医学における統計学の応用について

—(第2編) 平均値の差の検定を中心として—

新潟大学医学部衛生学教室（主任：山本正治教授）
遠 藤 和 男

An Application of Statistical Methods to Medical Science (Part II) on Tests for Two Means

Kazuo ENDOH

Department of Hygiene and Preventive Medicine,
Niigata University School of Medicine
(Director: Prof. Masaharu YAMAMOTO)

When the researchers evaluate the results of experiment by using statistical methods, the most important task for them is to understand whether or not their data have continuity. Biomedical data are roughly classified into two types; one is continuous variate (metric data), and the other is discrete one (categorical data), Depending on the type of data, the different statistical methods should be used.

The second important task is to understand whether their data have normality or not. When their data are not thought to follow normal distribution, nonparametric methods such as Wilcoxon's rank sum test should be applied.

Since the author has already reported how to deal with discrete variates elsewhere[4], the statistical methods to evaluate the continous variates are discribed in this paper. The present article is chiefly stressed on the tests of two means, because they have been widely used in the medical field.

Key words: tests of normality, paired t-test, Student's t-test, Welch's t-test, Wilcoxon's rank sum test
正規性の検定，対応のある t 検定，スチューデントの t 検定，ウェルチの t 検定，ウィルコクスンの順位和検定

Reprint requests to: Kazuo ENDOH, Department of Hygiene and Preventive Medicine, Niigata University School of Medicine, Niigata City, 951, JAPAN.

別刷請求先: 〒951 新潟市旭町通1番町
新潟大学医学部衛生学教室 遠藤和男

はじめに

研究者が統計学的方法を用いて，自分の実験成績を評価しようとする前に，最も重要なことは，自分のデータが連続性を持つかどうか，正しく認識することである．**図 1** に示すように医学的データは，定量的データ（連続変量）と定性的データ（離散変量）とに大きく分かれている．

定量的データの多くは，測定機器を用いて測定されたデータである．機器の精度さえ高ければ，有効桁数は増えていくので，連続性があると考えられる．これに対して定性的データでは，数値に変換されていても，その中間の値がとれないか，またはとれても意味がない．データが連続的か離散的かによって，用いる統計学的手法も異なるため，まず，データの型に対して十分に注意を払う必要がある．

次に重要なことは，データに正規性が認められるかどうか，正しく認識することである．**表 1** に正規性の検定方法を列記したが，どれも決め手にはならない．また，自然界に本当に正規分布に従うデータが存在するのかどうかについても議論[1]がある．

しかし，データが正規分布に従うと考えられる場合，正規分布やt分布等の分布のパラメーターを用いて検定でき，検出力[2]も高いと言われる．一方，正規分布に従わない場合は，Wilcoxon[3]の順位和検定等ノン・パラメトリック技法を用いなくてはならない．

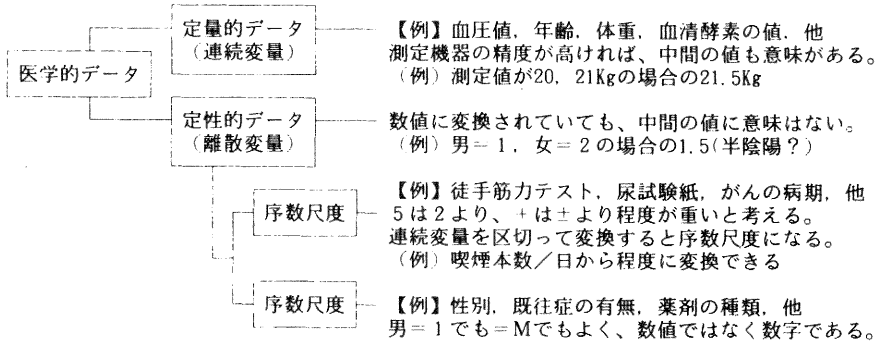


図 1　医学的データの区分

表 1　正規性の検定方法

（1）	ヒストグラムを描いて，左右の対称性，平均値付近の最頻性を確認する． ……直観的ではあるが，案外当たっている場合が多い． ただし，区間の数，幅によって印象が異なることもある．
（2）	累積相対度数について，正規確率紙で直線性を確認する． ……検定方法もあるが，教科書では直観的な判断でよいとされる．
（3）	ヒストグラムの各区間の中央値について，正規確率から χ^2 検定を行う． ……計算は煩雑だが，例数の多い場合によく用いられる．
（4）	歪み度（skewness）及び尖り度（kurtosis）を求めて検定する． ……3乗，4乗の計算が必要で，コンピュータでないと無理である．
（5）	例数が30未満の場合，標準化偏差（standardized deviate）を算出し， 1標本における適合度検定（Kolmogorov-Smirnov の検定）を行う． ……ノン・パラメトリック技法の一つで，標準正規分布への適合度をみる．
（6）	中心極限定理（central limit theorem）では，標本数が大きければ， 母集団が正規分布でなくとも，標本平均値は正規分布に従うと考えられる． ……データにランダム性が認められないと，成立しない． 全国成績が正規分布に従えば，実験データも正規分布と考えてよい．

第1編では，2×2表を中心とした離散変量の取扱いについて述べた[4]．本編では，様々な統計学的方法の中で最も使用頻度の高いといわれる[5]，2群の平均値の差の検定を中心として，連続変量の評価について述べることにする．医学的分野でも応用範囲が広いと考えられるので，若手の研究者の参考となるように若干の具体例を示して解説を加えた．

1. 飛び離れた値の棄却検定法（Grubbs-Smirnov の棄却検定）[6]

表 2 に例示したデータが得られたと仮定する．誰がみても，最後の8.3は他のデータから飛び離れているように思われる．あるいは転記ミスかも知れないが，今となっては確かめる方法がないとする．

表 2　飛び離れた値の棄却検定法

【例】　1.6, 1.9, 2.0, 2.1, 2.1, 2.4, 2.4, 2.5, 2.5, 2.6, 2.8, 2.9, 3.0, 3.5, 3.5, 8.3

$n=16, x=2.86, SD_{n-1}=u=1.533$

$$T_0=\frac{X_{max}-\bar{X}}{u}=3.549$$

$$>T_{16}\left(\frac{0.01}{2}\right)=2.747$$

したがって，8.3は飛び離れた値（$p<0.01$）として棄却できる．

このような場合，データ数が少ないからといって，全データを採用することは，かえって危険である．**表 2** に示した棄却検定法[6]を用いることによって，8.3を飛び離れた値として棄却できる．ただし，本来対数正規分布に従うようなデータについては，この方法を用いると誤りを犯す危険もある[7]ので，十分注意が必要である．

2. 標準正規分布への適合度検定（Kolmogorov-Smirnov の検定）[8]

標本数が30以上のデータについては，**表 1** (3) χ^2 の適合度の検定等を用いることができる．標本数が30未満の場合は，一変量の適合度検定である Kolmogorov-Smirnov 検定[8]を応用して，標準正規分布への適合度を検定すればよい．

それにはまず，標準化偏差［standardized deviate：$(x_i-\bar{x})/u$］を求めておくことが必要である．**表 2** のデータについて，8.3を棄却する前と後の検定結果を**表 3** に示した．また，正規確率紙に各データをプロットして，**図 2** 及び**図 3** に示した．

3. 対数正規分布

各データの対数をとると，正規分布に変換できる分布を対数正規分布（log-normal distribution）という．**表 4** に示すように，医学的なデータは，正規分布に従うものよりも，対数正規分布に従うものの方が多い．特に，数値の高い方が異常と判断される，血清酵素値はほとん

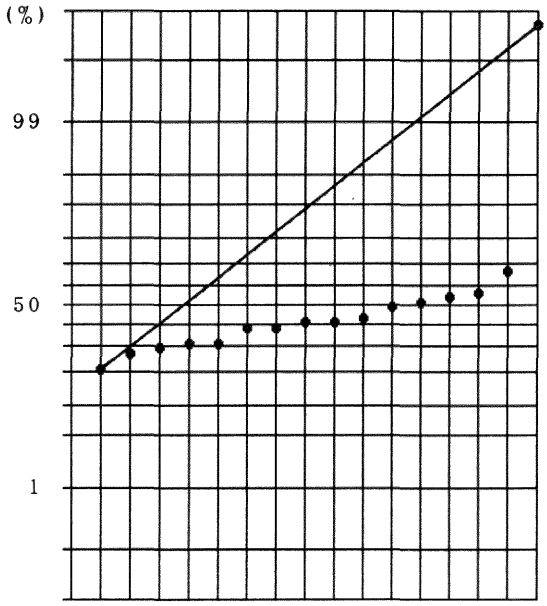


図 2　表2のデータの正規確率紙へのプロット

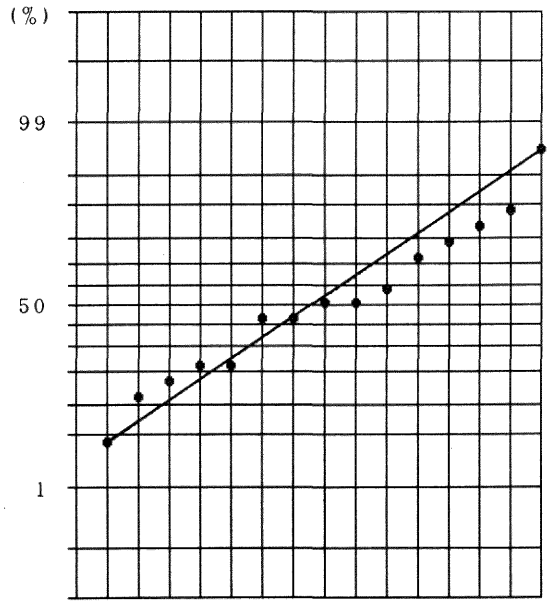


図 3　8.3 を棄却した後の正規確率紙へのプロット

表 3　標準正規分布への適合度の検定

（1）　表2のデータに Kolmgorov-Smirnov の検定を行うと，

番　号	観察値	標準化偏差	正規確率 Ux	理論値との差
(1)	1.6	−0.81961	0.206219	0.206219
(2)	1.9	−0.62388	0.266353	0.203853
(3)	2.0	−0.55864	0.288205	0.163205
(4)	2.1	−0.49340	0.310866	0.123366
(5)	2.1	−0.49340	0.310866	0.060866
(6)	2.4	−0.29767	0.382978	0.070478
(7)	2.4	−0.29767	0.382978	0.054522
(8)	2.5	−0.23243	0.408104	0.091896
(9)	2.5	−0.23243	0.408104	0.154396
(10)	2.6	−0.16718	0.433614	0.191386
(11)	2.8	−0.03670	0.485363	0.202137
(12)	2.9	0.02854	0.511385	0.238615
(13)	3.0	0.09379	0.537360	0.275140
(14)	3.1	0.15903	0.563177	0.311823
(15)	3.5	0.42000	0.662756	0.274744
(16)	8.3	3.55165	0.999808	0.062309

差の最大値＝0.311823＞5％点＝0.29472
したがって，$p<0.05$で正規性が棄却される．

（2）　表2のデータから8.3を棄却して再度検定すると，
差の最大値＝0.113431＜5％点＝0.30397
棄却できないので，正規分布に従うと考えてよい．

表 4　血液データの分布型

種　目	正規分布型	対数正規分布型
無機成分	Ca, Cl, 無機リン	Na, K
窒素化合物	総蛋白質, 尿酸, アルブミン	グロブリン, TTT, ZTT, 尿素, ビリルビン
糖・脂質	ブドウ糖	総コレステロール, リン脂質
酵　素	コリンエステラーゼ	GOT, GPT, LDH, アルカリフォスファターゼ
血　算	赤血球	白血球

*なお，Fe, Hb, Ht は指数正規分布型に近い．

表 5　血清総コレステロールの度数分布（仮想例）

区間(mg/dl)	人数	相対度数	累積相対度数
100 ～ 119	13	13.0	13.0
120 ～ 139	18	18.0	31.0
140 ～ 159	33	33.0	64.0
160 ～ 179	15	15.0	79.0
180 ～ 199	8	8.0	87.0
200 ～ 219	7	7.0	94.0
220 ～ 239	4	4.0	98.0
240 ～ 259	2	2.0	100.0
計	100	100.0	

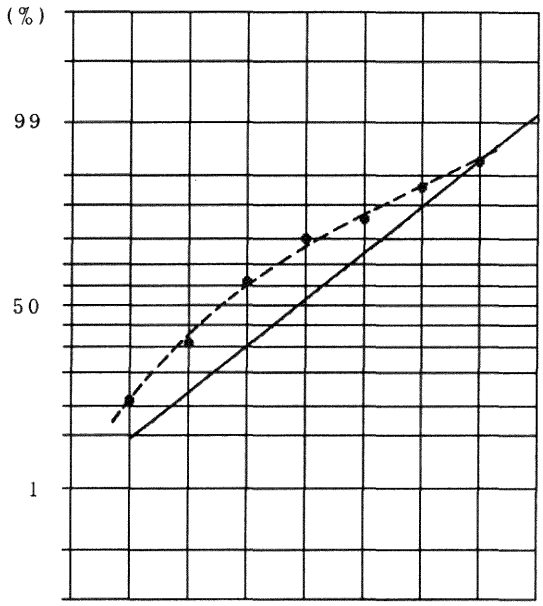


図 4　表5のデータの正規確率紙へのプロット
理論分布（直線）よりも曲線（点線）に近い

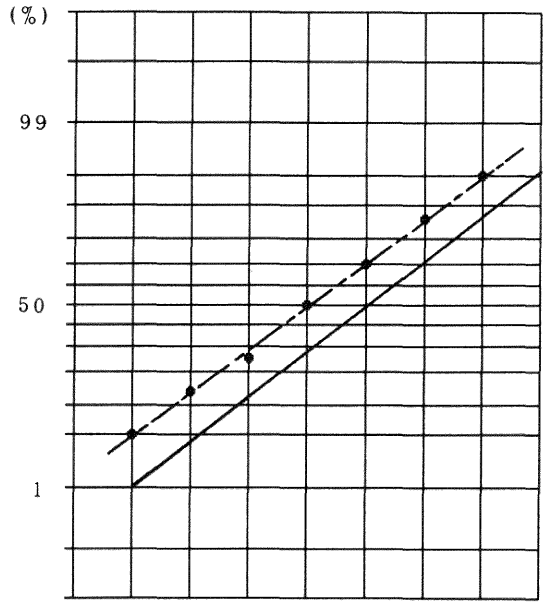


図 5　対数に変換した後の正規確率紙へのプロット
一点鎖線が直線性を示し，正規分布に従うと考えてよい．

ど対数正規分布に従うことが知られている[9)10)]．

表 5 のデータを正規確率紙にプロットすると，図 4 に示したように，曲線となってしまう．対数変換してからプロットすると，図 5 のようにほぼ直線となり，正規性を持つと考えてよい．自分のデータが正規分布に従うのか，あるいは対数正規分布に従うのかは，文献等を参照して検定の前に十分検討されなければならない．

4.　標準偏差と標準誤差について

統計学的検定では，母集団を仮定して検定を行うため，表 6 (1) 標本標準偏差（SD_n）ではなく，(2) 標準偏差の不偏推定量（SD_{n-1}，以下，不偏偏差という）を用

表 6　標準偏差と標準誤差の比較

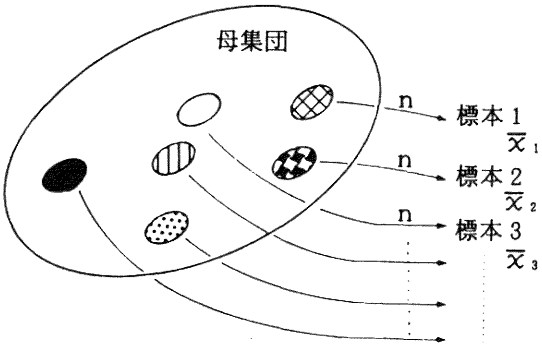
(1)　標本標準偏差　$s=\sqrt{\dfrac{\Sigma(x_i-\bar{x})^2}{n}}$

(2)　不偏標準偏差　$u=\sqrt{\dfrac{\Sigma(x_i-\bar{x})^2}{n-1}}$

(3)　標準誤差　$SE=\dfrac{s}{\sqrt{n}}=\dfrac{\sqrt{\Sigma(x_i-\bar{x})^2}}{n}$

(3)′　$SE=\dfrac{s}{\sqrt{n-1}}=\dfrac{u}{\sqrt{n}}=\sqrt{\dfrac{\Sigma(x_i-\bar{x})^2}{n(n-1)}}$

※u の代わりに誤って SE を用いると，本来検出できない差を「有意」と判断する。



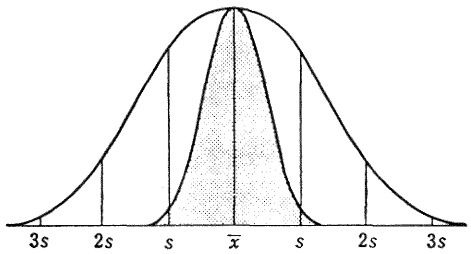


図 6　標準偏差と標準誤差の関係
下方白い部分は標本の分布で，網の部分は標本平均値 $\bar{x}$ の分布である．

いることが多い．(3) 標準誤差（SE：Standard Error）は (1) と式の形が似ているためか，標準偏差の代わりに誤って標準誤差を用いた例が知られている[11]．(3)' の式を用いれば，誤用は少なくなると考えられる．

図 6 に両者の関係を示したが，標準誤差の方が標準偏差より小さい．標準誤差は標本 x のバラツキの指標ではなく，標本平均値 $\bar{x}$ のバラツキの指標である．何回か同じ数の実験を繰り返さない限り，標準誤差は余り意味を持たない．不適当な使用は避けるべきである[12]．

5. 対応のあるデータと対応のないデータ

二群の標本平均値の差の検定には，種々の方法が知られている．**図 7** に示すように，まず第1に，対応のある2標本と対応のない2標本に大別される．対応のあるデータでは，対応のある t 検定（paired t-test）[13] を用いる．ただし，ペアといっても，主として同一人の時間的な変化をみる検定であり，例えば，両眼，両耳，夫婦等に対応のある t 検定を用いるべきではない．

表 7 のような対応のあるデータに対して，誤って対

表 7　降圧剤投与による最大血圧値 (mmHg) の変化

No.	投　与　前	投　与　後
1	174	178
2	166	166
3	190	192
4	170	172
5	204	184
6	184	162
7	182	174
8	182	162
9	194	182
10	198	188
11	196	178

表 8　対応のある t 検定とStudent の t 検定の比較

表7　のデーターについて

（1）　対応のある t 検定を用いると，

	No.	平均値	S.D.(n−1)
A群－前	11	185.4550	12.1685
A群－後	11	176.1820	10.0976
差(A−B)		9.2727	9.9709

前後の差の検定を行います

$t_0=3.0844$　　〈 d.f.＝10 〉

(p＝0.011554)→詳しくは統計表で

p＜0.05で有意差が認められる．

（2）　対応のない2群として，　Student の t 検定を用いると，

	No.	平均値	S.D.(n−1)
A群	11	185.4550	12.1685
B群	11	176.1820	10.0976
差(A−B)		9.2727	11.1811

分散が等しいと考えて t 検定を行います

$t_0=1.9443$　　〈 d.f.＝20 〉

(p＝0.065980)→詳しくは統計表で

（1）と異なり，有意差なしとなってしまう．

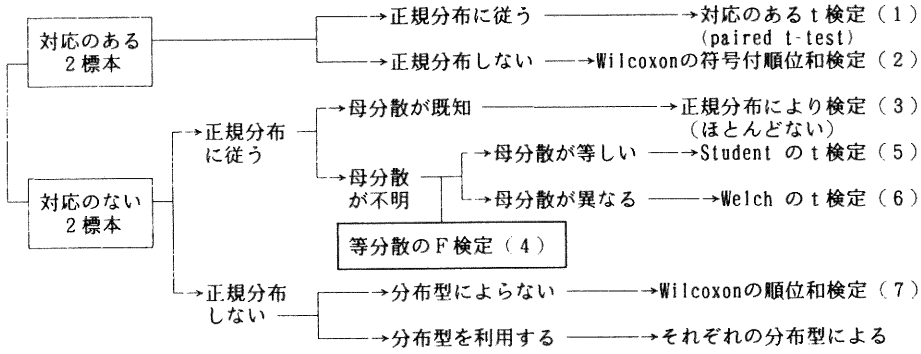


図 7　二つの平均値の差の検定

応のない検定法（ここでは，Student のｔ検定法[14]）を用いると，検出できる有意の差を検出できなくなってしまう．(表 8).

6. 分散の等しいデータと分散の異なるデータ

正規分布に従うと考えられる場合は，第2に，分散が等しいか異なるかを検討する必要がある．この検定法は等分散のＦ検定[15]と呼ばれる（図 7 (4)).

分散が等しいと考えられる場合に，Student のｔ検定法[14]を用いることはよく知られている．しかし，分散が異なると考えられる場合に，直ちに Welch のｔ検定法[16]に飛びつくことは，早計に過ぎる．**1. 飛び離れた値の棄却検定法**[7]を用いると，分散が等しくなる場合がある．もともと比較しても意味のないものを比較している場合もあり，十分検討してから実施すべきである．

また，既存のソフトの中には，等分散のＦ検定[15]を実施せずに，いきなり Welch のｔ検定法を行うプログラムもある．２群の平均値及び不偏偏差が全く同じで，標本数だけを入れ換えて Student 及び Welch の両法を用いた結果を**表 9** に示した．等分散のＦ検定[15]を実施しないと，微妙なところで誤りを犯すことになる．

7. 正規分布に従うデータと従わないデータ

２群の標本数が30未満と少なく，**2. 標準正規分布の適合度の検定**[8]でも，正規分布が仮定できないデータに対しては，Wilcoxon の順位和検定法[2]が用いられる．

この場合，誤って正規分布を仮定して Student または Welch のｔ検定法を用いても，結果が等しいことも

表 9　Student と Welchの ｔ 検定の比較

下の(１)と(２)は標本数を入れ替えただけで，平均値及び不偏偏差はそれぞれ等しい．

(１)

	No.	平均値	不偏偏差
A群	32	150.000	31.340
B群	16	132.000	26.110
合計(A＋B)	48	18.000	29.736

等分散性があるので

Studen で，　$t_0=1.977$ (df=46)　$p=0.054053$

Welch の ｔ 検定では，　$t_0=2.1024$(df=35.5)　$p=0.045095$

(２)ここで n_1 とn_2 だけを入れ替えると，

	No.	平均値	不偏偏差
A群	16	150.000	31.340
B群	32	132.000	26.110
合計(A＋B)	48	18.000	27.923

等分散性があるので

Student で，　$t_0=2.1053$ (df=46)　$p=0.040753$

Welch の ｔ 検定では，　$t_0=1.9794$ (df=25.7)　$p=0.062279$

※(１)と全く逆の結果を得てしまうことに注意.

表 10　B-cell の割合(％)

透析患者		健常者
10	14	7
13	6	8
7	11	8
16	12	6
12	7	7
12	9	11
10	12	10
10	9	8
4		9
(17)		

表 11　Wilcoxon と Welch の比較

【例１】 表10のデータをそのまま使用する．

（１） Wilcoxon の順位和検定

（２） Welch の ｔ 検定

のどちらでも $p<0.05$と同じ結果となる．

【例２】 表10の透析患者中(17)の値を除外する．

（１） Wilcoxon の順位和検定

数の少ない健常者の順位和を U とすると，

$U=5.5+9+9+2.5+5.5+18.5+15.5+9+12=86.5$

＞下側2.5％点＝84

したがって，５％では有意差なし

（２） 分散が異なるので Welch の ｔ 検定では，

$t_0=2.224>t_{29}(0.05)=2.045$

したがって，５％で有意差ありで，両者の検定結果は異なってしまう．

ある．しかし，**表10** の例[17]から1つのデータを除くだけで，決定結果が異なる場合もある（**表11**）．したがって，データに正規性があるかどうかについては，**表 1** に示した方法を用いて，事前に十分検討しておく必要がある．

おわりに

第450回新潟医学会の特別講演のうち，連続量の取扱いについて，正規性の検定及び2群の平均値の差の検定を中心にまとめた．本稿で述べた検定方法については，N88Basic で開発したプログラムを利用できる．

稿を終えるにあたり，医統計学について御教授いただいた上村桂博士に深く感謝申し上げます．また，日頃ご指導いただき，当日も司会の労をお取りいただいた山本正治教授に深甚なる謝意を表します．

参考文献

1) **Krishnaiah, P.R., Kanal, L.N., ed.:** Handbook of Statistics, Vol. 1, pp.279～320, North-Holland Pub. Co. (Amsterdom), 1980.
2) **Wilcoxon, F.:** Individual comparisons by ranking methods. Biometrics Bull., **1:** 80～83, 1945.
3) **橘　敏明:** 医学・教育学・心理学にみられる統計的検定の誤用と弊害．pp．50～57，医療図書出版社（東京），1986.
4) **遠藤和男:** 医学における統計学の応用について（第1編）2×2表を中心として．新潟医誌，**102:** 147～154．1988.
5) **丹後敏郎:** 医学で要求される統計計算．応用統計学，**9:** 141～158，1981.
6) **上村　桂:** 医統計学．pp．79～81，文永堂（東京），1979.
7) **丹後敏郎:** 医学への統計学．pp．55～58，朝倉書店（東京），1985.
8) **統計数値表編集委員会，編:** 簡約統計数値表．pp．91，167～168，日本規格協会（東京），1985.
9) **北村元仕:** 正常値とは何か．日本臨床（秋季増刊号），**34:** 3～13，1976.
10) **丹後敏郎:** 正規分布か対数正規分布か．臨床病理，**27:** 1151～1155，1979.
11) **臼井敏明，訳:** 臨床検査統計学．p．233，医学書院（東京），1987.
12) **Feinstein, A.R.:** Clinical biostasistics XXXVII. Clin. Pharmacol. Ther., **20:** 617～631, 1976.
13) **上村　桂:** 医統計学．pp．69～72，文永堂（東京），1979.
14) **Student (Gosset, W.S.):** The probable error of a mean. Biometrika, **6:** 1～25, 1908.
15) **Fisher, R.A.:** Statistical methods for research workers, Oliver and Boyd (Edinburgh), 1934.
16) **Welch, B.L.:** On the comparison of several mean values: an alternative approach. Biometrika, **38:** 338～336, 1951.
17) **丹後敏郎:** 医学への統計学．pp．71～73，朝倉書店（東京），1985.