ARTICLES

歯科疫学統計
－第10報　自由記載質問票の質的分析結果の量的評価の試み－
－KH Coderと統計Rの組合せによるシナジー分析－

瀧　口　　徹

A review of oral epidemiological statistics
－ Part X：A trial to adopt quantitative statistical evaluations for results of qualitative analyses using free description type questionnaires －
－ Synergy analyses using KH Coder and statistics R －

Toru Takiguchi

キーワード：質的研究、量的研究、シナジー分析、KH Coder、統計R

はじめに

本誌の「歯科疫学統計」シリーズにおいて9回（GD01：google driveダウンロードサイト）に渡って疫学統計の様々な概念と手法を解説してきた。しかし、質的研究（分析）については名義尺度、順序尺度という質的尺度の統計上の取扱い以外は扱ってこなかった。質的研究は量的研究の対極にあり、その重要性は論を待たないが、その定義と手法の妥当性と再現性および母集団との関連、更に量的分析との関連等の一般化が大変難しい。この難しさがこれまで解説の対象としてこなかった理由である。しかし、実はわが国では質的研究に関して先駆的な理論と手法が複数開発されてきている。これらのうち、本邦で開発された複数のソフトウェアを組み合わせて非常に使いやすいフリーソフトにまとめられたKH Coder（解説2）－（4）参照）を中心に、その使用による質的研究と量的研究のシナジーの試みを紹介したい。質的研究の対象となるデータの種類は多々あるが、調査質問紙法（以下、アンケート調査）で特記事項として設けてきた自由記載部分については体系的な分析を行ってこなかった反省から今回の分析の対象データを自由記載質問回答とした。

今回、この分野の解析に大変使い勝手のいいテキスト解析フリーソフト：KH Coderの利用と量的解析の併用を試みた。この試みは部分的であるが、いわば質的解析と量的解析のシナジー（synergy）[Web01]であり、それぞれの持っている新規性、確実性が相乗的に働き、セレンディピティ（Serendipity）[Web01]と称される思わぬ発見の可能性を秘めていると考えられる。そこで、本稿ではKH Coderを利用した直近の2つの質的調査研究（事例1、事例2）を例にとって質的分析をする際の基本的な要点と質的分析と量的分析のシナジーの試みを紹介したい。

なお、参考文献表記に類似したGD01～03表記は、今回からの試みとしてGoogle driveから文献をダウンロードできるフォルダー、もしくはファイルを示しているので利用願いたい。

【著者連絡先】
〒341-0003　埼玉県三郷市彦成3-86
深井保健科学研究所
主席研究員　瀧口　徹
TEL&FAX：048-957-3315
E-mail：taki8020@mth.biglobe.ne.jp

1．事例のプロフィール

社会福祉系大学医療情報管理学科の卒業研究として著者の指導のもと行われた大学学生を対象にしたアンケート調査（事例1：GD02）[1] と大学院博士後期課程論文となった東日本大震災時の福島県からの県外避難者のアンケート調査（事例2：GD03）[2] の2つを対象とした。

2．解　説

下記の手順で行う。

＜解説＞
1）質的分析の定義と量的研究との関係
2）質的分析の代表的手法
3）質問紙（アンケート）の取り方
4）「茶筌」による品詞の整理の注意点
①複合語と強制抽出語、非抽出語の扱い
②同義語、類似語、近似語、関連語の扱い
5）文（センテンス）と段落（パラグラフ）の扱い
6）よく使われる分析法
(1) 対応分析
(2) 階層クラスター分析
(3) 共起ネットワーク分析
7）その他の質的解析法の紹介
(1) 多次元尺度構成法
(2) 自己組織化マップ
8）量的分析と質的分析のシナジー
9）質的分析における注意と課題
①質的分析の目的の明確化
②妥当性と再現性との関係

1）質的研究の定義と量的研究との関係

本邦の研究者4人の質的研究の定義が志村健一[Web02] の総説において次のように紹介されている。それによると、①木下康仁は「質的研究とは質的データを用いた研究とし、質的データとは当該の研究テーマに関してディテールの豊富なデータであることと規定」、②西條剛央は「質的研究とは、現象をうまく言い当てる（構造化する）言語ゲームの一種」であり、「質的研究法とは、研究者の関心に応じて現場に入ったり、 観察したり、インタビューしたり、分析したり、解釈したりするために体系化されたツー ル」、③能智正博は、「質的研究とは、データの収集、分析、結果の提示のために、数値的表現ではなく、言語的表現を使用する研究方法の総称」、そして本人④志村健一は、研究方法を質的研究、量的研究と二分して論ずることに違和感を覚える立場から、質的研究を暫定的に、「質的分析とは、対象となった文脈（Context）からコーディング（Coding）によってコンセプト（Concept）を創発し、それらのコンセプトを構成（Construct）する一連の過程である」と定義している。

一方、田中博晃[Web03] は質的研究を評価する研究者には3つの立場があるとしている。それによると、『(1) 量的研究と同じ基準と用語で質的研究を評価する、(2) 質的研究独自の認識論に基づいて質的研究を評価する、(3) 質的研究の評価は不可能、の3つの立場がある。しかし、これら3つのいずれの立場でも、研究を遂行上深刻な問題を惹起するとしている。例えば (1) の立場は質的研究が主観的で客観性の無い劣った研究になる恐れがある。また (2) の立場は、質的研究と量的研究は共通の物指しで測れない、いわゆる共約不可能性（incommensurability）という状態に陥る。一方、(3) の立場は「評価の放棄」という最悪の状態になってしまう。』と指摘している。かように質的研究と量的研究の関係は曖昧で相性（あいしょう）は極めて悪い。そこで田中は第4の立場[Web03] として『量的研究と質的研究の両方に共通した科学論に基づき、量的研究と質的研究で共通の評価を行う』というアプローチを推奨している。ここでは両者に共通の科学論を用いることで科学論の相違による不公平さが回避可能になる。また (2) とは異なり共通の科学論を土台にするため、量的研究と質的研究の統合も可能になる。そして (3) のような、なんでもアリの相対主義も回避できるとしている。このような両者に共通の科学論として、構造構成主義[Web02] が挙げられている。構造構成

主義に則(のっと)った評価を行えば、量的研究でも質的研究でも公平に評価可能であるとされる。

このように質的研究（分析）および量的研究（分析）との関係は研究者によって様々な定義と手法があり両者の関係の統一的な解釈と説明が難しく、いずれの断定的な定義も問題が生じるので本稿では単純化して下記の英辞典の定義を暫定的に用いることとした。以下はその和訳である。

> 科学英辞典Reference*[Web04]によれば「質的研究（分析）は観察されることができるが、測ることができないデータの科学的研究」と定義している。また質的分析と量的分析の違いは
>
> 質的分析は、研究されることの性質のカタログ（目録）を作ることを主眼とし、一方、定量分析は、計測可能なデータのみを用いた研究である。質的、量的分析はそれぞれ物事や事象の本質を知るための重要な洞察に関連しており、いずれか一方だけで完成出来ない科学的研究に貢献している。従って、研究者が単独で余りに偏って量的データに頼ることになれば、逃してしまう多くの情報があると思われる。

2）質的分析の代表的手法

代表的な手法は４種類あり順次概説する。

(1) KJ法

その一つは文化人類学者、川喜田二郎氏が1960年代に開発し1990年代まで逐次バージョンアップされてきたKJ法[Web05, 3]である。企業研修や学校教育、各種のワークショップ等において与えられたテーマ（例えば、「医療過誤をなくすにはどうしたらよいか」）に関して付箋（ポストイット）に自由に原因や対策、問題点等を必要なだけ記載させ、集められた付箋を参加者の主観と直感を重視して類型化していく手法である。本法は実践的であり、卓上で行う質的分析法の定番として国内に広く普及している。KJ法の再現性（別のグループによる同一テーマに関する試行結果の同一性）について議論がある。参加者の直感的な主観を重視する本法が異なった時点で、あるいは異なった参加者で同様の結論が出るとしたら参加者がランダムに選ばれるとかの統計的な前提が必要と考えられる。しかしながら、実際の異なった参加者群による再現試験[Web06]で高い再現性が確認されている。そもそも参加者全体のその時その場面での各自の主観を体系的に整理し、参加者の多くが納得する少数の共通概念に統合するということを主目的にするのであれば再現性の議論は重要でなくなるであろう。

(2) 数量化理論

次は、文部科学省統計数理研究所所長であった林知己夫氏が1940年代から50年代に掛けて開発した数量化理論Ⅰ類～Ⅳ類である。Ⅰ類からⅣ類[4, 5, Web07, Web08]がそれぞれ量的分析における重回帰分析、判別分析、対応分析（コレスポンデンス分析）[4]および多次元尺度構成法（MDS）[4, 5]、に対応しており、様々な分野で用いられてきている。本法の特徴は重回帰分析に相当するⅠ類、判別分析に相当するⅡ類において説明変数に順序尺度あるいは名義尺度を用いることが出来ることである。この場合、一般の重回帰分析、判別分析においては順序尺度、名義尺度を0、1パターンなどの2値のダミー変数とすれば説明変数に用いることができるが、Ⅰ類、Ⅱ類では3値以上のカテゴリーでも説明変数として用いることができる。しかしながら現在判別分析の一種であるロジスティックス回帰分析ではSPSS等のソフトは順序尺度、名義尺度を3値以上のカテゴリーでも用いることが出来る利便性が付与されている。また外的基準（目的変数）のある数量化理論Ⅰ類、Ⅱ類では係数の有意差が算出されないので相関比[Web08]を用いて相関比が0.25以上の基準で変数選択を行うのが一般的とされる。

(3) 自然言語処理ソフト「茶筌」:

三番目は自然言語処理フリーソフト「茶筌(ちゃせん)」[5]である。形態素解析器と言われ奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座（松本裕治研究室）が開発した入力文を単語単位に分割し品詞を付与するツールである。

(4) KH Coder

最後に四番目として、第10報の主要キーワードであるKH Coder（ケイエイチ　コーダ）[6] が登場する。KH Coderは立命館大学の樋口耕一氏らが開発した質的分析フリーソフトである。構成は①プログラミング言語Perl（パール）、②関係データベース管理システム（RDBMS）であるMySQL（マイ・エスキューエル）、③統計Rの数量化理論等の多様な統計プログラムおよび④自然言語処理ソフト「茶筌」、を統合した質的分析用多機能フリーソフトである。その機能は出現語の分類機能としては、品詞別（名詞、動詞、形容詞等）頻度分布、同義語、類似語、関連語の一括化（コーディング）機能、階層的階層クラスター分析、多次元尺度構成法（MDS）、対応分析（コレスポンデンス分析）、共起ネットワーク、自己組織マップ等と多岐に渡っている。

なお、多言語対応有料ソフトウェアとしてはmaxqda [Web09]、nvivo [Web10]、atlasti [Web11] 等がある。

3）質問紙（アンケート）の取り方

表１に質的解析のための自記式自由記載用のフォーマット例（事例１：GD02）を示す。事例1の対象は医療情報管理学科学生１～4年生である。質問は「あなたは25年後の自身の就業についてどのように予測、あるいは希望していますか？特に重要と思う事項を３項目（各5行以内）以内で自由記載してください。」と尋ね、①冗長な記載を抑制し、②重要性の高い項目を意識して記述することを誘導している手法である。しかし、いつでもこの形式をとる必要があるということではない。また、「調査表は無記名（但し、性別、学年および取得資格分類は記載）で行い、回答用紙は無記名封筒に入れて封をして提出してください。」として匿名性を保証する手立てが徹底されなければならない。

4）「茶筌」による品詞の整理の注意点

①複合語と強制抽出語、非抽出語の扱い

文法的に単語をその特性から類型化すると名詞、サ変名詞、動詞、形容詞、形容動詞等になるがこれらの総称（表2）が品詞である。「茶筌」のシステムは表3のように文章中の各単語を自動的に品詞分類し、その頻度を提示する。この際、茶

表1　調査協力していただく学生各位へ

下記項目の該当番号をマルで囲んでください。
入学年:　2010, 11, 12, 13, 14
男女：　1:男　2::女
次に、下記の質問に自由記載で回答してください。なお、コンピュータ入力する関係で判読出来るよう出来る限り丁寧に書いてください。

【質問】
あなたは25年後の自身の就業についてどのように予測、あるいは希望していますか？特に重要と思う事項を３項目（各５行以内）以内で自由記載してください。

調査表は無記名（但し、性別、学年および取得資格分類は記載）で行い、回答用紙は無記名封筒に入れて封をして提出してください。

筌に装備されている辞書に無い新語は未知語として示される。また2語以上の単語の組み合わせからなる複合語についても自動的にリストアップされる。ここは最も慎重かつ丁寧に処理しなければならないステップである。例えば今回の事例1においては段落中に「紙カルテ」、「電子カルテ」がそれぞれ登場するが最初の自動品詞分類では複合語とは認識されず、紙、カルテ、電子、それぞれ名詞に分類されるが「紙カルテ」、「電子カルテ」は認識されない。例えば「電子」は物理学の分野では単独で文章中に頻回に出てくるであろうが、医療情報管理学の分野では単独で出現することはまずない。また「紙カルテ」は医療情報分野では頻回に出現する。そこで「電子カルテ」、「紙カルテ」という複合語を強制抽出する処理が必要で「電子」は非抽出扱いとなる。このための方法が表4に示した強制抽出語と非抽出語の登録システムである。

表2　品詞

名詞:　名詞、名詞B、名詞C、固有名詞、地名

サ変名詞:動詞「する」に接続してサ行変格活用の動詞となりうる名詞

形容詞:形容詞、形容詞B

形容動詞:平安時代に形容詞が不足したとき、形容詞で表現できない意味を持つ名詞を語幹として「なり」または「たり」をつけることによって成立した。(ナリ活用とタリ活用。　形容詞は、言い切りの形(辞書に載っている形)が「い」で終わります。大きい、小さい、長い、短い、青い、賢い などが形容詞。形容動詞は、言い切りの形(辞書に載っている形)が「だ」で終わる。

副詞:副詞、副詞B、副詞可能

未知語

感動詞

否定助動詞　否定を表す助動詞には「ず」「じ」「まじ」がある。「ず」は最も一般的に用いられている否定の助動詞であり、次のような活用形を持つ。基本形, 未然, 連用, 終止, 連体, 已然, 命令.
ず, ＊, ず, ず, ぬ, ね, ＊. ざら, ざり, ＊, ざる, ざれ, ざれ ...

動詞:動詞、動詞B

タグ:強制抽出語

表3　25年後の自身の就業状態の予測に関して使用された語の出現頻度（茶筌）

「する」に接続してサ行変格活用の動詞となる名詞

未知語や連結後を新たに定義した場合、タグに登録される

名詞		サ変名詞		形容動詞		副詞可能		タグ		動詞		動詞B		名詞C		否定助動詞	
病院	33	仕事	42	安定	7	今	11	就職	22	働く	59	する	102	職	12	ない	20
医療	21	関係	12	必要	4	現在	3	パート	14	思う	38	なる	53	人	8	ぬ	6
給料	8	希望	10	重要	3	その後	2	診療情報管理士	13	就く	18	できる	11	口	3	ん	3
自分	8	勤務	10	十分	2	時間	2	正社員	11	考える	12	いる	6	先	3		
事務	6	就業	6	普通	2	たくさん	1	医療事務	9	増える	12	ある	5	率	3		
会社	5	結婚	5	プライベート	1	今後	1	ＩＴ	7	続ける	6	もらう	4	幅	2		
職業	5	出世	4	ホワイト	1	場合	1	一般企業	7	勤める	5	つく	3	力	2		
立場	5	経営	3	可能	1	昔	1	転職	6	携わる	5	やる	3	案	1		
技術	4	貢献	3	健康	1	多く	1	社長	4	辞める	4	しれる	2	位	1		
資格	4	取得	3	新た	1	日々	1	事務長	3	上がる	4	なくなる	2	家	1		
場所	4	受付	3	大事	1	来春	1	病院勤務	3	変わる	4	もらえる	2	系	1		
状態	4	生活	3	馬鹿	1			部長	3	学ぶ	3	よる	2	才	1		
年収	4	卒業	3	無関係	1			民間企業	3	感じる	3	いう	1	事	1		
役職	4	介護	2	様々	1			メディカルクラーク	2	取り組む	3	いく	1	上	1		
公務員	3	給与	2					医師	2	出来る	3	いれる	1	場	1		
情報	3	作業	2					医療機関	2	進む	3	かかる	1	親	1		
職種	3	就学	2					医療分野	2	得る	3	こなす	1	身	1		
職場	3	充実	2					課長	2	扱う	2	すぎる	1	土	1		
人間	3	出産	2					高齢化	2	稼ぐ	2	ためる	1	道	1		

□ 頻度が10件以上を矩形で表示

表4　強制抽出語と非抽出語（茶筌）

<強制抽出語>
ＩＴ、ＩＴ関連、ＩＴ関係、PC、ＣＰ、レセプト、ロボット、ペーパーレス、見読性、電子カルテ、クラウド、データーベース、ビックデータ、自動化、一元化、一元的、遠隔医療、 個人の情報、個人情報、医療情報、電子化、カルテ、紙カルテ、医療機関、医療機器、 診療情報管理士、ドクターズクラーク、メディカルクラーク、医師、看護師、看護婦、医療事務、 医療分野、福祉分野、医療･福祉分野、医療･福祉の分野、病院勤務、県立病院、福祉関係、福祉施設、民間企業、一般企業、社長、部長、事務長、課長、中堅社員、正社員、派遣社員、パート、就職、無職、失業、転職、発展、進歩、進化、高齢化、核家族、キツくなる

<非抽出語（単語として扱わない語）>
電子、ペーパ、福祉、機関、個人、読、ロボ、ベース、自動、一つ、端末、二次、三次、民間、一般、社員、派遣、企業、キツ、ビック、クラーク、メディカル、回答、回答せず

注）医療情報管理を扱う学科学生のコメントでは「電子」を独立して語として出現することは（まず）ないので、 「電子」を非抽出語に扱っている。しかし「電子カルテ」、「電子化」は強制抽出語扱いとしている。また「カルテ」は独立して出現するので非抽出語としては扱わず「強制抽出語」扱いしている。

②同義語、類似語、近似語、関連語の扱い

同義語、類似語および近似語は文学の世界や日常会話において意味やニュアンスに深みと幅をもたらす必須要素である。例えば同義語、類似語は、友人、友、友達、友だち、知友、盟友、朋友、友朋、同朋、竹馬の友、幼馴染（おさなじみ）、親友、フレンド、友垣、仲よし、だち、だち公、馴じみ、などである。近似語としては、同僚、仲間、味方、身方、グループ、などが挙げられる。しかし、質的分析においては同義語、類似語、あるいは近似語と同一に扱う処理を行って分析結果の解釈を単純化する必要がある。表5にコーディング・ルール（コード化）の作成法[6]を例示した。ここでコーディングルールは同義語、類似語および近似語をorの関係（記号：｜）で結び、いずれか１語が出現した場合をプラス、1語も出現しない場合をマイナスと扱うコーディング、語間をandの関係（記号：＆）で結ぶコーディング、notの関係（記号：！）で結ぶコーディング、およびそれらの組み合わせ（&!：｜！）から構成されている。これらのコーディングは後述の対応分析、階層クラスター分析、および共起ネットワーク分析で利用されている。

５）文（センテンス）と段落（パラグラフ）の扱い

表6に文と段落の扱いを示す。文（sentence）は主語＋述語からなる句点（。）までの文書を指し、一方、段落（paragraph）は論じている意味が一括（ひとくく）りになる複数の文を指す。KH Coderの処理では、この文と段落の区別が極めて重要である。アンケート調査の場合、段落＝個人毎の記述全体とすることが一般的である。さもないと回答者が多数の文を作成した場合、あるいは主張を強調するため同じ語を繰り返し使用した場合、結果に反映する度合いが強くなり回答の比率にバイアスがかかってしまう。つまり、１回だけ語「重要」を使用した場合と複数回語「重要」を使用した場合と後者の重みが強くなってしまう。ただし、KJ法のように同一者が内容の異なった多数の文を書き、それを整理する場合は想起する単語、句など文より短い単位に集計するのが望ましい。

６）よく使われる分析法

(1) 対応分析

対応分析は数量化理論Ⅲ類（前項2）－（2））と同一のもので別名コレスポンデンス分析[Web12, Web13]と呼ばれる手法である。対応分析（数量化Ⅲ類）

表5　コーディング・ルールの作り方

#を行頭に付けると説明行となりデータとしては無視される
#　注1コード項目間の　|　は　or の意味　　&は　and の 意味
#　注2コード項目間の　|　は半角で前後に半角スペースが必要
#　★★★　or　★★★の関係
*ITコンピュータ関連orコード
IT | コンピュータ | PC | CP | パソコン | 情報処理
| 電子カルテ | ロボット | 電子化 | クラウド
*常勤役職関連orコード
社長 | 部長 | 課長 | 事務長 | 正社員 | 中核社員 | 公務員
*非常勤職員関連orコード
派遣社員 | パート | アルバイト | 非常勤 | 日雇い
*給料関連orコード
給料 | 年収 | 収入 | 万円
*病院(診療所)関連orコード
病院 | 大学病院 | 県立病院 | 市民病院 | 医療機関 | 診療所
*会社企業関連orコード
会社 | 企業 | 民間企業 | 関連企業
*医療事務関連職種関連orコード
診療情報管理士 | メディカルクラーク | ドクターズクラーク | 医療秘書 | 医療事務
*就職関連orコード
就職 | 就業 | 勤務

★★★　and ★★★の関係
*病院&就職 関連andコード
病院 & 就職
*病院&勤務 関連andコード
病院 & 勤務
*病院&(就職or勤務)関連andコード
病院 & (就職 | 勤務)

★★★　and,or,not ★★★の関係
*会社or企業&就労関連and,orコード
(会社 | 企業 | 民間企業 | 関連企業) & (就職 | 勤務)
*診療情報管理士のみに関心and,notコード
診療情報管理士 &! (メディカルクラーク | ドクターズクラーク | 医療秘書)
*医療秘書のみに関心and,notコード
(メディカルクラーク | ドクターズクラーク | 医療秘書) &! 診療情報管理士
*会社のみに就業関心関連and,notコード
(会社 | 企業 | 民間企業 | 関連企業) &! (病院 | 大学病院 | 県立病院 | 市民病院 | 医療機関 | 診療所)
*病院、診療所のみに就業関心関連and,notコード
(病院 | 大学病院 | 県立病院 | 市民病院 | 医療機関 | 診療所) &! (会社 | 企業 | 民間企業 | 関連企業)

表6　文の構成　文と段落の違い

KH Coderによる分析では対象となる文章の構成単位が重要
①語：word　名詞（サ変名詞）、動詞、形容詞（形容動詞）、副詞、助詞
②複合語（連結語）：connective word　医療福祉大学、電子カルテ、臨床医
③節（句）：　phrase　節：2語以上＋S+V　　句：2語以上
<u>形容詞節</u>（<u>句</u>）：
娘は<u>私が娘に買ってやった</u>　（<u>花の絵柄がついた</u>）着物を着て茶会に行った。

<u>副詞節</u>（句）：
<u>兄が買ってくれた</u>（<u>時間があるときに用の</u>）クラシック音楽CD

④文：　sentence　　主語＋述語からなる句点（。）までの文書
⑥段落：paragraph　　論じている意味がひとくくりになる複数の文
質問紙票調査（アンケート調査）の場合、段落＝個人毎の記述全体とすることが一般的。文の意味から統一個人を複数の段落に分ける場合は者率の計算をする場合の分母に注意。

の基本的考え方[Web13]は、分割表において、行の項目と列の項目の相関が最大になるように、行と列の双方を並び替えることである。表7にその手法を特定の植物の光合成量の光の波長に対する2極性[7]の例を用いて示す。左表は通常の相関係数であるSpearmanの積率相関係数:rであるが、光の波長と光合成量の単相関はr=0.47と比較的低い値となり、光合成量の強い波長特性が見かけ上低くなり実際の波長特性が低く判定(underestimate)されてしまう。一方、光合成量の大小関係に合わせて光の色を並び替えた右表ではr=0.93となり「光合成量は強い波長特性がある」という実態にあった数値を得ることができる。この例のように名義尺度であっても数量化理論を用いることにより適切に数量化されより正確に実態を捉えることができる。

数量化理論Ⅲ類を用いた対応分析の問題解決のアプローチは、計量値を用いた主成分分析、因子分析とほぼ同じである。そこで、対応分析を制約つき主成分分析、正準相関分析、質的因子分析として見なす研究者もいる。事例1の場合、表5のコーディングルールによる14項目の該当者数（段落数）および頻度％を表8に示す。ここで11､12項目は25年後の予測（希望、期待）として自身の就業している業種に関して診療情報管理士と医療秘書（メディカルクラーク、ドクターズクラーク、医療秘書）のどちらを意識しているか、また13、14項目は就職先が会社と病院、診療所のいずれを意識しているかを明らかにするためのコーディングである。

対応分析結果を図1に示す。外部変数である男女、学年（1～4年）の位置が矩形で示されている。この図から25年後の自身の就業している業種に関しての男女差、学年差を解釈すると次のようになる。

男女差：男性は会社就職および常勤役職に関心が強い。一方、女性は病院、診療所勤務および診療情報管理士、非常勤役職に関心が強い。

学年差：1～3年と4年生間には4年生の会社志向が強い特徴がある。1～3年生間では2年生が診療情報管理士、医療秘書関連資格に志向が強い傾向がある。

表7　量的変数の相関と質的変数の相関　－光の波長と光合成、光の色名と光合成との関係を例にとって－

表7-a　光の波長と光合成量の相関係数

計量値の相関

No.	光色名	x 波長nm	y 光合成量mg
1.	赤外線	800	15
2.	赤	700	55
3.	橙	600	35
4.	黄	590	18
5.	黄緑	570	12
6.	緑	530	5
7.	青	440	17
8.	藍	420	30
9.	紫	400	10
10.	紫外線	300	0
		単相関係数	0.4703

※波長nmと植物の光合成量mgの仮想データ:光合成量のピークは700nmと420nmの2峰性。単相関係数は0.4703→光の波長と光合成量はある程度の正の相関関係がある。

表7-b　光の波長と光合成量の相関係数
数量化した光の色のと光合成量の相関係数

名義尺度の数量化による相間

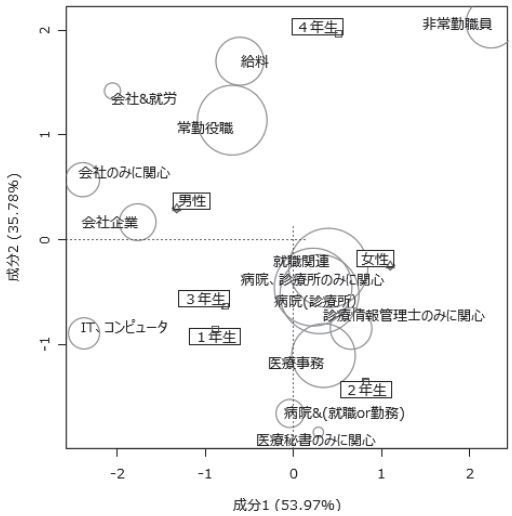
No.	色名	x 波長nm	x 光色名番号	y 光合成量mg
2.	赤	700	10	55
3.	橙	600	9	35
8.	藍	420	8	30
4.	黄	590	7	18
7.	青	440	6	17
1..	赤外線	800	5	15
5.	黄緑	570	4	12
9.	紫	400	3	10
6.	緑	530	2	5
10.	紫外線	300	1	0
			単相関係数	0.9317

※光合成量を降順ソートして色名番号を降順につける。色名番号と光合成量の単相関係数は0.9317→光色名と光合成量間には高い正の相関関係がある。

表8　コーディングルールによる関連コード頻度

使用	コーディング名	頻度	頻度%
○	1.*ITコンピュータ関連orコード	6	3.21%
○	2.*常勤役職関連orコード	26	13.90%
○	3.*非常勤職員関連orコード	14	7.49%
○	4.*給料関連orコード	13	6.95%
○	5.*病院(診療所)関連orコード	33	17.65%
○	6.*会社企業関連orコード	8	4.28%
○	7.*医療事務関連職種関連orコード	22	11.76%
○	8.*就職関連orコード	33	17.65%
×	*病院&就職 関連andコード	2	1.07%
×	*病院&勤務 関連andコード	3	1.60%
○	9.*病院&(就職or勤務)関連andコード	5	2.67%
○	10.*会社or企業&就労関連and,orコード	2	1.07%
○	11.*診療情報管理士のみに関心関連and,noコード	10	5.35%
○	12.*医療秘書のみに関心関連and,notコード	1	0.53%
○	13.*会社のみに就業関心関連and,notコード	7	3.74%
○	14.*病院、診療所のみに就業関心関連and,notコード	32	7.11%
×	#コード無し	74	39.57%
	（文書数）	187	

関連コードの種類:
論理演算子　or　記号　|
意味:段落にいずれかの単語が1つ以上存在する
論理演算子　and　記号　&
意味:段落に全ての単語が存在する
論理演算子　not　記号　!
意味:段落に単語が存在しない
論理演算子　or,not　記号　|!
意味:段落に左の語群が存在、または右の語群が存在しない
論理演算子　and,not　記号　&!
意味:段落に左の語群が存在、かつ右の語群が存在しない



読み取れること
①男女差：男性は会社就職および常勤役職に関心が強い。女性は病院、診療所勤務および診療情報管理士、非常勤役職に関心が強い
②学年差：1~3年と4年生間には4年生の会社志向が強い特徴。1~3年生間では2年生が診療情報管理士、医療秘書関連資格に志向が強い傾向。

図1　関連コードの対応分析（コレスポンデンス分析）

(2) 階層クラスター分析

コーディングルールによる関連コード頻度（表8）に示された14項目を用いて階層クラスター分析[Web12, 6]を行い図2のデンドログラム（樹形図）を得た。医療情報管理学科の学生が25年後の自身の就業を予測するとき4つのクラスターに分類された。すなわち、クラスター1：会社重視型、クラスター2：病院・診療所重視系、クラスター3：診療情報管理士重視系、および病院・診療所での処遇、勤務形態、医療秘書重視系の4つである。6)－(3) でこの結果と男女別の共起ネットワーク分析結果を重ねて考察することとする。

なお、項目数が数十個を超えるような非常に多くのデータを対象とするクラスター分析では、非階層型クラスター分析[Web14]を用いるのが一般的である。

(3) 共起ネットワーク分析

三省堂の大辞林によれば共起（きょうき）（collocation）とは、アメリカの言語学者ハリス（Z. S. Harris 1909～1992）が定義した言語学の概念で、複数の言語

現象が同一の発話・文・文脈などの言語的環境において生起することを指す。例えば、「しとしと」は「雨が降る」とは文・文脈の中で合理的意味を持って使われるが、「雪が降る」とは連動して文、文脈を構成しない。表9に共起性が高い語の組合せと共起性が低い組合せ例を示した。

図3は男女の25年後の自身の就業状況の予想の共通性と違いを示している。ここで検出条件は最小出現語数=2、品詞は新たに定義した名詞であるタグ、固有名詞、地域名を含む名詞のみ、描画数を規定値の60とした。この図の男性、女性区分のような比較のための区分情報を外部変数という。事例１では外部変数は性（2区分）と学年（4区分）の2つである。外部変数を用いることにより対応分析、共起ネットワーク分析等で外部変数のカテゴリー間の違いを比較できるようになる。共起性が強い組合せは太い線で結ばれ、また各語の登場頻度が高い場合は語が入っている円が大きくなっている。図は、男性と女性を示す矩形の間に男女共通語、男性の矩形の外側に放射状に男性

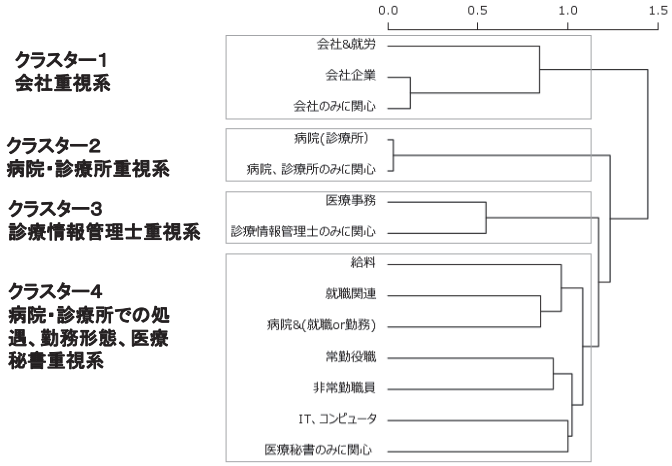


図2　関連コードの階層クラスター分析によるデンドログラム（樹形図）

表9　共起とは？

共起：きょうき（collocation ）（大辞林）
複数の言語現象が同一の発話・文・文脈などの言語的環境において生起すること。「しとしと」は「雨が降る」とは共起するが，「雪が降る」とは共起しないといえる。アメリカの言語学者ハリス（Z. S. Harris 1909～1992）の用語

共起性が高い語の組合せ

①雨　vs　しとしと、ザーザー
②雪　vs　しんしん
③焼芋　vs　アツアツ、ホクホク
④医師　vs　診断、治療、健診、手術、注射
⑤病院　vs　入院、退院、治療、保険証、見舞い
⑥選挙　vs　立候補、投票、当選、落選
⑦星　vs　恒星、惑星、衛星、彗星、星雲、銀河
⑧雲　vs　入道雲、積乱雲、きのこ雲、霧、霞
⑨和食　vs　御飯、味噌汁、漬物、焼魚、箸、醤油
⑩洋食　vs　パン、牛乳、バター、チーズ、ステーキ

共起性が低い語の組合せ

①雨　vs　しんしん
②雪　vs　しとしと、ザーザー
③焼芋　vs　ツルツル、ピカピカ
④医師　vs　占い、介護、調理、駐車
⑤病院　vs　入会、退会、訪問、埋葬
⑥選挙　vs　受験、試験、合格、不合格
⑦星　vs　入道雲、積乱雲、きのこ雲、霧、霞
⑧雲　vs　恒星、惑星、衛星、彗星、星雲、銀河
⑨和食　vs　パン、牛乳、バター、チーズ、ステーキ
⑩洋食　vs　御飯、味噌汁、漬物、焼魚、箸、醤油

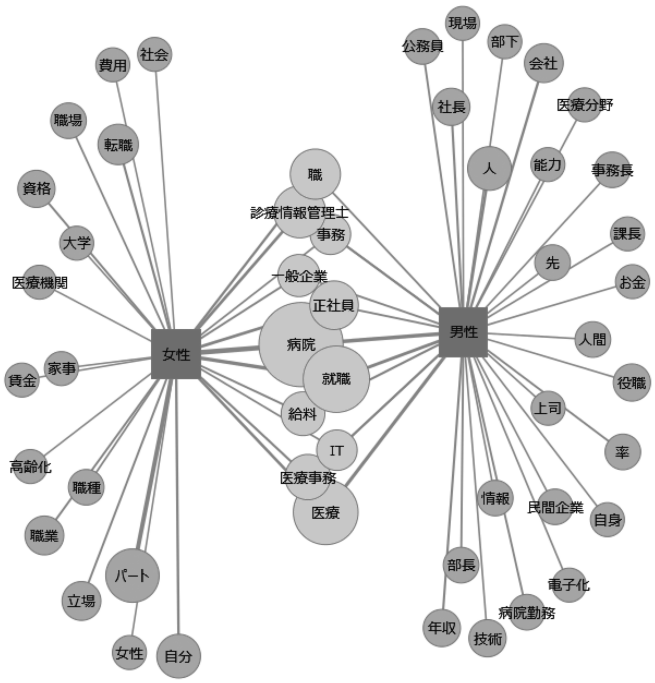


図3　25年後の就業に関する共起ネットワーク分析　－男女比較－
条件：最小出現数=2　描画数=60

固有語、女性の矩形の外側に女性固有語の3群の分布の、バタフライ（蝶）様の構造を示した。「病院」、「就職」、「医療」、「診療情報管理士」、「給料」等11個の語が男性女性の両方と線で結ばれており、男女に共通する語群であった。

一方、男女を示す矩形の外側に並ぶ語は男女間で著明な差が見られた語群である。男性は「社長」や「部長」、「公務員」、「年収」といった語が頻出していた。一方、女性は「パート」や「転職」などの語が特徴的であった。

＜Jaccard係数[Web15)] の活用＞

次に、図4および図5は同じデータで共起性の抽出条件を描画数60からJaccard係数≧0.2、0.3にそれぞれ変更した場合の共起ネットワークを示す。Jaccard係数（表10）は語群からなる複数の集合の類似度を示す代表的な指標で0-1の値で1にちかいほど集合の類似度が高いことを示す。図4および図5は図3と比較して描画数が60から25および10に減少しており、どの条件で共起ネットワークを確定し論文等に掲載するかという問題は悩ましい問題となる。いずれの図も正しいのだが本事例から言えば対応分析（図1）との関連からいうとJaccard係数≧0.2あるいは0.3の条件が言い過ぎや議論のブレがすくないように思われる。

＜KWICコンコーダンス機能の使い方＞

図6において男性固有有の語は「会社」と「人」が残ったが「人」について意味合いを解釈できない状況となる。また女性において残った「パート」の意味することも曖昧である。これらについて、より深い解釈を支援するための手法がKWICコンコーダンス機能6）である。本事例では「人」がどのような文脈に使われ、それは性差や学年差があるかを図6に、同様に「パート」を図7に示した。語「人」を文脈で用いていた8名のうち男性は6名であった。25年後の自身の就業を予測の中で「人」は電子化、ロボット化等で業務が減るという文脈と役に立つ「人」になっていたいという文脈で用いられていた。また語「パート」を用い

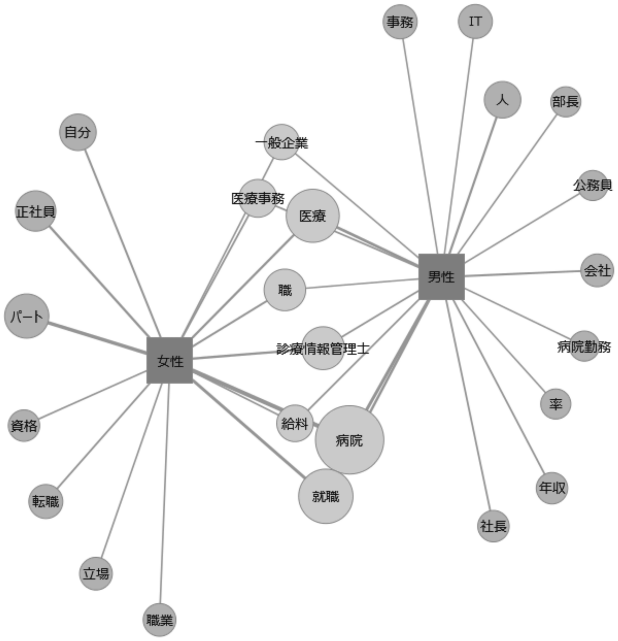


図4　25年後の就業に関する共起ネットワーク分析　－男女比較－
条件：①最小出現数=2　②Jaccard係数≧0.2

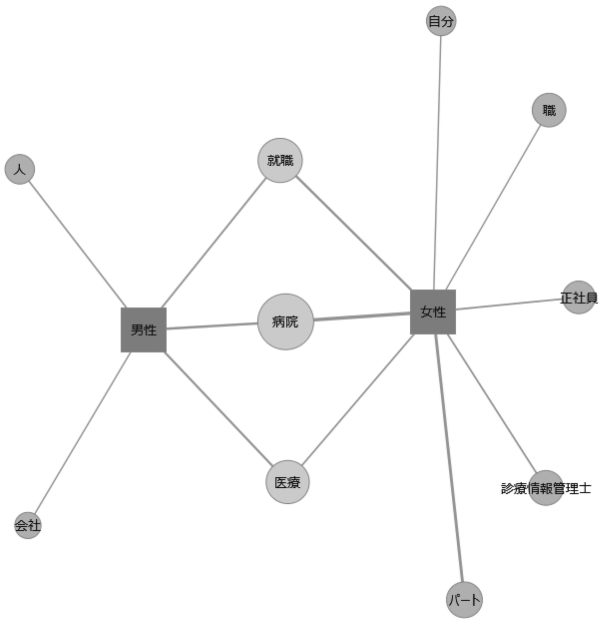


図5　25年後の就業に関する共起ネットワーク分析　－男女比較－
条件：①最小出現数=2　②Jaccard係数≧0.3

ていた14名の全員が女性、うち12名が4年生であった。このことから語「パート」は25年後の自身の就業を予測する言葉のうち男性と女性の違いを最も際立たせる語であり、女性は最終学年において自身の未来の就業を子育て中あるいは子育て後の家事と家計の状況のバランスをとることを想定していると考察された。

このようにKWICコンコーダンス機能を用いると共起ネットワーク等で示された語の文脈上の意味合いを外部変数とともに解釈したり再分類したりすることが格段に容易になる。

また、学年間の違いを表す共起ネットワークを図8に示す。4つの学年表示の内側に全学年共通の使用頻度が高い語が位置していた。その周辺に、

表10　Jaccard係数

4つの語群集合の類似度判定

C_1=パン、牛乳、バター、チーズ、ステーキ、ビール、ワイン、ウイスキー
C_2=パン、牛乳、チーズ、ラーメン、ソース、ワイン
C_3=御飯、味噌汁、漬物、焼魚、醤油、日本酒、ウイスキー
C_4=お茶漬け、味噌汁、漬物、日本酒、ビール

$$Sim(C_i, C_j)=\frac{|C_i \cap C_j|}{|C_i \cup C_j|}$$

2集合間で一致した語数

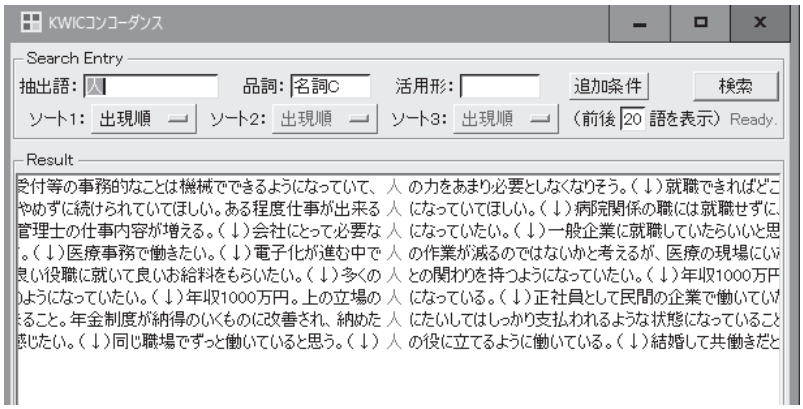
2集合を構成する全語数（二重カウントしない）

$Sim(C_1, C_2)=4/10=\underline{0.40}$

分子：　4種→パン、牛乳、チーズ、ワイン
分母：10種→パン、牛乳、チーズ、ワイン, ステーキ、ビール、バター、ウイスキー、ラーメン、ソース

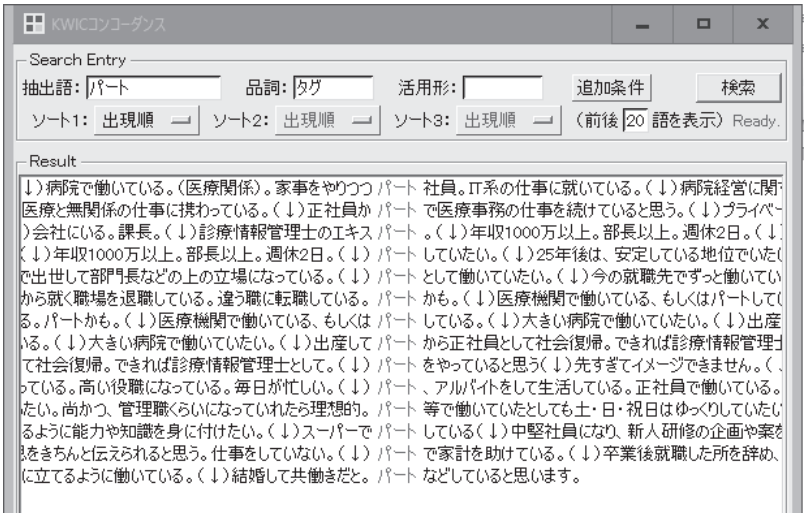
$Sim(C_1, C_3)=1/14=0.07$
$Sim(C_1, C_4)=1/12=0.08$　　$Sim(C_2, C_3)=0/13=0.00$
$Sim(C_2, C_4)=0/11=0.00$　　$Sim(C_3, C_4)=3/9=\underline{0.33}$

注1）Jaccard係数=0~1、KH Coderの初期設定値はJaccard係数>=0.2
注2）その他集合間類似度を表す係数はDice係数、Simpson係数等がある。
http://media.accel-brain.com/jaccard-dice-simpson/



KWICコンコーダンスの解釈：　語「人」を文脈で用いていた8名のうち男性は6名であった。25年後の自身の就業を予測の中で「人」は電子化、ロボット化等で業務が減るという文脈と役に立つ「人」になっていたいという文脈で用いられていた。

図6　25年後の就業に関する共起ネットワーク分析で男性の特性として挙がった語「人」の文脈での使われ方
KWICコンコーダンス機能



KWICコンコーダンスの解釈：14名の全員が女性、うち12名が4年生であることから語「パート」は25年後の自身の就業を予測する言葉のうち男性と女性の違いを最も際立たせる語であり、女性は最終学年において自身の未来の就業を子育て後あるいは子育て中の家事と家計の状況のバランスをとることを想定していると考えられる。

図7　25年後の就業に関する共起ネットワーク分析で女性の特性として挙がった語「パート」の文脈での使われ方　KWICコンコーダンス機能

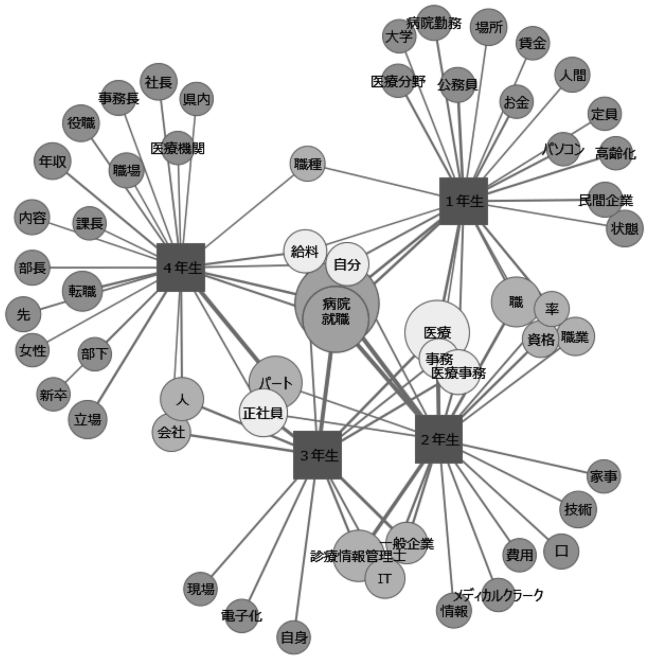


図8　25年後の就業に関する共起ネットワーク分析　－学年比較－

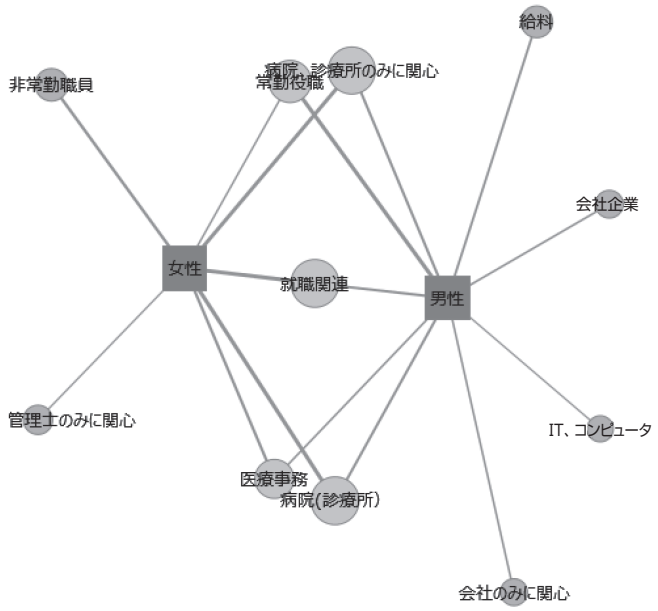


図9　25年後の就業に関する共起ネットワーク分析　－男女比較－
条件：①コーディングルール14項目　②Jaccard係数≧0.2

2または3学年に共通した語が位置する3重の同心円状の構造を示した。時計回りに学年固有の語が配置されており、学年が進むにつれ授業や病院、会社訪問等により関心が移っていくことを示していると考えられる。詳細は原文[1]を参考にされたい。

＜コーディングルールと共起ネットワーク分析＞

表5に示したコーディングルールによる14項目を用いるとこれまでの一つ一つの語を用いた共起ネットワークを同義語、類似語、近似語、および関連語を一括(くく)りにすることにより共起性の検証をより確実にすることができる。図9から男性は「会社のみに関心」、「会社企業」、「IT、コンピュータ」、「給料」、一方、女性は「非常勤職員」、「診療情報管理士のみに関心」が固有の特性であった。コーディングルールを活用することにより男女差がより明確になった。

7）その他の質的解析法の紹介

(1) 多次元尺度構成法：MDS（Multi Dimensional Scaling）

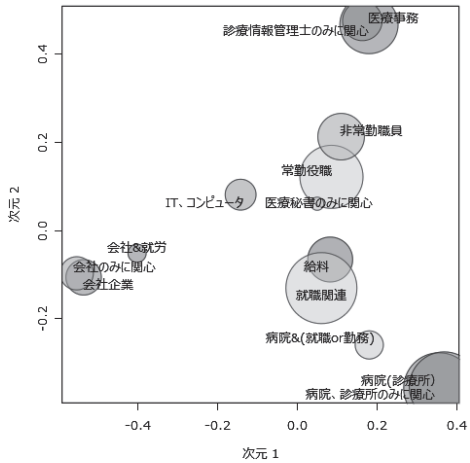
多次元尺度構成法は下記の特徴がある。

①多次元尺度構成法は計量法と非計量法に大別[Web16]されClassical法（計量）、Kruskal法（非計量）、およびSammon法（非線形マップ）の3種[6]がある。

②多次元尺度構成法は因子分析（対応分析）と同じくデータの潜在的構造を空間的（座標上）に表現する[Web17]がその次元は最大3次元である。座標上で視覚的に個体間の親近性データを、2次元あるいは3次元空間に類似したものを近く、そうでないものを遠くに配置する[Web16]。

③因子分析は構成変数の相関関係による相間行列を基本データとするが多次元尺度構成法はユークリッド距離、非ユークリッド距離等の距離行列を基本データとする点が異なる[Web17]。

事例1の14個のコーディングルールを用いた多次元尺度構成法分析結果を図10（二次元）および図11（三次元）に示す。図10は対応分析結果の図1に類似した配置に見えるが図1は語間の相関に基づく分類、図10は語間の距離に基づく分類である。図11では待針(まちばり)が立っているようにみえる

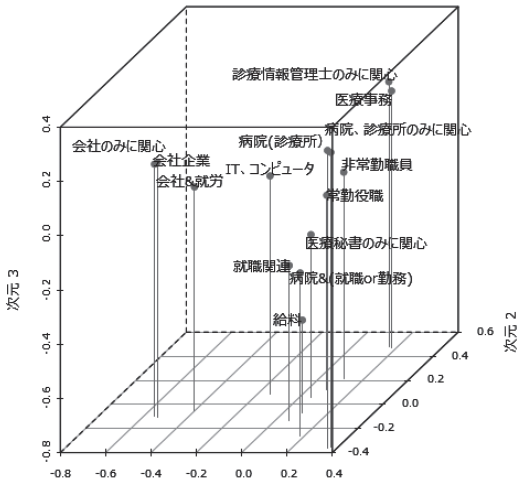


①対応分析（図FFF01）のように男女差、学年差は判定できない

読み取れること

①コード間の意味が近似しているかどうかが二次元座標において視覚的にわかる
②自身の25年後の就業分野は病院、診療所のみに関心、会社のみに関心に2分されている
③自身の25年後の業務は診療情報管理士のみに関心、医療秘書のみに関心に2分されている
④IT、コンピュータは実際には上記②、③のいずれとも必須技能関連が強いが位置の独立性が高く上記②、③に加えて第3の選択軸と考えられる
⑤常勤、非常勤は②~④のように未来の就業の対立軸にはなっていない

図10　25年後の就業に関する多次元尺度構造分析　－2次元－
条件：①コーディングルール14項目　②Classical法　③Jaccard係数≧0.2



二次元図に加えて読み取れること

①コード間の意味が近似しているかどうかが3次元座標において視覚的にわかる
②⑤常勤、非常勤は3次元でみても近接しており未来の就業の対立軸にはなっていない
③IT,コンピュータの位置は第3軸でも会社と病院、診療所、また診療情報管理士と医療秘書のいずれとも等距離的な位置にある
④給料は第3軸においてより独立した要素になっている

図11　25年後の就業に関する多次元尺度構造分析　－3次元－
条件：①コーディングルール14項目　②Classical法　③Jaccard係数≧0.2

14種類のコーディングルールの高さの不揃いが各コードの三次元での遠近関係を示しており、給与に関するコードの独立性が第3軸の特徴である。

(2) 自己組織化マップ

この分類アルゴリズムは、開発したTeuvo Kohone（1982）の名前を取って、「コホネンの自己組織化マップ（Kohonen's self-organizing MAP, SOM）[Web18, Web19]と呼ばれている。自己組織化のアルゴリズムは社会性昆虫といわれる蟻、シロアリ、蜂等の巧みな巣作りを例に説明されることが多い。これらの社会性昆虫の巣作りには設計図やリーダーは存在しない[Web20]。それぞれの個体は認知できる狭い範囲の情報のみに反応し、互いに行動を与え合うことだけで、全体として自律的に秩序だった構造を作り上げる現象が生じる。これを自己組織化あるいは自発的秩序形成という。高等生物の頂点に立つ人では蟻のような自己組織化は行わないが脳内細胞のニュウロンで同様なことが起こる[Web19]とされる。このことにより幼児が母国語を習得するときには、「言葉を学べ」と指示されたり文法等の語学教育を受けなくても、ただ言葉に暴露されるだけで、自然に言語を身につけていく過程が自己組織化だとされる。親や周囲の人の会話を聞くだけで、自発的に単語の意味や使い方を分類していく学習は自己組織化マップと呼ばれ、別名、教師なし学習（a type of unsupervised learning）[Web21]と呼ばれる。この理論は人工知能（AI）の必須条件である自己学習、自発的秩序形成と重なっている。

前置きが長くなったが、この自己組織化機能を用いて14種類のコーディングルールを処理した結果を図12に示す。①対応分析、②階層クラスター分析、③多次元尺度構造分析結果等とこの④自己組織化マップ結果が似通っているが①は前述の対応分析の項で解説した語間の相関係数を基に分析、②、③は語間の距離を基に分析、④は①～③と異なり人工知能のアルゴリズムの一端を用いた類型化ということになる。

図から読み取れること
①コード間の意味が近似しているかどうかが3次元座標において視覚的にわかる
②⑤常勤、非常勤は3次元でみても近接しており未来の就業の対立軸にはなっていない
③IT,コンピュータの位置は第3軸でも会社と病院、診療所、また診療情報管理士と医療秘書のいずれとも等距離的な位置にある
④給料は第3軸においてより独立した要素になっている

図12　25年後の就業に関する自己組織化マップ
条件：①コーディングルール14項目　②クラスター=7

8）量的分析と質的分析のシナジー

事例2（GD02）東日本大震災避難者関連のデータ分析である。2013年の東日本大震災とそれに続く福島第一原発事故により近隣市町村の住民は強制避難勧告または避難勧告を受け、一部は自主避難を余儀なくされた。これら福島県からの避難世帯859からアンケートを回収し主観的なストレス（10段階）の震災直後と3年後の変化と世帯の特性との関連を解析した。ストレスの変化の基礎統計量を表11と図13に示す。3年間のストレスの変化は平均値=-1.68、SD=2.25の凸型の分布を示した。ストレス度が減少していない回答者が34.81％存在していた。

＜共起ネットワーク＞

事例1のコーディング：表5と同様に同義語、近似語、関連語で20項目のコーディングルールを作成し事例1（GD02）と同様なコーディングルールで20コード（C1-C20）に類型化された自由記載関連語群の共起ネットワークを図14に示す。図において中央の4つの矩形がストレス度増減区分を示す。また周辺の4種の円形または多角形の図形は避難民関連語群：C1からC20それぞれの出

表11　震災直後と3年後のストレス度および差の基礎統計量

変　数	ストレス度（10段階）		
	震災直後	震災3年後	差
対象者数	859	859	859
平　均	7.80	6.12	−1.68
不偏分散	4.90	4.72	5.07
標準偏差	2.21	2.17	2.25
最小値	1.00	1.00	−9.00
第1四分位	6.00	5.00	−3.00
第2四分位（中央値）	8.00	6.00	−2.00
第3四分位数	10.00	8.00	0.00
最大値	10.00	10.00	7.00
		平均値	-1.68
		SD	2.25

3年間の差：対応のあるt検定　t=21.86　p=1.50718E-84

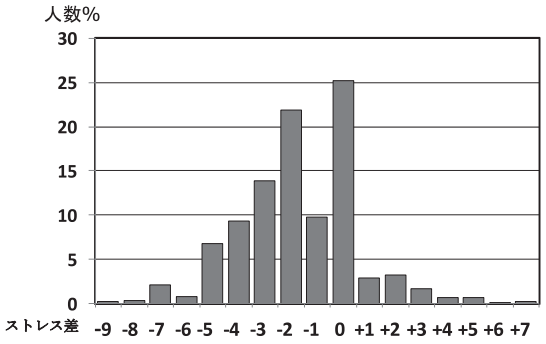


図1　避難者の被災時と３年後のストレス度差分布

ストレス度差

ストレス度差>=0：　34.81%
3年後のストレス度>=被災時ストレス度平均(7.80)　31.55%

図13　避難者の被災時と　3年後のストレス度差分布

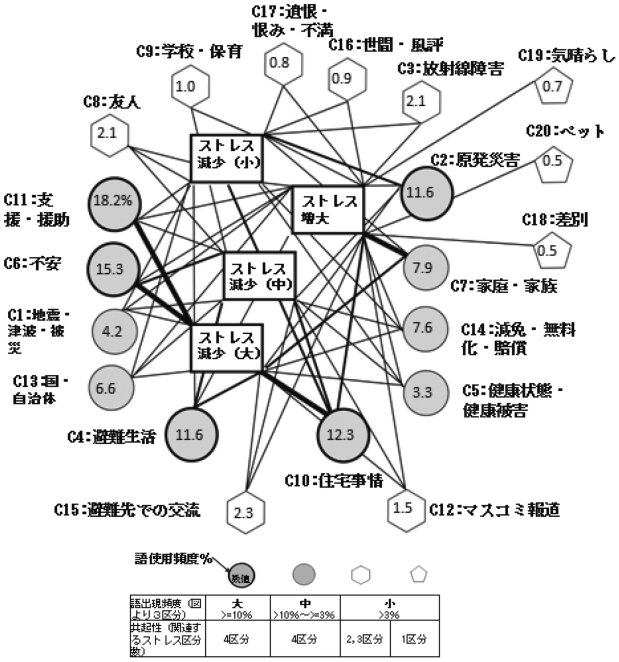


図14　被災直後とストレス度差と自由記載関連語群（同義語、類似語、関連語）との共起ネットワーク

現頻度（図形内の数値は％）を示す。図形と矩形間の線分の太さがストレス度増減区分との共起性を示す。更に４種の図形から出る線分数がストレス度の増減への係わりの多面性（複雑性）を示す。

図14中に太線の円で示す語出現頻度が大（10%以上）のコード（頻度の降順）は、C11：支援・援助、C6：不安、C10：住宅事情、C2：原発災害、およびC4：避難生活の5つであった。この他、語出現頻度が中（3%以上、10%未満）、小の5コードは図2中に細線の円で示した。共起性に関して注目されるのは住宅事情であり、4つのストレス度増減区分のいずれとも共起性が高く、住宅事情がストレス度減少の要因になったり、逆に増大の要因になったりする多面性を示した。また減免・無料化・賠償、国・自治体は4つのストレス度増減区分のいずれとも弱い共起性を示した。現象を単純に解釈すると、住宅事情や原発事故の賠償や保険料の減免、税金の減免等はストレス度増加に繋がったり減少に繋がったり様々に働くということであるが、この背景には深刻かつ複雑な事情があると考察された。

図14において4種類のストレス度と共起性が高い事象の間を太い実線で結んでいる。ここで、ストレス度増大コードはC7：家庭・家族、C10：住宅事情であった。ストレス度減少（大）コードはC11：支援・援助、C6：不安、C10：住宅事情、およびC4：避難生活であった。ストレス度減少（中）コードはC6：不安、C10：住宅事情、C4：:避難生活、であった。ストレス度減少（小）コードはC10：住宅事情、およびC2：原発災害であった。これにストレス増大を加えるとストレス度の減少度合いは4区分になる。

＜トレンド分析＞

トレンド分析の手法はGD01に収納されている歯科疫学統計　第4報、第5報を参照されたい。共起ネットワークで増減傾向に特徴がみられた6項目のトレンド分析結果を表12に示した。まずストレスの増減4パターンの凹凸傾向をFisherの正確確率法で検定した。また右肩上がり（下がり）の傾向をCochran Armitage傾向検定で確認した。

表12　震災直後と3年度のストレス度差と自由記載における高頻度使用語（名詞）との関連

n=859

ストレス度差（4区分） No　関連コード名	減少大 （第1四分位）	減少中 （第2四分位）	減少小 （第3四分位）	増大 （第4四分位）	家族形態6区分の合計 （実数、%）	Fisherの正確確率 （両側検定）	Cochran- Armitage 傾向検定（両側検定）
	288	188	301	82	859世帯数	p	p
2 原発災害	15.63	12.77	7.64	9.76	11.64%	0.0207	0.0043
3 放射線障害	3.82	2.13	1.00	0.00	2.10	0.0666	0.0057
4 避難生活	10.76	19.15	9.97	3.66	11.64	0.0011	0.1166
6 不安	14.24	25.53	11.96	7.32	15.25	0.0001	0.0669
11 支援・援助	17.71	35.11	9.63	12.20	18.16	0.0000	0.0024
14 減免・無料化・賠償	8.68	10.11	4.98	7.32	7.57	0.1452	0.1327

注1）家族形態のうち祖父母+子供は該当無しのため表から除外
注2）表はp>=0.14の語群（コード）を割愛
注3）本表の算出過程：①KH Coderを用いて回答者の自由記載から名詞（結合語を含む）抽出、②抽出した名詞からストレスに関係する可能性の視点で語を選択、 ③選択した名詞から同義語、類似語、関連語を20項目に分類（コーディング）、④回答者の避難直後と3年後の10段階で回答したストレス度区分（四分位区分）ごとに回答者単位の使用頻度（%）を示す。
注4）Fisherの直接確率（6×2表）は統計Rを用いて計算、判定は両側検定

Fisherの正確確率はストレス度増減区分の間で「支援・援助」（p<0.001）、「不安」（p<0.001）、「避難生活」（p<0.01）、および「原発災害」（p<0.05）において頻度%の凸凹が有意であった。一方、Cochran Armitage 傾向検定では「支援・援助」に関する語の使用頻度が高いとストレス度が少ない傾向の高度有意性（p<0.01）が確認された。また「原発災害」および「放射線障害」は共通の傾向が示され、いずれも関連語の使用頻度が高いとストレス度減少率が大きい傾向が高度に有意（p<0.01）であった。このように、質的分析結果のうち量的な確認が必要な項目は量的な統計量を適用すると事象の解釈に質的分析と量的分析とのシナジー効果を発揮出来ると考えられる。なお、これらの結果のより詳細な解釈は原本[2, GD03]を参照されたい。

9）質的分析における注意点と課題

①質的分析の目的の明確化

＜質的分析といえども母集団の想定が必要＞

質的分析でしばしば例数が極端に少ない（例えば、5例以下）報告に出会う。対象が古代人の化石やシーラカンスであれば致し方ない。また特異な希な疾患や人物（天才や極悪犯罪者）の分析は一義的には必ずしも母集団を意識したものではなく個々個別の特性を深く掘り下げることを目的としているため「少数例だからモノが言えない」とはならないであろう。しかし、調査対象の事象が稀有（けう）で無さそうなのに数例しか調査数がない質的分析に出会うと戸惑いを感じるのは私だけであろうか。量的研究の信頼性確保のツールは統計である。統計は悉皆（しっかい）調査で無い以上、標本調査から母集団を確率的に推定する。基本的には質的調査も同じと考える。例えば5例の質的研究の場合に確認した傾向が、6例目、7例目についても、100例目についても普遍的かどうかが重要であろう。その意味から質的分析を何らかの方法で量的にも把握して母集団を推定するというシナジー（相乗効果）が有用と考える。

②妥当性と再現性との関係

＜妥当性＞

アンケートを作成する際に内容の妥当性：validityと再現性：reliabilityが曖昧だと結果の質が確定できず調査の価値が失われる[Web23, Web24, Web25]。妥当性は「調べたいことを調べることができる指標で測っているか」で決まる。そのため、新たに指標を作成すると果たしてその指標が身体状況や事象を正確に代表しているかどうかの保証が必要となる。例えば皮下脂肪の代表指標としてBMI、血糖値を代表する指標ととしてHbA1cはそれぞれ妥当性があるが、それは既に先行研究で確立された指標であるからである。またWHOが開発し

表13　再現性の確認　クローンバックのα（Cronbach’α）

SPSS V21による計算
同一性が高い質問では0.7~0.8が一般的基準
1~4の質問全体のCronbachのα=0.5703であるが
質問2を除くと0.6712に上昇することを示している

Reliability Statistics

Cronbach's Alpha	N of Items
.5703	4

Item-Total Statistics

類似の質問 各2区分	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1.食品放射能汚染への心配	.8952	.8701	.3586	.4939
2.大切なペットの死	1.4703	1.1958	.0671	.6712
3.放射線被爆の心配	1.1700	.7100	.4948	.3638
4.内部被爆防止のための食材選別	1.3120	.7534	.5094	.3583

この矩形の項目をはずすと全体のChronbach’ αは0.5703から0.6712に上がる。
＜用途＞
①**類似の質問の回答の信頼性の確認**→虚偽回答の検出→アンケートの信頼性が疑われる
②**独立性の高い質問かどうかの確認**→ Chronbach’ αが低い方が好ましい

たうつ病の自己診断の20の質問票：SRQ20 [Web22] であればうつ病の指標としての妥当性は保証されているが、睡眠時間と食欲と外出時間を総合指標化してうつ病の指標にしたとすると、予（あらかじ）めうつ病を表すゴールドスタンダードとの関連を証明しておかなければならない。その手法はスクリーニング指標の精度を表す敏感度（感度）：sensitivityと特異度：specificity等が尺度となる。

＜再現性＞

アンケートの質を評価する際のもう一つの要点が再現性：reliability [Web23, Web24, Web25] である。再現性は内容の信頼性であり「同じ人に同じような（条件の）質問をして同じ答えを得ている」かの保証である。もし同じアンケートを２回やって全く同じ回答が出たら信頼性最大になるが非現実的なので近似した設問を入れてその回答の類似度で判定するのが現実的である。この指標がCronbachのα（クローンバックのアルファ）[Web23, Web24, Web25] である。Cronbachのαは0から1までの値をとるが0.7～0.8 [Web23] が信頼度の目安である。しかし、互いに独立性が高い指標、例えば主成分分析の主成分やや因子分析の因子のように互いに無相関な指標を意識して選択した場合はCronbachのαは小さいのが自然（0がベスト）[Web23] である。表13に事例2におけるCronbachのαの実際例を示す。SPSS V21による計算で放射能の影響に関する不安を訪ねた4種類の類似した質問全体のCronbachのα=0.5703であるが質問2を除くと0.6712に上昇し信頼性の目安である0.7の水準をほぼ達成することを示している。このことにより事例2のアンケートに対する回答者の回答が類似の質問に矛盾した答えをしていることは無く再現性（信頼性）が確認された。

＜自由記載データの妥当性と再現性＞

上記の妥当性と再現性の議論は設問が選択肢（カテゴリー）の場合に確認できるが自由記載の場合はいずれも確認できない。類似の設問で2ヶ所以上に自由記載欄を設け、両者の一致度を見るのは現実的でなく回答者の不信を招く恐れがあり好ましくない。十分と言える方法は無いが幾つかの選択からなる設問を合わせて設けてCronbachのαにより確認し、この値が基準を満たしていれば自由記載の信頼性も高いと見なす方法しかないと考えられる。もう一歩踏み込めば選択肢が示す特性と自由記載の各種の質的分析との関係が近似していることが確認できるであろう。ただし、自由記載は選択肢のある設問では捉えきれない特性を、設問では想定していない回答など、いわば生の言葉を拾うためにこそあるので再現性が確認しづらいから自由記載欄を設けないという選択は無い。妥当性、再現性は確認しづらくても対象となる集団の真の声を聴く意味で自由記載欄は重要で

ある。いままでこれといって体系づけて統計処理しなかった自由記載欄をKH Coder等の新兵器で調理し、それに量的分析手法を加味すれば関係者を唸らせる思わぬレシピを発見できるかもしれない。

おわりに

本稿で用いた「質的、量的分析はそれぞれ物事や事象の本質を知るための重要な洞察に関連しており、いずれか一方だけで完成出来ない科学的研究に貢献している。」という両者の違いと関連[Web26)]を調整して相乗効果（シナジー効果）により研究成果の質を向上させる手法はまだ一般的になっておらず開発途上にあると考えられる。今回自由記載アンケート調査にKH Coderによる品詞の頻度分析と共起ネットワーク分析を行い、その一部に定量的解析を試みた。類似の事例に対しての分析に参考になれば幸いである。

なお、第10報から初めて関連論文の一部をGoogle driveからダウンロードできるようにして本総説の利用の便を改善した。

謝　辞

本総説を執筆するにあたり各段階でご指導いただいた歯科疫学研究会の顧問である蓑輪眞澄、高江洲義矩、境脩、伊藤学而、佐々木英忠、平田幸夫、山本龍生の各先生、ならびにKH Coderを用いた統計解析に関して手解きいただいた立命館大学の樋口耕一先生に深謝申し上げます。さらに貴重なご助言をいただいた深井穫博同研究会会長をはじめとする幹事の方々に感謝申し上げます。

【Google driveからのダウンロードサイト】

GD01）総説：シリーズ歯科疫学統計第1報～第9報：
https://drive.google.com/drive/folders/0B3PYdxQv0EWiRll4Y1c5YmNiQ28?usp=sharing

GD02）原著：近未来の医療福祉分野でのIT技術の進展と自身の就業状況に関する大学医療情報管理学科生の意識調査
https://drive.google.com/file/d/0B3PYdxQv0EWienZLVTNMcDZmLTg/view?usp=sharing

GD03）原著：東日本大震災後福島県からの県外避難家庭の被災直後と3年後のストレス度差の要因分析

【参照 Web-site（2016/9/1現在）】

Web01）Whatls.com, definition of synergy:
http://whatis.techtarget.com/definition/synergy

Web02）志村健一：質的研究の動向と課題（日本社会福祉学会）：
http://www.jssw.jp/whatsnew/doc/archive_2012_02.pdf

Web03）田中博晃：質的研究のための評価基準：KJ法を用いた動機づけ研究での例,
http://www.mizumot.com/method/2011-10_Tanaka.pdf

Web04）Reference*：What is the difference between qualitative and quantitative analysis?
https://www.reference.com/web?q=Qualitative+Analysis+Versus+Quantitative+Analysis&qo=contentPageRelatedSearch&o=600606&l=dir&gc=1

Web05）田中博晃：KJ法入門：質的データ分析法としてKJ法を行う前に：
http://www.mizumot.com/method/tanaka.pdf

Web06）Jared M. Spool, The KJ-Technique: A Group Process for Establishing Priorities
https://articles.uie.com/kj_technique/

Web07）林知己夫の数量化理論：
http://www.orsj.or.jp/~wiki/wiki/index.php/数量化法

Web08）多変量解析の手法別解説：
https://istat.co.jp/ta_commentary/method1

Web09）質的分析ソフト：MAXQDA：
http://www.maxqda.com/lang/jp

Web10）質的分析ソフト：NVIVO：
http://www.nvivo.jp/

Web11）質的分析ソフト：ATLASTI：
http://atlasti.com/ja/

Web12）コレスポンデンス分析（対応分析）：
http://www.macromill.com/landing/words/b004.html

Web13）Rと対応分析：
https://www1.doshisha.ac.jp/~mjin/R/26/26.htm

Web14）クラスター分析：
http://www.albert2005.co.jp/technology/mining/cluster2.html

Web15）集合間類似度を表す3指標：Jaccard係数　Dice係数　Simpson係数
http://media.accel-brain.com/jaccard-dice-sirmpson/

Web16）多次元尺度構成法：
https://www1.doshisha.ac.jp/~mjin/R/27/27.html

Web17）多次元尺度構成法（MDS）の紹介：
https://www.jstage.jst.go.jp/article/jje1965/13/4/

13_4_137/_pdf
Web18）自己組織化マップ：
http://gaya.jp/spiking_neuron/som.htm
Web19）KohonAnts: A Self-Organizing Ant Algorithm for Clustering and Pattern Classification：
https://arxiv.org/pdf/0803.2695
Web20）シロアリの多様な巣の構造はいかにして生じるか？：
http://www.kyoto-u.ac.jp/ja/research/research_results/2015/documents/151104_1/01.pdf
Web21）Kohonen's Self-Organizing Map（SOM）：
http://www.willamette.edu/~gorr/classes/cs449/Unsupervised/SOM.html
Web22）User's Guide to The Self Reporting Questionnaire（SRQ20）：
http://webcache.googleusercontent.com/search?q=cache:bBnCu9gs8RIJ:apps.who.int/iris/handle/10665/61113+&cd=2&hl=ja&ct=clnk&gl=ca
Web23）Chronbachのα　～質問票の信頼性の指標～：
http://www.geocities.jp/niigataseisinkango/b02_Cronbach.pdf
Web24）Evaluating the Validity and Reliability of the Diversity Icebreaker Questonaire, Wolfgang K：
http://edoc.hu-berlin.de/master/sydorenko-tetyana-2012-02-20/PDF/sydorenko.pdf
Web25）妥当性と信頼性：
http://www.u-gakugei.ac.jp/~kishilab/validity-reliability.htm
Web26）Reference*：What is the difference between qualitative and quantitative analysis?
https://www.reference.com/web?q=Qualitative+Analysis+Versus+Quantitative+Analysis&qo=contentPageRelatedSearch&o=600606&l=dir&gc=1

文　献

1）東海林優，長崎佳苗，松田久義，ほか：近未来の医療福祉分野でのIT技術の進展と自身の就業状況に関する大学医療情報管理学科生の意識調査，新潟医療福祉学会誌，15（2）：7-19，2016.
2）小磯京子，本間美知子，関　千鶴，ほか：東日本大震災後福島県からの県外避難家庭の被災直後と3年後のストレス度差の要因分析，災害看護学会誌，日本災害看護学会誌，17（3）：14-29，2016.
3）佐藤郁哉：質的データ分析法　原理・方法・実践，新曜社，東京，初版，2008，3-16，53-58頁.
4）金　明哲：Rによるデータサイエンス　データ解析の基礎から最新手法まで，森北出版，東京，初版，2007. 87-106. 127-133頁.
5）野村義明，テキストマイニング，ヘルスサイエンス・ヘルスケア，552003，3（1），42-44.
6）樋口耕一：社会調査のための計量テキスト分析　内容分析の継承と発展を目指して，ナカニシア出版，京都，初版，2014. 1-28，137-190頁.
7）矢吹万寿，高　博：数種の蔬菜の光合成と光質との関係，農業気象，29（1）：17-23，1973.

A review of oral epidemiological statistics
− Part X : A trial to adopt quantitative statistical evaluations for results of qualitative analyses using free description type questionnaires −
− Synergy analyses using KH Coder and statistics R −

Toru Takiguchi
(Fukai Institute of Health Science)

Key Words : qualitative analysis, quantitative analysis, synergy analysis, KH Coder, statistics R

Various concepts and technique of epidemiology statistics have been discussed for nine times in the "A review of oral epidemiological statistics" series of this magazine. However, qualitative analysis had not been dealt other than the statistical handling of the qualitative scale as ordinal scale and nominal scale. There is no room to argue about the importance of qualitative study in the opposite end of quantitative study. However, the definition and confirmation of validity, reliability and population are difficult at all. It is the reason why this theme had not dealt in this series so far. However, in fact, pioneer theories and plural techniques on qualitative study have been developed in Japan. They are KJ method developed by Jiro Kawakita, Quantification Theory I～IV by Chikio Hayashi, text segmentation software packages Chasen(by Yuji Matsumoto and KH Coder by Koichi Higuchi. KH Coder is very useful and 3. easy-to-use free software comprising plural software developed in Japan. Unique trials were designed to evaluate synergistic effects of combination of qualitative analysis and quantitative analysis in this review.

Though there were many kinds of qualitative data, object data of this analysis was determined as freedom mention questionnaires out of regret that especially had not been dealt at all so far. r

Some quantitative studies using KH Coder throwing new light on qualitative analysis, and synergic studies between quantitative studies and qualitative studies using statistics R were performed. This trial is a small first step for synergy of qualitative analysis and quantitative analysis, but it is thought that possibilities of the unexpected discovery called “serendipity” will be increased. Therefore, two case studies dealing qualitative data were introduced to explain qualitative analyses and synergic procedures using KH Coder and statistics R in this review.

Various concepts and technique of epidemiology statistics have been discussed for nine times in the "A review of oral epidemiological statistics" series of this magazine. However, qualitative analysis had not been discussed other than the statistical handling of the qualitative scale as ordinal scale and nominal scale. There is no room to argue about the importance of qualitative study in the opposite end of quantitative study. However, the definition and confirmation of validity, reliability and population are difficult at all. It is the reason why this theme had not been explained in this series so far. However, in fact, pioneer theories and plural techniques on qualitative study have been developed in Japan. They are KJ method developed by Jiro Kawakita, Quantification Theory I～IV by Chikio Hayashi, text segmentation software packages Chasen by Yuji Matsumoto and KH Coder by Koichi Higuchi. KH Coder is very useful and easy-to-use free software

comprising plural software developed in Japan. Unique challenges were designed to evaluate synergistic effects of combination of qualitative analysis and quantitative analysis in this review.

Especially, freedom mention questionnaires had not been discussed at all so far out of many kinds of qualitative data in this statistical series of review. It was selected as object data of this analysis. Two quantitative studies using KH Coder throwing new light on qualitative analysis, and synergic studies with quantitative studies and qualitative studies using statistics R were performed. This challenge is a small first step for synergy with qualitative analysis and quantitative analysis, but it is thought that possibilities of the unexpected discovery called "serendipity" will be increased. Therefore, two case studies were introduced to explain qualitative analyses and synergic procedures using KH Coder and statistics R in this review.

Health Science and Health Care 16（1）：4－28，2016