

厚生労働科学研究研究費補助金
統計情報高度利用総合研究事業
患者調査の客体設定の在り方に関する研究

平成14年度 総括研究報告書

主任研究者 吉村 功
(東京理科大学工学部教授)

平成15 (2003) 年 3月

目次

第1章	はじめに	1
1.1	研究概要	1
1.2	研究目的	2
1.3	研究計画	2
第2章	データ	4
2.1	使用データとファイル	4
2.2	データの件数	5
第3章	解析方針	8
3.1	現在の標本抽出法	8
3.2	現状の問題点と解析方針	8
第4章	コレスポネンス分析の理論	10
4.1	解析における判定・評価	10
4.2	解析の算法	11
第5章	コレスポネンス解析の結果と考察	12
5.1	特異値, カイ二乗値, 寄与率, 累積寄与率の結果	12
5.2	座標値の散布図	14
5.3	各特異値に対応する疾病と診療科目	43
5.4	考察	44
第6章	患者数の推定精度についての試算	46
6.1	現状の患者分布の特徴と推定法の問題点	46
6.2	層化法変更の影響の試算	46
6.2.1	診療科目数の分布	46
6.2.2	層化の提案	47

6.2.3	母集団と有病率の算出	48
6.3	試算の結果	52
6.3.1	推定方法の違いによる推定精度の評価	52
6.3.2	層ごとの抽出率を変化させた場合の推定精度の評価	53
6.4	まとめ	56
第7章	結論	57
付録A	疾病中分類と診療科目番号の一覧	58
付録B	一般診療所の患者数推定方法	65

第1章 はじめに

1.1 研究概要

本報告書は、厚生労働科学研究費補助金を受けて行う「患者調査の客体設定の在り方に関する研究」の平成14年度の総括報告である。厚生労働科学研究費補助金交付申請書に記載した研究の概要を下記に示す。

患者調査は、医療施設（病院及び診療所）を利用する患者数を把握するとともに、疾病及び受療の種類並びに在院日数など国民の受療の実態を明らかにし、医療行政の基礎資料を得るための指定統計として、3年に一度行われている。この患者調査は、単に医療を受ける国民の性・年齢や疾病の有病や受療の状態を明らかにするのみならず、各都道府県において二次医療圏を単位として検討・策定される地域医療計画において、基準病床数の算定や各種疾病対策の検討の基礎資料として非常に重要な調査である。

患者調査は、現在、二次医療圏単位で、医療機関の種類、規模等で層化無作為抽出し、当該医療機関を利用した患者を客体として、一部には居住地や疾病等について詳細な調査を行い、残りの患者は性・年齢のみの簡易な調査を行っているが、医療機関の機能は多様であるため、ある特定の機能を有する医療施設が抽出されるか否かにより、結果に大きな偏りが生じているおそれがある。

本研究では、病院について、患者調査等のデータを用いて、医療施設の専門分化による患者特性の偏り、地理的な要因による患者の流出入の差等が与える影響を明らかにするとともに、それに基づいて、最適の層化・抽出方法を明らかにする。

1.2 研究目的

厚生労働科学研究費補助金交付申請書に記載した研究目的を下記に示す。

都道府県では、医療法に基づき二次医療圏を設定し、当該地域への入院患者の流入や流出も考慮しつつ、当該地域における基準病床数を算定し、さらに特定の病院が果たすべき機能についても記載した医療計画を策定する必要がある。

医療計画策定において、医療需要の把握については患者調査が担う部分が多く、二次医療圏毎についてのデータも非常に重要である。しかし、現状では層化・抽出調査であるため、例えば二次医療圏で唯一の二次救急病院や小児科標榜病院が客体から漏れるなどの不都合が生じている可能性がある。

一方、調査の費用対効果の観点、及び調査協力を依頼する医療機関の調査記入者負担も非常に大きく、出来る限り軽減する必要もある。

患者調査は3年に1度行われる調査であり、次々回は平成17年10月に予定されている。これに向けて費用対効果が高く、かつ、少ない患者記入者負担を実施するための、層化、抽出方法を確立するものである。

1.3 研究計画

厚生労働科学研究費補助金交付申請書に記載した、各年度ごとの研究計画を次に示す。

■ 1年目

- ① 当該データ提供を受けるために必要な目的外使用申請等の手続きを行う。
- ② 統計データを受領して、内容を確認する。
- ③ 受領した統計データを解析して、医療施設の特性等について検討を行う。
- ④ データ解析と並行して、研究班会議を開催し、以下の点などについて検討する。
 - ・ 医療施設ごとの患者特性の偏りについての作業仮説を明確化する。
 - ・ 最適な客体設定（調査対象となる医療施設の抽出方法と当該医療施設における患者の抽出方法）を検討する。

■ 2年目

- ① 考案した客体設定法の実行可能性について、次の3点を試みる。
 - ・ 計算機上でシミュレーションを行う。

- ・ 都道府県の担当者や医療関係者等から実行可能性についての意見聴取を行う。
- ・ 実行可能性に関する試行調査を実施する。

② 2年間の研究成果を総括し，報告書を作成する。

第2章 データ

2.1 使用データとファイル

厚生労働省から目的外使用を許可された調査票の名称を表 2.1, 表 2.2 にまとめた。ファイルはいずれも所要の事項を磁気テープに転写したもので、テキスト形式である。

患者調査の実施方法に伴い調査票の名称が異なっている。平成 5 年の「医療施設静態調査調査票 病院 (I)」, 平成 8 年の「医療施設静態調査 病院票 (I)」, 平成 11 年の「医療施設静態調査 病院票」がその例で。内容は本質的に同じである。

■ 医療施設静態調査

表 2.1: 医療施設静態調査票一覧

		平成 5 年	平成 8 年	平成 11 年
医療施設静態調査調査票	病院 (I)	○		
	病院 (II)	○		
	一般診療所	○		
	歯科診療所	○		
医療施設静態調査 病院票 (I)			○	
医療施設静態調査 病院票 (II)			○	
医療施設静態調査 一般診療所票			○	○
医療施設静態調査 歯科診療所票			○	○
医療施設静態調査 病院票				○

■ 患者調査

表 2.2: 患者調査票一覧

	平成 5 年	平成 8 年	平成 11 年
患者調査病院票	○		
患者調査病院（奇数）票		○	○
患者調査病院（偶数）票		○	○
患者調査一般診療所票	○	○	○
患者調査歯科診療所票	○	○	○
患者調査病院退院票	○	○	○
患者調査一般診療所退院票	○	○	○

■ その他のファイル

データセットの構造定義書として下記のファイル（厚生労働省作成）を使用した。ファイルは MS-Excel 形式である。

- 平成 5 年ファイルレイアウトフォーム
- 平成 8 年ファイルレイアウトフォーム
- 平成 11 年ファイルレイアウトフォーム

- 平成 5 年項目及び分類コード
- 平成 8 年項目及び分類コード
- 平成 11 年項目及び分類コード

2.2 データの件数

■ 医療施設静態調査

病院票、一般診療所票、歯科診療所票のデータの件数を表 2.3 にまとめた。

施設数を活動中と非活動中（「休止・1年以上の休診中」）に分けると、表 2.4 が得られる。これは、厚生省大臣官房統計情報部が作成した医療施設静態調査の数値（表 1 施設の種類別にみた施設数）と一致する。医療施設静態調査は活動中の施設についてまとめたものである。

表 2.3: 医療施設静態調査票 データの件数

	平成 5 年	平成 8 年	平成 11 年
病院	9,896	9,539	9,332
一般診療所	85,669	89,407	92,989
歯科診療所	56,654	60,056	63,119
合計	152,219	159,002	165,440

(解析プログラム : ncenter.sas)

表 2.4: 活動中・非活動中別の施設数

	平成 5 年		平成 8 年		平成 11 年	
	活動中	非活動中	活動中	非活動中	活動中	非活動中
病院	9,844 (99.5%)	52 (0.5%)	9,490 (99.5%)	49 (0.5%)	9,286 (99.5%)	46 (0.5%)
一般診療所	84,128 (98.2%)	1,541 (1.8%)	87,909 (98.3%)	1,498 (1.7%)	91,500 (98.4%)	1,489 (1.6%)
歯科診療所	55,906 (98.7%)	748 (1.3%)	59,357 (98.8%)	699 (1.2%)	62,484 (99.0%)	635 (1.0%)
合計	149,878 (98.5%)	2,341 (1.5%)	156,756 (98.6%)	2,246 (1.4%)	163,270 (98.7%)	2,170 (1.3%)

(解析プログラム : ncenter.sas)

■ 患者調査

患者調査病院票，一般診療所票，歯科診療所票のデータの件数を表 2.5 にまとめた。ただし，一般診療所は介輔診療所を含む。

表 2.5: 患者調査票 データの件数

	平成 5 年	平成 8 年	平成 11 年
患者調査病院票	1,206,207	-	-
患者調査病院 (奇数) 票	-	919,977	1,040,879
患者調査病院 (偶数) 票	-	875,251	997,224
患者調査一般診療所票	291,062	291,143	260,107
患者調査歯科診療所票	23,380	21,585	19,421
合計	1,520,649	2,107,956	2,317,631

(解析プログラム : npatients.sas)

患者調査病院退院票，一般診療所退院票のデータの件数を表 2.6 にまとめた。

表 2.6: 患者調査退院票 データの件数

	平成 5 年	平成 8 年	平成 11 年
患者調査病院退院票	584,997	635,484	787,130
患者調査一般診療所退院票	13,588	13,276	12,814
合計	598,585	648,760	799,944

(解析プログラム : npatients.sas)

第3章 解析方針

3.1 現在の標本抽出法

現在の標本調査¹では、全国の医療施設（休業中の医療施設等を除く）を利用するすべての客体を対象とし、層化無作為に抽出した医療施設（病院約7,000施設、一般診療所約6,000施設、歯科診療所約1,000施設）を利用する患者を調査の客体としている²。これを抽出率にすると、病院約70%、一般診療所約7.5%、歯科診療所2%となる。

実際のデータ（医療施設調査と患者調査）から抽出率を計算すると表3.1が得られる。ただし、表3.1の標本施設数は患者調査における施設数で、データには拒否施設数、休診施設数、患者なし施設数が含まれないので、厚生省大臣官房統計情報部が作成した患者調査の「I 調査概要」の施設数と値が一致していない。総施設数は医療施設静態調査における活動中の施設数（表2.4参照）である。一般診療所と歯科診療所では、調査日に患者が存在しないという状況が比較的発生しやすいので、実際の抽出率は、目標抽出率を1.4ポイント～2.2ポイント下回ることがわかる。また僅かではあるが、近年抽出率が低くなる傾向がある。

3.2 現状の問題点と解析方針

表3.1からわかるように、一般診療所の抽出率は病院の抽出率に比べて低い。また、一般診療所では開設診療科目が限定されることが多く、疾病の分布が診療科目によって大きく偏ることが予想される。したがって、本報告書では、層化法の改善効果がより大きいと考えられる一般診療所について検討することにした。また、歯科診療所は抽出率が最も低いと疾病が限定されるので、層化法の改善効果は低いと判断した。

一般診療所に対する層化の方法は年度ごとに変更されており、例えば、平成14年患者調査では、都道府県別に1.内科（内科、呼吸器科、消化器科（胃腸科）、循環器科、精神科、神経科、神経内科、心療内科、アレルギー科、リウマチ科）、2.小児科、3.外科（外科、脳神経外科、呼吸器外科、心臓血管外科、小児外科、こうもん科）、4.整形外科（整形外科、形成外科、美容外科）、5.産婦人科（産婦人科、産科、婦人科）、6.眼科、7.耳鼻いんこう科（耳鼻いんこう科、気管食道科）、8.皮膚科（皮膚科、泌尿器科、性病

¹平成14年患者調査の抽出率及び抽出施設数は、(1)病院（入院）：全国の約7/10（6,500施設）、（外来）：全国の約3.3/10（3,100施設）、(2)一般診療所：全国の約7/100（6,200施設）、(3)歯科診療所：全国の約2/100（1,300施設）である。

²平成5年、8年、11年患者調査実施要領による。

表 3.1: 実際の抽出率

	種別	標本施設数	総施設数	抽出率 (%)
平成 5 年	病院	6,840	9,844	69.5
	一般診療所	5,107	84,128	6.1
	歯科	868	55,906	1.6
平成 8 年	病院	6,593	9,490	69.5
	一般診療所	5,054	87,909	5.7
	歯科	860	59,357	1.4
平成 11 年	病院	6,420	9,286	69.1
	一般診療所	4,891	91,500	5.3
	歯科	833	62,484	1.3

(解析プログラム : sample.sas)

科), 9. その他の診療所 (リハビリテーション科, 放射線科, 麻酔科, 歯科, 矯正歯科, 小児外科, 歯科口腔外科, 不詳) で層化して抽出している. 類似性のある疾病のクラスとこの層化が対応していれば, 特定疾病の集積する医療施設が抽出率の小ささのゆえに漏れるということが起こりにくくなる. そうでない, その疾病の患者数の推定精度が落ちることになる.

層化においては, 患者の主疾病とその患者が通院・入院した主たる診療科の関係が重要である. 平成 5 年患者調査では診療科目数は 32 (不詳をいれると 33), 平成 8 年及び平成 11 年患者調査では診療科目数は 36 (不詳をいれると 37) である. 一方, 疾病分類は, 大分類, 中分類等いろいろであるが, 本研究では, 厚生労働省が採用した 106 分類 (平成 5 年) または 124 分類 (平成 8 年, 11 年) を考える (付録). これらの分類ごとに層化を行うのは無理なので, カテゴリーの併合による層の数のある程度の集約が必要である. そこで患者の通院・入院行動の類似性に着目した併合をコレスポネンス解析で行うことにした.

なお, 診療科目と疾病分類の詳細は付録の表 A.1~表 A.3 に示してある.

第4章 コレスポネンス分析の理論

4.1 解析における判定・評価

日本で林の数量化第3類と呼ばれているものと本質的に同じ手法が、欧米ではコレスポネンス解析と呼ばれている。統計ソフトウェアでは、SPSSが前者、SASが後者の用語を使っている。本研究では解析にSASを使う関係で、コレスポネンス解析という用語を使った。SASは、世界的に見て、規模が最も大きい統計ソフトウェアであることが知られている。

コレスポネンス解析は、2つの分類因子において、それぞれの因子の各カテゴリーに数値を付与し、その数値の相関係数がどの程度大きくできるかによって、2因子の関連性を評価する手法である。一方の因子のあるカテゴリーが他方の因子のあるカテゴリーと完全に対応していれば、適当に数値を付与することで、相関係数を1にできるし、そのような対応的関連が全くなければ、どのように数量化しても、相関係数を大きくすることができない。それを統計学的に調べる技法である。具体的には、統計データを表4.1のような分割表に整理する。ただし、 f_{ij} は診療科目*i*に来院した疾病*j*の患者数を表す。

表 4.1: 診療科目と疾病の分割表

	疾病 1	...	疾病 <i>j</i>	...	疾病 <i>b</i>	計
診療科目 1	f_{11}	...	f_{1j}	...	f_{1b}	$f_{1\cdot}$
...
診療科目 <i>i</i>	f_{i1}	...	f_{ij}	...	f_{ib}	$f_{i\cdot}$
...
診療科目 <i>a</i>	f_{a1}	...	f_{aj}	...	f_{ab}	$f_{a\cdot}$
計	$f_{\cdot 1}$...	$f_{\cdot j}$...	$f_{\cdot b}$	$f_{\cdot \cdot}$

診療科目 *i* に x_i 、疾病 *j* に y_j という値を付与し、頻度 f_{ij} を重みとして付与数値 (x, y) の相関係数を計算し、それが最大になるように $\{(x_i, y_j); i = 1, 2, \dots, a, j = 1, 2, \dots, b\}$ を定める。得られた値が近いものが類似したカテゴリーであり、両因子で対応する値のカテゴリーが関連の高いものとなる。

実際には、このような1次元的な数値化では十分な精度の数値化が無理なので、コレスポネンス解析では、多次元 (m 次元) の数量化を試みる。そのときの m をどう決めるか、その数量化に基づいてどの

ようにカテゴリーを併合するか、が解析における各種の判定・評価である。

4.2 解析の算法

具体的には、まず $e_{ij} = (f_{ij} - f_i \cdot f_j) / (f_i \cdot f_j)^{1/2}$ を要素とする行列 \mathbf{E} を特異値分解する。すなわち、 $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}'$ 。ただし、 \mathbf{D} は対角行列、 \mathbf{U} は $\mathbf{E}\mathbf{E}'$ の固有ベクトル、 \mathbf{V} は $\mathbf{E}'\mathbf{E}$ の固有ベクトル、「 $'$ 」は転置、という分解を行う。 \mathbf{D} の対角線要素（特異値）を大きさの順に d_1, d_2, \dots とすると、 $d_1^2, d_2^2, \dots, d_a^2$ が $\mathbf{E}\mathbf{E}'$ の固有値になる関係で、

$$\text{tr}(\mathbf{E}'\mathbf{E}) = d_1^2 + d_2^2 + \dots + d_m^2 + \dots + d_a^2 = X^2$$

という関係が成り立つ。

X^2 は2つの分類因子の独立性のカイ二乗統計量なので、これはカイ二乗統計量の加法的分解という意味を持つ。 $(d_1^2 + d_2^2 + \dots + d_m^2) / X^2$ が1に近ければ、 m 次元の数量化で2つの分類因子の対応関係が把握できることを意味する。

その上で、各固有値に対応する固有ベクトル $\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_m$ （あるいは $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_m$ ）に基づいて、カテゴリーの併合を行えばよい。併合においては、医学的意味も資料として利用し、試行錯誤的な層化を行うのが普通である。

第5章 コレスポネンス解析の結果と考察

5.1 特異値，カイ二乗値，寄与率，累積寄与率の結果

疾病と診療科目の分割表についてコレスポネンス解析を行った結果を表5.2～表5.2に示す。各年とも3次元までの累積寄与率は40%程度で，少数の次元で分割表の情報を要約することは困難であることがわかる。10次元までの累積寄与率は約94%で，11次元以降の寄与率は2%以下と無視できる大きさである。これより，疾病構造の複雑性を考慮すると，診療科目の特徴を表すには10次元程度必要といえる。コレスポネンス解析の結果は各年とも同様の傾向である。

表 5.1: 平成5年 特異値，カイ二乗値，寄与率，累積寄与率

次元	特異値	カイ二乗値	寄与率 (%)	累積寄与率 (%)
1	0.94	259908	16.3	16.3
2	0.84	205168	12.9	29.2
3	0.80	188321	11.8	41.0
4	0.78	178910	11.2	52.3
5	0.75	163529	10.3	62.5
6	0.71	146188	9.2	71.7
7	0.71	144705	9.1	80.8
8	0.60	106407	6.7	87.5
9	0.47	63098	4.0	91.4
10	0.29	25147	1.6	93.0

(解析プログラム: `corresp.sas`)

表 5.2: 平成 8 年 特異値, カイ二乗値, 寄与率, 累積寄与率

次元	特異値	カイ二乗値	寄与率 (%)	累積寄与率 (%)
1	0.94	258286	15.0	15.0
2	0.86	212977	12.4	27.4
3	0.85	210204	12.2	39.6
4	0.81	189974	11.0	50.6
5	0.79	179987	10.5	61.0
6	0.76	167953	9.8	70.8
7	0.74	159488	9.3	80.0
8	0.60	103816	6.0	86.1
9	0.51	75808	4.4	90.5
10	0.46	60774	3.5	94.0

(解析プログラム : corresp.sas)

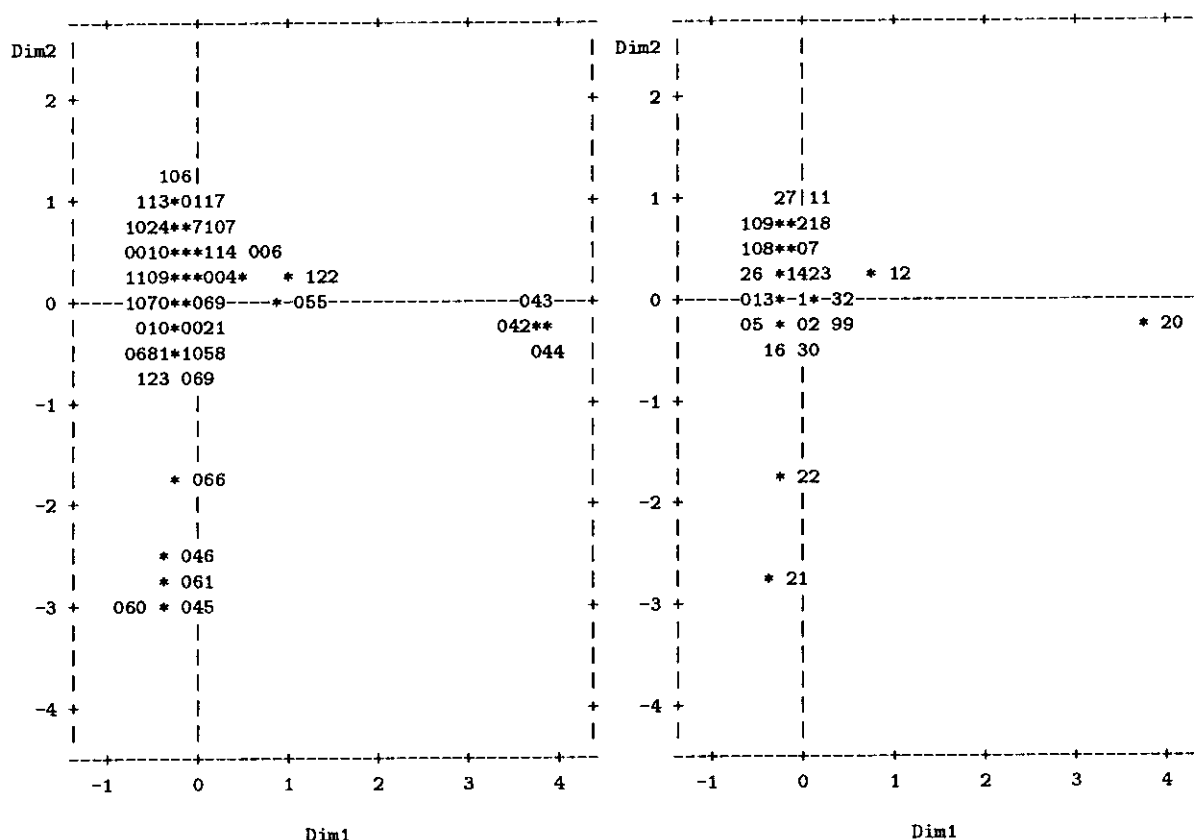
表 5.3: 平成 11 年 特異値, カイ二乗値, 寄与率, 累積寄与率

次元	特異値	カイ二乗値	寄与率 (%)	累積寄与率 (%)
1	0.94	231346	14.2	14.2
2	0.90	210492	12.9	27.2
3	0.85	188775	11.6	38.8
4	0.81	172351	10.6	49.3
5	0.80	164857	10.1	59.5
6	0.77	152497	9.4	68.8
7	0.75	146837	9.0	77.9
8	0.62	98614	6.1	83.9
9	0.59	91953	5.7	89.6
10	0.50	66210	4.1	93.6

(解析プログラム : corresp.sas)

5.2 座標値の散布図

10次元までの行、列カテゴリーの座標値に関する、次元1と次元*i* ($i = 1, 2, \dots, 10$)の散布図を図5.1～図5.27に示す。各図とも左側が疾病の座標値の散布図、右側が診療科目の座標値の散布図である。図中の番号はそれぞれ疾病（左図）または診療科目（右図）を表す。付録の表A.1～表A.3に示すように、疾病の番号と診療科目の番号はそれぞれ平成5年と平成8年および11年でが異なっているので解釈には注意が必要である。



NOTE: 211 OBS がありません。

NOTE: 20 OBS を表示していません。 11 chars hidden.

図 5.1: 平成5年 軸2と軸1の散布図 (左: 疾病, 右: 診療科目)

■ 軸1に特徴がみられる疾病

042: 白内障, 044: その他の眼の疾患, 043: 結膜炎

■ 軸1に特徴がみられる診療科目

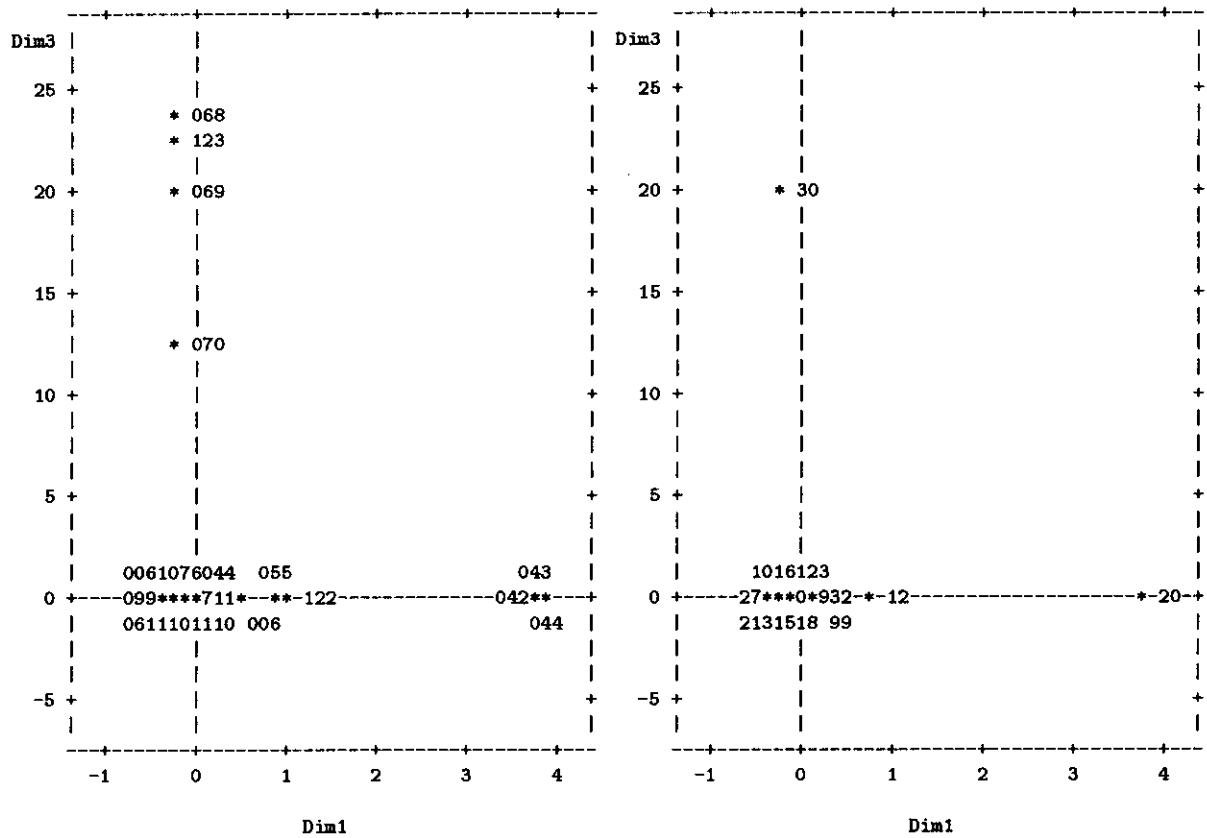
20: 眼科

■ 軸2に特徴がみられる疾病

060：慢性副鼻腔炎，045：中耳炎，061：アレルギー性鼻炎，046：その他の耳の疾患，066：その他の呼吸系の疾患

■ 軸2に特徴がみられる診療科目

21：耳鼻いんこう科，22：気管食道科



NOTE: 251 OBS がありません。

NOTE: 25 OBS を表示していません。 18 chars hidden.

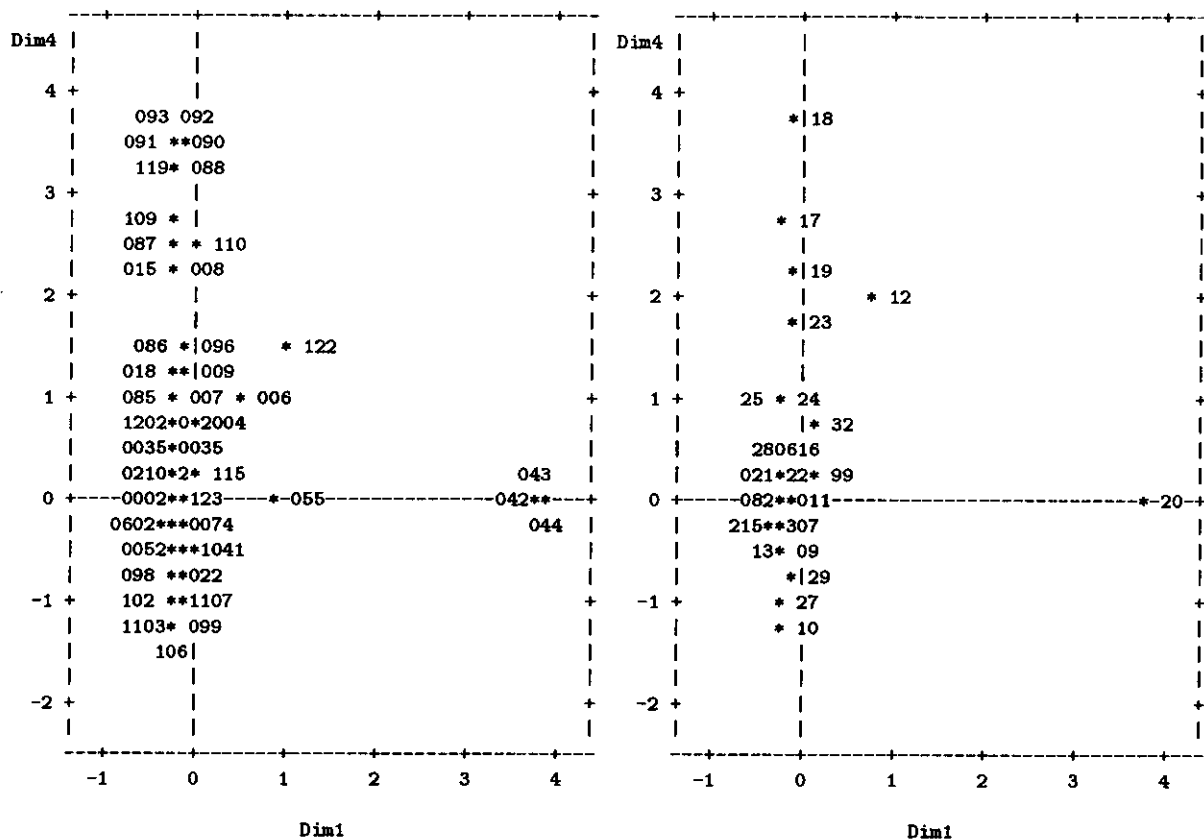
図 5.2: 平成5年 軸3と軸1の散布図 (左: 疾病, 右: 診療科目)

■ 軸3に特徴がみられる疾病

068: う蝕症, 123: 歯の補綴, 069: 歯肉及び歯周疾患, 070: その他の歯の疾患

■ 軸3に特徴がみられる診療科目

30: 歯科



NOTE: 175 OBS がありません。

NOTE: 14 OBS を表示していません。 5 chars hidden.

図 5.3: 平成5年 軸4 と軸1 の散布図 (左: 疾病, 右: 診療科目)

■ 軸4 に特徴がみられる疾病

093: その他の妊娠, 分娩及び産じょくの合併症, 092: 正常分娩, 091: 妊娠中毒症, 119: 正常妊娠・産じょくの管理, 090: 流産, 088: 乳房及びその他の女性生殖器の疾患, 109: 胎児発育遅延, 胎児栄養失調症及び未熟児, 110: その他の周産期に発生した主要病態, 087: 月経障害及び閉経期の障害, 015: 子宮の悪性新生物, 008: 真菌症

■ 軸4 に特徴がみられる診療科目

18: 産科, 17: 産婦人科, 19: 婦人科, 12: 美容外科, 23: 皮膚科