区分）別事業所数及び従業者数－都道府県，市区町村

| 市町村および産業大分類 | 総数 | | | | | | 民営 | | | | | 国・公共事業体・地方公共団体（非民営） | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 事業所数 | 従業者数 | | | | | 総数 | | 1～4人 | | （中略） | | |
| | | 総数 | 個人業主 | 家族従業者 | 雇用者 | | 事業所数 | 従業者数 | 事業所数 | 従業者数 | | 事業所数 | 従業者数 |
| | | | | | 総数 | 常雇 | | | | | | | |
| | | | | | | | | | | | | | |

b）変換後の標準形

①産業大分類，形態別従業者

| 市区町村 | 産業大分類 | 従業形態 | 従業者数 |
|---|---|---|---|
| 広島市 | 農林水産業 | 個人業主 | |
| 広島市 | 農林水産業 | 家族従業者 | |
| 広島市 | 農林水産業 | 雇用者総数 | |
| ・ | 農林水産業 | 常雇 | |
| ・ | 鉱　　業 | 個人業主 | |
| ・ | 鉱　　業 | 家族従業者 | |
| ・ | 鉱　　業 | 雇用者総数 | |
| ・ | 鉱　　業 | 常雇 | |
| ・ | ・ | ・ | |
| ・ | ・ | ・ | |
| ・ | ・ | ・ | |

②産業大分類，経営組織，従業者規模別事業所数，従業者数

| 市区町村 | 産業大分類 | 経営組織 | 従業者規模 | 事業所数 | 従業者数 |
|---|---|---|---|---|---|
| 広島市 | 農林水産業 | 民営 | 1～4人 | | |
| 広島市 | 農林水産業 | 民営 | 5～9人 | | |
| ・ | ・ | ・ | ・ | | |
| ・ | ・ | ・ | ・ | | |
| ・ | ・ | ・ | ・ | | |
| 広島市 | 農林水産業 | 民営 | 30人以上 | | |
| 広島市 | 農林水産業 | 非民営 | 総数 | | |
| ・ | ・ | 非民営 | 1～4人 | — | — |
| ・ | ・ | ・ | ・ | — | — |

を組み合わせた操作により行われる[2.10]．例えば，表題の変数を表側に移す操作はc）であり，表頭の変数と表側の変数を入れ替える操作にはこれらを組み合わせたいわゆる値－属性変換が相当する．図2.3のa）は市区町村別事業所統計表で，やや複雑な構造をしているが，これも次の手順によってb）の標準形に変換される．

2.10）ここで用いられているレコード，フィールドは，それぞれ関係モデルのファイル，属性にあたる．

a）この表の表側は市区町村×産業大分類の直積型抽象化として定義されている．一方表頭は従業形態，経営組織×従業者規模の２つの分類変数が併置されており，これらはそのまま表側に移行できないから，まず射影－結合変換により２つのファイルに分割する．

b）分割されたそれぞれのファイルで，表頭分類変数を属性－値変換により表側に移行し，一方を市区町村×産業大分類×従業形態，他方を市区町村×産業大分類×経営組織×従業者規模とし，これらを関係のキーとする．

しかし，このような変換によって導出された第３正規形の関係で DB を構成するとしても，以下のような問題点が残る．

a）キーには直積型抽象化が成立する分類変数が入り，タプルはそれぞれの分類変数の定義域全体から成る直積集合の全要素に対して定義される．ところが実際の統計表ではそのように定義された部分集団のすべてについてデータが存在することは稀であり，その結果関係は多くの欠損値を含むことになる[2.11]．

b）さらに，直積で定義されたタプルに分類変数の定義域の値が重複して現れるという冗長性がある．

c）クラスタ型抽象化における上位階層の既集約データのように論理的に導出可能なデータは，冗長性を除くために排除される．しかし，その導出に要する手間と時間の効率化の問題と秘匿数値を含む集約の問題が解決されていない．

d）さまざまな統計表中のデータを関係に格納する際に，その形式に適合させるためのデータ形式の組替え処理が莫大なものになりかねない．

本稿で提案するシステムではこれらを考慮して，上述の方法とは逆に表頭・表側の統計表構造を記述し，データを個々の統計表に収められた形式に沿って格納する方式を採用している．

---

2.11）図2.3の例では，非民営の事業所について従業者規模別の値が存在しない．

# 3. システムの設計

## 3.1 基本構想

### 3.1.1 システムの概要

(1) DBS の機能

一般に統計 DBS に要求される機能は，DBS を構築しデータを提供する側からは，

a）データの構造，その内容や特性などのいわゆる統計データに関するメタ＝データの管理，

b）統計データの DBS への格納と管理，

などであり，他方統計データを利用する側からは，

c）DB 中に存在する統計データに関する情報の提供，

d）必要なデータの検索，

e）集約を含め，各自が必要とする内容・型式へのデータ編集および変換，

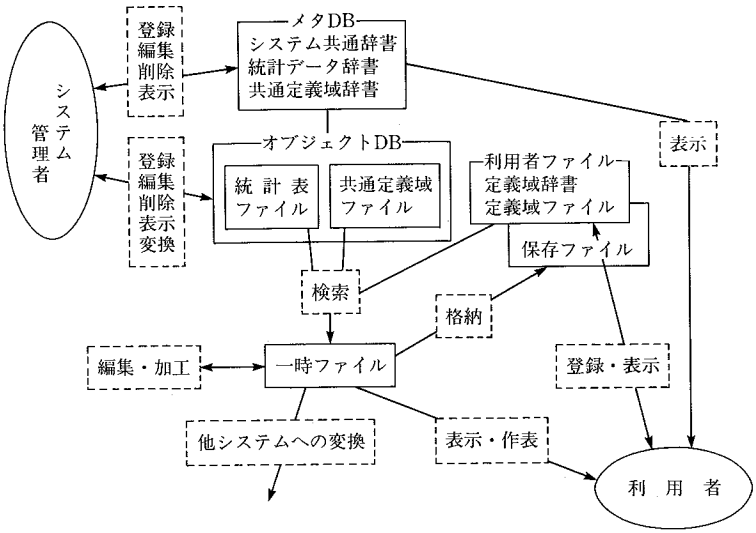



図3.1 システムの概要

f）システム外部への表示あるいは他システムへのデータの受け渡し，

g）統計解析，

などである．g）は通常各種の統計パッケージが使用できるため，DBSではa）～f）をサポートすることが中心となるであろう．これらを中心とした本システムの概要を図3.1に示す．

(2) DB 化の対象

本システムの DB 化にあたっての基本方針は，事業所統計表や工業統計表などの公表された個別統計表を一定の視点から再編成して総合的な DB を構築するのではなく，それぞれの既存統計表をできるだけそのままの形で格納・管理することである．これは以下の理由による．

a）統計表の表章形式や分類変数・解析変数の定義，表示単位などは各統計表間で統一性がなく，また年次ごとに場当り的な整理がなされていることもある．このようなデータに対して，これらデータ要素を論理的に統一化することはきわめて困難な作業である．

b）仮にそれが実現可能であるとしても，実際に DB 化するにはそれに合わせてデータを物理的に組み替えなければならない[3.1]．これは労力的にもコスト的にも多大な作業となることが予想されるため，個別統計表を対象として DB 化に着手することが現実的な方法であろう[3.2]．

c）DB 作成者のデータ再編成時の視点は必ずしも利用者のそれと一致しない．例えば，作成年次の違う統計表を1つにまとめる場合，その形式がまったく同一であっても，接続の基準や接続方法はデータの用途や分析法によって異なってくることがある．

このほか，既存の統計表の構造に沿って統計データを定義・格納するこ

3.1）磁気テープの形で公表されているものについても統計表と同一の形式で入力されたものがほとんどである．

3.2）例えば，同種のクロスセクション・データが数時点にわたって年次ごとに存在しているならば，たとえそれらが1つの時系列データとして再編成しうるとしても，それを行うことなく，年次ごとに1つの単位として DB に格納・管理する．

とにより，

d）表頭や表側，中間の集計値などの定義に見られる統計作成者の意図を DB の中に反映することができる．

e）利用者が使い慣れた統計表のイメージのまま DB に接することを可能にする．

などの利点がある．

(3) 統計表の基本形式

2.2.1節で述べたように，統計表ではデータの構造を規定する分類変数と解析変数を，表題，表側，表頭のいずれに配置するかは自由であるが，本システムでは一般性を大きく損なうことなく，対象とする統計表の形式に以下の制約を設ける．

a）解析変数は表頭にのみ置くものとする．これは，もし表側に置くならば，1つのフィールド中に単位の異なるものが混在するという，DBS ファイルとして扱いにくい事態が生ずるためである．

b）分類変数は表頭，表側のいずれにも置くことができる．この配置は原則として対象統計表に沿うものとするが，分類変数の性質も考慮する．分類変数は地域や産業等多くの統計表に共通する定義域を持ち，しかもそこに多段階の意味的階層性があるものと，名前が同一でもその定義域が各統計表ごとに異なったり，性別のように階層性があまり重要でないものとに分けられる．前者を共通分類変数とよび表側に置き，後者を個別分類変数とよび表頭に置くことにする[3.3]．

c）表側の構造は分類変数間の直積型抽象化だけで定義する．

現実の主要統計表の多くはこれらの制約を満たしているものと思われる．

3.1.2 DBS のファイル構成

システムを構成するファイルは図3.1に示すようにメタ DB とオブジェクト DB に分けられる．それらの概要を以下に示す．

---

3.3）対象となる統計表において共通分類変数が表頭に位置する場合は，いったんその形式のまま統計表ファイルに格納後，システムで変換して表側に位置づけることとする．

(1) メタ DB (DD/D)

DD/D ファイルはシステムに存在する各種のデータの構造と内容に関する情報（メタ＝データ）を管理するもので，その対象により，

a）DD/D を含むすべてのファイルに共通する属性を管理するシステム共通辞書，

b）各統計表の構造と内容を管理する統計データ辞書，

c）分類変数の定義域を管理する共通定義域辞書，

に分けられる．これらの内容については3.3節で詳述する．

(2) オブジェクト DB

オブジェクト DB は，

a）統計データを格納する統計表ファイル，

b）データの分類体系を具体的に記述する共通定義域ファイル

から成る．

統計表ファイルは，共通分類変数の値と解析識数の値の組から成る統計データを格納する．時系列表・クロス表を問わず，各統計表ごとに1個のファイルを作成し，その形は論理的にも物理的にもできるかぎり元の統計表に沿って，データを行列形式で配置したものである．その行および列を意味づける方法は次節で述べる．

共通定義域ファイルは，各統計表に共通して存在する分類変数の実現値を収めたものであり，その分類体系を記述する[3.4]．このファイルのキー変数は分類体系の最下位のレベル，すなわち最小分類単位である．例えば，市町村→県→地方という分類体系を記述するには，市町村をキーとする定義域ファイルを構築し，県，地方はそれに従属する変数として記述すればよい（図3.2）．また，同一のファイルに複数の分類体系を記述すれば，その間の変換表としての意味を持つ[3.5]．

### 3.1.3 統計表の表現方法

本システムでは，統計表の表現にあたり格納領域の節約のために表頭定

3.4）このファイルは横山＝椿[7]のカテゴリ関係に相当する．

3.5）ただし，キーが異なる場合には別ファイルとなる．

図3.2　共通定義域ファイル（地域）

| 市区町村コード（キー） | 市区町村名 | 県コード | 郡コード | 市・郡区分 |
|---|---|---|---|---|
| ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ |
| 34100 | 広島市 | 34 | — | C |
| 34101 | 広島市中区 | 34 | — | C |
| ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ |
| 34320 | 広島県佐伯郡 | 34 | 320 | G |
| 34321 | 五日市町 | 34 | 320 | G |
| ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ |

義行列の概念を用いる[3.6]．この方式によれば，対象となる統計表の構造は個々の分類変数と解析変数のグループを表す記号とそれらの結合状態を表す論理演算子から成る定義式によって端的に表現される．定義式で用いられる論理演筆子には，＋（論理和），＊（論理積）があり，演算の優先順位は括弧，(,) によって指示される．

これらを簡単な例によって示す．

(1)　地域別労働力状態（図3.3 a）

表側の分類変数は地域(Ｒ)，表頭の分類変数は性別(Ｓ)と労働力状態(Ｗ)であり，解析変数は人口(Ｐ)だけである．表側は一元分類であるから，Ｒだけで表す．表頭の分類は性別と労働力状態の二元分類であるからＳ＊Ｗ で表し，これの各カテゴリに対して人口の測定値が存在するから，定義式は Ｓ＊Ｗ＊Ｐ となる．ＳおよびＷのカテゴリ値を示すコードとＰを示すコードとしてそれぞれ１～２，１～３，１を与えた表頭定義行列を

3.6）表頭定義行列およびその導出方法については，本学部横山和典教授が「表頭定義行列とその導出アルゴリズムについて」と題して，本経済論叢に後日発表の予定である．

**図3.3a** 地域別，労働力状態

| 地　域 | 男 | | | 女 | | |
|---|---|---|---|---|---|---|
| | 就業者 | 完全失業者 | 非労働力 | 就業者 | 完全失業者 | 非労働力 |
| | | | | | | |

**図3.3b** 表頭定義行列（Ⅰ）

| | $H_1$ | $H_2$ | $H_3$ | $H_4$ | $H_5$ | $H_6$ |
|---|---|---|---|---|---|---|
| S | 1 | 1 | 1 | 2 | 2 | 2 |
| W | 1 | 2 | 3 | 1 | 2 | 3 |
| P | 1 | 1 | 1 | 1 | 1 | 1 |

**図3.3c** 地域別産業別，商店数，従業者数

| 地域産業 | 経営組織 | | | | 性別 | |
|---|---|---|---|---|---|---|
| | 法人 | | 個人 | | 男 | 女 |
| | 商店数 | 従業者数 | 商店数 | 従業者数 | 従業者数 | 従業者数 |
| | | | | | | |

**図3.3d** 表頭定義行列（Ⅱ）

| | $H_1$ | $H_2$ | $H_3$ | $H_4$ | $H_5$ | $H_6$ |
|---|---|---|---|---|---|---|
| Or | 1 | 1 | 2 | 2 | — | — |
| S | — | — | — | — | 1 | 2 |
| A | 1 | 2 | 1 | 2 | 2 | 2 |

（注）—は値が存在しないことを示す.

図3.3bに示す.

(2) 地域別産業別商店数，従業者数（図3.3c）

表側の分類変数は地域(R)と産業中分類(Im)，表頭は経営組織(Or)ごとに解析変数として商店数と従業者数（グループ名A）を持ち，さらに性別(S)の従業者数が加えられている．表側は二元分類であるから R＊Im で表す．表頭は Or の示す分類とSの示す分類が併置されており，前者についてはAの全体が，後者についてはその一部が存在している．この場合，それぞれを Or＊A，S＊A と表し，表頭全体としては両者を結びつけて (Or+S)＊A と表す．Or とSのカテゴリ値のコード，ならびに個々の解析変数を示すコードをともに1～2と与えた表頭定義行列を図3.3dに示す.

### 3.2 メタ＝データベース

メタ＝データとして以下の内容を管理する．またこれらを操作するための，入力，編集，削除，表示の各機能を提供する．

#### 3.2.1 システム共通辞書

(1) ファイル辞書

名前や種別，およびその他の記述的情報等，DBS に含まれるすべてのファイルの共通的属性を管理する．

(2) 項目辞書

DD/D ファイルのデータ項目について，名前，意味内容，長さ等の属性を管理する。これに基づいてメタ＝データの入力，編集，表示等の操作を行う。

#### 3.2.2 統計データ辞書

(1) 統計表ファイル辞書

統計表ファイルの種別に関する情報，作成年次，その他の記述的情報，および表頭・表側定義式で表した統計表の表現，表頭定義条件の有無等を記述する．

(2) 分類変数辞書

統計表ファイルに含まれる分類変数の名前，意味内容，配置，定義域ファイル名等を記述する．

(3) 解析変数辞書

統計表ファイルに含まれる解析変数の名前，意味内容，測定単位，集計方法等を記述する．

(4) 表頭定義コード辞書

表頭定義式は各分類変数名と解析変数の集合名とによって表わされ，実際の統計表の表頭との対応は分類変数の実現値と各解析変数をそれぞれ任意にコード化したものの組でとられている．ここではそれらのコードの意味を記述し，表頭の分類変数については定義域となる．

(5) 表頭定義条件辞書

表頭定義式から生成される表頭定義行列は，表頭の各分類変数の定義域

**図3.4a**　統計表ファイル辞書

| ファイル名 | 識別 | 内　容 | 年次 | 種類 | 表側仕様 | 表頭仕様 | 定義条件の有無 | 注記 |
|---|---|---|---|---|---|---|---|---|
| JI56 | 01010 | 市区町村別事業所統計表 | 1981 | 事業所統計 | R10＊I10 | (KIND+ORG*SIZE)*A | Y | |
| KO51 | 01020 | 工業統計表（市町村編） | 1976 | 工業統計 | R11＊I20 | A | N | |
| KO52 | 01030 | 工業統計表（市町村編） | 1977 | 工業統計 | R11＊I20 | A | N | |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ | ・ | |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ | ・ | |
| KO58 | 01090 | 工業統計表（市町村編） | 1983 | 工業統計 | R11＊I20 | A | N | |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ | ・ | |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ | ・ | |

**図3.4b**　分類変数辞書

| ファイル名 | 識別 | 分類変数コード | 内　容 | 配置 | 定義域ファイル名 |
|---|---|---|---|---|---|
| JI56 | 0101000010 | R10 | 地域（行政区分） | S | AREA |
| JI56 | 0101000020 | I10 | 産業大分類 | S | INDUST |
| JI56 | 0101000030 | KIND | 従業形態 | H | DDHEADC |
| JI56 | 0101000040 | ORG | 経営組織 | H | DDHEADC |
| JI56 | 0101000050 | SIZE | 従業者規模 | H | DDHEADC |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ |
| KO58 | 0109000010 | R11 | 地域（行政区分） | S | AREA |
| KO58 | 0109000030 | I20 | 製造業中分類 | S | INDUST |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | |

全体と解析変数群全体を用いた，いわゆる標準化されたものであるが，現実の統計表ではそれらの一部が欠落したものが多い．このため，現実の統計表の表頭定義との対応をとるため，その欠落部分を各変数値に対する論理条件式でここに表現する[3.7]．

(6)　表頭定義行列辞書

統計表の表頭定義式と(5)で記述された論理条件式にしたがって，(4)で

3.7）欠落条件を論理条件式だけでは十分に表現できない場合は，表頭定義行列辞書に格納後，DD/D の編集機能により修正する．

図3.4c　解析変数辞書

| ファイル名 | 識別 | 解析変数コード | 内　容 | 測定単位 | 集計方法 |
|---|---|---|---|---|---|
| JI56 | 0101000010 | EST | 事業所数 | — | — |
| JI56 | 0101000020 | PSN | 従業者数 | 人 | — |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ |
| KO58 | 0109000010 | EST1 | 事業所数（合計） | — | — |
| KO58 | 0109000020 | EST2 | 事業所数（内従業者10～299人） | — | — |
| KO58 | 0109000030 | EST3 | 事業所数（内従業者300人以上） | — | — |
| KO58 | 0109000040 | PSN | 従業者数 | 人 | — |
| KO58 | 0109000050 | WAGE | 現金給与総額 | 万円 | — |
| KO58 | 0109000060 | MATERI | 原材料使用額等 | 万円 | — |
| KO58 | 0109000070 | OUTPUT | 製造品出荷額等 | 万円 | — |
| KO58 | 0109000080 | ADDVL | 粗付加価値額 | 万円 | — |
| KO58 | 0109000090 | RESTATE | 有形固定資産（従業者10人以上） | 万円 | — |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ |

定義されたコードの値を組み合わせた表現により，表頭定義行列を記述する[3.8]．

これらを図3.4a～3.4fに例示する．

### 3.2.3　共通定義域辞書

(1)　定義域ファイル辞書

共通定義域ファイル名とそのキー変数（最小分類単位の変数）に関する情報を記述する．

(2)　定義域辞書

共通定義域ファイルに含まれる分類変数の名前，意味内容，値の範囲，コード表名等を記述する．

(3)　定義域変数コード表

共通定義域ファイルの分類変数の値とその意味内容の対応を記述する．

これらを図3.5a～3.5cに例示する．

3.8）これは3.1.3節で示した表頭定義行列に表側の定義を加えたものである．

**図3.4d** 表頭定義コード辞書

| ファイル名 | 識別 | 変数コード | 値 | 内　容 | 解析変数コード |
|---|---|---|---|---|---|
| JI56 | 0101000010 | R10 | * | — | — |
| JI56 | 0101000020 | I10 | * | — | — |
| JI56 | 0101000030 | KIND | 0 | 総数 | — |
| JI56 | 0101000040 | KIND | 1 | 個人業主 | — |
| JI56 | 0101000050 | KIND | 2 | 家族従業者 | — |
| JI56 | 0101000060 | KIND | 3 | 雇用者（総数） | — |
| JI56 | 0101000070 | KIND | 4 | 内常雇 | — |
| JI56 | 0101000080 | ORG | 1 | 民営 | — |
| JI56 | 0101000090 | ORG | 2 | 非民営 | — |
| JI56 | 0101000100 | SIZE | 0 | 総数 | — |
| JI56 | 0101000110 | SIZE | 1 | 従業者1－4人 | — |
| JI56 | 0101000120 | SIZE | 2 | 従業者5－9人 | — |
| JI56 | 0101000130 | SIZE | 3 | 従業者10－19人 | — |
| JI56 | 0101000140 | SIZE | 4 | 従業者20－29人 | — |
| JI56 | 0101000150 | SIZE | 5 | 従業者30人以上 | — |
| JI56 | 0101000160 | A | 1 | 事業所数 | EST |
| JI56 | 0101000170 | A | 2 | 従業者数 | PSN |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ |
| KO58 | 0109000010 | R11 | * | — | — |
| KO58 | 0109000020 | I20 | * | — | — |
| KO58 | 0109000030 | A | 1 | 事業所数（合計） | EST1 |
| KO58 | 0109000040 | A | 2 | 事業所数（内従業者10～299人） | EST2 |
| KO58 | 0109000050 | A | 3 | 事業所数（内従業者300人以上） | EST3 |
| KO58 | 0109000060 | A | 4 | 従業者数 | PSN |
| KO58 | 0109000070 | A | 5 | 現金給与総額 | WAGE |
| KO58 | 0109000080 | A | 6 | 原材料使用額等 | MATERI |
| KO58 | 0109000090 | A | 7 | 製造品出荷額等 | OUTPUT |
| KO58 | 0109000100 | A | 8 | 粗付加価値額 | ADDVL |
| KO58 | 0109000110 | A | 9 | 有形固定資産（従業者10人以上） | RESTATE |
| ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ |

（注）表頭仕様に現れる表側変数については値を＊とする.

図3.4e　表頭定義条件辞書

| ファイル名 | 識別 | 条件名 | 条　件 |
|---|---|---|---|
| J I 56 | 0101000010 | M1 | ('0'<KIND AND A='1') OR |
| J I 56 | 0101000020 | M2 | (ORG='2' AND '0'<SIZE) |
| ・ | ・ | ・ | ・ |

図3.4f　表頭定義行列辞書

| ファイル名 | 識別 | 変数コード | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 | H9 | H10 | H11 | H12 | H13 | H14 | H15 | H16 | H17 | H18 | H19 | H20 | H21 | H22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| J I 56 | 0101000010 | R10 | * | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| J I 56 | 0101000020 | I10 | - | * | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| J I 56 | 0101000030 | KIND | - | - | 0 | 0 | 1 | 2 | 3 | 4 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| J I 56 | 0101000040 | ORG | - | - | - | - | - | - | - | - | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | |
| J I 56 | 0101000050 | SIZE | - | - | - | - | - | - | - | - | 0 | 0 | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 | 5 | 5 | 0 | 0 |
| J I 56 | 0101000060 | A | - | - | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| ・ | ・ | ・ | | | | | | | | | | | ・ | | | | | | | | | | | |
| ・ | ・ | ・ | | | | | | | | | | | ・ | | | | | | | | | | | |
| K O 58 | 0109000010 | R11 | * | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| K O 58 | 0109000020 | I20 | - | * | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| K O 58 | 0109000030 | A | - | - | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | - | - | - | - | - | - | - | - | - | - | - |
| ・ | ・ | ・ | | | | | | | | | | | ・ | | | | | | | | | | | |

図3.5a　定義域ファイル辞書

| ファイル名 | 識別 | 内　容 | キー変数コード | キー変数の内容 | キー変数長 | キーラベル項目名 |
|---|---|---|---|---|---|---|
| AREA | 90010 | 地域分類 I (行政区分) | ACODE | 都道府県市区町村コード | 5 | ANAME |
| INDUST | 90020 | 産業分類 (JIS) | ICODE | 産業分類コード (JIS) | 6 | INAME |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ |
| ・ | ・ | ・ | ・ | ・ | ・ | ・ |

## 3.3　データの検索

検索は広く解釈すると，格納された統計表から必要とする集計水準と形式を備えた統計表あるいは資料行列を導出する操作であるといえる．その過程は

図3.5b 定義域辞書

| ファイル名 | 識別 | 変数コード | 内　容 | コード表名 | 値の範囲 |
|---|---|---|---|---|---|
| AREA | 9001000010 | ACODE | 地域コード(キー) | AREA | 01000−47999 |
| AREA | 9001000020 | KEN | 県コード | DDCODE | 01−47 |
| AREA | 9001000030 | GUN | 郡コード | DDCODE | 300−990 |
| AREA | 9001000040 | RC | 市・郡区分 | DDCODE | 1−2 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |
| INDUST | 9002000010 | ICODE | 産業コード | INDUST | 00000−149999 |
| INDUST | 9002000020 | LMS | 分類水準 | DDCODE | L−S |
| INDUST | 9002000030 | LCODE | 大分類コード | DDCODE | 01−14 |
| INDUST | 9002000040 | MCODE | 中分類コード | DDCODE | 01−99 |
| INDUST | 9002000050 | SCODE | 小分類コード | DDCODE | 011−969 |
| . | . | . | . | . | . |
| . | . | . | . | . | . |

図3.5c 定義域変数コード表

| ファイル名 | 識別 | 変数コード | 値 | 内　容 |
|---|---|---|---|---|
| AREA | 9001000010 | KEN | 01 | 北海道 |
| AREA | 9001000020 | KEN | 02 | 青　森 |
| . | . | . | . | |
| AREA | 9001000330 | KEN | 33 | 岡　山 |
| . | . | . | . | |
| AREA | 9001001480 | RC | 1 | 市　部 |
| AREA | 9001001490 | RC | 2 | 郡　部 |
| . | . | . | . | . |
| . | . | . | . | . |
| INDUST | 9002000010 | LMS | L | 大分類 |
| INDUST | 9002000020 | LMS | M | 中分類 |
| . | 9002000030 | LMS | S | 小分類 |
| . | 9002000040 | LCODE | 01 | 農　業 |
| . | 9002000050 | LCODE | 02 | 林業，狩猟業 |
| . | 9002000060 | LCODE | 03 | 漁業，水産養殖業 |
| . | . | . | . | . |
| . | . | . | . | . |

a）該当する範囲のデータの既存の統計表からの抜き出し，

b）それらの必要な集計水準への集約，

c）検索結果の表示や求める形式への変換と各利用者ファイルへの保存，に大別されるが，ここではa）を中心に考察する[3.9]．

### 3.3.1 範囲の検索

範囲の検索は統計表ファイル全体で構成される資料行列から部分行列を取り出す操作であり，これは表頭と表側のそれぞれの構造にしたがって行う．

(1) 表頭

表頭の検索は統計表の表頭定義行列の特定の列を決定し，それに対応する統計表ファイルの該当する列を取り出すことであり，以下のように行う．まず，各分類変数については分類カテゴリを表すコードを，解析変数群については個々の解析変数を表すコードを，値，またはその範囲で指定する[3.10]．次に，これに基づいて，システムが表頭定義式にしたがって仮想的な表頭定義行列を展開し，その結果と統計表ファイルの表頭定義行列とが重複するところを該当する列と認定する（例3.1）．

(2) 表側

表側の検索は共通分類変数の範囲を定め，統計表ファイルからそれに該当する行を取り出すことである．統計表ファイルにおける表側の分類変数は階層的意味関係を構成するもので，かつそれぞれの分類体系の最下層に位置づけられている．したがって範囲の指定については，単に格納されたレベルだけでなく利用者の必要とするそれぞれの分類水準に柔軟に対応しなければならない．このため表側の検索は以下の手順で行う．まず，利用者は共通分類変数がキー変数として存在する定義域ファイルを対象に，そこに含まれる任意の変数について値または値の範囲を指定し，これにより

---

3.9）ここでは個別統計表からの検索のみを検討対象とし，形式の異なる別種の統計表からの検索と結果の統合の問題については検索結果の編集機能で扱うものとする．

3.10）それぞれの変数における値または値の範囲の間は論理和で結ばれ，各変数ごとの条件の間は論理積で結ばれる．

**例3.1** 表頭の検索（事業所統計表）

《総数と事業所の規模別の従業者数を求める》

① 分類変数の指定

KIND（従業形態）：0（総数）

ORG（経営組織）：ALL（全体）

SIZE（従業者規模）：ALL（全体）

② 解析変数の指定

A（解析変数のグループ名）：2（従業者数）

（個々の検索条件は論理積で結びつけられて全体の条件となる.）

③ 検索条件から導出された表頭定義行列

| 変数コード | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | D10 | D11 | D12 | D13 | D14 | D15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| R10 | * | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| I10 | - | * | - | - | - | - | - | - | - | - | - | - | - | - | - |
| KIND | - | - | 0 | - | - | - | - | - | - | - | - | - | - | - | - |
| ORG | - | - | - | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| SIZE | - | - | - | 0 | 1 | 2 | 3 | 4 | 5 | 0 | 1 | 2 | 3 | 4 | 5 |
| A | - | - | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

④ 検索結果（③と図3.4 f 事業所統計表の表頭定義行列との重複部分）

| 変数コード | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | V10 |
|---|---|---|---|---|---|---|---|---|---|---|
| R10 | * | - | - | - | - | - | - | - | - | - |
| I10 | - | * | - | - | - | - | - | - | - | - |
| KIND | - | - | 0 | - | - | - | - | - | - | - |
| ORG | - | - | - | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| SIZE | - | - | - | 0 | 1 | 2 | 3 | 4 | 5 | 0 |
| A | - | - | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

定義域ファイル中のキー変数の値ベクトルを定める．システムは，これと統計表ファイル表側の共通分類変数の値の積集合をとり，該当する行を確定する．表側が複数個の共通分類変数による直積型抽象化となっている場合は，それら個々の変数についてそれぞれの定義域ファイルのキー変数の値ベクトルを定め，それらの直積を導出したうえで，統計表ファイルとの積集合を求める（例3.2）.

**例3.2** 表側の検索（工業統計表）

《広島県の市部，一般機械器具製造業および電気機械器具製造業について求める》

① 共通定義域ファイルにおける範囲の指定

AREA（地域）：KEN＝34 AND RC＝1

INDUST（産業）：MCODE＝34 OR MCODE＝35

② キー変数の値ベクトル

a 地域

| |
|---|
| 広 島 市 |
| 中　　区 |
| 東　　区 |
| ・ |
| ・ |
| 呉　　市 |
| ・ |
| ・ |
| 東広島市 |

b 産業

| |
|---|
| 一般機械器具製造業 |
| ボイラ・原動機製造業 |
| 農業用機械製造業 |
| ・ |
| ・ |
| 電気機械器具製造業 |
| ・ |
| ・ |
| ・ |

③ 統計表ファイルの表側

| R11 | I 20 |
|---|---|
| ・ | ・ |
| ・ | ・ |
| 広島市 | 製造業 |
| ・ | 食料品・たばこ製造業 |
| ・ | ・ |
| ・ | ・ |
| ・ | 一般機械器具製造業 |
| ・ | 電気機械器具製造業 |
| ・ | ・ |
| ・ | ・ |
| 中　区 | 製造業 |
| ・ | ・ |
| ・ | ・ |

④ 検索結果（②のａ，ｂの直積と③の積集合）

| R11 | I 20 |
|---|---|
| 広 島 市 | 一般機械器具製造業 |
| 広 島 市 | 電気機械器具製造業 |
| 中　　区 | 一般機械器具製造業 |
| 中　　区 | 電気機械器具製造業 |
| ・ | ・ |
| ・ | ・ |
| ・ | ・ |
| 東広島市 | 電気機械器具製造業 |

(3) 利用者定義の分類体系

共通定義域ファイルはシステム全体で共有されるべき性質のもので，シ

ステム管理者が設定・提供する．したがって，そこに記述される分類体系は既存の統計表を基礎にした共通性の高いもので，しかもかなり固定的なものである．この分類体系以外に，利用者がそれぞれ固有の概念に基づく独自の分類体系や，アド・ホックな分類を必要とする場合もある．これに応ずるために，共通定義域ファイルと同一のキー変数を基礎として，それぞれの分類体系を記述する定義域ファイルとそれを管理する定義域辞書を各利用者ごとに構築する手段を提供している．これらのファイルはシステムのそれと同じ構造を持っているために，検索過程における処理は最小限の変更にとどまる．

(4) 時系列

統計表ファイルが表頭または表側に時間を表す分類変数を配置した時系列表であれば，その検索において時間は他の分類変数と同様に扱うことができる[3.11]．他方調査年次ごとに作成されたクロス表の場合は，各ファイルごとに該当する範囲を抜き出したうえで，それらの結果を連結しなければならない．これには検索範囲の統計表ファイルの表頭・表側の仕様が同一であることが前提となるが，たとえそれが満足されたとしても，

a）異時点間で測定単位や表示単位が異なる時の換算，

b）町村合併のような定義域の内容の変化に応じた調整，

などの問題が残り，さらにデータによっては利用者の求めにしたがった調整が必要とされる場合もある．このため，システムで行う連結処理は，時間を表す分類変数を各検索結果の列に追加し，それをキーとして併合するだけの形式的なものに限定し，データの内容にかかわる処理は利用者の手に委ね，編集・加工機能で扱う[3.12]．

(5) 既集約データ

本システムでは統計表に掲載された値をそのまま格納するため，既集約

3.11）時間はその性質上共通分類変数として表側に位置することが望ましい．

3.12）ただし定義域の内容の変化は，共通定義域ファイルに年次ごとの分類変数の値を記述し，それらをキーを介して対応させることで検索処理に反映できる．

データが DB 中に存在していることが多い．また秘匿数値も数多く含まれることから，集約により上位レベルの値を求めるよりもそれらを直接的に検索するのが得策であろう．既集約データを示すカテゴリは，表頭では分類変数の1カテゴリ値として表頭定義コード辞書に記述されており，表側では共通定義域辞書中の分類変数として存在しているから，これらを用いれば容易に検索できる．

3.3.2 集約

既集約データが存在しない場合はその水準を指定して集約データを導出する．表側での集約は再分類と集計の2つの過程に分けられるが，検索機能では前者のみを扱い，後者は表頭変数間の演算により求める表頭での集約とともに，編集・加工機能で扱われる．再分類は集計水準を表す分類変数の値を各統計集団に割り当てる操作であり，共通定義域ファイルまたは利用者の定義域ファイルに存在する該当レベルの分類変数を取り出すことで行われる．他方集計は，秘匿数値が含まれていたり，解析変数の特性以外に利用者のアド・ホックな要求も考慮しなければならないため，システム側では自動的に行わず，集計方法に関する情報の提供を行うにとどめる．

3.3.3 表示と保存

資料行列として取り出された検索結果は一時的なファイルに収められており，表側の分類変数の値により行が，表頭定義行列の部分行列により列がそれぞれ意味づけられている．これに対するその後の処理としては，

a）そのままの形で表示するか，あるいは利用者ファイルへ保存する．

b）加工･分析用の他システムで利用しうる形式のファイルに変換する．

c）統計表として指示された形で表現する．

d）引き続き編集・加工操作を加える．

などがあるが，検索機能としては，表示・保存・システム外ファイルへの変換を持たせることとする．

## 4. むすび

現在，本稿で述べた設計に基づいた試作システムを SAS マクロ言語に

より構築中である．DD/D の管理およびデータの検索については既にその一部が稼働しており，時系列検索や集約におけるシステムの機能の強化，およびデータの編集・加工機能，統計表への表現などが今後の課題として残されている．

(86. 5 .31)

## 参 考 文 献


[1] Chan, P. and A. Shoshani, "SUBJECT: A Directory Driven System for Organizing and Accessing Large Statistical Databases," Proceedings of the 7th International Conference on Very Large Data Bases, 1981, pp. 553-63.

[2] Codd, E. F., "A Relational Model of Data for Large Shared Data Banks," Communications for the ACM, Vol. 13, No. 6, 1970, pp. 377-87.

[3] Kent, W., "Choices in Practical Data Design," Proceedings of the 8th International Conference on Very Large Data Bases, 1982, pp. 165-80.

[4] 石国直治，『研究と実務のための統計学教材』，柳盛社，1978年.

[5] 小菊一三，神尾視教，森元 逞，「リレーショナルデータベースを用いた統計検索機能と実現方式」，情報処理学会データベース・システム研究会資料 42−2，1984年.

[6] 佐藤和夫，穂高良介，「多目的統計データベース管理システムの設計」，情報処理学会「アドバンスト・データベース」シンポジウム予稿集，1984年.

[7] 横山和典，椿 康和，「地方統計のデータベース化について」，広島大学経済学部紀要 年報経済学，第5巻，1984年.

[8] 横山和典，椿 康和，「社会調査の管理および分析支援システム SSMS」広島大学経済論叢，第9巻，第2号，1985年.