

交互作用は相互作用？(2)

コレステロール低下剤Mは有効か？



広津 千尋 | Hirotsu Chihiro

明星大学理工学部教授

■1968年東京大学大学院工学系研究科計数工学専攻博士課程修了。工学博士。東京工業大学助手を経て、1971年東京大学工学部計数工学科講師、助教授、教授。2000年定年退官、東京大学名誉教授。2000年4月より現職。

1. 経時測定データ

表1 体内総コレステロール量 (mg/dL)

表1は薬剤1(実薬M)、2(プラセボP)を処方された被験者それぞれ $n_1=12$ 人及び $n_2=11$ 人について、体内総コレステロール量を4週間おきに6期測定したデータである。このようなデータは、血圧降下剤服用後の血圧推移、薬物の血中濃度の推移、尿中排泄量の推移等、数多く見られる。今回は、このようなデータに基づいて処理(薬剤)効果の違いを検定する方法について考える。

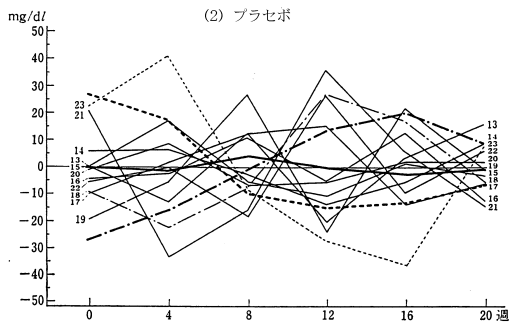
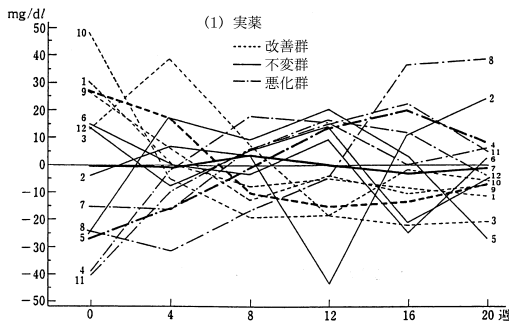
このようなデータの解析で最も陥りやすい誤りは、各期ごとに繰り返し n_1 、 n_2 のデータがあると見て2標本 t 検定を繰り返すことである。これは次のような点で適切でない。

- (1) 薬効は初期値からの下がり具合の差で評価すべきものなので、基本的に薬剤×時期の交互作用解析を行うべきである。
- (2) 繰り返し t 検定を行うとかなりの偽陽性が生じる。しかるべき多重比較法を用いなければならない。
- (3) 各時期ごとの比較では時期によって異なる

処理	被験者	時 期					
		1	2	3	4	5	6
1 実 薬 M	1 (男性)	317	280	275	270	274	266
	2 (男性)	186	189	190	135	197	205
	3 (男性)	377	395	368	334	338	334
	4 (男性)	229	258	282	272	264	265
	5 (男性)	276	310	306	309	300	264
	6 (男性)	272	250	250	255	228	250
	7 (女性)	219	210	236	239	242	221
	8 (女性)	260	245	264	268	317	314
	9 (女性)	284	256	241	242	243	241
	10 (女性)	365	304	294	287	311	302
	11 (女性)	298	321	341	342	357	335
	12 (女性)	274	245	262	263	235	246
2 プ ラ セ ボ P	13 (男性)	232	205	244	197	218	233
	14 (男性)	367	354	358	333	338	355
	15 (男性)	253	256	247	228	237	235
	16 (男性)	230	218	245	215	230	207
	17 (男性)	190	188	212	201	169	179
	18 (男性)	290	263	291	312	299	279
	19 (女性)	337	337	383	318	361	341
	20 (女性)	283	279	277	264	269	271
	21 (女性)	325	257	288	326	293	275
	22 (女性)	266	258	253	284	245	263
	23 (女性)	338	343	307	274	262	309

った結果が生じ、総合的な判断を下しにくいことがある。

さて、こう述べれば、これが被験者、あるいは薬剤の6ヶ月にわたる応答パターンを比較する(すなわち、行比較の)問題であることに気が付くと思う。従って前回の稲の国際適応試験と類似の問題ということになるが、それとの大きな違いは、列(時期)に自然な順序のあることである。つまり、様々な応答パターンの中で、時間軸に沿って上昇、下降、平坦というような推移パターンの違いに特に興味がある。図1に表1のデータに関する交互作用($y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}$)プロットを示す。一見したところ、実薬とプラセボ(偽薬)の間で特別な違いはないように見えるが、よく見ると実薬群では、上昇、下降のような推移パ



注) 改善群、不変群、悪化群は本論の解析法で得られた3つのクラスターで、太線は各クラスターの平均

図1 データの交互作用プロット

ターンが目につくのに対し、プラセボ群では概ねそのような系統的な変動ではなく、むしろ短期的な変動が目につく。特にこのような差異を検出するにはどうしたらよいだろうか。

2. 数学的定式化

差し当たり、薬剤1、2の区別は無視して、 $a = n_1 + n_2$ 人の被験者の応答パターンを比べるものとする。さて、第1節で述べたような興味は、数学的に被験者*i*と*i'*の応答の平行性帰無仮説

$$H_0: \mu_{i1} - \mu_{i'1} = \dots = \mu_{ib} - \mu_{i'b}$$

を傾向性対立仮説

$$H_1: \mu_{i1} - \mu_{i'1} \leq \dots \leq \mu_{ib} - \mu_{i'b}$$

に対して検定する問題として定式化される。 H_1 は被験者*i*と*i'*の応答の差が時間的に拡大していく、つまり相対的に*i*が上昇(悪化)、*i'*が下降(改善)傾向を示すことを表している。

H_1 は、差分行列

$$D_b' = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}_{b-1 \times b}$$

を用いて、

$$D_b' (\mu_i - \mu_{i'}) \geq 0 \quad (1)$$

と表される。ただし、 $\mu_i = (\mu_{i1}, \dots, \mu_{ib})'$ は被験者*i*の平均応答ベクトルを表す。

データから(1)式のような傾向を抽出するには、被験者*i*の観測値ベクトルを $y_i = (y_{i1}, \dots, y_{ib})'$ として

$$D_b' (y_i - y_{i'}) \quad (2)$$

という統計量を考えればよいように思える。ところが、(2)式は時間軸に沿った差分を取ることになり、これはむしろ時間軸に沿った系統的成分を除去してノイズを抽出するのに適した統計量となる。実は系統的成分を抽出するには D_b' の一般逆行列を用いて

$$(D_b' D_b)^{-1} D_b' (y_i - y_{i'}) \quad (3)$$

を基に行間の2乗距離を構成すればよいことが示される(参考文献[1])。ここで興味あることに、(3)式の係数行列は陽に

$$(D_b' D_b)^{-1} = \frac{1}{b} \begin{bmatrix} -(b-1) & 1 & 1 & \cdots & 1 \\ -(b-2) & -(b-2) & 2 & \cdots & 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & (b-1) \end{bmatrix}$$

と表され、それは変化点を順次 1, 2, ..., b-1 とずらしながら、変化点以前の平均とそれより後ろの平均を比べるオペレータとなっている。そこで行間の2乗距離として、(3)式の各要素が等分散になるように基準化した後、その2乗和を取ることとして、具体的に

$$S^*(i;i') = \frac{1}{2} \sum_{i=1}^{b-1} \left[\frac{bl}{b-l} \left\{ \frac{1}{l} \sum_{j=1}^l (y_{ij} - y_{i'j}) - (\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}) \right\}^2 \right] \quad (4)$$

が得られる[3]。この統計量(4)は、前回の $S(i;i')$ と異なり、累積和から構成されているため、累積 χ^2 統計量と呼ぶことにする。

3. 累積 χ^2 統計量の特徴付け

$S^*(i;i') = 0$ は $S(i;i') = 0$ と同じく、被験者 i と i' の応答が平行であることを意味する。一方、その最も良い特徴付けは、展開式

$$S^*(i;i') = \sigma^2 \left\{ \frac{b}{1 \cdot 2} \chi_{(1)}^2 + \frac{b}{2 \cdot 3} \chi_{(2)}^2 + \cdots + \frac{b}{(b-1) \cdot b} \chi_{(b-1)}^2 \right\} \quad (5)$$

与えられる[2]。ただし、 σ^2 は前回同様、誤差分散であり、 $\chi_{(k)}^2$ はチェビシェフの k 次選点

直交多項式に対応する χ^2 成分である。 χ^2 成分の荷重が高次において急速に減衰することから、 $S^*(i;i')$ が主に1次式、2次式で表されるトレンドを検出する統計量であることが分かる。これに対し $S(i;i')$ は

$$S(i;i') = \sigma^2 \{ \chi_{(1)}^2 + \chi_{(2)}^2 + \cdots + \chi_{(b-1)}^2 \}$$

と展開され、全く指向性のない統計量である。

$S^*(i;i')$ を2群間の2乗距離 $S^*(I_1; I_2)$ 、多群間の2乗距離 $S^*(I_1; I_2; \cdots; I_k)$ に拡張するアイデアは $S(i;i')$ の場合と同様である。このようなあらゆる拡張を行ったときの2乗距離の最大値の分布は、(3)式について先に述べた基準化を行った後の係数行列を P_b^* として、Wishart行列 $W(\sigma^2 P_b^* P_b^*, a-1)$ の最大根の分布に従うことが示される[3]。ちなみに、前回の最大値は、指向性のないことに対応して、 $W(\sigma^2 I_{b-1}, a-1)$ の最大根の分布に従う。ただし、 I_{b-1} は単位行列である。

さて、今回の最大値統計量の分布も攪乱母数 σ^2 を含んでいる。そこで σ^2 の推定を必要とするが、今回のデータの場合には、なるべく系統的成分を拾わないような工夫が必要である。そこで、ここでは累積 χ^2 統計量と逆特性を持つように構成した

$$S^{*+} = \frac{\sum_{i=1}^{b-1} \left\{ \frac{y_{ij+1} - y_{ij} - (\bar{y}_{\cdot j+1} + \bar{y}_{\cdot j})}{\rho_j} \right\}^2}{\rho_j = [b/|j(b-j)|]^{1/2} \text{(基準化定数)}} \quad (6)$$

を用いる。結局、行の群分けの有意性を評価するのに用いる分布は、 $W(\sigma^2 P_b^* P_b^*, a-1)$ の最大根を(6)式で除した統計量の分布である。

4. 適用例

表1の被験者23人について累積 χ^2 距離(4)を計算した結果から表2の左欄に示したような

表2 3群の交互作用パターン（上段）と平均の推移（下段）

群	時 期					
	1	2	3	4	5	6
I_1 (1, 3, 9, 10; 23)	27.1	17.2	-9.8	-14.7	-13.1	-6.7
I_2 (2, 5, 6, 12; 13, 14, 15, 16, 17, 19, 20, 21, 22)	336.2	315.6	297.0	281.4	285.6	290.4
I_3 (4, 7, 8, 11; 18)	-26.7	-15.7	-0.7	13.8	20.4	9.0
	259.2	259.4	282.8	286.6	295.8	282.8

表3 2薬剤の3群上での分布

薬 剤	改 善 度		
	I_1 (改善)	I_2 (不変)	I_3 (悪化)
M	4	4	4
P	1	9	1

3群が得られ、その群分けの有意確率は0.011である。これら3群それぞれの平均ベクトル、及び交互作用ベクトル (q_I は群Iのサイズ)

$$(\hat{\alpha}\beta)_{ij} = \frac{1}{q_I} \sum_{i \in I} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}), I = I_1, I_2, I_3$$

は表2右欄のようになる。また交互作用ベクトルを群別にプロットしたのが図2である。これから、明らかに群 I_1 が改善群、 I_2 が不変群、 I_3 が悪化群であることが分かる。すなわち、当初の目的に沿った群分けが行われている。

5. 薬効評価

ところで、本論本来の目的は経時データに基づく薬効評価であった。そこで、群 I_1 、 I_2 、 I_3 の薬剤ごとの頻度分布を求めると表3のようになる。

表3に対する適合度カイ2乗5.49は自由度2のカイ2乗分布として有意確率0.065に当たるが、列のカテゴリーに沿ったトレンド成分と2次成分に分解するとその全部が2次成分である。従って2次成分が有意であると解釈して差し支えない。つまりMとPの効果には有意差がある。Pが不変群に集中するのに対し、

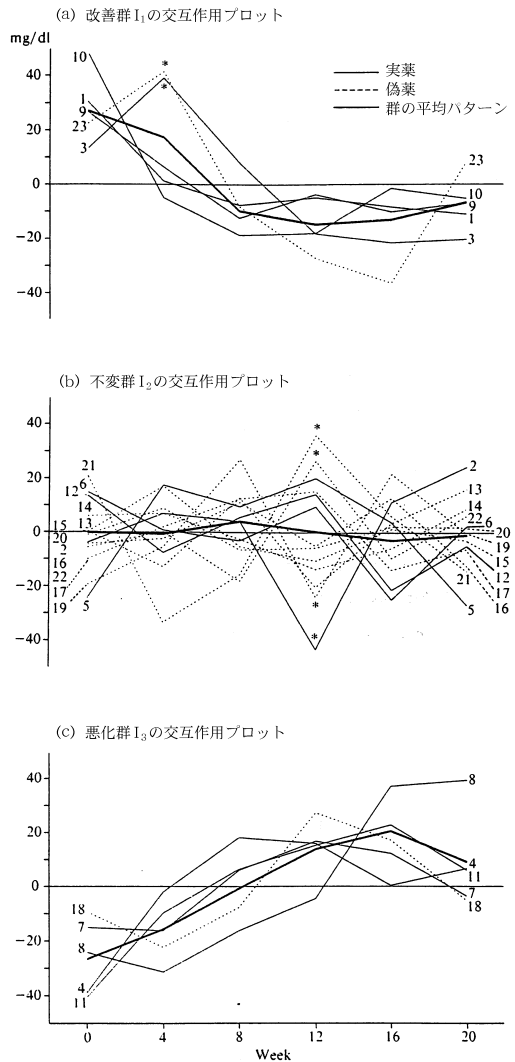


図2 交互作用プロット及び群の平均パターン

表4 5群の交互作用パターン（上段）と平均の推移（下段）

群	時 期					群の特徴付け
	1	2	3	4	5	
I ₁	79.7	-0.2	-40.9	-34.7	-3.8	非常に高度に改善
	402.3	292.0	250.3	257.7	289.3	
I ₂	22.6	8.6	-1.0	-12.1	-18.1	高度に改善
	296.6	252.2	241.6	231.5	226.3	
I ₃	14.3	-5.8	-7.9	-5.7	5.1	改善
	278.4	227.8	224.7	228.1	239.6	
I ₄	-17.0	5.5	10.4	6.8	-5.7	やや改善
	260.1	252.2	256.1	253.6	241.9	
I ₅	-23.4	-7.7	0.9	12.4	17.7	不変又はやや悪化
	263.4	248.7	256.4	268.9	274.9	

Mは変化を来すが、改善と悪化が同程度生じるのでPより有効であるという評価にはつながらない。もし、Mにより改善された被験者と悪化した被験者の医学的特徴付けができれば、薬として使える可能性がある。一方、プラセボが毒にも薬にもならないというのは、大変納得のいく結果である。

なお、注目すべきことに、多変量正規分布モデルを採用し、M、Pの平均レベルを比較すると、Mにおいて改善群と悪化群が相殺する結果、Pとの差が見えなくなる。MとPの差異はむしろ分散行列にある（表3で2次成分が有意になることに対応する：Mの方がばらつきが大きい）のだが、通常の分散行列の同等性の検定では時間順序が考慮されていないため、有意差は検出されない。すなわち、このデータは通常の変量解析の手法では何の特徴も見出せない。

6. まとめ

前節の薬効評価は、正に主治医がそれぞれの患者の治療経過から、いわゆる全般改善度を判定した結果の順序分割表に基づく薬効評価に対応している。通常が多施設臨床試験では、複数の主治医の主観評価はばらつきが大

表5 改善度の分布

薬 剤	群					
	I ₁	I ₂	I ₃	I ₄	I ₅	計
実薬M	3	28	35	14	9	89
対照薬	0	6	11	33	28	78

きく科学的でないとの批判がなされ、近年使用を控える傾向にあるが、本論の方法はそれに替わる客観的方法を与えたものと言える。なお、例題中の実薬Mは本解析結果ではやや特異的に見える。そこで、最後に、現在世界中で使用されている別の実薬Mの臨床試験の結果を示す。本論の手法を適用した結果である5分類と、その5群に対する実薬M及び対照薬の頻度分布を表4、5に掲げる。表5に対して順序分類データに対する2群比較の方法を適用すると、実薬Mが対照薬に対し高度に優れるという結論が得られる[3]。

*参考文献

- [1] Hirotsu, C. (1982) : Use of cumulative efficient scores for testing ordered alternatives in discrete models : *Biometrika* 69, pp.567-577.
- [2] Hirotsu, C. (1986) : Cumulative chi-squared statistic as a tool for testing goodness of fit. *Biometrika* 73, pp.165-173.
- [3] Hirotsu, C. (1991) : An approach to comparing treatments based on repeated measures : *Biometrika* 75, pp.583-594.