

## データ解析

<http://coconut.sys.eng.shizuoka.ac.jp/data/>

静岡大学工学部  
安藤和敏

2005.10.05

## 本講義の内容

本講義では、**多変量解析**について学ぶ。  
多変量解析は多変量データを解析するためのいろいろな手法の寄せ集めである。

それらの手法のうちで主なものに、

- 回帰分析法
- 主成分分析法
- 因子分析法
- 判別分析法

がある。本講義においては、これらの手法について学ぶ予定である。

## 多変量データの例(1)

あるコンサルタント会社の社員10人についてのデータ

社員No	社交性	勤勉性	企画力	判断力	給与評価
1	7	6	7	8	10
2	4	5	5	4	4
3	6	8	4	4	8
4	5	5	5	5	8
5	6	6	4	5	6
6	6	5	6	6	7
7	4	4	6	6	8
8	4	6	6	6	8
9	4	5	5	6	8
10	6	6	4	4	9

## 多変量データの例(2)

浜松駅周辺の中古マンションのデータ

マンションNo	広さ(平米)	築年数(年)	価格(千円)
1	51	16	3.0
2	38	4	3.2
3	57	16	3.3
4	51	11	3.9
5	53	4	4.4
6	77	22	4.5
7	63	5	4.5
8	69	5	5.4
9	72	2	5.4
10	73	1	6.0

## 多変量データ

個体	x	y	z	w	u
1	x1	y1	z1	w1	u1
2	x2	y2	z2	w2	u2
3	x3	y3	z3	w3	u3
4	x4	y4	z4	w4	u4
5	x5	y5	z5	w5	u5
6	x6	y6	z6	w6	u6
7	x7	y7	z7	w7	u7
8	x8	y8	z8	w8	u8
9	x9	y9	z9	w9	u9
10	x10	y10	z10	w10	u10

いくつかの変数をもつデータの集まり

## 重回帰分析

浜松駅周辺の中古マンションのデータ

マンションNo	広さ(平米)	築年数(年)	価格(千円)
1	51	16	3.0
2	38	4	3.2
3	57	16	3.3
4	51	11	3.9
5	53	4	4.4
6	77	22	4.5
7	63	5	4.5
8	69	5	5.4
9	72	2	5.4
10	73	1	6.0

## 重回帰分析によってわかること

1. 価格は、広さと築年数によってどのように予測できるか。
2. 予測できるとすれば、その精度はどれくらいか。
3. 同じ地区で広さ70m<sup>2</sup>、築年数10年、価格5.8千万円のマンションを提示された。この価格は妥当か。

1. 価格と広さと築年数は以下の関係にあると推定される。

$$\text{価格} = 1.02 + 0.0668 \times \text{広さ} - 0.0808 \times \text{築年数}$$

2. 寄与率は 0.933 で上式の精度は十分高い。
3. 広さ=70, 築年数=10を代入すると、価格=4.89となるので、5.8千万円は相場より高い。

## 判別分析

検査値のデータ

被験者No	健常者・患者	検査値1	検査値2
1	健常者	50	15.5
2	健常者	69	18.4
3	健常者	93	26.4
4	健常者	76	22.9
5	健常者	88	18.6
6	患者	43	16.9
7	患者	56	21.6
8	患者	38	12.2
9	患者	21	16.0
10	患者	25	10.5

## 判別分析によってわかること

1. 疾病にかかっているか否かを検査値1と検査値2から判別できるか。
2. 判別できるとすれば、その精度はどれくらいか。
3. 例えば、検査値1=70, 検査値2=19.0などどのように判別されるか。

1. 判別式  $Z = -8.843 + 0.158 \times \text{検査値1}$  が求まって、 $Z \geq 0$ ならば健常者、 $Z < 0$ なら患者と判別する。
2. 本当は健常者なのに患者と誤判別する確率は0.1075, 本当は患者なのに
3. 健常者と誤判別する確率も0.1075。
4.  $Z = -8.843 + 0.158 \times \text{検査値1}$ に、検査値1=70 を代入すると $Z = 0.05$ となるので、健常者と判別される。

## 主成分分析

試験の成績のデータ

学生No	国語	英語	数学	理科
1	86	79	67	68
2	71	75	78	84
3	42	43	39	44
4	62	58	98	95
5	96	97	61	63
6	39	33	45	50
7	50	53	64	72
8	78	66	52	47
9	51	44	76	72
10	89	92	93	91

## 主成分分析によってわかること

1. 各科目の点数を総合することによって、より少ない変数(主成分と呼ばれる)で表現できないか。
2. 各主成分は、どのように解釈できるか？

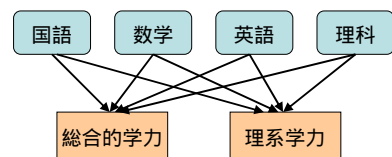
1. 主要な主成分として第1主成分z1と第2主成分z2を得る。

$$z1 = 0.487 \times \text{国語} + 0.511 \times \text{英語} + 0.508 \times \text{数学} + 0.493 \times \text{理科}$$

$$z2 = 0.527 \times \text{国語} + 0.474 \times \text{英語} - 0.481 \times \text{数学} - 0.516 \times \text{理科}$$

2. 係数の値より、z1は「総合的学力」を、z2は「理系と文系の学力の違い」を表すと解釈できる。

## 主成分分析のイメージ



# 因子分析

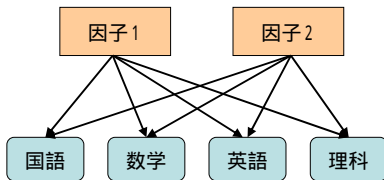
試験の成績のデータ

学生No	国語	英語	数学	理科
1	86	79	67	68
2	71	75	78	84
3	42	43	39	44
4	62	58	98	95
5	96	97	61	63
6	39	33	45	50
7	50	53	64	72
8	78	66	52	47
9	51	44	76	72
10	89	92	93	91

# 因子分析によってわかること

1. 各科目の点数を、1つあるいは、それ以上の共通した原因、(例えば、学力、理系的能力など)によって、表現できないか。これらの共通の原因は因子と呼ばれる。
2. これらの因子は、どのように解釈できるか？

# 因子分析のイメージ



# 教科書と参考書

## 教科書

涌井良幸, 涌井貞美「Excelで学ぶ多変量解析」ナツメ社, 2005年.

## 参考書

永田靖, 棟近雅彦「多変量解析法入門」サイエンス社, 2001年.

田中豊, 脇本和昌: 多変量統計解析法. 現代数学社, 1983年.

# この講義の履修に関する注意

- 出席は取らない。ただし、指名して答えてもらうことがある。
- 私語禁止(真面目に講義を聞いている人の邪魔をしてはいけない。あまりにひどい場合は、退室を命ずる。減点の対象するので名前を覚えてもらう。)
- 爆睡、内職禁止(講義に出る意味がないし、教員に対して失礼。減点の対象とするので、名前を覚えてもらう。)
- 教科書とMicrosoft ExcelがインストールされているノートPCを次回は持参してくること。