

### 第3章 薬剤師のための医療統計とデータ解析



#### 1. 臨床研究でのデータ解析とは

##### 1) はじめに

臨床研究における統計解析に関しては入門書から専門的なものまで多くの書籍がある。統計手法の手順を平易に、あるいは専門的に詳しく示したもの、代表的なソフトウェアの使用法とあわせて例示を多用したもの、実際の学術論文事例を紹介し、それを解説することによって統計手法の理解を深めることを目的としたもの、など多岐にわたる。

臨床研究を志そうと勉強を始めたとき、「統計の本を買って読んだけど、いつも最初のほうで行き詰ってしまう」とか、「数式がよくわからないのであきらめた」という悩みを耳にする。統計学なのである程度の数式への理解は必要ではあるが、容易に習得できるものでもないことも事実である。本章では、解析理論が中心ではなくこれから臨床研究を志す人たちが、それぞれの目指す研究の「流れ」の中で統計解析のイメージをつかめることを目標とする。

##### なぜ研究をするのか？

臨床研究の目的は、臨床疑問について実際の臨床データを調査して解決を試みることである。統計的有意差を出すためとか、学会で発表するからといった目的は間違っていないが第一目標ではない。やるべきことにしっかり取り組み、疑問を科学的に解決し、結果を治療に役立て患者に還元すること、が大切だと読者は承知されているはず。普段から業務に関する知識や技能が少しでも上達するように、いろいろな勉強をし、勉強会にも参加し、経験を積んでおられることと思うが、一步進んで自分らしいクリニカルクエスチョンを着想し研究を志すことは決して難しくない。日々勉強し、試行錯誤し、コツコツと臨床研究経験を重ねて自分を成長させる、そう心に決めたら臨床研究も楽しくなるのではないだろうか。「学問に王道なし」「仇花に実は実らず」である。自分の利益や都合ばかりを優先せず、他者のために貢献するという思いさえあれば、誰でも臨床研究に取り組むことができる。

とはいえ、何事でも自分の考えを行動に移す際には十分な計画が大切である。安易な計画はうまくいかないことは筆者も痛いほど経験している。臨床研究も同様である。手順としては、まずは「骨子(シノプシス)」とよばれる研究計画概要を作成し、いろいろな立場の人たちが内容の是非、妥当性について意見交換し内容を充実させる。自分ですべてのことを理解して骨子を磨き上げることも可能かもしれないが、いろいろな立場のひとから意見を聞くことができればなおよい。研究の目的の本質を理解し、前向きに背中を押してくれるひとたちと話したいものだ。

患者さんのデータは大切に扱う

臨床研究は、患者から提供いただく大切なデータを扱い、科学的に適切な方法で解析し、その結果を広く医療にフィードバックすることが目的である。そのためには、データのもつ性質に応じて適切な解析手法を用いることが必須であり、おのずと医療統計、データ解析の正しい方法を理解する努力が必要であろう。統計やデータ解析の分野は、医療に限らず心理学、教育学、天文学、その他さまざまな領域で検討され確立されてきたものである。近年のソフトウェアの発展に伴い、誰でも解析を行う環境になり便利にはなったが、その代償として学問の本質を軽視していないだろうか。「できた気になる」という勘違いをせず、先人の努力と統計学の本質に敬意を表しながら、常に謙虚に向き合いたいものだ。

世の中にはたくさんデータがあふれているが、一人ひとりの患者さんの貢献があってこそ得られるものである。一症例だけの症例報告でも、ビッグデータでも、すべて個人個人の患者さんの「思い」があることは忘れてはならない。医療現場で働く読者には今さら言うまでもないが。

## 2) クリニカルクエスチョンから研究計画へ

### ① 最初に研究タイトルを考えてみる

研究骨子を作成する最初の段階としては、研究のタイトルを文章化してみることをお勧めしたい。いきなりだが研究のタイトルを考えてみる。ぼんやりと頭の中で考えているだけでなく、「こうなってほしい」と思うことを文章に明示することで全体を俯瞰した内容を思い描くことができる。受益者である患者に何を還元できるかを文章で明確にしてみる。例えば「オピオイド使用時の効果と副作用に関する医療者評価と患者評価の乖離に関する研究」とすると、乖離を探索してその結果を治療に生かそうという意図が見えるし、実際には医療者と患者それぞれからの評価の数値やその差異を解析したいんだな、と予想がつくと思う。

### ② 背景および目的の文章化

タイトルができたら「背景(はじめに)」の内容を考えてみる。クリニカルクエスチョンを設定した段階で、ある程度関連する研究について自分たちの過去の経験や他者の研究論文などを参照していることと思う。研究開始時点で自分の知っている背景を整理し、今何がわかっていて、何を明らかにすべきなのかを抽出することで目的が明確になる。具体的には第2章の事例を参考にしてほしい。この段階でいくつかの論文を整理しておけば、将来自分の研究成果を発表する際の引用文献にもなる。すぐに手元に資料がなければ、PubMed、医中誌、などの検索サイトを活用して文献検索を行うことや、学会などの学術サイトを訪ねてみることも大切である。普段の調査を惜しまない地道な努力こそが良い成果につながる。

### ③ 具体的方法を考える手順を知る

方法には具体的に次の項目がある。いわゆる「PICO」とか「PECO」を含む内容である。

1. 対象患者の設定～「肺がんを有する患者」など
2. 選択基準～「肺がんと診断された 20 歳以上の患者」など
3. 除外基準～薬剤の効果を知りたい場合に「放射線治療患者を除外する」など
4. 試験デザイン～群分けの有無、データ収集の方法、盲検化、など
5. データ解析方法～得られるデータの量や性質に合った手法
6. その他～倫理審査対応、「説明と同意」、利益相反 (COI)、倫理教育受講、など

### 3) 方法を知る

ここでは具体的に方法について概説する。具体例がいちばんわかりやすい教材なので、読者自身で臨床論文をひとつ決め、「方法」の項と照らし合わせながら確認してほしい。ただし、必ず患者からのデータを得る手順が記載されている論文を参照してほしい。データベース解析などは「データありき」から研究が始まるのでここでの勉強目的にはそぐわない。

#### ① 対象患者と患者背景、選択・除外基準

臨床試験ではまず「どのような患者」を対象とするか、例えば、胃がんの患者、終末期がん患者、高血圧治療を受けている外来患者、など目的に沿った患者を定義する。次にその患者の「選択基準」「除外基準」を明確にする。成人患者を対象とするのであれば年齢制限が必要となり、腎機能を考慮する場合にはクレアチニンクリアランスといった患者の臨床検査値範囲に制限を設ける必要があるかもしれない。思いつく条件を列挙し、それぞれ基準として設定する必要があるかどうかの検討を行ってみる。

他職種の人たちからの意見を聞くことも大切である。医師や看護師なども詳しい領域があり、併用薬に制限を設けるなら薬剤師が詳しいであろう。臨床検査値や画像診断まで言及するのであれば各検査技師の意見も聞き、これをきっかけにチーム医療がより充実すれば一石二鳥である。

#### ② 試験デザインの概要

試験デザインとは研究を行う上でどのような方法でデータを得るのかという計画の一部で、患者をどのようにグループ分けするか、それぞれのグループでどのような治療を施すのか、どれくらいの期間の情報を得るのか、といった具体的な方法をいう。

次のような要素がある。

### 試験デザインを考える時の具体的項目

患者のグループ(群)分け	比較したい群を分ける、患者背景で分ける(層別する)、など。
各群の条件	治療方法の違い、プラセボ群の有無など。倫理的な面も考慮し実現可能な方法を決める。単群の場合もある。
研究期間 (データを取得期間)	効果が見られるまでの期間、副作用発症までの期間、患者が治癒するまでの時間、など、目的に即してデータ取得期間を考える。
必要症例数	何人の患者からデータを得るのか、統計的根拠で示すべきだが、必ずしも計算で求められるものでもない。現実的な条件から設定することもある。
データ取得の方法 (前向き、後ろ向き)	前向き:これから治療等を行うときにその都度必要なデータを得る場合。後ろ向き:既にデータが得られているものを後から調査し、さかのぼってデータを集める場合。
研究の分類 介入研究か観察研究か	介入研究:研究目的で「あえて」条件を設定する場合(未承認新薬の投与、特別な治療法の施行など)。 観察研究:治療の一環として得られるデータを活用する場合。

#### ③ 群分けの必要性を考える

一般には対照群を設定し群間比較を行うことが適切とされることが多い。しかしながら治験などとは異なり医療者が行う研究では、必ずしも比較が目的ではない場合もあり医療現場ならではの制限を受けることもある。

クリニカルクエストの内容によっては対照群を必要としない場合もある。第2章の例のように、患者群における倦怠感症状の程度が治療や予後日数の経過によってどのように変化するかを見たい場合、単一の患者群での症状を経時的に評価する。一方で治療法や病気の違いで結果を比較したい場合には比較群を設定する。また、倫理的側面から考えて対照群(あえてその治療を受けない患者群)を設定できない場合がある。教科書にあるデザインに縛られず、現状に応じた実施可能な試験デザイン、かつ自分の目標を達成するために必要なデザインを提案する必要がある。ただし、対照群が設定できないことはひとつの欠点ではあるので、謙虚に認めて研究限界のひとつとして受け入れる必要はある。

#### ④ 分割表で群わけの意味を考えてみる

因果関係を解析する方法のひとつとして分割表を用いた解析がある。オッズ比による検定やカイ二乗検定などが代表的である。研究目的を達成させるために得るべき測定項目のことをアウトカムという。降圧効果なら反復投与後の血圧低下値が考えられる。

2群を設定した試験の結果を整理する方法のひとつが分割表データの解析である。表のように、例えば薬剤1と薬剤2とで一定期間内に副作用が発生した症例数を比べる場

合、カルテ調査を行って患者データを得て、薬剤1を投与された患者 A+B 人のうち

	副作用あり	副作用なし	合計
薬剤1	A	B	A + B
薬剤2	C	D	C + D
合計人数	A + C	B + D	A + B + C + D

副作用を発症した患者は A 人、といったように具体的人数を調べて数値を記入し整理する。この分割表の数値を統計的に解析することで群間比較や因果関係の強さを評価できる。

A~D の人数を得る方法は、(方法1)それぞれの群の患者の今後の治療経過を追うことで「これから」得られる副作用情報を集計する場合と、(方法2)既に治療を終えている患者の「今までの」副作用情報調べて集計する場合がある。後者では副作用があった患者、なかった患者をカルテ等で先に調べて、実際にこれらの患者はどちらの薬剤を投与していたのかをさかのぼって調べることになる。

#### ⑤ 観察研究～ケース・コントロール研究とコホート研究

比較的实施しやすい研究のひとつとして、治療時に得られるデータを研究に活用する観察研究がある。観察研究では研究目的のために特別の処置などは行わず、日常診療・治療の中で得られるデータを活用する。一方で研究目的のために患者に特別な条件を設定する場合を介入研究という。

観察研究にはケース・コントロール研究とコホート研究とがある。ケース・コントロール研究とは、カルテ調査などの方法で評価対象となるイベント(例えば、関心のある副作用発症)が得られる症例(ケース)と、対照(コントロール)となる症例(関心のある副作用が発症しなかった症例)を選び出してそれぞれ集計して評価する方法である。治療が終わっている患者のデータを集計し条件をさかのぼって調べることが多く、通常は後ろ向き研究となる。

一方で、処置群、対照群といったコホート(集団)と呼ぶ群を設定してデータを集積し評価する方法をコホート研究という。コホート研究では多くの場合、治療開始前に計画を立てることで治療が進むにつれデータが得られることが多く、通常は前向き研究となる。既にコホートに分けられた成績を事後的に(後ろ向きに)調査するコホート研究もある。

#### ⑥ 比較条件以外は揃えることが原則

比較試験の場合、比較したい条件以外の患者背景(年齢、疾患、使用薬剤など)は同じという前提が必要になる。ケース・コントロール研究では、コントロール群の患者を選択する際にはケース群の患者背景と同じになるような症例を選ばなければならない。実際的には、観察研究であれば、まずは選択・除外基準を満たす症例をすべて集積し、その後データを精査して採否を決めることが無難であろう。

#### 4) 試験デザインと統計

##### ① 必要症例数(目標症例数)

臨床研究では、計画時に目標とするデータ数、すなわち必要症例数をあらかじめ算出する。これは解析的な面からだけでなく、被験者(患者)への配慮の意味も込められている。必要症例数とは目的とする「差」を統計的に適切に見出すために必要かつ最小限の患者数のことで、症例数が少なければ統計的な結論が明確にならず試験の目的が達成できないことがあり、一方症例数が多すぎると試験に参加する必要のない患者にまで負担をかける可能性がある。適切な症例数をあらかじめ計算しておくことが必要とされ、例えば2群間比較を行う場合の必要症例数は、(1) 想定されるデータのばらつき、(2) 「効果あり」などと判断するときの差(検出差という)、(3) 有意水準、(4) 検出力、を定めて計算する。

医療現場で行う研究では、事後的にカルテ等を調査して可能な限りの症例数のデータを得ることにする、あるいは対照群を設けた比較検定を行わない、という理由で必要症例数を算出できない場合もある。また、必要症例数を設定できたとしても治療が最優先であり思い通りにならないこともあろう。実際的な都合、例えば、対象患者がその施設で年間何例か、調査期間を長くすれば十分な症例数を確保できるか、といったことも考慮して設定根拠を示す場合もある。比較試験ではない場合には、過去の論文報告などを参考に現実的かつ統計的にも妥当だと思えるような必要症例数を理由とともに示すこともある。

数式通りには進まないことは特に医療にはよくあることで、計算が必須なのではなく症例数の設定理由が合理的であると説明できることが大切である。

##### ② 解析方法と欠損値の扱い、データどうしの相関

具体的な解析方法も計画の段階で決めておくべき事項のひとつである。目的に応じて、またデータの質、検査値などの連続的な値なのかカテゴリスコアといった離散的な値なのか、やその量も考慮した上で手法を正しく選択する必要がある。具体的な手法の選択についてはその都度調べる必要があるが、少なくとも計画段階で「自分はどんなデータを得て、どんな評価をしたいのか」を整理・理解しておくことは必要であろう。なんとなくデータを取ってみて、あとから専門家に相談するとなると、専門家も苦労が増してしまうし、そもそも科学的とは言えない。

臨床研究を進める上では、必ずしも必要なデータがすべての患者から得られるとは限らず、「欠損値」の問題は避けられないことが多い。理想的には被験者全員からすべての必要なデータが得られることが望ましいが、一部の検査値の測定値が正しくないとか、カルテ等の記録に残っていないこともある。欠損値がある場合、その症例を用いない(削除)、他の症例から推測して埋め合わせる(補完)といった操作がある。補完には、欠損した数値を他の被験者の同じ項目の平均値や中央値で代用する方法、直前に同じ患者で得られたデ

ータを欠損値に代用する方法、などがある。どの対処方法を取ったとしても本来得られる情報からは「欠損」した内容になるので、必ずバイアスが生じることを認めた上で計画の段階で欠損値の扱い方についてもきちんと定めておくべきである。

臨床研究ではデータ間に相関性があることが多い。新生児の年齢と体重の相関、検査値どうしの相関、患者間でのデータの相関、などがあり、具体的には実測値をプロットして視覚的に確認した上で相関係数を調べる。

③ その他計画段階で考慮すべきこと

臨床研究では必ず倫理委員会の承認を得る必要がある。病院内か薬剤師会等の団体かなど、実施可能で「信頼できる」倫理委員会を調査・選択し、審査に必要な資料を作成することが計画段階で必要になる。特に患者へのインフォームド・コンセントを得るための資料や、個人情報の保護については具体的な行動方針を明確に示さなければならない。

◇ **練習問題**

設問1) 試験デザインを決める上での具体的項目を列挙してみよう。

設問2) これから取り組んでみたいCQを設定し、それに必要な統計解析を考えてみよう。

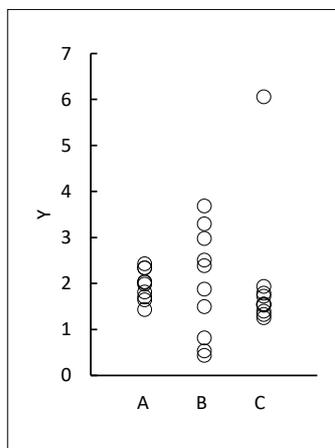
## 2. 臨床研究に必要なデータ解析の基本姿勢

### 1) データを得たらまず行うべきこと

患者からいただいた大切なデータである。データが得られるとすぐに統計ソフトを用いて統計解析を開始するのではなく、自分で得た「生の」データをじっくりと眺めて、データの語るところをよく読み取るように心がけたい。

#### ① データを得たらまず実測値をプロット

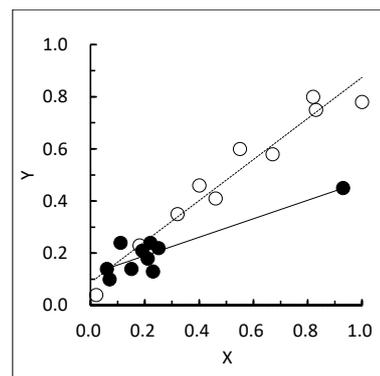
データから平均値や標準偏差といった要約統計量が計算できる。要約統計量はデータ全体の様子を知るための指標となるが、「データの内容を要約したもの」であり過信すると間違った判断をしてしまうこともある。統計解析の初期の段階でミス为了避免するためには必ず生データのプロットを眺めることから開始した上で要約統計量を算出し、隠れた情報がないかどうかを確認するとよい。



左図は3組のデータについて個々の値をプロットしたもので、A、B、C どの群でも平均値はほぼ同じであるが、実際にはCには極端に大きな数値、外(はず)れ値が含まれていて、平均値や標準偏差に影響する。外れ値を見落さないように個々のプロットが重要である。

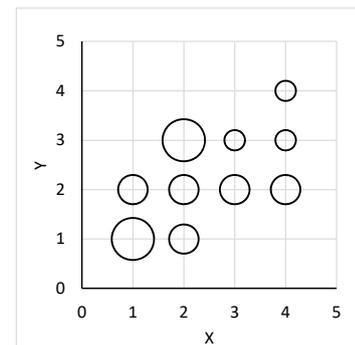
回帰分析を行う場合も同様に、右図で○は直線回帰に問題はないが、●では右端の1点が外れ値の可能性があり回帰分析の結果に影響を及ぼし、

この1点を除くと相関は無い可能性がある。プロットを見ずに正の相関があると結論付けると誤った考察につながる危険性が高い。自分のデータは大切に扱いたい。



#### ② 同じ数値のデータをグラフで識別する

カテゴリ変数といわれる測定値、例えば疼痛強度やQOL スコアなどの評価値は通常整数値として得られ離散型変数とも呼ばれる。これをプロットするとき、他の被験者のデータが重なってしまうことがあるため、右図のように同じ値のデータの個数を大きさで表すバブルチャートなどの工夫を行うことがある。



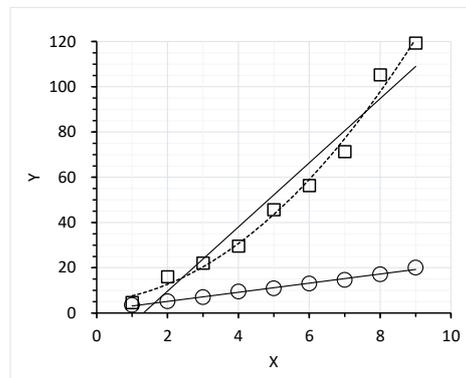
## 2) 知っておきたい統計の基本事項

### ① データの性質 ～連続値か離散値か～ で解析法は異なる

臨床研究で得られる2変量(X、Y)の相関性を示したいとき、もとのデータが連続的か、離散的かによって方法を使い分ける。相関係数としては、前者はピアソンの相関係数を、後者はスピアマンの相関係数を区別して使い分ける。相関の有意性の評価については、適切な回帰分析を行ってその回帰係数の有意確率(p値)や回帰係数の推定値の信頼区間から判断する。比較検定の場合でもデータの性質によって用いるべき方法が異なる。

### ② 相関は直線関係とは限らない

回帰分析を行うとき、理解しやすく操作も簡便なので直線を引く、すなわち直線回帰(線形回帰)を適用することが多いが、相関は必ずしも直線関係とは限らない。実測値をプロットして確認し、適切な回帰モデルを用いる。図に□、○の2組のデータについて直線回帰した結果を実線で示したが、□は中央が凹んでいるので二次関数が適切であった。直線以外の式、あるいは回帰分析を非線形モデルとか非線形回帰分析という。世の中の現象のほとんどは厳密には非線形モデルで説明されるので、非線形回帰分析にも慣れ、実測値のプロットに忠実な回帰式に用いる式を選ぶようにしたい。



### ③ 薬物速度論も非線形回帰分析のひとつ

薬物の体内での動き、すなわち薬物濃度推移を評価する回帰式は特に薬物速度論モデル解析とも呼ばれ、例えば TDM (Therapeutic Drug Monitoring) で活用されている。非線形モデル解析は医療の現場でも役に立っていることを知り、ぜひ馴染んでもらいたい。

## 3) 基本的な統計用語の整理

統計解析の基本となる事項について、特に重要なポイントを示した。

### ① データの分布～特に正規分布と対数正規分布

臨床研究に限らず実験で得られるデータは「ばらつき」を含むので、ばらつきを無視してデータを解釈することはできない。ばらつきの形状は分布として整理され、ヒストグラム(度数分布)でグラフ化できる。この分布の形状を数式で表せるように種々の統計分布が決められている。医薬領域で扱うデータは正規分布あるいは対数正規分布で説明される場合が多く、その他ポワソン分布や二項分布などの分布が用いられる。

## ② 要約統計量～平均値、分散、標準偏差、標準誤差

実測値の羅列やヒストグラムだけではデータの特徴を表すことが難しく要約統計量が用いられる。要約統計量には、平均値や中央値などの「データの中心」を示す「代表値」と、分散など「データのばらつきの程度」を示す「散布度」とがある。多くの場合は正規分布を想定した要約統計量を用いるが、対数正規分布を想定する時もある。

### 主要約統計量

算術平均 ～ 全データの和を個数で割った値。正規分布の平均値。

幾何平均 ～ すべてのデータの対数値を用いて得られる算術平均値。

中央値 ～ データを大小順に並べたときの真ん中の値。

偶数個ある場合には真ん中 2 個の算術平均値。

最頻値 ～ ヒストグラムで最もデータ数が多い「範囲(を代表する値)」。

分散 ～ 各データの平均値からのばらつきを示す値。

「標本分散」と「不偏分散」とがあり、通常は不偏分散を用いる。

標準偏差 (SD) ～ 分散の平方根。通常不偏分散の平方根。

標準誤差 (SE) ～ SD をさらにデータ数 N の平方根で割った値。

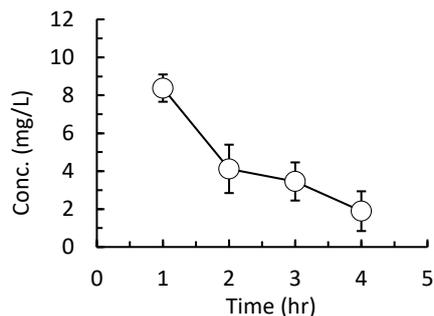
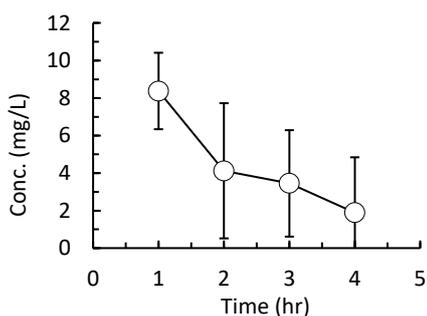
一般には「統計量のばらつき」を示すが、

「平均値のばらつき」と理解するとわかりやすい。

外れ値 ～ 分布から大きく離れた数値。

## ③ 標準偏差 (SD) と標準誤差 (SE) の使い分け

データをグラフ化するとき、SD あるいは SE の幅を「バー」で示す場合がある。平均値±SD (あるいは SE) である。図に同じ数値について、標準偏差 (SD) (左)、あるいは標準誤差 (SE) (右) で誤差幅を示した。どちらを用いても構わないが、グラフの説明にはどちらの数値を用いたのか明記するのがマナーである。SE が「見栄え」はよいといった理由で SE を選ぶことを優先するのは科学として間違っている。データの内容を他者に正しく(正確に)伝える姿勢があれば、おのずとデータ解析作業に真摯さが出るはずである。



### 3. 患者背景データの整理と比較検定

どのような目的の臨床研究でも、まずは対象患者の背景データ、すなわち年齢や体重、臨床検査値の値といった基本的な項目を集計することから始まる。臨床論文でも最初の表に患者背景データの集計結果と、必要であれば群間比較を行った結果が示される。

#### 1) データをそのまま整理する

データが得られたらまず、患者ごとの数値を表にまとめることから始める。このとき EXCEL などの表計算ツールを用いることが望ましく、またあとでデータの加工がしやすいように、一人の患者について一行にまとめると便利である。またセルの結合は後で表を修正するとき面倒になるので避ける。例を示す。

実際にはまだ右に長く続くが、ID が患者識別番号、Group はグループを区別する変数、Gender は性別

A	B	C	D	E	F	G	H
	ID	Group	Gender	Age (y)	BW (kg)	PS	Stag
	102	1	1	56	72	0	4
	103	1	1	63	59	1	4
	104	1	2	52	72	0	4
	105	1	2	66	38	1	3

(男性 1、女性 2 などと定める)、続いて年齢、体重、ECOG-Performance Status (PS) と続く。別途グループ番号 2 の行も定義すれば、グループ間での患者背景の比較も行える。

経験から、連続変数は正規分布に従うと仮定することが多いが、外れ値が存在する可能性もあるので必ずグラフで確認する。

#### 2) 連続変数の 2 群間比較

##### ① パラメトリック検定とノンパラメトリック検定

2 群間の比較を行う手法を選ぶとき、データに正規分布などの分布を仮定できる場合は パラメトリック検定 となり、t 検定が代表的である。一方で、NRS スコアなど整数値だけをとる順序データや、外れ値を含むデータの場合には ノンパラメトリック検定 に分類される方法を用いる。

##### ② 対応のある検定、対応のない検定

さらに 2 群間の比較検定では、2 つのデータに「対応」関係があるかどうかで方法が異なる。「対応」とは 2 つの条件で得られたデータが同じ患者から得たものか否か、という意味で、例えば反復投与の初回と最終回で同じ患者から臨床検査値を得て前後比較する場合には 対応のある検定 を用いる。2 群で異なる患者グループから得たデータであれば 対応のないデータ となる。

③ 2群間で「分散」が等しいか否かで手法を区別

さらに、対応のない、パラメトリック検定の場合にはまずは2群間でのデータの分散を比較し、分散が同じであれば Student の t 検定 を、分散が異なれば Welch の t 検定 を用いるとされている。分散の比較には F 検定 を用いる。

④ 片側検定と両側検定

2群間で平均値を比べる場合は、多くの場合 両側検定 (Two-sided Test) を用いる。平均値の大小があらかじめ仮定できる場合や、F 検定 や カイニ乗検定 では片側検定 (One-sided Test) を用いる。一般的には両側検定を用いる、と理解しておくのがよい。

⑤ 有意水準

有意差がある、ない、を判断するときには検定で得られる 有意確率 (p値) を用いる。多くの場合 有意水準 ( $\alpha$  と記載、差が無いのに差があると判断を間違ってしまう確率) を 0.05 (5%) 未満と定めて  $p < 0.05$  であれば「有意差あり」と結論付ける。 $\alpha = 0.05$  以外にも検定の目的によって  $\alpha = 0.01$  などを設定する場合がある。後述する 検定の過誤 についての理解が必要である。

⑥ 2群間比較検定手法の整理

検定法の名前を整理した。データの内容を吟味して正しい手法を選択する。パラメトリック検定の場合、F 検定に続いて t 検定を行うという順で2段階で行う。

2群間平均値比較検定方法の基本的手順

1) パラメトリック検定かノンパラメトリック検定かを決定する。		
2) パラメトリック検定の場合、F 検定により2群間の分散の差を調べる。		
3) 次表により適切な検定方法を選択する。		
	パラメトリック検定	ノンパラメトリック検定
対応がない&等分散	Student の t 検定	マンホイットニー検定 (ウィルコクソン検定)
対応がない&不等分散	Welch の t 検定	
対応がある	Paired-t 検定	ウィルコクソン符号付順位検定

ノンパラメトリック検定では、2群間のデータについて大小の順位をつけ、その和を比較することで分布の違いを評価するので、順位検定とよばれる。対応のあるデータの場合、パラメトリック検定は 対応のある t 検定 (Paired-t 検定) を用い、ノンパラメトリック検定では 対応のあるデータの前後差の符号を集計して判断するので「符号付き」と称される。

### 3) 基本統計量と検定結果を表にまとめる

#### ① 表の整理

Age データについて正規分布を仮定し t 検定を行った例を示す。2 群のデータの分散の差を F 検定で確認したところ  $p=0.756$  となり、分散に差はないことが確認できた。この結果をもとに、Student の t 検定により平均値の差を検定したところ、 $p=0.184$  で平均値に差はないことが確認できた（参考までに Welch の t 検定では  $p=0.1912$ ）。表のように各グループで得られた数値とともに、右端に p 値を、さらに表の脚注に用いた検定法を示すとわかり

やすい表となる。他にも中央値や変動係数などを示す場合もある。

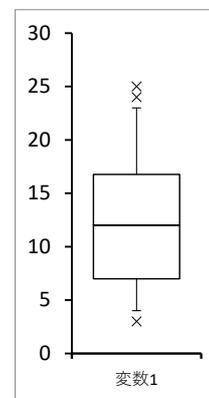
	Group 1				Group 2				p-value *
	N	Mean	S.D.	Range	N	Mean	S.D.	Range	
Age (y)	30	65.8	6.07	52 - 76	20	68.3	6.43	51 - 81	0.184

\* by t-test

#### ② データをグラフ化するときのグラフの種類

患者の背景データをグラフ化するとき、グラフの種類をデータの内容によって使い分ける。パラメトリック検定を用いたデータであれば、平均値±SD（標準偏差）のグラフが適切であろう。一方、ノンパラメトリック検定を用いたデータであれば、中央値や四分位点（25%点や75%点）を示すボックスプロット（箱ひげ図）を用いることが適切である。

例としてステロイドの投与日数をボックスプロットで示した（図）。中央値と四分位点を横線で示し、95%区間を「バー」で、外れ値を X で示している。こういった表記内容は任意に設定できる。



### 4) カテゴリ変数の群間比較

#### ① 2つの値（二値）のみをとるデータの比較

患者の性別について、グループ間での男女比が異なるかどうかを確認したいとき、カテゴリ変数なので右のような分割表を作成する。数値はそれぞれの項目に相当する患者数で、例えばグループ1の男性患者は17名となる。2グループ×男女 2 とおり、の分類なので 2×2 分割表となる。

	Male	Female	Sum
Group 1	17	13	30
Group 2	11	9	20
Sum	28	22	50

表から、各グループでの男性の割合は、 $17/30=0.567$ 、 $11/20=0.55$  で、この比に差があるかどうかを、ここではカイニ乗検定で検定した。その結果  $p=0.907$  となり、グループ間で男女人数に偏りはないと判断できる。結果を表に整理した。

	Group 1		Group 2		p-value *
	N	Male / Female	N	Male / Female	
Gender	30	17 / 13	20	11 / 9	0.907

\* by Chi-squared-test

分割表にある患者人数が少ないときにはカイニ乗検定のかわりにフィッシャーの正確確率検定を用いるとより正確な結果が得られる。同じデータでフィッシャーの検定では  $p=1.00$  となった。

### ③ 3つ以上の値を取るカテゴリ変数の場合

例としてがんステージを4段階評価したデータを考える。分割表に集計すると表のようになり、 $2 \times 2$  分割表ではないがカイニ乗検定が適用できる。ただし、度数(患者人数)に0が含まれると正しく計算できないことがあるので、ステージが1の場合を除外して  $2 \times 3$  分割表として解析すると、 $p=0.147$  となりグループ間でステージの値の分布に違いはないと判断できた。結果を表に整理した。

	1	2	3	4	Sum
Group 1	0	3	12	15	30
Group 2	0	5	10	5	20
Sum	0	8	22	20	50

Stage	Group 1		Group 2		p-value *
	Stage	N	Stage	N	
	1	0	1	0	0.147
	2	3	2	5	
	3	12	3	10	
	4	15	4	5	

\* by Chi-squared-test

ここで改めて、読者には何かひとつ臨床論文で実際の患者背景表を確認していただきたい。それぞれの項目や数値がどのようにして得られたのかが解れば論文を読みやすくなるし、今後自分自身のデータに迷うことなく適用できると思う。場合によっては、聞きなれない統計解析手法が出てくることもあるが、その都度調べて知識を増やしてほしい。

5) 検定の考え方～帰無仮説、検定の過誤

① 統計的有意差とは

統計的検定では、「差がない」、「因果関係がない」ことを帰無仮説として、この仮説が正しいかどうかを客観的に評価する。帰無仮説が正しくないとき対立仮説として「差がある」、「因果関係がある」という結論を採用する。結論は統計解析で得られる p 値で表す。P<0.05 なら有意差あり、という認識だけが先走る感があるが、p 値は「有意差があるという結論が間違っている確率」であり単純に差の有無を断言する指標ではない。

② t 検定を例に平均値の差の検定結果を見るポイント

2 群間での分散の差について EXCEL の分析ツールを用い F 検定を行った結果を示す(左図)。F 検定は片側検定のみで p=0.303 となり分散には差が無いと判断できる。分散に差がないとする Student の t 検定の結果を示す(右図)。両側検定の結果 p = 0.00320 (<0.05)

で平均値に統計的に有意な差がある。

	変数 1	変数 2
平均	71.2	50.2
分散	81.2	46.7
観測数	5	5
プールされた分散	63.95	
仮説平均との差異	0	
自由度	8	
t	4.15211166	
P(T<=t) 片側	0.00159984	
t 境界値 片側	1.85954804	
P(T<=t) 両側	0.00319968	
t 境界値 両側	2.30600414	

	変数 1	変数 2
平均	71.2	50.2
分散	81.2	46.7
観測数	5	5
自由度	4	4
観測された分散比	1.73875803	
P(F<=f) 片側	0.30260017	
F 境界値 片側	6.38823291	

③ 検定の過誤、検出力～第一種の過誤、第二種の過誤

仮説検定においては表に示す 4 通りの判断が想定され第一種の過誤、第二種の過誤が定義される。

	検定結果にもとづく「判断」	
事実	「(帰無) 仮説が正しい」と判断	「(帰無) 仮説が誤り」と判断
(帰無) 仮説は正しい	○(問題なし)	第一種の過誤 有意水準(危険率) = $\alpha$
(帰無) 仮説は誤り	第二種の過誤 検出力 = $1 - \beta$	○(問題なし)
「あわてんぼうの $\alpha$ 、ぼんやりの $\beta$ 」と覚える。通常 $\alpha = 0.05$ 、 $\beta = 0.2$ が用いられるが、特に $\alpha$ は目的によって例えば 0.01 などと定めることもある。		

第一種の過誤の確率( $\alpha$ )とは、「実際には差がないのに『あわてて』差があると間違ってしまう確率」で、第二種の過誤の確率( $\beta$ )とは、「実際には差があるのに『ぼんやりと』その差を見逃してしまう確率」で、 $1-\beta$ は「実際に差があるときに『きちんと』差があると判断できる確率」で「検出力」とよばれる。どちらの過誤もできるだけ小さくなるのが好ましいが、多くの有意差検定では第一種の過誤を犯す確率(危険率)を定めて判定を行い、第二種の過誤については詳細な評価は行わないこともある。

#### ④ 例数設定

臨床研究では事前に患者症例数を設定すべきで、必要症例数の算出を行うことがある。必要症例数の算出式は検定法により異なる。例えば、二群間比較であれば見出したい差(検出差)やデータのばらつきの大きさといった数値をもとに、あるいは生存時間データの解析でログ・ランク検定を行う場合には、2群それぞれの生存率やハザード比といった数値をもとに計算される、目的に合った解析を決めた上で症例数算出を行う。比較を行わない場合には別途合理的な理由で必要例数を設定することが望まれる。

#### ◇ 練習問題

設問1) パラメトリック検定とノンパラメトリック検定の違いについて説明してみよう。

設問2) 計画時に必要症例数が計算で求められない例を挙げてみよう。

#### 4. 多群間比較、信頼区間と生物学的同等性試験

##### 1) 多群間比較

##### ① 多重性の問題

3 つ以上の異なる条件での比較検定を多群間比較(多重比較)といい、2 群間比較とは区別して扱う。例として、異なる薬剤 A、B、C 投与後のある臨床検査値を測定した結果(表)を考える。

	1	2	3	4	5	6	Mean	S.D.
A	36.7	38.7	38.4	38.5	37.2	39.5	38.17	1.03
B	38.7	39.1	40.1	39.1	39.9	37.5	39.07	0.93
C	39.4	40.4	39.6	40.5	41.6	41.7	40.53	0.97

3 群以上のデータの場合、「対(つい)」ごとに 2 群間比較の検定方法を 2 回以上用いると(例えば t 検定を A-B、B-C、C-A の全組み合わせで行うと)、検定したときに有意差が出やすくなるという計算上の問題点が起こる。このように、何度も同じデータを使って検定を繰り返すことで有意水準が甘くなり「差がある」と誤って判断してしまう問題を検定の多重性(の問題)といい、多群間(多重)比較法を用いることで多重性を解決できる。同じデータは 2 回用いないこと、が重要なポイントである。

##### ② 多重比較法

代表的な方法として分散分析法(ANOVA)や、チューキー検定、ダネット検定などのパラメトリック多重比較検定や、クラスカルワリス検定やフリードマン検定といった ANOVA に対応するノンパラメトリック検定法がある。主なパラメトリック多重比較法を整理した。実際にどの方法を用いるかは、比較したいデータ対の種類や、その研究分野で汎用されていて信頼性のある方法を選択するなどして決定することが望ましい。

チューキー法	全ての群間の比較を対ごとに行う。シェフェ法も同じ目的で用いられる
ダネット法	ひとつの対照群と他の全ての群との比較を行う
ボンフェローニ法	有意水準を検定ペア数で「調整」して多重性を考慮 …例えば 3 群間(A、B、C とする)の比較では検定の組み合わせは 3 通りなので、有意水準を $0.05 \div 3 = 0.0167$ として 2 群間比較検定を繰り返し行う。直観的に理解しやすい方法である。群の数が増えすぎると検出力が悪くなり 3~4 群までにすべきとの説もある

主なノンパラメトリック多重比較法として、スティールデュワス法(チューキー法に相当)、スティール法(ダネット法に相当)、ダン法(ダネット法に相当)などがある。

方法によって解析結果が異なるので、あらかじめどの方法を用いるかを決めておく必要がある。いろいろ試してみても結果的に最も都合がよい方法を採用することはサイエンスとして不適切である。

## 2) 信頼区間

データを要約する数値として、平均値、標準偏差から信頼区間 (Confidence Interval, CI) が定義される。データに正規分布を仮定したとき、その「平均値の 95%信頼区間」は、

$$\text{平均値の 95\%信頼区間} = \text{平均値} \pm 1.96 \times \text{SE}$$

で計算できる。SE は標準誤差で標準偏差 (SD) をデータ数の平方根で割った値である。数値 1.96 は、正規分布を仮定した時の 95%幅に対応する数値であり、もし 90%信頼区間を求めるなら 1.96 ではなく 1.64 を用いる。データにどのような分布を仮定するかによって信頼区間の計算法は異なるので、上式がすべてのデータにあてはまるわけではない。

信頼区間を用いて検定を行うこともある。例えば 2 群間の平均値の差の信頼区間が  $-0.15 \sim 0.23$  と得られたとき、この幅は 0 (=平均値に差がない) を含むので、2 群間の平均値に有意差なしと結論付ける。もし信頼区間が  $0.25 \sim 0.76$  となれば 0 を挟んでいないので平均値に有意な差があると判断する。

比の信頼区間を考える場合もある。因果関係の程度を表すオッズ比や死亡の要因となる程度を表すハザード比では、2 群間のオッズあるいはハザードという値の比を評価するので、比が 1 (=比率に違いがない) を基準にする。例えば、比の信頼区間が  $0.75 \sim 1.25$  であれば 1 を挟むので比率に有意差がなく、比の信頼区間が  $0.23 \sim 0.84$  であれば 1 を挟まないため比率に有意な差があると判断する。

後発医薬品の開発では先発品と後発品との間の「生物学的同等性試験が」実施され、統計的に両製剤が「生物学的に同等」であることを証明しなければならない。生物学的同等性の判定基準は「信頼区間法」で行われ、ガイドラインには次の記載がある。

Cmax、AUC について、試験製剤 (後発品) と標準製剤 (先発品) の生物学的同等性判定パラメータの対数値の平均値の差の 90 %信頼区間が  $\ln(0.80) \sim \ln(1.25)$  の範囲にあるとき、試験製剤と標準製剤は生物学的に同等と判定する

\* Cmax: 最大血中濃度、AUC: 血中濃度曲線下面積

### ◇ 練習問題

設問1) 多重性の問題、および多重検定法の意義について説明してみよう。

設問2) 信頼区間を使って検定する場合の「差」と「比」の使い分けについて整理してみよう。

### 3) 検定方法の整理

いくつかの代表的な解析手法を整理した。必ずしもこれらの方法が全てではないので、どのようなデータを解析したいかを理解し適切な解析法を選択することが大切である。

#### ○ 解析するデータは

A: 計量値(連続変数)である ⇒ 1

B: 計数値(カテゴリ変数など不連続値)である ⇒ 2

#### 1 計量値データ

##### 1-1 外れ値を確認する(棄却検定、正規確率プロットなどを利用)

⇒ 外れ値がある ⇒ 実験をみなおして原因を考える

⇒ 実験に問題がある ⇒ 実験をやりなおす

⇒ 実験に問題がない ⇒ 外れ値の採否を判断する ⇒ 外れ値の扱いを把握し 1-2 へ

⇒ 外れ値がない ⇒ 1-2 へ

##### 1-2 分布の型を確認する

⇒ 正規性が仮定できる(パラメトリック検定を適用する)

⇒ 2 群間比較である

⇒ 対応がある ⇒ 対応のある t 検定 (paired-t 検定)

⇒ 対応がない ⇒ 分散が等しい ⇒ Student の t 検定

⇒ 分散が等しくない ⇒ Welch の t 検定

⇒ 多群比較である ⇒ 分散が等しい

⇒ 要因(因子)は 1 つで X に順序がない

⇒ 一元配置分散分析、多重比較法

⇒ 要因(因子)が 2 つ ⇒ フリードマン検定

⇒ 正規性が仮定できない

⇒ 適切な変換(対数変換など)で正規化できる ⇒ 正規化して 1-2 へ

⇒ 正規化できない(ノンパラメトリック検定を適用する)

⇒ 2 群比較である ⇒ 対応がある ⇒ ウィルコクソン符号付順位検定

⇒ 対応がない ⇒ ウィルコクソン順位検定、マンホイットニー検定

⇒ 多群比較である ⇒ 要因(因子)は 1 つ ⇒ ノンパラメトリック多重比較検定

#### 2 計数値データ

##### 2-1 度数(分割表など)か整数観測値かを区別する

⇒ 度数である ⇒ 分類区分は 2 区分(要因あり、なし、など)である

⇒ 2 群比較である ⇒ カイニ乗検定、フィッシャーの正確検定

⇒ 多群である ⇒ カイニ乗検定

⇒ 整数観測値である ⇒ 基本的には 1-2 のノンパラメトリック検定

#### (参考) 尺度の種類

- ・名義尺度 ~ 群を区別するためにだけ用いられる尺度。血液型を 1、2、3、4 と対応づける場合など。
- ・順序尺度 ~ 大小関係の順序に意味がある。効果を悪化(1)、不変(2)、改善(3)、著効(4)とするなど。
- ・間隔尺度(距離尺度) ~ 数値の差に関心がある場合。セ氏で表した気温、西暦など。
- ・比例尺度(比率尺度、比尺度) ~ 数値の差および比に関心がある場合。身長、物価など。

## 5. 回帰分析とモデル解析

### 1) 回帰分析と相関係数

#### ① 独立変数と従属変数、単回帰と重(多変量)回帰

回帰分析とは、従属変数(目的変数)と独立変数(説明変数)との関係を数式で説明する方法で、独立変数がどのように従属変数に影響するかを説明する。例えば検量線の式として、「物質濃度(従属変数) =  $\alpha \times$  吸光度(独立変数) ( $\alpha$  は直線の傾きを表す数値定数)」がある。あるいは、腎機能の指標であるクレアチニンクリアランスが、年齢 (Age)、体重 (WT)、血清クレアチニン (Scr) によって次の Cockcroft-Gault 式で説明できる(腎クリアランスを計算する式としては eGFR 計算式など他にもある)。

$$\text{Ccr(mL/min)} = (140 - \text{Age}) \times \text{WT} / (72 \times \text{Scr}) \quad (\text{女性にはこの値を } 0.85 \text{ 倍する})$$

前者の場合は独立変数がひとつで単回帰分析、後者は独立変数が複数で重回帰分析(多変量回帰分析)といい、それぞれの式が実際のデータを「うまく」説明する  $\alpha$  の値や 140、72 といった係数の値が決められる。臨床研究では複数の要因がひとつの結果に関係することが多いので重回帰分析を用いることが多い。

#### ② 線形回帰と非線形回帰

独立変数と従属変数との間に直線関係を想定する場合を線形回帰分析(線形モデル)、線形回帰以外を非線形回帰分析(非線形モデル)という。先の検量線は線形モデル、クレアチニンクリアランスは Scr の逆数に関係するので非線形モデルとなる。検量線の傾き  $\alpha$  といった回帰式の係数は回帰係数と呼ばれ、最小二乗法という数学的方法で求められる。

#### ③ 相関係数と回帰係数の有意性

変数間の相関性の強さを表す指標として相関係数が用いられ、記号  $r$  (あるいは  $R$ ) で表される。相関係数は次のことがらがポイントとなる。

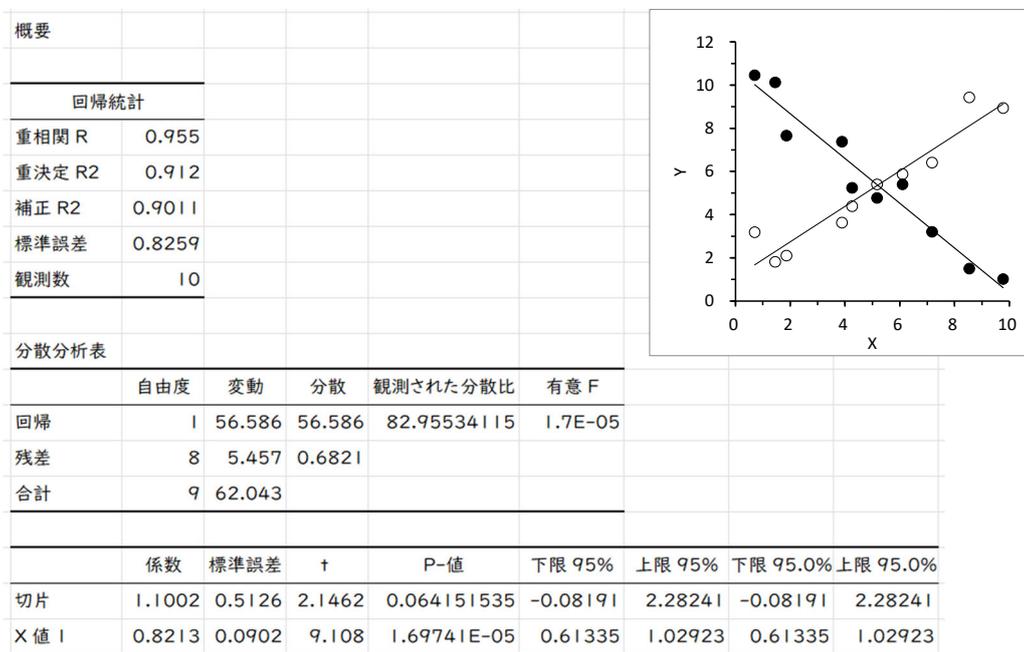
1.  $-1 \sim +1$  の範囲の数値である
2. 正の数なら「正の相関」(右上がり)、負の数なら「負の相関」(右下がり)の関係を表す
3. 絶対値が 1 に近いほど相関性が高い

相関係数から相関の強さを確認するひとつの目安を表に示した。あくまで目安であり、統計的有意性は回帰分析などの解析結果で得られる  $p$  値から評価する。

相関係数(絶対値、 $ r $ )	相関の強さの目安
0.7 ~ 1.0	強い相関
0.4 ~ 0.7	中程度の相関
0.2 ~ 0.4	弱い相関
0 ~ 0.2	相関なし

例として2組のデータ(○、●)に回帰分析を適用した結果を示す。EXCELなどのソフトウェアで回帰分析を行うと、回帰係数と相関係数が得られる。回帰係数の有意性、すなわち統計的に有意な相関であるかは回帰係数の有意性を示す p 値を参照して評価する。

○印のデータに対する直線回帰分析結果を示す。図には各データの結果を実測値と合わせて示した。



表に「重相関」と示されている。これは重回帰分析の場合も兼ねた表示であり、今の例は単回帰分析なので単相関の相関係数とみなす。決定係数は相関係数を二乗したものの、補正 R2とは自由度調整済み相関係数といわれ、データ数と説明変数の数を考慮して相関係数を補正したものである。「切片」は直線の式の切片(=0.821)、「X 値 1」とは「傾き=1.100」で、回帰式は  $Y = 0.821 \cdot X + 1.100$  となる。

回帰式が得られたのでそれぞれの係数の有意性を p 値で確認すると、切片では「0であるという帰無仮説に対して」 $p=0.064$ 、傾きは「0であるという帰無仮説に対して」 $p<0.001$ であり、有意水準を 5% とすると、切片は 0 と有意な差はない数値となり、傾きは有意に 0 ではない値であり、X と Y とには正の相関性がある、と判断できる。

●のデータのように、負の相関がある場合にも EXCEL では相関係数が正の数で表示されるので、自分で工夫して  $r = -0.970$  というふうに、傾きが負なら相関係数も負の値にして論文等に表示する。例えば必要な情報だけを取り出して次のように整理できる。

$$Y = -1.04 \cdot X + 10.8 \quad (r = -0.970, p < 0.001)$$

### ③ 重回帰分析

独立変数が 2 つ以上ある場合は、重回帰分析（多変量回帰分析）という。先のクリアチンクリアランスの式のように、医薬領域では多変量解析が必要な状況はたくさんある。

独立変数は、実務的に解りやすくするために、説明変数、変動要因、予測変数などと称することも多い。従属変数は目的変数ともよばれる。

重回帰分析でも単回帰分析の場合と同様に、それぞれの係数の p 値から有意性を評価する。複数個の独立変数があるので、どの変数が有意なのか、ソフトウェアでは一気に解析結果が得られるが、実際には「順序だてて独立変数をひとつずつ評価」する、逐次選択法を行っている。次のものがある。

- ・ 増加法～変数ひとつから開始し、有意であればそのまま式に残して次の変数を加え順次有意性を確認する。
- ・ 減少法～最初に全部の変数を入れた回帰分析を行い、ひとつずつ要因を減らしてみても有意なものだけを残す。
- ・ 増減法～増加法を行ったあとにさらに減少法を行う。あるいは減少法→増加法で確認する。

例を示す。  
偏回帰係数とは重回帰分析で得られた係数、推定値の標準偏差、

回帰式に含まれる変数(偏回帰係数・信頼区間等)							
変数	偏回帰係数	標準誤差	偏回帰係数の95%信頼区間		偏回帰係数の有意性の検定		
			下限値	上限値	F 値	t 値	P 値
X1	3.8278	0.1657	3.4877	4.1679	533.4234	23.0960	P < 0.001
X2	-2.0684	0.1078	-2.2896	-1.8472	368.2220	-19.1891	P < 0.001
定数項	4.4028	0.1484	4.0983	4.7072	880.3297	29.6704	P < 0.001

95%信頼区間と続き、右端には回帰係数の有意確率が示されている。

この結果では、 $X_1$ 、 $X_2$ は  $p < 0.05$  なので  $Y$  に対して統計的に有意な説明変数とわかる。一方で  $X_3$  は回帰係数が統計的に有意ではなかったので表には示されていない。以上の結果から回帰式として  $Y = 4.40 + 3.83 \cdot X_1 - 2.07 \cdot X_2$  が得られる。各係数の 95%信頼区間が 0

を挟んでいないことから係数は有意であることが確認できる。

回帰式に含まれる変数(偏回帰係数・信頼区間等)							
変数	偏回帰係数	標準誤差	偏回帰係数の95%信頼区間		偏回帰係数の有意性の検定		
			下限値	上限値	F 値	t 値	P 値
X1	3.787228	0.1701	3.437571	4.1368858	495.6833	22.26395	1.86E-18
X2	-2.08589	0.109	-2.309877	-1.861894	366.4079	-19.1418	7.59E-17
X3	0.069781	0.0675	-0.068892	0.208454	1.069888	1.034354	0.310496
定数項	4.33131	0.1635	3.995222	4.6673973	701.7468	26.4905	2.44E-20

重回帰分析では、統計的に有意な要因だけを残すのではなく、「あえて」すべての変数を含めたままで解析する全変数投入法が用いられることがある(上表)。すべての変数を組み入れることで、要因の有意性に関係なく、どの程度影響しているかを知ることができる。

結果を示す場合には、必ずしも有意な説明変数だけを示す必要はなく、全変数投入の場合の結果や、単回帰分析の結果と重回帰分析による有意な説明変数とを一つの表に併記するなど工夫できる。例を示す。

変数	単回帰分析			重回帰分析		
	偏回帰係数	標準誤差	p値	偏回帰係数	標準誤差	p値
X1	3.92	0.622	<0.001	3.83	0.166	<0.001
X2	-2.14	0.482	<0.001	-2.07	0.108	<0.001
X3	0.221	0.378	0.564	---	---	---
定数項	---	---	---	4.40	0.148	<0.001

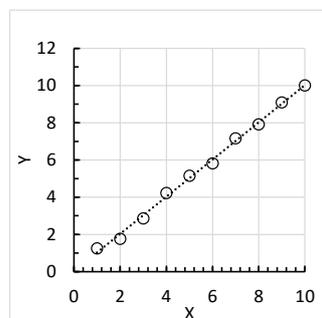
重回帰分析を行う前に、どの要因(説明変数)を変数として組み入れるのか、特に要因が多すぎる場合に判断に困ることがある。臨床的にみて目的変数に影響しそうな要因をあらかじめ考えて解析を優先する、すべての説明変数と目的変数との関係を単回帰分析で評価しておき、その相関性が高い要因を優先的に重回帰分析に組み入れる、などの工夫をすることもある。どのような理由で、どのようなプロセスで重回帰モデルを作成したのかがきちんと説明できることが大切である。

なお、一般には EXCEL 単独では非線形モデルを用いた解析を行うことは難しく、専用の統計解析ソフトウェアを用いると便利である。

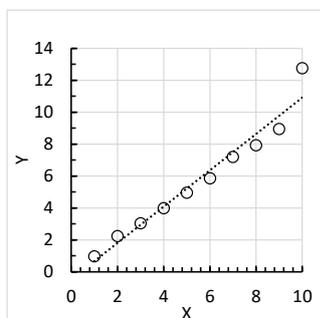
## 2) 回帰結果の診断と予測性評価

### ① 回帰結果の診断プロット

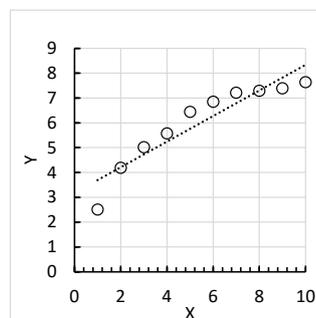
回帰分析の結果が適切かどうかは実測値とともに図に表して視覚的にも確認する習慣をつけたい。これを回帰結果の診断といい、いくつかの方法がある。例として、XとYの実測値と回帰直線(あるいは曲線)を重ね書きして確認する例を示す。



適切な回帰



はずれ値の可能性



直線回帰は不適

誤差の状況は回帰係数や相関係数の値からだけではわからないことがあるので、回帰分析のあとは各種プロットで結果の妥当性を確認する習慣が正しい判断につながる。

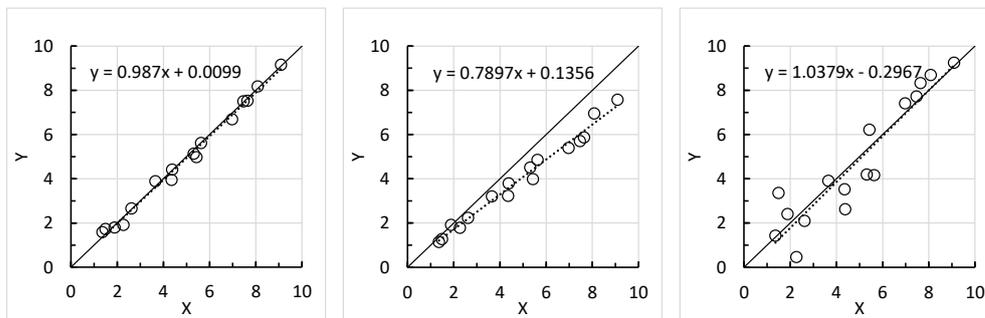
#### ④ 予測性の評価指標

回帰分析の結果を用いて新しいデータの予測を行うときがある。回帰式を求めるために用いるデータをテストデータと呼び、予測性を評価するために用いる（テストデータとは別の）データをバリデーションデータとよぶ。バリデーションデータの実測値と、回帰式から求められる予測値との関係から予測性が評価できる。

テストデータ → (回帰モデルを作る) → バリデーションデータで実測値と予測値を比較

回帰モデル(回帰式)が適切であれば、予測値と実測値が一致するが、実際には誤差が含まれるのでぴったりとは一致しない。誤差には、系統的にズレが生じる場合の誤差と、系統的ではないがばらつきにより生じる誤差がある。

図にいくつかのパターンを示した。横軸はバリデーションデータの実測値を、縦軸は予測値を示し、式および点線は回帰直線を示す。実線は 1:1 の関係を示す。



左図では回帰直線の傾きが 0.987 とほぼ 1 なので予測性は良好と考えられる。中央の図では偏りが傾向をもって大きく、全体に Y は X を過小評価しており、回帰直線の傾きも 0.788 と 1 より小さい。右図は左図と同じくほぼ 1:1 の関係だが、全体に誤差が大きいので分布が「ふくらんで」いるように見える。

予測性の評価指標として、予測の偏りの指標となる ME (mean error, 平均誤差)、予測精度の指標となる RMSE (root mean squared error, 平均平方和の平方根) がよく用いられる。ME、RMSE とともに 0 に近い値であることが望ましいが、実際には何らかの数値となる。上図の場合、左図:ME = -0.176、RMSE = 0.284、中図:ME = -1.03、RMSE = 1.12、右図:ME = -0.680、RMSE = 1.17 と得られた。

予測性の良さは、ME や RMSE の絶対値で判断するのではなく、相対的に小さいほうを優先的に用いるのが基本であるが、モデルの妥当性や簡便性なども含めて総合的に判断するのが望ましい。通常は ME によって偏りのないことを確認したり、RMSE によってばらつきがそれほど大きくないことを考察したりする、にとどめる場合が多い。

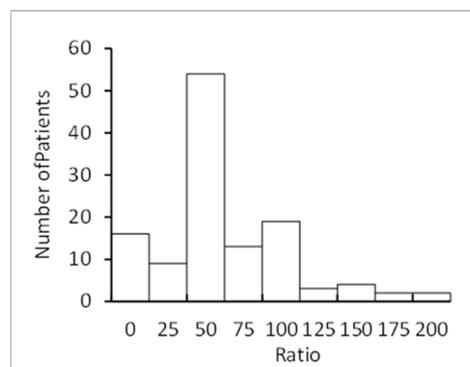
### 3) (発展)実務領域で用いる回帰分析の例

筆者らが行った臨床データを用いた事例を示す。図表は元の論文からは体裁を変更している。詳細は論文 (Am. J. Hosp. Palliat. Med., 33(2), 109-114 (2014)) を参照。

[概要]がん患者を対象に、経口オキシコドンからフェンタニル貼付剤へのオピオイド薬の変更 (オピオイドスイッチング) を行った症例を対象に、変更したときの投与量比 (換算比、Conversion Ratio, CR) について、変更理由との関係を多変量回帰分析で検討した。換算比とは変更前後の両薬剤の投与量の比で、図のようなヒストグラムが得られている。換算比に影響する要因として、年齢、性別、変更前のオキシコドン投与量 (低用量か高用量かの 2 値)、BMI、CRP 値、変更理由 (「副作用が強い場合」、「鎮痛効果が不十分な場合」、「嚥下困難」の 3 値) を検討した。なお、当時はガイドラインに従って「嚥下困難」を理由として含めていたが、現在はこの項目はガイドラインには含まれていない。

図のように、換算比のヒストグラムが対数正規分布に従うように見える。目的変数は厳密には正規分布を仮定すべきなので、ここでは換算比を対数変換した値、 $\ln(\text{CR})$  を用いた。

説明変数に二値あるいは三値データが含まれる。回帰式では、例えば変更理由については次のように、各理由に 0 か 1 かの変数を設定する。



$$\ln(\text{CR}) = \dots + b_{11} \times X_{11} + b_{12} \times X_{12} + b_{13} \times X_{13} + \dots$$

ここで、 $X_{11}$  は理由が副作用なら 1 で他の場合は 0、同様に  $X_{12}$  は理由が鎮痛効果なら 1 で他の場合は 0、 $X_{13}$  は理由が嚥下困難なら 1 で他の場合は 0 とするカテゴリ変数と定義する。係数  $b_{ij}$  は回帰分析で得られるそれぞれの理由の場合の回帰係数である。例えば、理由が副作用である患者の場合、

$$\ln(\text{CR}) = \dots + b_{11} \times 1 + b_{12} \times 0 + b_{13} \times 0 + \dots$$

となり、この患者のデータは  $b_{11}$  を求めるためのデータとして解析に用いられ、 $b_{12}$  や  $b_{13}$  の推定には関係しないことになる。

重回帰分析の結果、有意な要因のみを含めた式として、次式が得られた。

$$\ln(\text{CR}) = 4.56 - 0.613 (\text{低用量の場合}) - 0.407 (\text{鎮痛効果なしの場合})$$

換算比の対数値の式なので、この式を変形する (対数を外す) と、

$$\begin{aligned} \text{CR} &= \exp\{4.56 - 0.613 (\text{低用量の場合}) - 0.407 (\text{鎮痛効果なしの場合})\} \\ &= \exp(4.56) \times \exp(-0.613) \times \exp(-0.407) \\ &= 95.6 \times 0.542 (\text{低用量の場合}) \times 0.666 (\text{鎮痛効果なしの場合}) \end{aligned}$$

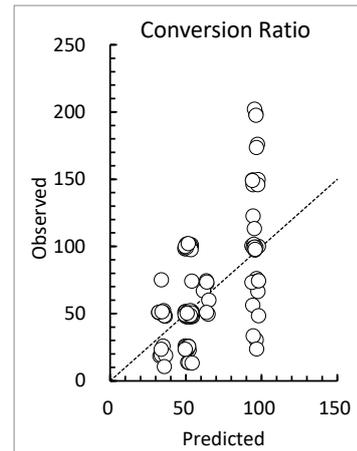
ここで用いている換算比とは、オキシコドンからフェンタニルに変更したときの、「オキシコドンの投与量をモルヒネの投与量に置き換えた値：フェンタニルの投与量」で、通常は100:1が適切とされている。今回求めた切片の値から、低用量でない、かつ理由が「鎮痛効果なし」ではない場合の換算比は95.6:1でほぼ100:1となる。さらに、

- ・オキシコドンが低用量の場合にはその0.542倍 → 換算比を51.8:1に、
- ・鎮痛効果なしの理由で変更する場合には0.666倍 → 換算比を63.7:1に
- ・両方の要因がある患者では $0.542 \times 0.666 = 0.361$ 倍 → 換算比を34.5:1と、

それぞれフェンタニル投与量を(100:1よりも多く)、約2倍、約1.5倍、約3倍に設定するのがよいとわかる。

この結果はあくまで患者全体の平均的な換算比の目安であり、患者の個人差が大きいのでこの式は個々の患者の換算比を必ずしも精度よく予測できない可能性があると考えた。そのことを確認するために、換算比について実測値とこの式から得られた予測値との比較を行い、モデル式の診断を行った。図に示す。

図の○は、各患者で得られた換算比の実測値(縦軸)と計算値(横軸)との関係を示し、点線は1:1の直線を示す。点線は○の「真ん中」を通っているが、点線の上下の広がり、すなわち各患者での実測値のばらつきが大きく、この回帰式だけでは換算比を予測することは必ずしも十分ではないことがわかり論文でも考察にも記述した。



特に臨床研究ではデータに誤差はつきものである。まずは回帰分析によりばらつきの原因をできる限り説明するように努力し、さらに統計解析では解決できないばらつきは、患者と接する医療者が患者を直接「みて」最適な医療が提供できるようにすべきと思う。

#### ◇ 練習問題

設問1) 身近なデータで、非線形な関係のありそうなデータを列举してみよう。

設問2) 回帰分析を行うときの、結果の読み方について確認すべき項目と、統計的な注意点をあげてみよう。

## 6. 分割表による因果関係の解析、ロジスティック回帰分析

### 1) 因果関係の解析とカイニ乗検定

#### ① 分割表と期待度数

因果関係を調べる際には、それぞれの要因（原因）と結果との対応を表に整理して解析する。例えば、喫煙者と非喫煙者の経過を調査し、それぞれの被験者が肺疾患を発症したか否か、あるいは薬剤 A あるいは B を服用した患者を追跡調査し、それぞれの被験者がある副作用を発症したか否か、などがある。このような因果関係を集計した表を分割表という。

分割表を作成するためには、まずは調査で得た個々の被験者の情報を整理する。EXCEL などで表形式とし、1 症例 1 行でデータをまとめると便利である。表は、被験ごとの

年齢、体重、性別とともに、喫煙の有無、肺がんになったかどうか、の集計結果を示す（架空データ）。

このままでは合計人数など集計結果がわからないので、喫煙の有無（1：有、0：無）、肺がんの有無（1：有、0：な

ID	Age (y)	BW (kg)	Gender (1/2)	Smoking (1/0)	Lung Cancer (1/0)
1	66.85	70.96	1	0	0
2	61.03	69.74	1	0	1
3	67.65	68.83	0	1	1
4	60.79	68.13	1	1	0
5	65.31	74.87	0	1	1
6	62.65	70.19	0	0	1
7	65.33	73.01	0	1	1
8	69.81	68.34	0	0	1
9	60.96	66.49	0	1	1

し)の人数を集計すると次のようになる。どちらも 2 値のカテゴリ変数である。

合計 100 人分のデータを想定して集計した結果を右表に示す。要因となり得る項目（喫煙）を縦に、結果となり得る項目（肺がん）を横に示して分割表を作成する（逆でも構わない）。この場合、要因のとり得る値が 2 通り、結果のとり得る値も 2 通りなので「2×2分割表」という。一般には m 個の要因と n 個の結果の場合には m×n 分割表が作成できる。

実測値	肺がん発症	肺がん非発症	合計
喫煙あり	42	9	51
喫煙なし	14	35	49
合計	56	44	100

#### ② カイニ乗検定

異なる要因の2群間に統計的な差があるかどうかを調べるひとつの方法としてカイニ乗検定がある。カイニ乗検定は統計ソフトウェアを用いると実行できるが、有意水準、および自由度という用語が出てくる。有意水準は他の検定法と同じく 5%、1%などを定める。自由度は、 $\text{自由度} = (\text{分割表の縦の項目数} - 1) \times (\text{分割表の横の項目数} - 1)$  で計算する。2×2 分割表の場合、自由度は  $(2-1) \times (2-1) = 1$  となる。

## 2) 相対危険度、絶対危険度、治療必要数

### ① 分割表に示される因果関係を正しく理解する

分割表を作成する上では、表が示す因果関係の内容をよく理解しておく必要がある。例えば、危険因子への暴露とその影響による罹患を評価する場合は次の分割表となり、原因が「危険因子への暴露」、結果が「罹患あり」なので「原因が『好ましくない』結果を及ぼすことを想定」した表になる。他にも、ある薬剤による処置とそれによる副作用発現との関係も同様となる。

	罹患あり	罹患なし	合計
危険因子への暴露あり	A	B	A + B
危険因子への暴露なし	C	D	C + D
合計	A + C	B + D	A + B + C + D

一方、薬剤あるいは治療による処置とそれによる効果との関係を見たい場合には分割表は次のようになり、「原因が『好ましい』結果を及ぼすことを想定」している。

	効果あり	効果なし	合計
処置あり	A	B	A + B
処置なし	C	D	C + D
合計	A + C	B + D	A + B + C + D

分割表を作成するコツとして、A のセルに、自分が注目する因果関係のペアを設定するように心がけるとよい。例えば次の表では処置あり／なしが逆に設定されていて、間違いではないが注目したい因果関係である「処置ありのときの効果ありの人数」が C の位置にくるので混乱する可能性がある。

	効果あり	効果なし	合計
★処置なし	A	B	A + B
★処置あり	C	D	C + D
合計	A + C	B + D	A + B + C + D

### ② 分割表から得られる指標

分割表の数値から、各種指標が得られる。

1. EER (Experimental Event Rate、介入群イベント発生率) =  $A / (A + B)$
2. CER (Control Event Rate、コントロール群イベント発生率) =  $C / (C + D)$
3. 群間での比率の差 (絶対危険度、Absolute Risk; AR)
 
$$= A / (A + B) - C / (C + D)$$

寄与危険度 (Attributable Risk) ともいう。介入群によるリスクが介入しない群と比べて絶対値がどれくらい大きいか (小さいか) ?
4. 絶対危険度 (リスク) 減少率 (Absolute Risk Reduction; ARR)
 
$$= C / (C + D) - A / (A + B)$$

5. 群間での比率の比 (相対危険度、Relative Risk; RR)

$$= A / (A + B) \div C / (C + D)$$

リスク比ともいう。介入群によるリスクが介入しない群と比べて相対的にどれくらい大きいか(小さいか)?

6. 相対危険度(リスク)減少率(Relative Risk Reduction; RRR) = 1 - RR

7. 治療必要数(Number Need to Treat; NNT) = 1/ARR。

一人の患者に「事象あり」と確認するために必要な例数(処置を治療として考える場合は Number Need to Treat、処置を危険因子として考える場合 Number Need to Harm (NNH)と使い分ける。

先の喫煙と肺がんのデータでは、

$EER=42/51=0.824$ 、 $CER=14/49=0.286$ 、 $AR=0.824-0.286=0.538$ 、 $ARR=0.286-0.824=-0.538$ 、 $RR=0.824/0.286=2.882$ 、 $RRR=1-2.882=-1.882$ 、よって  $NNH = 1 / |-0.538|=1.86$  となる。ARR や RRR が負の数となるが、介入群に該当する喫煙あり群でリスクの減少率が負、すなわちリスクが増大することを意味する。NNH は ARR の絶対値を用いて計算した。

特に RR (相対危険度) は、介入群とコントロール群の間のリスクの比率なので、 $RR > 1$  であれば介入群でリスクが大きい、 $RR < 1$  であれば介入群でリスクが小さいといえる。また  $RR = 1$  ならリスクに違いはないといえる。これを統計的に評価するためには信頼区間が 1 を挟むかどうかで判断する。

3) オッズ比

① オッズ比と信頼区間

2×2 分割表でオッズ比 (Odds Ratio, OR) は、 $OR=(A/B) \div (C/D)=(A \times D)/(B \times C)$  となり、各変数の相関の強さを表す指標となる。分割表の定義をどのように定めるかで A~D のセルの意味が変わるので、因果関係を確認したい項目(喫煙したら肺がんになる、薬剤を投与したら効果がある、など)が A の位置に来るように分割表を工夫するとよい。

OR は RR と同じく、群間で比率に違いがないなら 1 となるので、信頼区間から見て  $OR > 1$ 、 $OR = 1$ 、 $OR < 1$  のいずれであるかを判断し、信頼区間が 1 を含まないなら  $OR = 1$  と差がない、すなわち因果関係は無いと判断する。

先の喫煙と肺がんとの関係の場合、オッズ比は  $(42 \times 35) \div (9 \times 14) = 11.7$  となりその対数値は 2.46 となる。定義に従って 95% 信頼区間を求めると対数値で 1.51~3.41 で、OR の真数値では 4.51~30.2 となる。信頼区間が 1 を含まないことから、求めたオッズ比 11.7 は有意に 1 よりも大きい、すなわち喫煙と肺がん発症には有意な因果関係があると判断できる。

## ② オッズ比 (OR) を用いる時の注意

ケース・コントロール研究ではデータを遡って調査できるため、データを集めるという意味では比較的容易に試験を行えるが、一方で既存のデータを活用するので各種バイアスが生じる(あるいは、バイアスの原因となるデータが混じる)危険性も大きく、OR などの結果を決定的な判断に用いる際には十分に注意するべきである。また、交絡といわれる「他の要因にも結果にも両方に影響して複雑な関係性を生じる因子」への対応が必要となることもある。

## 4) 二値データとロジスティック回帰分析 (ロジットモデル)

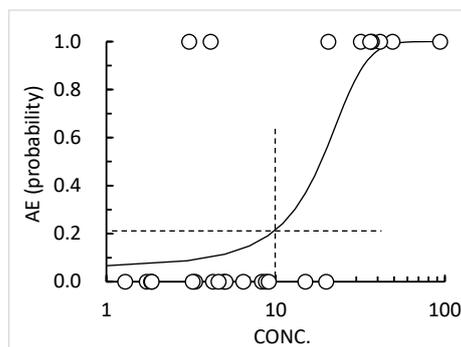
### ① ロジットとロジスティック解析

臨床研究においては「副作用あり(1)」「副作用なし(0)」といった結果がカテゴリ変数である二値データで得られることがある。一般には分割表を作成し、カイニ乗検定やオッズ比による検定を適用する。

二値データに回帰分析を適用する、即ち二値データを目的変数として用い、いくつかの説明変数の影響を回帰分析で説明する場合には、目的変数のデータをそのまま用いるのではなくロジットに変換してから用いる必要があり、このような回帰分析をロジスティック(重)回帰分析という。最尤推定法という方法で解析を行い、統計解析ソフトウェアを利用する。説明変数は連続変数であったりカテゴリ変数であったりする。

### ③ 実務領域でのロジスティック解析の使用例

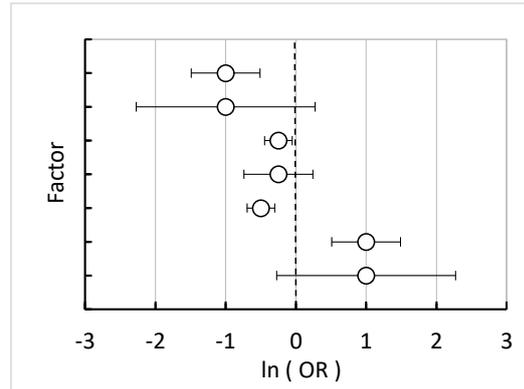
図は、ある薬物の血中トラフ濃度(次回投与前の最低血中濃度)と、ある副作用事象(adverse event、AE) 発症の有無との関係についてロジスティック解析を行った例である。副作用が発症した場合を 1、発声しなかった場合を 0 とし、トラフ濃度との関係をプロット○で示した。解析は専用のソフトウェアで行い、得られた回帰曲線を示した。



図のように回帰曲線は 0 あるいは 1 のみの値をとるのではなく、S 字型の滑らかな曲線になるのがこの解析の特徴である。図から、ある説明変数(血中トラフ濃度など)における目的変数(ある副作用発症確率)を読み取ることができる。例えば図から、この副作用の発症確率が 20% となる血中トラフ濃度を約 9.4 と読み取り、目標とする血中トラフ濃度設定に活用でき投与計画に有用となる。ここでは図の横軸は対数軸で示している。

### 5) フォレストプロット

各要因についてオッズ比とその信頼区間を図示したものをフォレストプロットという。重回帰分析のあと各要因について整理するときには便利で、要因ごとに同じ軸目盛に信頼区間を描くので、視覚的にも解りやすい。通常横軸はオッズ比の対数值 ( $\ln(OR)$ ) で表され、信頼区間を「バー」で表示する。



図に例を示した。点線は  $\ln(OR)=0$ 、すなわち  $OR=1$  の場合を示しており、信頼区間

がこの点線を挟む(またぐ)場合には有意な要因ではないと結論付ける。このように  $OR$  の有意性をいう場合には、 $OR$  そのものの値 (○印) の位置だけで判断するのではなく、信頼区間が  $OR=1$  ( $\ln(OR)=0$ ) とどのような関係があるかに着目する。

回帰分析で得られる回帰式から、その要因に関する  $OR$  が計算できる。回帰分析等で得られた  $OR$  の値を表に整理する場合は、回帰係数をそのまま示してもよいが、 $OR$  に換算した値を示すと実務的にはわかりやすくなる。

No.	OR	Lower	Upper
1	1.44	0.368	19.3
2	1.44	0.368	52.5
3	2.18	0.779	32.8
4	2.18	0.779	89.1
5	1.83	0.607	395
6	15.2	2.72	56.8
7	15.2	2.72	154

右の例では、各要因について  $OR$  と 95%信頼区間を示しており、この表と上の図とを一緒に論文等で示すことで読者にわかりやすい図表となる。

#### ◇ 練習問題

設問1) 分割表から得られる指標について整理してみよう。

設問2) ロジスティック回帰分析を用いるべき条件を整理してみよう。

## 7. 生存時間解析

### 1) 「時間-イベント型」データ

#### ① 定義

抗がん剤治療を受けた患者がその後どれだけの間生存できたか、といった「ある時点から開始して注目するイベントが発生するまでの時間」を扱う解析を生存時間解析といい、扱うデータを生存時間データという。

必ずしも「生存」という指標だけでなく、がんの再発や関心のある副作用の発現、ある病気から回復するまでの時間、といった注目するイベントが発生するまでのデータにも適用でき、必ずしもネガティブな（望ましくない）数値だけを扱うわけではない。いつ（の時点に）イベントが起こったか、というデータを解析するので時間-イベント型データともいわれる。

時間-イベント型データには二種類の打ち切りが存在する。被験者が臨床研究に参加して以降イベントが発生したらその時間を記録するが、打ち切りとしては、

1. 試験の途中で被験者が同意を撤回した、あるいは退院・転院などで試験を継続して追跡することができなくなった場合（途中で中止）
2. 臨床試験にはあらかじめ試験期間が定められるが、その試験期間が終了するまでにイベントが発生しない場合（最後まで継続）

がある。1 の場合にはイベントが発生したデータは得られないが、「試験開始後〇日まではイベントは発生しなかった」という情報は得られる。2 の場合、「試験終了までイベントは発生しなかった」という情報は得られる。

このように生存時間解析では、すべての被験者での生存時間データを得ることは実際的に不可能なことが多く、検査値データなどとは異なる特徴を持つので統計解析も特別な方法で行う必要がある。

#### ② 生存時間の分布（生存曲線）

生存時間の分布をグラフ化する代表的な方法としてカプラン・マイヤー法（プロット）がある。作図の前にまずデータを整理する。表は、患者から得た臨床研究参加日の年齢、イベントの発生の有無（0:打ち切りデータ、1:イベント発生日が確認されたデータ）、イベント発生日、などの値を示したものである（架空データ）。右側 3 列は途中で切れているが、この右にも患者から同時に得られた症状スコア値が並んでいる。このように EXCEL 等で、打ち切りの有無で 0、1 というふうに数値で区別し（0、1 を逆に定義しても構わない）、被験者ひとり分のデータを一行に整理するとわかりやすく、その後の解析にも便利になる。

カプラン・マイヤー法では打ち切りによりその後のデータが欠損する場合も考慮しており、試験途中で打ち切り時点で全体の生存率には変化はないが、症例数は減少することを反映している。

0		Survival data					
Patient No.	Age (y)	Event (Censored: 0, Death: 1)	Time to Event (days)	Cluster2 0:absence 1:presen	Q1: short walk	Q2: in bed	Q3: n hel
	62	1	42	1	2	4	2
	57	1	6	1	4	4	4
	81	1	18	1	4	4	4
	96	0	161	0	4	3	4
	73	1	11	1	2	4	3
	31	1	4	1	4	4	4
	72	1	56	1	2	1	4
0	64	1	25	1	2	4	3
1	81	1	14	1	4	4	4
2	75	1	20	1	4	4	4
3	83	1	24	1	4	4	4

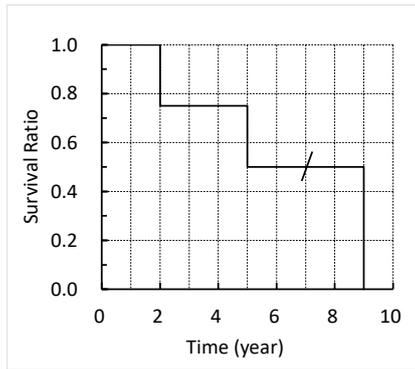
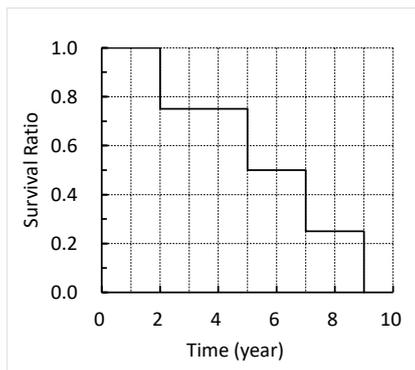
簡単な例として次のデータの Kaplan-Meier プットを作成してみる。

- i) 4 例の生存時間が、2、5、7、9 (途中打ち切りなし)
- ii) 4 例の生存時間が、2、5、7\*、9 (\*は途中打ち切り例)

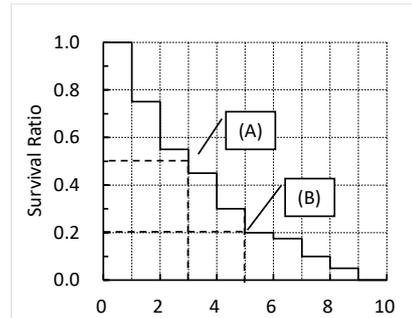
図のように、生存曲線では途中打ち切りが起こったことを示すためその時点に短い縦線を入れることがある。イベントが発生するたびに縦軸が減るので滑らかな曲線ではなく階段状のグラフになる。試験終了までイベントが発生しなかった場合には縦軸が 0 まで下がらない。「/」は打ち切り例を示す。

(発展) 計算方法を示す。i) では、2 年目まではすべて生存なので生存率は 1。2 年目に 1 例のイベントが発生するので  $1 - (1/4) = 3/4 (=0.75)$  となる。続いて 5 年目までは 0.75 が続き 5 年目で  $3/4 \times (1 - 1/3) = 1/2 (=0.5)$ 。7 年目までは 0.5 が続き、7 年目で  $1/2 \times (1 - 1/2) = 1/4 (=0.25)$ 、さらに 9 年目に  $1/4 \times (1 - 1/1) = 0$  となる。

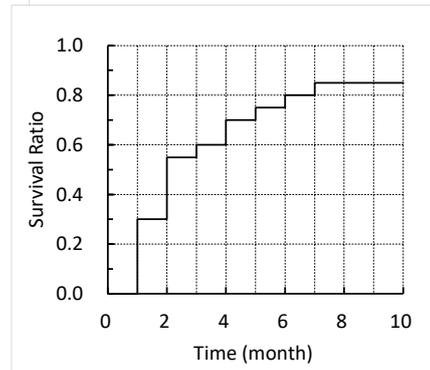
ii) 7 年目までは i) と同じ。7 年目に打ち切りがあるので生存率は変わらないので 9 年目までは 0.5 であるが、被験者数が減ることに注意する。9 年目で  $1/2 \times (1 - 1/1) = 0$  と分母が 2 ではなく 1 となる。



生存曲線から例えば 50%生存時間(中央値)や 5 年生存率などの指標を読み取ることができる。右図で A は、生存率が 0.5 となる生存曲線上の値を見つけることで 50%生存時間が 3 か月と読み取れる。また、B のように 5 か月生存率を読み取ると 0.2 となる。実際には被験者データ数が多くより細かな刻みで生存曲線が描かれ、指標はより正確に得られる。



生存曲線は右下がりとは限らない。例えば縦軸を副作用発生率とするとグラフは累積の形となり右上がりになる。どのようなイベントのデータを扱うか、結果をどのように伝えたいかによって形状を決めたらよい。右図の例の場合「副作用が発症したら人数(割合)が減る」という右下がりの図よりも、副作用が発症したらその累積人数を示す右上がりの図のほうが直観的にも理解しやすいと判断した。



関心のあるイベントがどの時間でも一定の割合で起こるとは限らない。病気による死亡や病気の再発、副作用の発症が試験開始後初期の段階で多く起こる場合には、生存曲線は初期に下がり方(上がり方)が急となることもある。治療から遅く出現するような副作用であれば初期の発生率の変化は少ないであろう。多くの生存時間データの解析では、生存率の減少が一定と仮定しているが、必ずしも成立しないことがあるので生存曲線を確認して、イベントの発生が時間によってどのように変化するかも確認する。

## 2) 2つの生存曲線の比較(ログ・ランク検定)

異なる条件(新規抗がん剤+従来抗がん剤を投与した群と、従来抗がん剤のみを投与した群、など)で得られた生存曲線を比較する方法としてログ・ランク検定が代表的である。生存曲線をカプラン・マイヤー法で描き、ログ・ランク検定による有意確率 p 値を示す、という手順になる。

### 3) コックス比例ハザード回帰

#### ① ハザードとハザード比

ある時間幅でのイベントの発生率をハザード (hazard) とし、調査期間中に時間によってハザードが変化することもある。2 群のハザードそれぞれの比をハザード比 (hazard ratio, HR) といい、「それぞれの群の危険な度合いの比率」である。対照群に対する処置群のハザード比が「1 よりも小さい」ほうが「好ましい=薬剤の効果あり」と判断できる。多くの生存時間データの解析では、このハザード比が時間によって一定で変わらないという比例ハザード性を仮定するが、この仮定が成立しないことも多い。

#### ② 生存時間のモデルによる解析

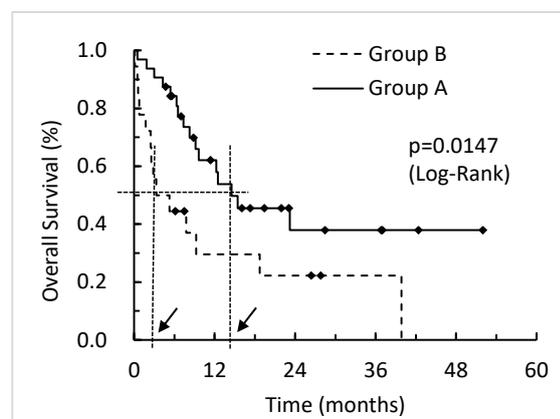
生存時間データに回帰分析を適用し要因解析を行う方法としてコックス比例ハザード回帰がある。比例ハザード性を仮定し、目的変数として「ハザード関数」といわれる生存時間データに仮定する関数を用い、各説明変数との関係を調べる。複数の要因について影響を検討する場合には重(多変量)回帰分析となり、ロジスティック回帰と同じような式になるが、目的変数がハザード関数となる部分が異なる。

コックス比例ハザード回帰で得られる回帰係数から各要因についてのハザード比とその信頼区間が得られ、オッズ比(OR)と同様に信頼区間を用いた検定が行える。

### 4) 実務領域における生存時間データの解析例

#### ① 一般的生存曲線図の見方

抗がん剤の治験成績報告などが論文や添付文書に示されるが、このとき生存曲線を見て結果を読み取る習慣をつけると便利である。図は2群A、Bそれぞれの生存曲線を示している、打ち切りデータを小さな◆で示している。図にはログ・ランク検定による検定結果も同時に示されているので「有意に差がある」ことがこの図からだけでも読み取れる。さらに自分で点線を引けば生存曲線を活用して50%生存率(生存時間中央値と同じ意味)のおおよその値も読み取れる(矢印)。



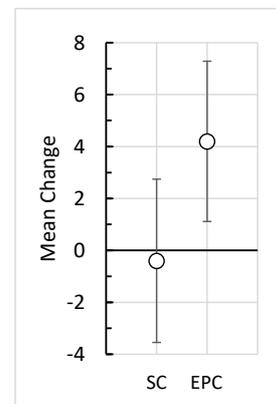
その他にも、図中に中央値の95%信頼区間や、各群の患者総数やイベント発生数などを表にして挿入している場合など、図にはいろいろな情報が掲載されていることが多い。

## ② 緩和医療における生存曲線の解釈例

緩和ケアの領域の論文で、「終末期がん患者において早期からの緩和ケアを施すことで生存時間が延長した」という報告がある (JS Temel, et al., *New Eng J Med*, 363(8), 733-742 (2010).)。試験デザインについて整理してみる (詳細は原著を参照)。

- ✓ 試験デザイン ~ ランダム化 2 群比較試験。処置群: 通常のがんケアに加えて早期緩和ケアも受けた患者群、対照群: 通常のがんケアのみを受けた患者群。
- ✓ 評価項目 ~ FACT-L スケールによる各種 QOL 評価項目。主要評価項目は 12 週後の QOL スコアの変化量。
- ✓ 結果 ~ FACT-L による 12 週後の QOL スコアが表に示されており、136点満点で、緩和ケア群で 98.0、対照群で 91.5 となり、両側検定で  $p=0.03$  と有意な違いがみられた。QOL スコアの変化量については表ではなくグラフで示されており、グラフの説明を見ると、緩和ケア群で 4.2 の増加、対照群で 0.4 の減少がみられこちらも有意な差であった ( $p=0.09$ )。変化量のグラフを示した。
- ✓ 生存曲線 ~ 論文にはさらに生存曲線が比較されており (著作権保護のためここには示していない)、図の説明文を読むと、生存期間中央値はログ・ランク検定の結果、緩和ケア群で 11.6 か月、対照群で 8.9 か月と有意な差があったとしている。

こういった論文を読む場合には、図表やその説明文をしっかり読むことで必要な情報がすぐにわかることが多い。英語論文を読む場合でも、論文全体をじっくり読むことが最もよいことではあるが、限られた時間で内容を把握するコツとして、要旨を理解し (要旨だけなら PubMed だけでも閲覧可能である)、次に図表とその説明文をチェックする、さらに興味があれば本文をしっかり読むといった手順もよいかもしれない。



### ◇ 練習問題

設問1) 生存時間データの特徴と作図、解析手法を整理してみよう。

設問2) 身近なデータで生存時間解析が適しているデータをあげてみよう。

## 8. 症状クラスター特定のための主成分分析とクラスター分析

### 1) 主成分分析でデータを理解しやすくする

#### ① 症状クラスター

第 2 章で、苦痛症状に関する患者報告型アウトカム (PRO) が同じ患者で複数同時に発症する「症状クラスター」の評価が重要であることを実務的観点から示した。この論文では EORTC QLQ-C15-PAL (以下 PAL15) で得られた PRO による各項目の評価を主成分分析で評価し症状クラスターを特定している。

原著論文の表 2 では、PAL15 のうち QOL を除く 14 項目についてスコア (1 から 4) の割合、平均値が示されており、これらの数値をもとに症状クラスターを評価した結果が表 3 に整理されている。

#### ② データの相関性

症状クラスターの特には、統計的には主成分分析、因子分析、クラスター分析といった手法が適している。EXCEL 統計 (Bell Curve、<https://bellcurve.jp/ex/sample.html>) では機能に制限はあるが、無料で解析を試すことができる。主成分分析は症状クラスターの特定に限らず広く用いられている手法で「主成分分析とは、複数変数のばらつき傾向から少数の合成された新たな変数 (主成分) を求める手法」とある。PAL15 のように項目数が多い場合にはデータを「集約しなおす」と全体の傾向を理解する上で便利になる。

#### ③ 相関行列

項目ごとの相関係数を一覧表にしたものを相関行列という。多くのデータ項目がある場合でも、それぞれの相関係数を相関行列として表示する。実際に論文で用いたデータの相関行列を示す。

Q1 から Q14 はそれぞれの項目で次の対応がある。カッコ内はよく用いられる略号を示しており、PA、FA は 2 つの項目で共通に示している。

**Q1:** Do you have any trouble taking a short walk outside of the house? (short walk)、

**Q2:** Do you need to stay in bed or a chair during the day? (in bed)、**Q3:** Do you need help with eating, dressing, washing yourself or using the toilet? (need help)、

**Q4:** Were you short of breath? (short of breath; dyspnea, DY)、**Q5:** Have you had pain? (pain, PA)、**Q6:** Have you had trouble sleeping? (SL)、**Q7:** Have you felt weak? (fatigue, FA)、

**Q8:** Have you lacked appetite? (AP)、**Q9:** Have you felt nauseated? (nausea and vomiting, NV)、**Q10:** Have you been constipated? (CO)、**Q11:** Were you tired? (fatigue, FA)、

**Q12:** Did pain interfere with your daily activities? (PA)、

**Q13:** Did you feel tense? 、 **Q14:** Did you feel depressed?

**(Q15:** How would you rate your overall quality of life during the past week? (QOL))

論文で用いた PAL15 データの相関行列

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
Q1	1.000	0.590	0.628	0.069	0.073	-0.004	0.197	0.165	-0.127	-0.040	0.076	0.088	0.107	0.119
Q2	0.590	1.000	0.593	0.032	0.057	0.065	0.347	0.267	-0.157	0.117	0.242	0.050	0.002	0.118
Q3	0.628	0.593	1.000	0.079	0.070	0.089	0.310	0.160	-0.010	-0.006	0.251	0.144	0.136	0.185
Q4	0.069	0.032	0.079	1.000	0.116	0.283	0.332	0.273	0.256	0.023	0.329	0.106	0.160	0.243
Q5	0.073	0.057	0.070	0.116	1.000	0.421	0.166	0.178	0.277	0.178	0.290	0.752	0.372	0.305
Q6	-0.004	0.065	0.089	0.283	0.421	1.000	0.251	0.184	0.371	0.103	0.360	0.450	0.291	0.295
Q7	0.197	0.347	0.310	0.332	0.166	0.251	1.000	0.416	0.145	0.196	0.528	0.235	0.065	0.320
Q8	0.165	0.267	0.160	0.273	0.178	0.184	0.416	1.000	0.381	0.181	0.410	0.257	0.142	0.314
Q9	-0.127	-0.157	-0.010	0.256	0.277	0.371	0.145	0.381	1.000	0.122	0.252	0.187	0.268	0.198
Q10	-0.040	0.117	-0.006	0.023	0.178	0.103	0.196	0.181	0.122	1.000	0.203	0.225	0.120	0.138
Q11	0.076	0.242	0.251	0.329	0.290	0.360	0.528	0.410	0.252	0.203	1.000	0.439	0.249	0.499
Q12	0.088	0.050	0.144	0.106	0.752	0.450	0.235	0.257	0.187	0.225	0.439	1.000	0.408	0.399
Q13	0.107	0.002	0.136	0.160	0.372	0.291	0.065	0.142	0.268	0.120	0.249	0.408	1.000	0.458
Q14	0.119	0.118	0.185	0.243	0.305	0.295	0.320	0.314	0.198	0.138	0.499	0.399	0.458	1.000

相関係数( $r$ )は  $-1 \leq r \leq 1$  の範囲をとり、数値の正負は正あるいは負の相関を表し、絶対値が1に近いほど相関性は強い。相関係数にもとづく相関の強さの目安として、  
 ~ -0.2: 相関なし 0.2~0.4: 弱い相関 0.4~0.7: 中程度の相関 0.7~1.0 強い相関

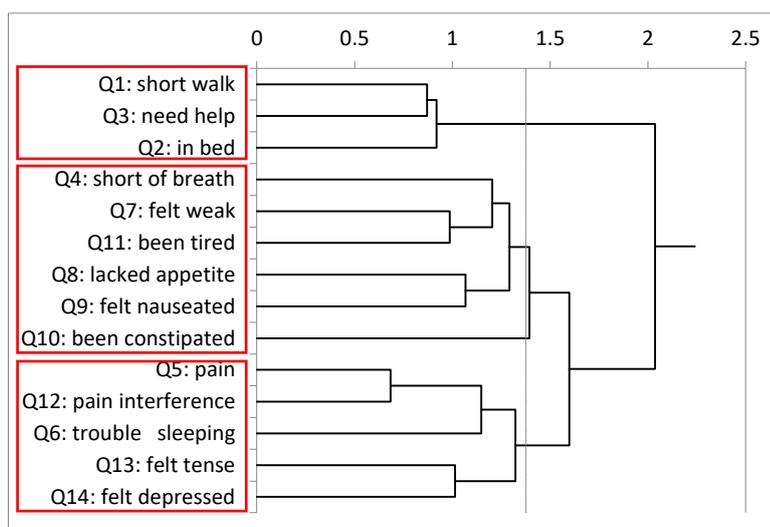
があるが、相関性について統計的な有意性を見る場合には、係数の有意確率( $p$  値)で評価すべきで、さらに臨床的有意性についても考える。

先の相関行列では、この分類に従って相関係数の大きさを色分けした。このように数値の大小で色分けした図をヒートマップという。PALI5 の場合、例えば、Q5 と Q12 はいずれも痛みに関する項目であり、相関係数が0.758と高く「傾向が似ている」と予想できる。Q7 とQ11との相関係数は0.534 でこちらも相関性が高そうである。このように相関係数が比較的高い項目どうしは同じクラスターに含まれるのではないかと予想できる。

主成分分析や因子分析といった手法は医療薬学領域では十分に広まっていない手法かもしれない。「因果関係」を明らかにし、患者の症状評価の「ウラ」にある要因(潜在因子などともよばれる)を知ることは医療現場で有用であり、患者志向の臨床研究では今後より必要となる解析手法であろう。主成分分析の詳細は触れないが、数学的には行列計算であり、固有値、寄与率、主成分負荷量などを駆使してクラスターを評価する。

## 2) クラスター分析

クラスター分析という手法もある。クラスター分析は主成分分析などと同じく「得られたデータをいくつかのグループに分類する」ことを目的とする方法で、「似た者同士をまとめる」イメージである。クラスター分析の結果はデンドログラムという図で得られ、視覚的にも症状クラスターの構成要素



がわかりやすい。今回のクラスターを四角で囲んでみた。計算方法が異なるので同じデータでも主成分分析の結果と異なる場合もある。

主成分分析や因子分析といった手法は、行動心理学の分野など医療薬学以外にも広く用いられている。統計学的にも古くからある手法で近年は計算機科学の発展にともない誰もが便利に解析を実行できるようになっている。

一方で、こういった手法はあくまで「全体的な傾向」を見出すだけの場合が多い。今回は「患者全体としてはこのように分類することで 9 個の症状項目が 3 つのクラスターでわかりやすく説明できる」可能性を示したに過ぎない。現実にはすべての患者でこのように分類できるとは限らず、今回特定した症状クラスター以外の複数の症状を発症する患者もある。患者と日々直面する医療従事者の現場感覚と統計解析結果がうまく融合し、新たな患者ケアの方法を考えることが患者志向の臨床研究の醍醐味だと思う。

◇ **練習問題**

設問1) 症状クラスターを評価することの臨床的な意義を考えてみよう。

設問2) クラスター分析が医薬領域以外で用いられている例を調べてみよう。