

シリーズ「医学統計処理の問題点」

医学統計処理の問題点

— 誤用のはなし I —

栗谷典量

(久留米大学医学部小児科)

はじめに

医学者の統計には誤りが多く信頼できるものは少ない、と統計学の専門家達からの評価は極めて手厳しい。ある経済新聞社発行の月刊誌に“統計手法が泣く医学論文の実際”という見出しで「誤解が乱用を生む統計の泥縄式利用」といかにもわびしいサブタイトル付きで厳しい批判記事が掲載されたことがある。

即ちほとんどの統計を利用した医学論文では、大なり小なりの何らかの誤りを指摘できると言うのである。統計の誤りを指摘する記事や書籍は少なくないが、事態の改善傾向は一向に伺えそうにない。

こうしたことから実情把握のため、某医学会雑誌1年分(Vol 94. 9~95. 8) 295論文について統計処理、手法のチェックを試みた。その結果からみると「多くの論文に誤りがある」とする月刊誌の指摘は妥当と判断せざるを得なかった。この現況を考えると医学者は統計学の専門家ではないのだから仕方ないこと、とばかりいってもおれない。記事は「根本的な解決には医学部の統計学の学習のあり方を見直す以外になさそうだ.」と結ばれていた。調査した295論文の誤用の内容を整理すると、

1. (Level I) 統計学の基礎を学んだ経験がないためにおこる統計技術以前の初歩の誤り。

2. (Level II) 独学のため、又は指導者(書)の誤り、プログラム解説書の誤りが原因で起こる誤り。(プログラムの欠陥(バグ)が原因の誤りを含む)

3. (Level III) 高水準の知識と技術の習得

を要する統計手法を学習不十分のまま実行する。

の3レベルに整理できよう。

以下、各種医学誌に見られる誤用の実態を観察してみよう。

1. Level I の問題。統計では運用、結果の記載に当たって、習慣、ルール、定義などの約束ごとが存在する。医学分野では、一般的にこの約束ごとに対して疎い人があり、この人達は約束ごとを簡単に無視する。代表的な例として有意差検定の結果の表現法があげられる。判定に対する表現法に統一性がなく勝手に統計的検定結果を無視(放棄)し、先入観に沿って身勝手な表現法を採る。「統計的には有意差はないが〇〇が優れている」と。科学的な表現では、使用する形容詞・副詞にはすべて厳格な定義に基づいて使用する。勝手に副詞を付加しない。厚生省の薬の副作用の発生率の表現法、天気予報の用語、株式市況の株価推移の表現法などいづれも一定の定義に基づいて使用されている。一方、ある医学者の著による統計プログラムのマニュアルの1節に有意差検定の結果、有意差なしとなったときの表現法として「統計学的に有意差は認められなかったが、一定の傾向が示唆された」と検定結果の無視を奨める記載があった。

検定結果の有意性マークとして常用されるアスタリスク(*)には世界的に統一された使い方があつた。演者が調査した某医学会雑誌の約20%の論文で基準的な使い方が無視されていた。ひどい例では[*]は検定の結果有意の場合用いるのが一般的なことはよく知られるところであるが、有意差なし [ns] の記号として用いる人

がある。統計用語の基礎知識の乱れも無視できない。例えば“母集団”の意味を総サンプル数(被験者群)と勘違いしている医学者は少なくない。この誤用例は有名指導者の著書・解説書などにさえ見られることがある。

2. Level II の問題. 統計の学習を始めるとき、最初に学ぶ基礎項目のひとつにデータの性質による分類法(尺度分類)がある。この分類法は統計を学ぶ上で非常に重要である。この分類法を理解してない人には統計を利用した確率論による論文を書くのは無理である。代表的尺度分類法は、①名義尺度(nominal scale)、②順序尺度(ordinal scale)、③間隔尺度(interval scale)、④比率尺度(ratio scale)の4種から成り立っている。(医学分野では、③間隔と④比率の両尺度をブールして計量値として扱うことも多い。)統計手法の選択は尺度の種類によって自動的に選定されるものだから、この分類法を無視した場合まともな解析結果が得られようはずがない。医学論文の基礎的誤用パターンで特に頻度が高いのは、a) 症例数とデータ数が一致しない。(登録データ数が症例によってまちまち) b) 順序データを計量値として処理する。(平均値、標準偏差を求める、t検定、相関分析を適用するなど)、c) データの対応の有無による分類(独立データ Unpaired, 対応データ Paired)の条件無視などが統計混乱の一因である。

3. 多変量解析などの誤りが Lever III に該当するが、高度のテクニックを用いるときは、それなりの理解者の協力と支えが必要で、これなしにはまともな論文は書けない。

以下誤用頻度が高いテーマから逐次取り上げながら考察していくことにする。

1. 症例数とデータ数のアンバランス。

医学論文、報告書などで症例数とデータ数が一致、または比例しない例は少なくない。臨床データ、特に外来患者を扱ったレポートなどで多くみられる。この場合、報告者は自分の処理に誤りがあることになかなか気付かない。その

過ちを指摘されてもすぐには納得しないのが普通である。この納得してもらえない人達に対する説明法を紹介する。「今から比処を通る人の体重を10例集めて下さい。」と言って体重計を持って街頭でデータ収集を始めさせる。そのうち、そこに284 kgの人(関取の小錦関)が通りかかったのでデータに組み入れた。ところが小錦関は何か忘れ物でもしたらしく急いで引き返して来た。そこで2回目のデータを登録した。その後、再度帰って来たとき3回目のデータを登録した。やがてデータ数が10になったので、平均値±SDの計算を行った。このデータは症例数8、データ数N=10のデータ群である。算出された統計量(平均値、SD)は全症例の情報が均等に提供されて得られた値ではなく、症例によってウェイトに違いがある。このようにウェイトが均一でない場合、あらかじめ、全症例のウェイトを均等にするための前処理を施した後、統計量を算出する必要がある。例えばデータが重複しているときは、①一回目のデータのみ採用する。②複数回の平均値を採用する。③中央値を採用する。などの対策法がある。

ある途上国に二重盲検法の実習指導に行った事がある。その国での話題である。

集められたデータをチェックしたところ、データ中に同一人らしい患者が入っていたので、確認のため問い正したところ、同一人であることが判明した。患者によって提供データ数が異なる例が一症例でも混入することは、致命的欠陥試験となり大変な損害になることを強調した。同一人を再度試験に組み入れた理由は一回めの治療が無効で再来した患者を二回めの登録をして例数稼ぎをねらったとのことである。無効例を繰り返し登録することは全体の有効率の低下を来す可能性のあること、群間比較であれば有意差なしに(P valueの上昇)近づくことを時間をかけて理解させた。日本でも症例の重複登録は1970年代初期頃の二重盲検試験で経験の浅い研究グループでたまに見られた現象である。

誤用例 1-1 市販統計ソフトのマニュアルにみる誤用例

ある学位論文の統計のチェックの際に発見された市販統計ソフトのマニュアルの誤用例である。即ち原稿に記載されている症例数とデータ数が一致せず、症例によりデータ数が異なっていたのでその点を指摘したところ、次のマニュアルを参考にしたのが原因だった事が分かった。指導書の印刷物にこうした基礎的な誤りが有ることは大変困った事である。

このマニュアルの問題点は、手術患者6人、データ数20の回帰分析を患者数20として処理していることである。著者は例数毎のデータ数の食い違いに全く関心がない。前に示した話での小錦のデータが2~15回組み込まれた状態である。

手術患者6人の散布図には20人分のプロットがあり、結果のまとめの表にも例数は20で表示されている。

例題12-1

手術患者6人を対象に直腸温と鼓膜温を同時に繰り返し測定したところ次のようなデータが得られた。鼓膜温から直腸温を推定する一次式を求めなさい。

<準備するデータ>

★列挙データ形式

鼓膜温(℃)	直腸温(℃)
37.1	37.6
35.1	35.8
35.5	36.4
34.9	35.4
35.4	36.0
37.0	37.3
35.0	35.7
35.4	36.3
35.2	35.9
35.8	36.5

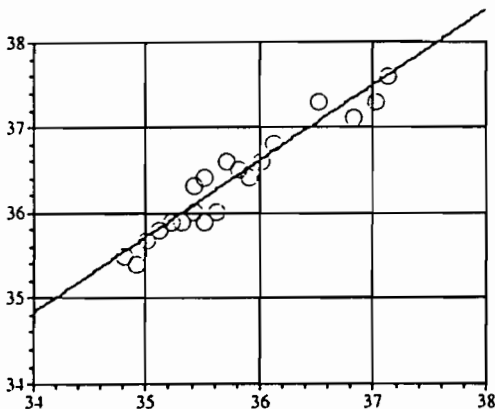
鼓膜温(℃)	直腸温(℃)
35.7	36.6
35.3	35.9
36.1	36.8
36.5	37.3
35.9	36.4
34.8	35.5
36.0	36.6
35.5	35.9
35.6	36.0
36.8	37.1

(第1列：説明変数，第2列：目的変数の設定)

<検定方法>

説明変数と目的変数の間に一次関数をあてはめて回帰関数を求める場合には、直線回帰分析(Simple regression)を用います。

求められた回帰関数のあてはまり具合については、回帰関数からの予測値と実測値の差を分散分析を行って検定します。Sta View 4.0では、回帰関数を求めると同時にあてはまり具合についても検定することができます。



回帰分析概要…データ数、欠損数、相関係数などが表示されます。

回帰分析概要

Rectal 対 Tympanic

例数	20
欠測値数	0
相関係数(R)	.963
R ² 乗	.927
自由度調整 R ² 乗	.923
RMS 残差	.174

誤用例 1-2

第1群の患者構成は症例数3, データ数5である. すなわち2件の重複症例が含まれている. 症例一覧表としては問題ないが, 検定に際し, 2群の群間比較をこのまま5例対5例で行うのは問題があり単純には行えない.

表1. 症 例

	No.	症 例	発症年齢	性別	診 断 名	ACTH 投与 開始年齢	ACTH 投 与 量	ACTH 投 与 方 法	臨床効果**
第 一 群	1	K.R.	3カ月	♀	West 症候群	4カ月	12.5 μg/kg	7週間で25回投与*	+++
	2	同 上	同 上	同上	同 上	7カ月	12.5 μg/kg	2週間連日, 3日後および1週間後各1回投与	+++
	3	I.M.	5カ月	♀	West 症候群	5カ月	12.5 μg/kg	7週間で25回投与*	+++
	4	F.Y.	5カ月	♂	West 症候群 Klinefelter 症候群 肺動脈弁狭窄症	6カ月	12.5 μg/kg	5日間連日, 隔日3回, 2日後および5日後各1回投与	+++
	5	同 上	同 上	同上	同 上	7カ月	12.5 μg/kg	隔日投与4回, 1投3休3回, 4日後, 1週および2週間後各1回投与	+++
第 二 群	6	T.S.	7カ月	♂	West 症候群	9カ月	25 μg/kg	7週間で25回投与*	+++
	7	K.H.	7カ月	♂	West 症候群	9カ月	25 μg/kg	7週間で25回投与*	+++
	8	F.S.	10カ月	♂	West 症候群	11カ月	25 μg/kg	連日投与9回, 隔日投与2回	+
	9	A.I.	1歳7カ月	♀	West 症候群	2歳8カ月	25 μg/kg	7週間で25回投与*	++
	10	E.M.	3歳2カ月	♂	Lennox 症候群	3歳5カ月	25 μg/kg	連日投与9回, 隔日投与6回, 2日後および1週間後各1回投与	-

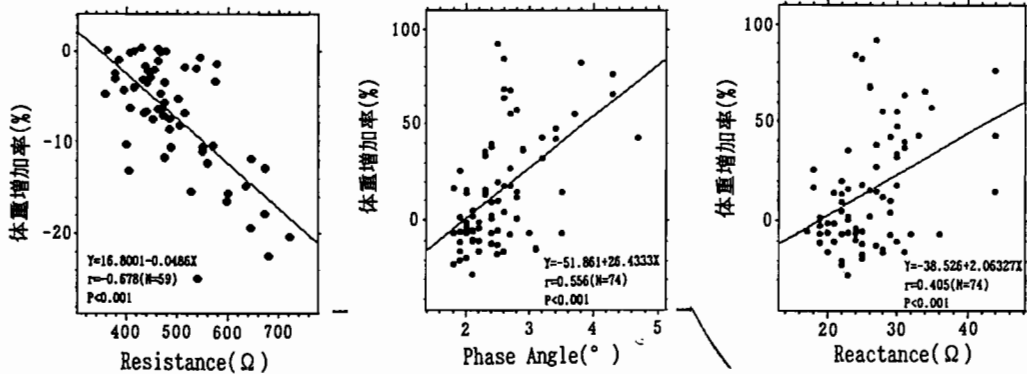
*: 福山方式による投与方法(本文参照)

** : 荻島による分類(+++: 発作の完全な消失, ++: 発作が1/3以下に減少, +: 発作が2/3以下に減少, -: 発作が不変または増悪)

誤用例 1-3

52名の児の生体電気と体重増加率の相関を検討した原著論文.

N=59, N=74, N=74 の散布図が描かれている.

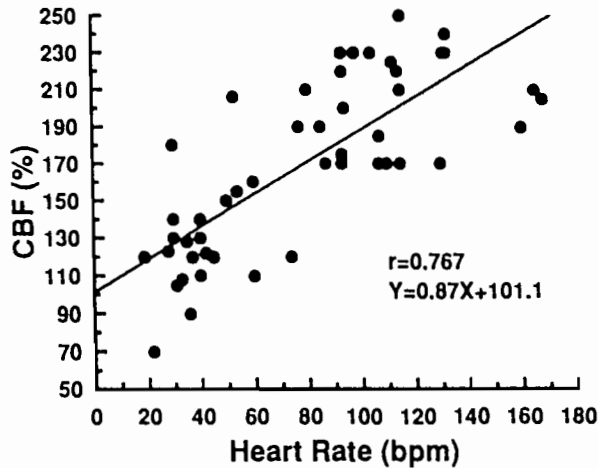


誤用例 1-4

仔猪8頭を対象にした原著論文 N=47 として回帰分析

方 法

成熟新生仔猪8頭(日齢0~1)を対象とし、ペントバルビタール麻酔下で2本の臍動脈にカテーテルを挿入し、1本のカテーテルを他方より深く挿入し、血圧、心拍数を連続的に測定した。他方はそれより浅く挿入し採血ルートとした。また、臍動脈より輸液を4 ml/kg/hrで行った。気管内挿管を行い臭化バンクロニウム(0.1~0.2 mg/kg/hr)を使用し自発呼吸を止め、人工呼吸器で呼吸管理を行った。リドカインにて局所麻酔後頭皮を剝離し、右側頭頂骨に直径10mm大

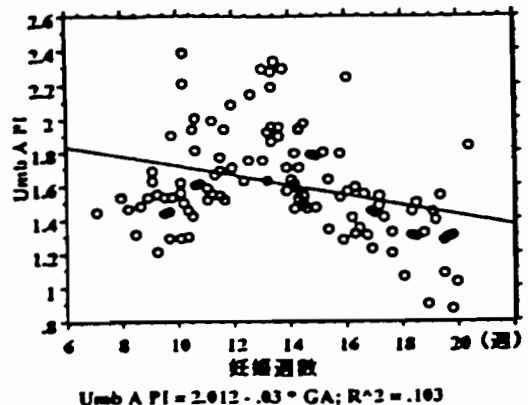
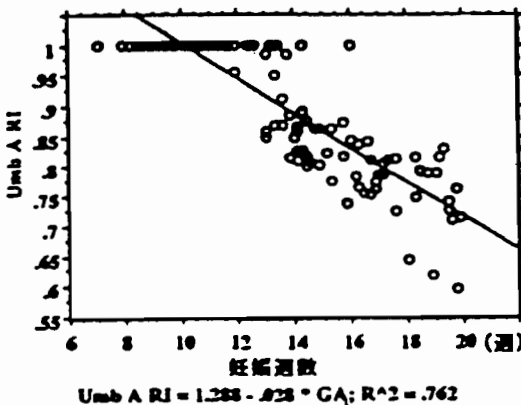


誤用例 1-5

106症例の臍帯動静脈血流波形を189回観察計測した。

【目的】 カラードプラー、パルスドプラーを使用し妊娠初期における臍帯動静脈波形の妊娠週数を伴う変化を明らかにする。

【対象および方法】 106症例の妊娠7週から19週における臍帯動静脈血流波形を経膈および経腹カラドプラーを用いて観察し(観察回数189回)、動脈血流については pulsatility index (PI) および resistance index (RI) を計測した。



誤用例 1-6

片肢・両肢の混在したデータ(鶴亀算型)の延べ上肢数の平均値±SDを用いた解析.

表2. 正常対象のSSEP

	N_1 - P_3 潜時 (msec)	N_1 持続時間 (d(msec))	N_1 振 幅 (a(μ V))	分 散 比 (a/d)
0~5カ月 (8名:n=15)	12.0±1.3	15.4±2.4	2.21±1.29	0.145±0.079
6~11ヶ月 (5名:n=8)	9.8±0.5	15.8±3.9	2.87±1.67	0.196±0.124
1~2歳 (9名:n=15)	9.6±0.8	14.0±2.7	1.58±0.81	0.112±0.060
3~4歳 (6名:n=12)	7.9±0.5	12.9±3.2	2.27±1.75	0.217±0.126
5~10歳 (17名:n=31)	6.0±0.4	10.1±2.5	2.90±1.63	0.301±0.191
11~16歳 (11名:n=18)	5.6±0.4	8.2±3.1	2.92±1.22	0.381±0.189
22~30歳 (11名:n=21)	5.9±0.7	9.0±2.4	1.99±0.78	0.240±0.123

持続時間(d)は N_1 陰性波が出現し陽性にふれるまでの時間, 振幅(a)は N_1 の基線からの振幅とし, 分散比はそれらの比(a/d)とした. 各値とも平均値±1SDで表している. 各年齢ごとの対象人数は()に示し, 両上肢をあわせた数をnで表した.

われわれは, 日常, 実例に示したような誤用例に遭遇する. データの処理に際し何げなく統計量として平均値, 標準偏差を用いているが, 平均値の使えない条件のあることを改めて認識しておく必要のあることを述べた. 今回は症例数とデータ数のアンバランスの問題をテーマとして取り上げた. 次回は, 知っておく必要のある平均値の使えない条件を二つ紹介する. 一つは, 順序尺度(著効, 有効, 無効, 悪化など)をスコアと称して平均値を求める例. 一つは歪みの大きい分布のデータ(白血球数, GOT, 貯蓄額など)に対する処理. これらには平均値を使った処理(統計的検定など)は不適当な条件であるにもかかわらず医学分野では多用される誤った用法である.