

実世界データベース

2L-4

(3) 情報表現

池田 一夫、倉科 周介、大橋 誠
東京都立衛生研究所

1. はじめに

国や地方自治体では、社会・経済の現状把握や政策立案の基礎資料として、各種の官庁統計が作成されている。著者らは、これらの情報を効果的に活用するためのデータベースの構築概念とその表記法を提案した¹⁻³⁾。

統計情報はその性質上、膨大な数値データより構成される。観測対象の全貌を的確に把握するためには数値群全体の精密な解析と適切な表現が必要である。膨大なデータのコンピュータグラフィックス(CG)による可視化は、そのための一助として多大な効果が期待される。

本論文では、先に提案した概念を基に作成した疾病構造データベースシステム(SAGE)で発生する情報(人口動態統計情報)を例として情報の効果的な表現法について論ずる。

2. 人口動態統計

わが国では、1899年から中央集査による人口動態統計が実施されている。このデータを利用することにより約100年にわたる日本人の死亡現象を解析することが可能である(1944-46年の一時期を除く)。人口動態統計死亡数表は、性別・年齢(5歳)階級別・死因別にまとめられた数値が、表頭を年齢、表側を死因・性別とする直交分割表の形式に編集されている(図1)。

国際基本分類番号 International Detailed List Code	死 因 ・ 性 Causes of Death and Sex	年齢						
		総 数 Total	0 歳 Year	1	2	3	4	0~4
総 数 Total	男 M.	793 014	6 265	1 071	593	495	392	8 816
	女 F.	428 094	3 434	606	328	297	235	4 900
(001-009)	感染症及び寄生虫症	11 336	183	55	30	24	16	308
	腸管感染症	753	25	7	2	-	-	34
001	コレラ	304	14	3	2	-	-	15
		449	74	25	11	12	5	127

図1. 人口動態統計死亡数表

3. 世代マップ

著者らは、以下のようなデータベース構築手法を報告した¹⁻³⁾。①実世界と同様の設定仮想空間上の該当するべき位置に既存のデータを配置する。②配置した個々の情報を分解・再配列する。③実世界を復元する。④実世界を観測する。この手法を人口動態統計に適用すれば「世代マップ」が得られる(図2)。

世代マップの縦軸は出生世代(3年世代)、横軸は年代(3年)である。個々のセルは死亡数を表している。世代マップを利用すれば、各種人口現象の分布パターンを、容易に観測することができる。

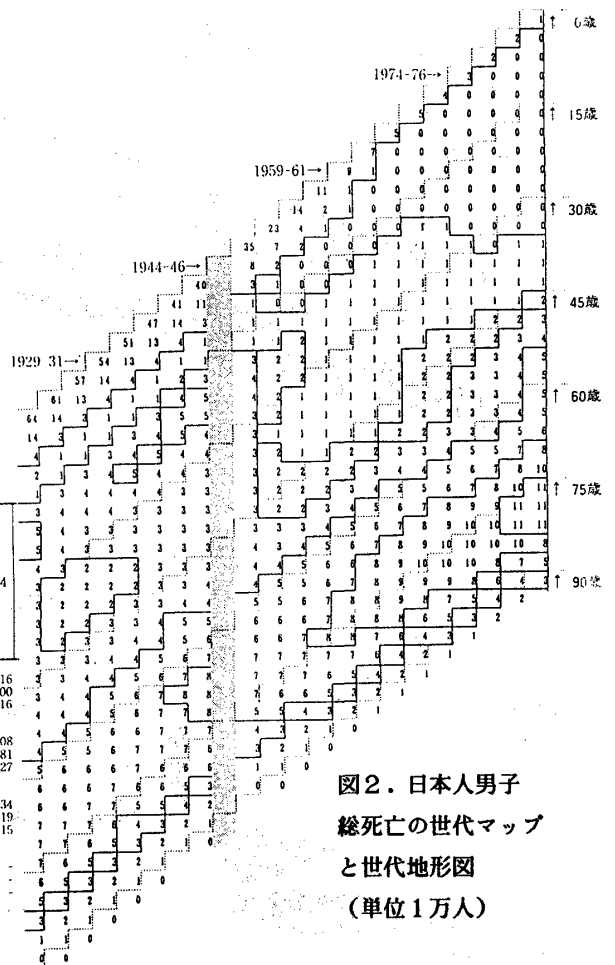


図2. 日本人男子
総死亡の世代マップ
と世代地形図
(単位1万人)

世代マップの上に、等死亡数線（通常の地形図の等高線と類似の概念）を描いたものが世代地形図である。

4. 世代地形図のCG表現（1）

CGを利用すれば対象の認識が容易になる。世代地形図の分布特性把握の一助として地形図の色分けを試みた（図3-1、3-2）。

色分けにはPC-9801を用い、モードを変えて、8色および16色で行った。8色モードでも、利用法を工夫すれば一定の効果をえられるが、16色表示の方が格段に情報表現力が高いことが明かとなった。

さらに、16色表示において色調の変化が情報の認識に及ぼす影響について検討した。その結果、数値の低い部分を青色とし、数値の増大につれて水色、緑色、黄土色、茶色と段階的に変化させると効果的であった。

5. 世代地形図のCG表現（2）

①地形図を各行（世代）および各列（時代）ごとに切断し、②切断面の形状を観測・分析し、③疾病動向の予測を行う。この様な手順で、世代地形図を巨視的な疾病の推移の認識と動向予測に利用することが可能である。

具体的には、①横軸に年齢、縦軸に死亡数を取り、②死亡数の世代別切断面を描き、③世代別死亡数が上昇傾向にあるのか、下降傾向にあるのかを観測し、④その情報を基礎に予測を行う。図4に男子肝硬変死亡の世代別切断面を示す。この図より、肝硬変による死亡の今後の推移を予測することが可能である。

6. むすび

統計情報の解析において、CGは大きな威力を発揮する。システム化に当たっては、統計情報の効果的なCG化に配慮することが肝要である。

統計情報は印刷物として公表されることが一般的であった。ために、情報処理技術を駆使したその高次利用を意図しても、データ入力に要する膨大な労力と経費など、多大な困難に直面することが多い。情報化社会といわれる現在、統計情報の新たな公表形態も考慮すべき時期にきていると考える。

参考文献

- 1) 倉科：病気の姿をデータで読む，公衆衛生，53，493（1989）
- 2) 倉科，池田，大橋：実世界データベース（1）基礎的検討，情報処理学会第40回全国大会（1990）
- 3) 池田，倉科，大橋：実世界データベース（2）生成，概念，表記法，情報処理学会第40回全国大会（1990）

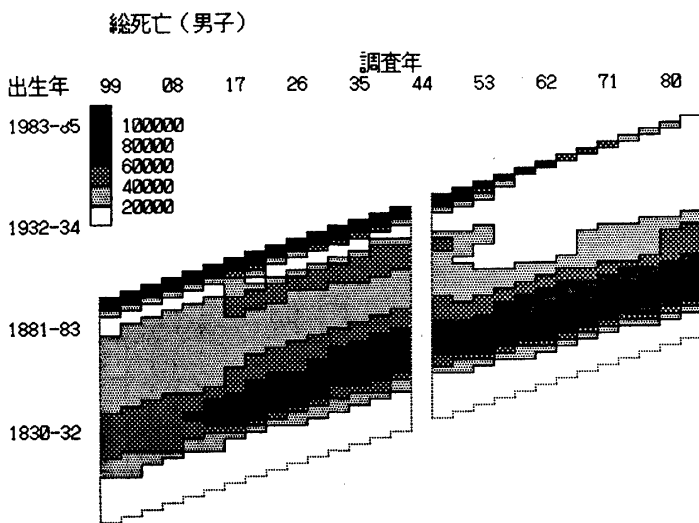


図3-1. 男子総死亡の世代地形図

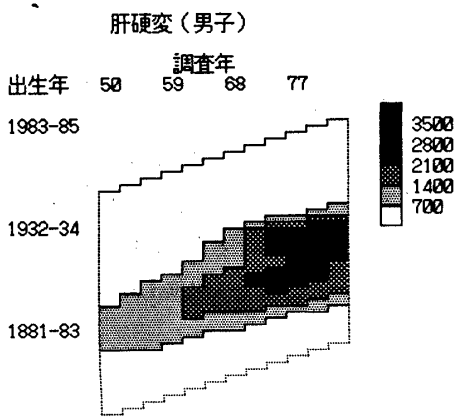


図3-2. 男子肝硬変死亡の世代地形図

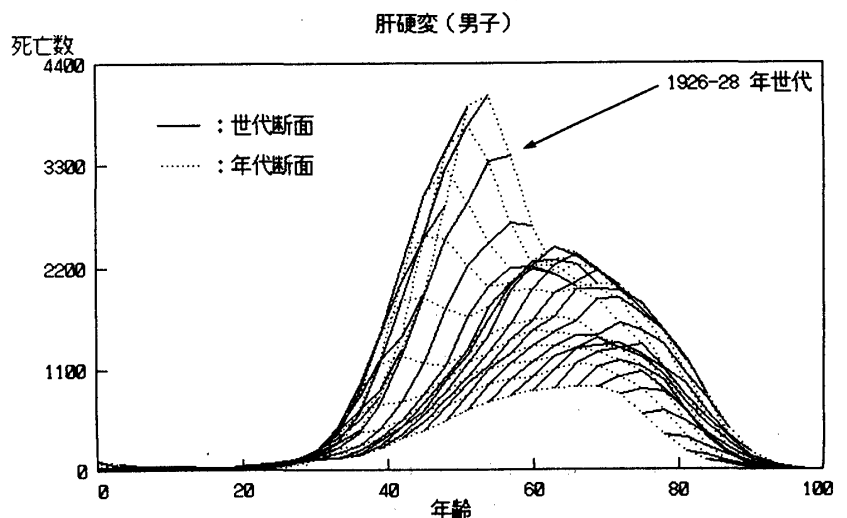


図4. 男子肝硬変死亡の世代別・年代別断面