

## 第3回 データ解析は素材に合った料理法選びから

横浜薬科大学医療統計学

奥田千恵子 Chieko OKUDA

### 医療データは初学者泣かせ

統計解析手法は様々な素材《データ》を料理するための道具やレシピにあたる。扱いやすい素材なら、手順通りに料理すれば十分おいしい一皿《解析結果》ができ上がるが、変った材料や、やや難点のあるものを扱うには、下ごしらえ《データ変換や外れ値、欠測値の扱いなど》を工夫したり、特殊な道具《高度な統計手法》を使ったり、でき上がった料理をお客に出す時にも細かい心遣い《解析結果の解釈、結論の導き方など》が必要である。

生理学的測定や生化学的検査、運動能力測定、評価スケールを用いた感覚や意識、知能の検査、日常生活活動や、生活の質(QOL)の評価……。高度な機器で測った検査値であれ、アンケート調査の○×式回答であれ、何でも研究データとなり得るが初学者向きの食材ばかりではない。

第1回で述べたように、研究対象となる「特定の疾患をもつ患者」には、推測統計の基礎となる母集団や標本概念を当てはめにくい。加えて、臨床データは属性や精度が多種多様であり、1人の患者から経時的に複数回採られることが多く、測定時点が不ぞろいになりやすく、欠測値が出やすいなどの特徴をもつ。

料理法を誤れば誰も食べてくれず棚の隅で朽ち果てる。

### t-検定

基礎統計学の教科書には必ず登場するためか、t-検定(t test)は医療研究者に多用される手法である。2群間で、連続量データの平均値の比較に用いる特定の検定法を指して、単に「t-検定」と呼ぶ人が多いが、t-検定とはt分布を利用する検定法の総称である。同様にF検定(F test)や $\chi^2$ 検定( $\chi^2$  test)もF分布や $\chi^2$ 分布を利用する検定法の総称である。

また、t-検定はパラメトリック検定法(母集団が正規分布に従うと仮定する検定法の総称)の1つである。ただし、ある程度データ数が多ければ、正規分布からの逸脱には頑健である(検定結果にあまり影響がない)とさ

れているのでそれほど厳密な正規性は要求されない。パラメトリック検定法では、正規性に加えて、各群のばらつきが等しく(等分散性)、データが別々の個体から採られていること(独立性)が前提条件となっている。この条件が満たされないデータにはそれぞれ異なったt-検定を用いる必要がある(表1)。

### 多重比較法

医療分野で、単純に2つの平均値を比較したいという状況は意外に少ない。コンピューターがなかった時代は、3群以上の平均値を比較したい場合、特定の群と群の対をつくってt-検定を繰り返すということが黙認されていた。しかし、検定を繰り返すことによって、5%だったはずの $\alpha$ 過誤率が研究全体ではもっと高くなり、真の差がないにもかかわらず偶然2群ずつの組み合わせのなかのどれかに有意差が出てしまう確率が増えてしまう。これを「多重性の問題」と呼び、統計ソフトでは様々な対比較デザインに対応できる多重比較法が用意されている(表2)。

多重比較法の多くはt-分布を利用するのでt-検定の一種である。慣習的に分散分析(analysis of variance:以下、ANOVA)の後に用いられてきたのでポストホックテスト(post hoc test)と呼ばれることがあるが、ほとんどの多重比較法は単独で用いることができる<sup>1)</sup>。しかし、一般的な統計ソフトではANOVAの後に対比較を行う設定になっていることが多い。

表1 t-検定：2群間での平均値の比較法

検定法名	各群のデータの前提条件		
	正規性	等分散性	独立性
対応のないt-検定 (unpaired t test)	○	○	○
ウェルチのt-検定 (Welch's t test)	○	×	○
対応のあるt-検定 (paired t test)*	—	—	×

\*：対応のある(同一個体から採られた)データの差が正規分布に従っていることが、本法を用いる前提条件。各群の正規性や等分散性は要求されない。

## ノンパラメトリック検定法

順序カテゴリデータは数値化しても等間隔とはならないので連続量データと同じ扱いはできない（例：健康状態 非常に悪い=1, 悪い=2, 普通=3, 良い=4, 非常に良い=5）。当然, このようなデータは正規分布には従わないため, データの分布型の仮定を必要としないノンパラメトリック検定法が用いる必要がある。本法は正規分布に従わない連続量データにも用いることができる。

群間比較に用いる主なパラメトリック検定法には, それぞれに相当するノンパラメトリック検定法がある(表3)。

表2 多重比較法：3群以上の間での平均値の対比較

検定法名	対比較の条件
ボンフェローニ補正法 (Bonferroni correction)	最も適用範囲が広い手法
チューキー検定 (Tukey test)	すべての群間の対比較
ダネット検定 (Dunnett test)	対照群との対比較
ウィリアムズ検定 (Williams test)	単調に増加 (減少) する場合の対比較

表3 群間比較に用いる主なパラメトリック検定法に相当するノンパラメトリック検定法

検定法名		群間比較のデザイン
パラメトリック検定法	ノンパラメトリック検定法	
対応のないt検定	マン・ホイットニー検定 (Mann-Whitney test)	対応のない2群間
対応のあるt検定	ウィルコクソン符号付き順位検定 (Wilcoxon's signed rank test)	対応のある2群間
(要因) ANOVA	クラスカル・ウォリス検定 (Kruskal-Wallis test)	対応のない3群以上 (系全体として検定)
多重比較法 (表2)	対比較の条件に合わせて, スティール・ドウラス検定 (Steel-Dwass test) やスティール検定 (Steel test) など	対応のない3群以上 (対比較)
反復測定分散分析	フリードマン検定 (Friedman test)	対応のある3群以上 (系全体として検定)

表4  $\chi^2$ 検定

検定法名	群間の順序の有無	各群のデータの独立性
独立性の $\chi^2$ 検定 ( $\chi^2$ test for independence)*	順序がない (例：試験薬A群, B群, C群, …)	独立データ
傾向性の $\chi^2$ 検定 ( $\chi^2$ test for trend)	順序がある (例：低用量群, 中用量群, 高用量群)	独立データ
マクニマ検定 (McNemar test)	順序のない2群間	対応のあるデータ

\* : 比較的少数データを比較すると $\chi^2$ 分布への近似が悪くなるのでイエーツの補正 (Yates' correction) が必要。

## カテゴリの出現度数の比較法

2値データ (例：男性=1, 女性=2) や, 順序のないカテゴリデータ (例：A型=1, B型=2, AB型=3, O型=4) は, 数字で表されていてもデータの値や順序に意味がない。従って, 群間で比較するには, 連続量データや順序カテゴリデータのようにデータの値そのものを比較するのではなく, 各カテゴリの出現度数 (人数, 個数など) を比較する。

群間の順序の有無や, 各群のデータが別々の個体から採られているか否かで, 異なった手法を選択しなければならない (表4)。

## 2群間での2値データの比較

2群間 (例：試験薬群, 対照群) での2値データ (例：有効, 無効) のカテゴリ出現度数の比較に限れば手法の選択肢はさらに増える。

フィッシャー直接確率法 (Fisher's exact test) は, 第2回で述べたように, 確率分布を利用せず直接p値を計算する手法であり, 少数データでも正確なp値が求まる。3群間以上の検定も可能であるが, 現在のところ2群間に限定されているがソフトが多い。

臨床研究では, データの出現度数を有効率や生存率などの比率で表して区間推定を行うことが多い。2群のリスク比 (risk ratio), あるいは, オッズ比 (odds ratio) を計算して, それらの95%信頼区間 (以下, 95% CI) を求める。例えば, 「試験薬群は対照群と比較して有効率が高い (リスク比: 0.500, 95% CI: 0.258~0.970)」という結論の仕方をする。有意水準5%で有効, 無効の出現度数

に差があるなら、リスク比やオッズ比の95% CIに1が含まれていない。

## 相関分析と回帰分析

相関分析 (correlation analysis) はしばしば回帰分析 (regression analysis) と混同される。相関分析とは1個体から得た2変数間の直線的な関係の有無を調べる手法である。回帰分析とは、1組の(複数の) データを、何らかのできるだけ単純な回帰モデル(関数) に当てはめる手法であり、回帰式は必ずしも直線とは限らない。

2変数間で直線に回帰した場合、モデルの適合度を表す決定係数 ( $R^2$ ) は相関係数 (correlation coefficient,  $r$ ) を2乗した値と同じになるため、相関係数を求めることのみを目的として回帰分析が行われている例をよくみかける。

相関係数も複数ある。単に相関係数 ( $r$ ) という場合は、2変数がともに正規分布に従う連続量データの直線関係を示す指標であるピアソン (Pearson) の積率相関係数を指す。少なくとも一方の変数が、正規分布に従わない連続量データ、あるいは、順位カテゴリデータの場合にはスピアマン (Spearman) の相関係数やケンドール (Kendall) の相関係数などのノンパラメトリックな相関係数を用いる。

## 交絡因子

相関分析や回帰分析を行う際にどうしても避けて通れないのが交絡因子 (confounding factor) の存在である。注目する2つの変数だけを取り上げて解析すると、交絡因子の影響を受けて、みせかけの因果関係が生じたり、逆に、あるはずの関係が検出できなくなったりする。

原因と思われる因子と、結果と思われる因子の、両方とかがわりのある因子が交絡因子となる。例えば、「年収が上がると血圧が上がる」という研究仮説を立てたとしよう。原因が年収、結果が血圧、と単純に2つの変数だけで回帰分析を行えば、かなり適合度の良い回帰式が得られるかもしれない。年収を上げるにはストレスが増えることを我慢して働かなければならないので血圧が上がるのだらうと、もっともらしい理由

も付けられるかもしれない。

しかし、これはみせかけの因果関係である。もし、第3の変数として対象の年齢のデータを得て、年収と年齢の関係、および血圧と年齢の関係を調べたとしたら、年齢が高くなるにつれて年収も血圧も高くなる傾向があることがわかるであろう。つまり、年齢は原因(年収)と結果(血圧)の両方とかがわりのある交絡因子なのである。

医療分野では、ごく限られた状況でしか2変数解析 (bivariate analysis) を行うことができない。血圧と年収の関係を調べるには、相関分析では、それぞれの値から年齢の影響を取り除いた後の相関関係を表す偏相関係数 (partial correlation coefficient) を求める。回帰分析では多変数解析 (multivariate analysis) を行う必要がある。

## 多変数解析とは

因果関係を求める医療研究では、一般に、対象の性別や年齢、基礎疾患や薬剤の使用の有無などの背景因子 (background factor) は、研究目的とする変数と多かれ少なかれ何らかのかかわりをもっており交絡因子となり得る。1つの従属変数 (dependent variable) に対して、複数の独立変数 (independent variable) を同時に扱うことができる手法(表5)では、交絡因子となり得る変数を独立変数として多変数解析に含めることで、その影響を除くことができる。

多変数解析には多くの種類があり、「多変数」の定義は必ずしも一致していない。因果関係を求めることを目的とする手法以外に、従属変数が2つ以上ある場合に用

表5 因果関係を求める多変数解析

検定法名	従属変数 (結果)	独立変数 (原因)
(多変数) 線形回帰 (linear regression analysis)	連続量変数	属性は何でもよい
数量化I類*	連続量変数	カテゴリ変数
2項ロジスティック回帰 (binary logistic regression analysis)	2値変数	属性は何でもよい
多項ロジスティック回帰 (multinomial logistic regression analysis)	3つ以上の順序のない カテゴリ変数	属性は何でもよい
順序ロジスティック回帰 (ordered logistic regression analysis)	順序のある カテゴリ変数	属性は何でもよい
判別分析 (discriminant analysis)	カテゴリ変数	属性は何でもよい
数量化II類*	カテゴリ変数	カテゴリ変数

\* : 1940年代後半に日本で開発された数量化理論 (Hayashi's quantification methods) を用いた解析法。I~VI類までであるが、異なる名称で呼ばれている標準的な手法と本質的に同一と考えられるものもある。

いる多変量分散分析 (MANOVA), 因果関係を求めず, 変数同士の関連性を調べて幾つかのグループに分類する手法である主成分分析 (principle component analysis), 因子分析 (factor analysis), クラスタ分析 (cluster analysis), 数量化Ⅲ類, 数量化Ⅳ類なども多変量解析と呼ばれている<sup>2)</sup>。

## 生存時間分析

それほど発生頻度が多くない致死的な疾患に対する治療の効果を評価する研究では, 研究期間が数年間に及ぶ場合があり, 途中で追跡ができなくなって生死不明になる対象が多くなる。また, 研究開始後, 順次対象が登録されるので, 終了時点近くに登録された対象は短期間の観察後に研究が終了してしまう。

このような研究では, 生死不明の対象, および研究の終了時点で生存している対象を打ち切り症例と呼び, 観察の打ち切り/非打ち切りという2値変数を用いて生存期間の解析を行う。

特定の時点での生存率 (がん患者の5年生存率など) を求めたり, 治療法の異なる複数の群の患者集団の生存率を比較したりすることができる。そのなかでも Kaplan-Meier 法 (Kaplan-Meier method) は少数例であっても正確に生存率が計算されるのでよく利用される。コックス比例ハザード回帰分析 (Cox proportional hazard regression analysis) は生存時間を扱う多変量回帰分析である。

## 高度な統計手法

臨床データを利用する研究では, それぞれの解析法を適用する際に求められる様々な前提条件から逸脱したデータを扱わなければならないことがある。そのような場合は様々なモデルに当てはめてもっとも当てはまりのよいモデルを選ぶ必要がある。t-検定やANOVA, 反復測定分散分析, 共分散分析, 線形回帰分析などのパラメトリック検定を体系的, 包括的に扱う一般線形モデル (general linear model: 以下, GLM) がよく利用される。

GLMは非常に拡張性が高い統計手法であるが従属変数は連続変数でなければならない。生存/死亡, 疾患の発症/非発症といったカテゴリ変数で表されるアウトカムを扱う研究ではGLMは使えない。そのような場合は, 対数線形モデル (log-linear model), あるいは, 一般化線形モデル (generalized linear model: GLIM) と呼ばれる体系を用いる。また, 測定時点が不ぞろいで, 欠測

データが多い経時的測定データには一般化推定方程式 (generalized estimating equation: GEE) を用いる<sup>3)</sup>。

## 見た目は良くても

医療研究では上述のような高度な解析手法が必要となるケースが少なからずあるが, いずれの解析法も理論が難解で解析結果には複雑な表現が用いられている。すでに論文として形を成したものであれば解析結果を理解できる程度の統計リテラシーをもつ研究者であっても, 自ら使いこなすのは容易ではない。見様見真似でいきなり懐石料理に挑戦すると, 見た目は良くても味が変わったり, ちょっと揺すっただけで崩れてしまう代物ができてしまうかもしれない。

今回は, 手にした材料をじっくりと眺めて特徴をつかみ, 素朴であっても素材の味を生かした逸品に仕上げるにはどうしたらよいかを考える。

## 統計手法の報告の仕方 To Do & Not To Do

必ず手法の区別ができる名称を用いる (例: t-検定や  $\chi^2$  検定ではなく, 対応のないt-検定や独立性の  $\chi^2$  検定)。

統計ソフトの出力のうち, 研究仮説を証明するのに必要かつ十分な情報を報告する。逆に, 不必要な情報はカットする。例えば, 相関係数のp値はデータ数 (n) が多ければ多いほど, 低い相関係数 (弱い相関関係) であっても相関は有意であると判定される。従って, p値には必ずnを添えて報告する。また, 相関係数 (例: 検査値Tと年齢の関係) を求めることを目的として回帰分析を行った場合は, 今後の研究や臨床に利用する価値のない回帰直線 (例: 検査値T = a + b × 年齢) は報告すべきではない。散布図上にも回帰直線を書き込まない。

多変量解析などの複雑な検定を行う場合は, 統計ソフトによって算出方法や統計用語が異なることがあるため, 利用した統計ソフト名 (バージョンも) を書く。できるだけ標準的な統計の教科書に載っている用語を用いて解析結果を報告する。

### 引用文献

- 1) 永田 靖, 吉田道弘: "統計的多重比較法の基礎", サイエンス出版社, 東京, 1997.
- 2) 内田 治: "すぐわかるSPSSによるアンケートの多変量解析 第3版", 東京図書, 東京, 2011.
- 3) A. Grafen, R. Hails: "一般線形モデルによる生物科学のための現代統計学", 野間口謙太郎, 野間口真太郎訳, 共立出版, 東京, 2007.