

論 説

ソフト・データの多変量解析による素診断

ー 国民医療費の抑制と被験者負荷の低減を目指して ー

平 井 孝 治

目 次

はじめに

§ 1 素診断用のソフトなデータ

§ 2 解析枠の概要 と 説明変数の設定

§ 3 目的変数の選定

§ 4 解析の Step と「判別」の定量化

§ 5 モデル変遷の概要 と 回帰式の P 値

§ 6 疾病の判定, 回帰精度と判別確度 D

おわりに

文末の注

付録 資料 2 各群の変遷

はじめに

過去 10 年ほど医療経営を研究して来た筆者は、かねてより被験者の「測定データ」や「アンケート・データ」、医療従事者が被験者を観察・問診して得られる「ダイアグ・データ」など、容易に入手できるソフトなデータを多変量解析して、素診断 (Primary Diagnosis) に活かすことが出来ないものかと考えて来た。

費用や時間の掛からないソフトな診療データを解析した結果が活用できれば、素診断の段階で、①被験者が罹患している可能性のある疾病を見出し、逆にその可能性のない無駄な治療を回避し、②肉体的にも精神的にも、あるいはまた経済的、時間的にも、被験者に負荷を強いる「ハードな検査」の必要十分性を事前に判定できるようになる。

①誤診を防ぎつつ疾病の早期発見に貢献し、②被験者に掛かる負荷を低減し、結果として、③高騰する国民医療費の抑制にも資する『ソフト・データの多変量解析による素診断』の手法の開発は、国民医療喫緊の課題である。

未だ、ライブなデータによる解析精度の議論には至っていないが、乱数による試行によって得られた知見を開示し、「素診断の豊かな可能性」を披露するのが、この小論の使命である。当該手法の開発は緒に就いたばかりだが、これまでに得られた研究成果を、協力医療機関のライブ・データでさらに発展させたいと痛切に思っている。該手法の開発完了までには長い期間を要すると思われるが、これは筆者の生涯最終の研究課題でもある。

§ 1 素診断用のソフトなデータ

国民医療費が高騰している。その国庫による負担額¹⁾を東京都の歳入総額、トヨタ自動車(単独)や電力9社²⁾の売上高と比較してみると、2000年度と2010年度とのここ10年間で、表0のごとくである。産業としてみると、筆者が医療経営に関心を抱いたわずか10年間で46.1%も成長し、今や国民医療は40兆円産業になっている。

表 0

代表的な機構の年度額, 10年間比較

| | 医療費の国庫負担 | 東京都の歳入総額 | トヨタ(単独)売上高 | 電力9社 電灯・電力 |
|--------|----------|----------|------------|------------|
| 2010年度 | 9兆9273億円 | 6兆1707億円 | 8兆2428億円 | 14兆2653億円 |
| 2000年度 | 6兆7956億円 | 6兆5143億円 | 7兆9036億円 | 14兆7506億円 |
| 増加率 | 46.1% | -5.3% | 4.3% | -3.3% |

このように、高齢化時代を迎え国民医療費の抑制は国家喫緊の課題であることは、論を俟たない。従来、国民医療費の抑制のためには、a. 予防と b. 早期発見の必要が説かれてきた。更に、筆者は c. ハードな検査の必要十分性の吟味も有効であると考えている。「必十性」のうち、よく理解できる「必要な検査」に対し、「十分な検査」とは「余計な検査をしない!!」ということである。

かねてから医療原価を主題として医療経営を研究して来たが、研究者として(時には患者として)、高額医療機器の減価償却(と医師の自己満足や保身)のため、「必要もないハードな検査」が実施されていると実感したことがしばしばある。国民医療について、かつて「三漬け」と呼ばれた時代があった。即ち、「薬漬け」「検査漬け」「入院漬け」の三つがそれである。

これらは医療機関を潤していたが、その後、「薬漬け」については「医薬分業」がなされ、医療機関が患者に余計な医薬品を処方することが少なくなったと云われている。また「入院漬け」についても DPC(疾病別定額報酬)が導入され、収益向上志向から費用低減志向に変わり、入院治療の様相が出来高払いの時代のそれとは一変したと喧伝されている。

病院などの医療機関で受診したり治療を受けたりした際には、現在では患者にも「診療明細」が開示され、診療区分別の保険点数を知ることが出来るようになった。そのなかで目立って点数の高い診区は「画像診断」である。医療における検査の区分³⁾にはいろいろな考え方があるが、この論文では便宜上「ソフトな検査」と「ハードな検査」に区分する。後者は PET など高額な医療機器(本文末の注 b. を参照)を用い、被験者に負荷を強いる医療検査である。

1) 国民医療費の算定法は各種あるが、ここでは、そのうちの一番確かな国庫負担額だけを示している。実際のそれは、概ね国庫負担額の約4倍と考えれば、まず間違いはない。

2) 敢えて沖縄電力を除いているのには理由がある。文末の注 a. を読んでもらいたい。

3) さしずめ、「診療報酬点数表」の「診療区分」における検体検査や生体検査がここである「ソフトな検査」、「画像診断」が「ハードな検査」と考えてもらえばよい。

ハードな検査によって被験者がこうむる負荷には、①被曝など肉体的なものから、②精神的なもの、そして三割負担に起因する③経済的な負荷などがある。検査に関わる時間的な負荷も看過できないであろう。ハードな医療検査を時系列でみると、予約して検査を受けるまでの待機期間、検査そのものにかかる時間、「画像診断」などの検査結果が出るまでの期間、これらはハードな検査にまつわる被験者の④時間的な負荷でもある。

筆者はハードな検査を決して忌み嫌っているわけではない。疾病の早期発見にそれが必要な場合も当然だろう。しかし高額医療機器の減価償却だけのために、必要もない検査を強いられるのは御免こうむりたい、と誰しも考えるだろう。そこで、入手容易なソフト・データによる素診断（Primary Diagnosis）が出来ないものかと思考する。

素診断用のソフト・データには、

- ① 体温や血圧・脈拍など、測定やソフトな検査によって容易に得られる「測定データ」、
- ② 既往症や体調・日頃の生活など、被験者や家族が即答できる「アンケート・データ」、
- ③ 看護師など、医療従事者が被験者を観察・問診しただけで得られる「ダイアグ・データ」

の三種類がある。これ等のソフト・データはいずれも入手するに、さして時間もかからず、費用も安く、被験者に掛かる負荷も軽くて済むデータ達である。

医療行政に携わっているわけでもなく、また医師や看護師などの医療従事者でも無いただの研究する筆者が、国民医療費の抑制に貢献できるとしたら、a. 予防と b. 早期発見と c. ハードな検査の必十性のうち、後二者についてであろう。すなわち、『ソフト・データの多変量解析による素診断』である。これが開発できれば、被験者の負荷低減のみならず、国民医療費の抑制にも大きく貢献することになる。

ためにこの小論は、診療科別ないし『疾病群別、判別重回帰分析』の可能性を論じようとするものである。そこには大きな壁が立ちふさがっている。その第一の壁は、研究・開発のためにライブなデータを提供していただける協力機関があるか、という問題である。第二の壁は、疾病群別に素診断する際、重回帰分析で出現する下記三つの「精度課題」である。

- ① 重回帰分析では精度を上げるために、残差絶対値の大きいサンプルを（5%を限度に）解析枠から除外するのを常とする。しかし、如何に残差が大きくとも「患者等」に、この5%ルールを適用して、解析枠から除外するわけにはいかない。
- ② 素診断のためには複数の疾病を同時に解析の対象にするので、同一の解析枠で複数の目的変数を扱わざるを得ない。ために、複数の目的変数をにらみながら残差絶対値の大きいサンプルを除外する Algorithm を確立する必要がある。
- ③ 来街者調査などでアンケート調査を実施し、知見を出す際の自由度修正済み決定係数 Q^2 は、 $Q^2 \geq 0.4096$ もあれば「実証」に使える。しかし、素診断に使えるモデルの精度はどの位あって然るべきなのか。

この小論では、これら三つの精度課題を究めるために、未だライブなデータが入手できていない段階で、乱数を用いて試行してみた⁴⁾。以下の各節は、その結果得られた知見の開示に充てている。この節の小括に替えて、次に共著論文「震災による消費者応善意識の変化」(『立命館経営学』第 51 巻第 2・3 号) の表 10 で提示した自由度修正済み決定係数 Q^2 のレベル区分を再掲しておく。

表 1 自由度修正済み決定係数 Q^2 と相関係数 r' のレベル区分

| | | | | | | | | | | | | | | | | | | |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|--------|--------|----|--------|----|--------|----|--------|----|----------------|
| 0 | 無相関 | 0.120 | 弱い相関 | 0.400 | 中程度の相関 | 0.640 | 強い相関 | 0.840 | 理論的な相関 | r' | | | | | | | | |
| {0.2i} | {0.120} | {0.265} | {0.400} | {0.525} | {0.640} | {0.745} | {0.840} | {0.925} | | | | | | | | | | |
| -0.04 | 無相関 | 0.0144 | 乱回 | 0.0702 | 若干 | 0.1600 | 参考 | 0.2756 | 説明 | 0.4096 | 実証 | 0.5550 | 実用 | 0.7056 | 規範 | 0.8556 | 近似 | Q ² |

ここに「i」は虚数

| r' の例 | | 修正済み決定係数 Q ² の解釈 | | アンケート・データ 重回帰の要求水準 | | |
|------------------------------------|--|-----------------------------|----------------|--|-----|-------------------------------------|
| サンプル数 n = 14 | | 乱回 | 乱数で回帰した程度の弱相関 | 要求の厳格度 | 項 目 | 要求水準 |
| 修正まえ相関 係数 r = 0.4 | | 若干 | 若干相関が認められる弱相関 | ① | ◎ | 自由度修正済み決定係数 Q ² ≥ 0.4096 |
| 修正済み決定 係数 Q ² = 0.09 | | 参考 | 参考にしか使えない程の相関 | ② | ◎ | 説明変数の数 q q ≤ 12 |
| 修正済み相関 係数 r' = 0.3 | | 説明 | 説明に使える程度の中相関 | ③ | ◎ | 符号マルチコ変数の数 m m = 0 |
| | | 実証 | 実証に援用できる強い相関 | ④ | ◎ | 各説明変数の p 値 p ≤ 5% |
| | | 実用 | 実用に供していい強力な相関 | ⑤ | ○ | 回帰式の P 値 P ≤ 10 ⁻¹² |
| | | 規範 | 規範として一線を画すレベル | ⑥ | △ | 各説明変数の F 値 t ² ≥ 6.000 |
| | | 近似 | 1 次式で代用していいレベル | 注 a) 回帰式の P 値では「☆」一つで 10 ⁻⁶ | | |

注 b) 自由度修正済み決定係数が $-1 \leq Q^2 < -0.04$ のときのレベル名は「論外」であるが、普通には惹起（回帰）しえないレベル。

注 c) 自由度修正済み相関係数 r' は、重回帰したときの重相関係数に元の符号を付したもの。ただし、 Q^2 が負ときは、 r' に虚数「i」を付す。

以下の乱数試行では、この表の「レベル区分」を基に、疾病群モデルの精度を論ずるが、未だライブなデータによる実証までには至っていない。素診断に求められる修正済み決定係数 Q^2 は、対象となる疾病群にもよるが、ことの性質上、かなり高い「実用」から「規範」の水準を要するものと思われる。(表 10 参照)

以下、説明変数の設定、目的変数の選定、判別の定量化、Epoch を示す「登録カード」、試行結果による知見、の順に書き進める。

§ 2 解析枠の概要と説明変数の設定

被験者数 $n = 220$ 名の下、疾病群には α, β, γ なる三つの疾病を想定し、その一方で説明変数には乱数を用いて、次のような三種類のソフト・データを設定した。

① 性別や「是否」のような名義的な変数を仮想した「0 と 1 だけのデジタルな A 型変数」

4) この試行には実に 8 ヶ月を要した。

を a1 から a21 まで 21 変数。

- ② アンケートやダイアグ・データを仮想した「1, 2, 3, 4 のスケールからなる B 型変数」も b1 から b21 まで 21 変数。
- ③ 測定値やソフトな検査データを仮想した元来アナログな「0 点～100 点の C 型変数」を c1 から c28 まで 28 変数。

以上、計 70 変数を仮想し、 $220 \times 70 = 15,400$ 個の乱数を発生させて、説明変数として用いている。A 型と B 型の説明変数の乱数発生実績は表 2 のごとくである。

表 2

A, B 型 乱数の発生実績

| A 型説変 | | 選択肢 | B 型説明変数 | | | |
|--------|--------|------|---------|--------|--------|--------|
| 0 | 1 | | 1 | 2 | 3 | 4 |
| 49.5% | 50.5% | 設定確率 | 29.2% | 39.3% | 20.3% | 11.2% |
| 49.52% | 50.48% | 発生割合 | 29.22% | 39.31% | 20.32% | 11.15% |

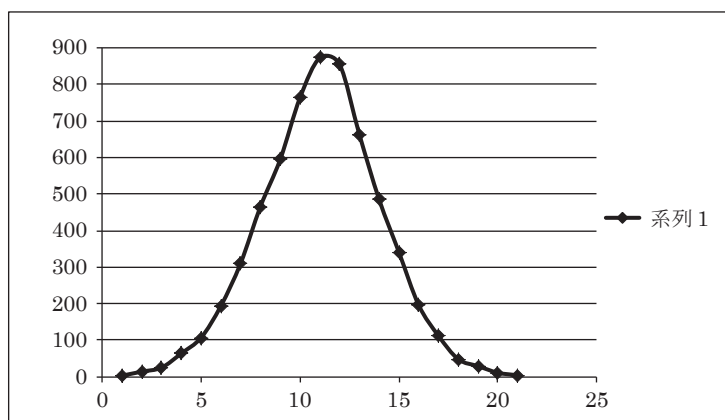
ここで A 型説明変数の発生割合を 50%, 50% より若干偏向したのは、過去の経験から、その方がいい回帰結果が得られるからである。また、B 型説変の選択肢「2」を尖頭に「4」を幾分低く設定したのも、過去のアンケート実績に合わせたためである。

C 型の 0～100 のアナログ型変数は、まず変数ごとに平均が $\mu = 50$ 、標準偏差が $\sigma = 12.5$ の下に正規乱数を発生させ、それを 28 変数全体で 0 点～100 点の自然数になるよう調整して丸め（四捨五入）た。その結果、 $\mu = 50.30$ 点、 $\sigma = 14.87$ 点の正規乱数が得られたが、その分布実績を表と図にしてみると、以下のようになった。

| データ区間 | 発生頻度 |
|--------|------|
| 0～2 | 3 |
| 3～7 | 14 |
| 8～12 | 25 |
| 13～17 | 64 |
| 18～22 | 107 |
| 23～27 | 194 |
| 28～32 | 311 |
| 33～37 | 466 |
| 38～42 | 595 |
| 43～47 | 764 |
| 48～52 | 873 |
| 53～57 | 855 |
| 58～62 | 663 |
| 63～67 | 485 |
| 68～72 | 341 |
| 73～77 | 196 |
| 78～82 | 112 |
| 83～87 | 48 |
| 88～92 | 29 |
| 93～97 | 11 |
| 98～100 | 4 |

図 1

C 型説明変数の分布実績



このようにして得た説明変数用の 15,400 個の乱数値は (わずか 3 個を例外として), その後一切手を加えていない。ただし, 後に変数ラベル名のみ, 解説の都合上, 付け替えを行っている。

なお, 実際の重回帰に供した説明変数は, 以上の 70 変数の他, あと 2 変数ある。というのは例えば, A モデル群の解析では, 後に選定する「目変 α 」を重回帰分析する際に, 残りの「目変 β 」と「目変 γ 」を説明変数に廻すので, 解析の出発時点 (Step 0) では説明変数の数 q が = 72 となるからである。

この節を閉じるに当たり, 後の議論に供するため, そもそも乱数変数間の相関係数はどのように分布しているのかを, 各変数の個数を $n = 220$ に設定して調べたが, その結果を以下に示しておく。Excel でいろんなタイプの乱数変数を $201 \times 200 \div 2$ を 25 組発生させ, 実に計 502,500 件の相関係数に当たってみた結果は, 次のごとくであった。

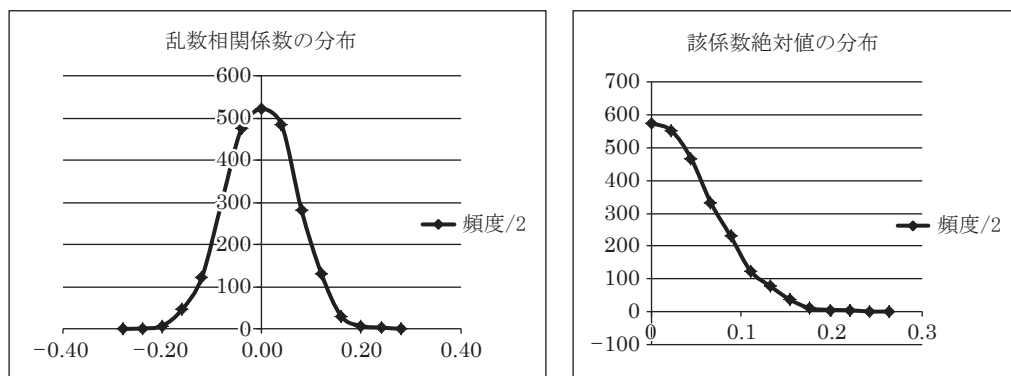
- ① 平均値 ϕ は限りなく 0 に近く, 誤差 ε は常に $\varepsilon \approx 5 / \text{相関係数の件数}$ であった。
- ② 標準偏差は $\sigma = 0.068180$ であった。
- ③ そのため, 該相関係数の絶対値の平均 μ と標準偏差 σ' は

$$\mu = \sigma \sqrt{2 / \pi} = 0.054400, \quad \sigma' = \sigma \sqrt{1 - 2 / \pi} = 0.041100$$
とあいなった。
- ④ 乱数による相関係数の絶対値の上限はほぼ $|r| \leq 0.2300$ で,
これを超えるものは 502,500 件のうち僅か 372 件 (0.0740%) であった。

既述の 70 の乱数説明変数に限って相関係数の分布を, 参考まで図 2 に示しておく。

図 2

乱数変数間の相関係数の分布



注) 左右のグラフで頻度数が整合していないのは, データ区間の区切り方が異なるからである。

§ 3 目的変数の選定

疾病群別の判別重回帰が開発したいので, 目変を以下のように三種類選定することにした。疾病 α は罹患しているか否かをデジタルに区分できる疾病で, α 値が「1」なら「疾病 α の罹患患者」, α 値が「0」なら「健常者」を意味している。そして被験者 220 名中 10 名が罹患者と

なるように乱数を発生させ、「目変 α 」を選定している。

二番目の疾病 β は罹患状態を重位から軽位へ四段階のスケールに分け、入院・加療を要する者4名を $\beta = 4$ とし、罹患している容疑のある者7名を $\beta = 3$ とし、健常な者115名を $\beta = 2$ とし、当該疾病に無縁な者94名を $\beta = 1$ とし、乱数を発生させ「目変 β 」を選定している。

以上、二つの疾病と後述する三番目の疾病 γ を、それぞれの疾病の程度（以下、本文ではこれを「当級」と称する）で分けた「当級区分」を、先に表3にして示しておく。

| 表 3 | | 疾病別 当級区分 | | | | | | | |
|------|-------------|------------|-------|------------|-----------|-------|-------------|------------|-------|
| | 疾病 α | α 値 | 人数 | 疾病 β | β 値 | 人数 | 疾病 γ | γ 値 | 人数 |
| 患者など | 罹 患 | 1 | 10 人 | 罹 患 | 4 | 4 人 | 重 篤 | 191 ～ 200 | 2 人 |
| | | | | 容 疑 | 3 | 7 人 | 入 院 | 171 ～ 190 | 4 人 |
| | | | | | | | 要観察 | 152 ～ 170 | 8 人 |
| 健常者等 | 健 常 | 0 | 210 人 | 健 常 | 2 | 115 人 | 健 常 | 00 ～ 151 | 206 人 |
| | | | | 無 縁 | 1 | 94 人 | | | |

注) No.10 のサンプルは $\alpha = 1$ で $\beta = 4$ の合併症、No.20 のサンプルは $\beta = 3$ で $\gamma = 194$ の合併症。

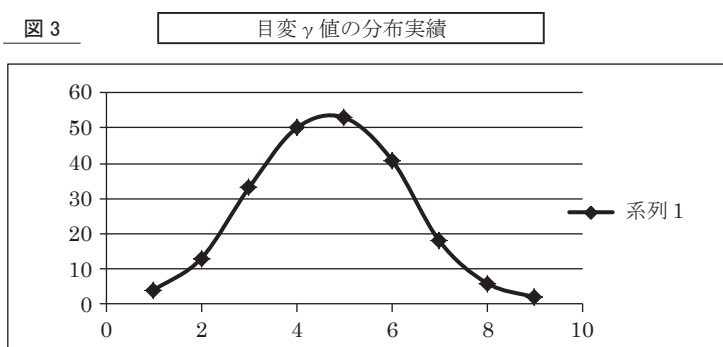
三番目の疾病 γ はICU（集中治療室）に入れなければならないほど重篤な患者2名と、入院・加療を要する患者4名と、通院させ、検査をし、今後も様子を見守る必要のある「要観察」者8名を、正規乱数を発生させて、次のように決めた。

まず、平均が $\mu = 100$ 、標準偏差が $\sigma = 25$ の下に正規乱数をサンプル数の $n = 220$ 個発生させ、それを「仮の γ 値」とし、「目変 γ 」とみなして選定に備えた。これら仮の「目変 γ 」を約1,000回も発生させ、その中から後述する基準で一つだけを選び、それを以下のような「 γ 操作」と称する手法で解析枠に繰り込み、正規の「目変 γ 」とした。

その「 γ 操作」であるが、発生させた220個の乱数を、まず少数第二位に丸め、最大値 $a = 171.37$ と最小値 $b = 28.51$ を得た。その両値の中点 $c = 99.94$ が100とさほど変わらなかったのので、先に得た仮の γ 値と c との差を $k = 1.4$ 倍して自然数に丸め、各サンプルの正規の γ 値とした。その結果、最大値が200点、最小値が0点、平均値が99.15の「目変 γ 」が得られた。この三番目の疾病に対応する γ 値の分布実績を、次に図3として示す。

| データ区間 | 発生頻度 |
|-----------|------|
| 0 ～ 22 | 4 |
| 23 ～ 46 | 13 |
| 47 ～ 70 | 33 |
| 71 ～ 94 | 50 |
| 95 ～ 118 | 53 |
| 119 ～ 142 | 41 |
| 143 ～ 166 | 18 |
| 167 ～ 190 | 6 |
| 191 ～ 200 | 2 |
| 計 | 220 |

図 3



このようにして得た γ 値で高位から順に、2 名、4 名、8 名と分け、それぞれ当級を「重篤」、「入院」、「要観察」患者とし、残りの 206 名を「健常」者とみなした。

以上で患者等は 220 名のサンプル中、疾病 α が 10 名、疾病 β が 11 名、疾病 γ が 14 名の延べ 35 名になった。が、 α と β の合併症、 β と γ の合併症が各 1 名存在するので、実際の患者等の実数はサンプル総数の 15% に当たる 33 名になった。なお、サンプルの付番 (renumbering) については、文末の注 c. を参照してもらいたい。

70 変数の説明変数の設定はいとも簡単に済んだが、三つの疾病にそれぞれ対応する三種の目的変数の選定には、実に半年以上にも及ぶ長期間を要した。今回の乱数による試行研究では、次の三つの困難に遭遇していた。

- ① 自由度修正済み決定係数 Q^2 のレベル区分
- ② 三つの目的変数「目変 α 」、「目変 β 」、「目変 γ 」の選定
- ③ 解析枠から除外するサンプル算定の Algorithm (以下、本文では「除外 Algo」と称す)

このうち②と③は、同時に決定せざるをえない性格のものであるが、三つの目的変数の選定に際し、次のことをその条件ないし基準とした。

- ① 解析の出発時点 (Step 0) における修正済み決定係数 Q^2 が、それぞれ後述の水準にあること。
- ② 疾病群別の判別重回帰分析の開発に役立つ資質を有する目的変数であること。
- ③ 解析の最終時点 (Step 34) までのストーリーが、判明した知見を明確に体现していること。

これらの条件ないし基準を満たす目的変数を選定するために、「目変 γ 」のみならず、「目変 α 」や「目変 β 」もそれぞれ約 1,000 回も乱数を発生させ、検討している。その結果、設定しておいた既述の説明変数に対し、自由度修正済み決定係数 Q^2 につき、次のような事実が生起した。

- a. 出発時点 (Step 0) におけるそれは $-0.040 \leq Q^2 \leq 0.120$ の範囲で概ね正規分布する。
- b. 0,1 型の「目変 α 」のそれは $0.010 \leq Q^2 \leq 0.070$ 、スケール型の「目変 β 」のそれは $-0.040 \leq Q^2 \leq 0.020$ 、アナログ正規型の「目変 γ 」のそれは $0.060 \leq Q^2 \leq 0.120$ の範囲で概ね正規分布する。
- c. どのような乱数目的変数でも、 Q^2 は説明変数の数 q が 36 変数 (Step 23) あたりでピークを迎え、 $q = 24$ (Step 27) あたりから (後述する) 破局的な事態 (Catastrophic Phenomenon) が現象する。

何故このようになるのか、今のところ定かではないが、ライブなデータならば間違いなく、出発時点の修正済み決定係数 Q^2 がそれなりに高く、その後、残差絶対値の大きいサンプルを解析枠から除外し、説明変数を削り込んでいけば、破局的な事態には至らず、疾病の当級識別が出来るものと思われる。

ともあれ、上記 b. の事実から、デジタル型の「目変 α 」については $Q^2 \approx 0.040$ 、スケール型の「目変 β 」については $Q^2 \approx -0.010$ 、アナログ型の「目変 γ 」については $Q^2 \approx 0.090$ に

なるように、三つの目的変数を選定した。

その後、回帰係数 k_i の変動が修正済み決定係数 Q^2 にどのような影響を及ぼすのかを調べる意味もあって、除外 Algo がなんであれ、いつも最初に除外されるサンプル No. 218 の三つの説明変数 c1, c2, c3 の小数第二位までを、以下のように手入れした。その結果を表 4 に示しておく。

表 4

一部の説明変数 数値変更の結果

| No. 218 | c 1 | c 2 | c 3 | Q^2 | 目変 α | 目変 β | 目変 γ |
|---------|-------|-------|-------|----------------|-------------|------------|-------------|
| 変更前 | 19 | 44 | 22 | 変更前 | 0.04015 | -0.01011 | 0.09029 |
| 変更後 | 18.71 | 43.55 | 21.83 | 変更後 | 0.04000 | -0.01000 | 0.09000 |
| | | | | 修正済み重相関係数 r' | 0.20000 | 0.10000 i | 0.30000 |

§ 4 解析の Step と「判別」の定量化

先の § 2 で設定した説明変数と、§ 3 で選定した目的変数とで構成された解析枠で、ここから先、「目変 α 」, 「目変 β 」, 「目変 γ 」をそれぞれ対象とした、A モデル群、B モデル群、C モデル群を、Step 0 ($q = 72$) から Step 34 ($q = 3$) まで、回帰分析して求めることになる。

診療科別ないし疾病群別の判別重回帰の開発が目的なので、解析枠から 5% 以内で除外するサンプルは、三群とも同じものでなければならない。従って（計 33 名いる）どの疾病の「患者など」も解析枠から除外するわけにはいかない。が、群によって目的変数が異なるので、当然のことながら削除して行く説明変数は、モデル群によって異なって来る。

過去の経験によれば、5% ルールによって外乱サンプルを除外し終わるタイミングは、説明変数の数 q が当初の半分になる直前がベストのようである。そこで Step を次のようにして、回帰分析を進めていく。

- ① $q = 72$ (Step 0) からスタートして、説変を三つ削除しては、一つのサンプルを除外する。これを 11 回繰り返し、Step 22 に至ると、説明変数の数が $q = 39$ で、サンプル数が 209 件となる。その後は $q = 3$ (Step 34) になるまで、三変数ずつ削除して行く。
- ② 重要性を示唆する F 値 (t 値の二乗) の小さいものから順に説明変数を削除するのを原則とする。ライブなデータなら F 値が小さくとも、意味論的に重要な説明変数を残すようにすべきだが、乱数変数に元より意味などない。そこでこの試行研究では、目的変数との相関係数 (の絶対値) が 0.137 を超える⁵⁾ 説明変数をそれと読み替えて、如何に F 値が小さくとも、外乱サンプル 11 件を除外し終わる Step 22 までは、当該変数たちを削除しないこととする。
- ③ 各モデル群とも Step 22 までは轡を並べて解析していくが、解析枠から除外するサンプル

5) なぜ 0.137 にするのかについては、文末の注 d. を参照。

を算定するには、除外直前の Step における各サンプル i の目的変数別の「標準化残差」⁶⁾の絶対値 R_α , R_β , R_γ 」につき、次のような加重平均 m_i を求め、

$$m_i = 0.2R_\alpha + 0.2R_\beta + 0.5R_\gamma + 0.1 \text{ Min } (R_\alpha, R_\beta, R_\gamma)$$

この値の最も大きなサンプルを、次の Step に進む前に解析枠から除外しておく。

この算定法を本論では「除外 Algo」⁷⁾ と略称している。繰り返しになるが、判別重回帰分析では「患者など」に属するサンプルはその対象外である。あくまでも「健常者等」にのみ限定し、適用されて然るべき手法である。この算定手法は、ここ数か月間、該解析枠から香り、筆者の嗅覚に訴えてきたものを Algorithm 化したものであるが、そのフレーバーについては、煩雑さを回避するため、文末の注 e. に譲ることとする。

この節の以上はいずれも修正済み決定係数 Q^2 , 即ち各モデル中の「回帰の精度」に関わるものであったが、判別重回帰で精度と同程度に重要な概念は、言うまでもなく「判別」そのものである。『判別の確度』を数値として定量化できなければ、「ソフト・データによる素診断」は始まらない。

そのため、判別に関わる用語を表 5 にまとめ、次に掲載しておく。なお、これらの用語は、各 Step の重回帰分析によって得られた回帰値につき、高い方から低い方へと降順に、全てのサンプルがソートされていることを前提に定義されている。

表 5 判別に関わる用語集

| 用 語 | 解説ないしコメント |
|----------|--|
| 当級 (クラス) | 各疾病の程度のこと。「クラス」と表現することもある。 |
| 当級名 | 当級を表すクラス名のことで、疾病 β の「罹患」や「容疑」などがそれ。 |
| 健常者等 | 当級名が「健常」または「無縁」のクラスに属するサンプル。 |
| 患者など | 健常者等以外のサンプルで、疾病 γ の重篤者や要観察者などがそれ。 |
| 一人クラス | 該当するサンプルが一つしかない当級。 |
| 重位クラス | 当該クラスより症状がより重い当級。 |
| 軽位クラス | 当該クラスより症状がより軽い当級。 |
| 上位 (下位) | 当該サンプルより回帰値が高い (低い) サンプル。 |
| 当級最下位 | 当該クラスで回帰値が最も低いサンプル。 |
| 直前クラス | 今議論している当級より疾病の程度が 1 級だけ重いクラス。 |
| 天井 | 最重症クラスの場合、「直前クラスの最下位」が存在しないので、回帰値で降順にソートしたサンプル列の 0 番目を仮想し、本論ではそれを解りやすく単に「天井」と表現している。 |
| 疾病最下位 | 当該疾病につき、「患者など」のうち、最も回帰値が低いサンプル。 最軽症クラスの最下位者とは限らないので、注意を要する !! |
| 敷居 | 疾病最下位と、それ以下の回帰値を有するサンプルとの境界。 |
| 敷居侵害 | 「健常者等」の回帰値が、疾病最下位のそれを上回ること。 |
| 敷居侵害件数 | 敷居を侵害している「健常者等」のサンプル数。 |
| 敷居識別 | 敷居侵害件数が 0 件になること。 |

6) 各目的変数で、残差を標準化し、算出した値。

7) この「除外 Algo」の考え方を、過去に実施した重要な重回帰分析に敷衍してみたところ、どの場合にも適合することが判明した。

| | |
|---------|--|
| 当級侵害 | 軽位クラスの「患者など」が、a「直前クラスの最下位」とb「当該クラスの最下位」との間に入る事象。但し、aの回帰値<bの回帰値のときは「侵害」とせず!! |
| 当級侵害件数 | 当級を侵害している「患者など」のサンプル数。「健常者等」をこれに算入すると、当該サンプルが「敷居侵害件数」と二重計上になるので、注意を要する!! |
| 丸のみ | あるクラスが「直前クラスの最下位」と「当級最下位」との間に、当該クラスよりも軽位なクラスのサンプルを全て含んでしまう事象。 そのときは、丸のみされたクラスの当級侵害件数は0件となる。 |
| 当級識別 | 当級侵害件数が0件になること。 |
| 判別侵害 | 敷居侵害と当級侵害を合わせた呼称。なお、当級が二つしか無い「目変 α 」のような疾病では、二種類の侵害を区別する意味は無い。 |
| 侵害総件数 d | 敷居侵害件数と当級侵害件数を合算した件数。 |
| 判別の確度 D | $1 - \sqrt[3]{\text{侵害総件数 } d / \text{当該 Step のサンプル数 } n}$ の 100 倍（文末の注 f. 参照） |
| 判別 | 侵害総件数が0件になっている事象。即ち、サンプルを回帰値でソートすると、重位クラスから軽位クラスまでサンプルが整列していて、疾病の程度が判る状態になっていること。 このとき、判別確度 D = 100.0 |
| 識別解消 | Step が進み（説明変数の数 q が少なくなつて）、今まで可能であった敷居識別や当級識別が出来なくなった時点で、該「識別」を撤回すること。 なお、両者とも撤回するときは「判別解消」 |

この用語集で理解されるように、時々の回帰式における「判別の確度」は「侵害の総件数」によって算定することが出来る。その際、「敷居侵害と当級侵害の重み」を対等に扱っているが、数千回行った試行研究の結果、それが妥当であるとの結論を得ている。

以上が「判別を定量化」した『判別の確度』である。

§ 5 モデル変遷の概要 と 回帰式の P 値

表 6

Step 0, 17, 34 の 相関係数 と 回帰の精度や判別の確度

| 説明変数との相関係数の絶対値 | | | | | | | | | |
|--------------------|---------|-----------------|--------|--------|-----------------|-----------------|--------------|-------------|-------------|
| | | | 最大 | 平均 | 0.137 超 説変割合 | 回帰の 精度 Q^2 | 回帰式 の P 値 | 侵害総 件数 d | 判別の 確度 D |
| cf. 乱数相関なら | | | 0.2300 | 0.0544 | | | | | |
| 目 変 α | Step 0 | n = 220, q = 72 | 0.1777 | 0.0636 | 8.3% | 0.040 | 27.0% | 2 | 79.1 |
| | Step 17 | n = 212, q = 45 | 0.2020 | 0.0811 | 20.0% | 0.300 | 0.0000% | 0 | 100.0 |
| | Step 34 | n = 209, q = 3 | 0.2052 | 0.1695 | 100% | 0.144 | 0.0067% | 68 | 31.2 |
| 目 変 β | Step 0 | n = 220, q = 72 | 0.1700 | 0.0527 | 4.2% | -0.010 | 55.0% | 9 | 65.5 |
| | Step 17 | n = 212, q = 45 | 0.1893 | 0.0708 | 13.3% | 0.257 | 0.0005% | 2 | 78.9 |
| | Step 34 | n = 209, q = 3 | 0.1962 | 0.1558 | 66.7% | 0.116 | 0.0521% | 95 | 23.1 |
| 目 変 γ | Step 0 | n = 220, q = 72 | 0.1691 | 0.0676 | 8.3% | 0.090 | 9.2% | 17 | 57.4 |
| | Step 17 | n = 212, q = 45 | 0.2187 | 0.0960 | 22.2% | 0.419 | 0.0000% | 6 | 69.5 |
| | Step 34 | n = 209, q = 3 | 0.2230 | 0.2013 | 100% | 0.185 | 0.0004% | 174 | 5.9 |

注) n はサンプル数で, q は説明変数の数である。なお, 乱数相関の場合, $r = 0.23$ を超える割合は 0.074% である。

A, B, C の各モデル群はそれぞれ Step 0 ～ 34 までの 35 steps あるが、概要を把握するため、まずは始点の Step 0 と、中間点の Step 17 と、終点の Step 34 を、相関係数との関係で前頁の表 6 に取り上げている。

目的変数との相関係数 (の絶対値) が最大なのは、始点では A 群の 0.178 で、外乱サンプルを除外し、説変を削り込んだ終点では C 群の 0.223 であった。これらは乱数相関の上限の目安である 0.230 に比して決して高いものではない。また、始点で目変との相関の平均が、最大なのは、C 群の 0.0676 で、最小なのは B 群の 0.0527 であった。これらの事実は、出発時における「回帰の精度」に大きく影響することになる。

次に、Step ごとの回帰の様子や判別の状況などを示す緒元をまとめた「登録カード」なるものを導入し、既述の解析枠で各モデル群がどのように変遷して行ったのかを示すことにする。そこで、登録カードの各欄に登場する項目や記号の意味を、次の表 7 にまとめておく。

表 7 「登録カード」記載項目の解説

| 項 目 | その意味や解説 |
|-------------------|--|
| 符号マルチコ | マルチコ (共線形性) の中でも説明変数と目的変数の相関係数の符号が、当該回帰係数 k_i のそれと異なる場合、このように称し、その数 m を登録する。なお $m = 0$ になれば、その旨を「判別のコメント」欄に記載する。 |
| R^2 と Q^2 | 自由度で修正する前後の決定係数 |
| 回帰式の P 値 | $P \leq 10E-12$ なら [☆☆], $10E-12 < P \leq 10E-06$ なら [☆] $10E-06 < P \leq 10E-03$ なら [***], $0.001 < P \leq 0.01$ なら [**] を次の回帰のコメント欄に登録。 |
| 回帰のコメント | 表 1 に示した修正済み決定係数 Q^2 で区分したレベル名や、上記「回帰式の P 値の記号」を初出に限って登録する他、当該モデル群で修正済み決定係数 Q^2 が最大に達した Step では「Peak」と記載する。 |
| 説変 p 値の最大 | 説明変数の p 値が全て $p \leq 5\%$ なら「all [**]」を、p 値が全て $p \leq 1\%$ なら「all [***]」をモデル・コメント欄に登録。 |
| $p \leq 5\%$ 説変の数 | 定数項も含め、回帰係数 k_i の危険率が 5% 以下である説明変数の数。これを説明変数の数 $q + 1$ で除した値は、「モデルの安定度」を示す。 |
| 説変 F 値の最小 | $0.250 \leq F < 1.000$ なら「T 0.5」, $1.000 \leq F < 4.000$ なら「T 1」, $4.000 \leq F < 9.000$ なら「T 2」をモデル・コメント欄に登録。 |
| モデル・コメント | 「all [***]」や、「T 1」などの登録欄。 |
| S 内外侵害件数 | 「S」は敷居 (Threshold) の略。A モデル群では不要な欄。 B モデル群では「敷居侵害件数 d_0 + 当級侵害件数 d_1 」と表示する。 C モデル群では「 $d_0 + d_1$ + 次の当級侵害件数 d_2 」のように表示する。 |
| 侵害総件数 d | 表 5 のとおり。なお、d は Disturbance の略。 |
| 判別の確度 D | 表 5 のとおり。なお、D は Distinction の略。 |
| 判別のコメント | 「敷居識別」など、疾病の判別に関わる Epoch (初出に限る) を登録。 なお、紙幅の関係で、「符号マルチ無」もこの欄に記載する。 |

次頁の群ごとに示した資料 1 の「登録カード列」は、0 ～ 34 までの計 35 steps のうち、Epoch となった 14 steps (40%) だけを挙げている。紙幅の関係で、残りの Step は付録の資料 2 に廻した。なお以下、グラフを描くときは、全ての Step の該数値を用いている。

資料 1

| Step | | 0 | 2 | 4 | 7 | 12 | 16 | 20 | 21 | 22 | 24 | 29 | 31 | 32 | 34 |
|------------|--|------------|-----------|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|--------------|
| サンプル数 | | 720 | 219 | 218 | 217 | 214 | 212 | 210 | 210 | 209 | 209 | 209 | 209 | 209 | 209 |
| 説変の数 | | 22 | 69 | 66 | 60 | 54 | 48 | 42 | 39 | 39 | 33 | 18 | 12 | 9 | 3 |
| 符号マルチコの数 | | 19 | 14 | 15 | 13 | 10 | 7 | 5 | 4 | 4 | 2 | 1 | 1 | 0 | 0 |
| 決定係数 R^2 | | 0.356 | 0.380 | 0.397 | 0.400 | 0.436 | 0.451 | 0.462 | 0.458 | 0.459 | 0.445 | 0.355 | 0.312 | 0.267 | 0.168 |
| 修正済み Q^2 | | 0.040 | 0.094 | 0.133 | 0.169 | 0.245 | 0.289 | 0.327 | 0.334 | 0.334 | 0.3411 | 0.294 | 0.259 | 0.222 | 0.144 |
| 回帰式の P 値 | | 27.0% | 7.9% | 2.1% | 3.7E-03 | 4.1E-05 | 7.6E-07 | 1.0E-08 | 2.2E-09 | 2.6E-09 | 1.9E-10 | 5.9E-11 | 8.5E-09 | 6.1E-08 | 6.7E-05 |
| 回帰のコメント | | 乱回 | 若干 | ***] | ***] | ***] | 説明, [☆] | | | | Peak | | 参考 | | ***], 若干 |
| 説変 F 値の最大 | | 99.5% | 96.2% | 98.4% | 84.3% | 82.2% | 71.8% | 54.2% | 51.5% | 51.2% | 37.7% | 8.8% | 4.8% | 3.6% | 0.95% |
| D ≤ 5% 説変数 | | 2 | 6 | 7 | 7 | 10 | 12 | 17 | 18 | 18 | 19 | 18 | 12 | 10 | 3 |
| 説変 F 値の最小 | | 0.000 | 0.002 | 0.000 | 0.039 | 0.051 | 0.131 | 0.374 | 0.426 | 0.432 | 1.013 | 2.948 | 3.951 | 4.459 | 6.845 |
| モデル・コメント | | | | | | | | T 0.5 | | | T 1 | | all [☆] | T 2 | all [***] |
| A 侵害総件数 d | | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 10 | 32 | 68 |
| 判別の確度 D | | 90.5 | 90.4 | 90.4 | 90.4 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 86.2 | 78.1 | 60.9 | 43.0 |
| 判別のコメント | | | | | | 判別 | | | | | | 判別解消 | | 符号マルチコ無 | |
| 符号マルチコの数 | | 10 | 9 | 7 | 7 | 5 | 2 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |
| 決定係数 R^2 | | 0.322 | 0.331 | 0.335 | 0.352 | 0.417 | 0.417 | 0.404 | 0.397 | 0.404 | 0.390 | 0.304 | 0.283 | 0.240 | 0.141 |
| 修正済み Q^2 | | -0.010 | 0.021 | 0.044 | 0.103 | 0.215 | 0.256 | 0.268 | 0.269 | 0.267 | 0.2745 | 0.238 | 0.210 | 0.193 | 0.116 |
| 回帰式の P 値 | | 55.0% | 36.5% | 20.9% | 4.7% | 1.9E-04 | 1.8E-05 | 4.2E-06 | 1.5E-06 | 5.8E-07 | 1.04E-07 | 2.4E-08 | 1.8E-07 | 1.9E-06 | 5.2E-04 |
| 回帰のコメント | | 無関 | 乱回 | 若干 | 若干 | 参考, [***] | | | | [☆] | Peak | | | ***] | 若干 |
| 説変 F 値の最大 | | 99.9% | 98.8% | 98.3% | 93.6% | 94.0% | 74.2% | 50.8% | 49.2% | 51.9% | 36.9% | 19.0% | 7.5% | 5.1% | 4.9% |
| D ≤ 5% 説変数 | | 6 | 6 | 6 | 8 | 14 | 14 | 15 | 14 | 13 | 14 | 15 | 12 | 9 | 3 |
| 説変 F 値の最小 | | 0.000 | 0.000 | 0.000 | 0.006 | 0.006 | 0.109 | 0.441 | 0.474 | 0.418 | 1.113 | 1.734 | 3.193 | 3.859 | 6.048 |
| モデル・コメント | | | | | | | | T 0.5 | | | T 1 | | | | T 2, all [☆] |
| S 内外侵害件数 | | 7 + 2 | 5 + 2 | 8 + 2 | 1 + 3 | 0 + 2 | 0 + 2 | 0 + 1 | 0 + 0 | 0 + 0 | 1 + 1 | 12 + 1 | 19 + 0 | 34 + 0 | 89 + 6 |
| B 侵害総件数 d | | 9 | 7 | 10 | 4 | 2 | 2 | 1 | 0 | 0 | 2 | 13 | 19 | 34 | 95 |
| 判別の確度 D | | 79.8 | 82.1 | 78.6 | 86.4 | 90.3 | 90.3 | 93.1 | 100.0 | 100.0 | 90.2 | 75.1 | 69.8 | 59.7 | 32.6 |
| 判別のコメント | | | | | | 数居識別 | | | 判別 | | 判別解消 | 符号マルチコ無 | 当数識別 | | 識別解消 |
| 符号マルチコの数 | | 14 | 14 | 14 | 10 | 10 | 6 | 4 | 3 | 4 | 3 | 0 | 0 | 0 | 0 |
| 決定係数 R^2 | | 0.389 | 0.424 | 0.459 | 0.482 | 0.522 | 0.546 | 0.557 | 0.546 | 0.555 | 0.524 | 0.383 | 0.345 | 0.305 | 0.209 |
| 修正済み Q^2 | | 0.090 | 0.158 | 0.223 | 0.282 | 0.360 | 0.412 | 0.445 | 0.442 | 0.452 | 0.434 | 0.324 | 0.294 | 0.263 | 0.185 |
| 回帰式の P 値 | | 9.2% | 0.98% | 4.5E-04 | 6.9E-06 | 7.3E-09 | 1.1E-11 | 3.7E-14 | 1.4E-14 | 5.3E-15 | 3.1E-15 | 6.3E-13 | 8.3E-11 | 3.1E-10 | 4.4E-06 |
| 回帰のコメント | | 若干 | ***] | ***] | 説明 | [☆] | 実証 | [☆☆] | | Peak | | 説明 | [☆] | 参考 | ***] |
| 説変 F 値の最大 | | 96.6% | 99.0% | 99.1% | 96.2% | 61.6% | 63.2% | 32.4% | 32.5% | 40.5% | 24.0% | 9.5% | 7.4% | 4.8% | 0.5% |
| D ≤ 5% 説変数 | | 7 | 10 | 13 | 13 | 17 | 22 | 28 | 29 | 29 | 29 | 18 | 12 | 10 | 4 |
| 説変 F 値の最小 | | 0.002 | 0.000 | 0.000 | 0.065 | 0.253 | 0.230 | 1.085 | 0.973 | 0.696 | 1.82 | 2.81 | 3.23 | 4.07 | 7.97 |
| モデル・コメント | | | | | | T 0.5 | | T 1 | | | | | | | all [☆], T2 |
| S 内外侵害件数 | | 11 + 1 + 5 | 9 + 1 + 5 | 8 + 1 + 3 | 10 + 2 + 2 | 6 + 1 + 1 | 5 + 1 + 1 | 0 + 2 + 4 | 4 + 0 + 2 | 5 + 0 + 3 | 5 + 1 + 1 | 57 + 6 + 1 | 66 + 2 + 4 | 69 + 2 + 2 | 168 + 6 + 0 |
| C 侵害総件数 d | | 17 | 15 | 12 | 14 | 8 | 7 | 6 | 6 | 8 | 7 | 64 | 72 | 73 | 174 |
| 判別の確度 D | | 0.722 | 0.738 | 0.765 | 0.746 | 0.807 | 0.818 | 0.831 | 0.831 | 0.804 | 0.817 | 0.447 | 0.413 | 0.409 | 0.088 |
| 判別のコメント | | | | | | | | 数居識別 | 識別解消 | | | 符号マルチコ無 | | | |

ところで、疾病別の判別重回帰で鍵となるのは、「回帰の精度」と「判別の確度」の両者であるが、回帰精度の指標 (Index) としては、「修正済み決定係数 Q^2 」と共に、忘れてならないのは「回帰式の P 値」である。これは回帰式の「危うさ」を確率で表現したものであるから、底が「2」の対数を取り「-」を付ければ、その値は「bit」を単位とする情報量になる。次の表 8 の数値は、各「モデルの情報量」を計算したものである。

市販のソフトで目的変数を重回帰した際、各説明変数の回帰係数 k_i の危険率 p 値が 5% 以下のときは「*」が付く。定数項を含め、この「*」ないし「**」が付いた説明変数の数を表 7 では「 $p \leq 5\%$ 説変の数」と称しているが、これも表 8 に一緒に示しておく。

表 8 各モデル、 $p \leq 5\%$ の説変の数と「回帰式の P 値」の情報量 (bit)

| Step | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|------------------|-----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| $p \leq 5\%$ 説変数 | 2 | 2 | 6 | 6 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 10 | 10 | 11 | 13 | 12 | 13 |
| A の情報量 | 1.9 | 2.5 | 3.7 | 4.6 | 5.5 | 6.7 | 6.8 | 8.1 | 7.9 | 9.2 | 10.8 | 12.3 | 14.6 | 16.5 | 18.0 | 20.1 | 20.3 | 22.4 |
| $p \leq 5\%$ 説変数 | 6 | 6 | 6 | 7 | 6 | 6 | 7 | 8 | 8 | 8 | 10 | 11 | 14 | 14 | 14 | 14 | 14 | 14 |
| B の情報量 | 0.9 | 1.2 | 1.5 | 2.0 | 2.3 | 2.7 | 3.5 | 4.4 | 5.5 | 6.6 | 8.3 | 9.7 | 12.3 | 14.1 | 14.1 | 16.0 | 15.8 | 17.6 |
| $p \leq 5\%$ 説変数 | 7 | 10 | 10 | 10 | 13 | 13 | 13 | 13 | 15 | 18 | 17 | 18 | 17 | 16 | 17 | 18 | 22 | 23 |
| C の情報量 | 3.4 | 4.3 | 6.7 | 7.9 | 11.1 | 12.7 | 15.2 | 17.1 | 21.7 | 24.1 | 24.6 | 26.9 | 27.0 | 29.4 | 31.8 | 34.5 | 36.4 | 39.1 |

| Step | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 単位 |
|------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|-----|
| $p \leq 5\%$ 説変数 | 15 | 16 | 17 | 18 | 18 | 19 | 19 | 19 | 19 | 20 | 19 | 18 | 14 | 12 | 10 | 6 | 3 | 個 |
| A の情報量 | 23.8 | 25.8 | 26.5 | 28.8 | 28.5 | 30.5 | 32.3 | 34.1 | 35.5 | 36.6 | 36.3 | 34.0 | 30.4 | 26.8 | 24.0 | 19.5 | 13.9 | bit |
| $p \leq 5\%$ 説変数 | 14 | 15 | 15 | 14 | 13 | 15 | 14 | 16 | 16 | 16 | 15 | 15 | 14 | 12 | 9 | 7 | 3 | 個 |
| B の情報量 | 16.8 | 18.5 | 17.9 | 19.4 | 20.7 | 21.9 | 23.2 | 24.0 | 24.5 | 25.2 | 24.9 | 25.3 | 23.9 | 22.4 | 19.0 | 15.3 | 10.9 | bit |
| $p \leq 5\%$ 説変数 | 28 | 27 | 28 | 29 | 29 | 28 | 29 | 27 | 27 | 23 | 21 | 18 | 15 | 12 | 10 | 7 | 4 | 個 |
| C の情報量 | 42.5 | 44.8 | 44.6 | 46.0 | 47.4 | 47.6 | 48.2 | 47.0 | 46.2 | 43.7 | 41.4 | 40.5 | 39.6 | 33.5 | 31.6 | 26.2 | 17.8 | bit |

次の表 9 は、A、B、C のモデル群ごとに、「 $p \leq 5\%$ 説変の数」と時どきの「Step」を説明変数とし、「 $-\log_2 P$ 」を重回帰分析した結果の緒元である。

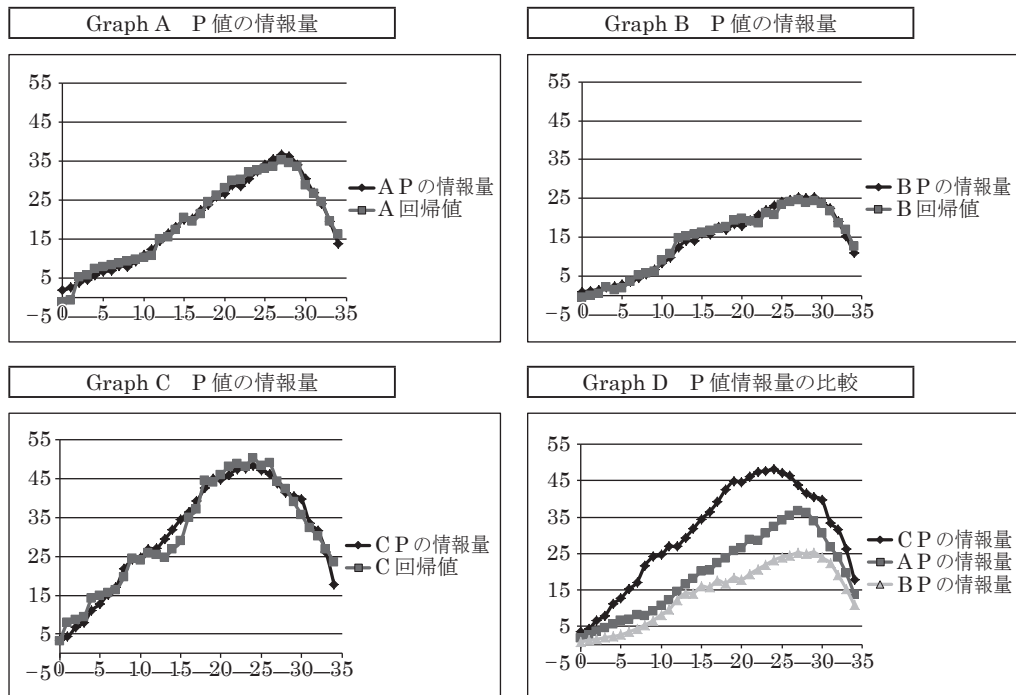
表 9 「回帰式の P 値」の $-\log_2 P$ に対する二変数重回帰分析

| A 群 重 回 | 相関行列 | Step | $p \leq 5\%$ 説変数 | P の情報量 | 分析精度 | | 回帰係数 | |
|------------------|------------------|-------|------------------|--------|------------|-------|------------------|--------|
| | Step | 1 | 0.599 | 0.834 | 決定係数 R^2 | 0.984 | Step | 0.467 |
| | $p \leq 5\%$ 説変数 | 0.599 | 1 | 0.930 | 修正済み Q^2 | 0.983 | $p \leq 5\%$ 説変数 | 1.327 |
| | A P の情報量 | 0.834 | 0.930 | 1 | 修正済み相係 | 0.992 | 定数項 | -3.822 |
| B 群 重 回 | 相関行列 | Step | $p \leq 5\%$ 説変数 | P の情報量 | 分析精度 | | 回帰係数 | |
| | Step | 1 | 0.478 | 0.865 | 決定係数 R^2 | 0.980 | Step | 0.490 |
| | $p \leq 5\%$ 説変数 | 0.478 | 1 | 0.837 | 修正済み Q^2 | 0.979 | $p \leq 5\%$ 説変数 | 1.175 |
| | B P の情報量 | 0.865 | 0.837 | 1 | 修正済み相係 | 0.990 | 定数項 | -7.555 |
| C 群 重 回 | 相関行列 | Step | $p \leq 5\%$ 説変数 | P の情報量 | 分析精度 | | 回帰係数 | |
| | Step | 1 | 0.272 | 0.714 | 決定係数 R^2 | 0.969 | Step | 0.720 |
| | $p \leq 5\%$ 説変数 | 0.272 | 1 | 0.846 | 修正済み Q^2 | 0.967 | $p \leq 5\%$ 説変数 | 1.353 |
| | C P の情報量 | 0.714 | 0.846 | 1 | 修正済み相係 | 0.983 | 定数項 | -6.293 |

「回帰式の P 値」の対数を取っただけで（説明変数との相関が上がり）、自由度修正済み決定係数 Q^2 が格段に高くなることが確認できるが、これをグラフに示すと、図 4 のようになる。Graph A, B, C は各モデル群で「P 値情報量の実値 U_i と回帰値 V_i 」を重ねたものであるが、これから次のような知見や示唆が得られた。

- ① 「P 値情報量の実値 U_i 」は「Step」と「 $p \leq 5\%$ 説変の数」とで、ほぼ正確に 1 次式で近似できる。
- ② 中間点直前あたりまでに外乱サンプルを既述の「除外 Algo」で解析枠から外すと共に、三変数刻みで残差絶対値の大きい説明変数を削除して行く Step 構成が妥当である。
- ③ A モデル群と B モデル群の Step 27 と C モデル群の Step 24 で、実値 U_i や回帰値 V_i が尖頭になるが、その 2 ～ 3 steps 前で「回帰の精度」もピークに達しそうである。

図 4

「回帰式の P 値」の $-\log_2 P$ の重回帰グラフ

§ 6 疾病の判定，回帰精度と判別確度 D

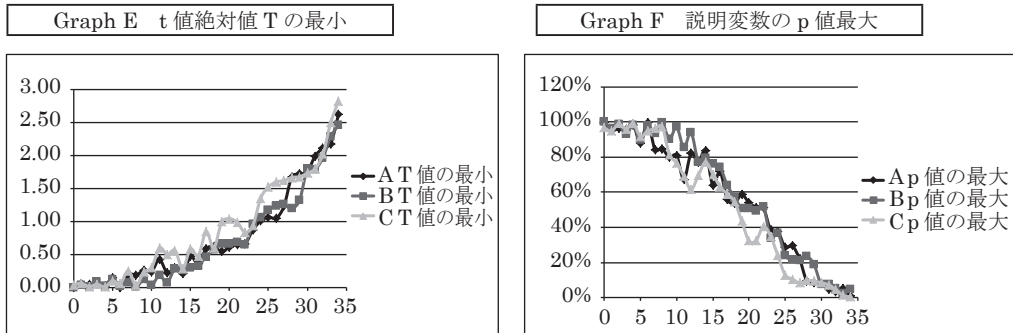
回帰精度のみならず判別確度にも，説明変数の F 値（や p 値）が関係していることは，その定義から容易に推定される。各説明変数の t 値は回帰係数 k_i をその標準誤差で除したもので，F 値はそれを二乗したものである。従って，F の平方根は t 値の絶対値になるが，本論ではこの値を単に該説明変数の「T 値」と呼ぶことにする。この T 値は，その回帰式における当該

説明変数の意義の大きさ（重要性）を表している。

そこで各モデル群別に、T 値の最小値と、それに大きく連動する説明変数の p 値の最大値を図 5 に示しておく。

図 5

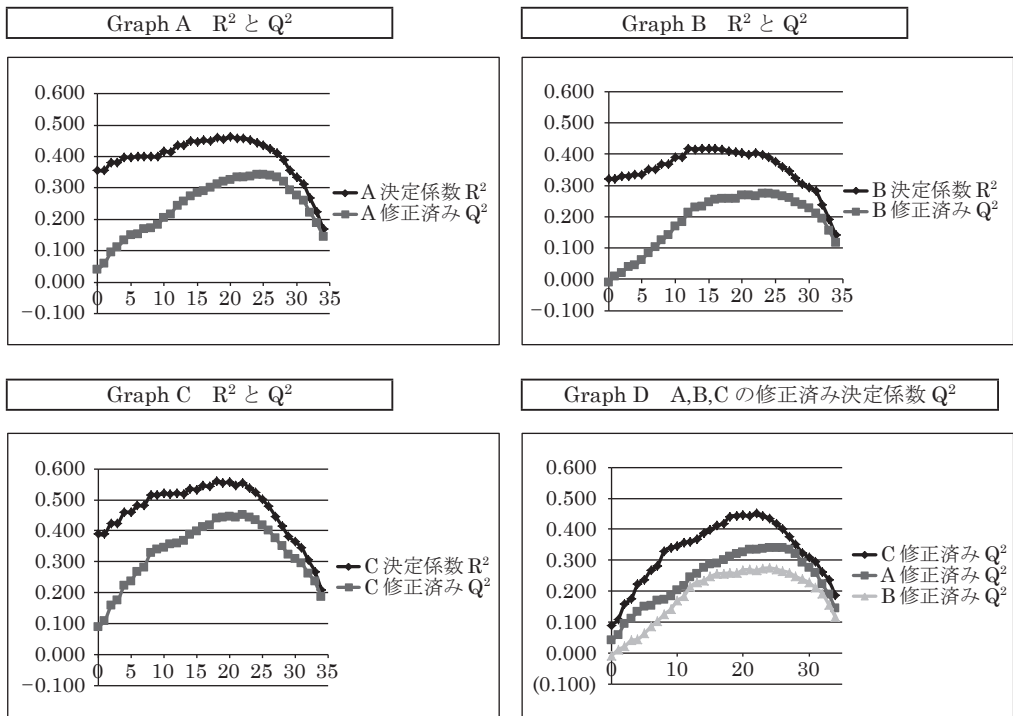
t 値絶対値の最小と説変 p 値の最大



いずれのモデル群においても、T 値の最小値は下に凹の右肩上がり、p 値の最大値はガウス曲線の右半分のような弧を描いている。

これらを前提に、続いて問題の「回帰の精度」を資料 1 と次の図 6 とで考察してみる。

図 6

自由度で修正する前後の決定係数 R^2 と Q^2 

自由度修正済み決定係数の最大なるものはCモデル群 Step 22 の $Q^2 = 0.4520$ で、ラベル名は「実証」である。「患者など」を除外できないという重大な制約がある中で、乱数モデルでここまで回帰できたのは、「目変 γ 」が正規乱数であったためと思われる。疾病の判定では、説明変数の数 q は多くても一向に構わない。 $q = 39$ を度外視すれば、アンケート・データなら十分に役立つ水準に達している。しかし、疾病の判定には到底及ばない。

未だライブなデータによる検証には至っていないが、疾病の難度によっては修正済み決定係数 Q^2 にかかなり高い値が要求されると思われる。そこで、筆者が予測している『判定基準』を、「軽病」、「中病」、「重病」と「難病」とに Rank 分け⁸⁾ して、次の表 10 に示しておく。

表 10

疾病の Rank による判別回帰の判定基準

| | | | | | | | | | |
|----------------|---------|----|---------|----|---------|----|---------|----|--------|
| 修正済み決定係数 Q^2 | 0.5550 | 軽病 | 0.6300 | 中病 | 0.7056 | 重病 | 0.7810 | 難病 | 1 |
| 判別の確度 D | 84.0 p. | | 90.0 p. | | 95.0 p. | | 98.0 p. | | 100 p. |

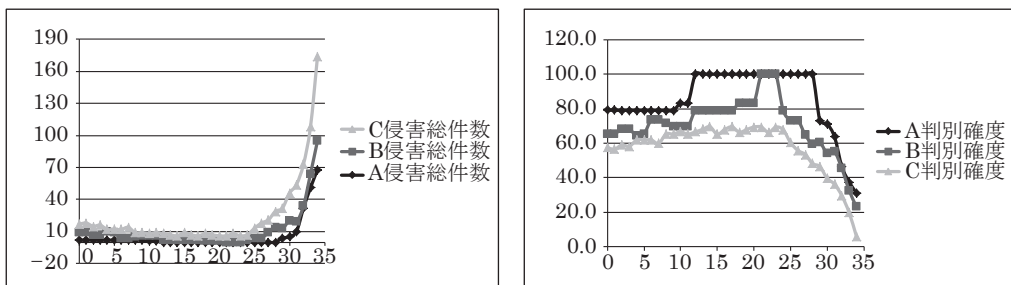
なお、Aモデル群やBモデル群で回帰精度が最大になるのは、いずれも Step 24 で、 Q^2 はそれぞれ 0.3411 と 0.2745 であった。

「回帰の精度」が如何に高くとも「判別の確度」が低ければ、『疾病の判定 (Judge)』には用をなさない。即ち、表 10 の Q^2 と D に対する判定基準は AND で要求されていると考えている。軽病の場合は 244 件に 1 件、中病の場合は 1,000 件に 1 件、重病の場合は 8,000 件に 1 件、難病の場合には 125,000 件に 1 件しか数居侵害ないし当級侵害を許容しないものと思われる。表 10 における各疾病 Rank の確度 D の基準は、この割合に準じて予測したものである。

判別確度 D は当然、 $X = \text{侵害総件数 } d / \text{サンプルサイズ } n$ の関数で、 $0 \leq X \leq 1$ である。

さらに、表 5 で紹介した $D = 100 \times (1 - \sqrt[3]{X})$ なる判別確度の定義は、筆者の疾病判別に対する感覚を表現した数式である。この定義式には X が 0 に近づくほど、違いが際立って来る利点がある。これを図 7 のグラフで見ることにする。

図 7

モデル群別にみた侵害総件数 d と判別の確度 D

8) 同じ疾病の中で「軽症」、「重症」と区分するのは別の概念。本論では「難病」でない疾病そのものを、さらに「軽病」、「中病」、「重病」に区分している。しかし、何がそれに該当する疾病なのか、医師でない筆者には判りかねます。

A モデル群では、実に約半数の Step 12 ~ 28 で侵害総件数が 0 となり、判別確度が D = 100 ポイントである。この疾病 α のような当級が罹患と健常の二つしか無い場合は、判別確度が比較的高くなるものと思われる。しかし、疾病の Rank が「軽病」であったとしても、先に指摘した回帰精度が低いために、罹患しているか否か判定できないことになる。(このように、判別重回帰では、「判別」と「判定」を使い分けていることに注意されたい!!)

B モデル群では、Step 12 で「敷居識別」に至り、Step 21 では更に「当級識別」も出来るようになって「判別」に至った。しかし、それから僅か 3 steps 後の Step 24 では「判別解消」のやむなきに至り、そこから後の Step では破局的な事態に落ちいつている。それは A モデル群でも Step 29 以降に生起しているが、この Catastrophic Phenomenon は乱数変数固有のことのように思われる。

C モデル群では、Step 20 で「敷居識別」になるが、直後の Step 21 には早くも「識別解消」になっている。また、破局的現象も Step 25 から生起しており、その落ち込みようは他のモデル群の比ではなく、終点の Step 34 における判別の確度 D は僅か 5.9 ポイントである。回帰の精度は他のモデル群より高かったが、判別の確度は低くなった。これはどうも目的変数が正規乱数であることに起因しているようである。目変が連続型の数値の方が回帰の精度が上がるようだが、判別の確度の方はそうも云えそうにない。

以上、『疾病の判定 (Judge of Diagnosis)』に、「回帰の精度 (Accuracy of Regression)」と「判別の確度 (Reliability of Distinction)」を使い、乱数を用いた試行研究により、各疾病 Rank に対する『判定基準 (表 10)』を予測したものである。

素診断はあくまでも素診断であって、正規には医師の診断によるが、本論で提起した素診断はその正規の診断のために、必要にして十分な解析情報の提供を目指すものである。

お わ り に

筆者は、多変量解析で有力な分析手法は、「主成分分析」と「重回帰分析」と「正準判別分析」だと考えている。平行する複数の判別平面で多次元空間を区分する訳だから、正準判別分析を疾病の判別に応用することが当然考えられる。しかし今のところ、複数の疾病を同時に解析する正準判別分析の手法を開発し得ていない。協力して頂ける医療機関が出てきたら、ライブなデータでこれを是非開発したいと思う。

高騰する国民医療費を抑制するためにも、国の事業として、疾病群別の判別重回帰を開発する必要がある。しかし、当面は「病院原価計算・原価管理研究会⁹⁾」に関係する方々の協力を得て、

9) 会長を岩崎榮氏 (卒後臨床研修評価機構専務理事) が、理事長を竹田秀氏 (竹田総合病院理事長) が、副理事長を相田俊夫氏 (倉敷中央病院理事長) と福島公明氏 (淀川キリスト教病院常任理事) が務める 2009 年 7 月に設立された研究会。

細々と研究を続けるよりない。ライブなデータについては理事の渡辺明良氏（聖路加国際病院，マネジャー）にご相談し，これからも，同研究会の理事兼事務局長である田原隆氏（元日本福祉大学教授，精神科医）や，会員の日月裕氏（同大学教授，麻酔科医），星雅丈氏（成美大学准教授，診療情報管理士）のご協力を願いたいと考えている。

この小論で示した重回帰分析はエスミ社の「多変量解析 ver.6」によるが，ライブなデータによる研究の継続には，独自のプログラム開発を要する。それには松本祐輔氏（ソシオアート代表取締役）の協力が不可欠である。また議論の相手として，山本友太氏（淀川キリスト教病院，職員），川瀬友太氏（関西大学，職員），奥山武生氏（松下記念病院，職員）のお三方にも引き続き協力を願えるものと確信している。

文末の注

- a. 一般電気事業者 10 社の電灯・電力の売上高から沖縄電力を除いているのは，同社が水力も原子力も持たず 100% 火力で，更に他社との地帯間融通も行っていないからである。次の表 11 は『電気事業便覧 2011 年版』から調整したものである。

| 表 11 | | 一般電気事業者の供給単価 | |
|---------|-------------|---------------------|---------------|
| 一般電気事業者 | 電灯・電力料（百万円） | 使用電灯・電力量（100 万 kwh） | 供給単価（円 / kwh） |
| 9 社計 | 14,265,324 | 898,896 | 15.87 |
| 沖縄電力 | 149,683 | 7,521 | 19.90 |

ちなみに，経済三団体は原子力を全廃すると電力料金が倍増すると主張しているが，これで見限り，16 円 / kwh が 20 円 / kwh になる位のものである。

- b. 高額な医療機器は装置そのものだけでなく，施設の建設・維持にも資金を要する。PET（陽電子断層撮影）や MRI（磁気共鳴装置）などがその代表的な例で，前者は人工的に作ったプラスの電荷を持った反電子（Positron）と普通の電子の対消滅を利用して，がん診療に役立てる機器。後者は Magnetic Resonance Imaging の略であるが，医療界ではかつてプロパーと呼んでいた製薬会社の Medical Representative と略字が似ている。
- c. 疾病 α の罹患者を No.1 ～ 10 まで，疾病 β の患者と容疑者を等級ごとに No.10 ～ 20 まで，疾病 γ の「患者など」を No.20 ～ 33 まで，それぞれ γ 値の降順に付番し直している。なお，No.10 と No.20 は合併疾患であるが，確率から見て，それがこの程度の割合で発生するものと思われる。
- d. § 2 末に記述した乱数相関の標準偏差 $\sigma = 0.068180$ と，その絶対値の平均 $\mu = 0.054400$ と標準偏差 $\sigma' = 0.041100$ との間で， $2.01 \sigma = \mu + 2.01 \sigma' = 0.1370$ となる。この値は乱

数相関の上限である $r = 0.2300$ に比してさほど高くなく、頃合いであるので、当該乱数変数を「意義ある説変」とみなすことにした。

- e. 解析枠で「健常者等」から外乱サンプルを除外する Algorithm につき、有力な手法として、標準化残差ないしその絶対値に対する主成分得点が考えられる。いくつかの主成分得点の加重平均なら、どんな疾病群でも妥当する「除外 Algo」を開発したが、煩雑を究めるので、本論ではその代理的な算定法に依存した。
- f. 侵害総件数 d / サンプル数 n を X とすると、判別確度 D を定義する際、誰しも $1 - X^r$ または、 $(1 - X)^r$ をベースにするであろう。ここに r は然るべき有理数であるが、筆者は後に研究する予定の正準判別分析の関係で、 D を $1 - \sqrt[3]{X}$ (のスカラー倍) としたが、 X の 0 近辺で据わりのいい判別関数になった。
- g. 疾病の Rank (重度) に依存して、判別に漏れる件数割合が決まってくる。軽、中、重病の場合はそれぞれ 244.14 件、1,000 件、8,000 件中の 1 件で肯えるが、「難病」の場合は乱数による試行では判然としない嫌いがある。しかし、125,000 件に 1 件という判別漏れは、素診断ではやむを得ない、と筆者は考えている。

資料 2・2/3

各群の変遷 (登録カード 12 ~ 23 Step)

| Step | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| サンプル数 | 214 | 214 | 213 | 213 | 212 | 212 | 211 | 211 | 210 | 210 | 209 | 209 |
| 説変の数 | 54 | 51 | 51 | 48 | 48 | 45 | 45 | 42 | 42 | 39 | 39 | 36 |

| | | | | | | | | | | | | |
|-----------|------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| A モデル群 | 符号マルチロの数 | 10 | 9 | 9 | 7 | 7 | 5 | 4 | 5 | 4 | 4 | 3 |
| | 決定係数 R^2 | 0.436 | 0.435 | 0.449 | 0.447 | 0.451 | 0.449 | 0.456 | 0.462 | 0.458 | 0.459 | 0.452 |
| | 修正済み Q^2 | 0.245 | 0.258 | 0.274 | 0.286 | 0.289 | 0.300 | 0.320 | 0.327 | 0.334 | 0.334 | 0.338 |
| | 回帰式の P 値 | 4.1E-05 | 1.1E-05 | 3.7E-06 | 9.2E-07 | 7.6E-07 | 1.8E-07 | 1.7E-08 | 1.0E-08 | 2.2E-09 | 2.6E-09 | 6.7E-10 |
| | 回帰のコメント | | | | 説明, [☆] | | | | | | | |
| | 説変 p 値の最大 | 82.2% | 76.9% | 83.7% | 64.0% | 71.8% | 55.8% | 58.7% | 54.2% | 51.5% | 51.2% | 38.5% |
| | $p \leq 5\%$ 説変数 | 10 | 10 | 11 | 13 | 12 | 13 | 16 | 17 | 18 | 18 | 19 |
| | 説変 F 値の最小 | 0.051 | 0.086 | 0.042 | 0.220 | 0.131 | 0.345 | 0.296 | 0.374 | 0.426 | 0.432 | 0.760 |
| | モデル・コメント | | | | | | T 0.5 | | | | | |
| | A 侵害総件数 d | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 判別の確度 D | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | 判別のコメント | | | | | | | | | | | |

| | | | | | | | | | | | | |
|-----------|------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| B モデル群 | 符号マルチロの数 | 5 | 5 | 4 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 |
| | 決定係数 R^2 | 0.417 | 0.416 | 0.418 | 0.417 | 0.417 | 0.415 | 0.411 | 0.406 | 0.404 | 0.397 | 0.399 |
| | 修正済み Q^2 | 0.215 | 0.229 | 0.234 | 0.247 | 0.256 | 0.257 | 0.259 | 0.258 | 0.268 | 0.267 | 0.2735 |
| | 回帰式の P 値 | 1.9E-04 | 5.6E-05 | 5.6E-05 | 1.5E-05 | 1.8E-05 | 4.9E-06 | 8.5E-06 | 2.8E-06 | 4.2E-06 | 1.5E-06 | 5.8E-07 |
| | 回帰のコメント | [***] | | | | | | | | | | [☆] |
| | 説変 p 値の最大 | 94.0% | 77.3% | 79.3% | 76.1% | 74.2% | 64.1% | 57.0% | 50.8% | 50.8% | 49.2% | 33.9% |
| | $p \leq 5\%$ 説変数 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 15 | 15 | 14 | 15 |
| | 説変 F 値の最小 | 0.006 | 0.083 | 0.069 | 0.092 | 0.109 | 0.218 | 0.323 | 0.440 | 0.441 | 0.474 | 0.919 |
| | モデル・コメント | | | | | | | T 0.5 | | | | |
| | S 内外侵害件数 | 0 + 2 | 0 + 2 | 0 + 2 | 0 + 2 | 0 + 2 | 0 + 2 | 0 + 1 | 0 + 1 | 0 + 1 | 0 + 0 | 0 + 0 |
| | B 侵害総件数 d | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 |
| | 判別の確度 D | 78.9 | 78.9 | 78.9 | 78.9 | 78.9 | 78.9 | 83.2 | 83.2 | 83.2 | 100.0 | 100.0 |
| | 判別のコメント | 数居識別 | | | | | | | | | 判別 | |

| | | | | | | | | | | | | |
|-----------|------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| C モデル群 | 符号マルチロの数 | 10 | 10 | 9 | 7 | 6 | 5 | 4 | 4 | 3 | 4 | 3 |
| | 決定係数 R^2 | 0.522 | 0.520 | 0.535 | 0.534 | 0.546 | 0.543 | 0.561 | 0.556 | 0.557 | 0.546 | 0.539 |
| | 修正済み Q^2 | 0.360 | 0.369 | 0.388 | 0.397 | 0.412 | 0.419 | 0.441 | 0.445 | 0.445 | 0.442 | 0.442 |
| | 回帰式の P 値 | 7.3E-09 | 1.5E-09 | 2.7E-10 | 4.2E-11 | 1.1E-11 | 1.7E-12 | 1.6E-13 | 3.2E-14 | 3.7E-14 | 1.4E-14 | 5.3E-15 |
| | 回帰のコメント | | | | | 実証 | | [☆☆] | | | | Peak |
| | 説変 p 値の最大 | 61.6% | 69.4% | 77.2% | 70.2% | 63.2% | 59.0% | 54.8% | 43.6% | 32.4% | 32.5% | 35.0% |
| | $p \leq 5\%$ 説変数 | 17 | 16 | 17 | 18 | 22 | 23 | 28 | 27 | 28 | 29 | 28 |
| | 説変 F 値の最小 | 0.253 | 0.315 | 0.084 | 0.344 | 0.230 | 0.714 | 0.363 | 0.986 | 1.085 | 0.973 | 0.878 |
| | モデル・コメント | | | | | | | | | T 1 | | |
| | S 内外侵害件数 | 6 + 1 + 1 | 4 + 1 + 2 | 4 + 1 + 1 | 5 + 2 + 2 | 5 + 1 + 1 | 4 + 1 + 1 | 3 + 3 + 2 | 1 + 2 + 4 | 0 + 2 + 4 | 4 + 0 + 2 | 2 + 0 + 4 |
| | C 侵害総件数 d | 8 | 7 | 6 | 9 | 7 | 6 | 8 | 7 | 6 | 6 | 6 |
| | 判別の確度 D | 66.6 | 68.0 | 69.6 | 65.2 | 67.9 | 69.5 | 66.4 | 67.9 | 69.4 | 69.4 | 69.4 |
| | 判別のコメント | | | | | | | | | 数居識別 | 識別解消 | |

資料 2・3/3

各群の変遷（登録カード 24 ～ 35 Step）

| Step | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 31 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| サンプル数 | 209 | 209 | 209 | 209 | 209 | 209 | 209 | 209 | 209 | 209 | 209 | 209 |
| 説変の数 | 33 | 30 | 27 | 24 | 21 | 18 | 15 | 12 | 9 | 6 | 3 | 12 |

| | | | | | | | | | | | | |
|---------------|------------|---------|---------|---------|----------|---------|---------|---------|---------|---------|---------|----------|
| A モデル 群 | 符号マルチロの数 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| | 決定係数 R^2 | 0.445 | 0.436 | 0.425 | 0.411 | 0.389 | 0.355 | 0.333 | 0.312 | 0.267 | 0.224 | 0.168 |
| | 修正済み Q^2 | 0.3411 | 0.3410 | 0.339 | 0.334 | 0.320 | 0.294 | 0.2761 | 0.259 | 0.1888 | 0.144 | 0.259 |
| | 回帰式の P 値 | 1.9E-10 | 5.6E-11 | 2.0E-11 | 9.90E-12 | 1.2E-11 | 5.9E-11 | 6.8E-10 | 8.5E-09 | 6.1E-08 | 1.3E-06 | 6.7E-05 |
| | 回帰のコメント | Peak | | | | | | | | ***] | 若干 | 参考 |
| | 説変 p 値の最大 | 37.7% | 28.7% | 29.5% | 22.0% | 9.5% | 8.8% | 8.1% | 4.8% | 3.6% | 5.5% | 0.95% |
| | p ≤ 5% 説変数 | 19 | 19 | 19 | 20 | 19 | 18 | 14 | 12 | 10 | 6 | 3 |
| | 説変 F 値の最小 | 1.013 | 1.138 | 1.104 | 1.513 | 2.808 | 2.948 | 3.077 | 3.951 | 4.459 | 4.739 | 6.845 |
| | モデル・コメント | T1 | | | | | | | all [*] | T2 | | all [**] |
| | A 侵害総件数 d | 0 | 0 | 0 | 0 | 0 | 4 | 5 | 10 | 32 | 51 | 68 |
| | 判別の確度 D | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 73.3 | 71.2 | 63.7 | 46.5 | 37.5 | 31.2 |
| | 判別のコメント | | | | | | 判別解消 | | 符号マルチ無 | | | |

| | | | | | | | | | | | | |
|---------------|------------|----------|---------|---------|---------|---------|---------|---------|---------|-------------|---------|---------|
| B モデル 群 | 符号マルチロの数 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 決定係数 R^2 | 0.390 | 0.376 | 0.361 | 0.345 | 0.323 | 0.304 | 0.294 | 0.283 | 0.240 | 0.193 | 0.141 |
| | 修正済み Q^2 | 0.2745 | 0.271 | 0.265 | 0.260 | 0.247 | 0.238 | 0.227 | 0.210 | 0.193 | 0.156 | 0.116 |
| | 回帰式の P 値 | 1.04E-07 | 5.9E-08 | 4.1E-08 | 2.6E-08 | 3.1E-08 | 2.4E-08 | 6.5E-08 | 1.8E-07 | 1.9E-06 | 2.5E-05 | 5.2E-04 |
| | 回帰のコメント | Peak | | | | | | | | ***] | 若干 | |
| | 説変 p 値の最大 | 36.9% | 24.0% | 21.6% | 21.1% | 23.4% | 19.0% | 7.5% | 7.5% | 5.1% | 2.4% | 4.9% |
| | p ≤ 5% 説変数 | 14 | 16 | 16 | 16 | 15 | 15 | 14 | 12 | 9 | 7 | 3 |
| | 説変 F 値の最小 | 1.113 | 1.388 | 1.539 | 1.575 | 1.426 | 1.734 | 3.211 | 3.193 | 3.859 | 5.193 | 6.048 |
| | モデル・コメント | T1 | | | | | | | | all [*], T2 | | |
| | S 内外侵害件数 | 1 + 1 | 2 + 2 | 1 + 3 | 12 + 2 | 6 + 3 | 12 + 1 | 19 + 1 | 19 + 0 | 34 + 0 | 61 + 3 | 89 + 6 |
| | B 侵害総件数 d | 2 | 4 | 4 | 9 | 14 | 13 | 20 | 19 | 34 | 64 | 95 |
| | 判別の確度 D | 78.8 | 73.3 | 73.3 | 64.9 | 59.4 | 60.4 | 54.3 | 55.0 | 45.4 | 32.6 | 23.1 |
| | 判別のコメント | 判別解消 | | | | | 符号マルチ無 | | 当破識別 | | 識別解消 | 当破識別 |

| | | | | | | | | | | | | |
|---------------|------------|-----------|-----------|------------|------------|------------|------------|------------|------------|-------------|-------------|-------------|
| C モデル 群 | 符号マルチロの数 | 3 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 決定係数 R^2 | 0.524 | 0.501 | 0.479 | 0.447 | 0.415 | 0.383 | 0.364 | 0.345 | 0.305 | 0.268 | 0.209 |
| | 修正済み Q^2 | 0.434 | 0.417 | 0.401 | 0.375 | 0.350 | 0.324 | 0.309 | 0.294 | 0.263 | 0.235 | 0.185 |
| | 回帰式の P 値 | 3.1E-15 | 7.0E-15 | 1.2E-14 | 7.0E-14 | 3.4E-13 | 6.3E-13 | 1.2E-12 | 8.3E-11 | 3.1E-10 | 1.3E-08 | 4.4E-06 |
| | 回帰のコメント | | | 説明 | | | | [☆] | | 参考 | | ***] |
| | 説変 p 値の最大 | 24.0% | 12.9% | 10.7% | 8.4% | 9.8% | 9.5% | 8.5% | 7.4% | 4.8% | 1.4% | 0.5% |
| | p ≤ 5% 説変数 | 29 | 27 | 27 | 23 | 21 | 18 | 15 | 12 | 10 | 7 | 4 |
| | 説変 F 値の最小 | 1.82 | 2.32 | 2.55 | 2.62 | 2.77 | 2.81 | 2.999 | 3.23 | 4.07 | 6.21 | 7.97 |
| | モデル・コメント | | | | | | | | | all [*], T2 | | all [**] |
| | S 内外侵害件数 | 5 + 1 + 1 | 8 + 2 + 3 | 12 + 3 + 3 | 16 + 2 + 3 | 25 + 2 + 2 | 28 + 3 + 1 | 42 + 3 + 1 | 50 + 2 + 2 | 69 + 2 + 2 | 100 + 8 + 0 | 168 + 6 + 0 |
| | C 侵害総件数 d | 7 | 13 | 18 | 21 | 29 | 32 | 46 | 54 | 73 | 108 | 174 |
| | 判別の確度 D | 67.8 | 60.4 | 55.8 | 53.5 | 48.2 | 46.5 | 39.6 | 36.3 | 29.6 | 19.8 | 5.9 |
| | 判別のコメント | | | | | 符号マルチ無 | | | | | | |

