

データマイニングとモデルベース マネジメントを融合する試み

関東学園大学2007年度
プロジェクト型授業成果報告書

犬童ゼミ(演習 I・II)

プロジェクトの概要

- データマイニング(DM)とは、データにうずもれた法則性やパターンをコンピュータを駆使して掘り起こす技術です。一方、モデル分析(MA)によって、個人やグループの判断や行動について説明したり、予測することができます。
- 本プロジェクトでは、ExcelとR言語を活用し、DMとMAの実践的応用を試みました。

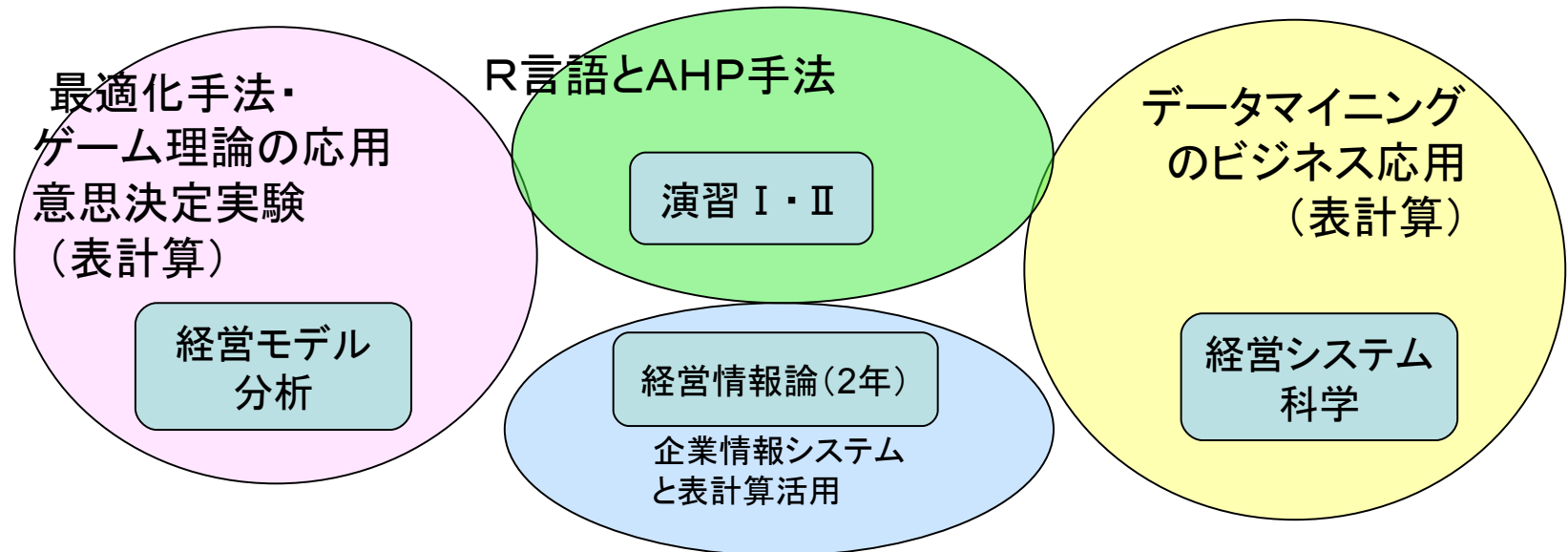


図1. 本プロジェクトに関係している科目群

成果

- 演習および関係する科目(図1参照)を履修し、最終的には次の作業を行いました。
- 「好みを計る」を統一テーマとして、以下のような4つの作業課題を設定しました。
 - A) ブランドについての意識調査
 - B) 資格志向の分析
 - C) 学食における麺類の新メニュー提案
 - D) 確率についての心理実験
- アンケート調査票をエクセルで作成し、図1の講義科目およびゼミ内において、アンケート調査*を実施しました(12/11~12)。
*ただしDにかんしては小テストの設問として出題。
- この調査データを集計し、以下のような分析を行いました。

手法・理論

- 単純集計以外に、それぞれの作業課題において用いた、主な手法や理論は以下のとおりです。
- 手法・理論
 - A) 主成分分析(およびクラスター分析)
 - B) AHP法
 - C) コンジョイント法(および決定木学習)
 - D) 確率、ベイズの法則
- 手法の説明 主成分分析やクラスター分析は、データを少数の主成分に分解し、好みの「基準」や回答者をグループ化します。AHP法は複数の基準の下での「好み」を、一対比較に基づき評価します。コンジョイント法は実験計画法を用い、とても少ないデータで「好み」を回帰分析します。決定木学習は「好み」を言い当てる質問を作ります。ベイズの方法は確率と「好み」を結びつけます。

数学的には、
固有値・固有ベクトル
問題を解く。

回答者のデータ

- 以下の表は、回答者数を、属性およびアンケートを実施した科目に分けて集計しました。
- 回収計43名、有効回答36名(女性3名、留学生1名を含む)

表1. 回答者の内訳※

- 性別
 - 1 男性
 - 2 女性
- 学年
 - 1 2年次
 - 2 3年次
 - 3 4年次以上
- 出身
 - 1 日本・関東
 - 2 日本・関東
 - 3 日本・関東
 - 4 海外・中国
 - 5 海外・それ
 - 6 その他

回答者数	学年			
出身	1	2	3	総計
1	6	14	6	26
2	2	3	2	7
3	1		1	2
4			1	1
総計	9	17	10	36

回答者数	学年			
科目	1	2	3	総計
演習 I・II		2	4	6
経営モデル分析		8	4	12
経営情報論	9	7	2	18
総計	9	17	10	36

※表1は、Excelで集約したアンケートの元データを、ピボットテーブルレポートを使って集計した。

各論

- それでは、各作業チームの成果を順次紹介いたします。なお時間が限られておりますので、詳細は後ほど添付資料をご覧ください。また質問等、頂ければ幸いです。
- まず、Aチームが「ブランド意識調査」の結果を発表します。これはどのような観点が、ブランドへの主観的評価を高めるのか調べたものです。(・・・スライドAへ)
- BチームはAHP法と言う手法を用い、「資格志向」について分析しました。(・・・スライドBへ)
- 次に、Cチームが「学食アンケート調査」の結果に基づき、新しい麺メニューを提案します。(・・・スライドCへ)
- 最後に、「確率」にかんする興味深い心理実験をDチームが報告します。(・・・続きはスライドDへ)

A) ブランドについての意識調査

- 仮説として、以下の6つの観点が、ブランドの評価に影響するのではないかと考えました。
 - A. 信頼できるから
 - B. 高級感があるから
 - C. 品質が良いから
 - D. 人に自慢できるから
 - E. 優越感をもてるから
 - F. 人気があるから

回答選択肢:

1. よく当てはまると思う
 2. やや当てはまると思う
 3. あまり当てはまらないと思う
 4. まったく当てはまらないと思う
- アンケート調査の結果、以下のことが分かりました。

分析結果(A-1)

- 単純集計 6つの観点全てについて、未記入(0)の8人を除く大部分の回答者が、あてはまると答えている(図A-1参照)。
- また平均や評価1と2の比率で見ると、B(高級感)とC(品質)、ついでF(人気)が最も重視されていることが分かる。

表A-1. 単純集計と平均

経営情報論のみ

評価\観点	A	B	C	D	E	F
0	6	6	6	6	6	6
1	6	12	13	7	5	11
2	9	6	4	8	10	6
3	3	0	1	3	3	1
4	0	0	0	0	0	0
計	24	24	24	24	24	24
1と2の比率	83%	100%	94%	83%	83%	94%
平均	1.8	1.3	1.3	1.8	1.9	1.4

3科目全体

評価\観点	A	B	C	D	E	F
0	8	8	8	8	8	8
1	10	24	18	17	15	22
2	17	9	13	11	11	10
3	7	1	4	3	4	3
4	1	1	0	4	5	0
計	43	43	43	43	43	43
1と2の比率	77%	94%	89%	80%	74%	91%
平均	2.0	1.4	1.6	1.8	2.0	1.5

※これらの表は、Excelで集約したアンケートの元データを、countif関数を使って集計した。

分散共分散と相関係数

- 変数間で値の動きが似ているかどうかを数値化するのが、共分散や相関係数である。
- 相関係数は0から1の範囲に収まり、各変数の持つスケールの影響を受けない。
- 図はExcelの分析ツールで計算し、条件付書式で色分けして示したものの。
- 観点Bは他のすべての観点と0.55以上の正の相関があり、かつ分散も小さいことが分かる。
- これらは次のA-2で用いる主成分分析の基礎となる。

分散共分散行列

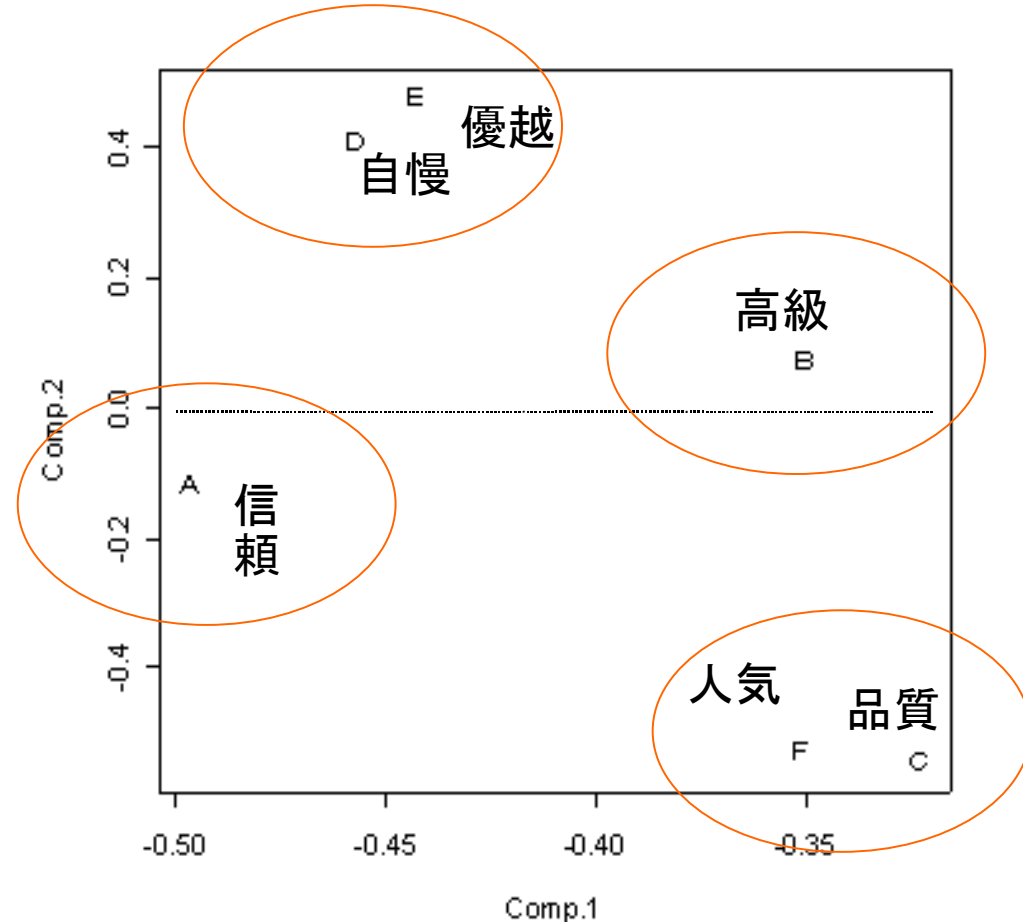
	A	B	C	D	E	F
A	1.05					
B	0.39	0.49				
C	0.56	0.39	0.72			
D	0.37	0.64	0.46	1.2		
E	0.35	0.58	0.38	1.03	1.39	
F	0.32	0.39	0.49	0.53	0.56	0.6

相関係数行列

	A	B	C	D	E	F
A	1					
B	0.55	1				
C	0.67	0.68	1			
D	0.38	0.84	0.55	1		
E	0.34	0.72	0.44	0.82	1	
F	0.44	0.76	0.77	0.67	0.66	1

分析結果(A-2)

- 次に主成分分析により、観点間の類似性を把握した(図A-2参照)。
- まず経営情報論のみを対象とすると、図から、6つの観点を4グループに分類できることが分かる。

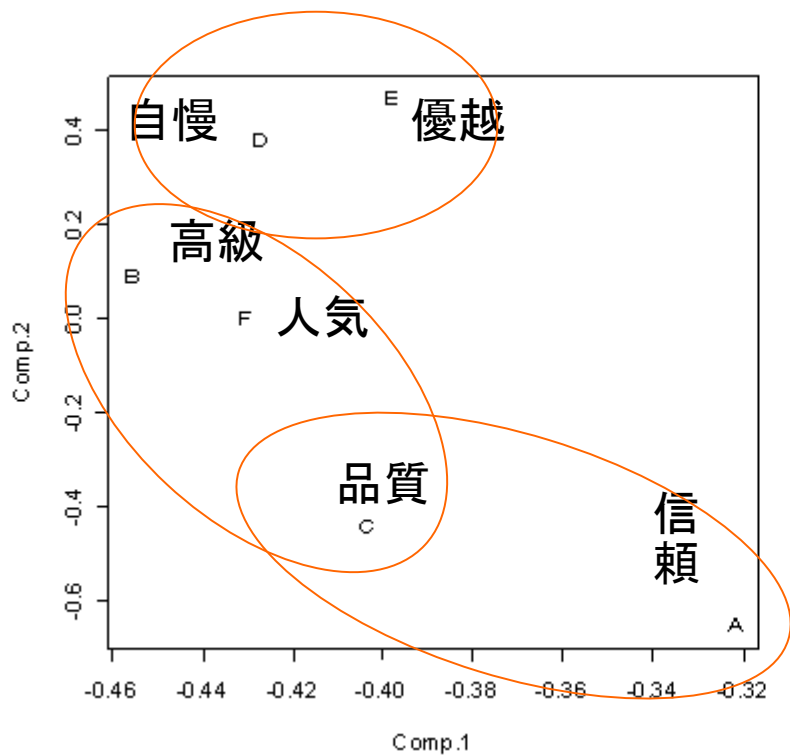


図A-2. 経営情報論のみの主成分分析

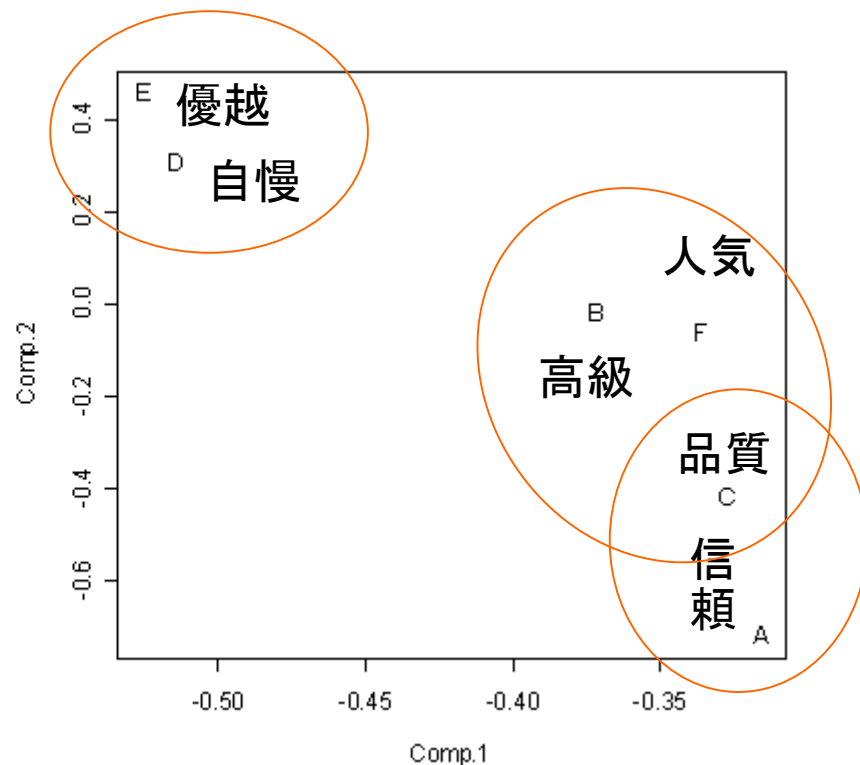
- Rのprcomp関数による。第2主成分までの累積寄与率は約84.5%)。ここでは共分散行列を用いたが、相関係数行列でも類似の結果を得る。ただしBはAに近づき、3グループにまとまる。

3科目分すすべての場合

- 3科目全てのデータを使うと、以下のような主成分(固有ベクトル)のマップが得られる。



相関係数行列を用いたときの第1～第2主成分(累積寄与率84.2%)



分散共分散行列を用いたときの第1～第2主成分(累積寄与率89.6%)

Rのコマンド(A-2)

- R(バージョン2.6.1)を用い、Excelからデータを読み込み、主成分分析を実行。

```
> x <- read.table("clipboard", header=TRUE)
```

```
> xp<-princomp(x[,-1])
```

```
> summary(xp)
```

Importance of components:	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Standard deviation	1.831366	0.783056	0.596746	0.454153	0.315445	0.252513
Proportion of Variance	0.714705	0.130666	0.075885	0.043952	0.021204	0.013588
Cumulative Proportion	0.714705	0.845371	0.921256	0.965208	0.986412	1

```
Loadings:
```

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
A	-0.497	-0.120	0.273	0.735	-0.270	-0.228
B	-0.349	0.106	0.160	0.832	0.379	
C	-0.323	-0.549	-0.219	-0.341	0.655	
D	-0.457	0.409	0.553	-0.534	-0.144	-0.111
E	-0.442	0.478	-0.737	-0.167		
F	-0.351	-0.532	-0.255	-0.318	0.265	-0.598

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
SS loadings	1.000	1.000	1.000	1.000	1.000	1.000
Proportion Var	0.167	0.167	0.167	0.167	0.167	0.167
Cumulative Var	0.167	0.333	0.500	0.667	0.833	1.000

```
> plot(xp$loading[,1:2],type="n")
```

```
> text(xp$loading,colnames(x[,-1]))
```

```
> plot(xp$scores[,1],xp$scores[,2],type="n")
```

```
> text(xp$scores[,1],xp$scores[,2],x[,1])
```

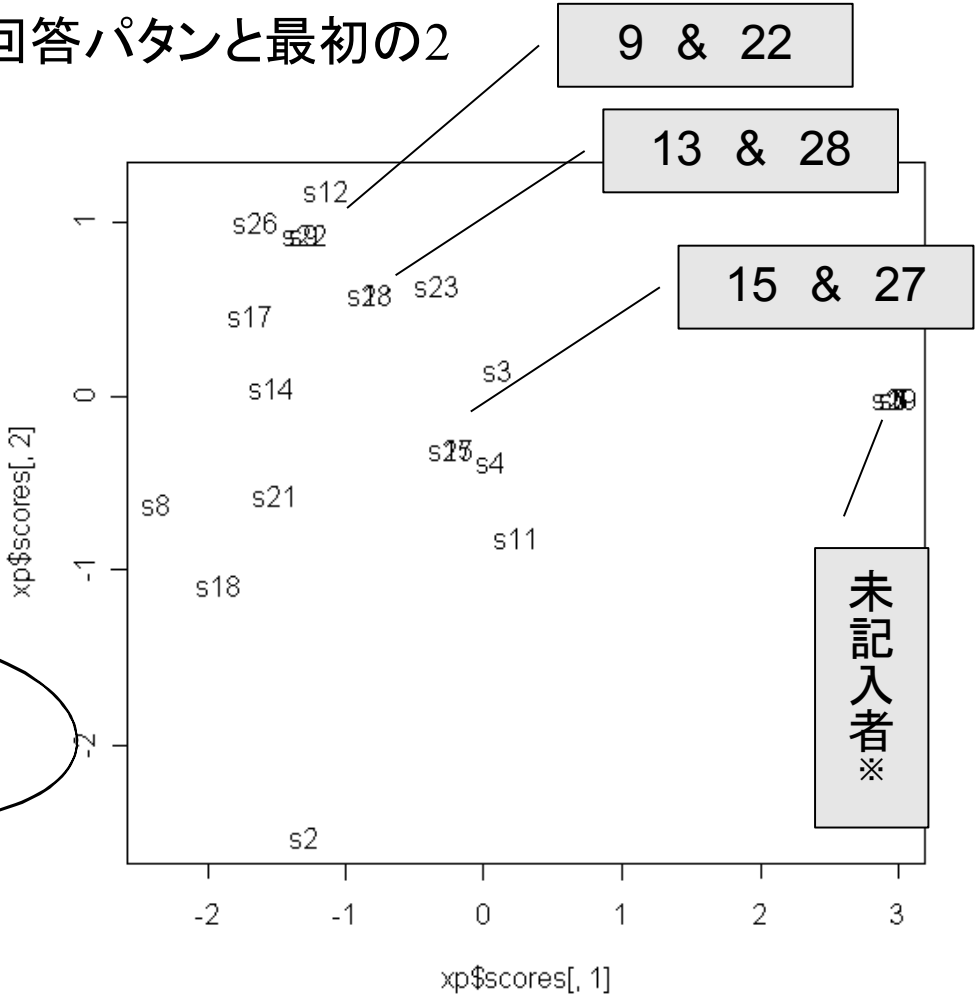
まずExcelでデータ範囲をコピーしておき、それからread.tableを実行することによってデータフレームxを作成する。



分析結果(A-3)

- 図は経営情報論の各回答者の回答パターンと最初の2つの主成分得点をプロットした。

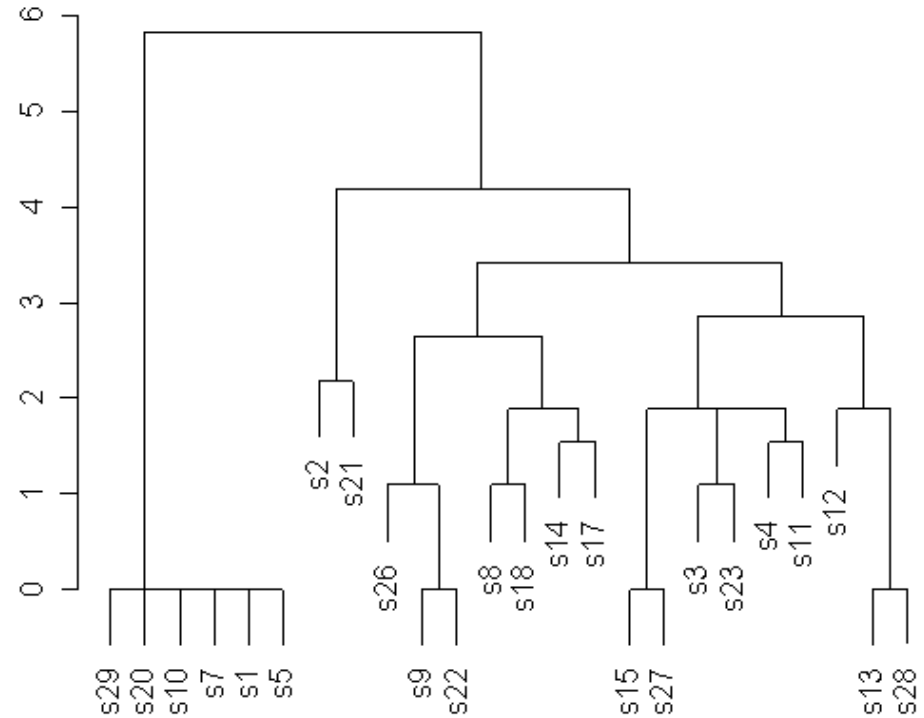
ブランド嗜好	A	B	C	D	E	F
s1	0	0	0	0	0	0
s2	2	1	3	1	1	3
s3	1	1	1	2	1	1
s4	2	1	1	1	1	1
s5	0	0	0	0	0	0
s7	0	0	0	0	0	0
s8	3	2	2	2	2	2
s9	2	1	1	3	2	1
s10	0	0	0	0	0	0
s11	1	1	2	1	1	1
s12	1	2	1	2	3	1
s13	2	1	1	1	3	1
s14	2	2	1	2	2	2
s15	1	1	1	1	2	2
s17	3	2	1	2	2	1
s18	3	2	2	2	1	2
s20	0	0	0	0	0	0
s21	2	1	2	2	2	2
s22	2	1	1	3	2	1
s23	1	1	1	2	2	1
s26	2	2	1	3	2	1
s27	1	1	1	1	2	2
s28	2	1	1	1	3	1
s29	0	0	0	0	0	0
average(s)	1.38	1	1	1.33	1.42	1.08
stdevp(s)	0.99	0.71	0.76	0.99	1	0.81



図A-3. ブランドについての回答パターン(経営情報論)と回答者の主成分得点プロット(第2主成分まで)

分析結果(A-4)

- 最後にクラスター分析による回答者の階層的グループ化を試みた(図A-4参照)。
- 例えば9と22は全く同じ回答パターン(距離0)、26はBの回答が1段階異なるだけ。
- 前の図の主成分得点で見ると、彼らは第1主成分得点が低く、第2主成分得点が高い(図A-3参照)。
- 同じく13と28、15と27も同一回答であり、この順に第2主成分の重みが下がる。



図A-4. クラスター分析

- Rのhclust関数による。距離はユークリッド距離に基づく。主成分分析のときと同じく経営情報論の受講者のみを分析している。左端の6人は回答未記入者(0)である。

Rのコマンド(A-4)

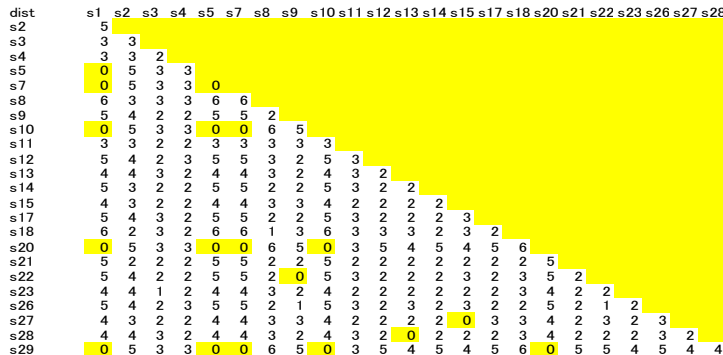
- 手順としては、まず距離行列を作ってから、クラスタ分析を実行する。

1) 距離行列作成

```
> x.d<-dist(x)
Warning message:
In dist(x): 強制変換により NA が生成されました
```

2) ファイルへ書き出し、Excelで視覚化する(これらの作業は必須ではない)

```
> sink("dist.txt")
> xd
> sink()
```



3) クラスタ分析実行

```
cluster analysis
> xc<-hclust(xd)
> xc
```

Call:
hclust(d = xd)

Cluster method : complete
Distance : euclidean
Number of objects: 24

```
> summary(xc)
      Length Class Mode
merge    46  -none- numeric
height   23  -none- numeric
order    24  -none- numeric
labels    0  -none- NULL
method    1  -none- character
call      2  -none- call
dist.method 1  -none- character
> plot(xc, x[,1])
>
```

分析のまとめ(A)

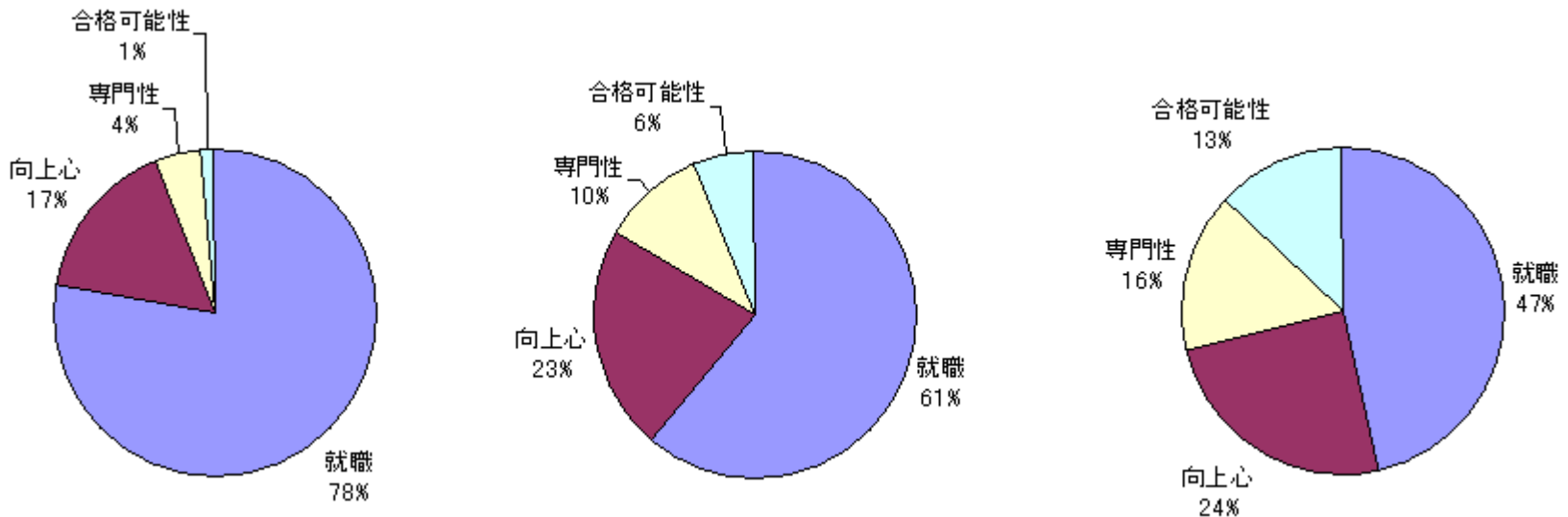
- 一見すると、6つの観点はブランドの共通イメージを言い換えたただけのように思えた。しかし主成分分析によって以下の4点が明らかになった。
 - ① 「全く重視しない」(4)を選んだ人はおらず、どの観点一定の重要性がある。
 - ② ブランドの価値を最もよく代表するのは、高級感(B)である。これに次いで品質(C)や人気(F)も、重視される傾向がある。
 - ③ 信頼感(A)はこれらほど重視されないが、相対的には品質に近い。
 - ④ 自慢(D)や優越感(E)はやや重視されないが、相関で見ると高級感に近い。

B) 資格志向の分析

- 目的 以下の4種類の資格について、学生の資格取得に対する関心の持ち方を調べる。
英検 簿記 P検 教職
- 手法 以下の4つの観点(評価基準)に着目し、AHP法を用いて、資格志向をモデル分析する。
 - 1 就職に有利になる
 - 2 向上心を高める
 - 3 専門知識やスキルが身に付く
 - 4 合格可能性がある
- 得られた知見 平均で見ると、まず就職、次いで向上心、専門性、合格可能性の順に重視されている。この傾向は3科目を通じて変わらないが、ゼミは就職重視の傾向が最も強かった。また、4つの資格中、英検が最も重視されている。

分析の結果(B-1)

- 図B-1に、AHP法によって得られた観点間の相対的重要性(科目ごとに平均したもの)を示す。

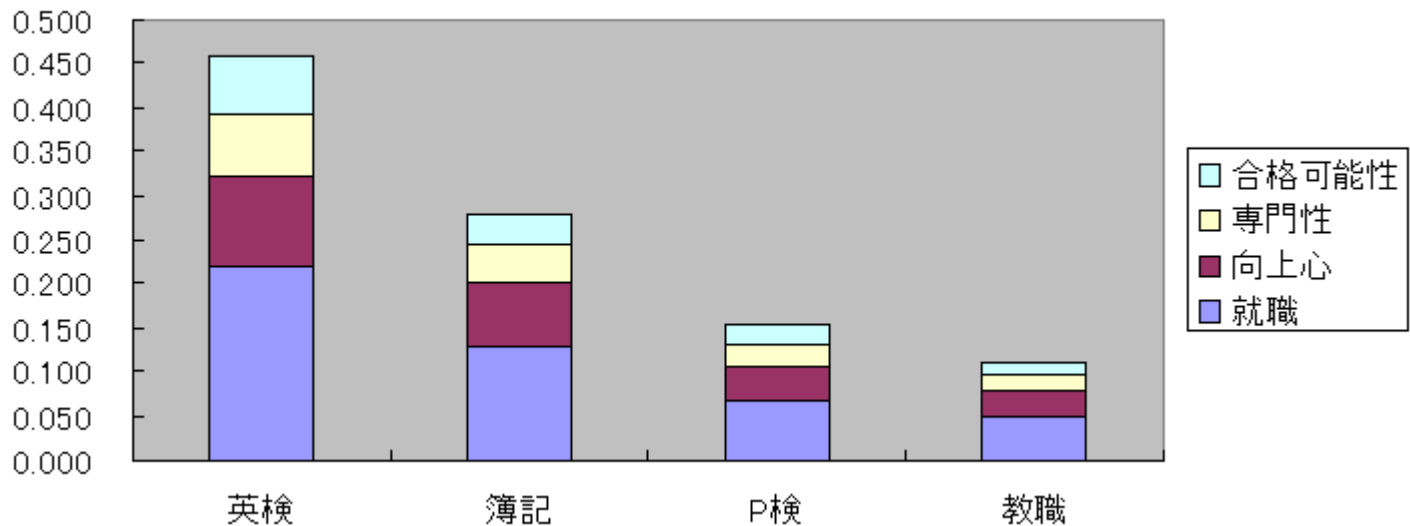


図B-1. AHP法による要因間の相対的重要性

アンケート実施3科目について、それぞれ回答者の評価ベクトルを平均した。
左からゼミ(演習)、経営モデル分析(3年次開講)、経営情報論(2年次開講)

分析の結果(B-2)

- 平均で見ると、英検が最も重視されている。それ以下は、簿記、P検、教職の順に重視されている。またこの順位は3科目を通じて変わらない。
- 一例として、回答者数が一番多かった経営情報論の結果を図B-2に示そう。



図B-2. AHP法によって得られた4資格の評価値(経営情報論)

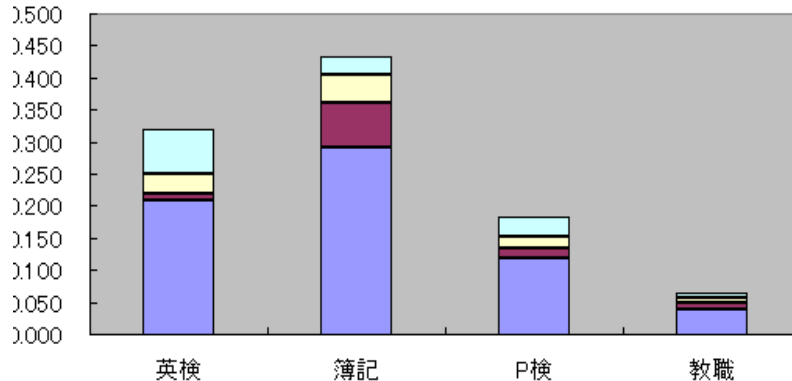
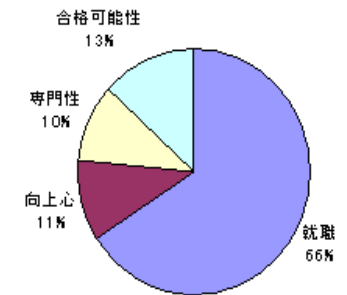
分析の結果(B-3)

- 個別回答者の資格志向をAHP法に基づき分析するためのRのプログラムおよび表計算モデルを開発した。
- 図は表計算(Excel)を用い、簿記を重視する学生の一例を示している(図B-3参照)。

data No.	name	的中度	整合性	max	average	stdev	要因間重み
12	s12	5	0.04196	0.36108	0.199737361	0.099538	

得点	英検	簿記	P検	教職	要因間重み
就職	0.318	0.445	0.180	0.057	0.655
向上心	0.105	0.635	0.148	0.112	0.109
専門性	0.307	0.439	0.181	0.073	0.102
合格可能性	0.505	0.212	0.222	0.061	0.134
AHP(階層)	0.319	0.434	0.182	0.065	
直接AHP	0.103	0.623	0.190	0.084	

総合評価	英検	簿記	P検	教職
就職	0.208	0.292	0.118	0.038
向上心	0.011	0.069	0.016	0.012
専門性	0.031	0.045	0.018	0.007
合格可能性	0.068	0.028	0.030	0.008



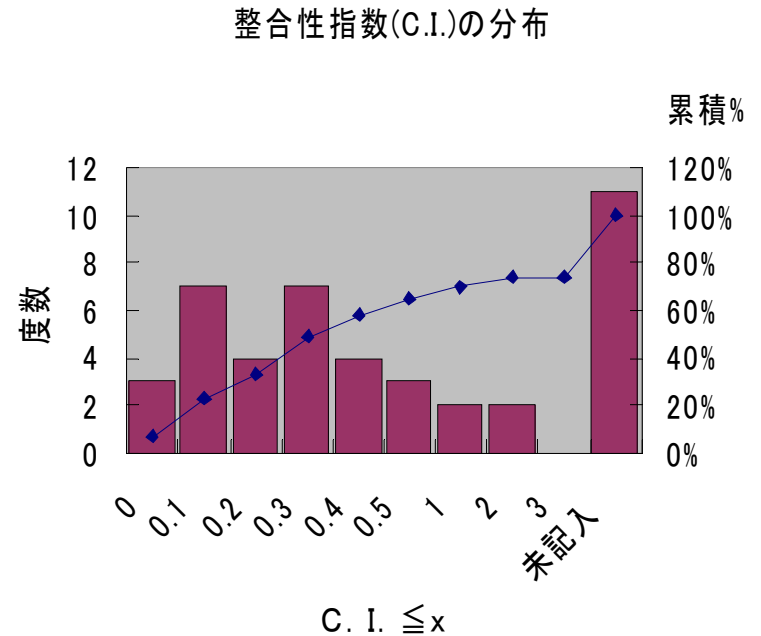
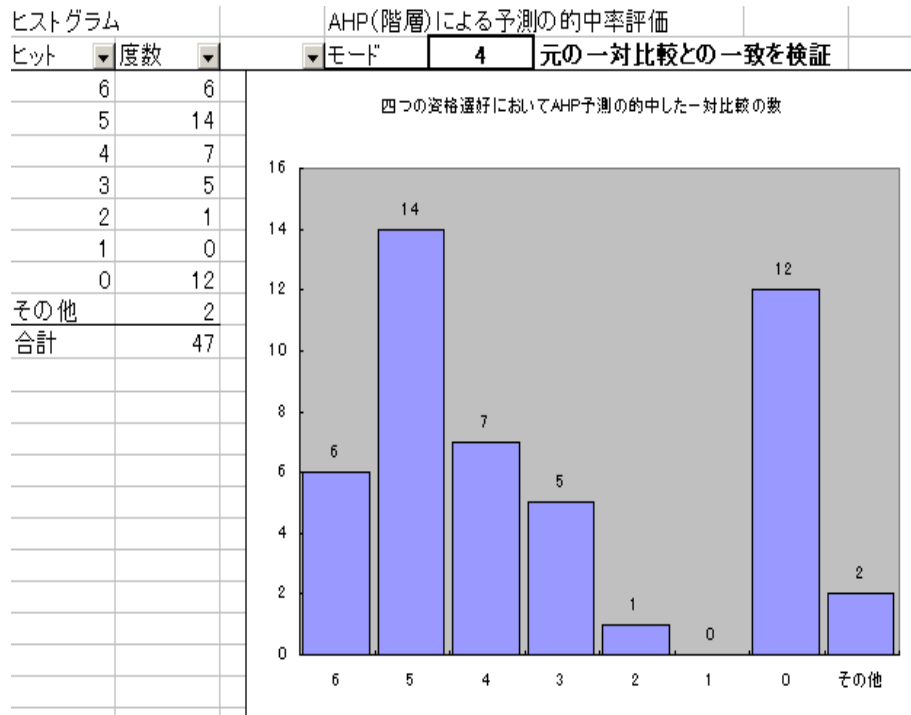
入B/クロスC/整合性/AHP/資格分析/クロス表分析/回答データ/集計/

図B-3. 個別回答者の資格志向を分析するワークシートモデル

分析の結果(B-4)

- ところで、AHP法の問題点として、(1)一対比較の入力が面倒で、評価に矛盾が生じること、(2)代替案の追加によって評価逆転が生じることなどが知られている。
- (1)についてはSaatyの整合性指数(C. I.)を求め、0.1~0.2程度なら許容するという経験則が用いられることが多い。
- 一方、(1)や(2)を心理学的に捉え、人間や組織の「認知的な歪み」や「意思決定のアノマリー」を計る指標として活用することが研究されている。
- 実際、今回のアンケートでは4つの評価基準からの一対比較だけでなく、資格間の直接比較も合わせて行い、順位の逆転がどの程度起きているのかについても分析している(図B-4参照)。

的中度および整合性

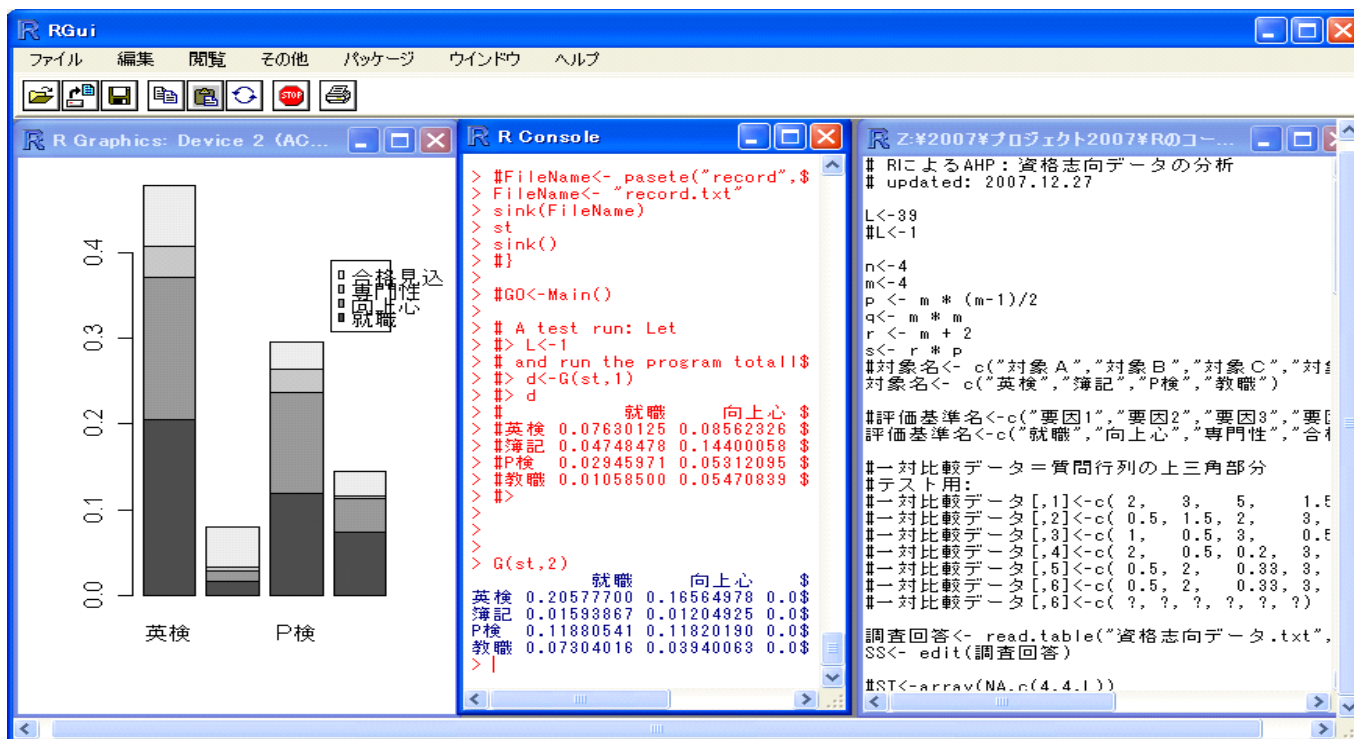


最重視資格	CI平均値
1 英検	0.141504
2 簿記	0.107387
3 P検	0.294249
4 教職	0.206304

図B-4. AHP法による一対比較予測的中度(左)および整合性指数の分布(右)

分析の結果 (B-5)

- 本プロジェクトでは、Excelの他、文献[2]におけるAHP法のR用スクリプトを一般化したプログラムを開発し、分析した。図B-5にその実行画面を示す。



図B-5. Rによる資格アンケートのAHP分析用システム

まとめ(B)

- 就職への意識が、年次を追うにつれ資格志向に強く作用していることが分かった。また、やや意外なことに、英検を重視する人が多かった。
- AHPによる予測は、直接資格同士を比較した場合と比べると、1か2つのペアまでの逆転が生じるものの、おおむね回答者の資格志向を再現していた。
- また最重視する資格で回答者をグループ化すると、簿記を重視するグループが最も整合性(C.I.)が良く、P検重視のグループが最も悪かった。
- ところで、AHP法ではC. I. の値が悪い場合、一対比較をやり直す。本実験では時間の都合で行っていないが、にもかかわらず一定の再現性が得られた。これはAHPが「好み」のモデル化に有用かもしれないことを示唆する。

C) 学食新メニューの提案

- 目的 学食に新しい麺類メニューのコンセプトを提案するとしよう。コンセプトを成功させるため、どのような麺類メニューを学生が欲しているかを予測したい。
- 手法 表C-1に示す8つの要因、およびそれぞれの要因に対する2つないし3つの水準値を考えた。またExcelでアンケート調査を作成、コンジョイント法と決定木学習を用いて分析した。

表C-1. 新メニューのコンセプト開発のための要因と水準

要因	第1水準	第2水準	第3水準
アレルギーに配慮する		しない	-
デザート	100円で付く	50円で付く	なし
ボリューム	少なめ	多め	普通
具財	肉	魚介	野菜
スープの濃さ	濃い	薄い	普通
調理時間	1分以内	3分以上	長くて3分
価格	350円以内	450円まで	500円前後
タイプ	イタリアン	中華風	和風

実験計画(C-1)

- 麺メニューの新しいコンセプトを決めるには、各要因につき一つの水準を選べばよいが、その組み合わせは膨大である。直交表 L^{18} を用いれば、ごく少量のアンケートデータで精度の高い回帰モデルが得られる。表C-2に表C-1の要因水準を割り当てた結果を示す。

メニュー1
 アレルギーに配慮した
 デザートは100円で付く
 ボリューム多めの
 具財に肉を多く使った
 スープが濃い
 調理時間に1分以内程かかる
 価格350円以内の
 和風の麺料理

メニュー10
 アレルギーに配慮しない
 デザートは50円で付く
 ボリューム多めの
 具財に野菜を多く使った
 スープが普通
 調理時間に1分以内程かかる
 価格350円以内の
 中華風の麺料理

メニュー11
 アレルギーに配慮しない
 デザートはなし
 ボリューム普通の
 具財に肉を多く使った
 スープが濃い
 調理時間に長くて3分程かかる
 価格350円以内の
 イタリアンの麺料理

表C-2. 直交表 L^{18} に割付けした18種類の麺メニュー

No.	アレルギー	デザート	ボリューム	具財	スープ	調理時間	価格	タイプ
1	する	100円で付く	多め	肉	濃い	1分以内	350円以内	和風
2	する	50円で付く	普通	魚介	薄い	長くて3分	350円以内	中華風
3	する	なし	少なめ	野菜	普通	3分以上	350円以内	イタリアン
4	する	なし	多め	肉	薄い	3分以上	450円まで	中華風
5	する	100円で付く	普通	魚介	普通	1分以内	450円まで	イタリアン
6	する	50円で付く	少なめ	野菜	濃い	長くて3分	450円まで	和風
7	する	50円で付く	多め	魚介	濃い	3分以上	500円前後	イタリアン
8	する	なし	普通	野菜	薄い	1分以内	500円前後	和風
9	する	100円で付く	少なめ	肉	普通	長くて3分	500円前後	中華風
10	しない	50円で付く	多め	野菜	普通	1分以内	350円以内	中華風
11	しない	なし	普通	肉	濃い	長くて3分	350円以内	イタリアン
12	しない	100円で付く	少なめ	魚介	薄い	3分以上	350円以内	和風
13	しない	なし	多め	魚介	普通	長くて3分	450円まで	和風
14	しない	100円で付く	普通	野菜	濃い	3分以上	450円まで	中華風
15	しない	50円で付く	少なめ	肉	薄い	1分以内	450円まで	イタリアン
16	しない	100円で付く	多め	野菜	薄い	長くて3分	500円前後	イタリアン
17	しない	50円で付く	普通	肉	普通	3分以上	500円前後	和風
18	しない	なし	少なめ	魚介	濃い	1分以内	500円前後	中華風



ゼミ担当教員が開発したExcel教材を使い、 L^{18} 直交表の割付け、質問項目の設計、回帰分析が簡単に行える。

分析結果(C-2)

- 以上18種類の仮想メニューを4段階で評価してもらった。
 - 1 ぜひ食べたいと思う
 - 2 やや食べてみたい
 - 3 どちらともいえない
 - 4 食べたいとは思わない
- また「お好みメニュー」上位3つまで回答してもらった。
- 図C-2に、各メニューがノミネートされた度数および順位による荷重スコアを示す。

メニューNo.	お好み1	お好み2	お好み3	荷重スコア	rank
1	17	1	0	53	1
2	1	2	5	12	5
3	1	0	0	3	14
4	0	2	0	4	12
5	0	2	1	5	11
6	2	0	1	7	8
7	2	2	1	11	6
8	1	1	1	6	10
9	0	7	0	14	4
10	4	5	4	26	2
11	1	4	4	15	3
12	1	1	2	7	8
13	1	1	3	8	7
14	0	0	1	1	18
15	0	1	1	3	14
16	0	0	2	2	16
17	0	0	2	2	16
18	0	1	2	4	12

	score	menu no.
rank 1	53	1
rank 2	26	10
rank 3	15	11

図C-2. 各メニューをお好みメニューにノミネートしている人の数および荷重スコアに基づくランキング

分析結果 (C-3)

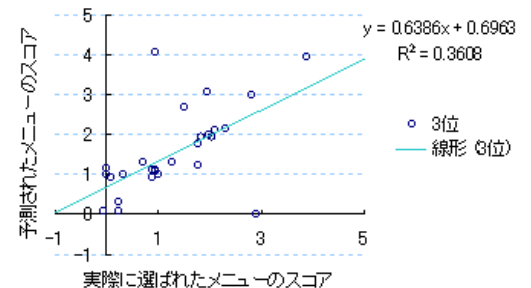
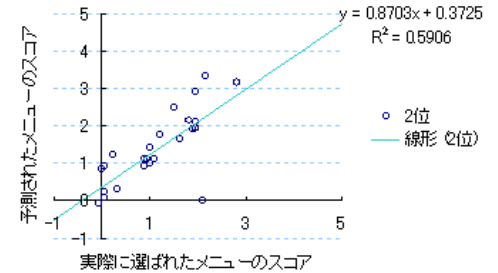
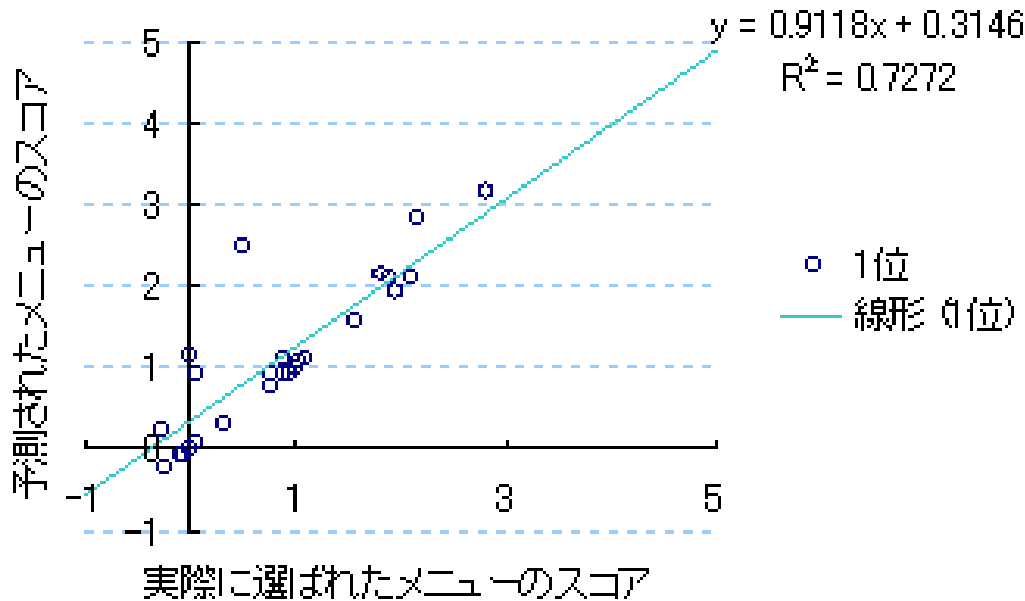
- 次に回帰分析によって各回答者の「好み」を予測する。図C-3は各科目の平均データに対する予測モデルである(Excelの配列関数で計算するが詳細は省く)。
- またそれが十分良いモデルであれば、C-1の要因水準に基づく、あらゆるメニューについて、評価を予測できることになる。

回帰係数	科目1平均	科目2平均	科目3平均
定数	2.032	1.617	1.601
する	-0.005	-0.060	-0.086
しない	0.005	0.060	0.086
100円で付く	-0.234	-0.160	-0.215
50円で付く	0.058	-0.016	-0.052
なし	0.176	0.176	0.267
少なめ	-0.060	-0.028	-0.071
多め	0.016	-0.087	-0.084
普通	0.044	0.115	0.155
肉	-0.046	-0.020	-0.132
魚介	0.058	-0.072	-0.061
野菜	-0.012	0.092	0.193
濃い	0.009	0.030	-0.010
薄い	0.023	0.101	0.098
普通	-0.032	-0.131	-0.088
1分以内	-0.019	0.002	0.009
3分以上	-0.005	0.023	0.024
長くて3分	0.023	-0.024	-0.032
350円以内	-0.025	-0.054	-0.084
450円まで	0.023	0.003	-0.010
500円前後	0.002	0.051	0.095
イタリアン	-0.081	-0.105	-0.103
中華風	0.037	-0.003	-0.003
和風	0.044	0.108	0.106

図C-3. 新メニューの「好み」を予測する回帰モデル(各科目平均に対して)

分析結果(C-4)

- モデルが「良い」基準は、実際に選ばれたお好みメニューのスコアを正しく予測できているかどうかで判断する。各順位の結果を図C-4にまとめた。
- 1位の予測は、まずまずである。2位以下はやや適合が劣る。



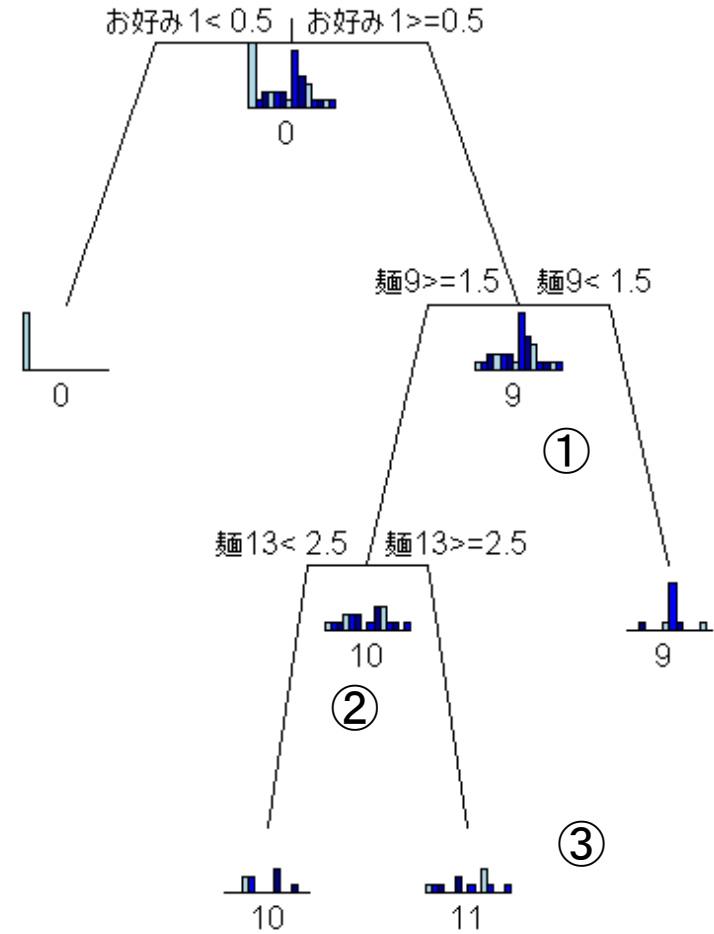
図C-4. お好みメニューの選択にかんしてモデルの予測力を評価する

分析結果(C-5)

- 最後に決定木学習によって、図C-5に示すような、お好みメニュー第2位を予測する質問ルール(決定木)を得た。

根から順に質問していくと、次のように推理できる。

- お好み1位未記入者以外(木の右半分)は、平均すると9番を第2位に選ぶ傾向がある…①
- また9を強く好まなければ、10番を選ぶ傾向がある…②
- さらに13番を好まなければ11番を選ぶ…③



図C-5. 決定木によるお好み3位予測モデル (mvpartパッケージrpart関数を使用)

まとめ(C)

- 以上の分析から得られた予測モデルを使って、私たちは、図C-5に示すような、各科目の回答者グループに対するベストコンセプトを提案する。
- なお、これらはいずれも18のメニュー中になかった組合せである。

要因	ベストコンセプトの水準			お好みメニューの水準		
	科目1平均	科目2平均	科目3平均	メニュー1	メニュー10	メニュー11
アレルギーに配慮	する	する	する	する	しない	しない
デザート	100円で付く	100円で付く	100円で付く	100円で付く	50円で付く	なし
ボリューム	少なめ	多め	多め	多め	多め	普通
具財	肉	魚介	肉	肉	野菜	肉
スープの濃さ	普通	普通	普通	濃い	普通	濃い
調理時間	1分以内	長くて3分	長くて3分	1分以内	1分以内	長くて3分
価格	350円以内	350円以内	350円以内	350円以内	350円以内	350円以内
タイプ	イタリアン	イタリアン	イタリアン	和風	中華風	イタリアン

D) 確率についての心理実験

- 以下は上記のアンケートと同時に行った小テスト課題の一部であり、①～③の3つの小問を含む。

問6

あるメーカーは主要製品のある重要な部品の調達にかんして、以下のような問題に直面している。これまでは系列のA社から調達してきたが、経営環境が変わり、B社とC社からも代替的に調達可能になった。今回、調達先は3社のうち、1社に絞られるが、どの調達先になるかは特定できない(いずれも確率1/3)。

- ① あなたはこのメーカーを担当するA社の営業課長であり、役員をゴルフ接待している。あなたがこの役員に、「B社とC社で取引されない方を1社教えてくださいませんか」とたずねたとする。「今回、社長はB社と取引する気がない。」と答えた。この役員の発言は確かに信じられるものとしよう。さて、A社が今回の調達先となる可能性は、役員の発言を聞いた後、どうなったと思いますか？

① 回答欄
選択肢

確率

上がった
下がった
変わらない

実験結果 (D-1)

- 結果は図D-1のようであった。この問いは、3ドア・クイズ問題、あるいは3囚人問題として知られるものをアレンジしたものである。
- 3つの箱に1つだけ賞品が入っており、回答者に一つを選ばせ、実験者が残りの2つのうち、空箱の方を示す。多くの人が、更新された確率を二分の一と思って答える。しかし、じつは条件付確率(ベイズの法則)にしたがって計算すると、正解は三分の一のまま変わらず、またもう一方の確率が三分の二になる。
- 本実験では、文献で知られているのと同様の傾向が観察された。

変化

下がった	2
上がった	22
変わらない	5

◀正解

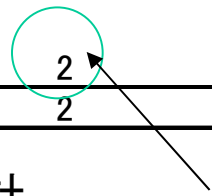
確率の値

0	7
0.5	20
1/3	2
300/1	1

◀正解

クロス集計

		問6③値				
問6①		無記入(又は0)	1/2	1/3	他(300/1)	総計
無記入		1				1
下がった		2				2
上がった		4	17		1	22
変わらない			3	2		5
総計		7	20	2	1	30



図D-1. 問6①の回答集計

本当の正解

実験結果(D-2)

- 続く小問②とその回答の集計結果を、図D-2に示す。これはエルズバーグの背理(2色問題)と呼ばれる問題をアレンジしたものである。
- 文献では、人々が曖昧な確率を嫌い、確率の分かっている方を選ぶ傾向が報告されている。
- 本実験では、B社とC社の確率が分からず、平均すればA社と同じであることだけ分かっている。それゆえ、件の背理からは、A社を選ぶ傾向が予想される。
- ところが、前の設問①で、B社の可能性が排除されていると考えた回答者が、C社を選ぶ傾向があった。注目すべきことに、その比率は2/3に近かった。

② あなたが組立てメーカーの経営管理者であり、調達先を決定できるとします。品質管理担当者に聞くと、「A社のこれまでの不良品率は3%でした。B社とC社のそれぞれ個別の不良品率については、取引実績がないため、分かりませんが、外部調査によってBとCを合わせて3%の不良品率であると考えてまちがいありません。」と答えた。さて、あなたはいずれを調達先を選びますか？

② 回答欄
選択肢

A社
B社
C社

選択

A社	10
B社	3
C社	16

図D-2. 問6②の回答集計

実験結果 (D-3)

- 最後の小問③とその結果を、図D-3に示す。
- 前の問②を少し変えると、再びエルズバーグの別の背理(3色問題)にアレンジしなおせる。
- この場合、文献で報告されているのと同様に、BとCを一括することにより、確率の曖昧さが解消されるため、この組合せを選ぶ傾向が見られた。

③ 次回の取引は3社の内、2社を選定することになっています。さて、メーカーの経営管理者であるあなたはどのようにしますか？

③ 回答欄
選択肢

- A社とB社を選ぶ。
- B社とC社を選ぶ。
- A社とC社を選ぶ。

理由 (自由に記述してください)

選択

A社とB社を選ぶ。	8
A社とC社を選ぶ。	6
B社とC社を選ぶ。	15

図D-3. 問6③の回答集計

まとめ(D)

- 本実験は、確率とベイズの法則について、それぞれのよく知られた心理学的実験であるエルズバーグの背理と、3ドア・クイズ問題を、同一の経営学的な題材を使ってアレンジした。
- 3ドア問題をアレンジした①では、文献どおり、ベイズの法則の正解に達する回答者は僅かであった。
- またエルズバーグの背理のアレンジ③では、文献どおり、曖昧な確率を嫌う傾向が見られた。
- ところが、エルズバーグの背理を3ドア問題に埋め込んで提示した②では、曖昧性の回避が消失し、また集団的傾向として、ベイズの解に近い選択確率が得られた。

全体のまとめ

- 本プロジェクトを通じて、講義で習った基本的な情報分析や意思決定の手法やツールが、その先にどのように具体的問題に応用できるかを、示すことができた。

参考文献・URL

- 金明哲, 『Rによるデータサイエンス』 森北出版, 2007.
- 竹内俊彦, 『S-Plus/R言語プログラミング』 オーム社, 2006.
- 上田和明ら, 『Excelで学ぶデータマイニング入門』 オーム社, 2005.
- 水田正弘ら, 『S-PLUSによるデータマイニング入門』 森北出版, 2005.
- 時永祥三, 『Sによる経営情報解析』 牧野書店, 1993.
- 岡田昌史, RjpWiki, <http://www.okada.jp.org/RWiki/>
- 青木繁伸, Rによる統計処理,
<http://aoki2.si.gunma-u.ac.jp/R/>
- 犬童健良, http://www.us.kanto-gakuen.ac.jp/indo/xls_sheets/ANOVA.xls (計画行列による回帰モデルの計算法を説明している)

ご清聴ありがとうございました。

プロジェクトメンバー一同

添付資料. アンケート調査の内容

はじめに——本アンケートについてのご説明

本アンケート調査は犬童ゼミ(演習Ⅰ・Ⅱ)のプロジェクト型授業および卒業研究の一環として行われるものです。

プロジェクトの研究テーマは「好みを測定する」です。(注)

(注)正式には「データマイニングとモデルベースマネジメントを融合する試み」ですが、分かりやすくするため、そのように称することにしました。

どうぞご協力をお願いいたします。

分析の結果は、プロジェクトの成果報告書を通じて公表します。
本調査のデータは本プロジェクトおよび卒業研究のためだけに用い、
公表資料に個人を特定できるデータは掲載しません。

2007年12月

「好みを測定する」プロジェクトメンバー一同(犬童の演習Ⅰ・Ⅱ受講生)

最初に属性についての質問です。

あなたの性別を教えてください(番号をお願いします)

- 1 男性
- 2 女性

回答

あなたの学年を教えてください(番号をお願いします)

- 1 2年次
- 2 3年次
- 3 4年次以上

回答

あなたの出身地を教えてください(番号をお願いします)

- 1 日本・関東
- 2 日本・関東より北
- 3 日本・関東より南
- 4 海外・中国
- 5 海外・それ以外のアジア
- 6 その他

回答

ありがとうございました。次にブランドにかんするアンケートに進んでください。

Q1.ブランド嗜好についてのイメージ

服飾関係ではチャネル、グッチなど有名なブランドがあります。
人がブランドを好むのはなぜだと思いますか？

以下の該当する選択肢の番号を選んでください。

- 1 よく当てはまると思う
- 2 やや当てはまると思う
- 3 あまり当てはまらないと思う
- 4 まったく当てはまらないと思う

- A. 値段できるから
- B. 高級感があるから
- C. 品質が良いから
- D. 人に自慢できるから
- E. 優越感をもてるから
- F. 人気があるから

選択肢の番号

ありがとうございました。次に資格にかんするアンケートに進んでください。

Q2.資格取得の難力について

現代の大学生は資格取得に熱心が強くなっているといわれます。
4つの観点から以下の資格を2つずつ比べて、どちらがどれだけ魅力的かと思われるかお聞きします。
該当する数字を選んでください。

観点	資格の種類	説明
1 就職に有利になる	英検	英語・英会話に関する検定。TOEFLやTOEFLはより上位の資格
2 向上心を奮める	簿記	自衛隊簿記は1級まである。税理士、公認会計士はその上位資格にあたる。
3 専門知識やスキルが身に付く	P検	パソコン検定。日商PC検定、国家資格である初級シニアドや情報処理技術者などが上位資格。
4 合格可能性がある	教職	卒業要件とは別に教員免許を取得するための大学のカリキュラム。

1. 就職に有利になることを考えたとき

- 問1 英検は簿記と比べて と思う
- 問2 英検はP検と比べて と思う
- 問3 英検は教職と比べて と思う
- 問4 簿記はP検と比べて と思う
- 問5 簿記は教職と比べて と思う
- 問6 P検は教職と比べて と思う

選択肢

- 大変難力がある
- ここそこ難力がある
- かなり難力がある
- 同じくらい
- 簡単
- 簡単すぎる

AHPで分析するときの二対比較評価値

数値	選択肢
9	大変難力がある
7	難力がある
5	ここそこ難力がある
3	かなり難力がある
1	同じくらい
1/3	やや難力が乏しい
1/5	あまり難力を感しない
1/7	難力がない
1/9	まったく難力がない
0	(未回答)

2. 向上心を高めることを考えたとき

- 問1 英検は簿記と比べて と思う
- 問2 英検はP検と比べて と思う
- 問3 英検は教職と比べて と思う
- 問4 簿記はP検と比べて と思う
- 問5 簿記は教職と比べて と思う
- 問6 P検は教職と比べて と思う

選択肢

- やや難力が乏しい
- あまり難力を感しない
- 難力がない
- 難力がない
- 難力がない

3. 専門知識やスキルが身に付くことを考えたとき

- 問1 英検は簿記と比べて と思う
- 問2 英検はP検と比べて と思う
- 問3 英検は教職と比べて と思う
- 問4 簿記はP検と比べて と思う
- 問5 簿記は教職と比べて と思う
- 問6 P検は教職と比べて と思う

選択肢

- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)

4. 合格可能性があることを考えたとき

- 問1 英検は簿記と比べて と思う
- 問2 英検はP検と比べて と思う
- 問3 英検は教職と比べて と思う
- 問4 簿記はP検と比べて と思う
- 問5 簿記は教職と比べて と思う
- 問6 P検は教職と比べて と思う

選択肢

- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)

5. 観点同士を比較したとき

- 問1 就職は向上心と比べて と思う
- 問2 就職は専門性と比べて と思う
- 問3 就職は合格見込と比べて と思う
- 問4 向上心は専門性と比べて と思う
- 問5 向上心は合格見込と比べて と思う
- 問6 専門性は合格見込と比べて と思う

選択肢

- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)

6. 最後に、総合的に資格同士を比較したとき

- 問1 英検は簿記と比べて と思う
- 問2 英検はP検と比べて と思う
- 問3 英検は教職と比べて と思う
- 問4 簿記はP検と比べて と思う
- 問5 簿記は教職と比べて と思う
- 問6 P検は教職と比べて と思う

選択肢

- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)
- (未回答)

ありがとうございました。最後に学食にかんするアンケートに進んでください。

学食に関するアンケートに回答する

Q3.学食メニューについて

もし学食で以下のような18種類の種類のメニューがあったとしたら、あなたはどの思われますか？
それぞれのメニューに対して、回答欄に以下から該当する選択肢を選んでください。

- 1 ぜい食べたいと思う
- 2 やや食べてみたい
- 3 どちらともいえない
- 4 食べたいとは思わない
- 5 (未回答)

1. メニュー1

アレルギーに配慮したデザートは100円で付く
ボリューム多めの具材に肉を多く使ったスープが濃い
調理時間に1分以内かかる
価格350円以内の和風の種料理

一つのメニュー(料理)を想像してください。

回答欄(メニュー1)
食べたいとは思わない

2. メニュー2

アレルギーに配慮したデザートは50円で付く
ボリューム普通の具材に魚介を多く使ったスープが薄い
調理時間に長く3分程かかる
価格350円以内の中華風の種料理

回答欄(メニュー2)
(未回答)

3. メニュー3

アレルギーに配慮したデザートはなし
ボリューム少なめの具材に野菜を多く使ったスープが普通
調理時間に3分以上程かかる
価格350円以内のイタリアンの種料理

回答欄(メニュー3)
(未回答)

・・・途中省略・・・

17. メニュー17

アレルギーに配慮しないデザートは50円で付く
ボリューム普通の具材に肉を多く使ったスープが普通
調理時間に3分以上程かかる
価格500円前後の和風の種料理

回答欄(メニュー17)
(未回答)

18. メニュー18

アレルギーに配慮しないデザートはなし
ボリューム少なめの具材に魚介を多く使ったスープが濃い
調理時間に1分以内かかる
価格500円前後の中華風の種料理

回答欄(メニュー18)
(未回答)

19. あなたの気に入ったメニューがあれば、番号が3つまであげてください。

回答欄(お気に入り)

1位	<input type="text"/>
2位	<input type="text"/>
3位	<input type="text"/>

参考情報(右の回答に合わせて自動表示され)

メニュー8

アレルギーに配慮したデザートは50円で付く
ボリューム少なめの具材に野菜を多く使ったスープが濃い
調理時間に長く3分程かかる
価格450円までの和風の種料理

ありがとうございました。これでアンケートは終わりですが、アンケート回収の手順を述べますので、指示に従ってください。