8. 推測

ここでは説明変数と目的変数の関係を求め、その関係を使って未知の目的変数を予測します。

8.1. 重回帰分析

重回帰分析(Multiple regression)とよばれる方法によって、次のような複数の説明変数($x_1, x_2, \ldots$)と 1 個の目的変数(y: Y_n)をもつデータから、未知の目的変数を予想する重回帰式を求めます。各説明変数に重み（負荷、ウェイト）W_p を掛けて重回帰式を作りますが、実際の結果 Y_n と重回帰式で求めた予測値ベクトル E_n の差が小さければ小さいほどその式が高く評価されます。そこで、実測値ベクトル Y_n と予測値ベクトル E_n の平方和が最小になるようにします。

たとえば、次のような成績表で、小テスト 3 回(x1, x2, x3)と、最終成績(POINT)の関係を見ます。

X	x1	x2	x3	y
d1	6	8	5	12
d2	7	10	6	11
d3	8	4	8	13
d4	9	7	2	7
d5	10	9	4	14

ここで、POINT に該当する予測値 E_n を、切片 W(0)と、各変数(X)に重みとしての係数(W_p)を掛けたものを加算して作った式から求めます。[i = 1, 2, …, n]

[1] $$E(i) = W(0) + W(1)\,X(i, 1) + W(2)\,X(i, 2) + \ldots + W(p)\,X(i, p)$$

この式の第 1 項 W(0)は回帰式の切片(intercept)を示します。この切片をすべての個体(1, 2, …, n)に共通に加えます。したがって、この列には単位ベクトル I_p を左積します。

$$E(i) = I_p\,W(0) + X(i, 1)\,W(1) + X(i, 2)\,W(2) + \ldots + X(i, {}_p)\,W({}_p) \qquad [i = 1\ldots n]$$

行列で示すと

$$E_n = X_{np}\,W_p \quad [X_{np} \text{の第 1 列は単位ベクトル}]$$

この式で求められた値と実測値 Y_n の間の残差のベクトルを R_n とします。

[2]　　$R_n = Y_n - E_n = Y_n - X_{np}\, W_p$

この残差 R_n の平方和 S を求めます。

$$
\begin{aligned}
S &= R_n^T R_n = (Y_n - X_{np}\, W_p)^T (Y_n - X_{np}\, W_p) && \leftarrow \text{上式[2]} \\
&= [Y_n^T - (X_{np}\, W_p)^T] (Y_n - X_{np}\, W_p) && \leftarrow \text{転置行列の性質(T)} \\
&= Y_n^T Y_n - Y_n^T X_{np}\, W_p - (X_{np}\, W_p)^T Y_n + (X_{np}\, W_p)^T X_{np}\, W_p && \leftarrow \text{展開} \\
&= Y_n^T Y_n - Y_n^T X_{np}\, W_p - Y_n^T (X_{np}\, W_p) + W_p^T X_{np}^T X_{np}\, W_p && \leftarrow \text{T} \\
&= Y_n^T Y_n - 2\, Y_n^T X_{np}\, W_p + W_p^T X_{np}^T X_{np}\, W_p && \text{2, 3 項を整理}
\end{aligned}
$$

この式中の W_p は未知数です。重回帰分析の目的は、この残差平方和 S を最小化することです。そのために、S を変数のベクトル W_p で微分し（→後述）、その値がゼロベクトル(O_p^T)になるときの W_p を求めることです（多変数空間中の変数が形成する「曲面」の最小値の位置座標をイメージしてください）。

ここで、$S = Y_n^T Y_n - 2\, Y_n^T X_{np}\, W_p + W_p^T X_{np}^T X_{np}\, W_p$ の第 1 項 $Y_n^T Y_n$ には W_p がないので、W_p で微分するとゼロになります。第 2 項の $-2\, Y_n^T X_{np}\, W_p$ と第 3 項の $W_p^T X_{np}^T X_{np}\, W_p$ の微分については後述します。第 3 項の中の $X_{np}^T X_{np}$ は対称行列です。

$$\frac{\partial S}{\partial W_p} = -2\, X_{np}^T Y_n + 2\, X_{np}^T X_{np}\, W_p = O_p^T$$

$$
\begin{aligned}
&2\, X_{np}^T X_{np}\, W_p = O_p^T + 2\, X_{np}^T Y_n && \leftarrow 2\, X_{np}^T Y_n \text{を移項} \\
&X_{np}^T X_{np}\, W_p = X_{np}^T Y_n && \leftarrow O_p^T \text{はゼロベクトル} \\
&(X_{np}^T X_{np})^{-1} (X_{np}^T X_{np})\, W_p = (X_{np}^T X_{np})^{-1} X_{np}^T Y_n && \leftarrow \text{（注[1]）} \\
&I_{pp}\, W_p = (X_{np}^T X_{np})^{-1} X_{np}^T Y_n && \leftarrow A\, A^{-1} = I_{pp} \\
&W_p = (X_{np}^T X_{np})^{-1} X_{np}^T Y_n && \leftarrow I_{pp}\, A = A
\end{aligned}
$$

このようにして求めたベクトル W_p が下に示す「係数」(Value)の列です。

Mr.w.	x1	x2	x3	Interc.	R.m.:Prp
Org.	.740	.462	1.157	-3.819	1.410
Std.	.433	.394	.957	.000	.591

予測値(E_n) は前述の式[1]で求めます。残差ベクトル(Residual: Res: R_n)は、次の式で求めます。

$$R_n = Y_n - E_n$$

なお、上表の R.m.(1.410)は残差の絶対値の平均を示します(Residual

[1] W_p を求めるためには左辺を Wp にします。そのために W_p の係数を単位行列 I_{pp} にする必要があるので両辺に$(X_{np}^T X_{np})^{-1}$ を左積します。

mean)。

Res.Ratio = Sm(AbsM(Rn)) / N

また、Prp (.591)は寄与率(Proportion)と呼ばれる数値で、重回帰式によって得られた分散(情報)が全体の分散(情報)に占める割合を示します[2]。

Prp. = Σ [En - Am(Yn)]^2 / Σ [Yn - Am(Yn)]^2

そして、上表の Std は説明変数行列と目的変数を標準化して行列について重回帰分析した結果を示しています。説明変数と目的変数をそれぞれ標準化すると（→「標準得点」）、それらの平均が 0 となり、その回帰直線は座標の原点(0, 0)を通るので、回帰式の切片がなくなります（→「相関係数」）。また、変数をその標準偏差で割っているので、次表のように、説明変数のバラツキをなくした標準化された「重み」が計算されます。変数の重みを比較するためには、この重みのほうが適しています。

このようにして求めた回帰式を先の[1]

$$E_n = X_{np} W_p$$

によって次の導出変数(Derived)を計算します。データの目的変数(y)と回帰式による導出変数(y^)を比較してください。そのとき残差(Res)が参考になります。

X	x1	x2	x3	y	y^	Res.
d1	6.000	8.000	5.000	12.000	10.104	-1.896
d2	7.000	10.000	6.000	11.000	12.926	1.926
d3	8.000	4.000	8.000	13.000	13.207	.207
d4	9.000	7.000	2.000	7.000	8.392	1.392
d5	10.000	9.000	4.000	14.000	12.371	-1.629

次に、先にもとめた係数ベクトル Wp を、目的変数が未知のデータ Dnp に右積して、次の予測変数 Zn を求めます。

Zn = Dnp Wp

ここで、求めた係数からなる重回帰式[1]を使って、目的変数(POINT)が未知のデータで、その目的変数を予測してみましょう。次のデータ X.e の e1 は先の d1 と同じです。よって、同じ係数を掛けた予測変数(Expected)は当然先の d1:Derived と同じになります(10.104)。e2, e3 は変数の値が異なります。よって、それに応じて予測変数が変化しています。

[2] Prp の分子も分母も分散であり同じ個数(N)で割りますから、N の割り算を省きます。

X.e	x1	x2	x3	Exp.
e1	6.000	8.000	5.000	10.104
e2	6.000	8.000	2.000	6.633
e3	5.000	5.000	4.000	6.821
e4	7.000	5.000	4.000	8.301
e5	8.000	5.000	9.000	14.826

●多重共線性

次は、それぞれの説明変数と目的変数を合わせた相関行列です。これを見ると、x3 と POINT の相関が他と比べて高いことがわかります[3]。

C.Cor	x1	x2	x3	POINT
x1	1.0000	-.0687	-.4243	.0000
x2	-.0687	1.0000	-.3885	-.0080
x3	-.4243	-.3885	1.0000	.6207
POINT	.0000	-.0080	.6207	1.0000

重回帰分析をするとき、このような変数間の相関係数を見る必要があるのは、説明変数と目的変数の相関が変数のポジティブな評価に役立つだけでなく、説明変数どうしの相関がネガティブに問題を引き起こすためです。係数間に強い相関があるときは、そのことが影響して異常な係数を生み出します。

このことは重回帰式が説明変数に重みを掛けた積の和になっていることから理解できます。たとえば説明変数 X(i, 1)と X(i, 2)の間に.98 などの強い相関があるとすると、回帰式の総和（積和）としての目的変数は一定なので、この 2 つの変数の値は競合して分け合うことになります。一方が強く働けば、他方を弱くしなければなりません。符号がプラスからマイナスに変わってしまうこともあります。もし、経験や直感から判断して、係数の符号（プラス・マイナス）が逆転していたり[4]、変数の係数がその変数の重要度を反映していないようなことなどが起きていれば、係数間の相関が高い可能性が高いのです。これは**多重共線性**(multicollinearity)とよばれる問題です。極端な場合は変数間の相関係数が 1 のときです。これでは、2 つの変数に固有の情報がなく 1 つの情報だけで十分になります。そして、重回帰式の中で使われる逆行列の計算（→後述）が不可能になります。相関係数が高い場合も情報が少ないので、同様の問題が起きます。そのとき

[3] それぞれのセルの右項は確率を示します。このデータでは、データ数が少ないので、どの相関係数もあまり確率は高くありません。

[4] 回帰式の係数 Wp の符号（プラス・マイナス）が、説明変数と目的変数の符号が異なっているときは、多重共線性の問題があります。この例では、回帰式の係数の符号はすべてプラスですが、x2:POINT の相関がマイナス値(-.008)になっています。

は、重要な変数だけを残し、回帰式を単純化して、残りの重要な変数に注目する、という手段がとられます。

回帰式に多くの係数を入れると、それだけあてはまりがよくなりますが、それは与えられたデータについてのあてはまりにすぎません。予測の一般性を高めるためには、実験を繰り返して適切な変数を選択し、なるべく少ない変数で予測式を求めるべきです。そうすれば目的変数を説明する変数が何なのかを的確に、そして「きれいに」示すことができるからです。また、複数の相関が高い変数群の中から 1 つを選ぶことによって変数のグルーピング（→「分類」）ができるので、変数間の関係の理解につながります。せっかく集めた変数の分布データを捨てるのが惜しい、ということでしたら、相関する変数の「どちらにも当てはまるケース」の頻度を計算し、これを新たな変数として使う、ということも考えられます。または、相関する変数の「どちらかに当てはまるケース」を数えて、比べてみるとよいでしょう。さらに後述する「主成分重回帰分析」を使えば、変数の情報をすべてつかった重回帰分析ができます。

●逆行列

(1) 逆行列の定義

正方行列(X_{pp})について

$$X_{pp}\ Y_{pp} = I_{pp}\ (\text{単位行列}) \rightarrow\ Y_{pp} = X_{pp}^{-1}$$

となる正方行列(Y_{pp})は、X_{pp} の「逆行列」(Inverse matrix: X_{pp}^{-1})とよばれます。逆行列が関係する次の演算は統計の計算によく使われます。

(a) $X_{pp}\ X_{pp}^{-1} = I_{pp}$

X_{pp}	1	2
1	7	8
2	9	10

×

X_{pp}^{-1}	1	2
1	-5	4
2	4.5	-3.5

=

I	1	2
1	1	0
2	0	1

(b) $X_{pp}^{-1}\ X_{pp} = I_{pp}$

X_{pp}^{-1}	1	2
1	-5.0	4.0
2	4.5	-3.5

×

X_{pp}	1	2
1	7	8
2	9	10

=

I_{pp}	1	2
1	1	0
2	0	1

(2) 逆行列の性質

(a) $(X_{pp}^{-1})^{-1} = X_{pp}$

X_{pp}	1	2
1	7	8
2	9	10

→

X_{pp}^{-1}	1	2
1	-5.0	4.0
2	4.5	-3.5

→

$(X_{pp}^{-1})^{-1}$	1	2
1	7	8
2	9	10

(b) $(X_{pp}\ Y_{pp})^{-1} = Y_{pp}^{-1}\ X_{pp}^{-1}$

X_{pp}	1	2
1	1	3
2	2	4

×

Y_{pp}	1	2
1	7	8
2	9	1

=

$X_{pp}Y_{pp}$	1	2
1	34	11
2	50	20

→

$(X_{pp}Y_{pp})^{-1}$	1	2
1	0.154	-0.085
2	-0.385	0.262

Y_{pp}^{-1}	1	2
1	-2.00	1.500
2	1.00	-0.500

×

X_{pp}^{-1}	1	2
1	-0.015	0.123
2	0.136	-0.108

=

$Y_{pp}^{-1}\ X_{pp}^{-1}$	1	2
1	0.154	-0.085
2	-0.385	0.262

(c) $(X_{pp}^{T})^{-1} = (X_{pp}^{-1})^{T}$

X_{pp}	1	2
1	7	8
2	9	10

→

X_{pp}^{T}	1	2
1	7	9
2	8	10

→

$(X_{pp}^{T})^{-1}$	1	2
1	-5.0	4.5
2	4.0	-3.5

X_{pp}	1	2
1	7	8
2	9	10

→

X_{pp}^{-1}	1	2
1	-5.0	4.0
2	4.5	-3.5

→

$(X_{pp}^{-1})^{T}$	1	2
1	-5.0	4.5
2	4.0	-3.5

(3) 逆行列の求め方

与えられた行列(X_{pp})と、初期値が単位行列である行列($Z_{pp}=I_{pp}$)を同時に変形していきます。X_{pp} が単位行列(I_{pp})になるように、X_{pp} と Z_{pp} に左から変形行列 T_{pp} を繰り返して掛けていきます。そのために

(a) 2 つの行を交換する T_{pp}
(b) 実数倍した 1 つの行全体に、実数倍した他の行を加算する T_{pp}

という 2 つの変換を使います。これらの変換を可能にする変形行列 T_{pp} を次々に左積すると、Z_{pp} が A_{pp} の逆行列になることを次の演算で確認しましょう(「Gauss の消去法」 Gauss reduction)。

0. $X^{(0)}, Z^{(0)} = I$ ←X, Z の初期状態$^{(0)}$
1. $X^{(1)} = T^{(1)}\ X^{(0)}, Z^{(1)} = T^{(1)}I$ ←$X^{(0)}$と $Z^{(0)}$=I に $T^{(1)}$を左積

2. $X^{(2)} = T^{(2)}\ T^{(1)}\ X^{(0)}$, $Z^{(2)} = T^{(2)}\ T^{(1)}I$ ←さらに $T^{(2)}$を左積
   (…) ←さらに $T^{(3)}$, …,$T^{(k)}$を順次左積
3. $I = T^{(k)} \ldots T^{(2)}\ T^{(1)}\ X^{(0)}$ ←$X^{(0)}$に T を順次左積し I に至る
4. $Z^{(k)} = T^{(k)} \ldots T^{(2)}\ T^{(1)}I$ ←$Z^{(0)}$ =I に T を順次左積し $Z^{(k)}$を得る
5. $I\ X^{(0)-1} = T^{(k)} \ldots T^{(2)}\ T^{(1)}\ X^{(0)}\ X^{(0)-1}$ ←3 の両辺に $X^{(0)-1}$を右積
6. $X^{(0)-1} = T^{(k)} \ldots T^{(2)}\ T^{(1)}I$ ←5. I A = A; A A^{-1} =I
7. $Z^{(k)} = X^{(0)-1}$ ←4.右辺= 6.右辺、よって $Z^{(k)}$は $X^{(0)}$の逆行列になる

たとえば次の行列 $X^{(0)}$の逆行列を求めることを考えましょう。以下の演算のために、作業用の行列 $T^{(1)}$ と出力用の単位行列 $Z^{(1)}$= I を用意します。目的は $T^{(1)}$, $T^{(2)}$, …, $T^{(k)}$の左積を繰り返して、$X^{(k)}$を単位行列にすることです。

X(0)	1	2	3
1	**0**	2	1
2	2	1	2
3	2	1	1

Z(0)	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1

はじめに、X(1, 1)を 1 にするために次の演算をします。

R1 ← R1 / X(1, 1)

これは X の第一行 R1 を X(1, 1)で割って新たな R1 にする、ということです。ここでは、X(1, 1)が 0 なので割り算ができません。そのときは、第一列 C1 が 0 でない行と交換します。その結果 $X^{(1)}$となります。

R1 ← R2, R2 ← R1

$X^{(1)}$	1	2	3
1	**2**	**1**	**2**
2	**0**	**2**	**1**
3	2	1	1

$Z^{(1)}$	1	2	3
2	**0**	**1**	**0**
1	**1**	**0**	**0**
3	0	0	1

こうして新たな X(1, 1)←2 で先の除算をします。

R1 ← R1 / X(1, 1) ← R1 / 2

$X^{(2)}$	1	2	3
1	**2/2=1**	**1/2**	**2/2=1**
2	0	2	1
3	2	1	1

$Z^{(2)}$	1	2	3
1	**0/2=0**	**1/2**	**0/2=0**
2	1	0	0
3	0	0	1

次に R2 と R3 を R1 を使って、それぞれの C1 の値を 0 にします。ここでは R2 の X(2, 1)が 0 なので R3 だけを次のようにして変えます。

R3 ← R3 – X(3, 1) R1 ← R3 – 2 R1

$X^{(3)}$	1	2	3
1	1	1/2	1
2	0	2	1
3	**2-2*1=0**	**1-2*(1/2)=0**	**1-2*1=-1**

$Z^{(3)}$	1	2	3
1	0	1/2	0
2	1	0	0
3	**0-2*0=0**	**0-2*1/2=-1**	**1-2*0=1**

これで C1 は完成です。次に同様なことを C2 で行います。

$X^{(4)}$	1	2	3
1	1	1/2	1
2	0	**2**	1
3	0	0	-1

$Z^{(4)}$	1	2	3
1	0	1/2	0
2	1	0	0
3	0	-1	1

今度は X(2,2)=2 は 0 でないので、そのまま R2 を 2 で割ります。

R2 ← R2 / X(2, 2) ← R2 / 2

$X^{(5)}$	1	2	3
1	1	1/2	1
2	**0/2=0**	**2/2=1**	**1/2**
3	0	0	-1

$Z^{(5)}$	1	2	3
1	0	1/2	0
2	**1/2**	**0/2**	**0/2**
3	0	-1	1

そして R1 と R2 の C2 を次の演算で 0 にします。

R1 ← R1 – X(1, 2) R2 ← R1 – 1/2 R2
R3 ← R3 – X(3, 2) R2 ← R3 – 0 R2

$X^{(6)}$	1	2	3
1	**1-(1/2)*0 =1**	**1/2-(1/2)*1 =0**	**1-(1/2)*(1/2)=3/4**
2	0	1	1/2
3	**0-0*0=0**	**0-0*1=0**	**-1-0*(1/2)=-1**

$Z^{(6)}$	1	2	3
1	**0-(1/2)*(1/2) =1/4**	**1/2-(1/2)*0 =1/2**	**0-(1/2)*0 =0**
2	1/2	0	0
3	**0-0*(1/2)=0**	**-1-0*0=-1**	**1-0*0=1**

これで C2 は完成です。次に同様なことを C3 で行います。

$X^{(7)}$	1	2	3
1	1	0	3/4
2	0	1	1/2
3	0	0	**-1**

$Z^{(7)}$	1	2	3
1	1/4	1/2	0
2	1/2	0	0
3	0	-1	1

R3 ← R3 / X(3, 3) ← R3 / -1

$X^{(8)}$	1	2	3

$Z^{(8)}$	1	2	3

1	1	0	3/4
2	0	1	1/2
3	**0/-1=0**	**0/-1=0**	**-1/-1=1**

1	1/4	1/2	0
2	1/2	0	0
3	**0/-1=0**	**-1/-1=1**	**1/-1=-1**

R1 ← R1 – X(1, 3) R3 ← R1 – 3/4 R3
R2 ← R1 – X(2, 3) R3 ← R1 – 1/2 R3

$X^{(9)}$	1	2	3
1	1-(3/4)x0 =1	0-(3/4)x0 =0	3/4-(3/4)x1 =0
2	0-(1/2)x0 =0	1-(1/2)x0 =1	1/2-(1/2)x1 =0
3	0	0	1

$Z^{(9)}$	1	2	3
1	1/4-(3/4)x0 =-1/4	1/2-(3/4)x1 =-1/4	0-(3/4)－1 =3/4
2	1/2-(1/2)x9 =1/2	0-(1/2)x1 =-1/2	0-(1/2)x-1 =1/2
3	0	1	-1

これらの演算の結果、次のようにXは単位行列になり、ZにXの逆行列が得られました。

$X^{(k)}$	1	2	3
1	1	0	0
2	0	1	0
3	0	0	**1**

$Z^{(k)}$	1	2	3
1	-1/4	-1/4	3/4
2	1/2	-1/2	1/2
3	0	1	-1

プログラムで実行すると確かにXの逆行列X^{-1}が得られ、XとX^{-1}の行列積を計算すると単位行列が得られます。

X	1	2	3
1	0	2	1
2	2	1	2
3	2	1	1

X^{-1}	1	2	3
1	-.250	-.250	.750
2	.500	-.500	.500
3	.000	1.000	-1.000

$X X^{-1}$	1	2	3
1	1	0	0
2	0	1	0
3	0	0	1

＊長谷川(2000:129-136)を参照しました。プログラムは縄田(1999:58-80)を参照しました。

(4) 逆行列演算の証明

次の演算はしばしば使われます。しっかりと理解しておくために証明をしておきましょう。

[1] $I^{-1} = I$

$I I^{-1} = I$ ←逆行列の定義: $X X^{-1} = I$, ここで $X = I$]
$I^{-1} = I$ ←$I X = X$, $X=I$

[2] $(A^{-1})^{-1} = A$

$A^{-1} (A^{-1})^{-1} = I$ ←逆行列の定義: $A A^{-1} = I$
$A A^{-1} (A^{-1})^{-1} = A I$ ←両辺に A を左積
$I (A^{-1})^{-1} = A I$ ←逆行列の定義: $A A^{-1} = I$
$(A^{-1})^{-1} = A$ ←$X I = X$; $I X = X$

[3] $(A B)^{-1} = B^{-1} A^{-1}$

$(A B) (A B)^{-1} = I$ ←$X X^{-1} = I$, $X = A B$
$(A B) (A B)^{-1} = A A^{-1}$ ←$A A^{-1} = I$
$(A B) (A B)^{-1} = A I A^{-1}$ ←$A = A I$
$(A B) (A B)^{-1} = A B B^{-1} A^{-1}$ ←$I = B B^{-1}$
$(A B)^{-1} = B^{-1} A^{-1}$ ←両辺から A B を削除

[4] $A A^{-1} = A^{-1} A$

$A A^{-1} = I$ ←逆行列の定義: $A A^{-1} = I$
$(A^{-1} A) (A A^{-1}) = (A^{-1} A) I$ ←両辺に $A^{-1} A$ を左積
$A^{-1} A A A^{-1} = A^{-1} A$ ←$X I = X$, $X = A^{-1} A$
$I A A^{-1} = A^{-1} A$ ←$A^{-1} A = I$
$A A^{-1} = A^{-1} A$ ←$I A = A$

＊[2], [3]は足立(2005:110-111)を参照しました。

プログラム[5]

```
Function Iv(ByVal Xpp) '逆行列(Gauss-Jordan 法. ver. 2013/06/28-2015/1/22)
  Dim TT$, P&, i&, j&, Tpp, Zpp, E: P = NC(Xpp): E = -15 'P=行数=列数
  TT$ = Xpp(0, 0): Zpp = Um(P) 'X 対象の行列 : Zpp 単位行列
  For i = 1 To P '1 列から P 列まで
    If Abs(Xpp(i, i)) < 10 ^ E Then '対角成分が 0 ならば行交換
      For j = i + 1 To P 'i+1 行から P 行まで
        If i < P And Abs(Xpp(j, i)) > 10 ^ E Then '非対角成分が 0
          Tpp = Um(P): Tpp(i, i) = 0: Tpp(j, j) = 0: Tpp(i, j) = 1: Tpp(j, i) = 1
                                                    '変形行列
          Xpp = X(Tpp, Xpp): Zpp = X(Tpp, Zpp) 'i 行と j 行を交換
          Exit For 'For j を脱出
        End If
      Next j
```

[5] Tpp(i, i) = 1 / Xpp(i, i)を、最終プロセスではなく各行のプロセスに置くことによって数値のオーバーフローを回避する方法は堀川遼太さんからいただいたアイデアです(2013)。

```
    End If
    If Xpp(i, i) = 0 Then '対角成分=0
      MsgBox Ln(29): Exit Function 'Msg「逆行列は存在しません。」
    End If
    For j = 1 To P '1 行から P 行まで、非対角成分=0, 対角成分=1
      If i <> j And Abs(Xpp(j, i)) > 10 ^ E Then
        Tpp = Um(P): Tpp(i, i) = 1 / Xpp(i, i) '変形行列  (Horikawa 2013)
        Xpp = X(Tpp, Xpp): Zpp = X(Tpp, Zpp) 'X(i, i) = 1
        Tpp = Um(P): Tpp(j, i) = -1 * Xpp(j, i) '変形行列
        Xpp = X(Tpp, Xpp): Zpp = X(Tpp, Zpp) 'Rj=Rj-X(j,i)*Ri→X(j,i) = 0
      End If
    Next j
  Next i
  Zpp(0, 0) = TT$ & "^":     Iv = Zpp '返し値
End Functio
```

●変形行列

単位行列の一部を変更した行変形用行列を作成し、これをある行列に左積すると、一定の行変形ができます。ここではそのような行列を「変形行列」(Transformation matrix)とよぶことにします。これらを逆行列の計算に使います。

(a) R1 ← 0

T_{pp}	1	2	3
1	0	0	0
2	0	1	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	0	0	0
2	4	5	6
3	7	8	9

(b) R1 ← R2

T_{pp}	1	2	3
1	0	1	0
2	0	1	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	4	5	6
2	4	5	6
3	7	8	9

(c) R1 ~ R2 (交換)

T_{pp}	1	2	3
1	0	1	0
2	1	0	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	4	5	6
2	1	2	3
3	7	8	9

(d) R1 ← 3 R1 (倍数)

T_{pp}	1	2	3
1	3	0	0
2	0	1	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	3	6	9
2	4	5	6
3	7	8	9

(e) R2 ← R2 + R1

T_{pp}	1	2	3
1	1	0	0
2	1	1	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	1	2	3
2	5	7	9
3	7	8	9

(f) R2 ← R2 + 2 R1

T_{pp}	1	2	3
1	1	0	0
2	2	1	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	1	2	3
2	6	9	12
3	7	8	9

(g) R2 ← 3 R2 + 2 R1

T_{pp}	1	2	3
1	1	0	0
2	2	3	0
3	0	0	1

×

A_{pp}	1	2	3
1	1	2	3
2	4	5	6
3	7	8	9

=

R_{pp}	1	2	3
1	1	2	3
2	14	19	24
3	7	8	9

最後の演算を見ると、変形行列の対角成分で自分の行を積算し、非対角成分でその列番にあたる行を積算していることがわかります。行のゼロ化[1]や行の移動[2][3]も同様です。

＊芝(1975: 197-199)を参照しました。

●行列の微分

多変量分析ではしばしば行列をベクトルで微分します。行列の積の成分を展開すればベクトルで微分した結果が行列とベクトルの積になることがわかります。

[1] はじめに、次のような行列 T_{pp} の W_p による微分について見ましょう。

$$T_{pp} = Y_n{}^T X_{np} W_p = [y_1, y_2, \ldots, y_n] \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_p \end{bmatrix}$$

を，「ベクトル Wp = $\begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_p \end{bmatrix}$ で微分する」ということの意味を理解するために T_{pp} を展開します。

$$T_{pp} = [y_1 x_{11} + y_2 x_{21} + \ldots + y_n x_{n1},$$
$$y_1 x_{12} + y_2 x_{22} + \ldots + y_n x_{n2},$$
$$\ldots,$$
$$y_1 x_{1n} + y_2 x_{2n} + \ldots + y_n x_{np}\ ] \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_p \end{bmatrix}$$

$$= (y_1 x_{11} + y2\ x_{21} + \ldots + y_n x_{n1}) * w_1$$
$$+ (y_1 x_{12} + y_2 x_{22} + \ldots + y_n x_{n2})\ w_2$$
$$+ \ldots$$
$$+ (y_1 x_{1n} + y_2 x_{2n} + \ldots + y_n x_{np}) * w_p$$

偏微分の記号 $\frac{\partial S}{\partial a}$ を Df(S, w)で示すと（「S を w で微分する」という意味）

$Df(T_{pp}, w_1) = y_1 x_{11} + y_2 x_{21} + \ldots + y_n x_{n1}$ (上式の 1 行目)
$Df(T_{pp}, w_2) = y_1 x_{12} + y_2 x_{22} + \ldots + y_n x_{n2}$ (上式の 2 行目)
…
$Df(T_{pp}, w_p) = y_1 x_{1p} + y_2 x_{2p} + \ldots + y_n x_{np}$ (上式の p 行目)

これらをまとめて示すと次のようになります。

$Df(T_{pp}, W_p) = Df(Y_n{}^T X_{np} W_p, W_p) = X_{np}{}^T Y_n$ [←縦ベクトル]

高等学校で既習の次の微分と比べてみてください。

Df(yxw, w) = yx

[2] 次は微分する項(Wp)が 2 乗されている場合です。たとえば

$$T_{pp} = W_p{}^T X_{pp} W_p = [w_1, w_2, \ldots, w_p] \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{12} & x_{22} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ x_{1p} & x_{2p} & \ldots & x_{pp} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_p \end{bmatrix}$$

を、ベクトル $W_p = \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_p \end{bmatrix}$ で微分します。ここでは X_{pp} を対称行列とします。

$$T_{pp} = [w1\ x_{11} + w1\ x_{12} + \ldots + w1\ x_{1p},$$

$$w1\ x_{21} + w2\ x_{22} + \ldots + w2\ x_{2p},$$
$$\ldots,$$
$$w1\ x_{n1} + w2\ x_{n2}\ w2 + \ldots + w_p\ x_{np}\ ] \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_p \end{bmatrix}$$

$$= w1\ x_{11}\ w1 + w1\ x_{12}\ w2 + \ldots + w1\ x_{1p}\ w_p$$
$$+ w2\ x_{12}\ w1 + w2\ x_{22}\ w2 + \ldots + w2\ x_{2p}\ w_p$$
$$+ \ldots$$
$$+ w_p\ x_{1p}\ w1 + w_p\ x_{2p}\ w2 + \ldots + w_p x_{pp}\ w_p$$

$$= x_{11}\ w1^2 + w1\ x_{12}\ w_2 + \ldots + w_1\ x_{1p}\ w_p$$
$$+ w_2\ x_{12}\ w_1 + x_{22}\ w_2^2 + \ldots + w_2\ x_{2p}\ w_p$$
$$+ \ldots$$
$$+ w_p\ x_{1p}\ w1 + w_p\ x_{2p}\ w_2 + \ldots + x_{pp}\ w_p^2$$

この式で w1 を含む成分は 1 行と 1 列の成分です。よって

$$Df(T_{pp}, w1) = 2w1\ x_{11} + 2(w2\ x_{12} + \ldots + w_p\ x_{1p}) = 2(w1\ x_{11} + w2\ x_{12} + \ldots + w_p\ x_{1p})$$

同様に、w2 を含む成分は 2 行と 2 列の成分です。よって

$$Df(T_{pp}, w2) = 2w2\ x_{12} + 2(w2\ x_{22} + \ldots + 2w_p\ x_{2p}) = 2(w2\ x_{12} + w2\ x_{22} + \ldots + w_p\ x_{2p})$$

…

同様にして

$$Df(T_{pp}, w_p) = 2w_p\ x_{1p} + 2(d2\ x_{2p} + \ldots + w_p\ x_{pp}) = 2(w_p\ x_{1p} + {}_2\ x_{2p} + \ldots + w_p\ x_{pp})$$

以上をまとめて示すと次のようになります。

$$Df(T_{pp}, Wp) = Diff.\ (Wp^T\ X_{pp}\ Wp, Wp) = 2\ X_{pp}\ Wp$$

次の微分と比べてみてください。

$$Df(wxw,\ w) = 2xw$$

●数量化 I 類

次のように、説明変数が数量ではなく質的なデータ(v)を扱うとき、これを 0-1 に変換して、同様に重回帰分析をすることができます。この方法は**数量化 I 類**(Quantification method of first type)とよばれます。

X	v1	v2	v3	POINT
d1		v		12
d2	v	v	v	11
d3	v		v	13
d4	v	v		7
d5	v	v		14

X	POINT	Expected	Residual
d1	12.000	12.000	.000
d2	11.000	11.000	.000
d3	13.000	13.000	.000
d4	7.000	10.500	-3.500
d5	14.000	10.500	3.500

Weight	P: Intercept	v1	v2	v3	Std res.
Value	14.000	-1.500	-2.000	.500	2.214

この方法を使用するにあたって注意しなければならないのは、次のようなケースです。

X	v1	v2	v3	POINT
d1	v	v		12
d2	v	v	v	11
d3	v		v	13
d4	v	v		7
d5	v	v		14

X	v1	v2	v3	POINT
d1		v		12
d2	v	v		11
d3	v		v	13
d4	v	v		7
d5	v	v		14

上左表では v1 がすべて選択されていますので、この v1 には弁別する情報がありません。また、右表では v2 と v3.が相補分布(complementary distribution)をしています。この場合は、どちらかを選択すれば他方が決まっているので、どちらか 1 つにしか弁別する情報がないことになります。このような行列ではすべて逆行列が存在せず分析ができないので、データから該当する行を取捨選択しなければなりません。

■文字頻度の変遷と年代

下左表は 13~19 世紀の文字母数を揃えたスペイン語文献の特定の文字の頻度と　文献の成立年代(Y)を示します。下右表は重回帰分析の結果です。<*>は文字が略されている箇所の頻度を示します。

Obra	<*>	ñ	è	á	τ	y	y^	Residual
Cid	836				144	1207	1396	-189
Fazienda	902				157	1220	1382	-162
Alcalá	921				444	1230	1249	-19
GE	1,349				301	1270	1266	4
Alexandre	877				78	1300	1421	-121
Lucanor	1,877				227	1330	1241	89
Troyana	1,105				399	1350	1249	101

LBA	1,366			146	1389	1335	54
Alba	464	156		543	1433	1485	-52
Especulo	1,024	52		215	1450	1419	31
Gramática	577	51	4	192	1492	1482	10
Celestina	573	41		131	1499	1491	8
Sumario	329	70		322	1514	1474	40
Diálogo	561				1535	1492	43
Lazarillo	297	33		142	1554	1505	49
Casada	139	40			1583	1598	-15
Quijote	165	57	3	2	1605	1621	-16
Buscón	93	47	7	1	1626	1617	9
Criticón	147	45	20		1651	1616	35
Instante	4	21	94	2	1677	1641	36
Austria	7	60	39		1704	1665	39
Autoridades		27	3	196	1726	1780	-54
Picarillo	4	123	108		1747	1798	-51
Delincuente		42		229	1787	1831	-44
Ortografía		35		93	1815	1694	121
Diablo		55		223	1841	1845	-4
Sombrero		89		222	1874	1894	-20
Perfecta		63		184	1899	1820	79

次は切片と変数の係数を示します。

Intercept	<*>	ñ	è	á	τ	Std res.
1554.853	- .112	1.475	.572	.936	- .457	70.948

略字<*>と接続詞の τ の係数がマイナスなので、年代の推移と逆相関していることがわかります。一方、スペイン語特有文字のエニェ ñ や、アクセント符号がついた母音文字は年代の推移と相関しています。しかし、標準残差が 70 なので、これらの文字の出現による予測はかなり困難です。

■スペイン語の ñ のバリアント

先に名義尺度の数量化をしたデータ（Y:文書の発行年代・P:文書の発行地・T:文書の類別; N:ñ のバリアント（目的変数）：→得点）で重回帰分析をすると、プログラムは次の結果（下左表）を出力しました。

M.Coef.	Value
Y	.255
P	.442
T	-.116

Correl.	Y	P	T	N
Y	1.000	.548	.435	.553
P	.548	1.000	.356	.434
T	.435	.356	1.000	.241

Intercept	578.629
Res.Ratio	.031

N	.553	.434	.241	1.000

上左表を見ると発行年代(Y)の係数(.255)が発行地(P)の係数(.442)より低くなっています。一方、相関係数（上右表）を見ると、発行年代(Y)が目的変数と一番大きく相関していますから(.553)、先の係数は納得できません。これは、発行年代(Y)のと発行地(P)の相関が高いため(.548)、多重共線性の問題が起きたためだと考えられます。そこで、主成分重回帰分析を実行しました。その結果が次の表です。

PCAe	#1	#2	#3
E.value	1.898	.662	.440

PCAv	#1	#2	#3
Y	.614	-.213	-.760
P	.585	-.523	.620
T	.530	.825	.196

PCA.Coef.	Value
#1	25.977
#2	-15.125
#3	-16.200
Intercept	1374.154
Res.Ratio	.031

固有値(E.value)を見ると第 1 主成分(#1)の働きがとくに強いことがわかります(1.898)。それに対応する固有ベクトルを見ると、発行年代(Y)、発行地(P)、文書類別(T)という重要性の順を示しています。次の表は、それぞれの主成分と目的変数(N)の相関係数を示します。

Correl.	#1	#2	#3	N
#1	1.000	.000	.000	.523
#2	.000	1.000	.000	-.180
#3	.000	.000	1.000	-.157
N	.523	-.180	-.157	1.000

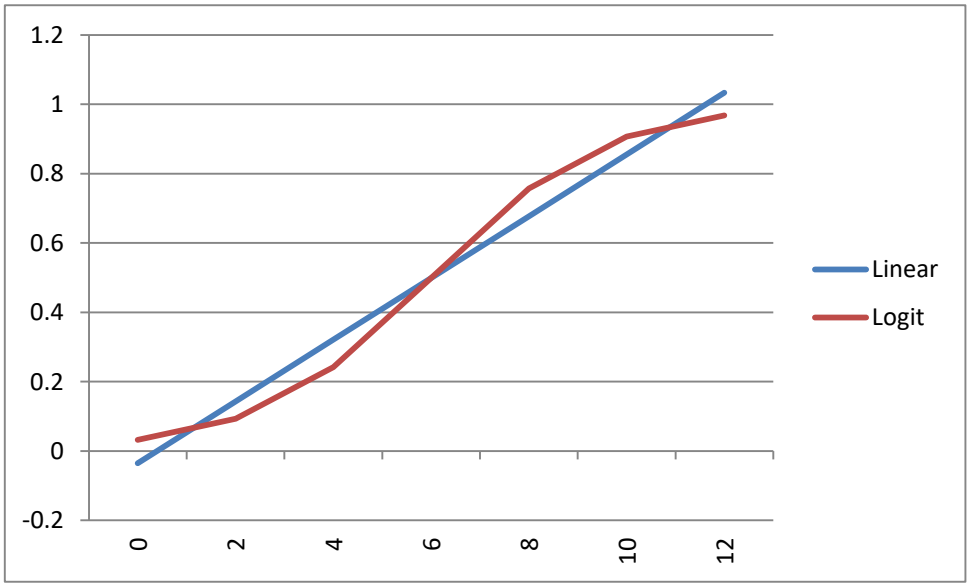
これは、第 1 主成分(#1)と N が強く相関し、他の主成分は N とほとんど相関していないことを示しています。あらためて、発行年代(Y)、発行地(P)、文書類別(T)が、それぞれ、.614, .585, .530 の割合で構成されている総合的な第 1 主成分(#1)の重要性が認められます。この割合は、先に見た重回帰分析による係数値と大きく異なります。また、N の実測値と、回帰式で予想される N 値の間の残差率(Res.Ratio)が少ないことにも注目すべきです(.031)。

8.2. ロジット重回帰分析

次の表は変数 X と、それに対応する確率（または何らかの比率:P）を示します。たとえば、1 週間の学習時間(X)と英語のテストの正解率のようなものを考えます。確率や比率の範囲は[0, 1]です。

L	X	P	Linear	Logit
d1	0	0.04	-0.035	.032
d2	2	0.07	0.143	.093
d3	4	0.24	0.321	.242
d4	6	0.50	0.499	.499
d5	8	0.77	0.677	.757
d6	10	0.90	0.855	.907
d7	12	0.97	1.033	.968

上の Linear は単回帰分析による導出変数です。これをグラフにすると、次図の直線のようになります。ここで、近似があまりよくないことと、X=0 で P がマイナスになり、X=12 で P が 1 を超えていることがわかりますが、これは率の範囲が[0, 1]であるので現実的ではありません。一方、上表の Logit はかなりよく P に近似しています。また、グラフを見ても[0, 1]の範囲を超えることはありません。



次の表と図は確率 P、その関数であるロジット(Logit: L)、そしてロジットから確率 P を導く逆関数 InvLogit を示します。ロジット(L)は

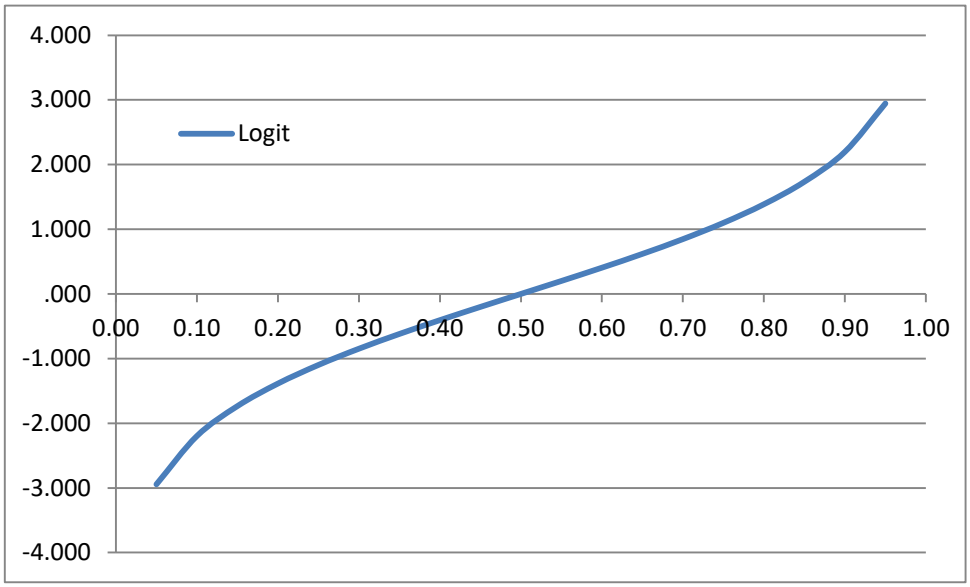
$$L = \mathrm{Ln}\,[P / (1 - P)]$$

つまり、ロジット(L)は、あることが起こる確率 P とそれが起こらない確率 1 − P の比率（P / (1 − P): **オッズ** Odds とよばれます）の自然対数(Ln)を示します。

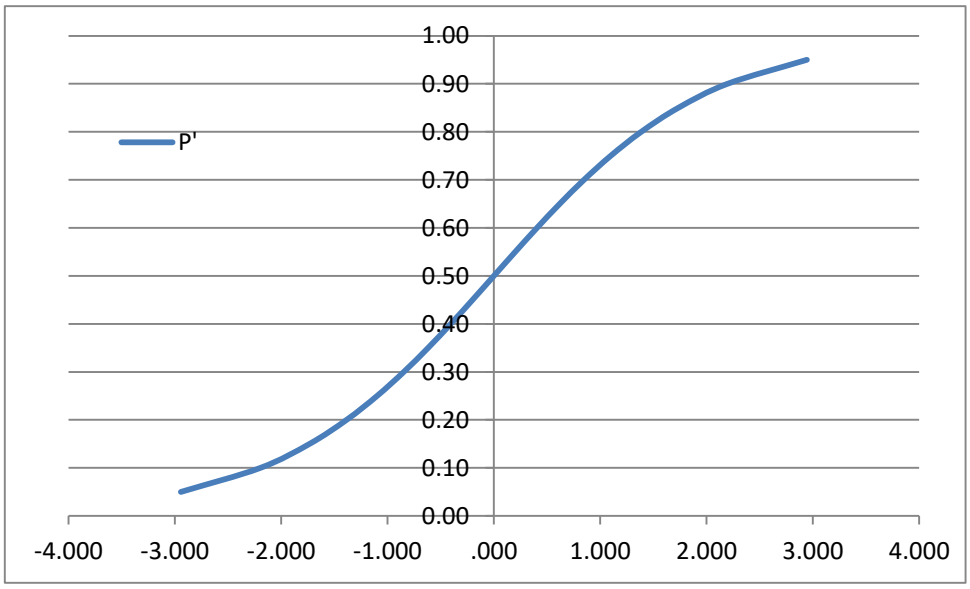
P	Logit	P'
0.05	-2.944	0.05
0.10	-2.197	0.10

0.15	-1.735	0.15
0.20	-1.386	0.20
0.25	-1.099	0.25
0.30	-.847	0.30
0.35	-.619	0.35
0.40	-.405	0.40
0.45	-.201	0.45
0.50	.000	0.50
0.55	.201	0.55
0.60	.405	0.60
0.65	.619	0.65
0.70	.847	0.70
0.75	1.099	0.75
0.80	1.386	0.80
0.85	1.735	0.85
0.90	2.197	0.90
0.95	2.944	0.95

下図は横軸が確率 P であり、それに応じてロジットがどのように変化するかを縦軸で示しています。P の範囲は[0, 1]ですが、ロジットは範囲が自由で P=0 のときに－∞、P=1 のときに＋∞になります。



次の図では横軸がロジット、縦軸が確率です。



上の確率 P はロジット(L)から次のようにして導出します(e: 自然対数の底)。

$$\mathrm{Ln}\,[P / (1 - P)] = L$$
$$P / (1 - P) = e^{L}$$
$$P = (1 - P)\, e^{L} = e^{L} - P\, e^{L}$$
$$P + P\, e^{L} = e^{L}$$
$$(1 + e^{L})\, P = e^{L}$$

よって

$$P = e^{L} / (1 + e^{L})$$
$$= 1 / [1 / (e^{L} + 1)] \quad \leftarrow \text{分子を分母に移動}$$
$$= 1 / (e^{-L} + 1) \quad \leftarrow \text{分母を整理}$$
$$= 1 / (1 + e^{-L}) \quad \leftarrow \text{分母を整理}$$

次の表は確率 P をロジットに変換して重回帰分析をし、その導出変数を確率に戻して出力した結果です。この方法を**ロジット重回帰分析**(Logit Multiple Regression: L.Regres)とよびます。

L.Regres.	P	Derived	Res.
d1	.040	.032	.008
d2	.070	.093	-.023
d3	.240	.242	-.002
d4	.500	.499	.001
d5	.770	.757	.013
d6	.900	.907	-.007
d7	.970	.968	.002

L.Coef.	Value
X	.569
Intercept	.033
Res.Ratio	.016

説明変数が複数の場合は次のモデルで回帰分析をします。

$$L = \mathrm{Ln}\,[P / (1-P)] = W(0) + W(1)\,X(i, 1) + W(2)\,X(i, 2) + \ldots + W(p)\,X(i, p)$$
$$= X_{np}\,W_p$$

上式で求めた Ln [P / (1-P)]から確率(P)を導くためには先の式を使います。

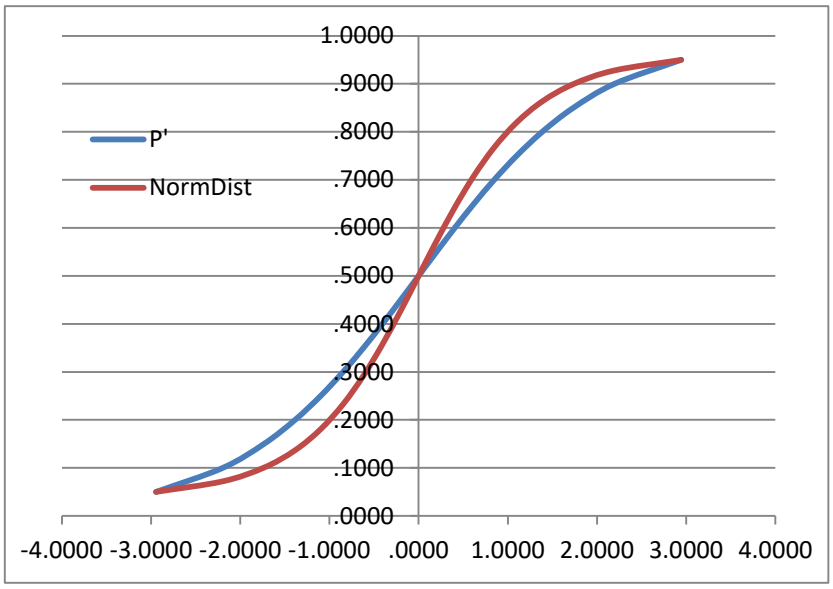
$$P = 1 / (1 + e^{-L})$$

8.3. 正規回帰分析

先の図（ロジットを横軸に、確率を縦軸にしたグラフ）は累積正規確率分布とよく似ています（→「確率」）。次の表と図が示すように、その中心の座標は、ロジットに対応する確率(P)でも、正規分布(NormDist)でも同じですが(0.5)、傾きが少し違っています[6]。

Logit	P'	NormDist
-2.9444	.0500	0.0502
-2.1972	.1000	0.0721
-1.7346	.1500	0.1006
-1.3863	.2000	0.1367
-1.0986	.2500	0.1807
-.8473	.3000	0.2326
-.6190	.3500	0.2919
-.4055	.4000	0.3575
-.2007	.4500	0.4276
.0000	.5000	0.5000
.2007	.5500	0.5724
.4055	.6000	0.6425
.6190	.6500	0.7081
.8473	.7000	0.7674
1.0986	.7500	0.8193
1.3863	.8000	0.8633
1.7346	.8500	0.8994
2.1972	.9000	0.9279
2.9444	.9500	0.9498

[6] NormDist 関数の引数である平均を 0.5 とし、 標準偏差を Logit の範囲で求めました。



ロジット回帰分析の回帰式の目的変数はロジットに対応する確率(P)を使いますが、その確率分布ではデータ（目的変数）の平均と分散（または標準偏差）が考慮されていません。どのようなデータの目的変数でも、すべて同じようにロジットに対応する確率分布をあてはめて一般化しています。

ここで、重回帰式の目的変数（確率）がこの変数の平均と分散によって求められる正規累積分布にしたがう、と見なし、正規累積分布の逆関数 NormInv で変換した数値を使って重さベクトルを算出する方法を**正規回帰分析** (Normal Regression: N.Regres) と名づけて提案します。導出変数(Derived)の計算では、もとの目的変数の平均(m)と分散(v)を使った正規累積分布関数 NormDist(x, m, sqr(v), 1)を適用します。

次の表が先のロジスティック回帰分析と同じデータを使った、正規回帰分析(N.Regres)の結果です。残差(Res)と残差比(Res.Ratio)がさらに小さくなりました。

L	X	P
d1	2	.017
d2	4	.023
d3	6	.037
d4	8	.050
d5	10	.070

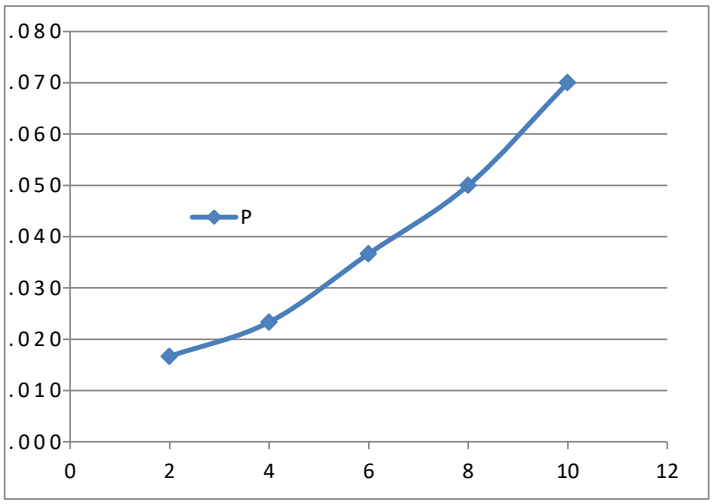
N.Regres.	P	Derived	Res.
1	.017	.016	.000
2	.023	.024	-.001
3	.037	.035	.001
4	.050	.050	.000
5	.070	.070	.000

N.Coef.	Value
X	.002
Intercept	-.005
Res.Ratio	.015

正規回帰分析は目的変数が直線式でなく、むしろ正規分布（の一部分）のような分布をしているときに有効です。そこで、次のように複数の説明変数(X1, X2)があるときはその相関係数行列を計算して、P と相関が高い変数(X1)を使って P の散布図を描きます。

L	X1	X2	P
d1	2	6	.017
d2	4	5	.023
d3	6	5	.037
d4	8	3	.050
d5	10	6	.070

C.Cor	X1	X2	P
X1	1.0000	-.2582	.9853
X2	-.2582	1.0000	-.1272
P	.9853	-.1272	1.0000



上の図を見ると分布は直線になっていないことがわかります。そこで、直線式による重回帰係数は適切でないことになります。次の 2 つの表によって重回帰分析と正規回帰分析の結果を比較すると（残差と残差比）、正規回帰分析のほうがこのデータに適していることが確認できます。

L	X1	X2	P
d1	2	6	.017
d2	4	5	.023
d3	6	5	.037
d4	8	3	.050
d5	10	6	.070

M.Regres.	P	Derived	Res.
d1	.0167	.0141	.0026
d2	.0233	.0255	-.0022
d3	.0367	.0393	-.0027
d4	.0500	.0484	.0016
d5	.0700	.0693	.0007

M.Coef.	Value
X1	.0069
X2	.0024
Intercept	-.0140
Res.Ratio	.0494

L	X1	X2	P
d1	2	6	.017
d2	4	5	.023
d3	6	5	.037
d4	8	3	.050
d5	10	6	.070

N.Regres.	P	Derived	Res.
d1	.0167	.0164	.0002
d2	.0233	.0244	-.0010
d3	.0367	.0355	.0012
d4	.0500	.0499	.0001
d5	.0700	.0705	-.0005

N.Coef.	Value
X1	.0016
X2	.0001
Intercept	-.0050
Res.Ratio	.0155

8.4. 主成分重回帰分析

次のデータ(D)は、英語(E)、ラテン語(L)、数学(M)の成績と、年間に読

んだ小説の冊数(NH)を示す架空のデータ例です。はじめに、重回帰分析をした結果(Mr)を見ましょう。

D	E	L	M	N
d1	58	34	90	3
d2	50	53	100	5
d3	45	48	66	6
d4	58	51	78	7
d5	43	44	32	9
d6	56	59	54	13
d7	77	72	20	28

Mr.	E	L	M	N	N^	Res.
d1	58.000	34.000	90.000	3.000	2.541	-.459
d2	50.000	53.000	100.000	5.000	4.030	-.970
d3	45.000	48.000	66.000	6.000	5.890	-.110
d4	58.000	51.000	78.000	7.000	8.703	1.703
d5	43.000	44.000	32.000	9.000	8.856	-.144
d6	56.000	59.000	54.000	13.000	13.523	.523
d7	77.000	72.000	20.000	28.000	27.457	-.543

Mr.w.	E	L	M	Interc.	R.m.:Prp
Org.	.279	.267	-.135	-10.591	.636
Std.	.373	.375	-.469	.000	.989

上右表(Mr)の N^は重回帰式を適用した予測変数で、Res は目的変数と導出変数の残差(Residual)を示します。その下の表(Mr.w)は、元の行列(Org.)の重みとなる負荷ベクトルと、元の行列を標準化した行列(Std)の負荷ベクトルです。どちらを適用しても同じ結果になりますが、同じ尺度で E, L, M の負荷を比較するときには、Std のほうが適しています。また、Std では切片(Interc.)がゼロになるので、負荷だけを考慮すればよいことになります。

次に、説明変数(E, L, M)と目的変数(N)の相関を見ましょう。

Correl.	E	L	M	N
E	1.000	.643	-.335	.771
L	.643	1.000	-.545	.871
M	-.335	-.545	1.000	-.799
N	.771	.871	-.799	1.000

上の相関行列を見ると、L:N の相関(.871)がとくに高く、それに続いて E:L の相関(.771)が高くなっています。しかし、E:L に相関があるので(.643)があるので、たとえば、E:N の相関に E:L の相関が影響していると考えられます。つまり、E:N の相関は、純粋に E と N の関係を示しているのではなく、そこには L も影響しているはずです。理想的には変数間の相関がゼロになっていれば、純粋に変数の負荷を比較できるのですが、どのような変数であっても、それらの変数の間の相関がゼロになることは、ふつうはありえません。

そこで、主成分得点の相関がゼロになるという特徴を利用すれば（→主成分分析）、相関のない変数（主成分）が示す純粋な重みを知ることができるはずです。はじめに、E, L, M の列だけを用いて主成分分析(PCA)をし

ます。下右表(Pc.Mr.)は標準化した主成分得点、主成分得点を使って行った重回帰分析の結果です[7]。

PcMr	#1	#2	#3	N	N^	Res.
d1	-.461	.263	-.342	3.000	2.541	-.459
d2	-.305	.255	.278	5.000	4.030	-.970
d3	-.274	-.181	.108	6.000	5.890	-.110
d4	-.059	.196	.002	7.000	8.703	1.703
d5	-.170	-.556	-.090	9.000	8.856	-.144
d6	.212	-.050	.123	13.000	13.523	.523
d7	1.057	.073	-.080	28.000	27.457	-.543

このように、導出変数(N^)と残差(Res)は、先の重回帰分析と同じです。しかし、標準化行列(Std)の第 1 主成分(#1)の負荷がとくに大きな数値(.990)を示していることに注目します。

PcMr.w.	#1	#2	#3	Intercept	R.m.:Prp
Org.	16.414	-2.418	-1.753	10.143	.636
Std.	**.990**	**-.084**	**-.041**	.000	.989

そして、次の相関行列(Correl.)で、主成分間の相関がゼロになることを確認した後で、各主成分と目的変数(N)の相関が、上の標準化行列の負荷ベクトルと同じになっていることに注目しましょう。このことは主成分間の相関がないことから、それらが理想的な軸（直角）になることを示しています。

Correl.	#1	#2	#3	N
#1	1.000	.000	.000	.990
#2	.000	1.000	.000	-.084
#3	.000	.000	1.000	-.041
N	**.990**	**-.084**	**-.041**	1.000

次は、各主成分変数の固有値(PcMr.e)と固有ベクトル(PcMr.v)を示します。このように第 1 主成分と第 1 主成分で 90%近く(.899)の累積寄与率を示しているので、もとのデータの分布は 2 次元で近似できることになります。

PcMr.e	#1	#2	#3
E.value	2.026	.672	.303
Ratio	.675	.224	.101
Ac.ratio	.675	.899	1.000

PcMr.v	#1	#2	#3
E	.569	.616	-.545
L	.635	.093	.767
M	-.523	.782	.338

[7] 後で、個体と変数の関係を同一の尺度を使って観察するために、主成分得点を（ゼロでない）正値の固有値の数(3)で割って平均化してあります。

主成分重回帰分析では、先に見た各主成分の負荷（重み：重要度）と、その主成分の構成（固有ベクトル）です。このデータでは言語：数学の軸を示す第 1 主成分が目的変数(N)と .990 という高い相関係数を示し、そしてその相関係数がそのまま標準重回帰式の負荷になります。第 2 主成分と第 3 主成分の負荷量は少ないのですが、第 2 主成分の固有値は全体の 22.4% を占めているので無視できません。

8.5. 名義主成分重回帰分析

言語研究では数値行列を扱うばかりでなく、次の下左表(N)のような文字（名義）行列を扱うことも多いので[8]、文字行列の重回帰分析法を考えます。

N	x1	x2	y
d1	A	C	X
d2	A	D	X
d3	A	D	Y
d4	A	E	X
d5	B	C	X
d6	B	D	Y
d7	B	E	Z

D1	A	B	C	D	E	X	Y	Z
d1	1	0	1	0	0	1	0	0
d2	1	0	0	1	0	1	0	0
d3	1	0	0	1	0	0	1	0
d4	1	0	0	0	1	1	0	0
d5	0	1	1	0	0	1	0	0
d6	0	1	0	1	0	0	1	0
d7	0	1	0	0	1	0	0	1

はじめに、この文字行列(N)を上右表(D1)のような 2 値の数値行列に変換します。たとえば、A があるセルは A 列に 1 を代入し、そればなければ 0 とします。このような変数は「ダミー変数」とよばれています。

たとえ 0 と 1 しかない行列でも、それらが数値であることに変わりはないので重回帰分析が可能です。しかし、これまでの重回帰分析とは異なり、目的変数がダミー変数のバリエーションの数だけ存在することになります。上の例では目的変数(y)には、X, Y, Z のバリエーションがあるので、ダミー変数の列は 3 列になります。

そこで、これまでの重回帰分析の方法に従えば、複数の説明変数＋単数の目的変数のケースを、それぞれの行で 3 回行って、その 3 つの導出変数の中から 1 に最も近い数値（実際には最大値）のあるケースに該当する目的変数（文字）を割り当てます。このときの重回帰分析の方法として、主成分重回帰分析を使います。この方法を**名義主成分重回帰分析**(Nominal Principal Component Multiple Regression:N.Pc.Mr)とよびます。

プログラムははじめに次を出力します。

[8] たとえば、x1 が性別(m, f)、x2 が地域コード（都市名）、y が検索された文字列などのデータです。また、年代の数値であっても適切な間隔で区分して名義化することにより、それぞれの年代の特徴が明らかになります。

N.Pc.Mr.	#1	#2	#3	y	y^	Ac:.857	X	Y	Z
d1	-.188	.514	-.522	X	X	Ok	1.250	-.100	-.150
d2	-.589	-.237	.000	X	Y	Ng	.500	.600	-.100
d3	-.589	-.237	.000	Y	Y	Ok	.500	.600	-.100
d4	-.188	.514	.522	X	X	Ok	.750	-.100	.350
d5	.652	.066	-.522	X	X	Ok	.750	.100	.150
d6	.251	-.686	.000	Y	Y	Ok	.000	.800	.200
d7	.652	.066	.522	Z	Z	Ok	.250	.100	.650

ダミー変数行列の説明変数は 5 列ありますが、A と B は互いに排除しあう関係にあるので、どちらかの値がわかれば、別の値は自動的に決まります。C, D, E の場合は、そのうちの 2 つの値がわかれば、残りの値が決まります。よって、それぞれの自由度は 1, 2 となるので、その和(3)が固有値（＝主成分）の数になります。次の列(y)が目的名義変数、次の y^が導出名義変数です。多くの場合、重回帰式で求めた値は目的変数と一致しますが、完全であることは稀で、ふつう残差が生じます。たとえば d2 は回帰式では Y が求められましたが、データは X なので一致していません。Ac 行の Ok は一致した場合を示し、Ng は一致しなかった場合を示します。Ac:857 は Ok の数をデータ数(N)で割った正答率(Acuracy)です。X, Y, Z の列は、それぞれのケースで導出された数値です。たとえば、d1 では、X(1)を目的変数とすると、1.250 になり、これが最大値なので、導出名義値を X とします。下左表が、それぞれの目的変数に対応する導出値を計算するために使われる負荷ベクトルです。下右表は、切片をなくし、全体の尺度を揃えるために、主成分得点と目的変数を標準化したときの負荷ベクトルを示します。

N.PcMr.w	#1	#2	#3	Intercept
X	-.092	.306	-.189	.571
Y	-.099	-.330	.000	.286
Z	.190	.024	.189	.143

N.PcMr.s	#1	#2	#3
X	-.185	.618	-.382
Y	-.218	-.730	.000
Z	.544	.067	.540

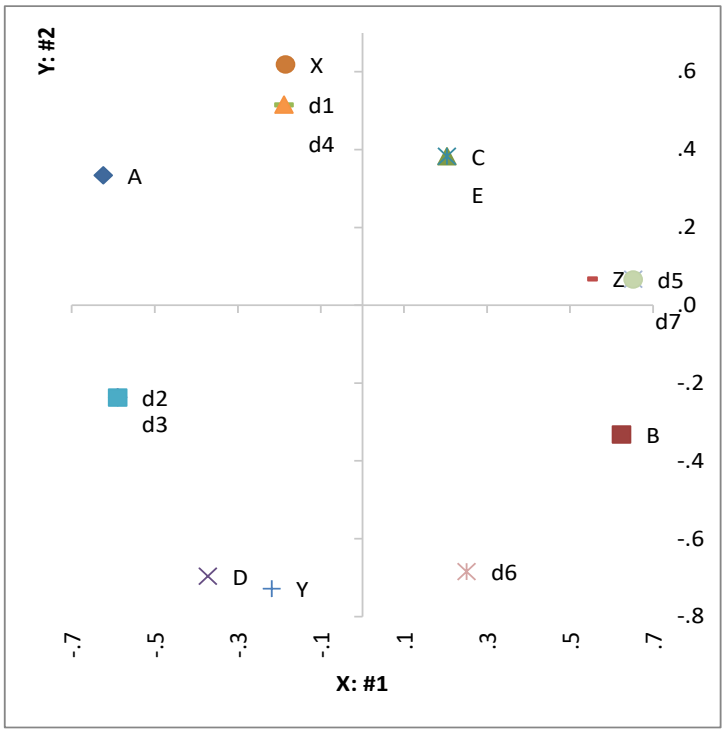
さて、先の不一致(Ng)のケース(d2)を少し追跡してみましょう。実は、d3 にも同じ説明変数(A, D)があり、この場合の目的変数は Y ですから、この回帰分析から導出された値 Y と一致します。つまり、データ全体から見て、説明変数が A, D ならば目的変数は Y になる、ということが予想されたわけです。このことは次の相関行列を見ると、たしかに A:X の相関(.417)よりも D:Y の相関(.730)が高くなっているので納得できます。

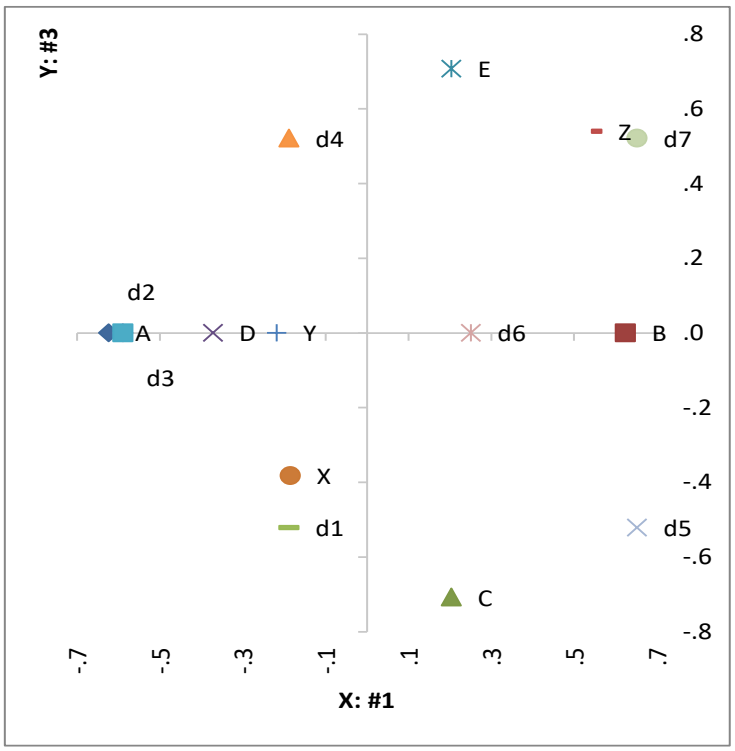
Correl.	A	B	C	D	E	X	Y	Z
A	1.000	-1.000	-.091	.167	-.091	.417	-.091	-.471
B	-1.000	1.000	.091	-.167	.091	-.417	.091	.471
C	-.091	.091	1.000	-.548	-.400	.548	-.400	-.258
D	.167	-.167	-.548	1.000	-.548	-.417	.730	-.354
E	-.091	.091	-.400	-.548	1.000	-.091	-.400	.645
X	.417	-.417	.548	-.417	-.091	1.000	-.730	-.471
Y	-.091	.091	-.400	.730	-.400	-.730	1.000	-.258
Z	-.471	.471	-.258	-.354	.645	-.471	-.258	1.000

次表は、それぞれの主成分の固有値と寄与率、累積寄与率を示します。#2 と#3 に大きな差がないので、#3 までを考慮すべきでしょう。

N.PcMr.e	#1	#2	#3
E.value	2.159	1.441	1.400
Ratio	.432	.288	.280
Ac.ratio	.432	.720	1.000

たとえば、d5, d7, B, Z の集団について、#1:#2 の図では見えない関係が#1:#3 の図でわかるようになります。





8.6. 群別分析

この節では、データ行列の右に 1 列にそれぞれのデータ行と連関する 1 つの実測値が示されている行列(データ行列＋実測値)を入力行列として、はじめにデータ行列と実測値の関係を分析します。続いて、実測値を持たないデータ行列について先の実測値に対応する予測値を求めます。データ行列と実測値には二名義(binominal)、多名義(multinominal)、数値(numerical)の 3 種が考えられます。先の二名義の「判別分析」(Discriminant analysis)と区別して、ここで扱う多名義の分析を「群別分析」(Group analysis)とよびます。

8.6.1. 共起回数による群別

下左表の行列には v1-v3 のデータ列と右端の名義群(Group)があります。D の行列の中から X のそれぞれの行 x1, x2 に近い行を探し、既知の D の群にしたがってを未知の X を群別します。

簡単な方法は、データが一致する回数が多い個体を探し、その群を推定群として未知のデータに与えることです。この方法を「共起回数による群別」(Grouping by cooccurrence)とよぶことにします。下右表(Co.g)は既知群(Grp)と、推定群(Grp.i)を出力し、それが一致したときは評価列(Eval.)に Ok を出力します。この場合既知データの評価はかならず Ok になり、一致度（値）は 1 になります。

D	v1	v2	v3	Grp

Co.g	v1	v2	v3	Grp	Grp.i	Eval.	値

d1	A	D	H	a
d2	A	D	I	b
d3	A	F	H	b
d4	A	E	H	c
d5	B	F	G	c
d6	B	F	I	c

d1	A	D	H	a	d1: a	Ok	1.000
d2	A	D	I	b	d2: b	Ok	1.000
d3	A	F	H	b	d3: b	Ok	1.000
d4	A	E	H	c	d4: c	Ok	1.000
d5	B	F	G	c	d5: c	Ok	1.000
d6	B	F	I	c	d6: c	Ok	1.000

次左表では群が未知のデータであり、右表の Grp.i がその推定群です。x1 では d1 と D が 1 つ一致しているので、値は 1/3 ≒ .333 となります。複数一致するときは最初に一致したデータの群を採用します。

X	v1	v2	v3
x1	B	D	J
x2	B	E	H

Co.g	v1	v2	v3	Grp.i	値
x1	B	D	J	d1: a	.333
x2	B	E	H	d4: c	.667

次に、個体ではなくて群全体の平均と比較する、という方法を考えます。たとえば、X の x1 は D の d1 と v2:D を共起させているので、係数 1/3 となり、これが d1-d6 のそれぞれの係数と比較して最大となるので群別を d1 の a とします。v2:D は d2 とも共起しますが、d2 は 2 成員の群なので、平均は 1 / (2 * 3)になります。

D	v1	v2	v3	Grp
d1	A	D	H	a
d2	A	D	I	b
d3	A	F	H	b
d4	A	E	H	c
d5	B	F	G	c
d6	B	F	I	c

Co.g	v1	v2	v3	Grp	Grp.i	Eval.	値
d1	A	D	H	a	a	Ok	1.000
d2	A	D	I	b	a	No	.667
d3	A	F	H	b	a	No	.667
d4	A	E	H	c	a	No	.667
d5	B	F	G	c	c	Ok	.556
d6	B	F	I	c	c	Ok	.556

X	v1	v2	v3
x1	B	D	J
x2	B	E	H

Co.g	v1	v2	v3	Grp.i	値
x1	B	D	J	a	.333
x2	B	E	H	c	.444

x2 と c 群全体の共起回数は、v1: B*2, v2:E, v3:H の 4 回です。c 群全体全体が 9 個あるので係数は 4/9≒.444 になります。これが他の群と比べたときの最大値です。

8.6.2. 距離による群別

次のような多名義の群別値(a, b, c, …)が既知のデータ D から、ベクトル間の「距離」によって、X のような群別値が未知のデータを分析し、D の

中の個体(d1, d2, ...)または群(a, b, c)に近い横ベクトルを探し、その群別値を X に与えます。

D	v1	v2	v3	Group
d1	5	2	7	a
d2	3	3	2	b
d3	2		2	b
d4	4	2	2	c
d5	2	4	3	c
d6	1	8	7	c

X	v1	v2	v3
x1	4	2	5
x2	3	7	6

たとえば、d1 と x1 の距離(Distance: Dist)を次のように定義します。このように算出される距離は「ユークリッド距離」とよばれます。

$$Dist(d1, x1) = \{\Sigma i\ [D_{np}(1, i) - X_{np}(1, i)]^2\}^{1/2}$$

d1 の成分は(5, 2, 7), x1 の成分は(4, 2, 5)なので、両者間の距離は次のように計算されます。

$$Dist(d1, x1) = [(5 - 4)^2 + (2 - 2)^2 + (7 - 5)^2]^{1/2}$$
$$= (1^2 + 0^2 + 2^2)^{1/2} = 5^{1/2} \fallingdotseq .236$$

このような計算を d2, ..., d6 でも行い、これら 6 つの距離の最小値が得られたときの群別値(a, b, c)を x1 の群とします。x2 についても同様です。その結果、以下のように x1 は d1 と一番近く、x2 は d6 に一番近いという結果になります。

Dt.g	v1	v2	v3	Grp.i	Val.
x1	4	2	5	d1:a	2.236
x2	3	7	6	d6:c	2.449

●群平均値などによる群別

次に D の個々の行ではなく、それぞれの群全体と比較します。そのとき群の代表値としてここでは次のように平均値を使います。

D	v1	v2	v3
a	5.000	2.000	7.000
b	2.500	1.500	2.000
c	2.333	4.667	4.000

上の 3 行と先の X の 2 行の間のそれぞれの距離を比較すると、結果は次のようになります。

Dt.g	v1	v2	v3	Grp.i	Val.
x1	4	2	5	a	2.236
x2	3	7	6	c	3.145

群の代表値として平均値のほかに、データの分散の状態によって中央値、中間値、大数平均値を使うことも考えられます。次は大数平均値を使ったときの結果です。

Dt.g	v1	v2	v3	Grp.i	Val.
x1	4.0	2.0	5.0	a	2.236
x2	3.0	7.0	6.0	c	3.446

●標準化距離による群別

次の v3 のように平均･標準偏差が大きく異なるデータを使うときには注意が必要です。

D2	v1	v2	v3	Group
d1	5	2	56	a
d2	3	3	33	b
d3	2		21	b
d4	4	2	22	c
d5	2	4	45	c
d6	1	8	72	c

X2	v1	v2	v3
x1	4	2	50
x2	3	7	60

上の v3 のような変数が群別に過大に影響することを防ぐために、データ D と X を合体したデータ X_{np} を標準得点に変換します。

$$X_{np} = [I_{np} - MeC(I_{np})] / SdC(I_{np})$$

D2	v1	v2	v3
d1: a	1.633	-.588	.649
d2: b	.000	-.196	-.693
d3: b	-.816	-1.373	-1.393
d4: c	.816	-.588	-1.335
d5: c	-.816	.196	.007
d6: c	-1.633	1.765	1.583

次が群別の結果(群平均)です。

Dt.g	v1	v2	v3	Grp.i	Val.
x1	.816	-.588	.299	a	.888
x2	.000	1.373	.883	c	1.330

●マハラノビス距離による群別

主成分得点を使って、各変数の標準偏差だけでなく変数間の相関もゼロになるように変換し、個体間のマハラノビス距離を計算して群別します。以下がその結果です。

D2	v1	v2	v3
d1: a	-.493	1.811	-.907
d2: b	-.389	-.397	.750
d3: b	-1.009	-1.636	-1.030
d4: c	-1.095	-.031	1.681
d5: c	.341	-.732	-.277
d6: c	2.010	-.485	-.454

Dt.g	v1	v2	v3	Grp.i	Val.
x1	-.399	.866	-.981	a	.953
x2	1.036	.603	1.218	c	1.493

●既知データの再評価

次は、先のデータ(D2)を使って、既知データを標準化距離と群平均を使って再分析した結果です。群は既知ですが、群の平均をとるため、一部推測値が合わないケースが出ました。これは重回帰分析や判別分析と同様に、既知データを分析して得られたパラメータを改めて、既知データにあてはめた結果です。

Dt.g	v1	v2	v3	Disc.	Grp.i	Eval.	Val.
d1	1.633	-.588	.649	a	a	Ok	.000
d2	.000	-.196	-.693	b	b	Ok	.797
d3	-.816	-1.373	-1.393	b	b	Ok	.797
d4	.816	-.588	-1.335	c	b	No	1.274
d5	-.816	.196	.007	c	c	Ok	.385
d6	-1.633	1.765	1.583	c	c	Ok	2.267

Grp.	Pos.	Neg.	AR.
Val.	5.000	1.000	.833

上右表の Pos[sitive]は一致した数(Ok の数)を示し、Neg[ative]は一致しなかった数(No)の数を示します。AR(Acuracy rate)は全体の中の Pos の割合です。この結果から、d4 は、c よりも b に近い、ということがわかります。たしかに、d4 は c の他のメンバーと大きく異なっています。分析の過程では、a, b, c の分類が便宜的な分類であるならば、d4 を改めて b 群においてもよいでしょう。先験的な分類を守る方法を前範疇化とよび、データを再評価して新たな分類を作る方法を後範疇化とよびます。言語研究では前者の方法をとることが多いのですが、柔軟な後者の方法が行われることもありま

す。

8.6.3. 確率による群別

群別値が既知のデータ行列の群内の列相対頻度を、それが該当する事象が起きる確率と見なして、行全体の確率を計算し、これを群別値が未知のデータ（横ベクトル）にあてはめて、一番大きな確率を示すデータの群別値を示す群の群別値を得ます。

次の左表(Q)のような質的データの既知の群別値から、右のような未知の群別値を確率を使って予測します。

Q	v1	v2	v3	Group
d1	v	v		a
d2	v		v	a
d3	v			a
d4		v		a
d5	v		v	b
d6		v	v	b

Y	v1	v2	v3
x1	v	v	
x2			v

次が各群(a, b)の確率表です。それぞれの v1, v2, v3 が群内の列の中で使用された率を示します。a:v1 = .667 は(3 + 1) / (4 + 2)の結果です。分子には 1 を加算し、分母には群数(2)を加算します[9]。

Likel.	v1	v2	v3
a	.667	.500	.333
b	.500	.500	.750

D.data	v1	v2	v3	Disc.	Grp.i	Eval.	mx:mn
d1	v	v		a	a	Ok	.711
d2	v		v	a	b	No	.006
d3	v			a	a	Ok	.711
d4		v		a	a	Ok	.495
d5	v		v	b	b	Ok	.006
d6		v	v	b	b	Ok	.339

Grp.	Pos.	Neg.	AR
Val.	5.000	1.000	.833

D.pred	v1	v2	v3	Grp.i	mx:mn
x1	v	v		a	.711
x2			v	b	.339

[9] ここで確率がゼロのとき積がすべてゼロになるため、すべての生起回数に 1 を加えました。

ここで、たとえば、x1 (v, v, x)の確率は

P(X=a|Y=x1) = (4/6) * (.667) * (500) * (1 － .333)
P(X=b|Y=x1) = (2/6) * (.500) * (.500) * (.750)

P(X=a|Y=x1)の最後の(1 － .250)で確率を逆転させるのは、v3 が選択されていないため、それが起きない場合の確率を示すためです。

D.pred	Ct(mx, mn)	Group
x1	.711	a
x2	.339	b

●ベイズの定理

2つの事象 X と Y が同時に起こる確率 P(X, Y)を次のように計算します。

P(X, Y) = P(X) P(Y|X)
P(X, Y) = P(Y) P(X|Y)

上の最初の式は、同時確率 P(X, Y)が、X が起こる確率 P(X)と、X が起きたとき Y が起こる確率 P(Y|X)の積になる、ということを示しています。たとえば、X がトランプのスペード、Y がエースであるとすると、スペードのエースが出る確率は(1/4) x (1/13) = 1/52 になります。2 番目の式も同様です。そこで、どちらも左辺が同じなので、1 つの式にまとめます。

P(X) P(Y|X) = P(Y) P(X|Y)

よって、次の式（「ベイズの定理」Bayes' theorem）が導かれます。

P(X|Y) = P(X) P(Y|X) / P(Y)

この定理は重要なので簡単な応用例を説明します。次の表は多数の文書からなる資料を A 地方と B 地方の割合 P(X)と、それぞれの地方の資料の中で観察される、ある言語現象（たとえば語末母音の脱落）がそれぞれの文書に起こる割合 P(Y/X)を示しています。たとえば、A 地方の文書は全体の 17 文書の中で 4 文書あり P(X)、その A 地方 4 文書の中で、3 文書で語末母音の脱落があった P(Y|X)、ということを示します。

資料(X)	P(X)	P(Y|X)	P(X) P(Y|X)	P(X) P(Y|X) / P(Y) = P(X|Y)
X=A	4/17	3/4	4/17 x 3/4 = 3/17	(3/17) / (8/17) = 3/8
X=B	13/17	5/13	13/17 x 5/13 = 5/17	(5/17) / (8/17) = 5/8
和	1		8/17 = P(Y)	1

上表の P(X)は P(Y)を考慮しないので「事前確率」(prior probability)とよ

ばれ、P(Y|X)は、それぞれの群内での確率を示すので「尤度(ゆうど)」(likelihood)とよばれます。事前確率と尤度の積 P(X) P(Y|X)は、先に見たように、X と Y の同時確率(joint probability)です。たとえば A の同時確率 3/17 は資料全体の中での A 地方の該当文書（現象のある文書）の割合を示します。B の P(X) P(Y|X) = 5/17 も同様です。この同時確率の計算で、積の第 1 項の分子(4)が第 2 項の分母と同じであることに注意してください。これは、群内で占める該当文書の割合（事前確率）を計算するときの分子が、尤度を計算するときのベース（分母）になる、と考えるとわかりやすいと思います。ここでそれぞれの確率を分数で示し小数やパーセント表示にしなかったのは、それぞれの分母と分子がどのような意味を持っているのかを確認したかったためです。

さて、A と B の尤度の和(3/17 + 5/17=8/17)になりますが、これがベイズの定理の分母 P(Y)にあたります。つまり、全文書数 17 の中で現象(Y)が起きている文書数(8)の確率(8/17)を示します。

最後に、上表の右端の列でベイズの定理にしたがって P(X|Y)を求めます。これは、先に求めた地方(X)と現象(Y)のそれぞれの同時確率 P(X) P(Y|X)を、その和である、文書全体で現象が起こる確率で割った割合を示します。

Y の事象が複数のときは条件付き確率(尤度)を次のように拡張します。

$$P(Y|X) = P(Y_1|X)\ P(Y_2|X)\ \ldots\ P(Y_p|X)$$

＊高村(2000: 99-117), 加藤・羽室・矢田(2008: 111-115)を参照しました。

■アンダルシア方言の前範疇化と後範疇化

アンダルシアの市町村 230 で調査された言語地理資料から 164 の音声特徴を選んでデータ行列を作り、各県を郡として質的確率による群別分析にかけました。行政区画と言語特徴による群別がどの程度一致するかを見ることが目的です。次の表は音声特徴と地点の一部を抜粋したものです。Huelva 県(H)のほとんどが H に群別されていますが、中には Cádiz 県(Ca)や Sevilla 県(Se)に郡別されている地点もあります。しかし、Ca と Se は H に隣接します。

154	155	156	157	158	159	160	161	162	163	164	Grp.d	Grp.i	Eval.
v											H	H	Ok
v											H	H	Ok
v									v	v	H	H	Ok
v									v	v	H	H	Ok
v											H	H	Ok
	v		v						v	v	H	H	Ok
v			v								H	H	Ok
v											H	H	Ok

v			H	H	Ok
			H	H	Ok
v			H	H	Ok
	v		H	H	Ok
			H	H	Ok
		v	H	H	Ok
			H	H	Ok
			H	H	Ok
v			H	Ca	No
v			H	Ca	No
v			H	Ca	No
v			H	Ca	No
	v		H	Ca	No
v			H	H	Ok
	v		H	Se	No
	v		H	Se	No

全体を見ると、正しい県に群別された地点は 173, 異なる県に郡別された地点は 57 であり確率による正答率は 75%でした。

Grp.	Pos.	Neg.	Pres.
Val.	173	57	.752

個体との距離による群別以外の方法であれば、どの方法を使っても、完全に一致することはありません。個体との距離による群別は既知のデータどうしの同定ということなので、正しくは「群別」ではありません。未知のデータの群別では、既知データの中に近いデータがあるか否かがポイントになるので、全体を見渡した言語分析ができません。

上の表に戻って、2 つの県名群(H, Ca, Se)の列を見ると、左の列は行政区画による「前範疇化」を示します。後の列は群別分析による「後範疇化」を示します。前範疇化によっておおよその区分ができたら、言語特徴から後範疇化をし、言語とは直接関係のない行政区画とは別の言語区画を探求する方法を考えてみたいと思います。

8.6.4. 判別分析

次のようなデータの説明変数行列に、適当な重みベクトルを与え、最終列の質的目的変数(v)を予想する方法は**判別分析**(Discriminant Analysis)とよばれます。そのとき求められた重みベクトルは、それぞれの変数の重みの意味を探るのに役立ちます。また、その重みベクトルを使えば未知のデータの目的変数を一定の精度で予想することもできます。

English	Read	Write	Vocab.	POINT
d1	6	8	5	
d2	7	10	6	
d3	8	4	8	v
d4	9	7	2	
d5	10	9	4	v

この例では d1, d2, ..., d5 という生徒の小テスト(x1:Read, x2:Write, x3:Vocab[ulary])の得点と、最終試験の評価(v:優)が示されているとします。

はじめに説明変数行列を次のように標準化します。

[1] $X_{np} = \mathrm{Std}(X_{np})$ ...標準化: (X – 列平均) / 列標準偏差

Std.s.	Read	Write	Vocab.
d1	-1.414	.194	.000
d2	-.707	1.166	.500
d3	.000	-1.748	1.500
d4	.707	-.291	-1.500
d5	1.414	.680	-.500

この X_{np} に未知の重みベクトル W_p を右積した合成ベクトルを Z_n とします。

[2] $Z_n = X_{np} W_p$

重みベクトル W_p が求められれば、上の式で Z_n が求められます。Z_n の平均 M は次のようにゼロ(0)になります。

$M = (\Sigma_{(i:N)} Z_n) / N$ ← 平均の定義

$= \Sigma_{(i:N)} (X_{np} W_p) / N$ ← [2]

$= \Sigma_{(i:N)} (X_{i1}W_1 + X_{i2}W_2 + \ldots + X_{ip}W_p) / N$ ← 行列積の成分

$= (\Sigma_{(i:N)}X_{i1}W_1 + \Sigma_{(i:N)}X_{i2}W_2 + \ldots + \Sigma_{(i:N)}X_{ip}W_p) / N$ ← Σを分配

$= (W_1 \Sigma_{(i:N)}X_{i1} + W_2 \Sigma_{(i:N)}X_{i2} + \ldots + Wp \Sigma_{(i:N)}X_{ip}) / N$ ←定数を前に

ここで X_{np} は標準化されているので、それぞれの縦和は 0 です。

$$\Sigma_{(i:N)}X_{i1} = \Sigma_{(i:N)}X_{i2} = \ldots = \Sigma_{(i:N)}X_{ip} = 0$$

よって、Zn の分子の項がすべて 0 になるので、Zn の平均 M は

[3] $M = 0$

Z_n の全変動 S は

$$S = \Sigma_{(i:N)} (Z_i - M)^2 \quad \leftarrow \text{変動の定義}$$
$$= \Sigma_{(i:N)} Z_i^2 \quad \leftarrow [3]\ M = 0$$

合成ベクトル Z_n 全体を、優をとった学生群 Zv と、そうでない学生群 Zc に分けて考え、それぞれの群の個数（人数）を Nv, Nc、群内の平均を Mv, Mc とします。

Zv 内の変動と Zc 内の変動の和は「群内変動」(within-groups sum of squares: Sw)とよばれます。

$$Sw = \Sigma_{(i:Nv)}(Zv_i - Mv)^2 + \Sigma_{(i:Nc)}(Zc_i - Mc)^2$$

M は 0 ですが、Mv と Mc は 0 になるとは限りません。なぜならば列全体を標準化しているので M=0 になるのですが、Zv, Zc はそれぞれの群内で標準化しているわけではないからです。

それぞれの群の成分がすべて同じだと仮定して、それと全体の平均 M (= 0)との編差の 2 乗和は「群間変動」(between-groups sum of squares: Sb)とよばれます。群間変動はそれぞれの群が全体（平均は M=0）の中でどのように変動するかを示します。群間変動は次のような式になります。

$$Sb = \Sigma_{(i:Nv)}(Mv - M)^2 + \Sigma_{(i:Nc)}(Mc - M)^2$$
$$= \Sigma_{(i:Nv)}Mv^2 + \Sigma_{(i:Nc)}Mc^2 \quad \leftarrow [3]\ M = 0$$
$$[4] \quad = Nv\ Mv^2 + Nc\ Mc^2 \quad \leftarrow \text{定数の倍数}$$

このとき、全変動(S)が群内変動と群間変動の和 S = Sw + Sb であることが、次のようにして確かめられます。

$$Sw = \Sigma_{(i:Nv)}(Zv_i - Mv)^2 + \Sigma_{(i:Nc)}(Zc_i - Mc)^2$$
$$= \Sigma_{(i:Nv)}(Zv_i^2 - 2\,Zv_i\,Mv + Mv^2) \quad \leftarrow \text{展開}$$
$$+ \Sigma_{(i:Nc))}(Zc_i^2 - 2\,Zc_i\,Mc + Mc^2)$$

$$= \Sigma_{(i:Nv)}Zv_i^2 - \Sigma_{(i:Nv)}2Zv_i\,Mv + \Sigma_{(i:Nv)}Mv^2 \quad \leftarrow \Sigma\text{を分配}$$
$$+ \Sigma_{(i:Nc))}Zc_i^2 - \Sigma_{(i:Nv)}2Zc_i\,Mc + \Sigma_{(i:Nv)}Mc^2$$

$$= \Sigma_{(i:Nv)}Zv_i^2 - 2\,Mv\,\Sigma_{(i:Nv)}Zv_i + Nv\,Mv^2 \quad \leftarrow \text{定数を前に}$$
$$+ \Sigma_{(i:Nv)}Zc_i^2 - 2\,Mc\,\Sigma_{(i:Nv)}Zc_i + Nc\,Mc^2$$

$$= \Sigma_{(i:Nv)}Zv_i^2 - 2\,Mv\,Nv\,Mv + Nv\,Mv^2 \quad \leftarrow \Sigma_{(i:Nv)}Zv_i = Nv\,Mv$$
$$+ \Sigma_{(i:Nv)}Zc_i^2 - 2\,Mc\,Nc\,Mc + Nc\,Mc^2 \leftarrow \Sigma_{(i:Nv)}Zv_i = Nv\,Mc$$
$$\leftarrow \text{和＝個数*平均}$$

$$= \Sigma_{(i:Nv)}Zv_i^2 - 2\,Nv\,Mv^2 + Nv\,Mv^2 \quad \leftarrow Mv\text{を合体}$$
$$+ \Sigma_{(i:Nv)}Zc_i^2 - 2\,Nc\,Mc^2 + Nc\,Mc^2 \quad \leftarrow Mc\text{を合体}$$

$$[5] \quad = \Sigma_{(i:Nv)}Zv_i^2 - Nv\,Mv^2 + \Sigma_{(i:Nv)}Zc_i^2 - Nc\,Mc^2 \quad \leftarrow -2^* + ^* = -^*$$

よって

$$Sw + Sb = \Sigma_{(i:Nv)} Zv_i^2 - Nv\,Mv^2 + \Sigma_{(i:Nc)} Zc_i^2 - Nc\,Mc^2 \quad \leftarrow [5]\ Sw$$
$$+ Nv\,Mv^2 + Nc\,Mc^2 \quad \leftarrow [4]\ Sb$$

$$= \Sigma_{(i:Nv)} Zv_i^2 + \Sigma_{(i:Nc)} Zc_i^2 = St$$

次に、群間変動(Sb)が全変動(S)の中で占める割合を問題にします。この割合は「相関比」(correlation ratio: CR)とよばれます。

$$CR = \text{群間変動(Sb)} / \text{全変動(S)}$$
$$[6] \quad = \text{群間変動(Sb)} / (\text{群内変動(Sw)} + \text{群間変動(Sb)})$$

たとえば、各群のすべての成分が群内の平均と等しいときは（Zv = Mv, Zc = Mc のとき）、群内変動(Sw)はゼロになり、すべての成分が 1 点に集中し群を完全に判別でき、上の式[6]から相関比(CR)は最大の 1 になります。また、それぞれの群内の平均（Zv の平均と Zc の平均）が全体の平均と同じときは(Mv = M, Mc = M), 群間変動 Sw はゼロになるので（群を判別できないので）、相関比(CR)は最小のゼロ(0)になります。

相関比 CR の分母の Z_n の全変動 S を W_p を含む式にします。

$$S = Z_n^T Z_n$$
$$= (X_{np} W_p)^T (X_{np} W_p) \quad \leftarrow [2]\ Z_n = X_{np} W_p$$
$$= W_p^T X_{np}^T X_{np} W_p \quad \leftarrow \text{行列演算}$$

ここで

$$[7] \quad S_{pp} = X_{np}^T X_{np}$$

とすると

$$[8] \quad S = W_p^T S_{pp} W_p, \quad \leftarrow [1]$$

相関比 CR の分子の Z_n の群間変動 Sb を W_p を含む式にします。

$$[9] \quad Sb = Nv\,Mv^2 + Nc\,Mc^2 \quad \leftarrow [4]$$
$$= Nv\,(Sv_p^T/Nv\,W_p)^2 \quad \leftarrow Sv_p: X_{np} \text{ の v 群縦和ベクトル}$$
$$+ Nc\,(Sc_p^T/Nc\,W_p)^2 \quad \leftarrow Sc_p: X_{np} \text{ の c 群縦和ベクトル}$$

$$= Nv\,(Sv_p^T W_p)^2 / Nv^2 \quad \leftarrow Nv \text{ はスカラー}$$
$$+ Nc\,(Scp^T W_p)^2 / Nc^2 \quad \leftarrow Nc \text{ はスカラー}$$

$$= (Sv_p^T W_p)^2 / Nv \quad \leftarrow Nv \text{ はスカラー}$$
$$+ (Scp^T W_p)^2 / Nc \quad \leftarrow Nc \text{ はスカラー}$$

$= (Sv_p{}^T W_p)^T (Sv_p{}^T W_p) / Nv$ ← 行列演算
$+ (Scp^T W_p)^T (Scp^T W_p) / Nc$ ← 行列演算

$= W_p{}^T Sv_p Sv_p{}^T W_p / Nv$ ← 行列演算
$+ W_p{}^T Scp Scp^T W_p / Nc$ ← 行列演算

$= W_p{}^T (Sv_p Sv_p{}^T / Nv + Sc_p Sc_p{}^T / Nc) W_p$
$= W_p{}^T Bpp W_p$ ← Bpp は以下の式

[9b] $Bpp = Sv_p Sv_p{}^T / Nv + Sc_p Sc_p{}^T / Nc$

よって、先の[6]の相関比 CR = Sb / S は次のようになります。

[10] $CR = Sb / S = W_p{}^T B_{pp} W_p / W_p{}^T S_{pp} W_p$

$S_{pp} = X_{np}{}^T X_{np}$ ←[7]
$B_{pp} = Sv_p Sv_p{}^T / Nv + Sc_p Sc_p{}^T / Nc$ ←[9]

この相関比[10] CR が最大になるときのベクトル W_p を求めるのが判別分析の目的です。つまり、もっとも良く 2 群を判別するときの W_p を探すことになります。そこで、相関比の式を未知数の W_p で微分しますが、このような分数の微分については、分母(S: $Wp^T S_{pp} Wp$←[8])を 1 とする条件をつけて、相関比が最大化する値を求めます。そこで、ラグランジュの未定乗数 L と、全変動 ST = 1 → ST－1 = 0 という条件をつけた関数 F (Wp)を考えます。

$F (W_p) = Sb – L (S – 1)$ ←ラグランジュの未定乗数法
$= W_p{}^T B_{pp} W_p － L (W_p{}^T S_{pp} W_p － 1)$ ←[8], [9]

この F を W_p で微分した式がゼロ(0)であるときの W_p を求めます。

$Diff.(F, W_p) = 2 B_{pp} W_p － 2 L S_{pp} W_p = 0$ ←行列の微分

よって

[11] $(B_{pp} － L S_{pp}) W_p = 0$

$S_{pp}^{-1} (B_{pp} － L S_{pp}) W_p = S_{pp}^{-1} 0$ ← S_{pp}^{-1} を左積
$(S_{pp}^{-1} B_{pp} － S_{pp}^{-1} L S_{pp}) W_p = 0$ ← S_{pp}^{-1} をそれぞれの項に
$(S_{pp}^{-1} B_{pp} － L S_{pp}^{-1} S_{pp}) W_p = 0$ ← スカラーL を移動
$(S_{pp}^{-1} B_{pp} － L I_{pp}) W_p = 0$ ← $S_{pp}^{-1} S_{pp} = I_{pp}$ (単位行列)
$S_{pp}^{-1} B_{pp} W_p － L I_{pp} W_p = 0$ ← W_p をそれぞれの項に
$S_{pp}^{-1} B_{pp} W_p － L W_p = 0$ ← $I_{pp} W_p = W_p$

となり、これが固有値問題の形($R_{pp} A_p － L A_p = 0$)になります。ここで、S_{pp}^{-1}

B_{pp}から固有値 L と固有ベクトル W_pを求めることができます。

また、先の式[11]から、次のようにして固有値が相関比であることがわかります。

$(B_{pp} - L S_{pp}) W_p = 0$ ← [11]
$W_p^T (B_{pp} - L S_{pp}) W_p = W_p^T 0$ ←両辺に W_p^Tを左積
$W_p^T B_{pp} W_p - W_p^T L S_{pp} W_p = 0$ ←展開
$W_p^T B_{pp} W_p - L W_p^T S_{pp} W_p = 0$ ← スカラーL を移動
$Sb - L S = 0$ ← $W_p^T B_{pp} W_p = Sb$, $W_p^T S_{pp} W_p = S$
$Sb = L S$ ← L ST を右辺に移動
$L = Sb / S$ ← Sb / ST = 相関比

相関比は分母も分子も変動を使い 2 次関数になるので、その根をとったほうがわかりやすく、それが使われることもあります。ここではそれを「根相関比」(Root Correlation Ratio: RCR)とよぶことにします。

根相関比 $RCR = (Sb / St)^{1/2}$

Std.s.	Read	Write	Vocab.	POINT	Expect.	Score	Eval.
d1	-1.414	.194	.000			-1.090	Ok
d2	-.707	1.166	.500			-.297	Ok
d3	.000	-1.748	1.500	v	v	1.088	Ok
d4	.707	-.291	-1.500			-.408	Ok
d5	1.414	.680	-.500	v	v	.707	Ok

上表（標準得点 Standard score: Std. s.）の期待値(Expect[ed value])の列では Zn の成分が正であれば v を出力します。実測値（ここでは POINT）と期待値が一致したときに評価列(Eval[uation])に Ok を出力します。

得点列(Score)は[2]の合成ベクトル Z_nです。

次の変数表(Var[iable])の重み(Weight)は求められた固有ベクトル W_pであり、その下にそれぞれの変数の和(Sum)、平均(M.)、標準偏差(St[andard] dev[iation])を出力します。

Var.	Read	Write	Vocab.
Weight	.761	-.070	.644
Sum	40.000	38.000	25.000
M.	8.000	7.600	5.000
DT	1.414	2.059	2.000

T. eval.	Ac. R.	R.C.R.
Value	1.000	.927

最後の表、総合評価(T[otal] eval[uation])には正答率(Ac[curacy] R[atio])と根相関比(Root Correlation Ratio: RCR)を出力します。正答率は上の評価の Ok の数を行数で割った値です。

＊三野(157-161)、石井(2014: 140-149)を参照しました。

●未知の判別値

既知のデータ(X_{np})で得られた重みベクトル(W_p)を、判別値が未知のデータ(D_{np})に適用するときは、先に得られた平均 $M(X_{np})$と標準偏差 $Sd(X_{np})$を使って、判別値が未知のデータ行列を標準化し、これに重みベクトルを左積します。

$$Y_{np} = [D_{np} - Me(X_{np})] / Sd(X_{np})$$
$$E_n = Y_{np} W_p$$

●数量化２類分析

次のような説明変数が質的データの場合は、チェック(v)を 1 に変換して数量化し同じ判別分析をします。この方法は「数量化２類分析」とよばれます。

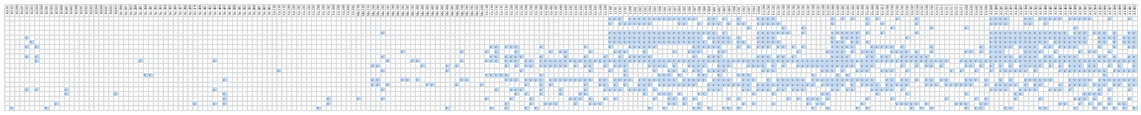
English-5	Read	Write	Vocab.	POINT
d1		v		
d2	v	v	v	
d3	v		v	v
d4	v	v		
d5	v	v		v

■東西アンダルシア方言の判別

次の表はアンダルシア地方を西(H, SE, CA, MA)と東(CO, J, GR, AL)に分ける音声特徴の相対的な頻度を示します。両側相対値の大小順にソートしました。

Ñ1000	H	SE	CA	MA	CO	J	GR	AL	West	East	Cntr
1602B:disgusto:-sg->x						710	217	600	0	1527	1.000
1660B:unos granos:s=g>x						581	174	233	0	988	1.000
1694C:clavel:-él>ér						32	109	33	0	174	1.000
1663B:las juergas:s=xwe>xwe				38		774	348	833	38	1955	.961
1577A:naranja:-nx->nx	42					839	196	900	42	1934	.958
1647A:las lentejas:-x->x	42					871	217	833	42	1922	.958
1624A:decir:-ír>í+l	83				280	516	413	633	83	1843	.913
1626C:tos:o++				77	280	323	391	400	77	1394	.895
1623A:beber:-ér>é+l	83			38	400	355	413	667	122	1835	.875
1627C:nuez:e++				192	560	581	565	600	192	2306	.846
1632D:los árboles:s=a>a			59	77	80	387	370	767	136	1603	.844
1581C:carne:-rn->ln		65			240	355	65	33	65	693	.830
1695B:claveles:e-es>-e+-e+		32		269	720	774	717	700	301	2912	.812
1631C:los ojos:s=o>o				269	240	742	783	667	269	2431	.801
1620A:mar:-ár>ál	125	32		38	320	452	370	500	196	1641	.787
1616A:árbol:-ol>o+	83	32			240	258	130	200	116	828	.755
1614A:peregil:-íl>í+(:)		32	59	77	320	258	239	267	168	1084	.732
1613A:zagal:-ál>á+(:)	42	97	59		360	194	348	267	197	1168	.711

これらの県別の音声特徴を使って、下図の原資料（個人の回答の集計）の判別分析をすると、下表の結果になりました。



Var.	Weight	Sum	M.	St.dev.
1613A:zagal:-ál>á+(:)	.147	44.000	.191	.393
1631C:los ojos:s=o>o	.137	92.000	.400	.490
1614A:peregil:-íl>í+(:)	.120	39.000	.170	.375
1635A:las vacas:s=b>ph	.107	41.000	.178	.383
1623A:beber:-ér>é+l	.102	63.000	.274	.446
1581C:carne:-rn->ln	.100	23.000	.100	.300
1620E:mar:-ár>ár	.090	36.000	.157	.363
1627C:nuez:e++	.073	81.000	.352	.478
1620A:mar:-ár>ál	.072	59.000	.257	.437
1632D:los árboles:s=a>a	.072	57.000	.248	.432
1695B:claveles:e-es>-e+-e+	.061	104.000	.452	.498
1602B:disgusto:-sg->x	.040	50.000	.217	.412
1616A:árbol:-ol>o+	.034	29.000	.126	.332

1663B:las juergas:>xwe	.031	66.000	.287	.452
1693A:redes:redes>rede	.028	49.000	.213	.409
1624A:decir:-ír>í+l	.018	63.000	.274	.446
1626C:tos:o++	.015	49.000	.213	.409
1660B:unos granos:s=g>x	.011	33.000	.143	.351
1694C:clavel:-él>ér	.002	7.000	.030	.172
1647A:las lentejas:-x->x	-.003	63.000	.274	.446
1577A:naranja:-nx->nx	-.030	63.000	.274	.446

T. eval.	Ac. R.	R.C.R.
値	.943	.910

とくに末尾子音の脱落による開母音化現象が東方言の特徴であることがわかりますが、さらに<j>の強い摩擦音[x]と語末の-s と続く語頭の g が融合した[x] (unos granos > unoxranos)もその顕著な特徴です。

＊資料：『アンダルシア言語民俗地図』(Manuel Alvar y Antonio Llorente: *Atlas lingüístico y etnográfico de Andalucía*, 1973)

●分散分析

次のようなデータから変数（M1, M2, M3）間の分散の差の有意性を調べるときに分散分析(Analysis of Variance: Anova)が使われます。

Xnp	M1	M-2	M-3
A	44	34	33
B	39	29	32
C	42	33	35
D	45	36	32
E	48	30	31

ANOVA	Variation	D.f.	Variance	F.ratio	P. 5%:1%:
Among g,	410.800	2	205.400	28.137	3.885
Within g.	87.600	12	7.300		6.927
All	498.400	14	35.600		0

この分析のために変数間の変動（Sb: 群間の偏差平方和）と、各変数の中での変動（Sw: 郡内の偏差平方和）、そして全体の変動(S: 全体の偏差平方和)を求めます。目的は群間の偏差平方和と郡内の偏差平方和の比(「分散比」）を計算し、それが有意であるかどうかを判定することです。

はじめに列（群）の縦平均行(Mp)と全体の平均（M）を求めます。個数を N, 変数を P とします。

$$Mp = I_p^T X_{np} / N$$
$$M = \Sigma (X_{np}) / (N * P)$$

次にそれぞれの変動 Sb, Sw, S を求めます。

$$Sb = N \sum (Mp - M)^2$$
$$Sw = \sum (X_{np} - Mp_i)^2$$
$$S = \sum (Xnp - M)^2$$

全体の自由度(Degree of freedom: D.frd.)はすべての成分数-1 で計算されます(N * P - 1)。1 を引くのは、総和と 1 つの成分を除く全成分が決定されていれば、その成分は自動的に決まるので自由がないからです。同様に群間の自由度は P - 1 になります。郡内の自由度は同様にして求めた各群の自由度(N - 1)に群の数(P)を掛けた値です。それぞれの分散(Variance)は変動を自由度で割って求めます。フィッシャー比率(Fisher ratio: F. ratio)は群間の分散を郡内の分散で割った値です。

全体変動：$V = S / (N * P - 1)$
群間変動：$Vb = Sb / (P - 1)$
群内変動：$Vw = Sw / [(N - 1) * P]$
$F.ratio = Vb / Vw$

このフィッシャー比率があらかじめ決めた基準(5%, 1%)を超えていれば、群間の分散に差がない、という帰無仮説を棄却できます。上図の最後の列は、F の基準値(5%, 1%)と、確率を示します。