

質疑応答

臨床医学における統計解析の方法を教えてください?

解答：小川 龍*

質問は「医学論文における統計解析の方法」である。現在パーソナルコンピュータのパッケージプログラムを用いてデータベースを作れば、いかなる統計処理も一瞬で完了する。しかし統計処理

する前のデータの性質を見極め、それに適した統計処理方法を適用することが重要である。そこで、フローチャート(図1)に基づいて統計解析方法の選択と実際について述べる。

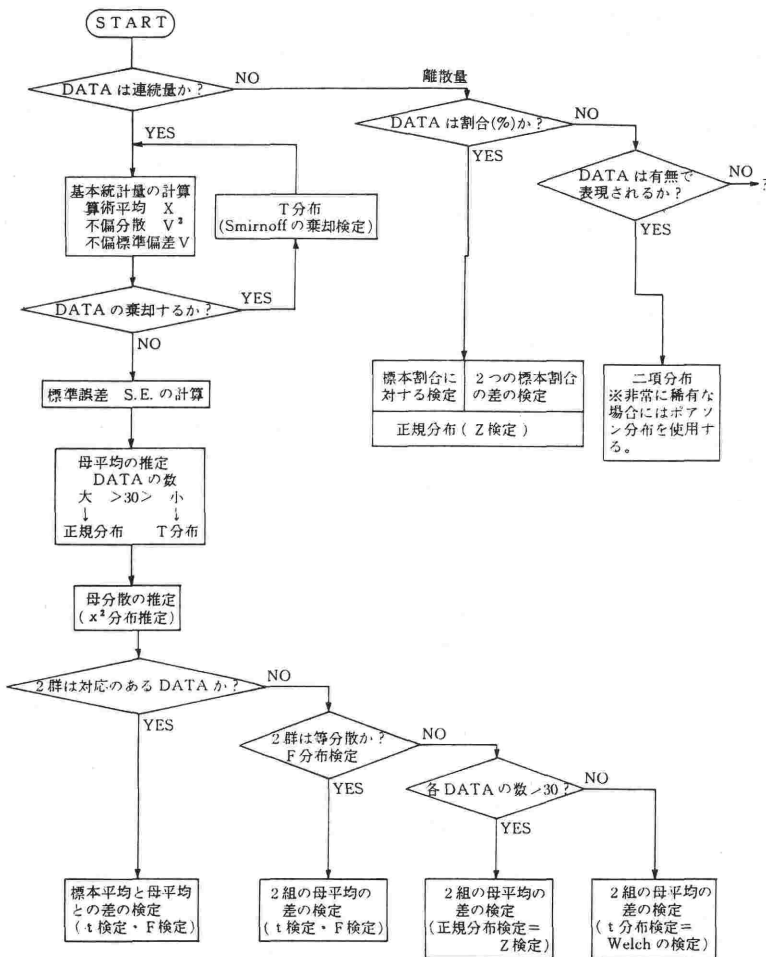


図1 統計処理のフローチャート

*日本医科大学麻酔科学教室

1. データの性質と分布の型

a. データとは何か

データとは標本の特性を表現したものである。これを臨床医学に当てはめれば、患者の属性を数値その他で表現したものとなる。

b. 母集団と標本

筆者の病院を受診する患者は、たまたま近くに住んでいるためである。全国の患者のほんの一部である。全国の患者（母集団）の中からたまたま抽出されたことになる。このような患者を対象として行った研究結果（例えば麻酔前の CPK 平均値）は、全国の患者の平均（母集団の平均すなわち母平均）とは一致しない。B病院で同じ研究をすれば異なった結果が得られる。したがって筆者が得た CPK 平均は無数に得られる CPK 平均値の一つである。母集団から抽出された標本の平均（データ平均）である。

しかし筆者が得た平均値は全国の麻酔前患者の CPK 平均値からそれほど隔たっていない。そこでデータ平均は母集団の平均（母平均）の推定値を意味をもつ。

c. データの種類

データとは患者の属性の表現である。表現からは、分類的表現と数量的表現に分けられる。

1) 分類データと数量データ

前者は男・女、生存・死亡、などの区別する表現である。後者は白血球数（個数/mm³）、罹病者の数（人）、体重（kg）、血糖値（mg/g^l）などであり、数または量で表現される。前者を分類データ、後者を数量データとよぶ。

2) 連続量と離散量

数量データも2種類に分けられる。ある病気に罹った患者を数える場合、何人中何人となり端数ではない。このような数データは離散量と呼ばれる。体重はどのような端数も取り、連続的に変化する。このようなデータは連続量と呼ばれる。

3) 尺度

さらにデータは表現の内容より4つの尺度に分けられる。

i. 名目尺度

違いを表現するデータであり、前述の男・女、生・死などがこれに当たる。

ii. 順序尺度

大小、優劣、長短など不等号（<、>）で区別されるデータである。

iii. 間隔尺度

不等号で区別されるデータであるが、それらの間隔がおおよそ等しい場合である。たとえばペインスケール（VAS）の60点と70点の距離は70点と80点との距離とほぼ等しい。

IV. 比例尺度

ゼロがはっきりしており、加減乗除に耐えうるデータである。通常の離散量、連続量はこれに当たる。

このような尺度によって、代表値の選び方や統計処理が異なってくる。名目尺度、順序尺度、間隔尺度は割合（%）で代表され、統計処理には二項（多項）分布が使われる。比例尺度（時に間隔尺度も）になってはじめて分布型を確認した後、算術平均（中位数、最頻値）で代表させ、正規分布やt-分布の性質が適用される。

d. 分布の型

研究を行いデータを収集したら、統計解析に移ることとなる。まず行うべき事は分布の型の確認である。このためには必ず度数分布表を作り、ヒストグラム（柱状図）を描く。度数分布表の階級の数、データ数により定まり表1のように設定

表1 データ数と階級との関係

データ数	10	50	100	1,000	10,000	100,000
階級数	4	7	8	11	15	18

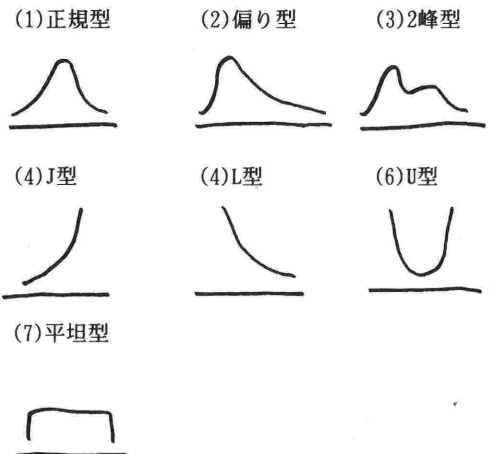


図2 ヒストグラムの様々な型

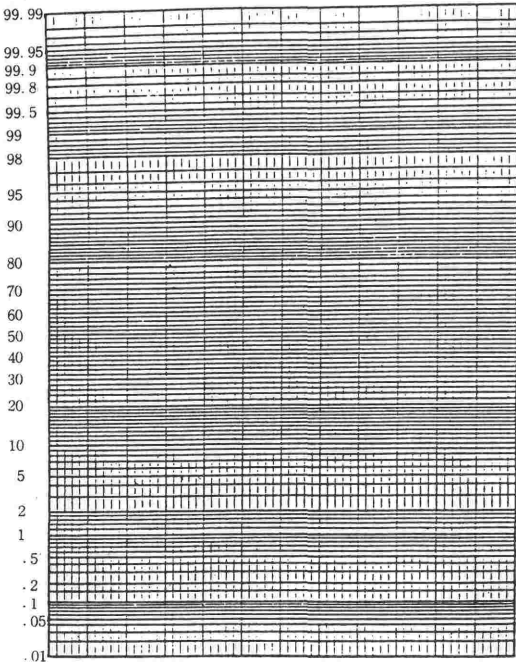


図3 正規確率紙

する。

ヒストグラムを描くとさまざまな型(図2)が見られる。左右対称の釣形鐘型は正規分布型とよばれる。ヒストグラムが正規分布型のものに限って、通常(古典的)の統計処理が可能となる。

データが正規分布かどうかを確認するには正規確率紙(図3)が用いられる。もし正規分布型でなければ、種々の転換法により正規分布に修正する。正規分布に修正できなければ、通常的手法と異なった統計処理法を使わなければならない。これは一括してノンパラメトリック統計処理法と呼ばれる。

e. 代表値の選択

データを代表させる数値が必要となる。名目尺度、順序尺度、間隔尺度では最頻値が代表値となる。臨床医学ではデータ代表として算術平均を使うことが多いが、データが比例尺度で正規分布型と確認された場合に限られる。比例尺度でも図2のようにさまざまな分布型があり、J型やL型では算術平均より中位数が代表値となり、偏り型や2峰型では最頻値が代表値となる。ヒストグラムを描き、分布の型を確認せず、算術平均を求めても無意味である。

2. 正規分布データの標準化

正規分布しているデータのみが、古典的(通常の)統計処理の対象となる。その理由は正規分布していれば、総ての分布が単一の分布に統一できる。そのため次ぎの操作を行う。①平均値が0となるよう平行移動する、②標準偏差が1となるよう横軸のスケールを変える、③分布曲線の下面積が1となるよう縦軸のスケールを変える。標準正規分布曲線は数式的には、

$$F(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

となる。新しい横軸(Z)は、 $Z = (X - \mu) / \sigma$ となり、これをZ変換と呼ぶ。

教科書の巻末に載っている正規確率表はこの分布の左右斜線部分の面積(図4)である。古典的統計学はこの表に基づき、判断を下している。

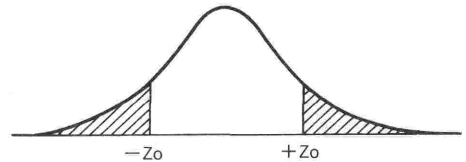


図4 標準正規分布曲線と確率の関係。正規確率表にはZ軸の目盛りと斜線部分の面積(確率)を表示している。

3. 統計解析の第一歩：連続量が離散量か

データが比例尺度であれば連続量が離散量かをまず考える。連続量で正規分布していれば、チャート(図1)にしたがって下方へ進む。離散量であればチャート(図1)を右へ進む。

連続量で正規分布しているデータでは次ぎの項へ進む。

4. 連続量の基本統計

算術平均と標準偏差は正規分布データの重要な指標で、統計学ではパラメータと呼ばれる。パラメータを使った統計処理が通常(古典的)の方法であり、パラメトリック検定と呼ばれる。

a. 平均値の計算

算術平均(X)は次式で計算されるが、後に説明するように母平均(μ)の点推定値でもある。

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 \cdots + X_n) = \frac{1}{n} \sum (X_i - \bar{X})$$

b. 標準偏差の計算

標準偏差 (V, variance の略) は通常不変標準偏差を求める。その理由は標準偏差は母標準偏差の点推定値の意味があるからである。

$$V = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2}$$

c. SD と SE の違い

標準偏差は正規分布の重要な指標である。 $\bar{X} \pm SD$ の間にデータの68%が、 $\bar{X} \pm 2 \times SD$ の間にデータの95%が入る。

多数の母集団より標本を取り出し、その算術平均を求める。このような作業を重ねれば、多数の算術平均 ($X_i; i=1, 2, \dots, n$) が得られる。この平均値の群は必ず、正規分布をなし母平均 (μ) の周辺集まる。この集団の標準偏差を標準誤差 (SE) と呼ぶ。標準誤差は次式で求められる。

$$\text{標準誤差} = \text{標準偏差} / \sqrt{n}$$

以上のように標準誤差は「平均値の標準偏差」であり、散布度とは異なる。したがってデータの散らばりを示す目的で標準誤差を使うのは誤りである。研究者の陥りやすい誤りのひとつである。

d. 棄却の問題

研究者はデータの散らばりが気になる。またとんでもないデータが現れることがある。このようなデータを棄却してよいかどうかを検定するのが、棄却検定である。棄却検定の原理は、抜き出したデータが、平均値 $\pm 2 \times$ 標準偏差より外れていれば、95%以上の確率で、このデータが同じ母集団から抽出されたものではないとの考えの上に立つ。

実際にはスミルノフ、トンプソンの方法があり、前者は棄却したいと考えるとき、後者は保留したいと考えるとき適応すると安全性が高い。

e. データ数の問題

多少の意見の相違はあるが、データ数が30以上では大数データ、30未満では少数データとする。大数データでは分布の型を確認するのが容易であり、正規分布の性質を適用して統計処理を追及できる。少数データでは、データ平均を母平均の代わりに使うには誤差が大きくなる可能性がある。このためゴセットが考えた t-分布を使う。t-分布を用いた検定が t-検定である。

5. 推 定

基本統計量を計算したら母平均の推定に移る。データ数により、大数であれば正規分布、少数であれば t-分布に基づく。

a. 確率の考え

統計学の判断はすべて確率に基づく。確率は、難しく言えば、ある事象が現れる「相対度数の極限」である。私共が統計判断に用いる確率 (有意水準) は5%である。この5%は研究者の約束事であり、科学的な根拠はない。

b. 推定の実際

筆者の病院で全身麻酔下に胃切除手術をおこなっている成人患者25例の血液を採取して、血漿中のコルチゾールを定量した。これらのデータが正規分布しており代表値として算術平均を計算し、チラバリの指標として標準偏差を求めた。結果は $69 \pm 15 \mu\text{g/dl}$ であった。ではなぜ平均値を知りたいのか?。それは全身麻酔下に手術を受ける患者の「真の、普遍的な」血中濃度を知りたいためである。

ただし測定したデータの平均が必ずしも母平均と等しいとは限らない。他の病院で同じ研究をすれば異なった値が得られる。全国の n 病院で研究を行えば、n 個の平均値が得られる。それらの平均値 ($X_i; i=1, 2, 3, \dots, n$) は真の値 (母平均 μ) の周辺に分布する。そして正規分布の性質から、n 個の内の95%は、 $\mu \pm 1.96 \times$ 分布の標準偏差 (σ) の間に含まれる。当然私共の測定値も95%の中に入る。

$$\mu - 1.96 \times \sigma < 69 < \mu + 1.96 \times \sigma$$

上の式より、次ぎの式が導かれる。

$$69 - 1.96 \times \sigma < \mu < 69 + 1.96 \times \sigma$$

ここで σ は知ることができない。データ数が30以上と多ければ、 $\sigma =$ データの標準偏差 (V) より推定してよい。n はデータ数である。

$$\sigma = \frac{V}{\sqrt{n}}$$

したがって、

$$69 - 1.96 \times 15 / \sqrt{25} < \mu < 69 + 1.96 \times 15 / \sqrt{25}$$

計算すると、 $67 < \mu < 71$ にあると推定される。そして67-71派信頼区間と呼ばれる。以上の経過をまとめると次のようになる。

母平均は平均値 $\pm 1.96 \times$ 標準誤差の間にある。
ここで、標準誤差 $= 1.96 \times$ 標準偏差 $/\sqrt{n}$

これより標準誤差 (standard error, SE) は平均の変動範囲を示す指標である事が理解される。標準偏差と標準誤差は全く異なった意味をもつ。

6. 検定

推定に次いで重要なのは母集団に関する統計学的仮説の検定である。検定は確率に基づく。統計学的仮説とは、AがBと等しい ($A=B$) の形を取る。そして $A=B$ が成立する確率を求める。もし確率が5%未満であれば、仮説を棄却する。すなわち、 $A \neq B$ となる。

1) 第1種の誤差と第2種の誤差

統計仮説の検定では、二つの過誤を犯す可能性がある。第1種の過誤は、 $A=B$ なのにこれを、否定する誤りである。その確率を α (有意水準) という。これが私共がよく使う、P である。第2種の過誤は、 $A=B$ ではないのに、 $A=B$ を受け入れる誤りで β という。 α と β は裏腹の関係にあり、 α を小さくすれば β が大きくなる関係にある。

2) 片側検定と両側検定

統計仮説は、 $A=B$ である。一方 $A=B$ に対立する仮説は $A \neq B$ である。対立仮説と呼ぶ。さらに対立仮説は $A > B$, $A < B$ に分解される。これを正規分布よりみると、左右の裾野にあたり、右の裾野が $A > B$ を意味すれば、左の裾野が $A < B$ である。左右の裾野部分の面積が、全体の5% (0.05) に当たる。この時の Z 値は1.96である。通常は対立仮説 ($A > B$, $A < B$) の両方を考慮する。

まれに $A > B$ あるいは $A < B$ が自明な場合、片側の裾野部分を考慮せず、片側の面積を5% (0.05) と考える。この場合は Z 値が1.64となる。すなわち片側検定では $Z > 1.64$ で仮説を棄却する。しかし研究者の間で、 $A > B$ あるいは $A < B$ が了解されていない限り、片側検定は使われない。

3) 2群の平均値の差の検定

関係のない2群 (それぞれ n_1 , n_2 個のデータ数よりなる) のデータを得た。それらのデータ平均値 (\bar{X}_1 , \bar{Y}_1) とデータ標準偏差 (V_{x1} , V_{y2}) のがそれぞれ得られた。これらのパラメータはたまたま

今回の研究で得られたものである。同じような研究を多数回繰り返せば、多数の組のデータ平均が得られる。2組の平均値の差 ($\bar{X}_i - \bar{Y}_i$) は近似的に正規分布することが知られている。 ($\bar{X}_i - \bar{Y}_i$) の分布の標準偏差 (σ_z) は、下記の式で示される。

$$\sigma_z = \sqrt{\frac{V_1^2}{n_1} + \frac{V_2^2}{n_2}}$$

ここで σ_x , σ_y は二つのデータの母標準偏差である。もし X_1 と Y_1 が同じ母集団からのデータであれば、その差はゼロに近いである。

仮説: $\bar{X} = \bar{Y}$

ここで ($\bar{X}_i - \bar{Y}_i$) は95%の確率で分布の標準偏差の2 (1.96) 倍に入る。したがって、測定値の平気の差 ($\bar{X}_i - \bar{Y}_i$) が分布の標準偏差の1.96倍より大きければ、 \bar{X}_i と \bar{Y}_i が同じ母集団から抽出されたとの仮定が棄却される。

そこで、 $Z = (\bar{X}_1 - \bar{Y}_2) / \sigma_z$ を計算して、 $Z_{0.05} = 1.96$ より大であれば仮説を棄却する。

7. 離散量の取り扱い

データが離散量であれば、チャート (図1) を右に進む。

a. 割合で表現されるデータ

離散量の多くは割合 (%) として表現される。次ぎの手順で推定と検定をおこなう。

1) 母割合の推定

母割合 (π) の推定には正規分布が使われる。データ割合 (p) より母割合の分布の標準偏差 (σ_π) は、 $\sigma_\pi = \sqrt{\frac{p(1-p)}{n}}$ であるからデータ数が30以上であれば、95%信頼限界は次式で推定できる。

$$\pi = p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

データ数が30未満では上式は用いられない。

2) 2つの割合の差の検定

母割合 (π) の母集団から n_1 個, n_2 個の標本を抽出し、2つの測定割合 (p_1 , p_2) を得た。その差 ($p_1 - p_2$) は正規分布をなす。その正規分布の標準偏差は次式で求められる。

$$\sigma = \sqrt{p(1-p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

したがって ($p_1 - p_2$) が $1.96 \times \sigma$ より大であれば、仮説: $p_1 = p_2$ が5%の有意水準で棄却できる。すなわち $Z_0 = (p_1 - p_2) / \sigma > 1.96$ であれば仮説を

棄却する。

8. 非正規分布のデータ

前述の古典的統計学では、データの分布が正規である事が大前提であった。そしてパラメータを用いて、推定と検定をおこなった。ところが、医学領域では正規分布していないデータも存在する。例えば、肝臓腫瘍根治術後の生存時間(週)などが挙げられる。またデータ数が少なく、分布の型が判明しない場合もある。

このようなデータでは、算術平均は代表値とはならず、標準偏差も意味がない。そこでパラメータを使わない統計処理方法が必要となる。この場合、代表値には中位数、最頻値を使い、チラバリの指標には範囲(最大値と最小値)、4分1範囲(全体の3/4が含まれる範囲)などを示す。そして2群の差の検定、3群以上の差の検定、相関なども順位を用いて行う。そこでノンパラメトリック統計処理を古典的統計処理と対比して示す。

古典的手法：ノンパラ的手法

独立 t-検定：ウイルコクソンU検定，マン・ホイットニー検定

対応 t-検定：ウイルコクソンT検定

分散分析：クリスカル・ウォーレス検定，フリードマン検定

相関関係：スペアマン，ケンドールの相関係数
最近では分布の型に敏感になっているので、ノンパラメトリック手法が広く採用されている。実際の方法は成書に譲る。

9. 論文への記載事項

研究を終了しデータを解析したら論文にまとめる。通常論文中の「対象と方法」の最後に統計手法を記載する。これには①群の分け方と対照群の明記，②群中のデータ数，③分布の型，④代表値，⑤チラバリの指標，⑥検定方法，⑦有意水準が含まれる。

おわりに

以上統計処理方法の選択と実行について述べたが、臨床医学における統計学の役割は研究計画の立案にあることを強調したい。その中心的課題は、研究計画を統計学的管理下に置くことである。この点は本稿では省かせていただいた。