福山平成大学経営学部紀要
第１３号（2016），*－**頁

社会システム分析のための統合化プログラム２８
－メタ分析・ロジスティック回帰分析－

福井正康[*1]、小玉一樹[*1]、尾崎誠[*1]
[*1] 福山平成大学経営学部経営学科

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は複数の研究資料をまとめてより強固な結論に導くための手法であるメタ分析や、2 項分布や多項分布の確率を説明変数の多項式の関数として求めるロジスティック回帰分析について、プログラムを作成した。この論文ではこれらの分析の理論を説明し、具体的なプログラムの利用法を紹介する。
キーワード：College Analysis，多変量解析，一般化線形モデル，メタ分析，ロジスティック回帰分析
URL：http://www.heisei-u.ac.jp/ba/fukui/

1. はじめに

我々はこれまで社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は、有意差検定を行った複数の研究資料を元に、主旨の同じ検定を集めて検定結果をより強固なものにしようと試みるメタ分析及び、一般化線形モデルに基づくロジスティック回帰分析についてプログラムを作成した。

メタ分析で問題となるのは、主旨の同じ検定でも論文によっては、検定方法が異なる場合があることである。例えば、2 つの量の関係を見る検定では、2 つの量からそのまま相関係数を求めて検定することもできるし、一方を２つの群に分けて 2 群の差の検定として見ることもできる。また、効率は悪いが両方を２つの群に分けて χ^2 検定等を行うこともできる。メタ分析では、これらの異なる検定から効果量を取り出し、それらの効果量を変換して一種類の効果量にする。その後、それらをまとめて、複数の検定結果を統合した検定結果を導出する。

このプログラムは参考文献[1] を元に作成したが、効果量として標準化平均値差、バイアス修正標準化平均値差、対数オッズ比、相関係数が取り上げられている。これらの他に、論文で扱われる指標には t 統計量や χ^2 統計量などがあるが、t 統計量は標準化平均値差に簡単に変換できる。χ^2 統計量は少なくとも分割表の非対角要素の値が分からなければ変換が困難である。

一般化線形モデルの 2 項分布モデルは、2 項分布の確率パラメータを説明変数の線形結合の関数で予測するモデルで、連結関数と呼ばれる予測の関数の形によっていろいろなモデルがあるが、ここでは、ロジスティックモデル、プロビットモデル、極値モデルを取り上げている。

分析名としては代表的なロジスティック回帰分析の名前を用いている。同様に一般化線形モデルの多項分布モデルは、多項分布の確率パラメータを説明変数の線形結合の関数で予測するモデルで、プログラムには名義尺度に対する名義ロジスティックモデルと順序尺度に対する累積ロジットモデルを取り上げている。ここでは今後のプログラム作成者の利用を考えて、一般化線形モデルの基礎から話を始めて、具体的なモデルを解説している。

この報告では、各章が独立しているため、式番号、図表番号は章ごとに付けるものとする。章を越えて参照する場合は、その旨を示す。

2. メタ分析

メタ分析は、多くの研究資料から同一の調査内容を選び出し、それらを統合して結果をより強固なものにしようとする分析手法である。１つの研究資料からは、効果量と呼ばれる統計量と、その分散またはデータ数を取り出す。代表的な効果量には標準化された平均値差、オッズ比、相関係数などがある。しかし、研究資料ごとにこれらが同じである保証はないので、必要があれば、これらを統一的な効果量に変換する。その後、各研究資料にウェイトをかけて、研究で与えられた結果が保証されるかどうか検討する。この一連の手法をメタ分析という[1]。

我々はこの過程を計算するプログラムの開発を考えた。ここでは、参考文献 [1] に従い、効果量の入力、効果量の変換、統計的分析に分けて、理論的にどのような式が使われているのかをまとめて紹介する。

2.1 効果量とその入力

我々がプログラムの中で扱う効果量は以下にまとめて示す。種々の研究資料には効果量（または検定値）とデータ数は記載されているが、効果量の分散が記載されていないことが多い。また、参考文献[1]では、後の統計的分析のために分散は記載されているが、データ数が記載されていない。これらの状況に対処するために、我々は結果表示に必要なデータは何か、またそれを得るためにはどのようなデータが必要かを検討した。結論は、比較的良い近似として、結果表示に必要なデータは、効果量と全データ数または、効果量と分散であった。ここでは、効果量と、全データ数または分散のどちらかが分かっているものとして、他方を求める近似式を与えておく。但しこの結果には関数 $y = 1/x(1-x)$ の性質を利用している。

１）標準化平均値差 d （ヘッジスの g とも呼ばれる）

効果量： $d = \dfrac{\overline{x}_1 - \overline{x}_2}{u_{pooled}}$, $u_{pooled} = \sqrt{\dfrac{(n_1 - 1)u_1^2 + (n_2 - 1)u_1^2}{n_1 + n_2 - 2}}$ （pooled 標準偏差）

分散： $V_d = \dfrac{n_1 + n_2}{n_1 n_2} + \dfrac{d^2}{2(n_1 + n_2)}$

全データ数： $N = n_1 + n_2$

変換近似式

$$d,\,N \to V_d \quad \text{のとき、} V_d \simeq \frac{4\alpha + d^2/2}{N}$$

$$d, V_d \to N \quad \text{のとき、} N \simeq \frac{4\alpha + d^2/2}{V_d}$$

ここで、分散の $(n_1+n_2)/n_1n_2 = N/n_1(N-n_1)$ の項については、例えば、$N=100$ とすると、n_1 の関数として図１のようなグラフとなる。

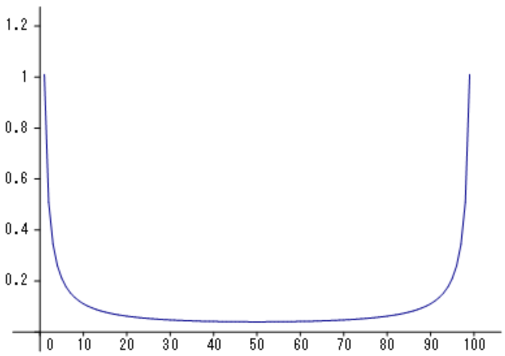


図１　$y = 100/x(100-x)$

このグラフは、中央部で $4/N$ に近いほぼ安定な値を取っており、変動は少ないと考えられる。そこで我々は、この関数の $x = N/2$ の値を中心とした正規分布による加重平均を考え、その結果を $(n_1+n_2)/n_1n_2 \simeq 4\alpha/N$ とした。

α の値については、以下のように計算した。

$$\alpha = \frac{1}{4A}\int_{0.1}^{0.9} \frac{1}{x(1-x)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-0.5)^2}{2\sigma^2}\right] dx$$

$$A = \frac{1}{\sqrt{2\pi}\sigma}\int_{0.1}^{0.9} \exp\left[-\frac{(x-0.5)^2}{2\sigma^2}\right] dx$$

この場合、例えば、$\sigma = 0.2$ （約 95%の研究で $0.1N < n_1 < 0.9N$ ）とすると、$\alpha = 1.187$ となる。我々はこの値を利用する。実際の調査では、2 群間の差の検定におけるデータ数について、群間にあまり極端な差はないものと考えられる。

　標準化平均値差の代わりに、研究資料で t 統計量とデータ数が使われている場合は、以下のように簡単に標準化平均値差に変換することができる。

$$t=\sqrt{\frac{n_1n_2}{n_1+n_2}}\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\dfrac{(n_1-1)u_1^2+(n_2-1)u_2^2}{n_1+n_2-2}}}=\sqrt{\frac{n_1n_2}{n_1+n_2}}d\simeq\sqrt{\frac{N}{4\alpha}}d$$

２）バイアス修正標準化平均値差 g

効果量：$g=J\times d$，　$J=1-\dfrac{3}{4(n_1+n_2-2)-1}$，　分散：$V_g=J^2\times V_d$

全データ数：$N=n_1+n_2$

変換近似式

$N\rightarrow V_g$　のとき、$J=1-\dfrac{3}{4(N-2)-1}$　として、$V_g=J^2V_d\simeq\dfrac{4\alpha J^2+g^2/2}{N}$

$V_g\rightarrow N$　のとき、$J\simeq 1$　であると考え、$N\simeq\dfrac{4\alpha+g^2/2}{V_g}$

３）対数オッズ比 LOR

以下の２次元分割表を考える。

	効果あり	効果なし	合計
介入群	a	b	$a+b=n_1$
統制群	c	d	$c+d=n_2$

効果量：$LOR=\ln\left(\dfrac{a/b}{c/d}\right)=\ln\left(\dfrac{ad}{bc}\right)$，　分散：$V_{LOR}=\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}$

全データ数：$N=a+b+c+d=n_1+n_2$

変換近似式

$N\rightarrow V_{LOR}$　のとき、$V_{LOR}\simeq\dfrac{16\alpha^2}{N}$　　$V_{LOR}=\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}\simeq\dfrac{4\alpha}{n_1}+\dfrac{4\alpha}{n_2}\simeq\dfrac{16\alpha^2}{N}$

$V_{LOR}\rightarrow N$　のとき、$N\simeq\dfrac{16\alpha^2}{V_{LOR}}$

ここで、$1/a+1/b\simeq 4\alpha/n_1$，$1/c+1/d\simeq 4\alpha/n_2$　などとした部分については a,b や c,d の値が大きく異なっている場合は誤差が大きい。

対数オッズ比の代わりに、資料で以下の χ^2 統計量が使われていた場合は、簡単に対数オッズ比に変換することができない。分割表の度数から効果量を計算する必要がある。

$$\chi^2=\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

４）相関係数r

効果量：$r = \dfrac{s_{xy}}{s_x s_y}$，　分散：$V_r = \dfrac{(1-r^2)^2}{n-1}$，　全データ数：$N = n$

変換近似式

$N \to V_r$ のとき、$V_r = \dfrac{(1-r^2)^2}{N-1}$

$V_g \to N$ のとき、$N = \dfrac{(1-r^2)^2}{V_r} + 1$

2.2　効果量の変換

効果量は相互に変換可能である。ここではプログラムで用いられる変換について式を与える。

変換$d \leftrightarrow g$

効果量：$g = J \times d$，　分散：$V_g = J^2 \times V_d$

ここに、$J = 1 - \dfrac{3}{4(N-2)-1}$　（入力の際にNは設定済みとする）

変換$LOR \leftrightarrow d$

効果量：$d = LOR \times \dfrac{\sqrt{3}}{\pi}$，　分散：$V_d = V_{LOR} \times \dfrac{3}{\pi^2}$

変換$r \to d$

効果量：$d = \dfrac{2r}{\sqrt{1-r^2}}$，　分散：$V_d = \dfrac{4V_r}{(1-r^2)^3}$

変換$d \to r$

効果量：$r = \dfrac{d}{\sqrt{d^2+a}}$，　分散：$V_r = \dfrac{a^2 V_d}{(d^2+a)^3}$

ここに、$a = \dfrac{(n_1+n_2)^2}{n_1 n_2} \simeq 4\alpha$

2.3　統計的分析

研究を統合する場合は、以下の２つの場合に分けて行われる。１つは研究間に差がなく、研究iの効果量d_iが独立に$d_i \sim N(0, V_i)$に従うと考えられる場合である。もう１つは、研究iの効果量d_iが広く拡がり、$d_i \sim N(0, V_i + \sigma^2), \sigma^2 > 0$に従うと考えられる場合である。

前者を固定効果モデル、後者を変量効果モデルと呼ぶ。どちらかを調べるには、帰無仮説を $\sigma^2 = 0$ として、以下の関係を利用する。

$$Q = \sum_{i=1}^{n} w_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i} \sim \chi^2_{n-1}$$

ここで、w_i や d については、後の固定効果モデルのところで定義を示す。

固定効果モデル

固定効果モデルでは、研究間の差はなく、研究 i の効果量 d_i は独立に $d_i \sim N(0, V_i)$ に従うと仮定し、以下の集計を考える。

$$d = \sum_{i=1}^{n} d_i / V_i \Bigg/ \sum_{i=1}^{n} 1/V_i \sim N\left(0, 1\Bigg/\sum_{i=1}^{n} 1/V_i\right)$$

ここで、$w_i = 1/V_i$ として、これをウェイトと考え、$w = \sum_{i=1}^{n} w_i$ とすると、以下となる。

$$d_i \sim N(0, 1/w_i), \qquad d = \sum_{i=1}^{n} w_i d_i \Bigg/ w \sim N\left(0, 1/w\right)$$

この性質より、研究を結合した検定は、検定統計量 $z = d\Big/\sqrt{1/w} \sim N(0,1)$ を使って行う。

変量効果モデル

変量効果モデルでは、研究間に差があり、研究 i の効果量 d_i は広く拡がり、$d_i \sim N(0, V_i + \sigma^2)$ に従うと考える。$w'_i = 1/(V_i + \sigma^2)$ とおくと、$d_i \sim N(0, 1/w'_i)$ より、$\sqrt{w'_i} d_i \sim N(0,1)$ となり、以下を得る。

$$Q' = \sum_{i=1}^{n} w'_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i + \sigma^2} \sim \chi^2_{n-1}$$

一方、

$$Q = \sum_{i=1}^{n} w_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i}$$

は元の分散で測った量である。その差は、以下で与えられる。

$$Q - Q' = \sum_{i=1}^{n} \frac{(d_i - d)^2 \sigma^2}{V_i (V_i + \sigma^2)} = \sum_{i=1}^{n} (d_i - d)^2 w'_i w_i \sigma^2 = C\sigma^2$$

ここで、Q' と C は期待値を使って求める。即ち、

$$E(Q') = n - 1$$

$$C = E[\sum_{i=1}^{n} (d_i - d)^2 w'_i w_i] = \sum_{i=1}^{n} E[(d_i - d)^2] w'_i w_i$$

C については、

$$E[d_i^2] = V_i' = 1/w_i'$$

$$E[d_i d] = E[d_i \sum_{j=1}^{n} d_j w_j' / w'] = V_i' w_i' / w' = 1/w'$$

$$E[d^2] = E[\sum_{i=1}^{n} d_i w_i' / w' \sum_{j=1}^{n} d_j w_j' / w'] = \sum_{i=1}^{n} V_i w_i'^2 / w'^2 = \sum_{i=1}^{n} w_i' / w'^2 = 1/w'$$

より、以下となる。。

$$C = \sum_{i=1}^{n} (1/w_i' - 1/w') w_i' w_i = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i' w_i / w' \simeq \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / w$$

これらより、σ^2 が以下のように求められる。

$$\sigma^2 = \frac{Q-(n-1)}{C}$$

以後、ウェイトは $w_i' = 1/(V_i + \sigma^2)$ ， $w' = \sum_{i=1}^{n} w_i'$ を用いて、計算を行えばよい。即ち、研究を結合した検定は、検定統計量 $z' = d\Big/\sqrt{1/w'} \sim N(0,1)$ を使って行う。

2.4　研究群間の比較

何らかの指標の違いにより、研究が k 個のグループに分けられるとする。各グループの研究の数を n_i，全体の研究の数を n とするとき、そのグループ間の効果量の差を検定するには、以下の性質を用いる。

$$Q_{Total} \sim \chi^2_{n-1},\ \ Q_i \sim \chi^2_{n_i-1},\ \ n = \sum_{i=1}^{k} n_i \quad \text{より、}$$

$$Q_{Total} - \sum_{i=1}^{k} Q_i \sim \chi^2_{df},\ \ df = (n-1) - \sum_{i=1}^{k} (n_i - 1) = k-1$$

この計算には、固定効果モデルではウェイト $w_i = 1/V_i$ を用い、変量効果モデルではウェイト $w_i' = 1/V_i'$ を用いる。

2.5　プログラムの利用法

メニュー［分析－多変量解析等－メタ分析］を選択すると、メタ分析の分析実行メニューが図１のように表示される。

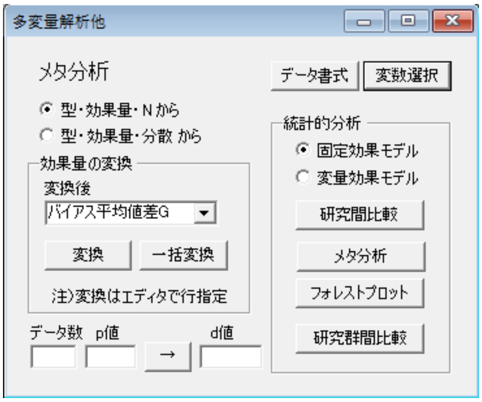
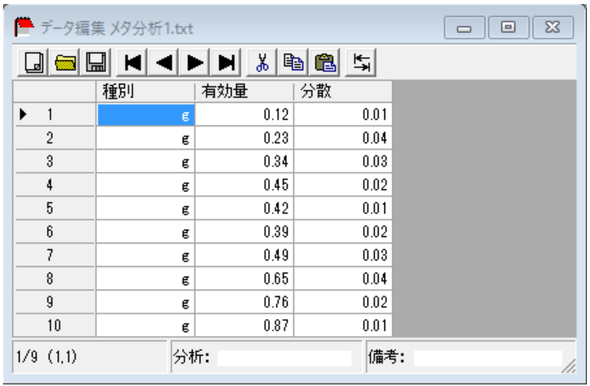


図１　分析実行メニュー

ここでは、参考文献[1] で与えられた図２のデータ[3] を元にプログラムの利用法を説明する。



	種別	有効量	分散
1	g	0.12	0.01
2	g	0.23	0.04
3	g	0.34	0.03
4	g	0.45	0.02
5	g	0.42	0.01
6	g	0.39	0.02
7	g	0.49	0.03
8	g	0.65	0.04
9	g	0.76	0.02
10	g	0.87	0.01

図２　メタ分析データ

このデータではデータ型、効果量（有効量）、分散が用いられているので、分析実行メニューのデータ型を、「型・効果量・分散」に変えて、分析を実行する。しかし、一般には分散が与えられる場合は少なく、むしろ、データ型、効果量、データ数が与えられることが多いと思われる。その場合には、分析実行メニューのデータ型を、「型・効果量・N」にして実行する。データ型には、G：バイアス修正標準化平均値差、D：標準化平均値差、LOR：対数オッズ比、R：相関係数が指定できる。なお、指定する文字は大文字でも小文字でも同じである。

また、2 群の差の検定などでは、検定統計量を省略し、検定確率だけを表示している場合もあるので、その際には、標準化平均値差 D の値を簡易的に計算できる機能をメニューの下に設けている。その他の対数オッズ比や相関係数では、殆どの場合、値を記述するので、ここでは標準化平均値差 D に限定している。また、ノンパラメトリック検定の確率から近似的に D を求めても、少し乱暴ではあるが、経験上特に大きな差は出ないように思う。

一般に各研究では効果量が同一とは限らない。異なる効果量の場合は、効果量の変換を行い、同じ効果量に合わせて分析する。そのためにプログラムには効果量の変換機能を付けている。変数選択で 3 つの変数を選択し、「変換後」コンボボックスで変換先の型を選び、「一括変換」

ボタンをクリックすると、図３のような結果が得られる。ここでは、相関係数として出力している。

効果量変換

	種別	効果量	分散
1	R	0.0551	0.0021
2	R	0.1056	0.0083
3	R	0.1549	0.0059
4	R	0.2029	0.0037
5	R	0.1896	0.0019
6	R	0.1767	0.0039
7	R	0.2204	0.0055
8	R	0.2875	0.0066
9	R	0.3302	0.0030
10	R	0.3713	0.0014

図３　相関係数への変換

グリッドの一部分のデータについて変換をしたい場合は、種別・効果量・分散の必要な行を連続的に選択して「変換」ボタンをクリックする。出力結果は省略する。

すべての研究結果を統合して検定を行いたい場合、研究間の効果量の値にばらつきがあるかどうか知らなければならない。それを調べる場合は、「研究間比較」ボタンをクリックする。結果を図４に示す。

研究間効果量比較

```
研究間効果量比較

研究数       10
研究間分散   0.054
Q値          35.839
自由度       9
有意水準     0.05
確率         0.0000
研究間の効果量に差があるといえる。
```

図４　研究間効果量の差の比較検定

この結果から、研究間の効果量に差が見られたので、分析には「変量効果モデル」を用いる。これは「固定効果モデル」に比べて差が検出しにくい検定である。

変量効果モデルを用いた最終的な分析結果を得るには、「変量効果モデル」ラジオボタンを選択し、「メタ分析」ボタンをクリックする。結果を図５に示す。

フォレストデータ

	種別	効果量	N	分散	標準誤差	p値	2.5%下限	2.5%上限
1	G	0.1200	476	0.0100	0.1000	0.2301	-0.0760	0.3160
2	G	0.2300	119	0.0400	0.2000	0.2501	-0.1620	0.6220
3	G	0.3400	160	0.0300	0.1732	0.0496	0.0005	0.6795
4	G	0.4500	242	0.0200	0.1414	0.0015	0.1728	0.7272
5	G	0.4200	484	0.0100	0.1000	0.0000	0.2240	0.6160
6	G	0.3900	241	0.0200	0.1414	0.0058	0.1128	0.6672
7	G	0.4900	162	0.0300	0.1732	0.0047	0.1505	0.8295
8	G	0.6500	124	0.0400	0.2000	0.0012	0.2580	1.0420
9	G	0.7600	252	0.0200	0.1414	0.0000	0.4828	1.0372
10	G	0.8700	513	0.0100	0.1000	0.0000	0.6740	1.0660
結合		0.4747	2773	0.0076	0.0870	0.0000	0.3041	0.6453

図５　変量効果モデルを用いた分析結果

各研究の結果がまとめて表示され、一番下の行に結合された結果が表示されている。

さらに、この結果を分かり易く表す図がフォレストプロットである。「フォレストプロット」ボタンをクリックすると、図６のような結果が表示される。

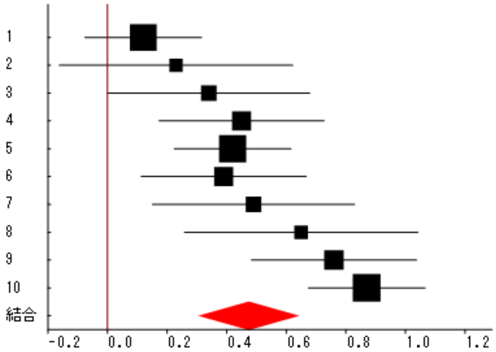
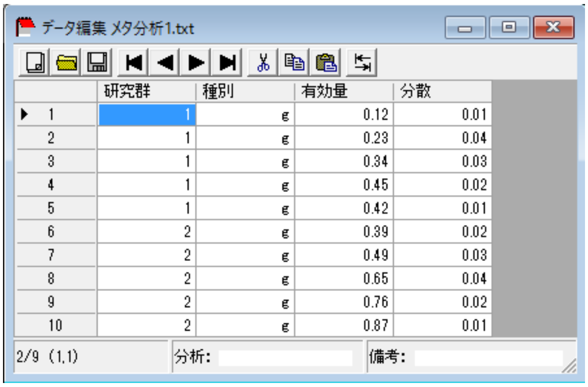


図６　フォレストプロット

一番下のひし形が、0 をまたいでいないことから、この結果では、有意に差があるといえるということになる。

次に、研究がいくつかの特徴に分かれ、その研究群間に差があるかどうか調べてみたいと考えたとする。その際には、先頭列に分類変数を加えた図 7 のようなデータ[3] を用いる。

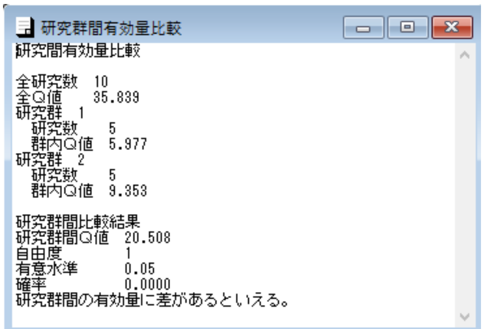

データ編集 メタ分析1.txt

	研究群	種別	有効量	分散
1	1	g	0.12	0.01
2	1	g	0.23	0.04
3	1	g	0.34	0.03
4	1	g	0.45	0.02
5	1	g	0.42	0.01
6	2	g	0.39	0.02
7	2	g	0.49	0.03
8	2	g	0.65	0.04
9	2	g	0.76	0.02
10	2	g	0.87	0.01

2/9 (1,1)　分析:　備考:

図７　２つの研究群による比較データ（メタ分析 1.txt）

すべてのデータを並んだ順に選択し、分析実行メニューの「研究群間比較」ボタンをクリックすると、図８のような結果が得られる。


研究群間有効量比較

研究間有効量比較

全研究数　10
全Q値　　35.839
研究群　1
　研究数　　5
　群内Q値　5.977
研究群　2
　研究数　　5
　群内Q値　9.353

研究群間比較結果
研究群間Q値　20.508
自由度　　　　1
有意水準　　　0.05
確率　　　　　0.0000
研究群間の有効量に差があるといえる。

図８　研究群間の比較結果

3.　2値ロジスティック回帰

2値ロジスティック回帰分析は2項分布の確率を説明変数の1次式の関数で予測する分析手法である。この分析は、確率の大きさによって事象の出現、非出現を区別することにも使え、質的データの予測手法としても利用できる。同様の分析に判別分析があるが、これは説明変数によるマハラノビス距離を用いた判別方法で、理論的な根拠という点ではロジスティック回帰分析の方が優れている。ここでは今後のために、参考文献[2] に従って、一般化線形モデルの基礎からロジスティック回帰分析について説明しておく。

3.1　一般化線形モデル

1)　指数型分布族

ある単一のパラメータ θ を持つ確率変数 Y が以下の密度関数に従うとき、その分布を指数型分布族という。

$$f(y\,;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

指数型分布族には、ポアソン分布、正規分布、2項分布等が含まれる。特に $a(y) = y$ のとき分布は正準形であると言われ、$b(\theta)$ は分布の自然パラメータと呼ばれる。

確率変数 $a(\mathrm{Y})$ については

$$\begin{aligned}\frac{d}{d\theta}\int f(y;\theta)dy &= \int [a(y)b'(\theta) + c'(\theta)] f(y;\theta)dy \\ &= E[a(Y)]b'(\theta) + c'(\theta) = 0\end{aligned}$$

より、

$$E[a(Y)] = -c'(\theta)/b'(\theta)$$

$$\frac{d^2}{d\theta^2}\int f(y;\theta)dy$$

$$= \int [a(y)b''(\theta) + c''(\theta)] f(y;\theta)dy + \int [a(y)b'(\theta) + c'(\theta)]^2 f(y;\theta)dy$$

$$= E[a(Y)^2]b'(\theta)^2 + E[a(Y)][b''(\theta) + 2b'(\theta)c'(\theta)] + c''(\theta) + c'(\theta)^2$$

$$= E[a(Y)^2]b'(\theta)^2 - \frac{c'(\theta)}{b'(\theta)}[b''(\theta) + 2b'(\theta)c'(\theta)] + c''(\theta) + c'(\theta)^2$$

$$= E[a(Y)^2]b'(\theta)^2 - E[a(Y)]^2 b'(\theta)^2 + \frac{1}{b'(\theta)}[-b''(\theta)c'(\theta) + c''(\theta)b'(\theta)]$$

$$= V[a(Y)]b'(\theta)^2 + \frac{1}{b'(\theta)}[-b''(\theta)c'(\theta) + c''(\theta)b'(\theta)] = 0$$

より、

$$V[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

という性質がある。

対数密度関数 $l(y\,;\theta)=\log f(y\,;\theta)$ の θ に関する微分の y を確率変数とみなした

$$U(Y\,;\theta)=a(Y)b'(\theta)+c'(\theta)$$

は、スコア統計量とも呼ばれ、その分布の期待値と分散は上式を使うと以下となる。

$$E[U]=0$$

$$V[U]=V[a(Y)]b'(\theta)^2=\frac{b''(\theta)c'(\theta)-c''(\theta)b'(\theta)}{b'(\theta)}$$

さらに、

$$V[\mathrm{U}]=E[U^2]-E[U]^2=E[U^2]$$

$$E[U']=E[a(Y)]b''(\theta)+c''(\theta)=-\frac{b''(\theta)c'(\theta)-c''(\theta)b'(\theta)}{b'(\theta)}=-V[U]$$

の関係より、以下も成り立つ。

$$V[U]=E[U^2]=-E[U']$$

スコア統計量の分散 $V[U]$ は情報量とも呼ばれる。

2)　**正準形の一般化線形モデル**

正準形の指数型分布族の分布に従う確率変数 $Y_i\ (i=1,2,\cdots,N)$ が、パラメータ θ_i の同じ形の以下の独立な密度関数の分布に従うと考える。

$$f(y_i\,;\theta_i)=\exp[y_i b(\theta_i)+c(\theta_i)+d(y_i)]$$

対数密度関数 $l(y_i;\theta_i)$ は以下で与えられる。

$$l(y_i;\theta_i)=y_i b(\theta_i)+c(\theta_i)+d(y_i)$$

確率変数 Y_i の平均と分散は前節の議論より、以下のように与えられる。

$$E[Y_i]=-c'(\theta_i)/b'(\theta_i)\equiv\mu_i$$

$$V[Y_i]=\frac{b''(\theta_i)c'(\theta_i)-c''(\theta_i)b'(\theta_i)}{b'(\theta_i)^3}$$

ここで、θ_i は μ_i の関数であるとみることができる。

我々はこの μ_i に対して、ある説明変数 ${}^t\mathbf{x}_i=(1,x_{i1},\cdots,x_{ip})$（$i=1,\cdots,N$）とパラメータ ${}^t\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_p)$ を用いて以下のような仮定をする。

$$\eta_i\equiv g(\mu_i)=\sum_{j=1}^{p}\beta_j x_{ij}+\beta_0={}^t\boldsymbol{\beta}\mathbf{x}_i$$

この仮定により、θ_i は $\boldsymbol{\beta}$ の関数と見ることができる。またこの関係を与える関数 $\eta_i=g(\mu_i)$ を連結関数という。

確率変数Y_iの同時密度関数（尤度関数）は以下で与えられる。

$$f(\mathbf{y};\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{N}\{y_i b(\theta_i) + c(\theta_i) + d(y_i)\}\right]$$

また対数尤度関数$l(\mathbf{y};\boldsymbol{\theta})$は以下のようになる。

$$l(\mathbf{y};\boldsymbol{\theta}) = \sum_{i=1}^{N} l_i = \sum_{i=1}^{N}[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \tag{1}$$

この対数尤度関数のβ_jによる微分をスコアベクトルと呼び、U_jとすると、スコアベクトルU_jは以下のようになる。

$$U_j = \sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}$$

ここで、

$$\begin{aligned}\frac{\partial l_i}{\partial \theta_i} &= y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)\left[y_i + c'(\theta_i)/b'(\theta_i)\right] \\ &= b'(\theta_i)(y_i - E[Y_i]) = b'(\theta_i)(y_i - \mu_i)\end{aligned}$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{b'(\theta_i)^2}{-c''(\theta_i)b'(\theta_i) + c'(\theta_i)b''(\theta_i)} = \frac{1}{b'(\theta_i)V[Y_i]}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \eta_i}{\partial \beta_j}\frac{\partial \mu_i}{\partial \eta_i} = x_{ij}\frac{\partial \mu_i}{\partial \eta_i}$$

となることから、以下の関係を得る。

$$U_j = \sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{(y_i - \mu_i)x_{ij}}{V[Y_i]}\frac{\partial \mu_i}{\partial \eta_i} \tag{2}$$

また、以下も成り立つ。

$$E[U_j] = 0 \tag{3}$$

さらにU_jのβ_kによる微分をU_{jk}とすると、U_{jk}は以下のようになる。

$$\begin{aligned}U_{jk} &\equiv \frac{\partial U_j}{\partial \beta_k} = \frac{\partial}{\partial \beta_k}\sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_k}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial}{\partial \theta_i}\left(\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}\right) \\ &= \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \mu_i}{\partial \beta_k}\left(\frac{\partial \theta_i}{\partial \mu_i}\right)^2\frac{\partial}{\partial \theta_i}\left(\frac{\partial l_i}{\partial \theta_i}\right) + \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_k}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}\frac{\partial}{\partial \theta_i}\left(\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\right)\end{aligned}$$

$$=\sum_{i=1}^{N}\frac{\partial\mu_i}{\partial\beta_j}\frac{\partial\mu_i}{\partial\beta_k}\left(\frac{\partial\theta_i}{\partial\mu_i}\right)^2[y_i b''(\theta_i)+c''(\theta_i)]$$

$$+\sum_{i=1}^{N}\frac{(y_i-\mu_i)x_{ik}}{V[Y_i]}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial}{\partial\theta_i}\left(\frac{\partial\mu_i}{\partial\beta_j}\frac{\partial\theta_i}{\partial\mu_i}\right)$$

また、

$$E[Y_i b''(\theta_i)+c''(\theta_i)]=E[Y_i]b''(\theta_i)+c''(\theta_i)=-c'(\theta_i)b''(\theta_i)/b'(\theta_i)+c''(\theta_i)$$

$$=\frac{-c'(\theta_i)b''(\theta_i)+b'(\theta_i)c''(\theta_i)}{b'(\theta_i)}=-b'(\theta_i)^2V[Y_i]$$

$$E\left[\sum_{i=1}^{N}\frac{(Y_i-\mu_i)x_{ik}}{V[Y_i]}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial}{\partial\theta_i}\left(\frac{\partial\mu_i}{\partial\beta_j}\frac{\partial\theta_i}{\partial\mu_i}\right)\right]=0$$

であることから、(2) 式を求める際の計算により、U_{jk} の変数の値を確率変数で置き換えて計算すると以下となる。

$$E[U_{jk}]=-\sum_{i=1}^{N}\frac{\partial\mu_i}{\partial\beta_j}\frac{\partial\mu_i}{\partial\beta_k}\left(\frac{\partial\theta_i}{\partial\mu_i}\right)^2 b'(\theta_i)^2V[U_i]=-\sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{V[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$$

一方、(2) の表式より、

$$E[U_jU_l]=\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{E[(y_i-\mu_i)(y_k-\mu_k)]x_{ij}x_{kl}}{V[Y_i]V[Y_k]}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\mu_k}{\partial\eta_k}$$

$$=\sum_{i=1}^{N}\sum_{k=1}^{N}\frac{V[Y_i]\delta_{ik}x_{ij}x_{kl}}{V[Y_i]V[Y_k]}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\mu_k}{\partial\eta_k}=\sum_{i=1}^{N}\frac{x_{ij}x_{il}}{V[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2$$

であることから、以下となる。

$$E[U_{jk}]=-E[U_jU_k] \tag{4}$$

ここで、$(\Im)_{jk}=-E[U_{jk}]=E[U_jU_k]$ とすると、行列 $\Im$ は情報行列と呼ばれる。

$$(\Im)_{jk}=E(U_jU_k)=\sum_{i=1}^{N}\frac{x_{ij}x_{ik}}{V[Y_i]}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \tag{5}$$

今、(1) で与えられる対数尤度関数が最大となる $\boldsymbol{\beta}$ の値を求めてみよう。これには、

$$\frac{\partial l}{\partial\beta_j}=U_j=0$$

という方程式を解くことになる。

（$f(x)=0$ を解くには $y_m=f'(x_{m-1})(x_m-x_{m-1})-f(x_{m-1})=0$ を計算することから）

解はニュートン・ラフソン法によると、

$$U_j^{(m)}=\sum_{k=1}^{p}U_{jk}^{(m-1)}(\beta_k^{(m)}-\beta_k^{(m-1)})+U_j^{(m-1)}=0$$

のように、$\beta_k^{(m)}$ の値を逐次求めて行くことになるが、実際の計算では$U_{jk}^{(m-1)}$ の代わりに、期待値を取った情報行列$-(\mathfrak{I}^{(m-1)})_{jk}$ を用いる。この式を書き変えると、以下となる。

$$\boldsymbol{\beta}^{(m)}=\boldsymbol{\beta}^{(m-1)}+(\mathfrak{I}^{(m-1)})^{-1}\mathbf{U}^{(m-1)}$$

(3)式と(5)式を元にして、大標本においては、スコアベクトルの分布は漸近的に

$$\mathbf{U}\sim\mathbf{N}(\mathbf{0},\mathfrak{I}),\qquad {}^t\mathbf{U}\mathfrak{I}^{-1}\mathbf{U}\sim\chi^2(p)$$

であることも示される。

最尤推定量$l(\boldsymbol{\beta})$ の推定値$\mathbf{b}$ の近傍でのテイラー展開近似は以下となり、

$$l(\boldsymbol{\beta})=l(\mathbf{b})+{}^t(\boldsymbol{\beta}-\mathbf{b})\mathbf{U}(\mathbf{b})-\frac{1}{2}{}^t(\boldsymbol{\beta}-\mathbf{b})\mathfrak{I}(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})$$

スコアベクトルの推定値$U(\mathbf{b})$ の近傍でのテイラー展開近似は以下となる。

$$\mathbf{U}(\boldsymbol{\beta})=\mathbf{U}(\mathbf{b})-\mathfrak{I}(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})=-\mathfrak{I}(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})$$

ここでは$E[\partial U_j/\partial\beta_k]=-\mathfrak{I}_{jk}$ や$\mathbf{U}(\mathbf{b})=\mathbf{0}$ を使っている。これらより、

$${}^t(\boldsymbol{\beta}-\mathbf{b})\mathfrak{I}(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})\sim\chi^2(p)$$

も示される。また、同様にして以下も示される。

$$\mathbf{b}=\boldsymbol{\beta}+\mathfrak{I}^{-1}\mathbf{U}\sim N(\boldsymbol{\beta},\mathfrak{I}^{-1})$$

後に説明するが、モデルの最適値からのずれを表す逸脱度D を以下のように定義する。

$$D=2[l(\mathbf{b}_{\max};\mathbf{y})-l(\hat{\mathbf{b}};\mathbf{y})]\sim\chi^2(N-p)$$

ここに、$l(\mathbf{b}_{\max};\mathbf{y})$ はパラメータ数N の飽和モデルでの対数尤度、$l(\hat{\mathbf{b}};\mathbf{y})$ は現在考えている、パラメータ数p のモデルでの対数尤度である。同じパラメータ数では、この値が小さい連結関数のモデルほど適合が良いと判断する。但し、分布は漸近的に成り立つものである。

モデルに意味があるかどうかの検定では、以下の尤度比χ^2統計量が使われる。

$$C=2[l(\hat{\mathbf{b}};\mathbf{y})-l(\mathbf{b}_{\min};\mathbf{y})]\sim\chi^2(p-1)$$

ここに$l(\mathbf{b}_{\min};\mathbf{y})$ は定数パラメータ１つの最小モデルの対数尤度、$l(\hat{\mathbf{b}};\mathbf{y})$ は現在考えている、パラメータ数p のモデルでの対数尤度である。これは、帰無仮説として最小モデルが正しい（回帰式は意味がない）とする検定である。

２項分布の場合、対数尤度関数は以下で与えられる。

$$l(\boldsymbol{\beta};\mathbf{y})=\sum_{i=1}^{N}\left[y_i\log p_i+(n_i-y_i)\log(1-p_i)+\log\binom{n_i}{y_i}\right]$$

$l(\mathbf{b}_{\max};\mathbf{y})$ と $l(\hat{\mathbf{b}};\mathbf{y})$ は、確率 p_i の中に、それぞれ実測値を用いた $p_i = y_i / n_i$ と予測値を用いた $\hat{p}_i = \hat{y}_i / n_i$ を代入して求める。$l(\mathbf{b}_{\min};\mathbf{y})$ は確率 p_i の中に同じ以下の値を代入して求める。

$$\tilde{p} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} n_i$$

実測値と推測値の関係を与える指標として､決定係数からの類推である以下の擬似 R^2 も利用される。

$$\tilde{R}^2 = \frac{l(\mathbf{b}_{\min};\mathbf{y}) - l(\hat{\mathbf{b}};\mathbf{y})}{l(\mathbf{b}_{\min};\mathbf{y})}$$

さらにプログラムでは、実測値と予測値の相関係数も求めている。

3)　2 項分布モデル

2 項分布のパラメータを説明変数の線形結合で推測する場合、密度関数、密度関数の対数、逸脱度、目的変数の平均と分散は以下のようになる。ここで、密度関数の対数の最後の項はパラメータに依存していないので､計算上は考えないことにする（参考文献[2] の数値に従っている）。

$$f(y_i;p_i) = {}_{n_i}C_{y_i}\, p_i^{y_i}(1-p_i)^{n_i - y_i}$$

$$\begin{aligned}
l(y_i;p_i) &= \log[{}_{n_i}C_{y_i}\, p_i^{y_i}(1-p_i)^{n_i - y_i}] \\
&= y_i \log p_i + (n_i - y_i)\log(1-p_i) + \log {}_{n_i}C_{y_i} \\
&= y_i[\log p_i - \log(1-p_i)] + n_i \log(1-p_i) + \log {}_{n_i}C_{y_i} \\
&\to y_i[\log p_i - \log(1-p_i)] + n_i \log(1-p_i)
\end{aligned}$$

$$D = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{n_i p_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - n_i p_i}\right)\right] \sim \chi^2(N-p)$$

$$C = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{\hat{y}_i}{n_i \tilde{p}}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - n_i \tilde{p}}\right)\right] \sim \chi^2(p-1)$$

$$\tilde{p} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} n_i$$

$$E[Y_i] = n_i p_i \equiv \mu_i$$

$$V[Y_i] = n_i p_i (1-p_i)$$

ここでは n_i 回の試行に対して、y_i 回の事象が起こったとしているが、1 回の試行で起こったか起こらないかにする場合は、$n_i = 1,\ y_i = \{0,1\}$ とすればよい。

逸脱度と同様に最適値からのずれを表す統計量に以下のピアソン χ^2 統計量がある。

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{n_i \hat{p}_i (1-\hat{p}_i)} \sim \chi^2(N-p)$$

これは逸脱度と漸近的に同じ指標であるが、逸脱度と比べてこちらの方が分布によく適合するという意見もある。

これまでは、2 項分布に基づく一般論であったが、これ以降は、説明変数との関係を与える連結関数の部分に仮定が入る。連結関数の仮定でよく利用されるモデルが、ロジスティックモデル、プロビットモデル、極値モデル等である。以下に最終的な計算で用いられる式を与えておく。

ロジスティックモデル

$$\eta_i = \log \frac{p_i}{1-p_i} = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} = \frac{1}{1+e^{-\eta_i}}, \quad 1-p_i = \frac{1}{1+e^{\eta_i}}$$

$$\mu_i = n_i p_i = \frac{n_i e^{\eta_i}}{1+e^{\eta_i}}, \qquad \frac{\partial \mu_i}{\partial \eta_i} = \frac{n_i e^{\eta_i}}{(1+e^{\eta_i})^2} = n_i p_i (1-p_i)$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{n_i p_i (1-p_i)} n_i p_i (1-p_i) = \sum_{i=1}^{N} (y_i - \mu_i) x_{ij}$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij} x_{ik} n_i^2 p_i^2 (1-p_i)^2}{n_i p_i (1-p_i)} = \sum_{i=1}^{N} x_{ij} x_{ik} n_i p_i (1-p_i)$$

プロビットモデル

$$\eta_i = \Phi^{-1}(p_i) = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = \Phi(\eta_i)$$

$$\mu_i = n_i p_i = n_i \Phi(\eta_i), \qquad \frac{\partial \mu_i}{\partial \eta_i} = \frac{n_i}{\sqrt{2\pi}} \exp(-\eta_i^2 / 2)$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{p_i (1-p_i)} \frac{\exp(-\eta_i^2 / 2)}{\sqrt{2\pi}}$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{p_i (1-p_i)} \frac{n_i \exp(-\eta_i^2)}{2\pi}$$

極値モデル

$$\eta_i = \log[-\log(1-p_i)] = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = 1 - \exp[-\exp(\eta_i)] \qquad (1 - p_i = \exp[-\exp(\eta_i)])$$

$$\mu_i = n_i p_i, \quad \frac{\partial \mu_i}{\partial \eta_i} = n_i \frac{\partial p_i}{\partial \eta_i} = n_i \exp(\eta_i) \exp[-\exp(\eta_i)]$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{p_i(1 - p_i)} \exp(\eta_i) \exp[-\exp(\eta_i)]$$

$$= \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{p_i} \exp(\eta_i)$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{p_i(1 - p_i)} n_i \exp(2\eta_i) \exp[-2\exp(\eta_i)]$$

$$= \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{p_i} n_i (1 - p_i) \exp(2\eta_i)$$

極値モデルの計算には以下の性質を利用する。

$$\lim_{p \to 0} \exp(\eta) / p = 1$$

プロビットモデルと極値モデルの場合、$p_i \to 0$ や $p_i \to 1$ のときに、計算機のまるめ誤差や分布関数の近似誤差から、除算のエラーが生じることがある。そのため、プログラムではある程度のところで、これらの極限を止めるようにしている。また最終結果でも対数尤度の計算で同様のことが起こる可能性があるので、同じように極端な値を避けるようにしている。現在のプログラムでは、$0.000001 \le p_i \le 0.999999$ の範囲に設定している。

3.2　プログラムの利用法

メニュー［分析－多変量解析他－判別手法－2 値ロジスティック回帰］を選択すると、図 1 のような、2 値ロジスティック回帰分析の分析実行メニューが表示される。

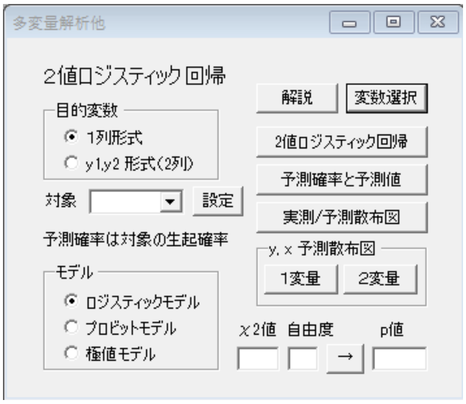
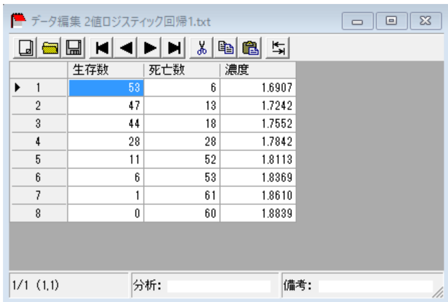


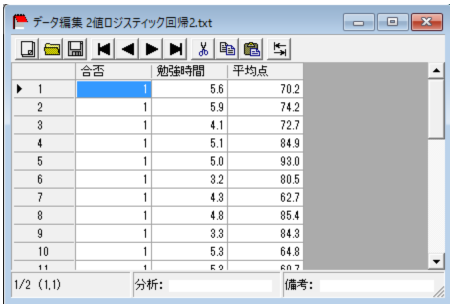
図１　分析実行メニュー

利用するデータの形式は、「y1,y2 形式（2 列）」と「0/1 形式（1 列）」があり、それぞれ図 2a と図 2b のように、目的変数が 2 列で表されるか、1 列で表されるかの違いである。



	生存数	死亡数	濃度
1	53	6	1.6907
2	47	13	1.7242
3	44	18	1.7552
4	28	28	1.7842
5	11	52	1.8113
6	6	53	1.8369
7	1	61	1.8610
8	0	60	1.8839

図 2a　目的変数 2 列データ[3]



	合否	勉強時間	平均点
1	1	5.6	70.2
2	1	5.9	74.2
3	1	4.1	72.7
4	1	5.1	84.9
5	1	5.0	93.0
6	1	3.2	80.5
7	1	4.3	62.7
8	1	4.8	85.4
9	1	3.3	84.3
10	1	5.3	64.8

図 2b　目的変数 1 列データ[3]

目的変数が 2 列で表される場合は、事象 1 が何回起きて、事象 2 が何回起きたかの重複のあるデータで、1 列で表される場合は、1 回の試行で事象が起きるかどうかの重複のないデータである。2 列の場合、対象変数と非対象変数を入力し、対象変数をコンボボックスで選択しておく。対象変数の出現確率が予測確率となる。1 列のデータを事象が起きないと事象が起きたにして 2 列で表現することも可能である。目的変数が 1 列の場合は、2 列の特別な場合と考えてもよい。以後データ形式を分けて、プログラムの出力について説明する。

図 2a のデータのとき、「ロジスティックモデル」ラジオボタンを選択し、「2 値ロジスティック回帰」ボタンをクリックすると図 3 の結果が表示される。

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
濃度	34.2703	2.3117	2.9121	0.0000	28.5625	39.9781	7.646E14
切片	-60.7175	0.7438	5.1807	0.0000	-70.8716	-50.5633	
対数尤度値	-186.235						
逸脱度D	11.232	自由度	6	上側確率	0.0815	<-n小注意	
ピアソンχ2	10.027	自由度	6	上側確率	0.1235	<-n小注意	
C尤度比	272.970	自由度	1	上側確率	0.0000		
擬似R2	0.423						
実測予測R2	0.989						

図 3　2 値ロジスティック回帰結果

ここでは回帰パラメータの値とその検定値、対数尤度値、逸脱度などが表示される。

また、「予測確率と予測値」ボタンをクリックすると、個別の実測値、予測確率、予測値が図 4 のように表示される。

	実測値	予測確率	予測値
1	6	0.059	3.457
2	13	0.164	9.842
3	18	0.362	22.451
4	28	0.605	33.898
5	52	0.795	50.096
6	53	0.903	53.291
7	61	0.955	59.222
8	60	0.979	58.743

図 4　予測確率と予測値

「実測/予測散布図」ボタンをクリックすると、この実測値と予測値が、図５のようにプロットされる。

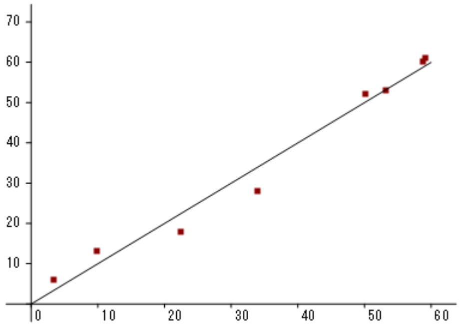


図５　実測/予測散布図

予測の説明変数が１つまたは２つの場合、実測値と確率の予測関数（連結関数の逆関数）の関係を表示することができる。ここでは説明変数が１つであるので、「y, x 予測散布図」グループボックス内の「1 変量」ボタンをクリックする。結果は横軸を濃度として図６のようになる。

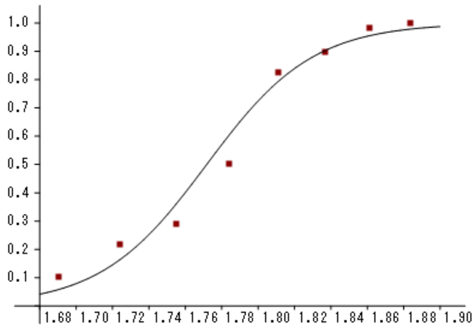


図６　予測関数とデータ（ロジスティックモデル）

但し、ここでは軸設定を使ってグラフの軸を変更している。

この図と同様に、プロビットモデルと極値モデルの予測関数についても図７と図８で当てはまりを見てみる。

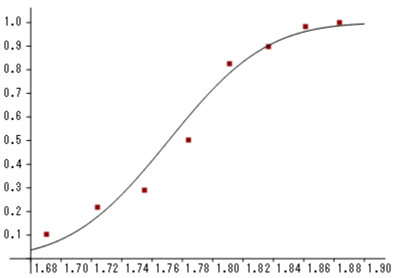


図７　プロビットモデル

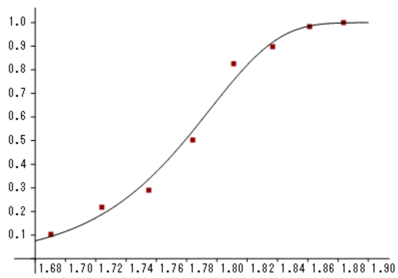


図８　極値モデル

これらを比べると極値モデルの当てはまりが良いことが分かる。このことは、「2 値ロジスティック回帰」ボタンで表示される、対数尤度値、逸脱度 D、R^2 の値でも確認できる。

次に図 2b のデータを用いた場合のロジスティックモデルの実行結果を示す。目的変数は「0/1 形式（1 列）」を選択し、「2 値ロジスティック回帰」ボタンをクリックすると図 9 のような結果が表示される。

ロジスティック回帰分析結果

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
勉強時間	5.0765	5.9238	2.5049	0.0427	0.1670	9.9861	1.602E02
平均点	0.4581	5.2007	0.2231	0.0400	0.0208	0.8955	1.581E00
切片	-52.7491	-20.9681	24.9005	0.0341	-101.5540	-3.9442	
対数尤度値	-5.569						
逸脱度D	11.137	自由度	27	上側確率	0.9969	<-n小注意	
ピアソンχ2	13.422	自由度	27	上側確率	0.9863	<-n小注意	
C尤度比	29.917	自由度	2	上側確率	0.0000		
擬似R2	0.729						
実測予測R2	0.864						
誤判別確率	0を1と	0.059	1を0と	0.077			

図 9　2 値ロジスティック回帰結果

このデータ形式では、確率の予測による 0/1 の判別についての誤判別確率が追加されている。

また、「予測確率と予測値」ボタンをクリックすると、個体別の実測値、予測確率、予測値が図 10 のように表示される。

予測確率と予測値

	実測値	予測確率	予測値
9	1	0.932	1.000
10	1	0.979	1.000
11	1	0.877	1.000
12	1	1.000	1.000
13	1	0.991	1.000
14	0	0.000	0.000
15	0	0.374	0.000
16	0	0.070	0.000
17	0	0.005	0.000
18	0	0.000	0.000

図 10　予測確率と予測値

ここでは、予測値として、予測確率が 0.5 未満なら 0、予測確率が 0.5 以上なら 1 が与えられている。この予測値と実測値との違いを表すのが、図 9 の誤判別確率である。

「実測/予測散布図」をクリックすると、この実測値と予測確率が、図 11 のように表示される。ここに、図 11 と図 5 の違いは、図 11 はデータが個別であるので、実測値と予測値の代わりに実測値と予測確率を用いているところである。

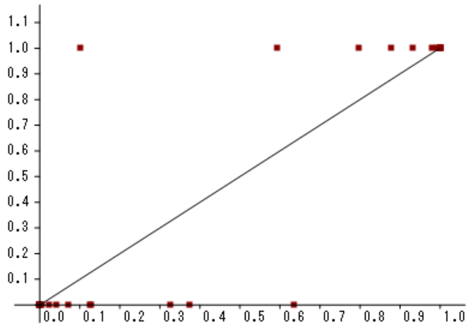


図 11　実測/予測確率散布図

このデータでは説明変数が 2 つであるので、「y, x 予測散布図」グループボックス内の「2 変量」ボタンをクリックする。結果は図 12 のようなグラフになる。

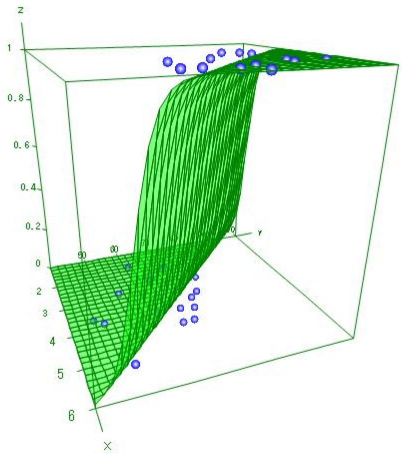


図 12　予測関数とデータ（ロジスティックモデル）

最後に、分析実行メニューの下部に、利用する可能性のある χ^2 分布の値と自由度から確率を求めるボタンを追加しておいた。必要に応じて利用してもらいたい。

4.　多値ロジスティック回帰

4.1　多項分布モデル

多項分布の密度関数、対数密度関数は以下で与えられる。ここで、対数密度関数の最後の項はパラメータに依存していないので、計算上は考えないことにする（参考文献[2] の計算に従っている）。

密度関数

$$f(y_i; p_i) = n_i! \prod_{\alpha=1}^{J} \frac{p_{i\alpha}^{y_{i\alpha}}}{y_{i\alpha}!}, \qquad \sum_{\alpha=1}^{J} y_{i\alpha} = n_i, \ \sum_{\alpha=1}^{J} p_{i\alpha} = 1$$

これより、y_i 及び p_i の中の１つは他の変数で規定される。

対数密度関数

$$l(y_i;p_i) = \log f(y_i;p_i) = \sum_{\alpha=1}^{J} y_{i\alpha} \log p_{i\alpha} + \log n_i ! - \sum_{\alpha=1}^{J} y_{i\alpha} !$$

$$\to \sum_{\alpha=1}^{J} y_{i\alpha} \log p_{i\alpha}$$

以下この関係を利用して計算過程を考えてみる。

4.2　名義ロジスティックモデル

名義ロジスティックモデルでは、一般性を失わず、他の変数で規定される基準となるカテゴリを１とすると、$\alpha \neq 1$ として、

$$\frac{\partial l_i}{\partial p_{i\alpha}} = \frac{\partial}{\partial p_{i\alpha}} \left[y_{i1} \log p_{i1} + \sum_{\beta=2}^{J} y_{i\beta} \log p_{i\beta} \right] = -\frac{y_{i1}}{p_{i1}} + \frac{y_{i\alpha}}{p_{i\alpha}}$$

$$E[Y_{i\alpha}] = n_i p_{i\alpha} \equiv \mu_{i\alpha}$$

$$Cov[Y_{i\alpha} Y_{i\beta}] = n_i p_{i\alpha} (\delta_{\alpha\beta} - p_{i\beta})$$

名義ロジスティックモデルは、基準となるカテゴリに対する他のカテゴリのロジットを説明変数の線形結合で推測する。

$$\eta_{i\alpha} = \log \frac{p_{i\alpha}}{p_{i1}} = \sum_{k=1}^{p} \beta_{k\alpha} x_{ik} + \beta_{0\alpha} \quad \text{for} \ \ \alpha = 2, \cdots, J$$

これより、$p_{i\alpha} = p_{i1} e^{\eta_{i\alpha}}$

また、$1 = \sum_{\alpha=1}^{J} p_{i\alpha} = p_{i1} + \sum_{\alpha=2}^{J} p_{i\alpha} = p_{i1} [1 + \sum_{\alpha=2}^{J} e^{\eta_{i\alpha}}]$　より、

$$p_{i1} = \frac{1}{1 + \sum_{\beta=2}^{J} e^{\eta_{i\beta}}}, \quad p_{i\alpha} = \frac{e^{\eta_{i\alpha}}}{1 + \sum_{\beta=2}^{J} e^{\eta_{i\beta}}}$$

定義より、

$$\mu_{i1} = n_i p_{i1} = \frac{n_i}{1 + \sum_{\beta=2}^{J} e^{\eta_{i\beta}}}, \quad \mu_{i\alpha} = n_i p_{i\alpha} = \frac{n_i e^{\eta_{i\alpha}}}{1 + \sum_{\beta=2}^{J} e^{\eta_{i\beta}}}$$

以下この関係を利用して計算過程を考えてみる。$\alpha \neq 1, \beta \neq 1$ に対して

$$\frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = \frac{n_i e^{\eta_{i\beta}} \delta_{\alpha\beta} (1 + \sum_{\gamma=2}^{J} e^{\eta_{i\gamma}}) - n_i e^{\eta_{i\alpha}} e^{\eta_{i\beta}}}{(1 + \sum_{\gamma=2}^{J} e^{\eta_{i\gamma}})^2} = n_i p_{i\beta} (\delta_{\alpha\beta} - p_{i\alpha})$$

$$\frac{\partial \mu_{i\beta}}{\partial \beta_{j\alpha}} = \frac{\partial \eta_{i\alpha}}{\partial \beta_{j\alpha}} \frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = x_{ij} \frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = x_{ij} n_i p_{i\beta} (\delta_{\alpha\beta} - p_{i\alpha})$$

以上より、

$$\begin{aligned} U_j^\alpha &= \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_{j\alpha}} = \sum_{i=1}^{N} \sum_{\beta=2}^{J} \sum_{\gamma=2}^{J} \frac{\partial \mu_{i\beta}}{\partial \beta_{j\alpha}} \frac{\partial p_{i\gamma}}{\partial \mu_{i\beta}} \frac{\partial l_i}{\partial p_{i\gamma}} \\ &= \sum_{i=1}^{N} \sum_{\beta=2}^{J} \sum_{\gamma=2}^{J} x_{ij} n_i p_{i\beta} (\delta_{\alpha\beta} - p_{i\alpha}) \frac{\delta_{\beta\gamma}}{n_i} \left(-\frac{y_{i1}}{p_{i1}} + \frac{y_{i\gamma}}{p_{i\gamma}} \right) \\ &= \sum_{i=1}^{N} \sum_{\beta=2}^{J} x_{ij} p_{i\beta} (\delta_{\alpha\beta} - p_{i\alpha}) \left(-\frac{y_{i1}}{p_{i1}} + \frac{y_{i\beta}}{p_{i\beta}} \right) = \sum_{i=1}^{N} x_{ij} (y_{i\alpha} - n_i p_{i\alpha}) \end{aligned}$$

$$\begin{aligned} \mathfrak{I}_{jk}^\alpha &= -\frac{\partial U_j^\alpha}{\partial \beta_{k\alpha}} = -\frac{\partial}{\partial \beta_{k\alpha}} \sum_{i=1}^{N} x_{ij} (y_{i\alpha} - n_i p_{i\alpha}) = \sum_{i=1}^{N} x_{ij} n_i \sum_{\beta=2}^{J} \frac{\partial \mu_{i\beta}}{\partial \beta_{k\alpha}} \frac{\partial p_{i\alpha}}{\partial \mu_{i\beta}} \\ &= \sum_{i=1}^{N} x_{ij} n_i \sum_{\beta=2}^{J} x_{ik} n_i p_{i\beta} (\delta_{\alpha\beta} - p_{i\alpha}) \frac{\delta_{\alpha\beta}}{n_i} = \sum_{i=1}^{N} x_{ij} x_{ik} n_i p_{i\alpha} (1 - p_{i\alpha}) \end{aligned}$$

これらのスコアベクトルと情報ベクトルより、$\beta_{i\alpha}$（$\alpha \neq 1$）は推定される。

最適値からのずれを表す、逸脱度、ピアソンのχ^2統計量及び、最小モデルからのずれを表す尤度比χ^2統計量は以下のようになる。

$$D = 2 \sum_{i=1}^{N} \sum_{\alpha=1}^{J} y_{i\alpha} \log \frac{y_{i\alpha}}{n_i \hat{p}_{i\alpha}} \sim \chi^2((J-1)(N-p))$$

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{J} \frac{(y_{ij} - \hat{y}_{ij})^2}{n_i \hat{p}_{ij}} \sim \chi^2((J-1)(N-p))$$

$$C = 2 \sum_{i=1}^{N} \sum_{\alpha=1}^{J} y_{i\alpha} \log \frac{\hat{y}_{i\alpha}}{n_i \tilde{p}_{\alpha}} \sim \chi^2((J-1)(p-1)), \quad \tilde{p}_\alpha = \sum_{i=1}^{N} y_{i\alpha} \Big/ \sum_{i=1}^{N} n_i$$

ピアソンのχ^2統計量は逸脱度と漸近的に同じ指標であるが、逸脱度と比べてこちらの方が分布によく適合するという意見もある。ここではn_i回の試行に対して、y_i回の事象が起こったとしているが、１回の試行で起こったか起こらないかにする場合は、$n_i = 1,\ y_i = \{0, 1\}$とする。但し、分布はデータ数が無限大のときの極限であるので、注意が必要である。

4.3　順序ロジスティックモデル

順序ロジスティックモデルには、累積ロジットモデル、隣接カテゴリロジットモデル、連続比ロジットモデルなどがあるが、ここではプログラムに取り入れている累積ロジットモデルについて説明する。

累積ロジットモデルでは、以下の比の対数を線形関数で予測する。

$$\frac{p_1}{p_2+\cdots+p_J}=e^{\eta_1},\quad \frac{p_1+p_2}{p_3+\cdots+p_J}=e^{\eta_2},\ \cdots,\ \frac{p_1+\cdots+p_{J-1}}{p_J}=e^{\eta_{J-1}}$$

これは、連続した複数のカテゴリの出現確率と残りのカテゴリの出現確率のオッズ比を説明変数の線形関数で予測することに相当する。

上の関係を以下のように書き換え、

$$\frac{p_2+\cdots+p_J}{p_1}=e^{-\eta_1},\quad \frac{p_3+\cdots+p_J}{p_1+p_2}=e^{-\eta_2},\ \cdots,\ \frac{p_J}{p_1+\cdots+p_{J-1}}=e^{-\eta_{J-1}}$$

$q_\alpha=p_1+p_2+\cdots+p_\alpha$ と定義すると、以下の関係が示される。

$$1-p_1=p_2+\cdots+p_J=p_1e^{-\eta_1}\quad \text{より、}\quad p_1=\frac{e^{\eta_1}}{1+e^{\eta_1}}=q_1$$

$$p_2=p_1e^{-\eta_1}-(p_1+p_2)e^{-\eta_2}\quad \text{より、}$$

$$p_2=\frac{1}{1+e^{-\eta_2}}-\frac{1}{1+e^{-\eta_1}},\quad p_1+p_2=\frac{e^{\eta_2}}{1+e^{\eta_2}}=q_2$$

同様にして、

$$p_{J-1}=\frac{1}{1+e^{-\eta_{J-1}}}-\frac{1}{1+e^{-\eta_{J-2}}},\quad p_1+\cdots+p_{J-1}=\frac{e^{\eta_{J-1}}}{1+e^{\eta_{J-1}}}=q_{J-1}$$

また、

$$p_J=1-(p_1+\cdots+p_{J-1})=1-\frac{1}{1+e^{-\eta_{J-1}}}=\frac{1}{1+e^{\eta_{J-1}}}=1-q_{J-1}$$

これらより、q_α について考えれば、各カテゴリ α について独立に、q_α と $1-q_\alpha$ の２項分布として β_α の値を推定できることが分かる。そのためこれは２値ロジスティック回帰の拡張として捉えることができ、各カテゴリ p_α（$\alpha=1,2,\cdots,J$）については以下のように与えることができる。

$$p_1=q_1,\ p_\alpha=q_\alpha-q_{\alpha-1},\ p_J=1-q_{J-1}$$

4.4　プログラムの利用法

メニュー［分析－多変量解析等－判別手法－多値ロジスティック回帰］を選択すると図１のような多値ロジスティック回帰分析の分析実行メニューが表示される。

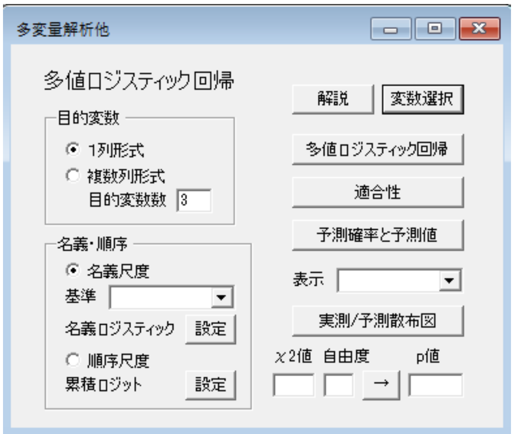


図１　分析実行メニュー

複数列形式のデータの例を図２に示す。


データ編集 多値ロジスティック回帰1.txt

	重要でない	重要	とても重要	性別	年齢	職業
1	26	12	7	0	0	0
2	9	21	15	0	1	0
3	5	14	41	0	0	1
4	40	17	8	1	0	0
5	17	15	12	1	1	0
6	8	15	18	1	0	1

1/2 (1,1)　分析:　備考:

図２　複数列形式のデータ

「目的変数」グループボックスの「複数列形式」を選択し、「目的変数数」テキストボックスに「3」を入力する。変数選択ですべての変数を選択し、「名義ロジスティック」の設定から図３のように基準に「重要でない」を選択する。

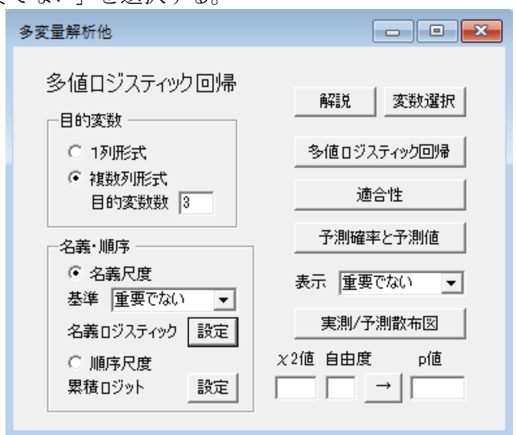


図３　複数列目的変数の名義ロジスティック設定

ここでは、「重要でない」カテゴリの確率で、他のカテゴリの確率を割った対数オッズを説明変数の線形関数で推定することになる。

「多値ロジスティック回帰」ボタンをクリックすると図 4 のような分析結果が表示される。

log(確率/重要でない確率)の線形予測

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
重要/重要でない							
性別	-0.3881	-0.2126	0.2528	0.1247	-0.8835	0.1073	0.6783
年齢	1.1283	0.5826	0.3064	0.0002	0.5277	1.7289	3.0903
職業	1.5877	0.8199	0.3108	0.0000	0.9785	2.1969	4.8925
切片	-0.5908	-0.0616	0.2595	0.0228	-1.0993	-0.0823	
とても重要/重要でない							
性別	-0.8130	-0.4453	0.2710	0.0027	-1.3442	-0.2818	0.4435
年齢	1.4781	0.7633	0.3640	0.0000	0.7647	2.1916	4.3846
職業	2.9167	1.5062	0.3463	0.0000	2.2380	3.5955	18.4805
切片	-1.0391	-0.0668	0.3072	0.0007	-1.6412	-0.4370	

図 4　対数オッズの推定

ここでは、オッズ比推定の偏回帰係数、標準化偏回帰係数、偏回帰係数の標準誤差、偏回帰係数が 0 となる検定確率、偏回帰係数の下限と上限、説明変数の単位量変化によるオッズ比の変化量が表示される。

「適合性」ボタンをクリックすると、図 5 のように各種の適合性指標が表示される。

適合性

対数尤度値	-290.351				
逸脱度D	3.939	自由度	4	上側確率	0.4144
ピアソンχ2	3.927	自由度	4	上側確率	0.4160
C尤度比	77.842	自由度	6	上側確率	0.0000
擬似R2	0.118				
実測予測R2	0.981				

図 5　適合性指標

「予測確率と予測値」ボタンをクリックすると、図 6 のような結果が表示される。

予測確率と予測値

重要でない	予測確率	予測値	重要	予測確率	予測値	とても重要	予測確率	予測値
26	0.524	23.589	12	0.290	13.066	7	0.185	8.345
9	0.235	10.556	21	0.402	18.069	15	0.364	16.375
5	0.098	5.855	14	0.264	15.865	41	0.638	38.280
40	0.652	42.411	17	0.245	15.934	8	0.102	6.655
17	0.351	15.444	15	0.408	17.931	12	0.241	10.625
8	0.174	7.145	15	0.320	13.134	18	0.505	20.721

図 6　予測確率と予測値

これには 3 つのカテゴリについての実測値、予測確率、予測値が表示される。「表示」カテゴリを 1 つ選んで「実測/予測散布図」ボタンをクリックすると、図 7 のような散布図が表示される。

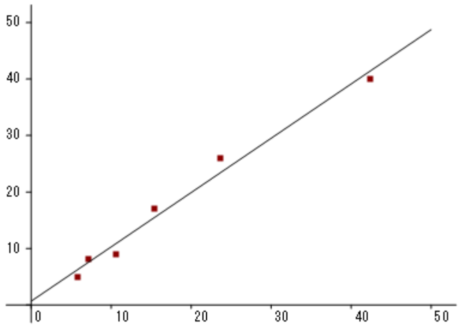


図７　実測/予測散布図

同じデータを順序尺度として、順序ロジスティックの累積ロジットモデルで分析すると図８のような結果を得る。

log(累積確率/他累積確率)の線形予測

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
重要でない/重要～							
性別	0.5723	0.3135	0.2708	0.0346	0.0416	1.1031	1.7724
年齢	-1.2597	-0.6505	0.3064	0.0000	-1.8602	-0.6591	0.2838
職業	-2.2490	-1.1614	0.3577	0.0000	-2.9501	-1.5479	0.1055
切片	0.0746	-0.6751	0.2492	0.7647	-0.4139	0.5631	
～重要/とても重要							
性別	0.5930	0.3248	0.2710	0.0286	0.0619	1.1241	1.8095
年齢	-0.9711	-0.5015	0.3638	0.0076	-1.6841	-0.2581	0.3787
職業	-2.1137	-1.0915	0.3462	0.0000	-2.7923	-1.4351	0.1208
切片	1.5266	0.8221	0.3087	0.0000	0.9216	2.1317	

図８　累積ロジットモデルでの結果

これは最初が「重要でない」を「重要」と「とても重要」を足したカテゴリで割った対数オッズ、次が「重要でない」と「重要」を足したカテゴリを「とても重要」で割った対数オッズについて、説明変数の線形関数での推定である。

最後に目的変数が同じファイル２頁目の「1 列形式」で与えられる場合、「適合性」の結果に図９のように誤判別確率の値が表示される。

適合性

対数尤度値	-8.299				
逸脱度D	16.597	自由度	54	上側確率	1.0000
ピアソンχ2	14.657	自由度	54	上側確率	1.0000
C尤度比	45.054	自由度	4	上側確率	0.0000
擬似R2	0.731				
実測予測R2	0.848				
	Aを他と	Bを他と	Cを他と		
誤判別確率	0.000	0.231	0.250		

図９　１列形式の場合の適合性結果

5. おわりに

この報告では、メタ分析、2 値ロジスティック回帰分析、多値ロジスティック回帰分析のプログラムについて、背景となる理論も含めて紹介した。特に後者 2 つの分析については一般化線形モデルの基礎から解説したので、理論の部分が非常に長くなった。

メタ分析については、効果量とデータ数及び、効果量と分散という 2 種類のデータから分析が実行できるようにした。これらのデータ間の変換はプログラム開発中に考えたものであるが、これで非常に便利に使えるようになったものと思う。ただ、効果量として非常によく利用される t 統計量や χ^2 統計量を用いていないことが気になる。t 統計量については、全データ数が分かると、近似計算を用いて標準化平均値差に変換することができるが、χ^2 統計量については全データ数だけでは対数オッズ比に変換できず、分割表の実数が必要になる。通常分割表は文献に記載されていることが殆どなので、問題は少ないと思われるが、理論的統一性の面では気にかかる部分である。また、メタ分析は、実用としてどの程度役に立つものか、未知の部分が多い。今後何回か分析を実行することによってプログラムの改善が必要になるものと思われる。

2 値ロジスティック分析については、一般化線形モデルの 2 項分布のパラメータの推定で、3 種類の連結関数を用いたモデルを提供している。分析の名前はその中の代表的なロジスティックモデルを元にしている。データは 2 値の度数データと 2 値の個体別データを利用できる。個体別データの場合は、2 群の判別分析と同じような使い方ができる。但し、群の推測において、個体別の判別関数値ではなく、事象の出現確率が表示される。判別の分点に相当するものは出現確率で 0.5 の値である。

多値ロジスティック分析については、一般化線形モデルの多項分布のパラメータ推定で、目的変数が名義尺度の場合と順序尺度の場合に分けて結果が表示される。名義尺度の場合は名義ロジスティックモデル、順序尺度の場合は累積ロジットモデルが使われる。他のモデルについては必要が生じたら追加する。データは多値の度数データと多値の個体別データを利用できる。個体別データの場合は、多群の判別分析と同じような使い方ができるが、この分析では、事象の出現確率が最大と予測される事象が判別値となる。

参考文献

[1] 山田剛史, 井上俊哉編, メタ分析入門 心理・教育研究の系統的レビューのために, 東京大学出版会, 2012.

[2] Annette J. Dobson 著, 田中豊他訳, 一般化線形モデル入門 原著第 2 版, 共立出版, 2008.

[3] ホームページ http://www.heisei-u.ac.jp/ba/fukui/analysis.html 内のサンプルデータ Samples.zip 内のファイル

社会システム分析のための統合化プログラム２８
－メタ分析・ロジスティック回帰分析－

Multi-purpose Program for Social System Analysis 28

- Meta-Analysis, Logistic Regression Analysis -

Masayasu FUKUI[*1], Kazuki KODAMA[*1] and Makoto OZAKI[*1]

[*1] *Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we make program of meta-analysis which is a method for leading stronger conclusion using with past research materials, and of Logistic regression analysis to calculate the probability of binomial distribution or multinomial distribution as a function of polynomial of independent variables. We explain the theory of these analyses and introduce our programs in this paper.

Keywords: College Analysis, multivariate analysis, generalized linear model, meta-analysis, logistic regression analysis

URL: http://www.heisei-u.ac.jp/ba/fukui/