

2. 組織細胞化学シリーズ (若手研究者へのヒント)

病理標本と臨床因子解析の基礎 (7)

赤城 一郎¹ 宮下 正夫¹ 内田 英二¹ 瀧澤 俊広²

¹日本医科大学大学院医学研究科臓器病態制御外科学

²日本医科大学大学院医学研究科分子解剖学

2. Histocytochemistry Series

The Basic Knowledge for the Analysis of Clinicopathological Features (7)

Ichiro Akagi¹, Masao Miyashita¹, Eiji Uchida¹ and Toshihiro Takizawa²

¹Division of Surgery for Organ Function and Biological Regulation, Graduate School of Medicine, Nippon Medical School

²Division of Molecular Medicine and Anatomy, Graduate School of Medicine, Nippon Medical School

Abstract

The basic knowledge of treating clinicopathological features is essential for analysis. In this note, we describe procedures for the analysis of clinicopathological features, with a focus on statistics.

(日本医科大学医学会雑誌 2011; 7: 24-26)

Key words: clinicopathological features, statistics

はじめに

本シリーズの最終回は、実験によって得られたデータを実際にどのように臨床統計解析をするかということに焦点を絞って解説する。

臨床症例のデータベース化

研究を行う場合、必ずといってよいほど実験データをコンピュータに入力する必要がある。コンピュータによるデータベース化は、データ収集と分析の確実性とスピードを上げ、ミスの検出を容易にし、またデータをグラフ化したり、新たな変数を作成するうえで非常に有用である。しかしながらどのような研究においても、測定、収集、転写、コンピュータへの入力の際

にデータセットにミスが生じる可能性がある。これらすべてのミスを排除することは難しい。しかし、そのことを念頭において慎重にデータをチェックすることによって、入力時、転写時のミスを減らすことができる。

入力誤差

データ入力時のミスが誤差の最も多い原因である。誤差がカテゴリーデータ内に存在した場合、そのカテゴリー内に定められた範囲から外れた値は誤差に違いないので判別が比較的容易である。しかしながら、誤差が数値データの場合はしばしば判別が難しい。それぞれの変数に上限、下限が指定されている場合はその範囲内かどうかでチェックできる。誤差のチェックにおいては、ミスがあったという根拠があった場合は数値を修正すべきであるが、単にそれらが異常に見えるからといって安易に数値を変更してはならない。

Correspondence to Toshihiro Takizawa, Division of Molecular Medicine and Anatomy, Graduate School of Medicine, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113-8602, Japan

E-mail: t-takizawa@nms.ac.jp

Journal Website (<http://www.nms.ac.jp/jmanms/>)

欠損値の処理

若干のデータが欠ける可能性は常に存在する。データが非常に広範囲に欠けているなら結果は信頼できそうにない。欠損値が特定の変数や個体群の特定のサブグループの集団に偏っているなら、なぜデータが欠けているかを必ず調べるべきである。それは変数がその個体群に不適切であったか、あるいはその個体群を測定したことがないことを示しているのかもしれない。このような場合は、その変数もしくはその個体群を分析から除外する必要がある。

免疫組織染色の判定法

免疫組織染色で得られた結果を陽性、陰性と判断するためには必ず事前に陽性コントロールと陰性コントロールを用意し、これらと常に比較しながら慎重に判定することが必要である。一般的に判定法は二種類に分類され、染色される部位にも注意が必要である。

染色率による判定

目的とする細胞群（例えば癌細胞）を同定し、そのうち何パーセントの細胞が染色されているか (rate)、ある一定の割合以上染色されていれば陽性と判定する方法である。いわゆる cut-off 値をいくつに設定するかに関する規定はないため比較的自由に設定することができる。もし自分が使用した抗体と同じ抗体を使用した論文がすでに公表されている場合はそれを参考にすることができる。

染色強度による判定

目的とする細胞群の染色の強さ (intensity) で判定する方法である。染色強度をスコア化して(例えば 0, 1+, 2+, 3+) 群分けするのが一般的である。染色強度の判定は独自の基準で行うのではなく病理専門医と相談しながら慎重に行う必要がある。

染色部位

目的とする抗原が、細胞質にあるのか、核にあるのか、存在する部位を想定しながら判定することが大切である。例えば本来、核に染まるべき抗体が細胞質に強く染まっているときは、偽陽性 (false positive) の可能性があるため染色条件を変更したり、同様の抗体を使用している文献を検索するなど慎重に判定する必要がある。また、標本が癌の検体であった場合、腫瘍のどの部分に強く発現しているかも検討する必要がある。例えばある抗原が腫瘍の先進部で強く発現してい

れば、腫瘍の進展と関係があるかもしれない。いずれの場合も病理専門医と協議しながら注意深く判定するのが望ましい。

統計解析

ここでは、病理標本と臨床因子を解析するために必要な統計学の基礎と、どのような場合にどのような統計学的手法を選択すべきかに焦点を絞って説明する。

統計学の基礎

量的データの大きさを表す最も分かりやすい代表値は平均値 (mean) であろう。標本の平均はサンプル全体を一つの数で代表する役割があり、平均以外の代表値として中央値 (median) がある。中央値とは、有限個のデータを小さい順に並べたとき中央に位置する値のことである。同じ平均値をもつデータでも、その散らばり方が異なると意味も違ってくる。例えば、3人の年齢が39歳、40歳、41歳でも、30歳、40歳、50歳でも平均年齢は同じ40歳である。平均値はすべてのデータ値を対象に均等に配分した値なので、個々のデータはその周りに分布する。平均値だけでは、データの様子はつかみきれたとは言えない。この点を克服するために観測値と標本平均の差を2乗した値を足し合わせてばらつきの目安とする。この値を偏差平方和 (sum of squares) という。さらに対象の数の異なるものも比較できるように偏差平方和を対象数で割ったものを分散といい、この平方根を標準偏差 (standard deviation) という。この標準偏差はデータのばらつきを表現する上できわめて重要な値であり平均値と共に解析した図表やグラフに記載されることが一般的である。一方、標準誤差 (standard error) とは母集団からある数の標本を選ぶ時、選ぶ組み合わせによって統計量がどの程度ばらつくかをすべての標準偏差で表したものをいう。標準偏差と標準誤差の区別は、標本平均をサンプルの代表値と見るか、母平均の推定値と見るかの違いに対応している。n数が多いと見かけ上、このエラーバーは小さく見えてしまうので、必ずn数を明記することが誤解を招かないようにするために重要である。なお、図表のエラーバーを標準偏差で表示しようが標準誤差で表示しようが検定結果は同じである。

検定法には大きく分けて2種類 (パラメトリックとノンパラメトリック) がある。パラメトリックな検定法では、母集団の分布を一定の形 (正規分布など) であると仮定し解析する手法である。また、臨床病理学

表1 クロス集計表

	A (+)	A (-)
B (+)	a	b
B (-)	c	d

的因子などの解析の際は、母集団の分布を仮定しないノンパラメトリックな検定法を選択するのが無難である。この方法は母集団が正規分布している必要がない反面、検出力はパラメトリックな手法に比べやや劣る。

免疫組織染色による有意差検定の仕方

免疫組織染色の判定が完了したら、それぞれの群(例えば陽性群と陰性群)と臨床病理学的因子との関係を解析する。癌細胞を染色したとすれば、腫瘍の深達度(T factor)、リンパ節転移の有無(N factor)、遠隔転移の有無(M factor)、病期(Stage)、腫瘍の分化度、脈管侵襲やリンパ管侵襲の有無などと、免疫組織染色の結果との間に有意差があるかどうか解析するのが一般的である。

次にさらに具体的に解説する。例えばある抗体による免疫組織染色で陽性、陰性に分けるとしよう。そこである要因によって群分けし(例えばリンパ節転移の有無など)陽性群と陰性群の間に有意差があるかどうかを調べるときは要因別に出現頻度を集計したクロス集計表を用いて検定する(表1)。この表は2つの要因が各個体において独立に出現しているかどうか(有意差の有無)を検定するためのものである。ここで注意しなくてはならないのは、どのセル(a, b, c, dの欄)も6以上の場合は全体数、要因別の出現率から各セルの理論値を推定し、実際の値とかけ離れているかどうか検定する、カイ2乗検定(chi-square test for independence)を用いて検定する。一方、どれかのセルに5以下のものが含まれる場合にはカイ2乗検定では誤差が大きくなるため、フィッシャーの直接確率計算法(Fisher's exact probability test)を用いて検定する。この方法は実際の集計表よりも極端に偏る確率を直接計算して検定する方法である。

免疫組織染色以外の手法による有意差検定の仕方

免疫組織染色法以外の手法で(例えば定量的RT-PCRなど)、ある要因によって2群に分け量(value)を比較する場合にはMann-Whitney U検定法を用いる。この検定法は2群間の中央値が等しいか否かの検定に用いられるノンパラメトリックな手法の中で最も使われる検定法である。ただし3群以上の中央値を比較する場合に、この検定法を2群ごとに繰り返しては

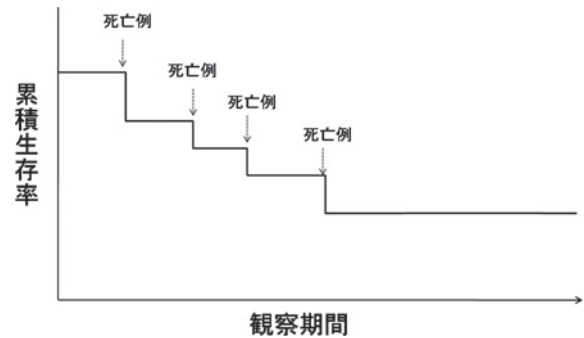


図1 累積生存率曲線

ならない。また、ある要因によって3群以上に分け量(value)を比較する場合にはKruskal-Wallis検定法を用いる。この検定法では、すべての群の中央値が等しいかどうか、つまり全体としての有意性を検出するのであり、個々の群間での有意差検定はできないため、多重比較と組み合わせて使用するのが望ましい。

Kaplan-Meier法による生存曲線

Kaplan-Meier法は、データを観察期間の順に並び替え、事象(死亡)が発生するたびに累積生存率を逐次計算する方法である。この解析方法では累積生存率が段階的に減少するため、累積生存率曲線は階段状のグラフで表示される(図1)。また、ある要因で群分けをして、それぞれの群における生存率を比較するときは、Log-rank検定やWilcoxon検定を用いることが多い。

おわりに

以上、臨床データ解析を中心に解説した。末節ながら本文が若手研究者の医学研究に少しでも役に立てれば幸いである。

文献

1. 吉田勝美監訳：一目でわかる医科統計学。第2版。2006; pp1-155, メディカル・サイエンス・インターナショナル 東京。
2. 三宅由子：臨床データのまとめかた。改訂第2版。2001; pp 1-162, 杏林書院 東京。
3. Byrne DW：国際誌にアクセプトされる医学論文 研究の質を高めるPowerの原則(木原正博, 木原雅子訳)。第1版, 2000; pp 1-233, メディカル・サイエンス・インターナショナル 東京。
4. 長田 理：StatView医学統計マニュアル。改訂第2版第5刷, 1999; pp 1-248, 真興交易医書出版部 東京。
5. 長田 理：StatView多変量解析入門。第1版第6刷, 2001; pp 1-133, オーエムエス出版 埼玉。

(受付：2010年5月28日)

(受理：2010年9月30日)