

平成 11 年 10 月 21 日(木)

「学会・論文発表に必要な統計の知識」

大阪府立母子保健総合医療センター 検査科 宮野 章
学術部 血清検査部会、情報システム部会、卒後教養部会

1. 統計ソフト

以前の紙による統計から、現在はパソコンによる統計に変わりつつあります。

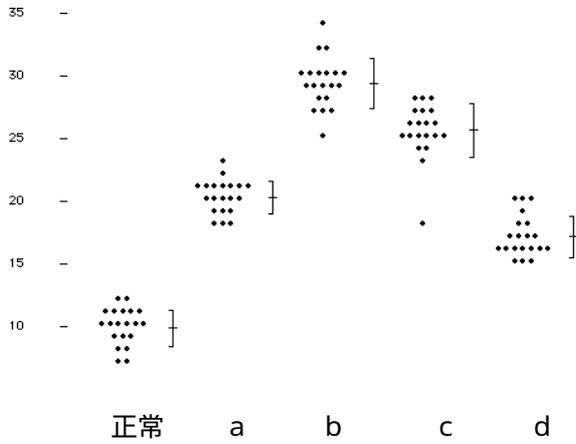
統計ソフトを使いこなす事が、重要になって来ました。

- Group Scatter(シェアウェアソフト)
NIFTY-Serve のバイオフォーラムで入手可能なマッキントッシュ用の群散布図をかけるソフト。HyperCard (マッキントッシュに標準添付されている) がある。(図1)
- StatFlex Plus(View Flex 社)
国産の市販汎用統計ソフトであるが、ROC 曲線の作成機能を持っている。DOS や WINDOWS 用。(図1 の様な事もできる)
- StatView(株式会社ヒューリンクス)
データ解析とプレゼンテーション統合化システム。マッキントッシュや WINDOWS 用。欠損値は自動的に統計から省かれる。(図2)
- SPSS(南江堂より購入可)
データ解析システム。マッキントッシュや WINDOWS 用。
- Excel(Microsoft)
表集計ソフトであるが、関数等を使いこなせばかなりの統計ができる。マッキントッシュや WINDOWS 用。
- Lotus 123(Lotus)
表集計ソフトであるが、関数等を使いこなせばかなりの統計ができるが、Excel に押され気味。WINDOWS 用。
- DeltaGraph (パラロイド)
グラフ作成ソフトであるが、簡単な統計はできる。マッキントッシュや WINDOWS 用。軸ブレイクを作れるので便利。(図3)

2. 分布を示す

1) 群散布図

図 1

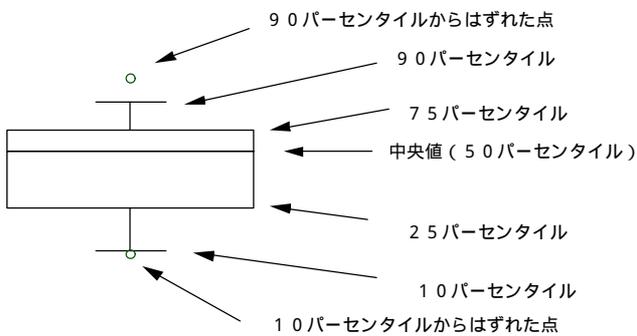


2) 箱ひげ図(Box Plots)

このグラフはデータが正規分布していない場合にも分布形態が把握しやすいため、実験から得られたデータをまずこのグラフで表示してデータの傾向をつかむものに向く。また、データが正規分布しないデータ(ノンパラメトリック・データ)の場合には、 $\text{平均値} \pm \text{標準偏差}$ という表示でなく、箱ひげ図を用いて表示する。

[箱ひげ図の見方]

図 2



3. 用語集

平均値(Mean)

標準偏差(Standard deviation : S.D.): 「データがどのくらいばらついているか」を表す指標。 データの分布状況を知る目安として有用。

標準誤差(Standard error : S.E.): 「データの平均値がどのあたりにあるのか」を表す指標。あくまで平均値がこの範囲にあることを示すだけであり、個々のデータの分布については一切情報が含まれない。

データ数(Count)

最小値、最大値(Minimum, Maximum)

データ欠損数(Number missing)

水準数(Number of levels)

分散(Variance)

変動係数(Coefficient of variation)

範囲(Range)

合計(Sum)

偏差平方和(Sum of square)

相乗平均、調和平均(Geometric and harmonic mean)

歪度(Skewness)

尖度(Kurtosis)

中央値(Median)

四分位数間範囲(Interquartile range : IQR)

最頻値(Mode)

10%調整平均(10% Trimmed mean)

中央絶対偏差(Median absolute deviation : MAD)

4. 有意差を求める

パラメトリック法とノンパラメトリック法

	パラメトリック法	ノンパラメトリック法
尺度水準	間隔・比例尺度	名義・順序尺度
対象の母数	連続量、定量値 平均値、分散、標準偏差 Pearson's 積率相関係数	離散量半定量値、定性値 中央値など、散布度 Sperman's 順位相関係数
母集団分布型 誤差分布	正規分布 正規分布に関する手法	問わない 分布に関わらない手法
標本サイズ 計算量	小さ過ぎは、いけない 多い	問わない 小さい
必要数表	Z, t, F表など 一般的によく利用される (t, F検定など)	特殊な数表が多く、 小標本に限定される。 (マンウィットニーのU検定など)

1) 2群比較

正規性が成立する場合、Student t 検定を選択し。成立しない場合は Wilcoxon 順位和検定 (Mann-Whitney 検定) を実施する。

おおよその判断基準としては、歪度 (skewness) の絶対値が 0.5 を越えておらず、尖度 (kurtosis) が 2.5 ~ 3.5 の間に入っていれば正規分布とみなしてよい。

2) 多群比較

質問: 図 1 の場合、Student t 検定を行い、正常群とそれ以外の 4 群についてそれぞれ有意差をもとめて報告した。これは、正しいだろうか？

答え: 否。2 群の比較を多群についても同じように繰り返してはならない。検定を幾度となく繰り返すばあい、設定した有意水準を低くしなければならないから。この事を“検定の多重性”と言う。

StatView で利用できる多重比較法

多重比較法	制限条件		検出力
	サンプル数	分布	
Fisher's PLSD	各群の数等しい	正規分布等分散	優れる
Scheffe's	なし	なし	劣る
Bonferroni/Dunn	各群の数等しい	正規分布等分散	比較的優れる

アドバイス: Scheffe's を使えば、サンプル数も分布制限もない。

5. 診断テストの特性

診断テストを行った場合の結果は、通常、陽性(T^+)と出るか陰性(T^-)のいずれかである。しかし、その結果が正しく疾患を診断(またはチェック)しているか否かによって4つのカテゴリーに分けられる。検査の精度やそれを用いた場合の有効性を示す指標となる。

		疾 病	
		有(D)	無(noD)
検 査 結 果	陽性 (T^+)	TP (true positive)	FP (false positive)
	陰性 (T^-)	FN (false negative)	TN (true negative)

a	b
c	d

TP : 真陽性
 FN : 偽陰性
 FP : 偽陽性
 TN : 真陰性

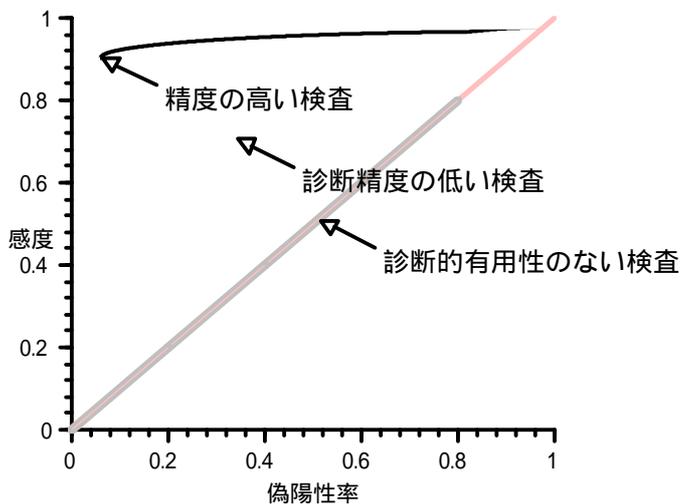
診断テストの精度と有効性を示す用語

用語	定義	計算法
感度 (sensitivity)	$\frac{T^+}{D}$	$\frac{a}{a + c}$
特異度 (specificity)	$\frac{T^-}{noD}$	$\frac{d}{b + d}$
偽陽性率 (false positive rate)	$\frac{T^+}{noD}$	$\frac{b}{b + d} = 1 - \text{specificity}$
偽陰性率 (false negative rate)	$\frac{T^-}{D}$	$\frac{c}{a + c} = 1 - \text{sensitivity}$
陽性反応適中度 (positive predictive value, PPV)	$\frac{D}{T^+}$	$\frac{a}{a + b}$
陰性反応適中度 (negative predictive value, NPV)	$\frac{noD}{T^-}$	$\frac{d}{c + d}$
有病率 (prevalence)	$\frac{D}{D + noD}$	$\frac{a + c}{a + b + c + d}$
陽性率 (test positive rate)	$\frac{T^+}{T^+ + T^-}$	$\frac{a + b}{a + b + c + d}$
有効度 (efficiency)		$\frac{a + d}{a + b + c + d}$

6 . ROC 曲線(receiver operating characteristic curve : 受信者操作特性曲線)

ROC 曲線は もともと 1950 年代にレーダーの性能評価を目的として考案された。医学への応用は、画像診断領域で先行したが、1980 年代中ごろから臨床検査領域でも広く利用され、診断精度評価の標準的手法として採用される様になった。

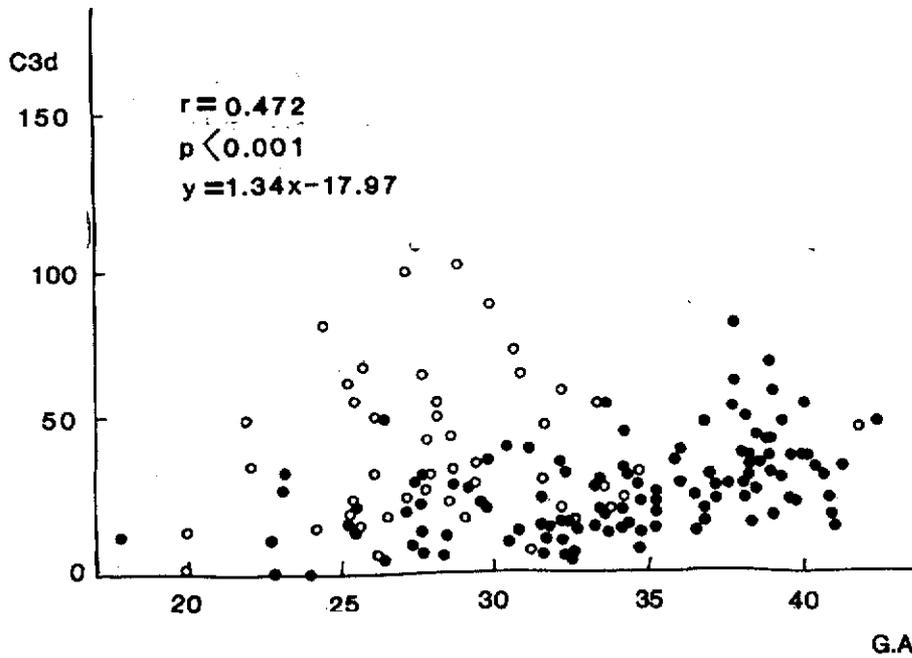
図 3



7. レフリーからの指摘と訂正後の具体例

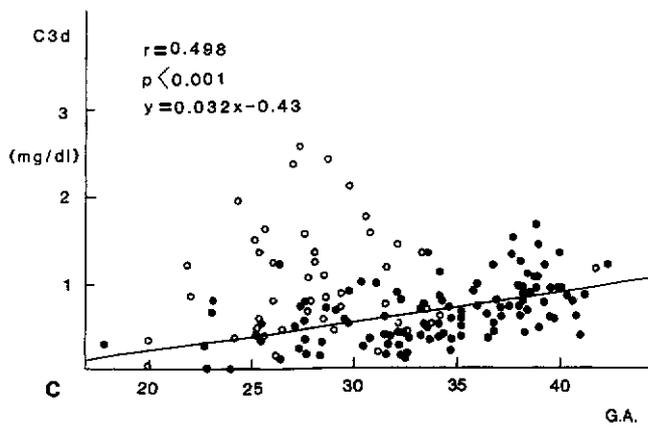
第一例

訂正前



レフリーからの指摘 : It would make the figures clearer if the regression lines were drawn.

訂正後



第二例

訂正前

Table 1. The Complement and Protein Levels in Cases Without and with CAM, and in Negative and Positive of Bacterial Culture

	CAM				Bacterial Culture					
	Without (n=56)	n	With (n=30)	n	P value	Negative (n=37)	n	Positive (n=49)	n	P value
C3d (mU/l)	23.2± 11.4	56	42.3± 28.1	30	<0.001	26.4± 12.1	37	32.5± 25.5	49	N.S.
C2 (%)	83.1± 26.5	49	67.5± 25.5	23	<0.05	80.9± 26.0	34	75.6± 28.0	38	N.S.
C3 (mg/dl)	40.6± 18.8	56	58.2± 22.9	30	<0.001	48.6± 18.0	37	45.3± 24.4	49	N.S.
C4 (mg/dl)	12.7± 9.0	56	17.6± 7.8	30	<0.02	15.3± 8.3	37	13.7± 9.4	49	N.S.
C5 (%)	67.9± 19.6	47	59.0± 21.1	20	N.S.	66.9± 17.3	33	63.7± 23.0	34	N.S.
C9 (%)	70.7± 22.2	49	66.5± 30.6	22	N.S.	73.4± 24.3	34	65.7± 25.3	37	N.S.
CH50 (U/ml)	13.5± 5.6	54	15.7± 6.7	29	N.S.	13.8± 4.6	35	14.6± 6.9	48	N.S.
C1q (mg/dl)	5.7± 2.3	54	5.3± 2.3	28	N.S.	5.6± 1.9	35	5.6± 2.6	47	N.S.
Factor B (mg/dl)	5.4± 2.6	53	8.0± 4.3	26	<0.001	5.8± 1.8	33	6.6± 4.3	46	N.S.
Albumin (mg/dl)	3692.3±599.9	52	3289.6±620.3	25	<0.01	3679.2±533.6	34	3468.4±691.5	43	N.S.
IgG (mg/dl)	740.4±310.3	50	549.5±241.3	28	<0.01	763.4±357.5	32	608.2±237.0	46	<0.05
IgA (mg/dl)	3.8± 6.0	50	4.6± 6.4	28	N.S.	4.4± 7.3	32	3.9± 5.2	46	N.S.
IgM (mg/dl)	16.5± 14.4	50	49.3± 80.1	28	<0.01	18.8± 16.0	32	34.8± 65.0	46	N.S.
Haptoglobin (mg/dl)	4.7± 11.8	49	18.0± 30.9	28	<0.01	4.1± 8.1	32	13.5± 26.9	45	N.S.
CRP (mg/dl)	0.0± 0.1	50	0.1± 0.2	28	N.S.	0.0± 0.2	33	0.1± 0.2	45	N.S.
Orosomuroid (mg/dl)	13.1± 6.2	21	17.5± 12.0	15	N.S.	12.9± 7.0	11	15.8± 10.0	25	N.S.
IL-6 (pg/ml)	5.2± 9.8	23	84.9±157.2	10	<0.02	7.0± 15.8	14	45.8±118.6	19	N.S.
Gestational Week	31.9± 3.9	56	28.4± 2.8	30		31.9± 3.6	37	29.7± 3.9	49	

Results are expressed as mean ± SD. For each complement or protein, the difference between the groups of newborns without and with CAM and those with negative and positive bacterial culture was analyzed using Student's *t* test.

レフリーからの指摘 : In table 1, it would appear that ANOVA was used for statistical analysis. Again this is an inappropriate statistical analysis especially given the fact that many values increase with gestational age. I believe that analysis of covariance is a more appropriate to statistical test.

All concentrations should be per liter. Use L, not l, for liter throughout.

訂正後

	Table 2.—The Complement and Protein Levels in Groups Without and With Chorioamnionitis and With Negative and Positive Bacterial Cultures*									
	Chorioamnionitis					Bacterial Culture				
	Without	n	With	n	P	Negative	n	Positive	n	P
C3d, mU/L	24.5 ± 15.1	141	34.6 ± 25.5	74	<.001	26.4 ± 11.9	37	32.8 ± 25.3	48	NS
C2 (%)	84.8 ± 27.6	120	61.7 ± 26.6	54	NS	80.9 ± 25.6	34	75.1 ± 27.8	37	NS
C3, mg/L	430.8 ± 171.0	141	500.9 ± 201.1	74	<.001	486.2 ± 178.0	37	454.6 ± 244.1	48	NS
C4, mg/L	132.4 ± 79.3	141	168.4 ± 80.3	74	<.001	152.8 ± 81.5	37	138.4 ± 93.5	48	NS
C5 (%)	73.2 ± 20.6	115	58.3 ± 23.5	50	NS	66.9 ± 17.0	33	63.7 ± 23.0	33	NS
C9 (%)	84.1 ± 32.0	119	60.0 ± 24.3	54	NS	73.4 ± 24.0	34	65.1 ± 25.0	36	NS
CH50, U/L	15324 ± 6243	121	18272 ± 8681	71	<.001	13750 ± 4542	35	14638 ± 6921	47	NS
C1q, mg/L	58.3 ± 27.2	120	44.1 ± 22.8	70	NS	55.6 ± 18.8	35	54.5 ± 22.8	46	NS
Factor B, mg/L	50.2 ± 22.9	131	77.6 ± 37.4	67	<.001	58.3 ± 17.4	33	65.6 ± 42.9	45	NS
Albumin, g/L	38.5 ± 8.5	129	29.1 ± 7.6	65	NS	36.8 ± 5.3	34	34.9 ± 6.8	42	NS
IgG, g/L	9.4 ± 4.2	97	5.9 ± 3.1	39	NS	7.6 ± 3.5	32	6.0 ± 2.3	45	NS
IgA, mg/L	31.3 ± 43.8	97	66.9 ± 139.4	39	NS	43.6 ± 72.1	32	39.0 ± 51.8	45	NS
IgM, mg/L	155.5 ± 120.3	97	586.4 ± 814.1	39	NS	188.4 ± 157.5	32	351.2 ± 649.2	45	NS
Haptoglobin, mg/L	24.9 ± 60.6	96	203.3 ± 310.1	39	<.01	41.0 ± 79.6	32	123.8 ± 259.1	44	NS
C-reactive protein, mg/L	0.02 ± 0.20	98	1.26 ± 2.96	39	<.001	0.30 ± 1.71	33	0.40 ± 1.14	45	NS
Orosomuroid, mg/L	125.8 ± 49.3	67	194.1 ± 123.4	25	<.001	129.5 ± 66.3	11	158.3 ± 98.1	25	NS
Interleukin-6, ng/L	2.4 ± 6.5	68	91.6 ± 160.8	27	<.001	7.0 ± 15.2	14	45.8 ± 115.5	19	NS
Gestational age (wk)	33.6 ± 5.0	136	28.0 ± 3.8	72		31.9 ± 3.5	37	29.6 ± 3.9	48	

* Results are expressed as mean ± SD. For each complement or protein, the difference between the groups without and with chorioamnionitis and those with negative and positive bacterial culture was analyzed using analysis of variance. NS indicates not significant (*P* > .05).

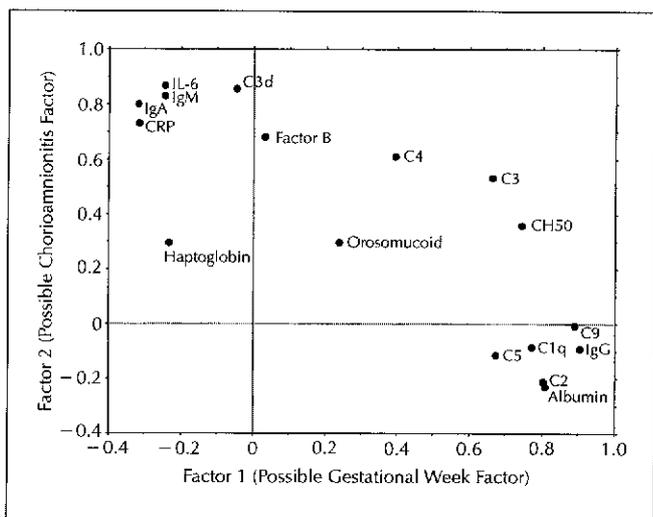
第三例

訂正前

前ページの表の項目を全てグラフでプロットした。

レフリーからの指摘: From reviewing your many graphs, it appears that you have some interesting findings. Your manuscript could be much improved if you could develop a more focused hypothesis directed toward these significant findings.

訂正後



第四例

訂正前: These data suggest that IL-6 is better than CRP to diagnose the acute phase reaction.

CAM	CAM			SNF
	ACUTE	PROLONGED	CAM WITHOUT SNF	
C3d		↑↑↑↑		↑↑↑↑
C3		↑↑↑↑		↑
C4		↑↑↑↑	↑	↑↑↑↑
CH50		↑↑↑		↑↑
C1q	↓↓↓↓	↓	↓↓↓↓	
Factor B	↑↑↑	↑↑↑↑	↑↑↑↑	↑↑↑
C2	↓↓↓↓	↓↓↓	↓↓↓↓	↓↓
C5		↓↓↓↓	↓↓↓	↓
C9	↓↓↓↓	↓↓↓	↓↓↓↓	↓↓
Albumin	↓↓↓↓	↓↓↓	↓↓↓↓	↓↓↓↓
Haptoglobin	↑↑	↑↑↑↑	↑↑↑↑	
Orosomuroid		↑↑↑	↑	
IgG	↓↓↓↓	↓↓↓↓	↓↓↓↓	↓
IgA		↑↑↑↑	↑↑	↑↑↑↑
IgM		↑↑↑↑	↑↑↑	↑↑↑↑
CRP	↑↑↑↑	↑↑↑↑	↑↑↑↑	↑↑↑
IL-6	↑↑↑↑	↑↑↑↑	↑↑↑↑	↑↑↑↑
GW	↓↓↓↓	↓↓↓↓	↓↓↓	↓

レフリーからの指摘: This has also been suggested, have they done a positive predictive value calc or ROC curve to show how much better IL-6 is?

訂正後

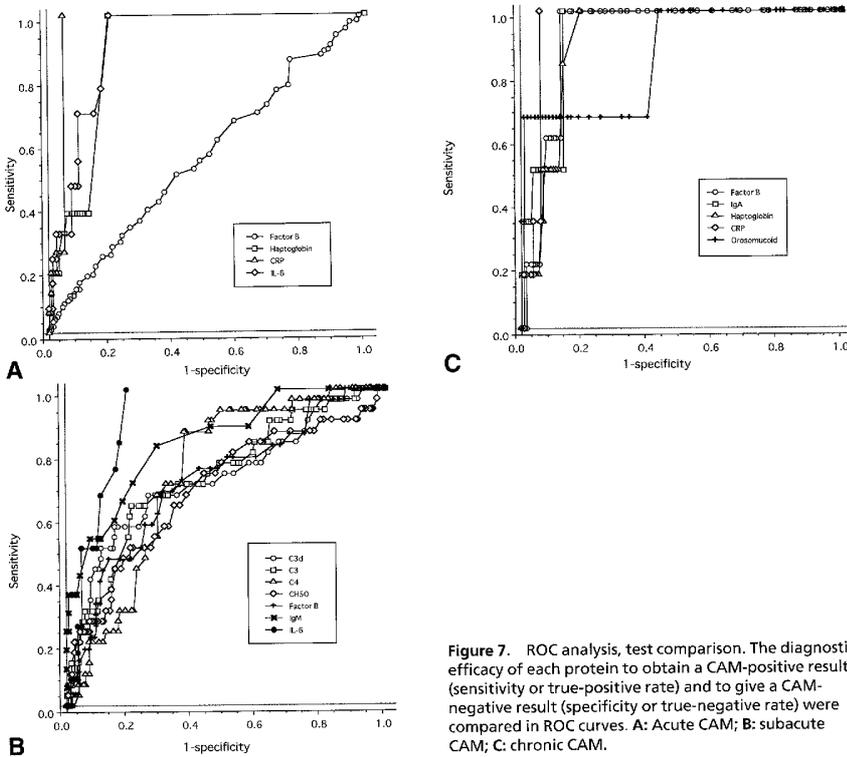


Figure 7. ROC analysis, test comparison. The diagnostic efficacy of each protein to obtain a CAM-positive result (sensitivity or true-positive rate) and to give a CAM-negative result (specificity or true-negative rate) were compared in ROC curves. A: Acute CAM; B: subacute CAM; C: chronic CAM.

第五例

訂正前

前ページの表に対して以下のコメントがあった。

レフリーからの指摘: If ROC plots are to be presented, I believe that the authors should choose a threshold for a few key proteins, which might be clinically useful, and state what the sensitivity, specificity, and predictive values are for these analyses using their data.

訂正後

Table 3. Comparison of sensitivity, specificity, and predictive values of a few key proteins in acute, subacute, and chronic CAM

	Factor B		Haptoglobin		CRP		IL-6		
	-	+	-	+	-	+	-	+	
Acute CAM	-	151	13	110	9	115	6	68	14
	+	24	10	10	6	12	4	4	9
Sensitivity	86%		92%		91%		94%		
Specificity	43%		40%		40%		39%		
Positive predictive value	92%		92%		95%		83%		
Negative predictive value	29%		38%		25%		69%		

	C3d		C3		C4		CH50		Factor B		IgM		IL-6		
	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
Subacute CAM	-	174	11	179	6	178	7	150	12	154	16	110	9	69	14
	+	22	8	24	6	28	2	23	7	21	7	10	7	3	9
Sensitivity	89%		88%		86%		87%		88%		92%		96%		
Specificity	42%		50%		22%		37%		30%		44%		39%		
Positive predictive value	94%		97%		96%		93%		91%		92%		83%		
Negative predictive value	27%		20%		7%		23%		25%		41%		75%		

	Factor B		IgA		Haptoglobin		CRP		Orosomuroid		
	-	+	-	+	-	+	-	+	-	+	
Chronic CAM	-	173	20	127	3	117	12	123	8	83	6
	+	2	3	4	2	3	3	4	2	1	2
Sensitivity	99%		97%		98%		97%		99%		
Specificity	13%		40%		20%		20%		25%		
Positive predictive value	90%		98%		91%		94%		93%		
Negative predictive value	60%		33%		50%		33%		67%		

8. パソコンによるデータ解析

1) 第二例 (SPSSによる)

```
ANOVA  従属変数名リスト
        BY  要因変数名リスト(最小値、最大値)
           [要因変数名リスト(最小値、最大値)]...
        [WITH  共変量変数名リスト]
        [/ 変数名リスト  ...]
```

従属変数名リスト 5変数以内、2変数以上の場合には個々に解析を行う。

要因変数名リスト キーワードBYの後ろには要因変数名を与える。10変数まで指定できるが、交互作用はたかだか5次までである。各変数名は整数値(水準)をとるものであり、解析に用いる範囲をカッコ内にコンマで区切って(最小値、最大値)で与える。範囲外のデータは解析から除外される。値が整数でない場合は、少数部を切り捨てて整数化した値が用いられる。

共変量変数名リスト 共分散分析を行う場合には、キーワードWITHの後ろに10個までの連続型の値をとる共変量を指定する。

解析リスト 従属変数名リストから共変量変数名リストまで1組の解析リストとし、1つの手続きにはスラッシュで区切って5組までの解析リストを指定できる。

2) 第三例 (SPSSによる; 他に StatView でもできる)

```
FACTOR VARIABLES = 変数1 変数2...
                /EXTRACTION = 因子の抽出方法
                /ROTATION = VARIMAX
```

因子の抽出方法

指示語	内容
PC	主成分分析(/ EXTRACTION = を指示しない場合も主成分分析を行う)
PA1	主成分分析(PCと同じ)
PAF	主因子法(共通性を反復推定)
PA2	主因子法(PAFと同じ)
ALPHA	アルファ因子法
IMAGE	イメージ因子法
ULS	重みなし最小自乗法
GLS	一般化最小自乗法
ML	最尤法
DEFAULT	デフォルトは主成分分析

参考文献

- 1) 陶山昭彦: 多群間の解析技法、整・災害、37: 757-764、1994
- 2) 長田理: Macintosh-医学-統計マニュアル、真興出版(東京)、1993
- 3) 棟近雅彦: 統計学の基礎と実践-III.データの数量的なまとめか方-、医学検査、42: 1015-1019、1993
- 4) 三宅一徳: ROC 曲線の求め方と注意点、臨床検査、40: 1409-1413、1993
- 5) 三宅一郎他: 新版 SPSS^x 解析編2、33、東洋経済新報社(東京)、1991
- 6) 司馬正次: やさしいデータ解析 SPSSXによる、第二版、206-207、東洋経済新報社(東京)、1992