

## 正準相関分析についての解説

### 0. 判別分析 (discriminant analysis)

多変量のデータを用い、重みづけた説明変数（独立変数）を合成して、個々人の所属する集団を分ける基準変数（従属変数）を予測（判別）する多変量解析法を、判別分析と総称する。

例 1：ある患者に対する多種類の検査結果を総合して（説明変数）、どのような病気かを診断する（基準変数）

例 2：発掘された石器や土器などを、材質や形態、地層などの諸要因を分析して（説明変数）、どの時代に帰属できるか（基準変数）を決定する

すなわち、判別分析は、分類の基準を与えてくれる説明変数の構成を目的にしている  
また、判別分析は、正準相関分析の特殊な場合と位置づけられている（表 1）

表 1. 多変量解析における論理展開

対象とする変数		解析法		数学モデル
X	Y			
単変数	単変数	相関分析	⇔ 単回帰分析	ベクトルの内積
変数群	単変数	重相関分析	⇔ 重回帰分析	1組の線型モデル
変数群	変数群	正準相関分析	⇔ 判別分析	2組の線型モデル

2変数群を分析するには、表 2 のような技法を選択できる

表 2. 2変数群を分析する際に用いる技法

分析技法	説明
正準相関分析	<p>a. 2つの正準変数が最大の相関を示すよう変数群を要約</p> <p>b. 相関が最大となるよう正準変数が合成されるものの、個々の変数群に対する最適な説明（説明変数の分散の最大化）になるとは限らない</p> <p>c. 正準相関の意義を解釈→個々の変数群が含む成分（正準変数の内容）の解釈</p> <p>d. 正準相関の解釈は、抽象的で困難、明快な解釈は困難</p>
主成分分析（または因子分析）	<p>a. 2変数群（または多変数群）を1群化し、主成分（因子）を探索する</p> <p>b. 主成分（因子）に変数間の影響関係と構成関係が混在するので解釈は難しい</p>

主成分分析＋重回帰分析	<p>a.2 変数群に対して、以下のような手順を実行(多変数群)では、2 変数群を組み合わせで実行)</p> <p>①説明変数群の主成分分析</p> <p>②基準変数群の主成分分析</p> <p>③説明変数群における 1 個の主成分得点に対して、基準へ数群における主成分得点のすべてを用いて重回帰分析を反復</p> <p>b.技法としては、単純明快であるが、各変数群が独自の構造を仮定している場合には妥当性が問題となる</p>
-------------	---

## 1. 正準相関分析 (canonical correlation analysis)

多変量データにおいて、多数の変数が 2 つの変数群を構成するとき、変数群間の相互関係を分析するために用いられる。

重回帰分析、判別分析、共分散分析、数量化 I, II, III類を、その特殊な場合として含む。

重回帰分析……外的基準変数が 1 個、説明変数が複数個存在；説明変数の重み付き合計と基準変数との間の相関が最大になるよう、解を求める

例：基準変数  $Y$  = 入学後の成績、説明変数  $X_1$  = 内申書得点、 $X_2$  = 入試得点

正準相関分析…… $Y$  にあたる変数が複数個、 $X$  にあたる変数も複数個存在する

変数群  $\{X_1, X_2\}$  と、もう 1 つの変数群  $\{Y_1, Y_2\}$  とが与えられている

各変数群内で重みつき合計を考えると、 $U_1 = a_1X_1 + a_2X_2$  と  $V_1 = b_1Y_1 + b_2Y_2$  という 2 つの重みつき合計  $U_1, V_1$  が求められる

しかし、このような重みつき合計  $U_1, V_1$  の値は、重み  $a_1, a_2$  と  $b_1, b_2$  の値の選び方によって千変万化する

そこで、 $U_1, V_1$  の相関  $r_{U_1V_1}$  が最大となるようにそれぞれの重みの値を決めようとする方法が、正準相関分析

その際の  $U_1$  と  $V_1$  との間の相関係数は、正準相関係数 (canonical correlation coefficient) と呼ぶ

したがって、 $Y$  にあたる変数群が 1 つの変数しかもたないときには、正準相関係数は、重相関係数に一致する (ただし、正準相関分析では、片方が外的な基準変数群で、他方が説明変数群であるというような区別は、一般には行わない←2 つの変数群は、対等に扱う)

## 2. 正準相関分析の手続きと要点

1) 正準相関分析は、2つの変数群を処理する

①図1では、以下のように変数群を模式化する

- ・変数群 X = {X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>}
- ・変数群 Y = {Y<sub>1</sub>, Y<sub>2</sub>}

②変数群の変数は、定量データで構成されていなければならない。測定単位が異なる定量データは、あらかじめ標準化しておく

2) 変数群の変数を合成した値を正準変数と呼ぶ

①相関係数が最大になるように、2つの正準変数をそれぞれ合成する

②相関係数が最大化された正準変数間の相関係数を正準相関係数と呼ぶ

③正準相関は、変数の個数が少ない変数群の個数だけ求められる

- ・図1の例では、変数群 X が 3 変数で、変数群 Y が 2 変数であるので、正準相関は第 2 まで算出される

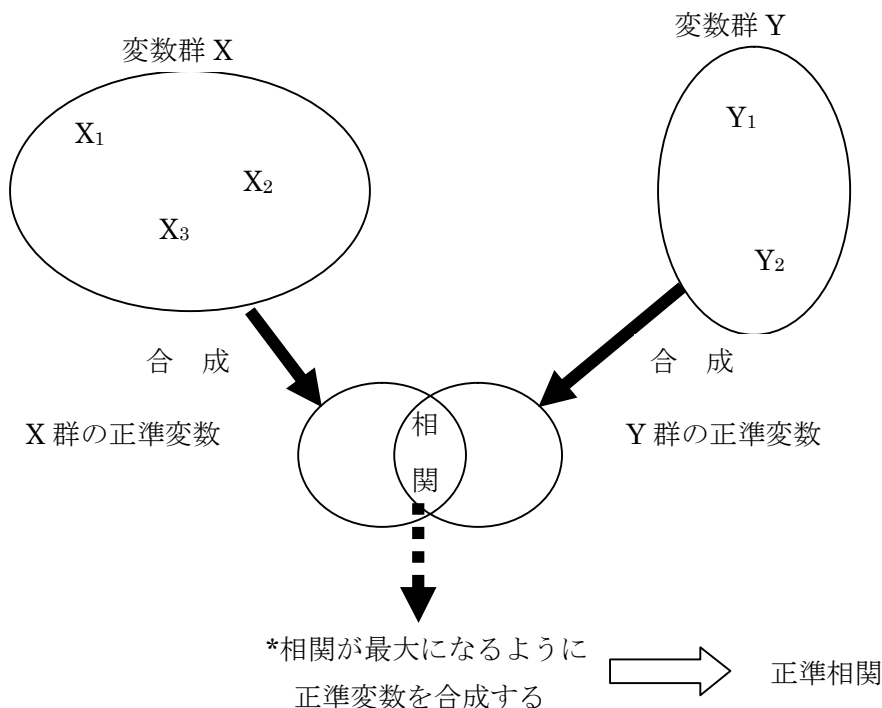


図1. 正準相関分析の手続きと要点

2. 正準相関分析の実際

ここでは、渡部（1988）にあげられている例をもとに正準相関分析の実際について確認する。

1) 用いるデータ

表 4-1 に示された、体格及び運動能力に関するデータを用いる。その相関行列は、表 4-4 にある。

## 2) 主成分分析

正準相関分析との違いを検討するために、Y と X とを込みにした主成分分析を実施する。相関行列から主成分分析を行った結果、固有値 > 1.0 の成分が 2 つ得られたので、その 2 成分について考察する。表 4-5 によれば、この 2 成分の累積寄与率は、約 76% となっている。図 4-1 には、2 成分の負荷量によってプロットしてある (varimax 回転を実施)。

図 4-1 から明らかなように、第 1 主成分は身長、体重、胸囲のように形態学的な大きさに関係する変動を示し、第 2 主成分は、運動能力と関係の大きい変動を示している。懸垂は第 1 主成分に負の負荷を示しており、体の小さい者の方がこの種目では多少有利であることを示していると思われる。ハンドボールや、走り幅跳びは、逆に体の大きい者に有利とみることができる。50m 走と 1500m 走の第 2 主成分の負荷量が負となっているが、これらの種目は成績が、所要時間で測定されるため、値が小さいほどよい成績となっていることによる。

## 3) 正準相関分析

次に、同じデータについて、正準相関分析を実施してみる。

この例では、Y 変数の数が少なく、4 であるので、最大 4 個の正準相関が得られることになる (表 4-6)。これらすべての正準相関が有意かどうかを検定するために、次の式による  $\chi^2$  統計量  $\chi_0^2$  を計算する

$$\chi_0^2 = - \left[ N - 1 - \frac{1}{2}(p + q + 1) \right] \sum_{i=1}^r \ln(1 - \lambda_i^2)$$

N (被験者数) = 46, p (第 1 群の変数の数) = 4, q (第 2 群の変数の数) = 5,  
第 1 固有値  $\lambda_1^2$  から第 4 固有値  $\lambda_4^2$  までを代入すると、 $\chi_0^2 = 52.86$  となる

この検定の自由度は、pq であるから、 $\chi_0^2$  は自由度 (df) = 20 で、99.9 パーセンタイルを越えており、少なくとも 1 つの正準相関係数は、統計的に有意である

次に、2 つめ以降の正準相関の有意性を検定するために

$$\chi_1^2 = \chi_0^2 + 40 \ln(1 - v_1^2)$$

を求めると、 $\chi_1^2 = 14.01$  となる。この値は、表 4-6 に示されたコンピューターからの出力とは若干異なるものの、まるめ誤差の範囲内である。

この  $\chi_1^2 = 14.01$  は、自由度 = (p - 1)(q - 1) = 12 で、70 パーセンタイルをようやく越

表 4-1 ある中学校における2年男子の体格と運動能力

No.	身長 (cm)	体重 (kg)	胸囲 (cm)	座高 (cm)	50m走 (sec)	走幅跳 (cm)	ハンドボール 投げ (m)	懸垂 (回)	1500m走 (min)
1	160.0	49.5	77.0	84.7	7.7	401	26	5	5.47
2	159.7	49.0	74.0	86.6	8.5	331	24	0	6.05
3	141.9	31.0	67.9	73.9	8.1	373	19	5	6.17
4	153.3	46.5	76.8	79.7	8.9	340	26	0	6.27
5	159.5	51.0	82.0	83.1	7.6	403	26	7	5.38
6	167.7	70.0	93.0	87.5	9.1	308	26	0	6.46
7	162.2	66.5	88.5	90.1	7.3	440	31	1	6.13
8	156.6	52.5	79.2	83.1	8.9	321	25	0	6.26
9	165.3	48.0	76.5	87.0	8.4	372	23	1	6.30
10	139.9	31.5	67.4	76.9	8.2	353	20	12	6.30
11	149.5	40.5	70.1	79.8	8.1	373	19	8	6.09
12	168.9	53.0	78.4	89.1	7.8	379	27	6	6.26
13	164.5	49.0	76.8	90.9	7.8	377	23	3	5.58
14	150.4	37.5	71.0	76.9	10.1	280	15	0	7.50
15	146.8	36.5	70.8	78.1	7.9	390	23	4	6.11
16	163.3	51.5	77.3	89.1	7.3	453	22	6	5.38
17	145.1	77.8	74.5	79.4	8.3	349	22	6	6.04
18	157.0	60.6	88.2	85.3	8.3	363	23	1	6.43
19	169.3	46.0	76.0	90.4	8.5	314	13	0	7.03
20	145.1	38.5	73.0	78.3	8.2	349	19	4	5.57
21	158.7	43.0	70.4	83.9	8.2	323	16	1	5.57
22	168.3	49.5	79.0	86.8	7.7	401	30	6	5.39
23	160.4	48.0	77.0	84.8	7.6	375	18	1	6.15
24	156.9	47.5	78.5	84.1	7.7	399	31	3	6.02
25	154.0	49.0	79.0	81.2	8.2	301	19	1	5.52
26	153.5	44.5	75.5	82.3	7.9	398	19	4	5.39
27	157.4	50.0	80.5	82.0	8.6	327	23	0	6.12
28	160.8	47.0	77.7	86.3	7.8	410	29	2	6.01
29	155.5	47.0	76.5	83.6	7.9	382	20	3	5.46
30	149.1	38.5	68.5	80.2	8.9	347	20	6	5.56
31	143.1	33.0	70.0	78.1	10.4	294	13	0	7.06
32	150.4	36.2	69.3	80.7	8.4	335	20	9	6.31
33	160.0	46.5	75.3	82.9	7.6	403	22	4	5.26
34	149.7	33.5	66.2	78.3	9.5	320	16	0	6.38
35	166.8	55.0	82.0	87.3	8.2	368	28	2	5.47
36	150.5	38.5	71.5	81.1	7.8	394	23	6	5.32
37	146.6	34.4	66.5	79.2	7.7	360	23	7	6.13
38	162.2	48.0	75.0	87.6	7.8	394	21	2	6.25
39	169.9	64.5	90.0	90.9	7.6	385	28	2	5.37
40	174.3	61.5	81.5	91.6	7.1	449	30	5	5.54
41	174.4	54.0	77.5	90.9	7.6	381	24	0	6.46
42	156.7	42.5	75.2	82.3	8.8	322	22	0	6.04
43	165.7	52.5	78.9	88.1	7.4	391	25	7	5.46
44	169.0	51.5	78.0	89.1	7.6	400	25	4	6.13
45	161.5	47.2	73.8	85.1	7.3	460	29	11	5.30
46	162.9	51.0	80.0	85.7	7.9	416	29	1	5.55

表 4-4 相 関 行 列

	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>
Y <sub>1</sub> (身長)		.60	.67	.94	-.45	.39	.51	-.25	-.19
Y <sub>2</sub> (体重)	.60		.83	.64	-.32	.26	.54	-.21	-.17
Y <sub>3</sub> (胸囲)	.67	.83		.67	-.27	.24	.59	-.34	-.16
Y <sub>4</sub> (座高)	.94	.64	.67		-.51	.45	.50	-.19	-.21
X <sub>1</sub> (50m走)	-.45	-.32	-.27	-.51		-.85	-.57	-.45	.66
X <sub>2</sub> (走幅跳)	.39	.26	.24	.45	-.85		.64	.45	-.60
X <sub>3</sub> (ハンドボール投げ)	.51	.54	.59	.50	-.57	.64		.15	-.43
X <sub>4</sub> (懸垂)	-.25	-.21	-.34	-.19	-.45	.45	.15		-.39
X <sub>5</sub> (1500m走)	-.19	-.17	-.16	-.21	.66	-.60	-.43	-.39	

表 4-5 主成分分析：回転後の成分負荷量(直交回転)バリマックス法

変数名	成分 1	成分 2
Y <sub>1</sub> 身長	.87	.18
Y <sub>2</sub> 体重	.84	.10
Y <sub>3</sub> 胸囲	.90	.03
Y <sub>4</sub> 座高	.87	.24
X <sub>1</sub> 50m 走	-.30	-.87
X <sub>2</sub> 走幅跳	.26	.87
X <sub>3</sub> ハンドボール投げ	.58	.56
X <sub>4</sub> 懸垂	-.45	.73
X <sub>5</sub> 1500m 走	-.09	-.78
成分負荷量 2 乗和	3.74	3.09
寄与率 (%)	41.60	34.33
累積寄与率 (%)	41.60	75.93

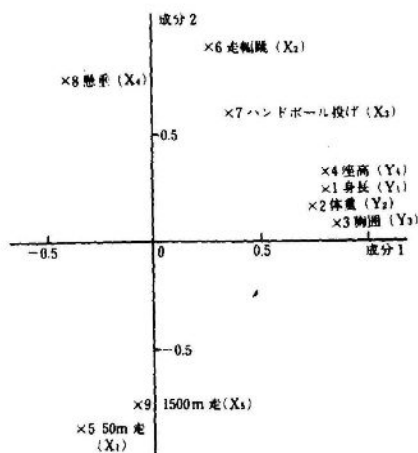


図 4-1 成分負荷量の散布図

表 4-6 正準相関係数とカイ 2 乗(χ<sup>2</sup>)統計量

成分(j)	固有値(λ <sub>j</sub> <sup>2</sup> )	正準相関係数(λ <sub>j</sub> )	カイ 2 乗値	自由度	上側確率
1	0.62	0.79	52.68	20	0.00
2	0.27	0.52	14.27	12	0.28
3	0.03	0.18	1.49	6	0.96
4	0.00	0.04	0.14	2	0.93

えるレベルであるので、帰無仮説は棄却できず、2 つめ以降の正準相関係数はすべて、統計的に有意ではないとみなせる。

そこで、第 1 正準相関を与える第 1 正準変数について、詳しく見ることにする。第 1 正準変数を与える重みベクトルの値を実線の枠で囲ったものを表 4-7 に示す。Y 変数の重みベクトルは、この正準変数 U<sub>1</sub> が身長と胸囲に大きな重みを持って作られていることがわかる。一方、X 変数の重みベクトルからは、50m 走、ハンドボール投げ、懸垂に大きな重み

表 4-7 正準相関分析：重み係数

	成分 1	成分 2	成分 3	成分 4
[変数群 X]	(a <sub>j</sub> )			
X <sub>1</sub> 50m 走	-0.60	0.89	0.77	1.52
X <sub>2</sub> 走幅跳	-0.16	-0.67	1.20	1.55
X <sub>3</sub> ハンドボール投げ	0.67	0.82	-0.70	0.30
X <sub>4</sub> 懸垂	-0.68	0.05	-0.89	0.33
X <sub>5</sub> 1500m 走	0.07	-0.43	-0.17	0.06
[変数群 Y]	(b <sub>j</sub> )			
Y <sub>1</sub> 身長	0.43	1.07	0.35	-2.79
Y <sub>2</sub> 体重	0.03	0.13	-1.76	-0.51
Y <sub>3</sub> 胸囲	0.59	0.90	1.18	1.06
Y <sub>4</sub> 座高	0.05	-2.18	0.00	2.19

が与えられていることがわかる。重みベクトルは、正準変数を作るための基本的なベクトルであるが、値の大小のバラツキが大きく、解釈が難しい。そこで、この性質をより明らかにするため、正準変数と各変数の間の相関を検討してみる。

例えば、Y 変数から得られた正準変数と、Y 変数の 1 つ Y<sub>j</sub> との相関係数を r<sub>cY<sub>j</sub>(l)</sub> であらわすと、

$$r_{cY_j(l)} = \frac{\sum_{i=1}^N u_{li} y_{ij} / (N-1)}{\sqrt{\sum_{i=1}^N u_{li}^2 / (N-1)} \sqrt{\sum_{i=1}^N y_{ij}^2 / (N-1)}}$$

となる。統計パッケージソフトでは、表 4-8 のような構造係数ベクトル（構造ベクトル）として出力される。表 4-8 では、第 1 正準変量の構造ベクトルを枠で囲ってある。

U<sub>1</sub> の構造係数は、いずれの Y 変数とも高い相関を示しており、第 1 正準変数 U<sub>1</sub> は、体の形態的大きさに関する変動を代表する変数と考えられる。

表 4-8 正準相関分析：構造係数

	成分 1	成分 2	成分 3	成分 4
[変数群 X]	r <sub>cx</sub> <sup>(1)</sup>			
X <sub>1</sub> 50m 走	-0.49	0.68	0.43	-0.07
X <sub>2</sub> 走幅跳	0.43	-0.62	-0.19	0.56
X <sub>3</sub> ハンドボール投げ	0.77	0.09	-0.43	0.45
X <sub>4</sub> 懸垂	-0.41	-0.35	-0.73	0.36
X <sub>5</sub> 1500m 走	-0.25	0.18	0.27	-0.12
[変数群 Y]	r <sub>cy</sub> <sup>(1)</sup>			
Y <sub>1</sub> 身長	0.89	-0.31	0.09	-0.33
Y <sub>2</sub> 体重	0.81	0.11	-0.57	0.10
Y <sub>3</sub> 胸囲	0.93	0.26	-0.04	0.24
Y <sub>4</sub> 座高	0.87	-0.50	-0.01	-0.06

一方、第1正準変数  $V_1$  は、50m 走、懸垂、1500m 走に負の相関を、走り幅跳び、ハンドボール投げに正の相関を持っている。50m 走と1500m 走、主成分分析のときと同様に、符号を反転して考えた方がよい。 $V_1$  は、運動能力の変数から作られる合成変数のうちで、体の形態の大きさを代表する変数  $U_1$  と相関が最大となるよう作られているので、 $V_1$  と相関の高い変数は、 $U_1$  と相関の高いことが期待される。したがって、運動能力のうちでは、ハンドボール投げ、50m 走、走り幅跳びの順で体の大きさと相関が高く、逆に懸垂は体の大きさと負の相関を示す変数であると考えられる。

#### 4)両者の比較

以上から、この例題データの場合、正準相関分析の結果と、 $Y$ 、 $X$  を込みにした主成分分析の結果とはかなり整合性が高いと言える。正準相関分析に適していると考えられるデータがあるときには、両者ともに試みて、それらの結果を細かく比較検討することが望ましい。

### 3. 変数の標準化

正準相関分析を実際に用いる場合、もとのデータは、さまざまなものが用いられ、その単位も異なっている。例えば、表 4-1 のように、身長、体重、胸囲、座高のような生徒の形態的側面を記述する変数群と、50m 走、走り幅跳び、ハンドボール投げ、懸垂、1500m 走は運動能力を記述する変数群である。すべての変数が同じ単位で測定されていれば、分散や共分散の大小が意味をもつが、この例のように変数がまちまちの単位で測定されているときには、分散の大小を比較しても意味がない。

また、正準相関分析では、重みつき合計点（合成変数）を求めるが、この合成変数についても、異なった単位で測定した数の重みつき合計点を持つ意味は判断できない。

そこで、正準相関分析では、測定値をいったん単位と関係のない量に変換して、分析を行う。その変換は、例えば次式で与えられるような標準得点に変換する（標準化）をおこなう。標準得点に変換した場合、平均=0、分散=1 となる。

$$\text{標準得点} = \frac{\text{測定値} - \text{平均値}}{\text{標準偏差}}$$

### 4. 判別分析と正準相関分析との関連

#### 1)正準相関分析と判別分析

基本的な関連性は、0. 判別分析 (discriminant analysis) で述べたとおりである。

正準相関分析から見ると、判別分析は、

変数群  $Y$  が、ダミー変数の場合



といえる。

## 2)判別分析

判別分析では、一方の変数群を説明変数  $X$  として設定して、他方の変数群を基準変数  $Y$  として設定する。そして、基準変数をダミーの変数によって群化する。

①変数群  $X$  では、それぞれの変数を重みづけて合成した正準変数を説明変数として構成する

②変数群  $Y$  では、例えば 3 基準を変数化したいときには、ダミーの 2 変数で定義することで、正準相関分析の範囲内で扱うことができる

・変数群で基準を定義する例

■3 基準の変数化→2 つのダミー変数

A 大学= [1, 0]    B 大学= [0, 1]    C 大学= [0, 0]

■4 基準の変数化→3 つのダミー変数

A 大学= [1, 0, 0]    B 大学= [0, 1, 0]    C 大学= [0, 0, 1]

D 大学= [0, 0, 0]

3 グループ以上の基準変数に適用する場合を、**正準判別分析**と呼ぶ

## 3) 判別分析の利用

観測対象にグループがあり、それぞれがどのグループに属しているのかがわかっている。また、同時にそれぞれの複数の特徴が連続量として観測されているときに、どのグループに属しているのかをこれらの特徴から予測する基準を作る場合に利用する。

この基準を作成すると、グループのわからない対象にであったときに、その基準によってどのグループに所属するかが予測できることになる。また、それぞれの特徴をどの程度重視すればよいのかも、この基準でわかることになる。

## 4) 判別分析しようにあたっての手順と留意点

### ①データの用意

データとしては、グループを示す基準変数、特徴を示す複数の説明変数を用意する。説明変数は、量的変数で、統計的にも内容的にも互いに独立であることが望ましい。

データ構造は、被験者×変数のプロフィールデータとし、基準変数と説明変数を横並びに整理しておく。

### ②説明変数の選択

既存の統計パッケージソフトを利用する場合にも変数選択の判断が求められる。

重要なことは、1 つひとつの変数の性質を、あらかじめ吟味しておくことである

統計的な側面……平均値、標準偏差から分布の偏りの有無

内容的な側面……変数の意味を簡潔な表現で整理し、判別に対する重要度の見通しを持つ

変数選択の一般的指針

- a. グループの予測に有用
- b. 解釈しやすい
- c. 他の説明変数との相関が小さい
- d. 信頼性が高い

### ③判別分析のあてはまり

必要な変数がそろったら、判別分析を実施し、データに対するあてはまりを評価する。

一般には、分析全体のウィルクスの $\lambda$ 値 ( $0 \leq \lambda \leq 1$  の値をとり、0に近いほど集団差異が大きい)を確認し、エフ値に変換することで ( $[1-\lambda]/\lambda$ に定数をかけたものがエフ値)、「説明変数群が判別に貢献しない」という帰無仮説を確率的に評価する。

また、ウィルクスの $\lambda$ 値は、変数の出し入れの変化から、各説明変数の貢献度の指標に変換できる。この場合、「説明変数が判別に貢献しない」帰無仮説を評価する。

いずれも有意水準以下の p 値で評価すればよい。

判別関数それぞれがどの程度判別に役立っているかを次に評価する。判別関数は、説明変数の数か、基準変数に示されるグループ数から 1 を引いた数かのいずれか小さい数だけ計算される。

この場合、次の 2 点に注目する：

- a. 統計的には、固有値や $\chi^2$ 検定に注目する  
固有値を見る……寄与率を順次加算した累積寄与率を並べ、値の増加が鈍る手前の判別関数までを利用する  
 $\chi^2$ 検定を見る……判別関数上で「すべてのグループ平均値が等しい」という帰無仮説を評価するもので、有意水準以下の p 値で有効とする
- b. 内容面では、意味の納得しやすさを重視する

### ④判別関数の計算と判別

利用する判別関数を定めたら、説明変数にあてる判別係数（重み係数）、標準判別係数、構造係数等の指標を計算する。

判別係数……説明変数の値をそのまま活かして判別関数値を求める重み；新たに出会う対象を予測するとき利用；判別の基準となる重要な指標

標準判別係数……説明変数の平均=0、標準偏差=1に変換したときの重み；判別係数と同じ用法に加え、説明変数の相対的貢献度を示す

構造係数……判別関数と説明変数との相関；判別関数の意味づけでもっとも重視

対象の判別……判別関数値を用いて行う；全対象について判別関数値を座標値にして判別空間に位置づける；実際のグループで判別関数平均値（重心値）を求め、判別空間に位置づける

判別空間……判別関数で構成した空間

判別の制度の評価……予測と実際とグループ一致率（百分率）で示す=的中率（正診率）  
 $100 - \text{的中率} = \text{誤判断率（誤診率）}$

なお、基準を求めたデータ（これをトレーニング・データと呼ぶ）の的中率は、過大評価に傾き、それを一般化するとさまざまなリスクをとともなうことになる。これには、交差妥当化で対処する

交差妥当化……同じ基準を使って別データの的中率を求める手順で、トレーニング・データによる「推定された（見かけの）的中率」、別データによる「実際の的中率」と区別する

## 5) 判別分析の結果の読み取り

### ①判別分析のあてはまり

データに対して判別分析を選択したことの適否について、ウィルクスの $\lambda$ 値を報告する

$\lambda = 0.0$  で完全に判別できる， $\lambda = 1.0$  で判別能力がないことを意味する

選択した変数群全体の適切さを反映しているとも解釈できる

的中率……80%を越えると、まずまずの判別とみなされる

ただし、4) ④で見たように、「推定された的中率」は過大評価される傾向があるので、「実際の的中率」もあわせて報告する

### ②判別関数の有効性

複数の判別関数が計算される正準判別分析では、各判別関数の有効性を報告する

判別関数のうち、いくつかを選んで報告した場合も同様

有効性の指標

固有値とバートレットの $\chi^2$ 値を報告する

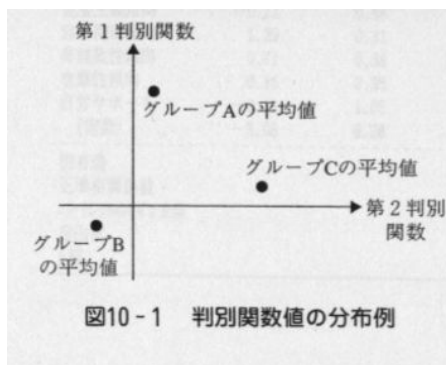
固有値……対応する判別関数の値によるグループ平均値間距離

大きな固有値に対応する判別関数ほど、相対的にグループ間の違いを明瞭に反映していると解釈される

バートレットの $\chi^2$ 検定で有意……判別関数がいずれかのグループの違いを見いだすのに役立っていると解釈する

### ③判別関数の解釈

解釈には、説明変数との関連から考えるものと、判別空間内の対象の分布から考えるものと2つの経路がある



説明変数との関連での解釈……説明変数と判別関数との相関である構造係数に着目  
たとえば、構造係数の絶対値が、0.4以上の説明変数に注目し、それらの意味の重なりから解釈する

正の構造変数……そのままの意味、負の構造変数……逆転した意味づけ

判別空間内の対象の分布による解釈……空間の軸である判別関数の役割を考慮した解釈

判別関数のグループ平均値を判別空間に示すことで、どのグループを見分けるのに有効な判別関数（軸）か、軸の方向でどのグループが分けられているか、視察により検討する

図10-1のように3グループの判別関数の平均値を示すと、第1判別関数はグループAとBとを、第2判別関数はグループBとCとを見分けるのに有効な判別関数（軸）と解釈する

## 5. SPSS での判別分析

### 1) 判別分析の実行

[分析 (A)] メニュー → [分類 (Y)] → [判別分析 (H)] を選択

- ・ [グループ化変数 (G) :] に基準変数を指定する
- ・ **範囲の定義 (D)** をクリックし、[最小 (I) :], [最大 (A) :] を指定する
- ・ **続行** をクリックする
- ・ [独立変数 (I) :] に、説明変数を指定する
- ・ **分類 (C)** をクリックし、[表示] の [交差妥当化 (V)] にチェックを入れる
- ・ **OK** をクリック

### 2) 出力の見方 (小塩, 2004)

固有値, 正準相関係数, ウィルクス (Wilks) の  $\lambda$ , 標準化された正準判別関数, グループ重心の関数, 交差妥当化の結果が算出される.

それぞれの出力内容の意味は, 本資料を参照するか, 小塩 (2004) を参照する.

## 6. 文献

- 1) 深谷澄男・喜田安哲 (2003) : SPSS とデータ分析 2. 北樹出版.
- 2) 小塩真司 (2004) : SPSS と Amos による心理・調査データ解析. 東京図書.
- 3) 渡部洋 (編著) (1988) : 心理・教育のための多変量解析法入門—基礎編—. 福村出版.
- 4) 渡部洋 (編著) (2002) : 心理統計の技法. 福村出版.