講演

研究における統計学の基礎

谷 浩明¹⁾

LECTURE: A Fundamental Knowledge of Statistics for the Researches
Hiroaki TANI, RPT¹⁾.

1)Dept. of Physical Therapy, Tokyo Metropolitan College of Allied Medical Sciences
7-2-10 Higashiogu Arakawa-ku Tokyo (173) Japan. TEL 03-819-1211.

J. Exerc. Physiol. 3(3):149-154, 1988. ISSN 0912-7100. Received June 30, 1988.

ABSTRACT

Presumptive Statistics that clarifies characteristics of a population using those of random samples of unlimited population is necessary for our researches. At present, presumption and certification representing major methods of this statistics are easy to be conducted by virtue of development of softwares for the statistics. On the other hand, in spite of this handiness, they also involve a risk of improper use without understanding the meaning of the method. Especially, although certification of difference in average values and analyses of correlation are frequently observed in articles, it is seldom that the results obtained were sufficiently discussed. This tendency is considered to result from lack of understanding of the concept of presumption or certification, not from that of meaning. Comprehension of the concept of presumptive statistics is indispensable to consider what kind of or to what accuracy presumption or certification should be performed at the preparation stage of the experiment preceding sampling.

Key Words: statistic, research,

要旨

我々の研究において必要とされる統計学は、無限母集団から無作為抽出した標本の特性を知ることによって母集団の特性を推測する推測統計学である。現在ではこの統計学の手法である推定や検定は、統計用ソフトウェアの普及によって容易に行うことが可能だが、それゆえ、手法の意味を理解せずに誤って用いる危険性もはらんでいるといえる。特に、平均値の差の検定、相関分析などは用いられる機会が多いにもかかわらず、その結果を記述するのみで解釈が充分になされていないということがよくあるが、これは、各手法の理解というよりも、推測統計学における推定や検定の考え方の理解の度合に起因すると思われる。この推定、検定の考え方は、標本抽出に先立つ実験計画の段階から、どのような推定や検定を、どのくらいの精度で行うのかを考慮しておかなければならないという意味を理解するうえで非常に重要である。

キーワード: 統計 研究

1) 東京都立医療技術短期大学 理学療法学科: 東京都荒川区東尾久7-2-10 (〒116)

TEL 03-819-1211.

運動生理 3(3):149-154, 1988. 受付日1988年6月30日 (第2回運動生理研究会夏期研修会)

I はじめに

研究を行う上で統計処理は必要不可欠なものだという話を耳にすることがある。実際、学会発表においても図1に見るような統計的手法が用いられている。この学会に限らず、発表される論文の多くになんらかの統計的手法が用いられており、場合によっては、どのような統計的手法が使われているかが問題にされることさえある。このように統計的手法が浸透してきたのは、理学療法士自身の研究方法に対する考え方の変化だけでなく、パーソナル・コンピュータの普及と統計用ソフトウェアの充実が、こうした統計処理をより容易にしているという背景もあるだろう。

しかし、統計処理の必要性が唱われる中で、我々は必要以上に統計を恐れてはいないだろうか。また、コンピュータの打ち出す「危険率××%で有意な差が認められました。」が最も重要な結果であると考えたり、「有意な差が認められない」ということで、考察することをやめてしまっていないだろうか。

そこで今回、「統計は自分の測定した人間データの意味を明瞭にする”道具”である」という立場から、統計処理の意味と役割を考えてみたいと思う。

II 母集団と標本

母集団とは研究、調査の対象となるものの全体を指している。たとえば、「脳卒中右片麻痺患者」「前十字靭帯損傷例」「20代の健康な男性」のように表されるもので、これらは研究者の考える条件によって規定されている。このため、「脳卒中右片麻痺患者」といっても、その患者の年齢、性別、発症からの時間など様々な要素を考えておくことが重要である。

このように設定された母集団からその一部

図1 学会発表にみる統計的手法
(第22回理学療法士学会)

図2 統計的推論

である標本を抽出し、この標本の情報から母集団の特性を理解しようとするのが、統計による推論である(図2)。

正確な推論を行うためには、標本が母集団を代表するような集団であることが望ましく、これを満足するために無作為抽出(random sampling)が行われる。ただし、年齢、性別などが問題となるような疾患の場合には、層別後、各層から無作為抽出を行う層別抽出(stratified sampling)が行われる。この無作為抽出、層別抽出を行うことによって、「脳卒中右片麻痺」を母集団としているにも関わらず、実験を頼みやすい人だけを集めた標本になってしまうといったような系統的な偏りを防ぐことができる。ここで問題になるのは抽出する標本の大きさであろう。これについては、後述する検定と推定の項で触れたいと思う。

Ⅲ ヒストグラムについて

標本の分布状態を知る方法のひとつとしてヒストグラムがある。これは、統計用ソフトウェアには必ずと言っていいほど組み込んであるプログラムであるが、実験の偏りを見つけないという点からは、コンピュータまかせにしないほうがいい場合もある。例えば、図3(a)のようにCRT（もしくはプリンター）に出力されたとしよう。確かに山が2つあり、分布に偏りのあることが予測される。ところが、自分の手で実験順序を書き込みながら作成すれば、図3(b)のような偏りか(c)のような偏りかを知ることができる。もし、(c)のような偏りが見られれば、まず最初に5番目から10番目の実験状態を調べてみるなど、同じ層別の必要性を感じてもその後にとる行動(action)のための情報が多いという利点がある。もちろんプログラムを改造すれば、表示後、あるクラス以上のものだけを出力させることは容易だが、改造の手間を考えると、膨大なデータ数でない場合には自分の手で作成してみるのが良いだろう。いずれにしても、ヒストグラムは、層別の必要

図3 ヒストグラム

な集団であることを見落として統計処理を行う危険を未然に防ぐ簡単にして有効な方法であるということができよう。

Ⅳ 平均値の差の検定について

我々が論文を書く際に最も使用頻度の高い統計的手法といえ、この平均値の差の検定であろう。しかし、その割にあまり正確に使われてないのも事実である。一般的に、平均値の差の検定と呼ばれるものには次の5つがある。

- 1) σ (母標準偏差) 既知で等しいとき
- 2) σ 既知だが等しくないとき
- 3) σ 未知だが等しいという確信のあるとき
- 4) σ 未知で等しいという確信のないとき
- 5) 対応のある平均値の差の検定

このなかで、我々がよく用いるのは3)~5)である。

まず、5)の対応のある場合について考える。これは、本によっては対をなすデータの場合とも書かれているが例としては次のような問題が考えられる。

五十肩の患者6人に対してA、B2種類の温熱療法を行った後の肩関節外転の可動域を表1に示す。A、B2種の治療により可動域に有意な差があるか。

この類の実験で、対応のない場合の検定を行っているのをみたことがあるが、問題したいのはA、B間の差であるから、A、B毎に6人の平均を求めてしまうと、各患者のばらつきでA、B間の差が見えなくなってしまう。そこで各患者毎にA、B間の差を求めて

表1 温熱療法後の肩関節外転

被験者	X1	X2	X3	X4	X5	X6
方法A	40°	50°	30°	55°	60°	70°
方法B	50°	50°	60°	60°	80°	70°

その平均をとる。この「差の平均」はA、B間に差がなければ0になると考えられる。これより、差の母集団の平均が0である ($\mu_d = 0$) とする帰無仮説ををたてて、検定を行うわけである。

次に、対応のない場合について考えてみる。この例として考えられるのは、右片麻痺患者と左片麻痺患者の比較、健常者と閉塞性肺疾患患者の比較などであろう。ここでは3)と4)の違い、すなわち母標準偏差が等しいという確信の有無による違いについて触れておきたい。3)の等しいという確信のあるときに用いられるのが、t検定であり、4)の等しいという確信のないときに用いられるのがWelchの検定である。では、等しいかどうかの確信はどうやって得られるかだが、一般的には、等分散の検定を行い有意でなければ等しいという確信が得られたとしてt検定を、有意であればWelchの検定を行う方法がとられている。

V 相関係数について

相関関係を分析する手法も平均値の差の検定に次いでよく用いられている。ここでは相関係数に関する問題点をいくつか挙げてみる。相関分析を行う際に最も間違いを冒しやすい点は2つの変量の間に高い相関係数が認められたとしても、それは必ずしも、2変量間に相関関係があるとはいえないということである。なぜなら、この高い相関係数は2変量のそれぞれが関係している別の因子によって引き起こされていることも考えられるからである。こうした考え方を踏まえたうえで相関係数rの実際上の扱いについて触れてみたい。

まず、相関係数を求めるうえで忘れてはならないのが散布図である。これもヒストグラムと同様、統計用ソフトウェアの多くに組み込まれているが、より多くの情報を得ようと

図4 散布図

する立場にたてば、一度は自分の手で描いてみるのが望ましい。または、プログラムを改造して、キーを押すと入力データの順に点が表示されて行くようにするのもひとつの方法である。言うまでもないが、散布図を描かないで数値計算のみで相関係数を求めたり、検定を行うなど論外である。今、仮に図4(a)のような散布図が得られたとしよう。この時、データL1、L2をどうするか、ということが問題になる。L1、L2をいれておいた方が相関係数が高そうだから外さないというのでは話にならないが、かと言って単純に異常値だといってこのデータを外すのも正しいだろうか。もちろん、異常値を統計的手法を使って外すことは可能であるが、まず、行わなければならないのは、実験を行ったとき、L1とL2に共通で他のデータとは異なる状況がなかったかということを調べることである。

これによって技術的な問題が確かになれば、それを改善する必要があるであろう。また、L1、L2というのは異常値ではなく、図4(b)のような分布の中で取りうる値として外さないでデータをさらにとってみるという考え方もある。どちらにしても、自らが実験で得たデータに対する処理の選択は正確に行うべきであり、根拠なく異常値と決めてかかり、データを外したりすることは避けるべきである。

次に、相関係数の検定についてであるが、無相関の検定($\rho = 0$)を行う場合に、検定の結果のみで区間推定値や標本数が示されていないことがたまにある。これでは、相関分析を行ったというにはいささか片手落ちといわざるを得ない。なぜなら、標本相関係数が同じ $r = .80$ であったとしてもその標本数が $n = 20$ であれば、母相関係数 ρ の95%信頼区間は $0.553 \sim 0.918$ 、 $n = 200$ であれば、 $0.716 \sim 0.861$ となるからである。このことから、相関分析を行った際には標本数と区間推定値を明記しておく必要があるだろう。

VI 検定と推定について

ここで、検定について考えてみたい。平均値の差の検定の場合、2つの標本が、実は同じ母集団から抽出されてきたものだとする帰無仮説($\mu_a = \mu_b$)をたてて、これが棄却できるほど標本平均値の差が大きいかどうかを確かめる、すなわち標本平均値の差が偶然生じたとするには大きすぎると判断した時、「統計的に有意な差がある」として帰無仮説を棄却するわけである。この「帰無仮説を棄却できる」と判断したときに、これが間違っている(=帰無仮説は正しい)可能性は危険率として表され、帰無仮説が正しいにもかかわらず、棄却する誤りのことを「第1種の誤り」と言い、 α で表される(図5斜線部)。この

図5 「第1種の誤り」と「第2種の誤り」

逆に帰無仮説が正しくないときにこれを棄却しない誤りのことを「第2種の誤り」と言い、 β で表される(図5横線部)。 α は研究者が有意水準(危険率)を決めることによって決定される。 α を決めたとえで β を小さくする(検出力 $1 - \beta$ を大きくすること)ためには標本数(n)を増やしてやればよい。多少、乱暴だが、標本数を増やしてやればやるほど検出力は上がり、有意な差も出やすいという言い方もできる。これは、我々が実験を行う前に、どのくらいの差をどのくらいの確率で検出したいのか、そのために、どのくらいの数の標本を抽出すればよいかを決めておかなければならないことを示唆している。つまり、標本の数を増やして有意な差が出ててもその後の行動(action)にとっては意味のない差だったという結果を招く危険もあるわけで、これは、例えば、Aという治療法による可動域の改善度と、Bという治療法による可動域の改善度との違いを見る場合、研究者はどのくらいの改善度の差があれば、治療として意味があるか(××度改善すれば、ADLの達成度が変わってくるなど)を考えて実験を組まなければならないということを意味している。

統計の講義などで、検定を行ってから区間推定を行う例題を解くような問題に触れるせいか、検定と推定はつながった手法のような印象を持つ人がいるが、推定と検定は全く異なる手法である。さきほどの関節可動域の治療を例にとって、研究者は改善度において10°差があれば意味があると考えていたとしよう。このとき検定で有意にならなくても区間推定

を行い、 $\mu_a - \mu_b$ の95%信頼限界で $3.5^\circ \sim 12.5^\circ$ と導きだされ、「 10° の差があるかもしれない」という情報を得ることができるとすれば、それは十分に意味のあることであろう。

VII おわりに

研究というものから統計のみを取り出して論ずることは困難である。それは、研究そのものが実験を計画する段階から統計と密接に結びついており、統計的手法の使い方や、解析結果の解釈というものが、その研究者の研究に対する取り組み方、研究者自身のセンスといったものと深く関わっているためであろう。「科学者は例外なしに、統計学者であるか、あるいは統計学者と親密かつ永続的な関係をもっているかのいずれかである。」¹⁾という言葉に示されるように、我々が研究を続けていくためには正確な統計の知識が必要とされているといえよう。“道具”が“よい道具”となるか“悪い道具”となるかは研究者自身に委ねられているといってもいいだろう。

本稿を終えるにあたり、御校閲いただいた梅村守先生（東京理科大学工学部）に深謝致します。

引用文献

- 1) Carl J. Sindermann. 著, 山本祐靖, 小林俊一訳: 成功するサイエンティスト. 丸善.

参考文献

- 1) T.D.V. Swinscow. 著, 西村昂三監訳, 大島邦夫訳: 医学・薬学・生物学のための統計処理, 共立出版.
- 2) 近藤良夫, 安藤貞一編: 統計的方法百問百答. 日科技連.
- 3) 増山元三郎監修, 吉川英夫編: 統計解析手順集, 日科技連.
- 4) Marcia K. Johnson, Robert M. Libert 著, 西平重喜, 村上征勝訳: 統計の基礎—考え方と使い方—. サイエンス社.
- 5) J.C. Miller 著, 村上正康訳: 統計学の基礎培風館.
- 6) 石川馨, 中里博明, 松本洋, 伊東静男著: 改訂版 初等実験計画法テキスト, 日科技連
- 7) A.L. Edwards 著, 並木博, 小林ポオル, 内藤俊史, 佐伯千鶴子訳: 直線回帰と相関, 慶應通信.
- 8) 久米均, 飯塚悦功著: 回帰分析, 岩波書店.
- 9) 蓑谷千風彦著: 推定と検定のはなし. 東京図書.
- 10) 丹後俊郎著: 臨床検査への統計学, 朝倉書店.
- 11) 駒澤勉著: 医学統計解析の基礎, 朝倉書店
- 12) 石川馨著: サンプルング法入門, 日科技連