

2020 年度 修士論文

論文題目

データマイニングによる

学生の通学状況及び目的意識が成績に与える影響の分析

(Analysis of the Influence of Student Attendance
and Sense of Purpose on Grades by Data Mining)

指導教員

舟橋	健司	准教授
伊藤	宏隆	助教

名古屋工業大学大学院 工学研究科 情報工学専攻

2019 年度入学 31414094 番

名前 長縄 龍風

目次

第 1 章	はじめに	1
第 2 章	データマイニング技術	3
2.1	相関分析	3
2.2	主成分分析	3
2.3	対応分析	4
2.4	決定木分析	5
2.5	クラスタリング分析	6
第 3 章	データの概要と変換	7
3.1	データの概要	7
3.1.1	通学状況に関するデータ	7
3.1.2	目的意識に関するデータ	9
3.2	単純集計	9
3.3	データの変換	12
3.3.1	データの数値化	12
3.3.2	新しい変数の追加	12
第 4 章	分析結果	14
4.1	通学状況に関する分析	14
4.1.1	相関分析	14
4.1.2	主成分分析	16
4.1.3	対応分析	19
4.1.4	クラスタリング分析	25
4.1.5	決定木分析	28
4.1.6	通学状況に関する分析のまとめ	28
4.2	目的意識に関する分析	29
4.2.1	相関分析	29
4.2.2	主成分分析	33
4.2.3	対応分析	36
4.2.4	決定木分析	42
4.2.5	クラスタリング分析	47
4.2.6	目的意識に関する分析のまとめ	50
4.3	成績以外の関連性の分析	51
4.3.1	相関分析	51
4.3.2	対応分析	53

4.3.3 成績以外の関連性の分析のまとめ	60
第5章 むすび	61
謝辞	62
参考文献	63
発表論文リスト	66
付録 分析ソフトについて	67

第1章 はじめに

近年，ソーシャルネットワークサービスなどの情報通信技術の普及と発達により，ビッグデータと呼ばれる大規模データが様々な分野で蓄積されている．これらを有効的に活用するための技術として，データを分析し有用な情報や傾向を見つけ出すデータマイニングの技術が注目を集めている．実際に商業や医療の分野では，顧客情報や臨床データなどに対してデータマイニング技術を用いることで，利益の向上や患者の治療に役立てられている．具体的には，売上データから顧客の需要の傾向を発見し，その時期のニーズに沿った商品提供につなげる試みや，既存薬の効果から新薬の副作用を予測する研究 [1] などがあげられる．その他にも，コミュニケーションパターンの解析 [2] やロボットの行動制御 [3][4] など，データマイニング技術は幅広い分野で応用されている．

教育の分野でも，データマイニング技術を用いることで学生の修学傾向を読み取り，その情報を活かして修学指導を行うという提案がされている．過去の研究では，学生の成績と授業アンケートから，成績と学生の行動の関係，成績と授業の進行の仕方との関係などを調査したもの [5] や，留学や退学をする学生を調査・分析し，事前にこれらの学生を予測する研究 [6] などがある．その他にも，授業評価アンケート結果を教員へより細かくフィードバックするシステムを開発した研究 [7] や，入試選抜区分ごとの学生の特徴を追った研究 [8]，理系学生と文系学生の違いを調査した研究 [9]，特定の教科に対する学生の意識を明らかにした研究 [10][11] など，様々な形で教育に対する支援の提案が行われている．

ところで，近年多くの大学では教員ひとりが受け持つ学生の数が多くなり，教員の修学指導における負担が大きくなっているという問題 [12] がある．この問題は教育支援の研究が注目を集めている要因のひとつでもある．これにより，学生ひとりひとりに対して十分な指導が行えなかったり，誤った方針の指導をしてしまうリス

クが発生する。すると、授業内容についていけず大学を留年・退学してしまう学生へのサポートが間に合わない場合や、行き違いにより学生と教師間の不和を生む原因となってしまう可能性がある。このような事態を避けるためにも、学生の成績の傾向を把握し、個々の学生に沿った適切な指導をおこなうことが重要である。

そこで本研究では、データマイニングの技術を用いて学生のデータを分析し、成績に関する傾向を把握する。この分析から得られた情報を活用し、教員の指導支援をすることを目指す。特に本研究の主題として、学生の通学状況及び目的意識の2つの要素に焦点を当てて分析を行う。通学状況とは、学生の自宅から大学までの距離や通学にかかる時間、通学に用いている交通手段の情報など、通学に関わる様々な情報である。さらに、その学生が下宿かどうかといった情報や、学生の出身高校所在地などの情報も含まれる。一方で目的意識とは、学生が普段の生活の上で何を重要視しているか、卒業後の進路をどう考えているかなど、学生の内面的な動機に関する情報である。

分析対象のデータとしては、名古屋工業大学の学生に対して任意で行ったアンケート調査の結果を用いる。アンケートの質問項目ごとの結果と回答者の成績及び学年を照らし合わせることで分析を行う。分析に際しては、複数のデータマイニング技術を用いることで様々な角度からのアプローチを試みる。これにより、得られた結果の正当性を高めると共に、小さな傾向を発見できる可能性を高める。また、並行して通学状況、目的意識の各変数間においても同様に分析を行う。これにより、成績以外に関しても学生の特徴を把握することを旨とする。

本論文では第2章で分析に用いたデータマイニング手法について解説する。第3章では分析に用いたデータの概要と分析のためのデータの変換等について述べる。第4章では行った分析の概要とその結果、及び結果から得られた考察について述べる。最後に、第5章では本研究のまとめと今後の課題について述べる。なお、研究を進める上で学生のデータを扱う際には、氏名、学籍番号などの個人の特定に繋がる情報は除いてある。そのため、本研究によって個人情報侵害されることがないことをここに明記する。

第2章 データマイニング技術

本研究では、分析の手法として様々なデータマイニングの技術を用いている。本章ではそれらの技術について説明する。

2.1 相関分析

相関分析 [13] とは、2つの変数間の相関、つまり関連性の強さを求める分析手法である。変数間の相関の強さは“相関係数”と呼ばれる値で表される。相関係数は-1から1の間の値を取る。相関係数が正の値を取るとき、2つの変数間には正の相関があり、一方の変数の値が増加するともう一方の変数の値も増加する傾向にある。一方で、相関係数が負の値を取るとき、2つの変数間には負の相関があり、一方の変数の値が増加するともう一方の変数の値は減少する傾向にある。また、相関係数の値の絶対値が1に近いほど相関が強く、0に近いほど相関がない。これらを踏まえて、相関係数の値から2変数間の関係を見る。

相関係数の計算方法には複数の種類があるが、本稿では最も一般的であるピアソンの積率相関係数を利用する。2つの変数 $x = \{x_1, x_2, \dots, x_n\}$, $y = \{y_1, y_2, \dots, y_n\}$ があるとすると、この2変数間の相関係数 r_{xy} は以下の式 2.1 で求められる。

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.1)$$

2.2 主成分分析

主成分分析 [18] とは、複数の変数を合成し、主成分と呼ばれる新しい変数を作り出す手法である。このとき、情報の損失が極力少なくなるように変数を圧縮する。より情報量の多い変数を作成することで、少ない変数からデータの構造を把握することが可能となる。

変数 $x = \{x_1, x_2, \dots, x_n\}$ があるとき，第1主成分 z_1 を式2.2のようにおく．

$$z_1 = a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (2.2)$$

このとき， z_1 の分散が最大となるように係数ベクトル $\mathbf{a} = (a_1, a_2, \dots, a_n)$ を変化させる．ただし， \mathbf{a} の大きさは1である．これと同様に第2主成分 z_2 を求める．ただし，このときの係数ベクトル \mathbf{b} は大きさが1であることに加えて \mathbf{a} と垂直になる必要がある．これを繰り返すことで主成分を作成する．

2.3 対応分析

対応分析 [14][15] とは，2項目の各要素間の関連性をグラフ化することでデータの理解を容易にする分析手法である．この手法では分析対象のデータとしてクロス集計表を用いる．このクロス集計表の各要素をそれぞれの特徴に基づいてグラフに配置する．このとき，同一項目内で特徴の似通った要素同士は近くに配置される．ただし，項目の違う要素同士の場合，互いの距離は意味を持たない．この場合，原点からの方向に注目し，互いの原点からの方向が同じであれば関連性がある．また，原点からの距離が遠いほど強い特徴がある．これらを踏まえ，配置した要素の位置からそれぞれの要素間の関連性を探る．

ここからは各要素をグラフに配置する手順について説明する．表2.1のような m 行 n 列のクロス集計表を対象とする．

表 2.1: m 行 n 列のクロス集計表

	X_1	X_2	\dots	X_j	\dots	X_n	合計
Y_1	f_{11}	f_{12}	\dots	f_{1j}	\dots	f_{1n}	f_{1+}
Y_2	f_{21}	f_{22}	\dots	f_{2j}	\dots	f_{2n}	f_{2+}
\vdots							
Y_i	f_{i1}	f_{i2}	\dots	f_{ij}	\dots	f_{in}	f_{i+}
\vdots							
Y_m	f_{m1}	f_{m2}	\dots	f_{mj}	\dots	f_{mn}	f_{m+}
合計	f_{+1}	f_{+2}	\dots	f_{+j}	\dots	f_{+n}	N

対応分析では，各変数間の関連性の特定のためカイ2乗統計量を利用する．

各セルのカイ2乗統計量 χ_{ij}^2 は式2.3によって求められる。

$$\chi_{ij}^2 = \frac{(f_{ij} - f_{i+}f_{+j}/N)^2}{f_{i+}f_{+j}/N} \quad (2.3)$$

ここで,

$$z_{ij} = \sqrt{\chi_{ij}^2/N} = \frac{f_{ij} - f_{i+}f_{+j}/N}{\sqrt{f_{i+}f_{+j}}} \quad (2.4)$$

となる z_{ij} を定義し, 表2.1の各セルの値を z_{ij} に変換したデータ行列 $Z_{m \times n}$ を考える。また, $Q = Z^t Z$, $Q^t = Z Z^t$ の固有ベクトルを (U, V) , f_{+j}/n , f_{i+}/n をそれぞれ要素とした対角行列を D_n , D_m とする。このとき, 各要素 X_j , Y_i はそれぞれ $(D_n^{-1/2}U, D_m^{-1/2}V)$ によって定められる値をもとにグラフに配置される。ただし2次元平面にプロットするため, 第1固有ベクトル及び第2固有ベクトルに基づいた値を利用する。

2.4 決定木分析

決定木分析 [16] とは, ある特徴に従って頻出するデータ群を発見し, その分類のルールを作成する分析手法である。説明変数と目的変数を設定し, 説明変数の条件分岐ごとの目的変数を推定した決定木と呼ばれる樹形図を作成する。この決定木より, 目的変数に定めた変数の要因やルールを可視化することができる。

決定木作成のアルゴリズムには複数の種類が存在するが, 本研究では C5.0 と呼ばれるアルゴリズムを利用する。このアルゴリズムでは分岐の基準としてエントロピーを用いている。これは以下の式2.5の通りである。

$$Info(t) = - \sum_j P(j | t) \log_2 P(j | t) \quad (2.5)$$

ただし, $P(j | t)$ はノード t 内のクラス j の割合である。分岐時, 親ノードのエントロピーと分岐後の子ノードの合計エントロピーの差を基準化した値が最大となる説明変数を分岐の変数として選択する。これを繰り返すことで決定木を作成する。

2.5 クラスタリング分析

クラスタリング分析 [17] は、データ間の類似度に基づいてデータをいくつかのグループに分割することで、データの特徴を捉える分析手法である。クラスタリング分析には階層的手法と非階層的手法があるが、ここでは非階層的手法について述べる。非階層的手法はデータの個数が多い場合の分析に適しており、代表的な手法として k-means 法がある。k-means 手法の手順としては以下の通りである。

1. k 個の初期クラスタの中心をランダムに設定する。
2. 各データと各クラスタ中心との距離を計算する。
3. 最も近いクラスタに各データを分類する。
4. 各クラスタの中心を再計算する。
5. クラスタ中心が変わらない、または一定回数に達するまで 2-4 を繰り返す。

このようにデータを k 個に分割したのち、各クラスタの中心座標や分類されたデータ数などをもとに考察をおこなう。

第3章 データの概要と変換

本章では、本研究で用いるデータの概要とその変換について述べる。データは大きく分けて成績、学年、通学状況、目的意識の4つから成る。また、これらのデータを分析に適した形へ変換する。

3.1 データの概要

本研究では、分析データとして名古屋工業大学の2年生から過年度生の446名に関するデータを用いる。彼らには学生生活に関するアンケートである学生生活実態調査に回答してもらった。なお、このアンケートは質問に対しこちらが提示した選択肢の中から回答を選ぶ選択式である。また、このアンケートを回答してもらう上で、回答内容は成績などに影響しないことを彼らには伝えている。このアンケートの内、通学状況、目的意識に関する項目の回答を利用する。また、成績を表す指標としてGPA (Grade Point Average) を用いる。GPAは0から4までの値を取る。アンケートの回答と、回答者のGPA及び学年をもとに分析を行った。各アンケート項目についての詳細を次項から述べる。

3.1.1 通学状況に関するデータ

学生生活実態調査の内、通学状況に関する項目は「通学距離」「通学時間」「通学手段第1位」「通学手段第2位」「入構手段」「住居」「同居人」「出身校」の8項目である。それぞれの項目に対する設問内容と回答の選択肢を表3.1に示す。なお、これらの回答は名古屋工業大学の所在地が愛知県名古屋市であることを考慮して作成してある。

表 3.1: 通学状況に関するアンケート項目

アンケート項目	設問内容	選択肢
通学距離	現住所は本学からどのあたりにありますか?	1km 未満 (1) 1km 以上 5km 未満 (2) 上記以外の名古屋市内 (3) 名古屋市以外の愛知県内 (4) 岐阜県または三重県 (5) その他 (6)
通学時間	通学の片道の所要時間は?	30 分未満 (1) 30 分以上 60 分未満 (2) 60 分以上 90 分未満 (3) 90 分以上 (4)
通学手段第 1 位 通学手段第 2 位 入構手段	住居から大学までの主な交通手段（主なもの 2 つ）と大学内に入構するときの手段（大学の門に入る手段）を選択してください。入構手段は、徒歩、自転車、原付・自動二輪、自動車のみから選択してください。	徒歩 (1) 自転車 (2) 原付・自動二輪 (3) 自動車 (4) JR (5) 地下鉄 (6) その他の鉄道 (7) バス (8) その他 (9)
住居	どこから通学していますか?	自宅から (2) 自宅以外から (1)
同居人	あなたは現在誰と暮らしていますか?	ひとりで暮らしている (1) 家族と同居している (2) 友人と同居している (3)
出身校	あなたの卒業した高校の所在地はどこですか?	愛知県 (1) 岐阜県 (2) 三重県 (3) 静岡県 (4) 関西 (5) 北陸 (6) 関東・甲信越 (7) 東北・北海道 (8) 中国・四国 (9) 九州・沖縄 (10) 日本国外 (11)

3.1.2 目的意識に関するデータ

学生生活実態調査の内，目的意識に関する項目は「優先順位第1位」「優先順位第2位」「優先順位第3位」「入学前進路希望」「現在の進路希望」の5項目である．それぞれの項目に対する設問内容と回答の選択肢を表3.2に示す．

表 3.2: 目的意識に関するアンケート項目

アンケート項目	設問内容	選択肢
優先順位第1位 優先順位第2位 優先順位第3位	現在の学生生活で、どの時間を優先していますか？時間の長さでなく、優先度で第1～3位を選んでください。	学業 (1) クラブ・サークル (2) 趣味・娯楽 (3) 友人との交流 (4) その他 (5)
入学前進路希望	大学入学時にどのような卒業後の進路を考えていましたか？	学部を卒業後、就職する (1) 本学の大学院博士前期課程へ進学する (2) 他大学の大学院博士前期課程へ進学する (3) 本学の大学院博士後期課程へ進学する (4) 他大学の大学院後期課程へ進学する (5) 海外の大学・大学院へ進学する (6) 未定 (-1)
現在の進路希望	現在どのような進路を考えていますか？	学部を卒業後、就職する (1) 本学大学院前期課程へ進学・修了し、就職する (2) 他大学の大学院前期課程へ進学・修了し、就職する (3) 本学の大学院博士後期課程へ進学する (している) (4) 他大学の大学院後期課程へ進学する (5) 海外の大学・大学院へ進学する (6) 未定 (-1)

3.2 単純集計

各アンケート項目及び学年について，分析の前におおまかな特徴を把握するため，単純集計を行い各要素における人数をまとめた．学年に関する集計結果を表3.3，通学距離に関する集計結果を表3.4，通学時間に関する集計結果を表3.5，通学手段第1位，通学手段第2位，入構手段の3項目に関する集計結果を表3.6，住居に関する集計結果を表3.7，同居人に関する集計結果を表3.8，出身校に関する集計結果を表

3.9, 優先順位第1位から第3位までに関する集計結果を表3.10, 入学前の進路希望に関する集計結果を表3.11, 現在の進路希望に関する集計結果を表3.12にそれぞれ示す. また, GPAは平均2.65である.

表 3.3: 学年集計

学年	人数
2年生	177
3年生	132
4年生	125
過年度生	12

表 3.4: 通学距離集計

通学距離	人数
1km未満	89
1km以上5km未満	71
上記以外の名古屋市内	46
名古屋市以外の愛知県内	190
岐阜県または三重県	45
その他	2

表 3.5: 通学時間集計

通学時間	人数
30分未満	153
30分以上60分未満	164
60分以上90分未満	104
90分以上	22

表 3.6: 通学手段第1位, 第2位, 入構手段の集計

回答	通学手段第1位 人数	通学手段第2位 人数	入構手段 人数
徒歩	33	131	292
自転車	131	92	94
原付・自動二輪	9	6	8
自動車	1	22	10
JR	87	98	5
地下鉄	68	32	5
その他の鉄道	110	27	1
バス	3	18	0
その他	1	0	4

表 3.7: 住居集計

住居	人数
自宅から	343
自宅以外から	96

表 3.8: 同居人集計

同居人	人数
ひとり暮らし	116
家族と同居	325
友人と同居	1

表 3.9: 出身校集計

出身校	人数
愛知県	290
岐阜県	46
三重県	28
静岡県	22
関西	21
北陸	6
関東・甲信越	14
東北・北海道	4
中国・四国	7
九州・沖縄	3
日本国外	5

表 3.10: 優先順位の集計

回答	優先順位第1位 人数	優先順位第2位 人数	優先順位第3位 人数
学業	213	92	84
クラブ・サークル	71	72	54
趣味・娯楽	97	138	114
友人との交流	44	108	125
その他	14	29	50

表 3.11: 入学前進路希望の集計

入学前進路希望	人数
学部を卒業後, 就職する	98
本学の大学院博士前期課程へ進学する	251
他大学の大学院博士前期課程へ進学する	43
本学の大学院博士後期課程へ進学する	19
他大学の大学院後期課程へ進学する	2
海外の大学・大学院へ進学する	2
未定	28

表 3.12: 現在の進路希望の集計

現在の進路希望	人数
学部を卒業後, 就職する	106
本学の大学院博士前期課程へ進学・修了し, 就職する	264
他大学の大学院博士前期課程へ進学・修了し, 就職する	13
本学の大学院博士後期課程へ進学する (している)	13
他大学の大学院後期課程へ進学する	4
海外の大学・大学院へ進学する	3
未定	35

3.3 データの変換

本研究では, 3.1 節で説明した学生のデータを用いて分析を行う。しかし, このままのデータ形式では分析は困難であるため, 2 章で説明した各分析手法に合わせてデータを変換する。また, より分析の精度を高めるために, いくつかの変更を行う。次項からこれらの詳細について記す。

3.3.1 データの数値化

2 章で説明した分析の内, 相関分析, 主成分分析, クラスタリング分析ではデータは数値であるほうが好ましい。そこで, 各アンケート項目ごとの回答選択肢それぞれに数値を当てはめる。このとき, 各選択肢の連続性を出来るだけ考慮する。この当てはめた数値は表 3.1 及び表 3.2 の各選択肢の後ろの () 内に示す。通学状況に関する選択肢は, 主に大学から離れるほど数値が大きくなるよう設定した。また, 目的意識の内, 進路希望の 2 項目では進路における勉学への取組意識が高いほど数値が大きくなるよう設定した。優先順位の 3 項目では選択肢に連続性を見出しにくかったが, 暫定的に数値を設定した。また, 未回答の項目の値は 0 とした。

3.3.2 新しい変数の追加

3.3.1 項で述べた通り, 目的意識における項目「優先順位」では, 選択肢に連続的な意味合いが見出しにくい。そこで, 選択肢の中でも「学業」に注目し, その生徒

がどれだけ学業を優先しているかを判断するための指標として新しい変数「学業意識レベル」を作成する。学業意識レベルは優先順位第1位から第3位までのなかで「学業」と回答した順位をもとに値を設定する。この変数の定義は以下の式3.1の通りである。

$$\text{学業意識レベル} = \begin{cases} 3 & (\text{優先順位第1位} = \text{「学業」}) \\ 2 & (\text{優先順位第2位} = \text{「学業」}) \\ 1 & (\text{優先順位第3位} = \text{「学業」}) \\ 0 & (\text{その他}) \end{cases} \quad (3.1)$$

第4章 分析結果

3章のデータ群に対し、2章で述べたデータマイニング手法を用いて分析を行った。本章では、その分析の概要と結果、及びその考察について述べる。なお、このデータの管理及び変換には Microsoft 社の Excel2016[19] を、決定木分析にはフリーの機械学習ソフトウェアである Weka[20] を、その他の分析にはフリーの統計解析向けソフトウェアである R[21] を用いている。

4.1 通学状況に関する分析

通学状況と成績との関連性について分析を行った。本節では、その分析の概要と結果、考察について述べる。

4.1.1 相関分析

GPA と学年間、及び GPA と通学状況に関する各アンケート項目間に対してそれぞれ相関分析を行った。分析に際しては、3.3.1 項で述べたデータの数値化を行っている。これにより得られた相関係数を表 4.1 に示す。また、各相関係数に対して t 検定を行い、それぞれの相関の有意性について確認した。検定によって有意であると判断された係数は、表において赤く表示している。

この表から、通学状況に関するすべての項目において GPA との相関係数が小さいことがわかる。よって通学状況と GPA の間には相関関係はないと考えられる。また、学年と GPA には有意な相関がみられた。これは小さな負の相関であるため、学年が増えるにつれて GPA はわずかに低くなる傾向がある。ただしこれは、過年度生のデータの GPA が全体的に低いことが影響しているためだと考えられる。これを確認するため、過年度生のデータを除いた状態で GPA と学年の相関分析を行ったところ、有意な相関はみられなかった。

表 4.1: 通学状況と GPA の相関係数

変数	GPA との相関係数
学年	-0.1447
通学距離	0.0333
通学時間	0.0011
通学手段第 1 位	0.0348
通学手段第 2 位	-0.0108
入構手段	0.0242
住居	0.0757
同居人	0.0303
出身校	-0.0437

さらに、学年別の傾向を探るため、データを学年ごとの 4 つに分割しそれぞれのデータに対して同様に GPA との相関分析を行った。この結果を表 4.2 に示す。前回の分析と同様に、検定により有意であると判断された係数は赤く表示している。

表 4.2: 通学状況と GPA の学年別相関係数

変数	GPA との相関係数			
	2 年生	3 年生	4 年生	過年度生
通学距離	-0.065	0.057	-0.075	0.346
通学時間	-0.017	-0.034	0.095	-0.089
通学手段第 1 位	0.066	0.019	0.041	-0.552
通学手段第 2 位	0.014	-0.138	0.062	-0.263
入構手段	0.146	-0.184	0.096	-0.211
住居	0.1135	-0.0572	0.1691	-0.3463
同居人	0.030	-0.011	0.115	-0.283
出身校	-0.092	-0.032	-0.008	-0.366

この表より、学年別の分析でもほとんどの項目で GPA との相関係数が小さいことがわかる。過年度生の相関係数は一見大きいように見えるが、これは過年度生のデータが少ないことが原因であり、有意な相関はない。ただし、3 年生の入構手段において有意な相関がみられた。これは小さな負の相関であるため、入構手段の値が大きくなるほど GPA が低くなるといえる。

この関連性について確認するため、入構手段ごとの GPA を箱ひげ図で表す。ただし、回答者数が少なく、回答として適切でない「JR」「地下鉄」「その他の鉄道」「バ

ス」「その他」の項目は除く。これを図 4.1 に示す。この図から、遠距離向きの交通手段になるにつれて GPA が全体的に低くなっていく傾向がみられる。よって、入構手段と GPA は関連性がある可能性がある。

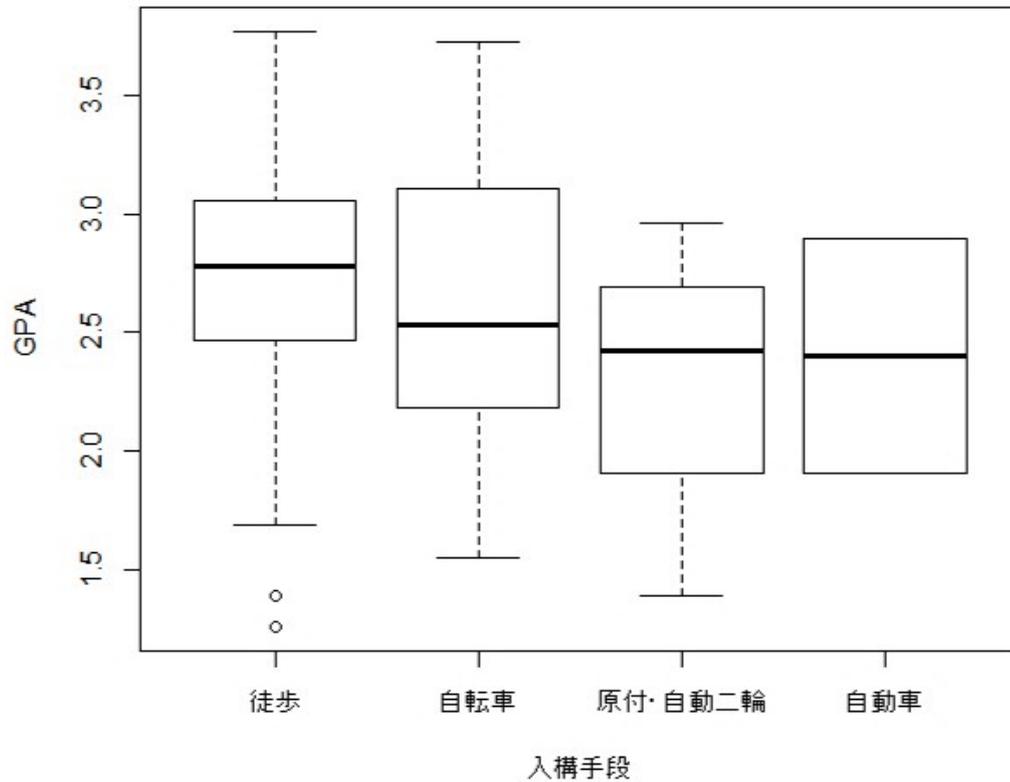


図 4.1: 3 学年の入構手段と GPA に関する箱ひげ図

4.1.2 主成分分析

通学状況に関するアンケート項目に対して主成分分析を行った。分析に際しては、3.3.1 項で述べたデータの数値化を行っている。このとき、各主成分の寄与率及び累積寄与率について表 4.3 に示す。なお寄与率とは、ある主成分がデータ全体においてどの程度意味を持つかを表した指標であり、累積寄与率は第 1 主成分からの寄与率の合算値である。また、PC とは主成分 (Principal Component) の略称である。

この表より第 4 主成分まででデータの 8 割以上を再現出来ていることがわかる。そこで、以後の分析では第 1 主成分から第 4 主成分を分析対象とする。

表 4.3: 各主成分の寄与率 (通学状況)

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
寄与率	0.558	0.126	0.094	0.080	0.051	0.039	0.032	0.017
累積寄与率	0.558	0.685	0.779	0.859	0.910	0.950	0.982	1.000

また、各主成分に対してそれぞれの元変数がどれだけ影響しているかを表した値 (主成分負荷量) をグラフにまとめた。第1及び第2主成分の主成分負荷量を図 4.2 に示す。横軸が第1主成分、縦軸が第2主成分を表す。赤い円は半径1.0の円であり、各変数の負荷量の大きさを測るための目安である。

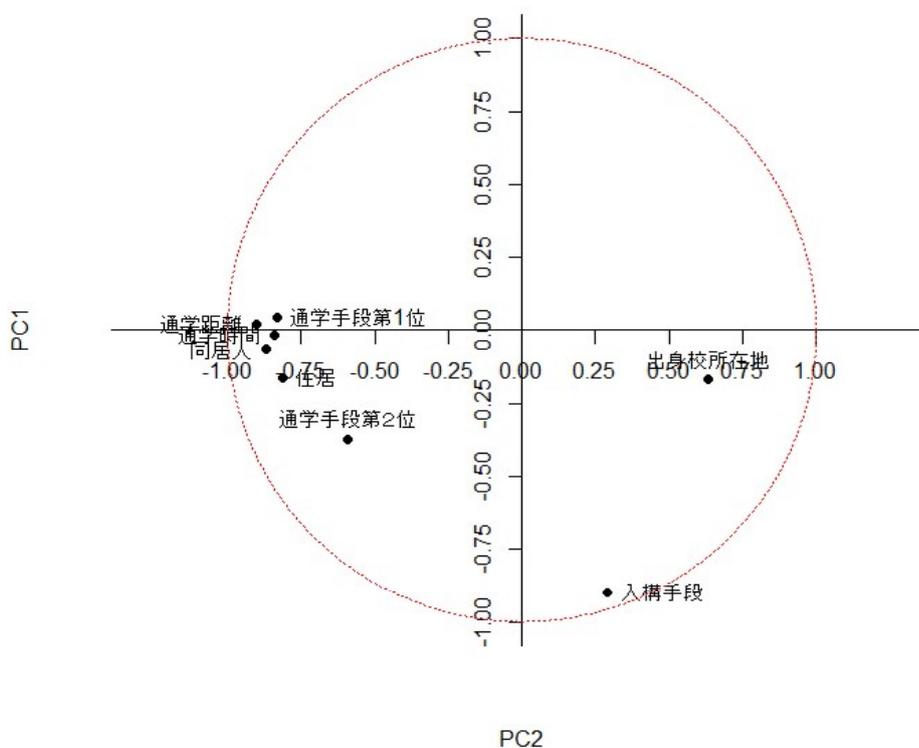


図 4.2: 第1第2主成分負荷量 (通学状況)

この図より、第1主成分に大きく影響しているのは通学距離や通学時間、通学手段などであり、第2主成分に大きく影響しているのは入構手段である。よって、第1主成分は主に通学距離、第2主成分は主に入構手段を表した成分であるといえる。

同様に、第3及び第4主成分の主成分負荷量を図 4.3 に示す。横軸が第3主成分、縦軸が第4主成分を表す。

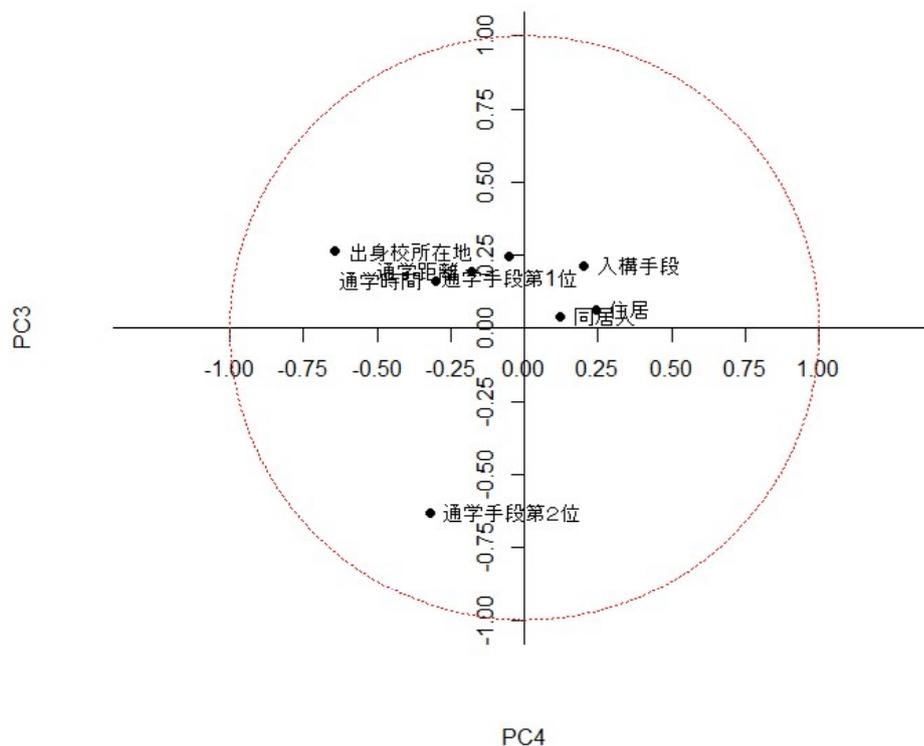


図 4.3: 第3 第4 主成分負荷量 (通学状況)

この図より、第3主成分に大きく影響しているのは出身校であり、第4主成分に大きく影響しているのは通学手段第2位である。よって、第3主成分は主に出身校、第4主成分は主に通学手段第2位を表した成分であるといえる。

その後、これらの主成分と成績との関連性をみるため、各主成分とGPAとの相関係数を計算する。この結果を表4.4に示す。

表 4.4: 主成分とGPAの相関 (通学状況)

変数	GPAとの相関係数
PC1	-0.0363
PC2	-0.0221
PC3	0.0664
PC4	0.0337

この表より主成分とGPAとの間に有意な相関はみられなかった。つまり、通学状況と成績は関連性がほとんどないと考えられる。

4.1.3 対応分析

GPA と通学状況に関する各アンケート項目間において対応分析を行う。まず、分析のため各変数間でクロス集計表を作成した。ただし、クロス集計表の作成にあたり、GPA を連続データでなくカテゴリデータにする必要がある。そこで、GPA を0.5 間隔で「1.0 未満」から「3.5 以上」までの7つのグループに分割した。その後、作成したクロス集計表に対して対応分析を行った。なお、住居及び同居人は回答の種類が少ないので今回の分析対象からは除外する。通学距離に対する分析結果を図4.4に示す。

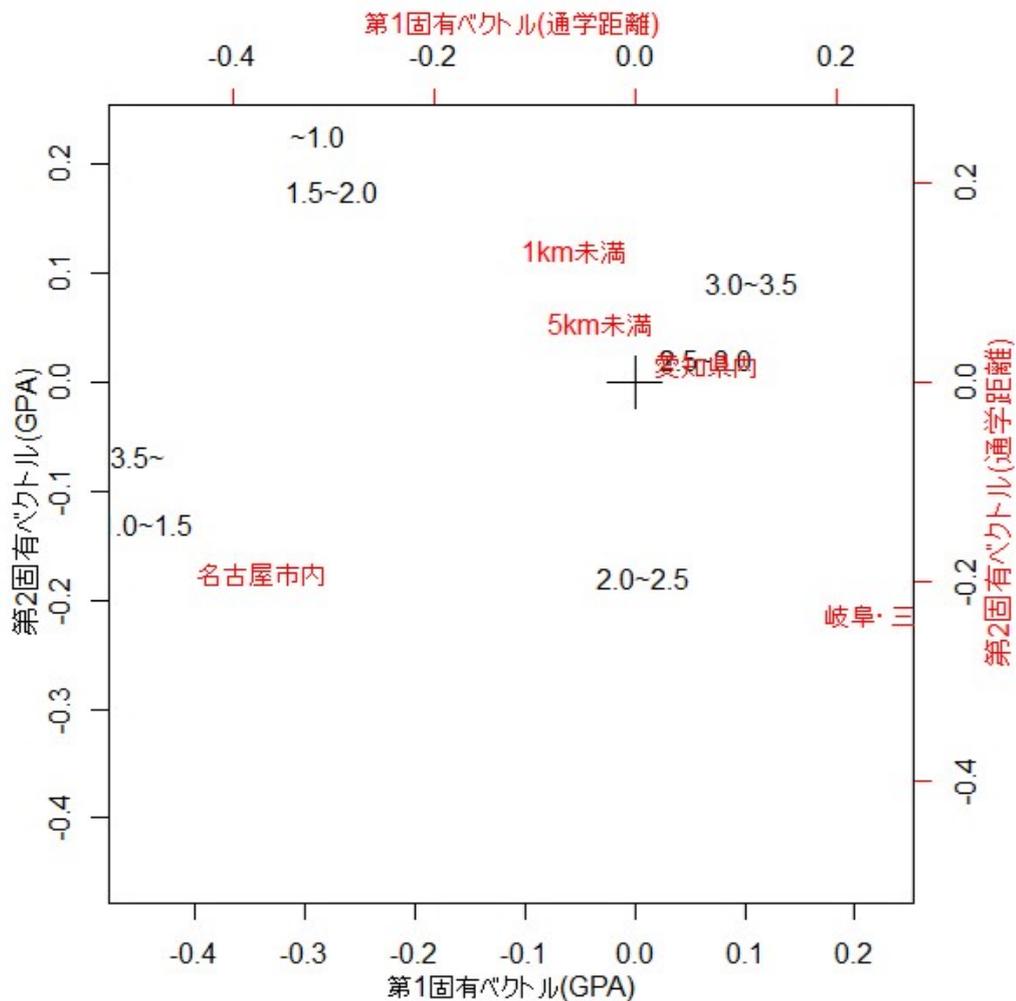


図 4.4: GPA と通学距離の対応分析

赤のラベルが各アンケート項目，黒のラベルがGPAを示す。また，「+」は原点を表す。また，データ数が少ない変数はノイズになってしまうため，回答者数が5人未満である項目は分析から除いてある。この図から，「名古屋市内」と「1.0~1.5」及び「3.5~」において，原点からの方向が等しく，どの要素も原点からの距離が大きい。つまり，これらの要素間には関連性がある。したがって，これらの成績を修める学生は大学から少し離れた名古屋市内に住んでいる傾向があるといえる。

次に，通学時間に対する分析結果を図4.5に示す。

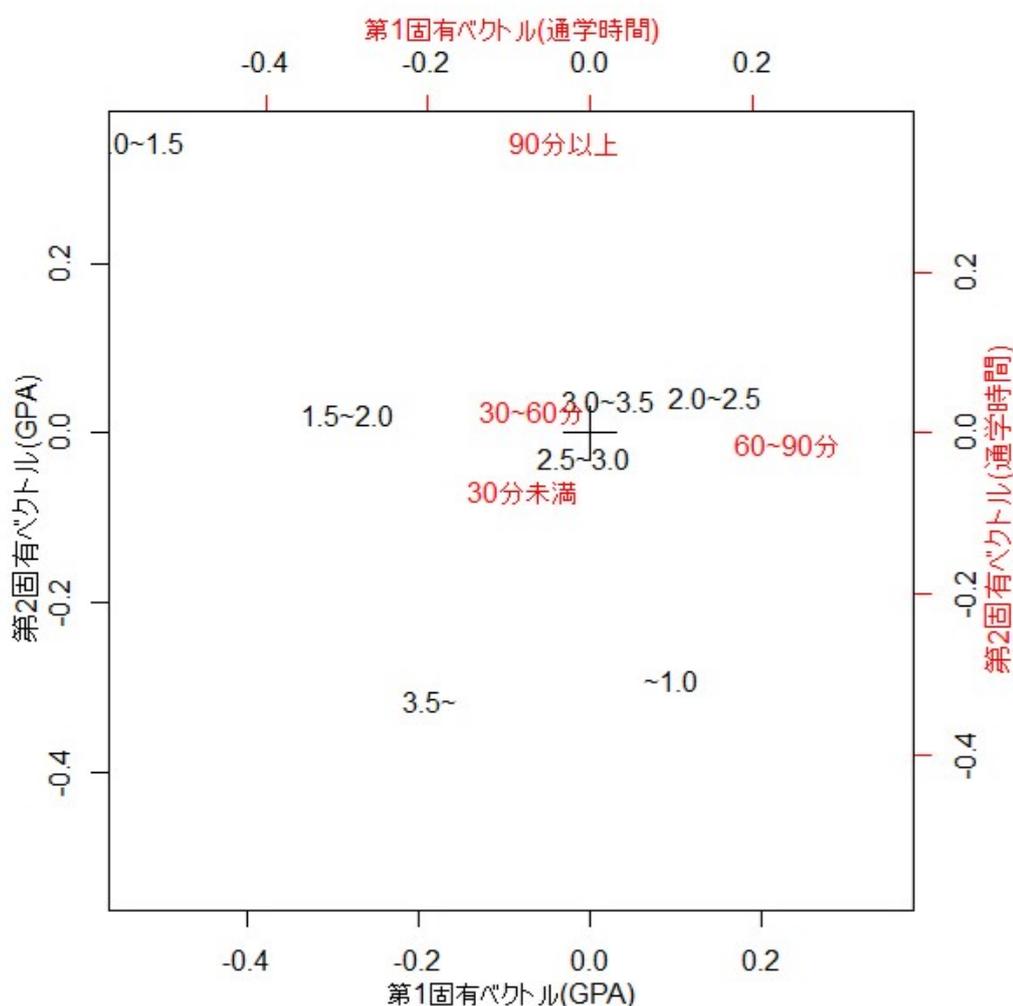


図 4.5: GPA と通学時間の対応分析

この図から，通学時間と GPA の要素間に大きな特徴はみられない。よって，通学時間と成績には関連性がほとんどないといえる。

次に、通学手段第1位に対する分析結果を図4.6に示す。

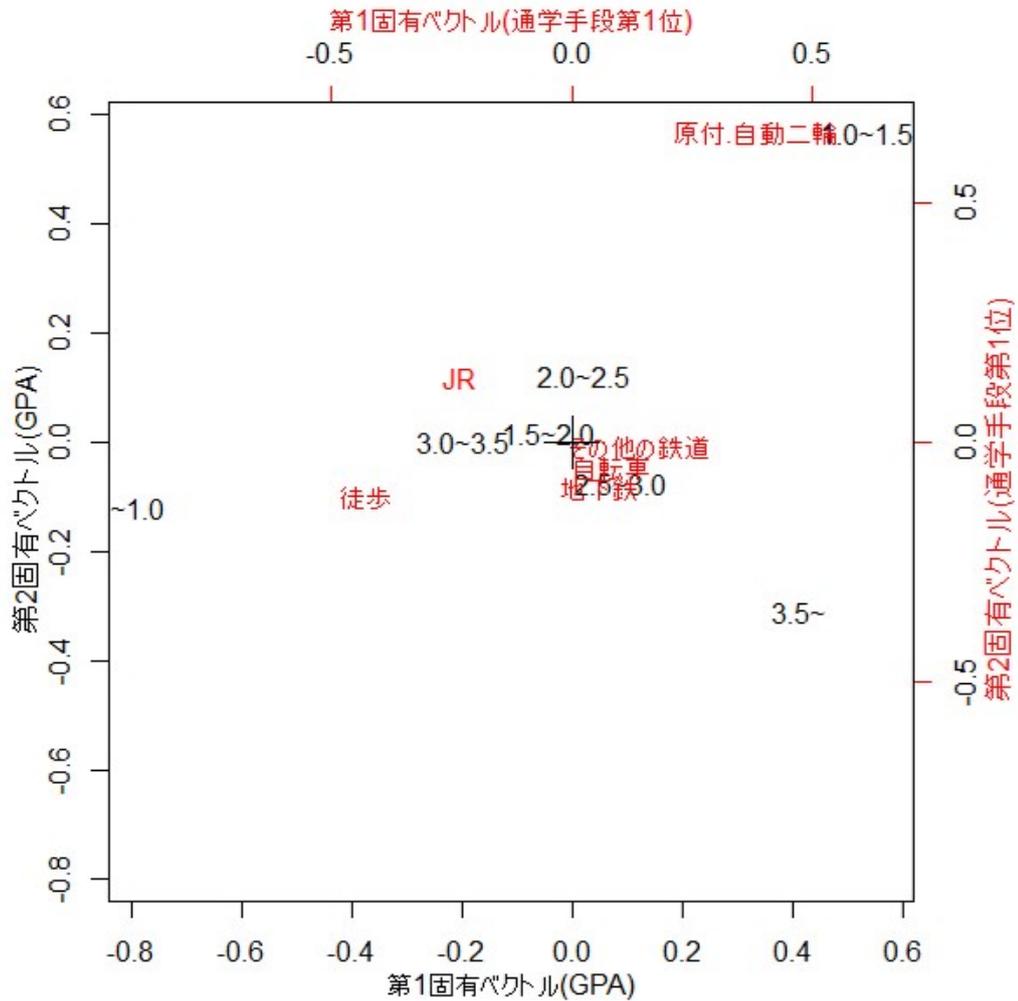


図 4.6: GPA と通学手段第1位の対応分析

この図から、「原付・自動二輪」と「1.0~1.5」について、原点からの方向が等しく、どちらも原点からの距離が大きい。つまり、この2つの要素間には関連性がある。したがって、原付・自動二輪で通学している学生には低成績の学生が多い傾向にある。ただし、3.2節にも示した通り、原付・自動二輪を使用している学生は少ないため、この傾向は慎重に判断する必要がある。

次に、通学手段第2位に対する分析結果を図4.7に示す。

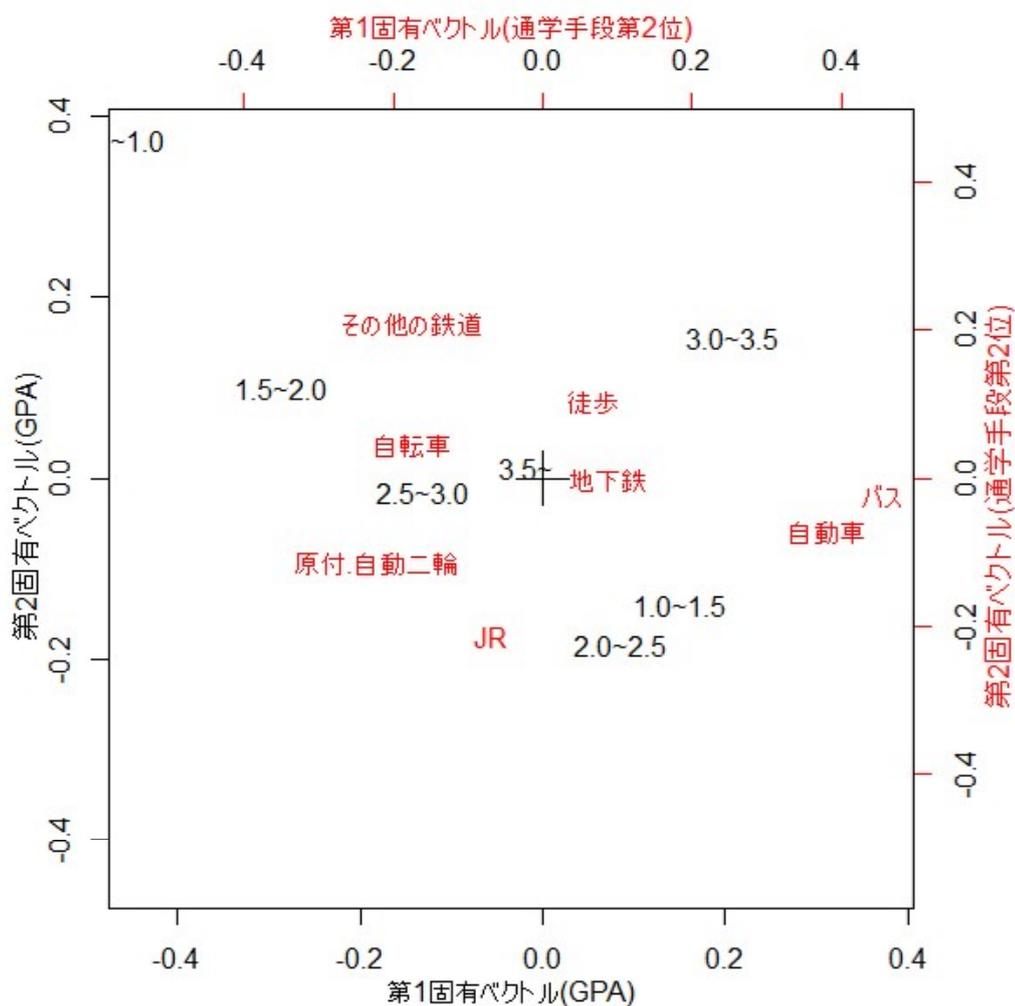


図 4.7: GPA と通学手段第2位の対応分析

この図から、通学手段第2位とGPAの要素間に大きな特徴はみられない。よって、2番目に多く用いる通学手段と成績には関連性がほとんどないといえる。

次に、入構手段に対する分析結果を図4.8に示す。

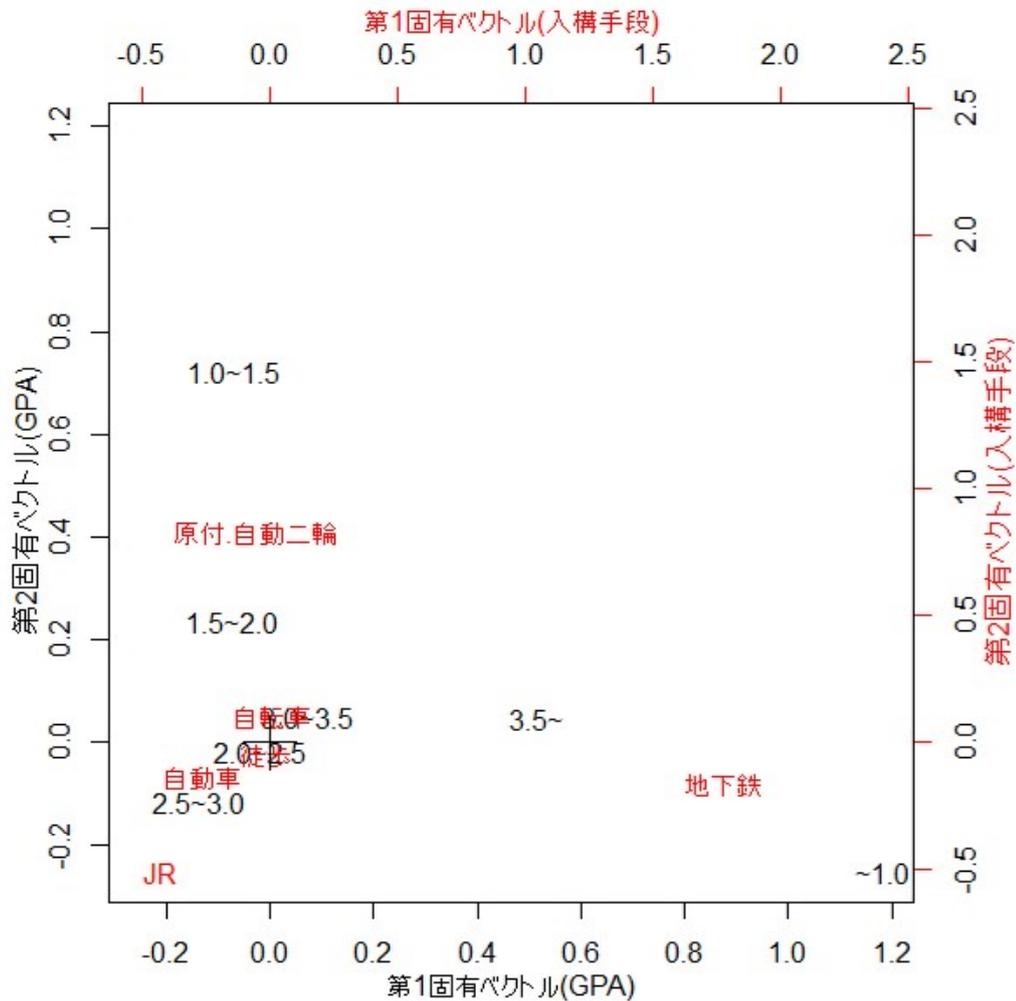


図 4.8: GPA と入構手段の対応分析

この図から、「原付・自動二輪」と「1.0~1.5」及び「1.5~2.0」、また「地下鉄」と「~1.0」及び「3.5~」において、原点からの方向がそれぞれ等しい。さらに、「原付・自動二輪」「1.0~1.5」「地下鉄」「~1.0」については原点からの距離も大きい。つまりこれらの要素間、特に「原付・自動二輪」と「1.0~1.5」、「地下鉄」と「~1.0」には関連性がある。通学手段第1位の結果からも「原付・自動二輪」と「1.0~1.5」の関連性が表れていたことから、この2項目には強い関連性があると考えられる。また、入構手段において「地下鉄」は不適切な回答である。つまり、不適切な回答をした学生は成績が悪い傾向にあると考えられる。

最後に、出身校に対する分析結果を図 4.9 に示す。

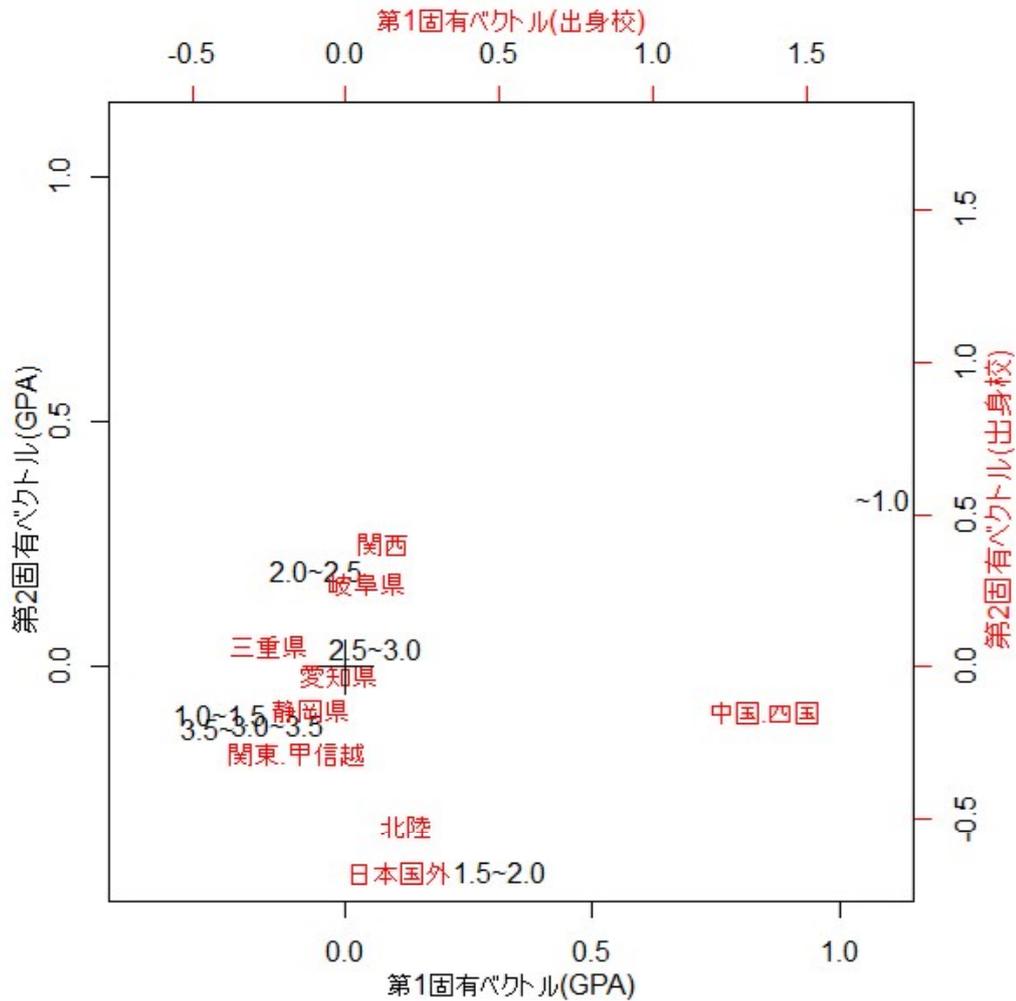


図 4.9: GPA と出身校の対応分析

この図から、「日本国外」「北陸」と「1.5~2.0」、また「岐阜県」「関西」と「2.0~2.5」において、原点からの距離はあまり大きくないが、原点からの距離が等しい。つまり、これらの要素間には小さな関連性がある。したがって、国外、北陸出身の学生は平均より少し低い成績を取る傾向にあり、岐阜、関西出身の学生は平均的な成績を修める傾向にあるといえる。

4.1.4 クラスタリング分析

GPA, 学年, 通学状況に関するアンケート項目に対して, 非階層クラスタリングである k-means 法を用いてクラスタリング分析をおこなった. また, 最適なクラスタ分割数を計算するため, k-means 分析におけるクラスタ数決定手法のひとつであるエルボー法 [22] を用いた. その結果からクラスタの分割数は2つに設定した. なお, 分析に際しては3.3.1項で述べたデータの数値化を行っている. 表4.5に得られた各クラスタの中心点の座標を示す.

表 4.5: 各クラスタの中心 (通学状況)

変数	クラスタ1の中心	クラスタ2の中心
GPA	2.64	2.65
学年	3.01	2.90
通学距離	1.48	3.84
通学時間	1.05	2.43
通学手段第1位	1.84	5.69
通学手段第2位	1.47	4.08
入構手段	1.85	1.24
住居	1.27	1.99
同居人	1.19	1.99
出身校	4.04	1.27

この表から, 各クラスタの全体像をとらえることができる. これらの項目の内, 通学距離, 通学時間, 通学手段, 出身校では2つのクラスタ間で違いが大きく出ている. また, 選択肢の少なさを考慮すると, 住居, 同居人においてもクラスタ間の違いが大きいといえる. 出身校以外のこれらの項目ではいずれもクラスタ2のほうが大きくなっている. つまりクラスタ2には大学からの距離が遠い学生が集まっているとわかる. また, 出身校はクラスタ1のほうが大きい, これは遠方の出身の学生が下宿して大学近郊に住んでいることが多いことが原因だと考えられる. これらのことから, 名古屋工業大学生は通学距離によって2つのグループに分けられることがわかる. 特に, 各項目の値を考慮すると, おもに大学がある名古屋市内に住んでいるかどうかで分けられるといえる. 加えて, クラスタ1に分類された学生は147人, クラスタ2に分類された学生は299人である. つまり, 大学近隣に住んでいる学

生のほうが少ないことがわかった。ただし、どちらのクラスでも GPA はほぼ同じである。したがって、大学からの距離は成績にはあまり影響しないと考えられる。

加えて、データとクラスをグラフにまとめることで、分析結果を視覚的にわかりやすくする。ただし、このデータは多次元変数であるため、2次元平面にそのままプロットすることは困難である。そこで、主成分分析によって GPA、学年、通学状況のすべての変数を合成し、寄与率の高い第1主成分及び第2主成分を軸に使用する。このグラフを図4.10に示す。青い円がクラス1、赤い円がクラス2、緑の数字が各データを表す。また、クラス1に分類される学生は丸、クラス2に分類される学生は三角が数字の下に記されている。

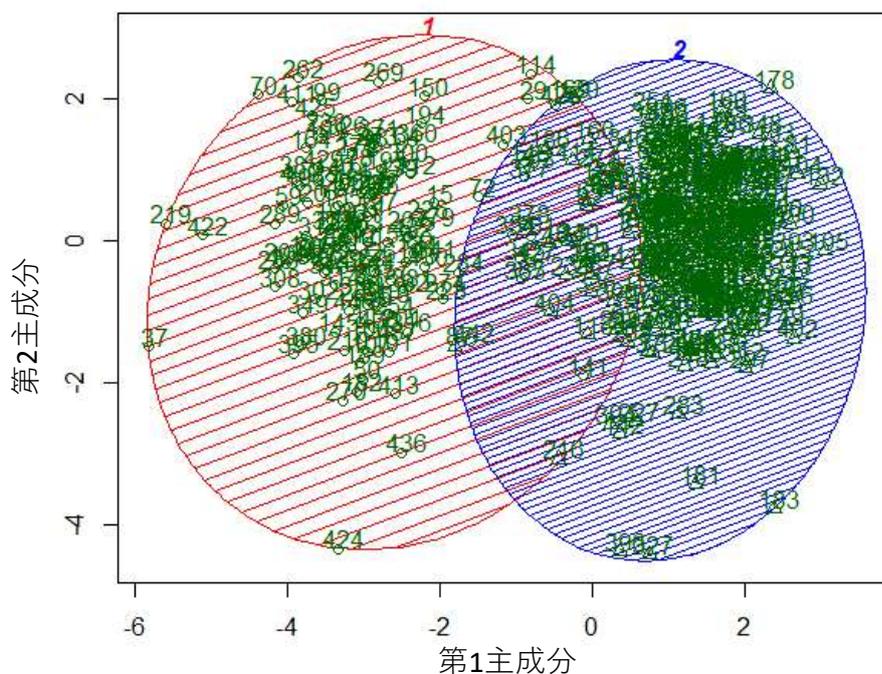


図 4.10: クラスタリング分析 (通学状況)

この図から、2つのクラスは主に第1主成分において2つに分かれていることがわかる。また、各学生においても第1主成分において2つの塊に分かれていることが視覚的にも見て取れる。また、第2主成分に注目すると、どちらのクラスにおいてもデータが全体的にばらけていることがわかる。

ここで、各主成分の意味を知るため、2つの主成分の主成分負荷量について図 4.11 に示す。

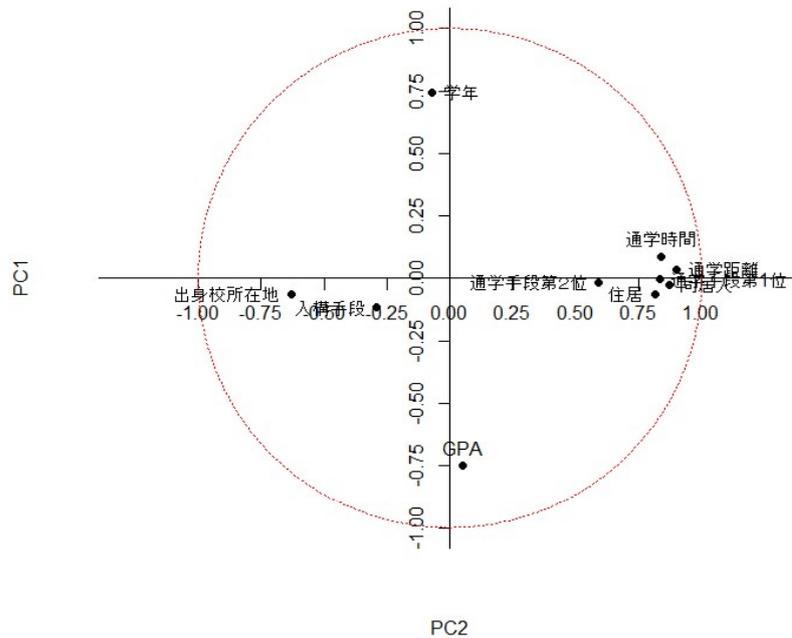


図 4.11: クラスタリング分析における第1第2主成分負荷量（通学状況）

この図から、第1主成分に大きく影響しているのは通学距離や通学時間、通学手段などの通学状況に関する要素すべてであり、第2主成分に大きく影響しているのはGPAと学年である。つまり、第1主成分は通学状況についてまとめた、大学からの遠さを表した成分であり、第2主成分はGPAについて表した成分であることがわかる。この結果と図 4.10 の結果を照らし合わせると、名古屋工業大学の学生は通学状況において2つのグループに分類でき、どちらのグループにも成績の良い学生も悪い学生も一定数存在していることがわかる。

さらに、この結果について確認するため、各クラスタにおけるGPAを箱ひげ図で表したグラフを図 4.12 に示す。

この図から、最大値、中央値、最小値など、すべての値がクラスタ間においてほとんど変わらないことがわかる。したがって、通学状況によって2つに分かれたどちらのクラスタにおいても、成績の良い学生も悪い学生も両方存在しているといえる。

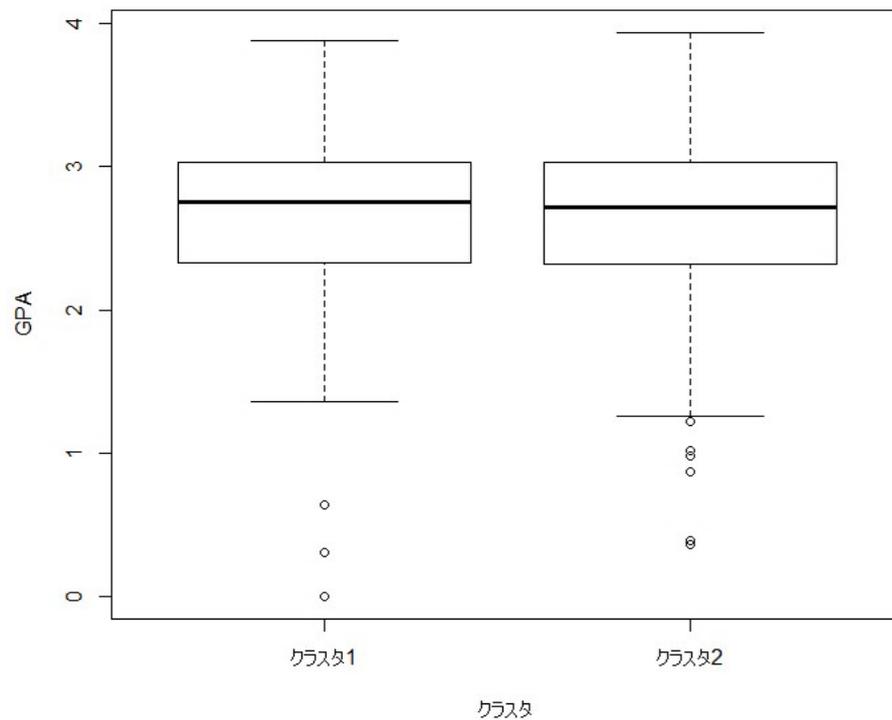


図 4.12: 各クラスタの GPA に関する箱ひげ図 (通学状況)

4.1.5 決定木分析

GPA, 学年, 通学状況に関するアンケート項目に対して決定木分析を行った。GPA を目的変数, 他の変数を説明変数とした。分析の結果, GPA を予測するに際し, 通学状況の変数はすべて不必要であった。この結果から, 通学状況と成績は関連性が極めて小さいといえる。

4.1.6 通学状況に関する分析のまとめ

これまでの結果から, 通学状況は成績とほとんど関連性がないといえる。ただし, 一部の分析においていくつかの小さな関連性がみられた。特に, 通学手段, 入構手段として原付・自動二輪を用いている学生は成績が低い傾向がある可能性がある。ただし, 原付・自動二輪を利用している学生の人数が少ないため, この傾向は慎重に判断する必要がある。

4.2 目的意識に関する分析

目的意識と成績との関連性について分析を行った。本節では、その分析の概要と結果、考察について述べる。

4.2.1 相関分析

4.1.1 項と同様に、GPA と目的意識に関する各アンケート項目間に対して相関分析を行った。また、分析対象として3.3.2 項で述べた新しい変数である学業意識レベルも採用している。これにより得られた相関係数を表4.6に示す。4.1.1 項と同様に、検定によって有意であると判断された係数は赤く表示している。

表 4.6: 目的意識と GPA の相関係数

変数	GPA との相関係数
優先順位第1位	-0.1662
優先順位第2位	0.0522
優先順位第3位	0.00002
入学前進路希望	0.0652
現在の進路希望	0.2133
学業意識レベル	0.2183

この表から、優先順位第1位、現在の進路希望、学業意識レベルにおいて有意な相関がみられた。優先順位第1位は負の相関である。つまり、優先順位第1位の値が大きくなるほどGPAが低くなるといえる。また、現在の進路希望及び学業意識レベルは正の相関である。つまり、これらの値が大きくなるほどGPAは高くなるといえる。さらに、これらの相関係数は優先順位第1位や4.1.1 項での学年の相関係数よりも大きい。したがって、より強い相関関係があるといえる。

これらの関連性について確認するため、それぞれの箱ひげ図を作成する。優先順位第1位に関するグラフを図4.13に示す。

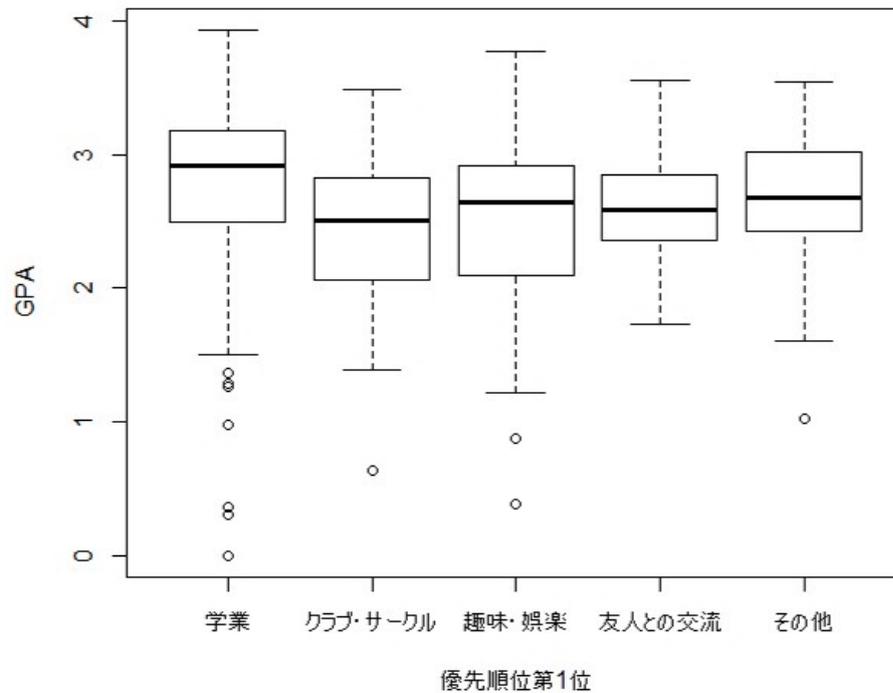


図 4.13: 優先順位第1位と GPA に関する箱ひげ図

この図より、中央値において学業とそれ以外の項目の間に大きな差があることがわかる。また、最大値についても学業の値が最も大きくなっている。したがって、学業を選んだ学生は他の項目を選んだ学生より GPA が高くなる傾向がみられた。一方で、学業以外の4項目間では目立った差はみられなかった。

次に、現在の進路希望のグラフを図 4.14 に示す。

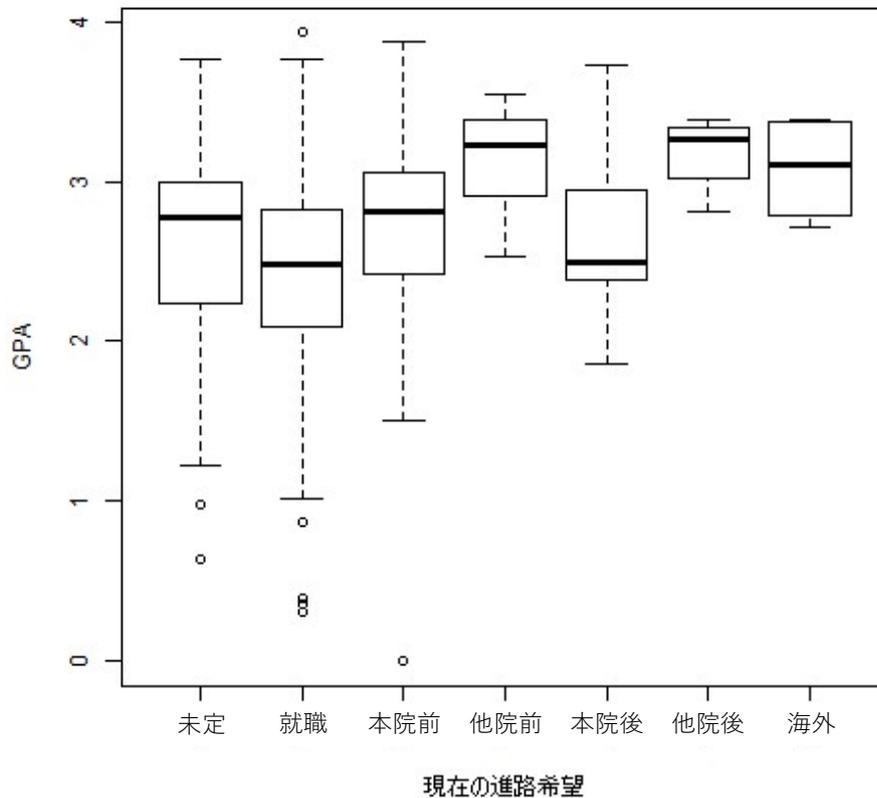


図 4.14: 現在の進路希望と GPA に関する箱ひげ図

この図より、全体的に右上がりに GPA が高くなっており、相関関係が見て取れる。つまり、本学大学院前期課程を志望する学生よりも他大学前期課程を志望する学生の方が GPA が高い。ただし、未定を選んだ学生の GPA は就職を選んだ学生のものより全体的に高く、本学大学院前期課程を選んだ学生よりやや低い。つまり、未定を選んだ学生は就職か本学大学院前期課程かで悩んでいる生徒が多いと考えられる。また、本学大学院後期課程を選んだ学生の GPA が低くなっている。これより、成績のそこまで高くない学生でも本学の後期課程を志望しているということがわかる。

最後に、学業意識レベルのグラフを図 4.15 に示す。

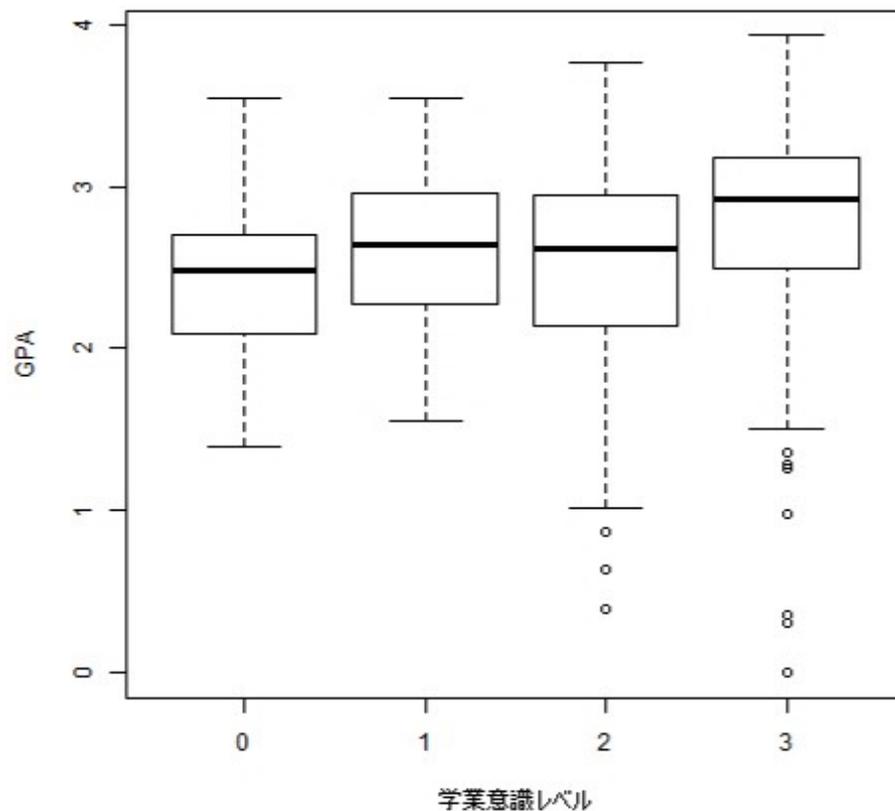


図 4.15: 学業意識レベルと GPA に関する箱ひげ図

この図より、学業意識レベルが大きくなるにつれ GPA も全体的に高くなっており、相関関係が見て取れる。ただし、GPA が極端に低い学生も学業意識レベルが高い傾向があるため、この点には注意が必要である。

さらに、学年別の傾向を探るため、データを学年ごとの4つに分割しそれぞれのデータに対して同様に GPA との相関分析を行った。この結果を表 4.7 に示す。これまでと同様に、検定により有意であると判断された係数は赤く表示している。

表 4.7: 目的意識と GPA の学年別相関係数

変数	GPA との相関係数			
	2 年性	3 年生	4 年生	過年度生
優先順位第 1 位	-0.0498	-0.2942	-0.2743	-0.1166
優先順位第 2 位	-0.0089	0.0346	0.0745	0.4601
優先順位第 3 位	0.0789	-0.0252	0.0472	-0.3288
入学前進路希望	0.0509	0.0886	-0.2004	-0.2503
現在の進路希望	0.1692	0.2139	0.3192	0.6904
学業意識レベル	0.1971	0.3807	0.2741	-0.1123

この表から、現在の進路希望はどの学年においても有意な相関がみられることがわかる。つまり、現在の進路希望は GPA と強い関連性があるといえる。同様に、学業意識レベルに関しても過年度生以外のデータで相関がみられたため、GPA との関連性が強い。また、優先順位第 1 位では 2 年生のデータで相関がほとんどみられなかったことから、入学後数年以内では優先意識は少し影響しにくいと考えられる。さらに、全体での分析ではみられなかったが 4 年生のデータにおいて入学前進路希望と GPA との有意な相関がみられた。つまり、現在の進路希望ほど密接に関連してはいないが、入学前の進路希望も成績に影響がある可能性がある。

4.2.2 主成分分析

目的意識に関するアンケート項目に対して主成分分析を行った。4.1.2 項と同様に、分析に際してはデータの数値化を行っている。このとき、各主成分の寄与率及び累積寄与率について表 4.8 に示す。この表より第 4 主成分まででデータの 8 割以上を再現出来ていることがわかる。そこで、以後の分析では第 1 主成分から第 4 主成分を対象とする。

表 4.8: 各主成分の寄与率 (目的意識)

主成分	PC1	PC2	PC3	PC4	PC5	PC6
寄与率	0.312	0.244	0.189	0.120	0.0837	0.0510
累積寄与率	0.312	0.556	0.745	0.865	0.949	1.000

また、各主成分の主成分負荷量をグラフにまとめた。第1及び第2主成分の主成分負荷量を図4.16に示す。横軸が第1主成分、縦軸が第2主成分を表す。赤い円は半径1.0の円であり、各変数の負荷量の大きさを測るための目安である。

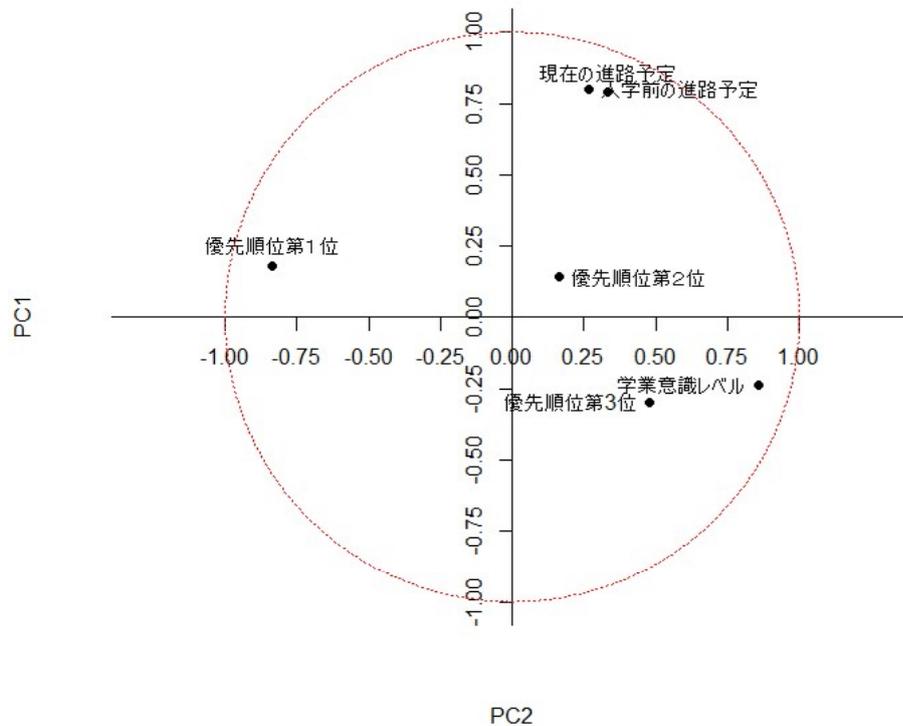


図 4.16: 第1第2主成分負荷量（目的意識）

この図より、第1主成分に大きく影響しているのは学業意識レベルや優先順位第1位であり、第2主成分に大きく影響しているのは現在の進路希望及び入学前の進路希望である。つまり、第1主成分は主に学業への取組意識、第2主成分は進路希望をそれぞれ表しているといえる。

同様に、第3及び第4主成分の主成分負荷量を図4.17に示す。横軸が第3主成分、縦軸が第4主成分を表す。

この図より、第3主成分、第4主成分に大きく影響しているのはどちらも優先順位第2位及び第3位であり、第3主成分では2つの変数が打ち消しあう形、第4主成分では2つの変数を合算する形である。つまり、第3主成分、第4主成分は主に優先順位第2位、優先順位第3位を表した成分である。優先順位第2位の値が高ければどちらの成分の値も大きくなり、優先順位第3位の値が大きくなると第3主成

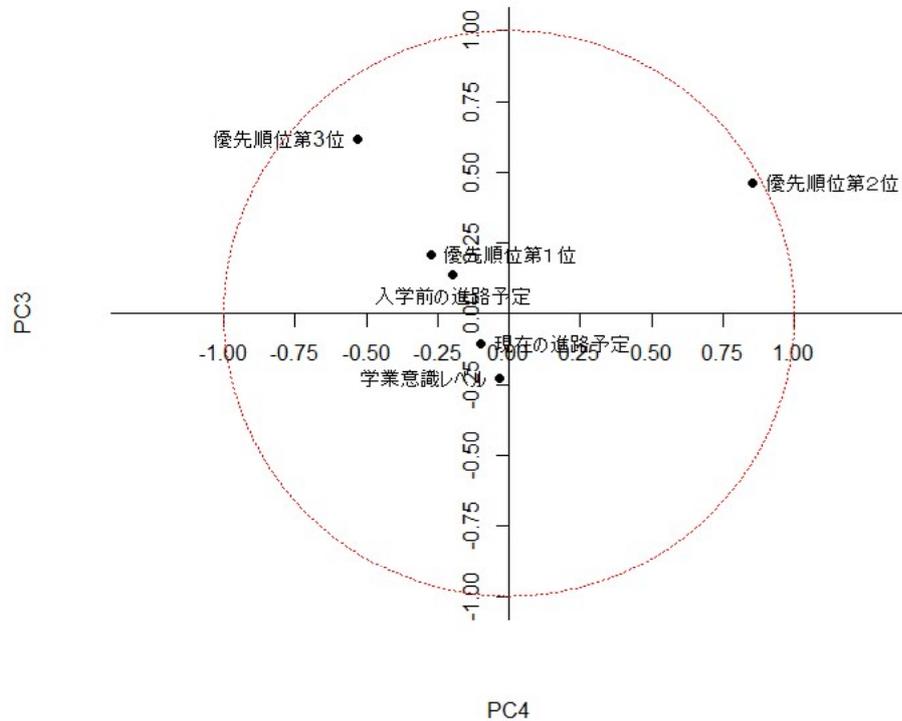


図 4.17: 第3第4主成分負荷量 (目的意識)

分の値は小さく、第4主成分の値は大きくなる。

その後、これらの主成分と成績との関連性をみるため、各主成分とGPAとの相関係数を計算した。この結果を表4.9に示す。

表 4.9: 主成分とGPAの相関 (目的意識)

変数	GPAとの相関係数
PC1	0.2260
PC2	0.0999
PC3	0.0431
PC4	-0.1028

この表より、第1、第2、第4主成分において有意な相関がみられた。特に第1主成分とは強い正の相関がみられたため、学業への取組意識が高いほど成績が良い傾向があるといえる。また、第2主成分も正の相関であるので、各進路の値が大きいほど成績が良い傾向がある。一方、第4主成分とは負の相関がある。第4主成分は優先順位第2位及び第3位の値が大きいほど値が大きくなるので、これも学業への

取組意識と成績との相関をあらわしていると考えられる。

4.2.3 対応分析

4.1.3 項と同様に、GPA と目的意識において対応分析を行う。GPA と目的意識の各変数間においてクロス集計表を作成し、作成したクロス集計表に対して対応分析を行った。優先順位第1位に対する分析結果を図4.18に示す。赤のラベルが各アンケート項目、黒のラベルがGPAを示す。また、「+」は原点を表す。ただし、回答者数が5人未満である項目は分析から除いてある。

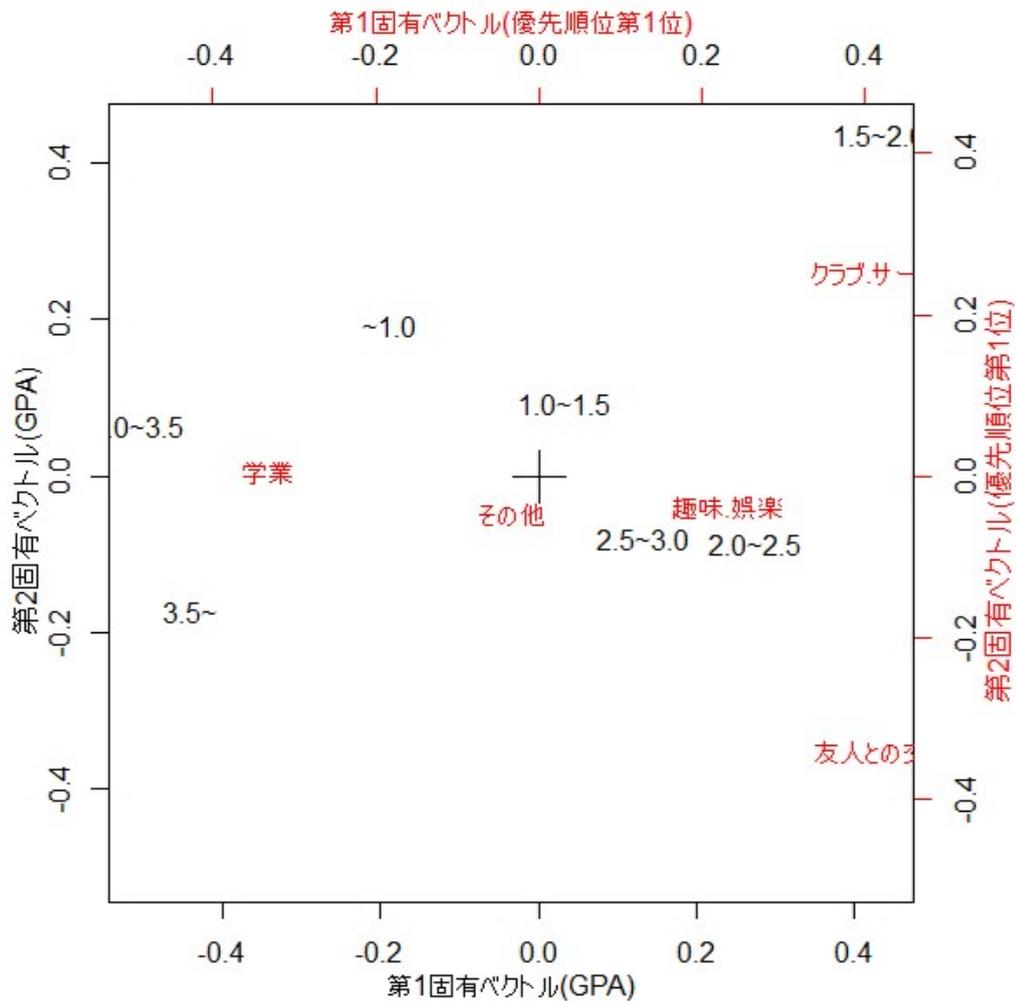


図 4.18: GPA と優先順位第1位の対応分析

この図より、「学業」と「3.5～」及び「3.0～3.5」，また「趣味・娯楽」「友人との交流」と「2.5～3.0」及び「2.0～2.5」において，原点からの方向が等しい．また，「学業」「3.5～」「3.0～3.5」「友人との交流」については原点からの距離が大きい．つまりこれらの要素間には関連性があり，中でも原点からの距離が大きい要素は強い特徴がある．したがって，高成績を修める学生は学業を最も優先する傾向にあり，趣味や娯楽を優先する学生は平均的な成績を修める傾向にあるといえる．

次に，優先順位第2位に対する分析結果を図4.19に示す．

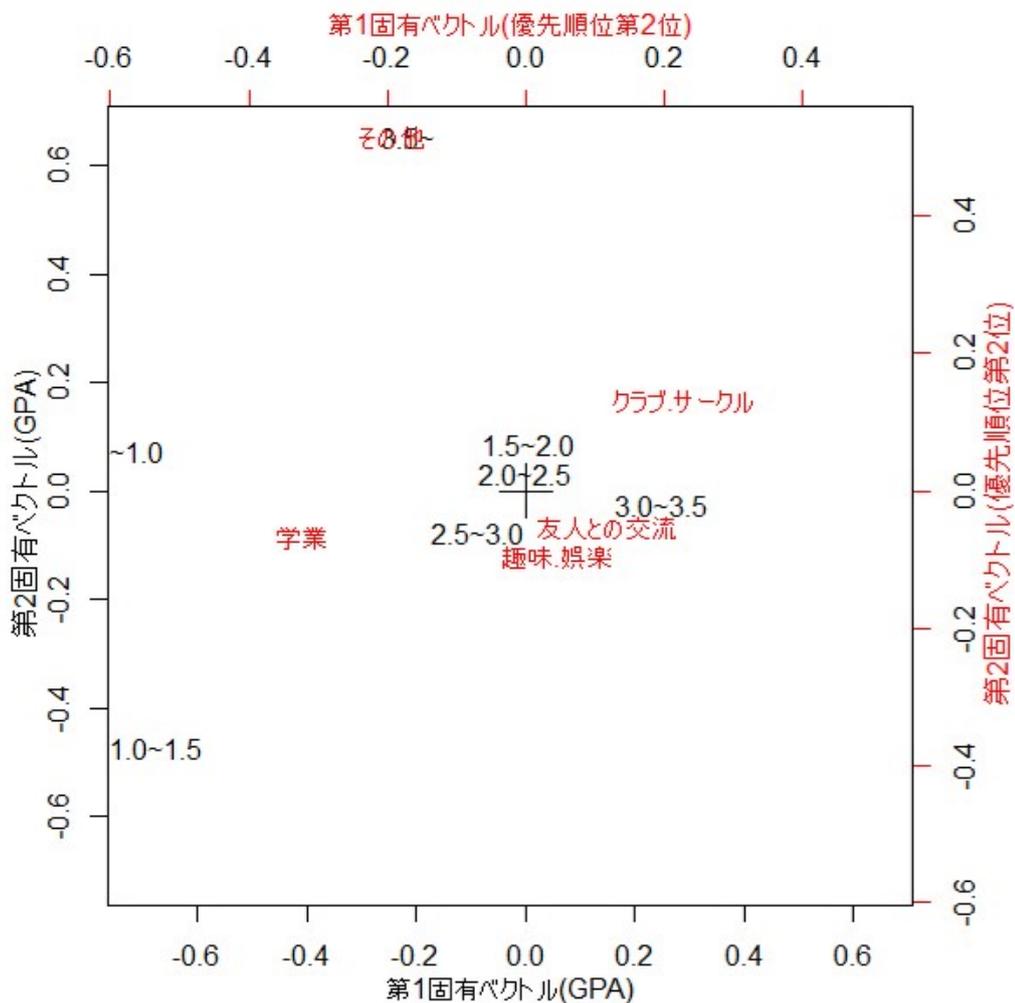


図 4.19: GPA と優先順位第2位の対応分析

この図より、「その他」と「3.5～」において，原点からの方向が等しく，どちらの要素も原点からの距離が大きい．つまり，この2つの要素間には関連性がある．「そ

の他」の内容として考えられる候補のひとつはアルバイトである。したがって、アルバイトを2番目に優先する学生は高成績を修める傾向がある可能性がある。

次に、優先順位第3位に対する分析結果を図4.20に示す。

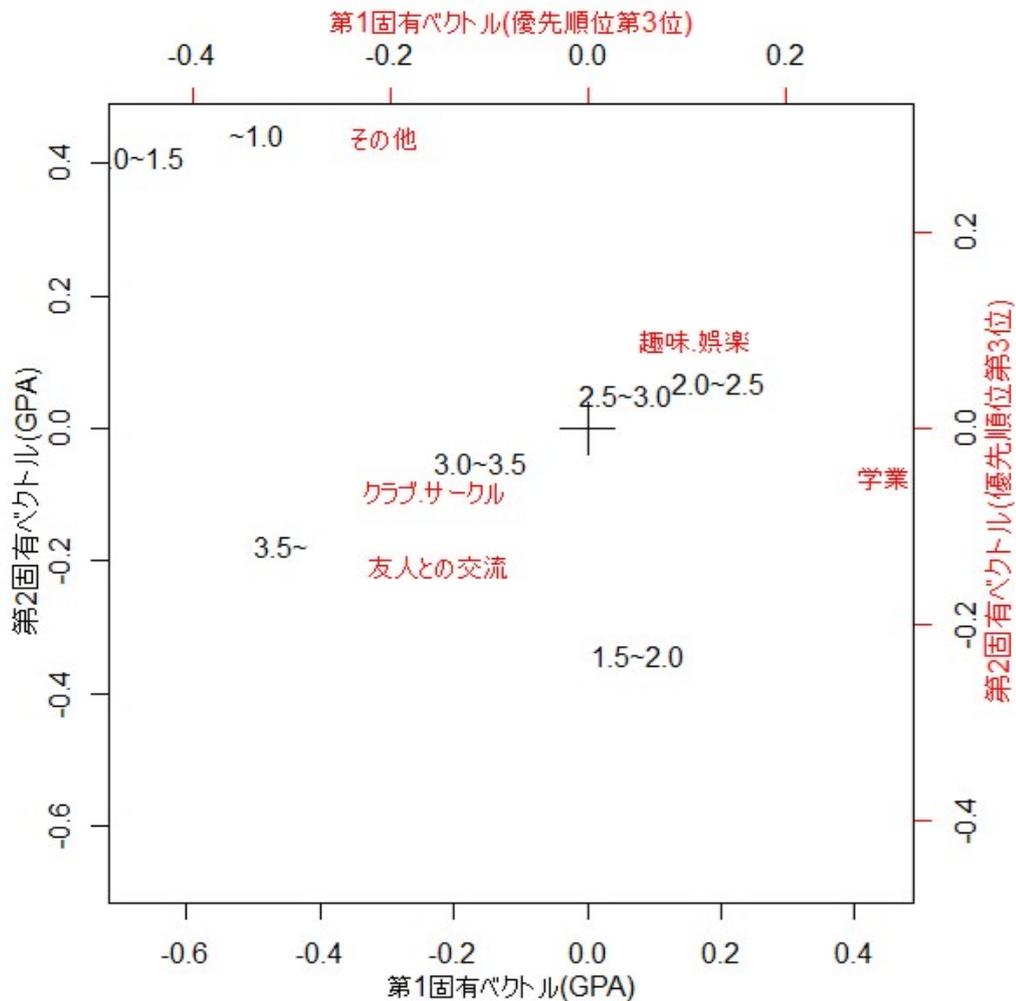


図 4.20: GPA と優先順位第3位の対応分析

この図より、「その他」と「~1.0」及び「1.0~1.5」において原点からの方向が等しく、どの要素も原点からの距離が大きい。また、原点からの距離は大きくないが、「クラブ・サークル」「友人との交流」と「3.0~3.5」及び「3.5~」においても原点からの方向が等しい。つまり、これらの要素間には関連性があるといえる。ただし、優先順位第3位には第1位と第2位の結果が大きく関わってくる。つまり、第1位において「学業」または第2位において「その他」を選んだ学生が第3位に「クラ

ブ・サークル」や「友人との交流」を選んでいるため、第3位でクラブや友人を優先する生徒は高成績である傾向があるといえる。また、第2位まででその他を選ばず、第3位で選んだ生徒は低成績である傾向がある。

次に、進路希望について入学前のものに対する分析結果を図4.21に示す。

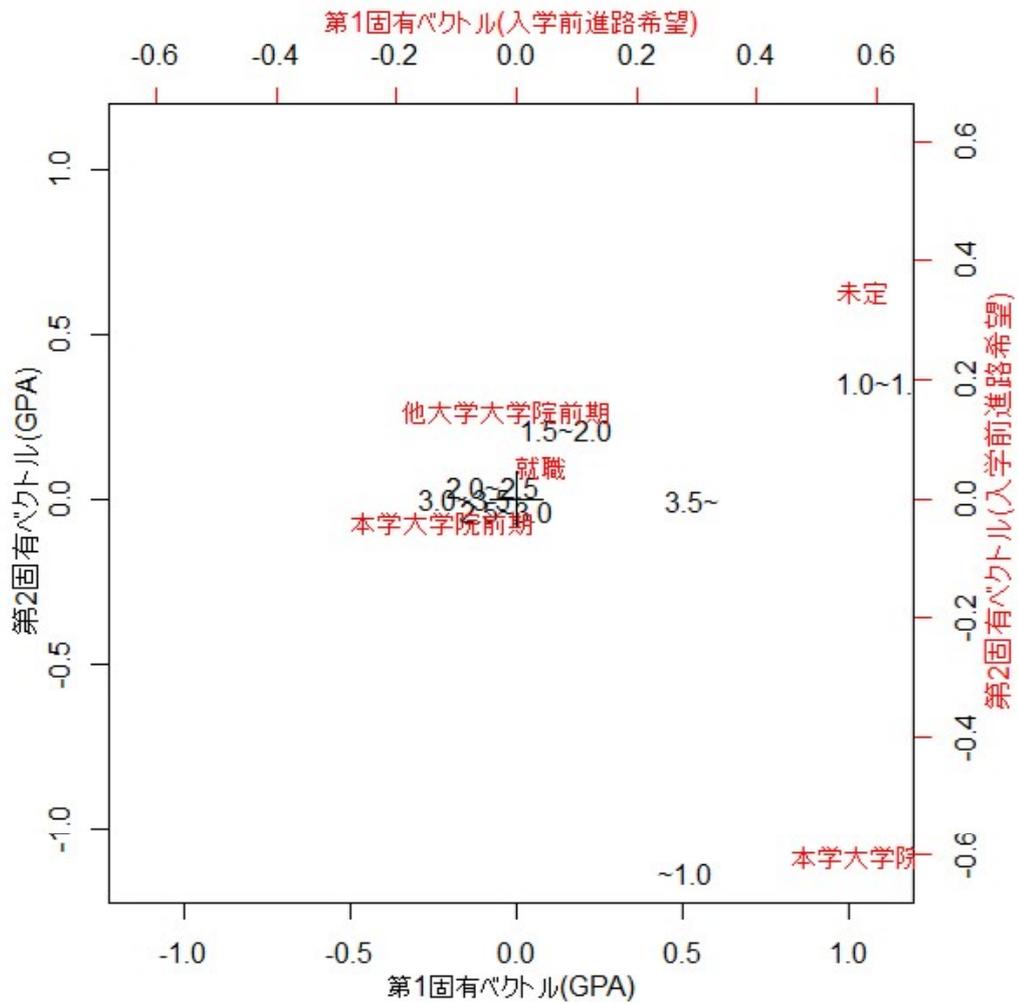


図 4.21: GPA と入学前進路希望の対応分析

この図から、「未定」と「1.0~1.5」、また「本学大学院後期」と「~1.0」について、原点からの方向が等しく、どの要素も原点からの距離が大きい。つまり、これらの要素間には関連性があるといえる。したがって、入学前の進路が未定である学生は成績が悪く、入学前に本学の後期課程を希望していた学生も成績が悪い傾向がみられた。

最後に、現在のものに対する分析結果を図 4.22 に示す。

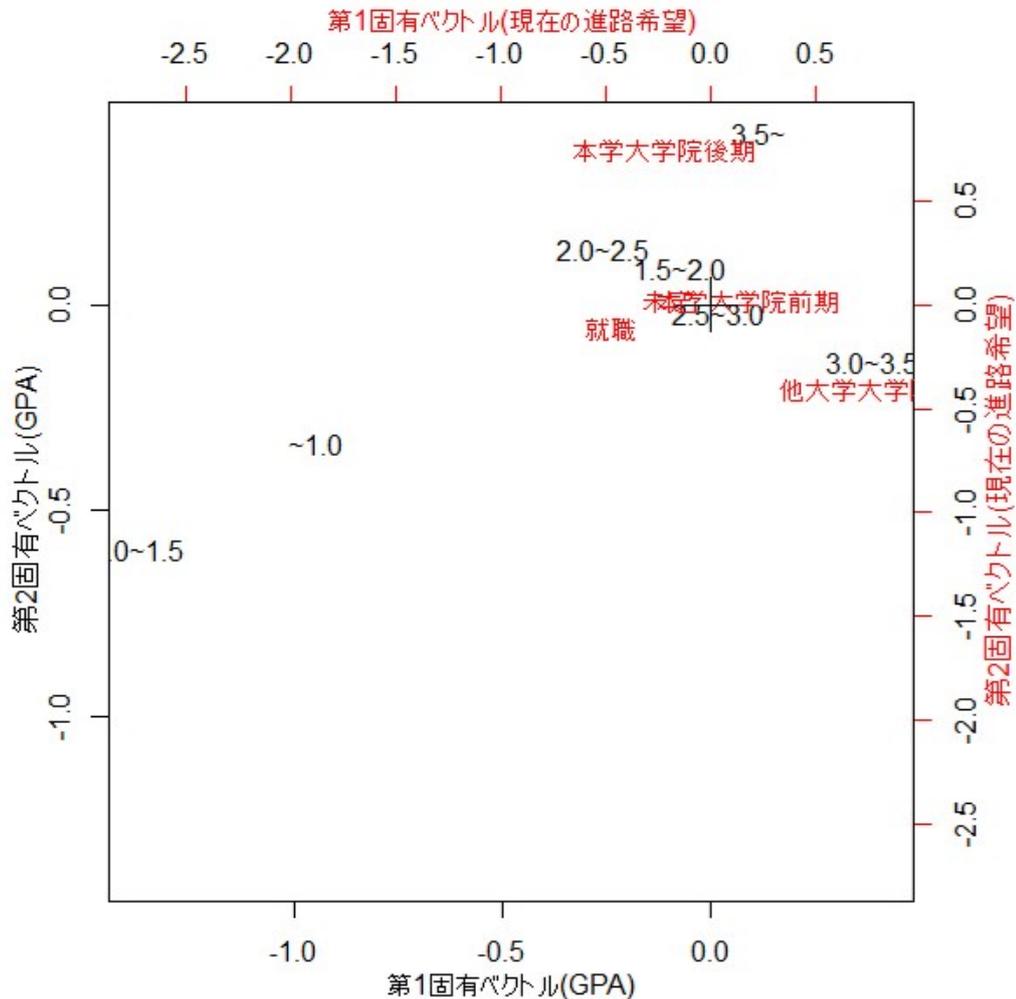


図 4.22: GPA と現在の進路希望の対応分析

この図から、「本学大学院後期」と「~3.5」,「他大学大学院前期」と「3.0~3.5」,「就職」と「~1.0」及び「1.0~1.5」において、原点からの距離はあまり大きくないが、原点からの方向がそれぞれ等しい。つまり、これらの要素間には関連性があるといえる。したがって、アンケート回答時において卒業後の進路として本大学院の博士後期課程や他大学院の前期課程などを希望する学生は成績が良い傾向にある。また、成績の悪い学生は就職を希望する傾向にあるといえる。

さらに、優先順位において第1位に「学業」を選んだ場合について注目し、優先順位1位で「学業」を選んだ学生のデータを取り出す。このデータにおける優先順位第2位とGPAについて対応分析を行った。この結果を図4.23に示す。

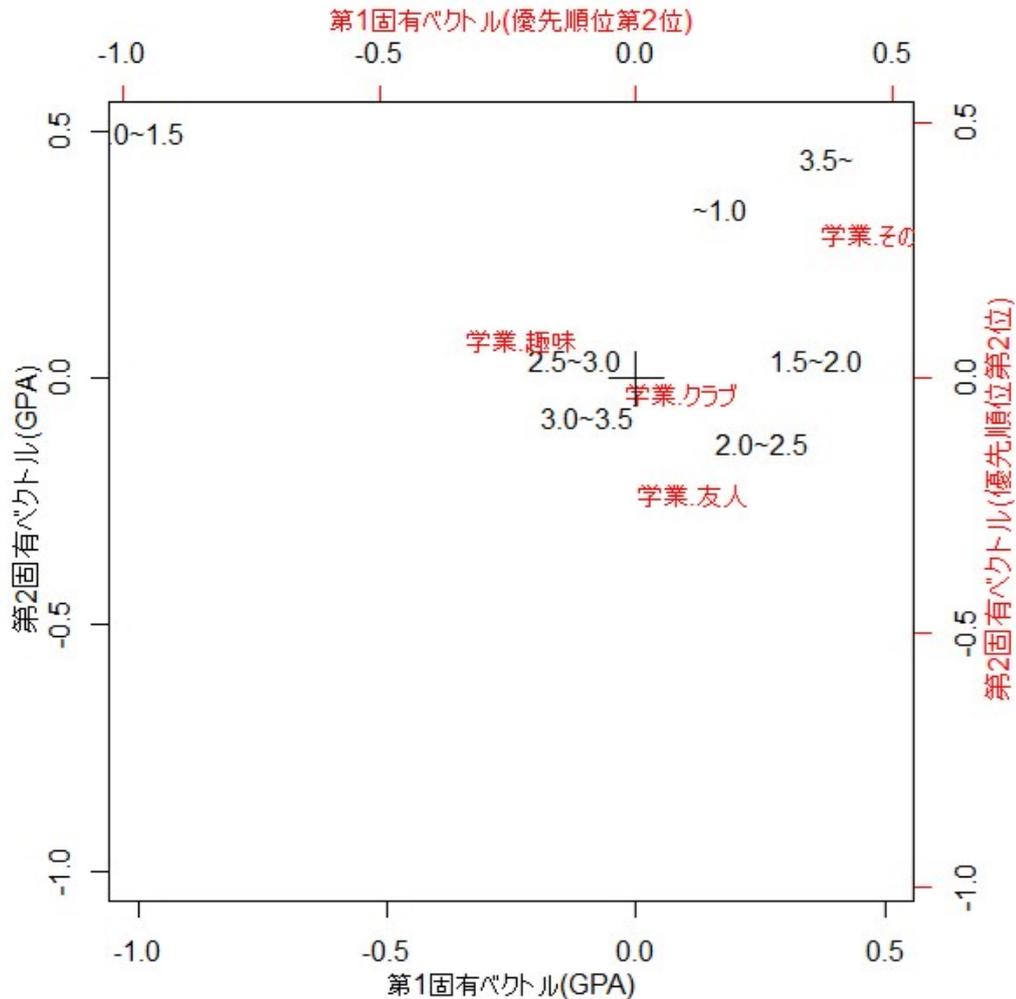


図 4.23: GPA と優先順位第1位が「学業」である第2位の対応分析

この図から、優先順位第1位で「学業」を選んだ学生のうち、第2位で「その他」と答えた学生について、原点からの距離が大きいことから特徴が強いことがわかる。これは「3.5～」及び「～1.0」と原点からの方向が同じであるので、これらと関連性があるといえる。優先順位第1位及び第2位の結果を踏まえると、第1位で「学業」、第2位で「その他」と回答した学生は、成績が良い傾向にあるといえる。

4.2.4 決定木分析

GPA, 学年, 目的意識に関するアンケート項目に対して, 決定木分析を行った. GPA を目的変数, その他の変数を説明変数に設定した. なお, 本研究で用いる C5.0 アルゴリズムでは目的変数はカテゴリデータである必要がある. したがって GPA を分割しカテゴリデータとした. 分割方法は, 1.0 台, 2.0 台, 3.0 台での 3 分割, 平均以上か以下かの 2 分割の 2 種類である. さらに, 優先順位の代わりに学業意識レベルを採用して再度分析を行った.

3 分割で優先順位を採用した分析の結果を図 4.24 に示す. この図より, 最初の分割において優先順位第 1 位が採用されていることがわかる. つまり優先順位第 1 位は強く成績に関連しているといえる. また, 優先順位第 1 位においてクラブや趣味, 友人を優先する学生は平均的な成績を修める傾向があることがわかる.

次に, 2 分割で優先順位を採用した分析の結果を図 4.25 に示す. この図より, 3 分割の決定木と同様に最初の分割において優先順位第 1 位が採用されていることがわかる. つまり優先順位第 1 位は強く成績に関連しているといえる. また, 優先順位第 1 位において学業を優先する学生は平均以上の成績を修める傾向にあることがわかる.

次に, 3 分割で学業意識レベルを採用した分析の結果を図 4.26 に示す. この図より, 最初の分割において現在の進路希望が採用されていることがわかる. つまり, 現在の進路希望が成績に強く関連しているといえる. また, 現在の進路希望において他大学院の前期課程や後期課程, 海外の大学を志望している学生は成績が良い傾向がみられた.

最後に, 2 分割で学業意識レベルを採用した分析の結果を図 4.27 に示す. この図より, 最初の分割において現在の進路希望が採用されていることがわかる. つまり, 現在の進路希望が成績に強く関連しているといえる. また, 現在の進路希望において就職を希望する学生は平均以下, それ以外を希望する学生は平均以上の成績を修める傾向がみられた.

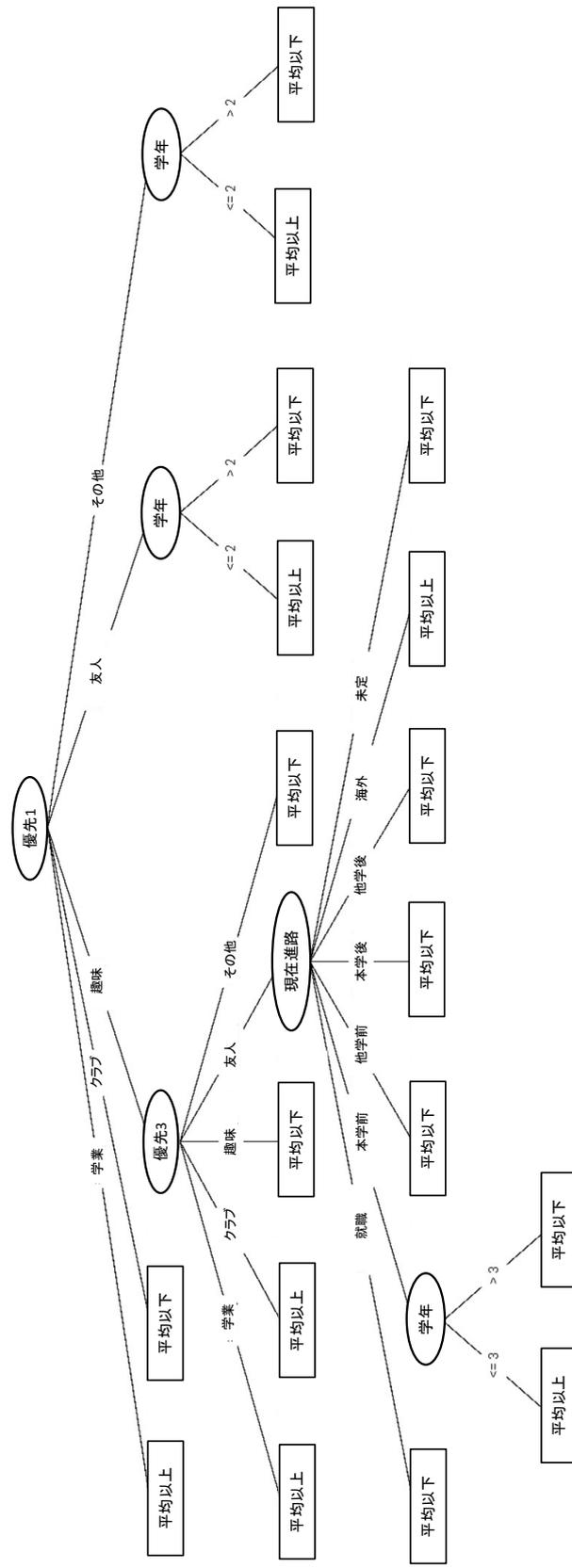


図 4.25: GPA2 分割での決定木

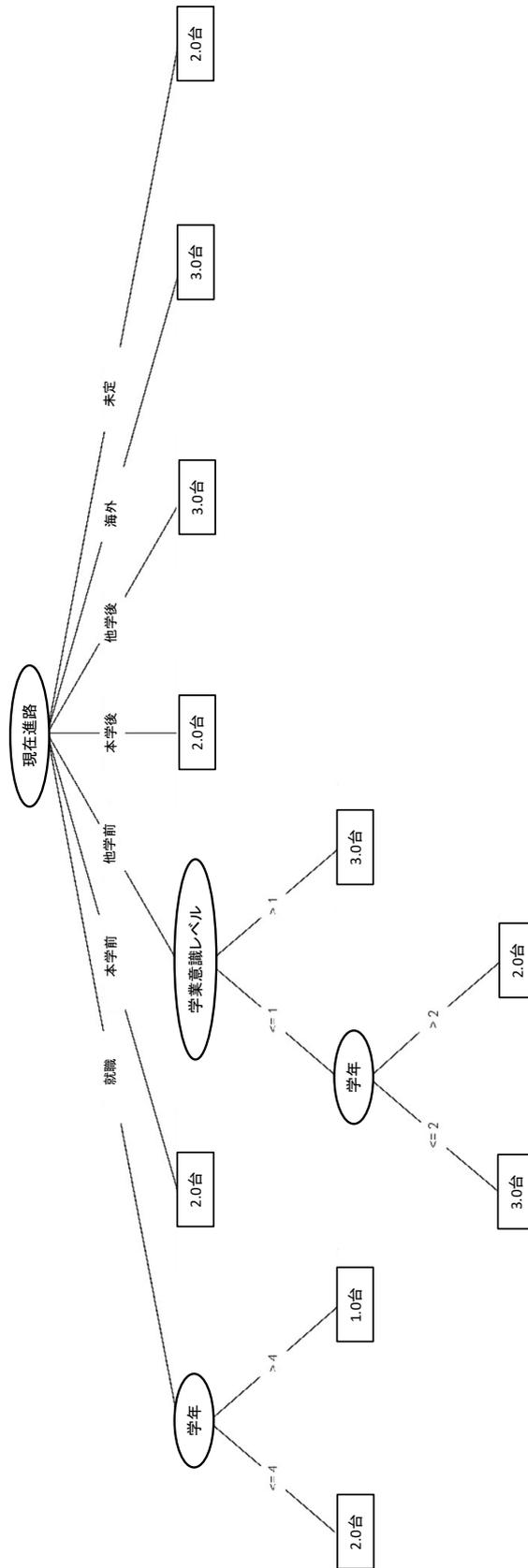


図 4.26: GPA3 分割での決定木

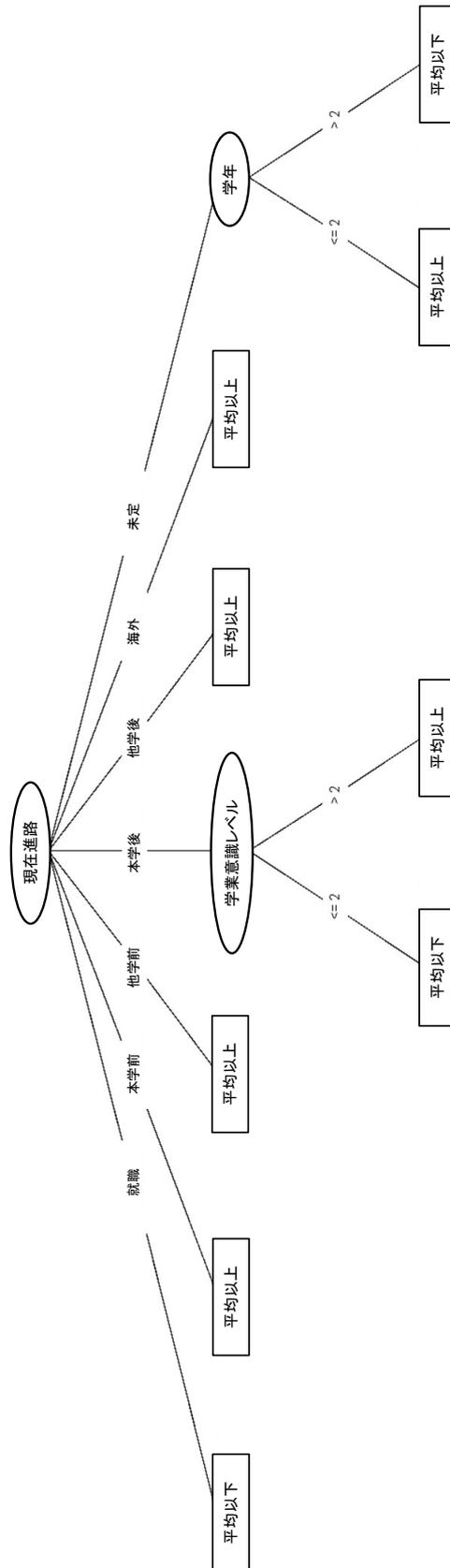


図 4.27: GPA2 分割での決定木

4.2.5 クラスタリング分析

GPA, 学年, 目的意識に関するアンケート項目に対して, 非階層クラスタリングである k-means 法を用いてクラスタリング分析をおこなった. また 4.1.4 項と同様に, エルボー法により最適なクラスタ数を計算した. その結果からクラスタの分割数は2つに設定した. また, 分析に際しては 3.3.1 項で述べたデータの数値化を行っている. 表 4.10 に得られた各クラスタの中心点の座標を示す.

表 4.10: 各クラスタの中心 (目的意識)

変数	クラスタ 1 の中心	クラスタ 2 の中心
GPA	2.51	2.73
学年	2.81	3.02
優先順位第 1 位	3.02	1.34
優先順位第 2 位	2.80	2.72
優先順位第 3 位	1.93	3.49
入学前進路希望	1.69	1.86
現在の進路希望	1.50	1.76
学業意識レベル	0.88	2.76

この表から, 各クラスタの特徴を捉える. この変数のなかで, 優先順位第 1 位及び第 3 位, さらに学業意識レベルの値が大きく異なっている. つまり, このデータはどれほど学業を優先しているかによって 2 つに分けられるといえる. 各変数の値から, クラスタ 2 のほうが学業を優先する傾向にある. また, 各クラスタに含まれるデータ数はクラスタ 1 が 176, クラスタ 2 が 270 でありクラスタ 2 のほうが多い. つまり, 学業を優先的に考えている学生が多いことを表す.

加えて, 4.1.4 項と同様に, データとクラスタをグラフにまとめることで, 分析結果を視覚的にわかりやすくする. このグラフを図 4.28 に示す. 赤い円がクラスタ 1, 青い円がクラスタ 2, 緑の数字が各データを表す. また, クラスタ 1 に分類される学生は丸, クラスタ 2 に分類される学生は三角が数字の下に記されている.

この図から, 各クラスタは第 1 主成分および第 2 主成分においてデータを斜めに分割していることがわかる. つまり, 第 1 主成分得点 = 第 2 主成分得点 となる直線によってデータを分割でき, 第 1 主成分得点 \leq 第 2 主成分得点 となるグループがクラスタ 1, 第 1 主成分得点 \geq 第 2 主成分得点 となるグループがクラスタ 2 である.

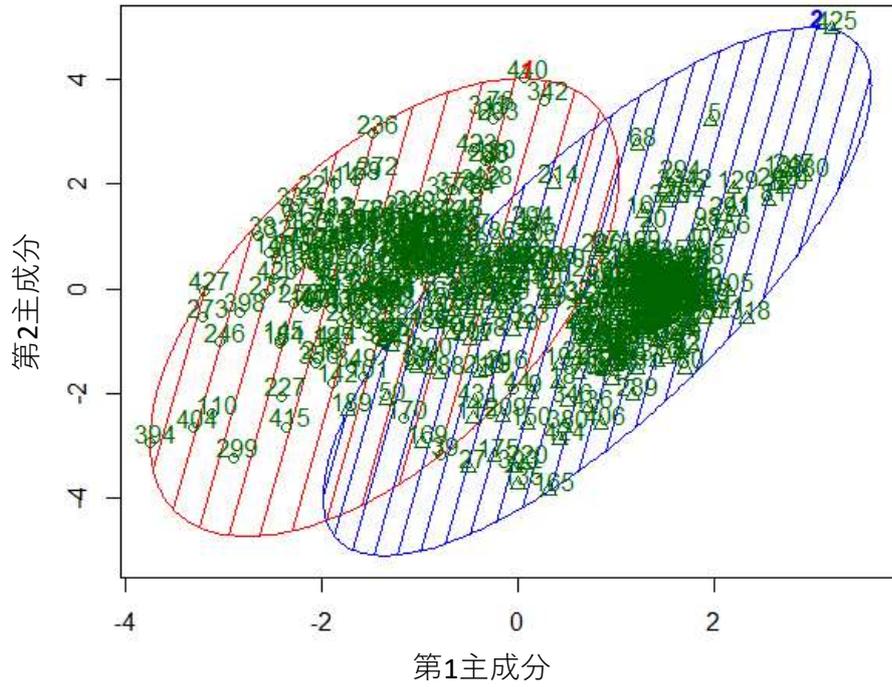


図 4.28: クラスタリング分析 (目的意識)

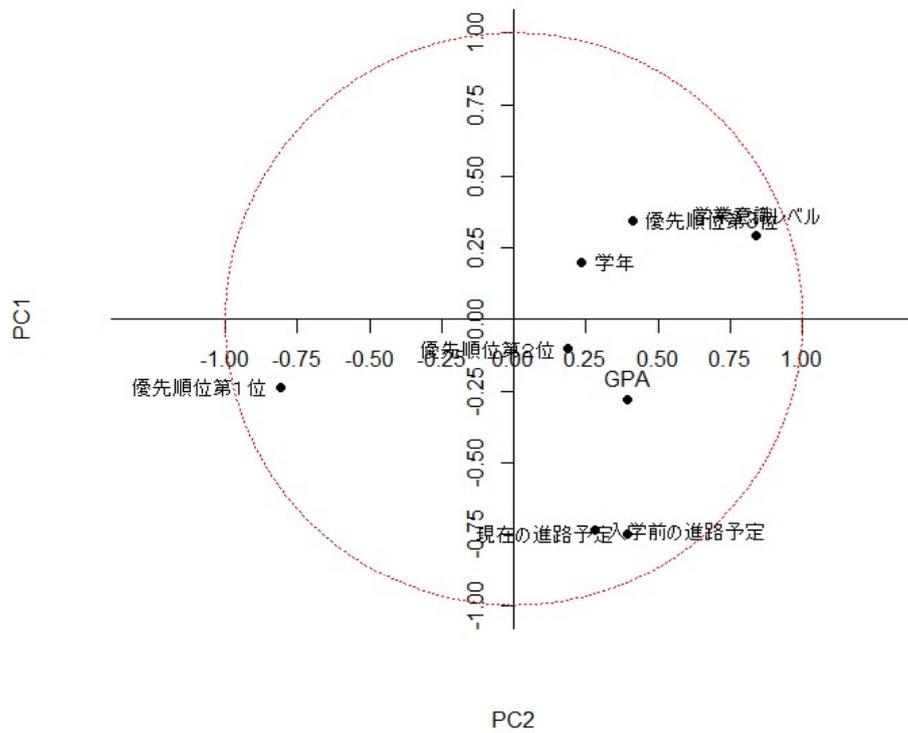


図 4.29: クラスタリング分析における第1第2主成分負荷量 (目的意識)

ここで、2つの主成分の主成分負荷量について図 4.29 に示す。

この図から、第1主成分に大きく影響しているのは学業意識レベルや優先順位第1位であり、第2主成分に大きく影響しているのは入学前の進路希望と現在の進路希望である。よって、第1主成分は学業への取組意識の高さ、第2主成分は進路希望の値の小ささを表しているといえる。つまり、クラスタ1には学業への取組意識が低く、進路希望の値が小さい学生が、クラスタ2には学業への取組意識が高く、進路希望の値が大きい学生が含まれる。

さらに、各クラスタにおける GPA を箱ひげ図で表したグラフを図 4.30 に示す。

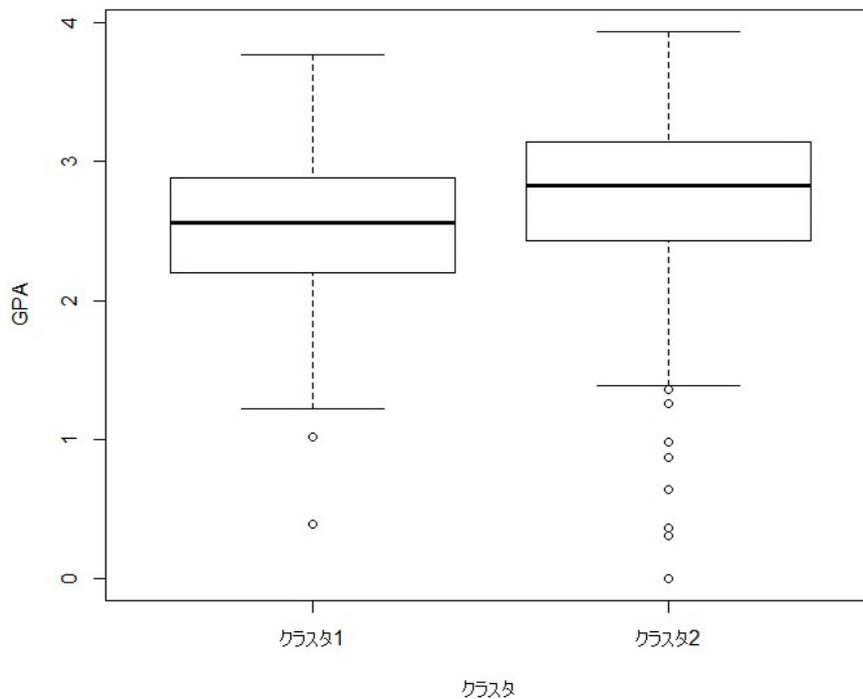


図 4.30: 各クラスタの GPA に関する箱ひげ図 (目的意識)

この図からクラスタ2のほうが全体的に GPA が高い傾向にあることがわかる。クラスタ2の中央値や最大値はクラスタ1より高いが、同時に GPA が極端に低い学生も外れ値としてクラスタ2に分類されている。したがって、クラスタ2には高成績の学生と低成績の学生が両方多く含まれているといえる。つまり、成績の良い学生は学業への取組意識が高く、進路希望の値が大きい傾向があるが、一定以下の悪い成績を修める学生もこの傾向にある。

4.2.6 目的意識に関する分析のまとめ

これまでの結果から、現在の進路希望と成績は密接な関連性があることがわかった。より高度な教育課程への進学を志望する学生は、成績も高い傾向がみられた。また、優先順位と成績とも関連性があり、学業への取組意識が高い学生ほど成績が高い傾向がみられた。ただし、一定以下の成績を修める学生も学業への取組意識が高い傾向がみられたため、この点には注意が必要である。

4.3 成績以外の関連性の分析

通学状況及び目的意識の各変数間の関連性について分析を行った。本節では、その分析の概要と結果、考察について述べる。

4.3.1 相関分析

通学状況の各変数間、目的意識の各変数間、通学状況と目的意識の各変数間それぞれに対して相関分析を行った。通学状況の各変数間における相関係数を表 4.11 に示す。検定によって有意であると判断された係数は、表において赤く表示している。

表 4.11: 通学状況の変数間の相関係数

	距離	時間	手段 1	手段 2	入構	住居	同居人	出身校
通学距離	1	0.8305	0.7402	0.4538	-0.2594	0.6621	0.7477	-0.4371
通学時間	0.8305	1	0.6815	0.4779	-0.2218	0.5718	0.6328	-0.3746
通学手段第 1 位	0.7402	0.6815	1	0.3782	-0.2262	0.6087	0.6626	-0.4522
通学手段第 2 位	0.4538	0.4779	0.3782	1	-0.0346	0.4229	0.4634	-0.2684
入構手段	-0.2594	-0.2218	-0.2262	-0.0346	1	-0.0940	-0.1844	0.2182
住居	0.6621	0.5717	0.6087	0.4229	-0.0940	1	0.7315	-0.5131
同居人	0.7476	0.6327	0.662	0.4634	-0.1844	0.7315	1	-0.5285
出身校	-0.4371	-0.3745	-0.4522	-0.2684	0.2183	-0.5131	-0.5285	1

この表より、ほぼすべての項目間において有意な相関がみられた。通学状況の要素はすべて距離によって関連付けられるため、この結果は自明なものである。

次に、目的意識の各変数間における相関係数を表 4.12 に示す。

表 4.12: 目的意識と GPA の相関係数

	優先 1	優先 2	優先 3	入学前進路	現在進路	意識レベル
優先順位第 1 位	1	-0.2077	-0.1972	-0.0157	-0.1132	-0.6556
優先順位第 2 位	-0.2077	1	-0.1344	0.0336	0.0470	0.0069
優先順位第 3 位	-0.1972	-0.1344	1	0.0192	-0.0272	0.3240
入学前進路希望	-0.0157	0.0336	0.0192	1	0.4871	0.0474
現在の進路希望	-0.1132	0.0470	-0.0272	0.4871	1	0.0843
学業意識レベル	-0.6556	0.0069	0.3240	0.0474	0.0843	1

この表より、優先順位間、進路希望間ではそれぞれ有意な相関がみられたが、優先順位と進路希望間には有意な相関はみられなかった。つまり、優先するものと進路の間には直接的な因果関係はないと考えられる。

最後に、通学状況と目的意識間での相関係数を表 4.13 に示す。

表 4.13: 通学状況と目的意識間の相関係数

	距離	時間	手段 1	手段 2	入構	住居	同居人	出身校
優先順位第 1 位	0.0013	0.0251	0.0679	-0.0415	0.0649	-0.0128	0	-0.0025
優先順位第 2 位	0.0090	0.0305	0.0174	-0.0573	-0.0092	0.0211	0.0205	0.0215
優先順位第 3 位	0.0295	0.0601	0.0127	0.0346	0.0010	-0.0337	0.0173	-0.0383
入学前進路希望	-0.0126	0.0134	-0.0454	0.0759	0.0250	0.0113	0.0609	0.0335
現在の進路希望	0.0159	0.0004	-0.0323	0.1134	0.0096	0.0457	0.0589	-0.0019
学業意識レベル	-0.0395	-0.0200	-0.0729	-0.0266	-0.0134	0.0849	-0.0238	0.0566

この表より、通学状況と目的意識間ではほとんど有意な相関はみられなかったが、現在の進路希望と通学手段第 2 位の間には小さな正の相関がみられた。つまり、現在の進路希望の値が大きい学生は、2 番目に用いている通学手段が遠距離向きのものとなる傾向がある。

4.3.2 対応分析

通学状況及び目的意識内の一部の2変数間に対して対応分析を行った。通学手段第1位と第2位についての分析結果を図4.31に示す。

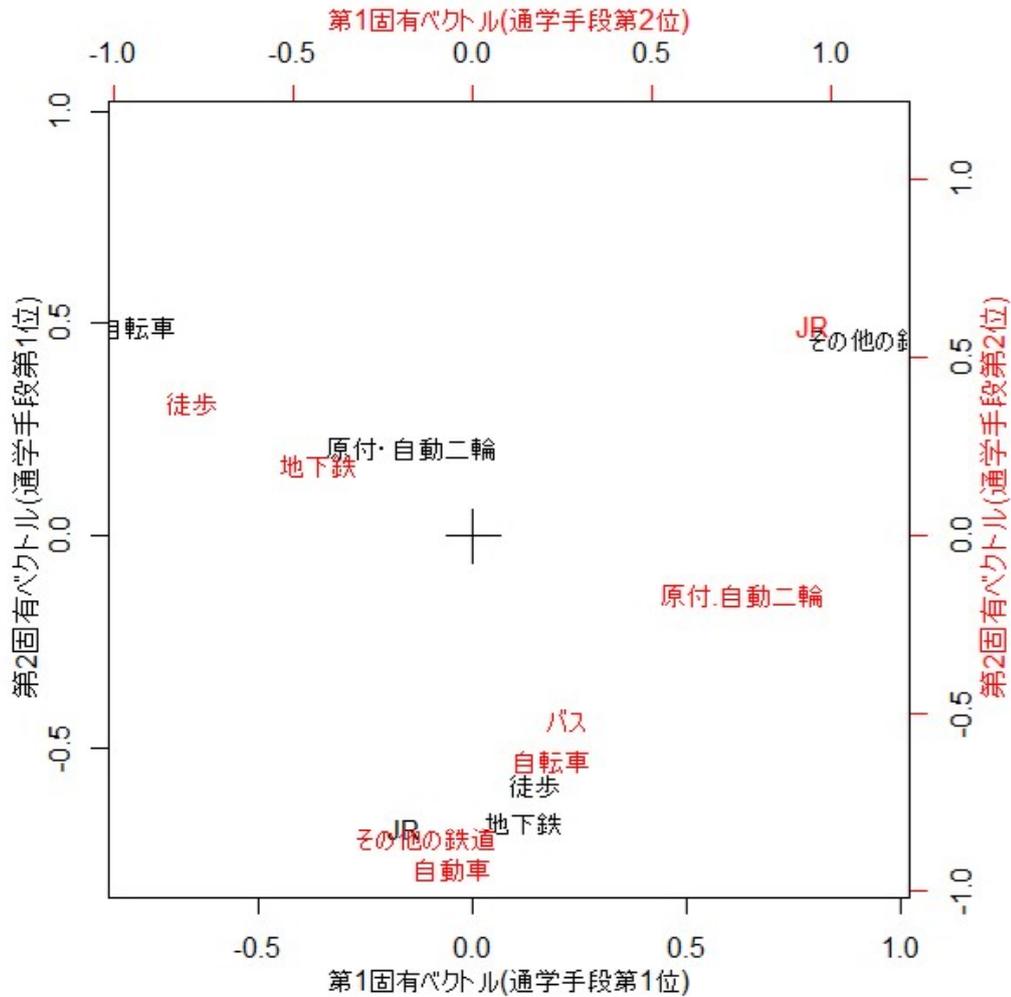


図 4.31: 通学手段第1位と第2位の対応分析

この図より、原点からの方向が等しく、原点からの距離が大きい要素に注目すると、JR とその他の鉄道を組み合わせて通学している学生が多いことがわかる。また、電車やバス、自動車などと徒歩を組み合わせている学生も多い傾向にある。

次に、優先順位第1位と現在の進路希望についての分析結果を図4.32に示す。

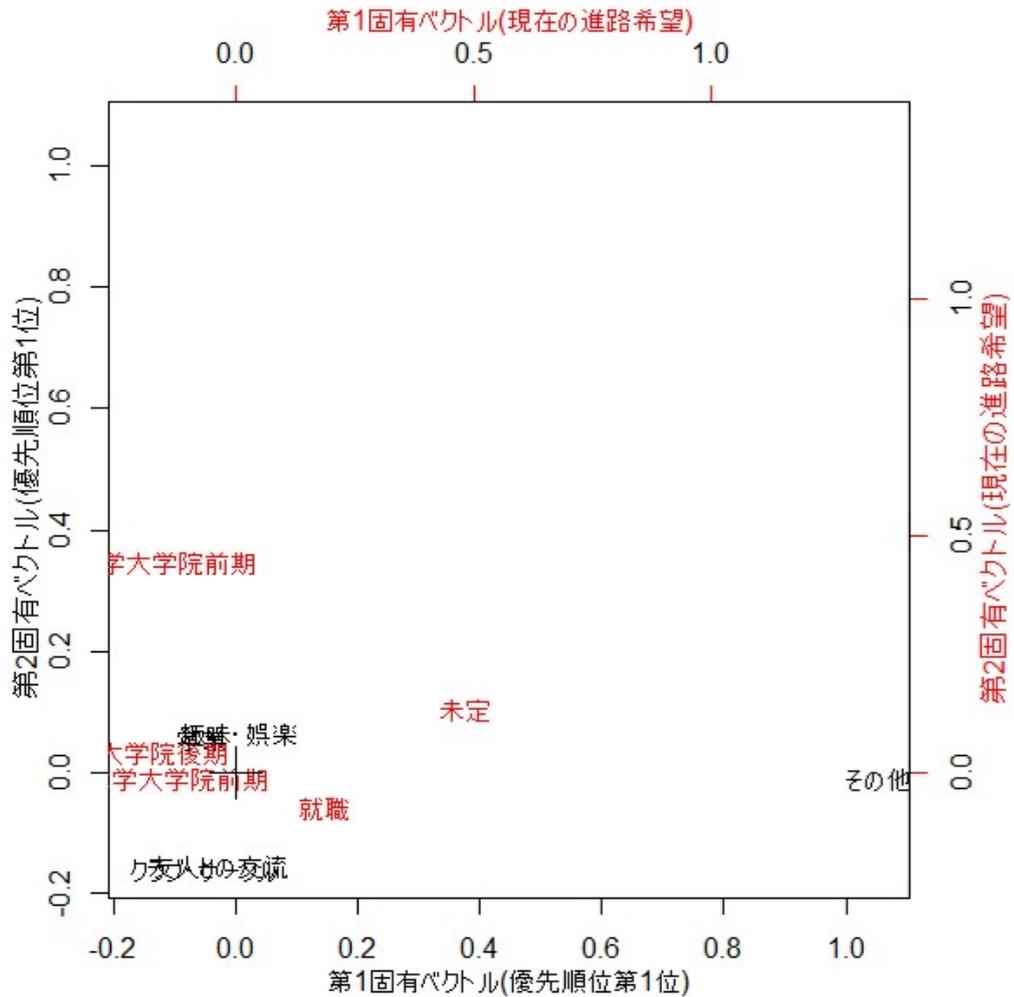


図 4.32: 優先順位第1位と現在の進路希望の対応分析

この図より、ほとんど特徴はみられないが、「その他」と「未定」の原点からの方向が等しいため、これらに微弱な関連性があるといえる。したがって、優先順位と現在の進路希望の間にはほとんど関連性はないが、優先順位でその他と回答した学生は進路が定まっていない傾向がある。

次に、通学距離と優先順位第1位についての分析結果を図4.33に示す。

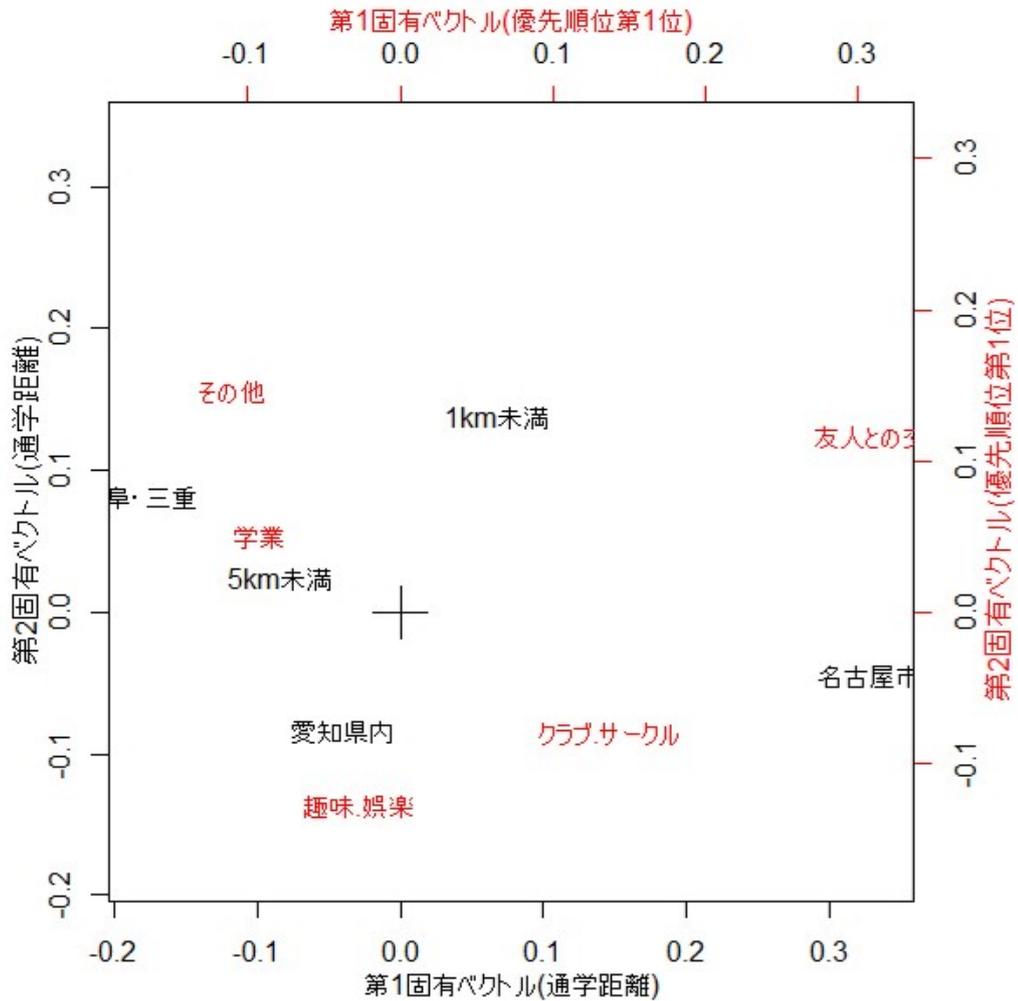


図 4.33: 通学距離と優先順位第1位の対応分析

この図より、「岐阜・三重」「5km未満」と「学業」、また「愛知県内」と「趣味・娯楽」は、原点からの距離はあまり大きくないが、原点からの方向が等しい。つまり、岐阜・三重または大学の周囲5km以内に住んでいる学生は学業を優先する傾向があり、名古屋市外の愛知県内に住んでいる学生は趣味・娯楽を優先する傾向がある。

次に、通学距離と現在の進路希望についての分析結果を図4.34に示す。

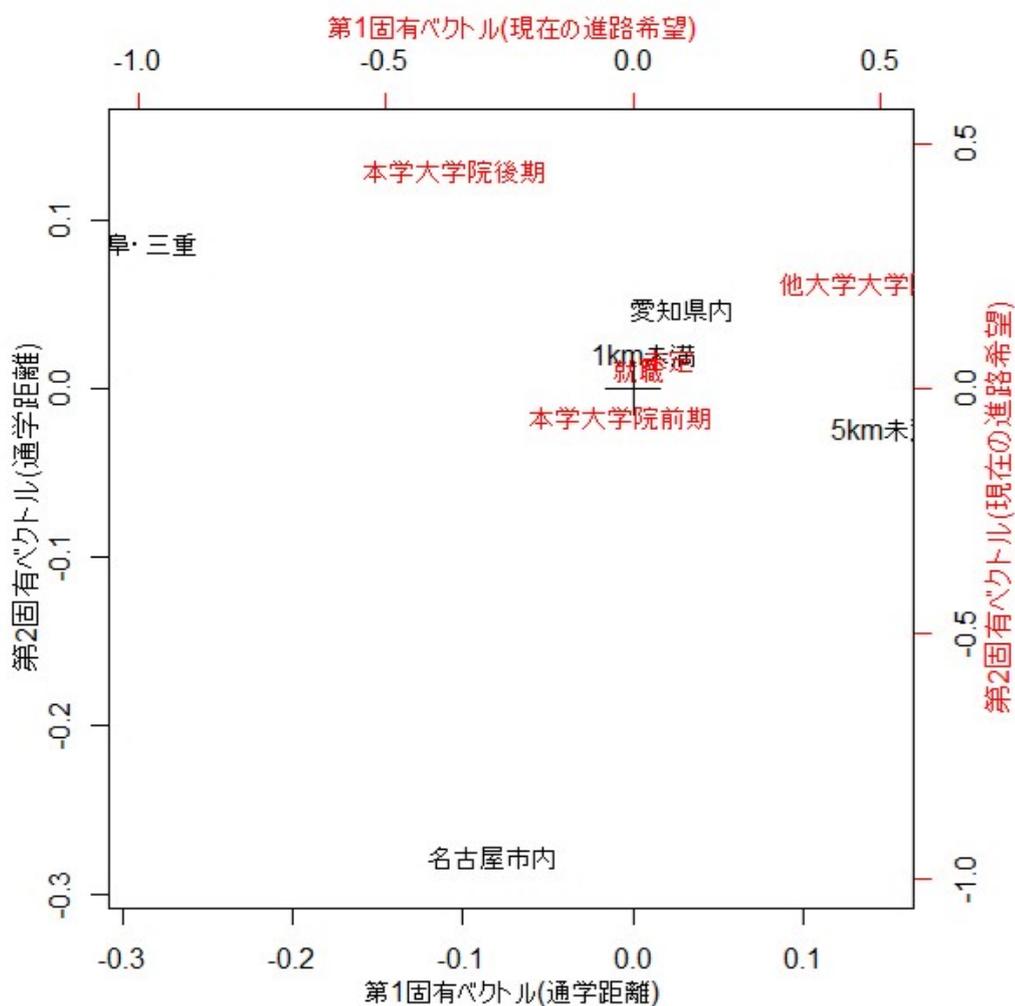


図 4.34: 通学距離と現在の進路希望の対応分析

この図より、目立った特徴はみられなかった。したがって、通学距離と現在の進路希望の間にはほとんど関連性がないといえる。

次に、出身校と優先順位第1位についての分析結果を図4.35に示す。

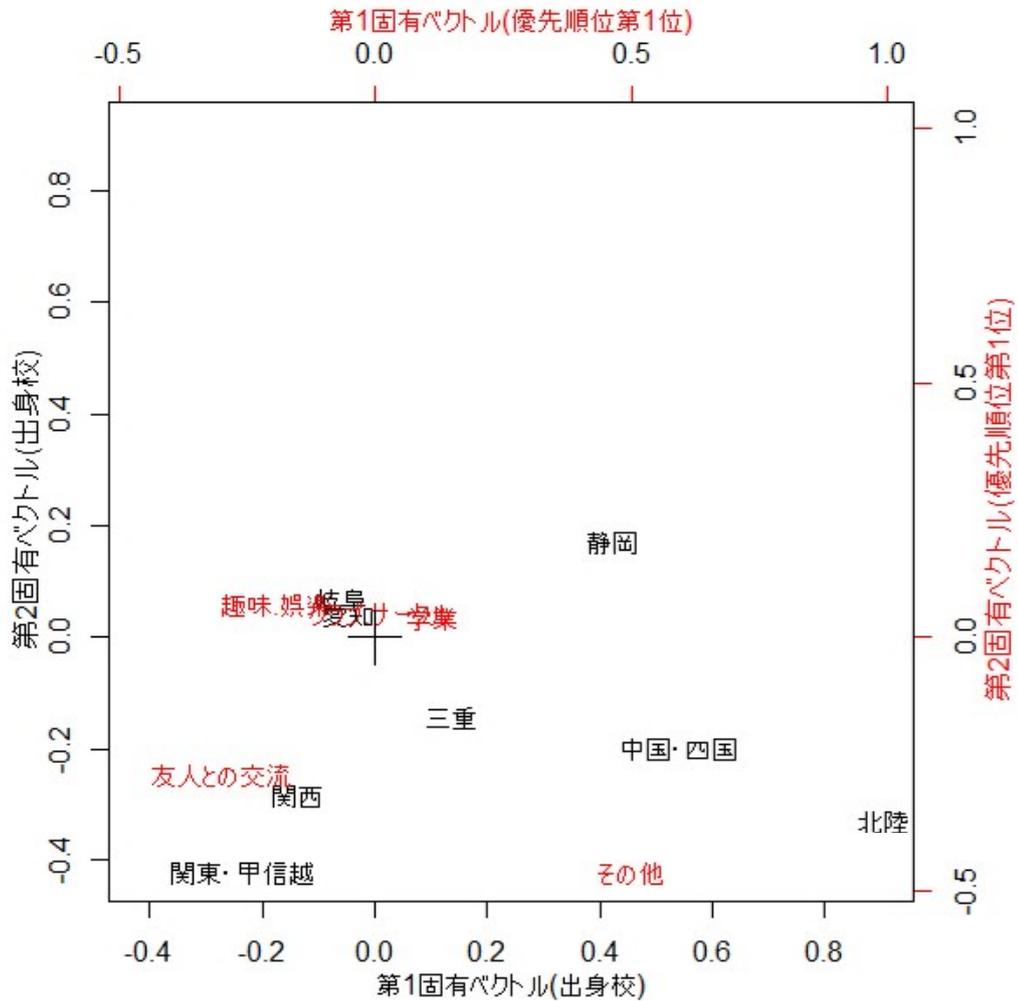


図 4.35: 出身校と優先順位第1位の対応分析

この図より、「関西」「関東・甲信越」と「友人との交流」について、原点からの距離はあまり大きくないが、原点からの方向が等しい。つまり、これらの要素間には関連性がある。したがって、関西及び関東・甲信越出身の学生は友人との交流を優先する傾向があるといえる。

次に、出身校と現在の進路希望についての分析結果を図4.36に示す。

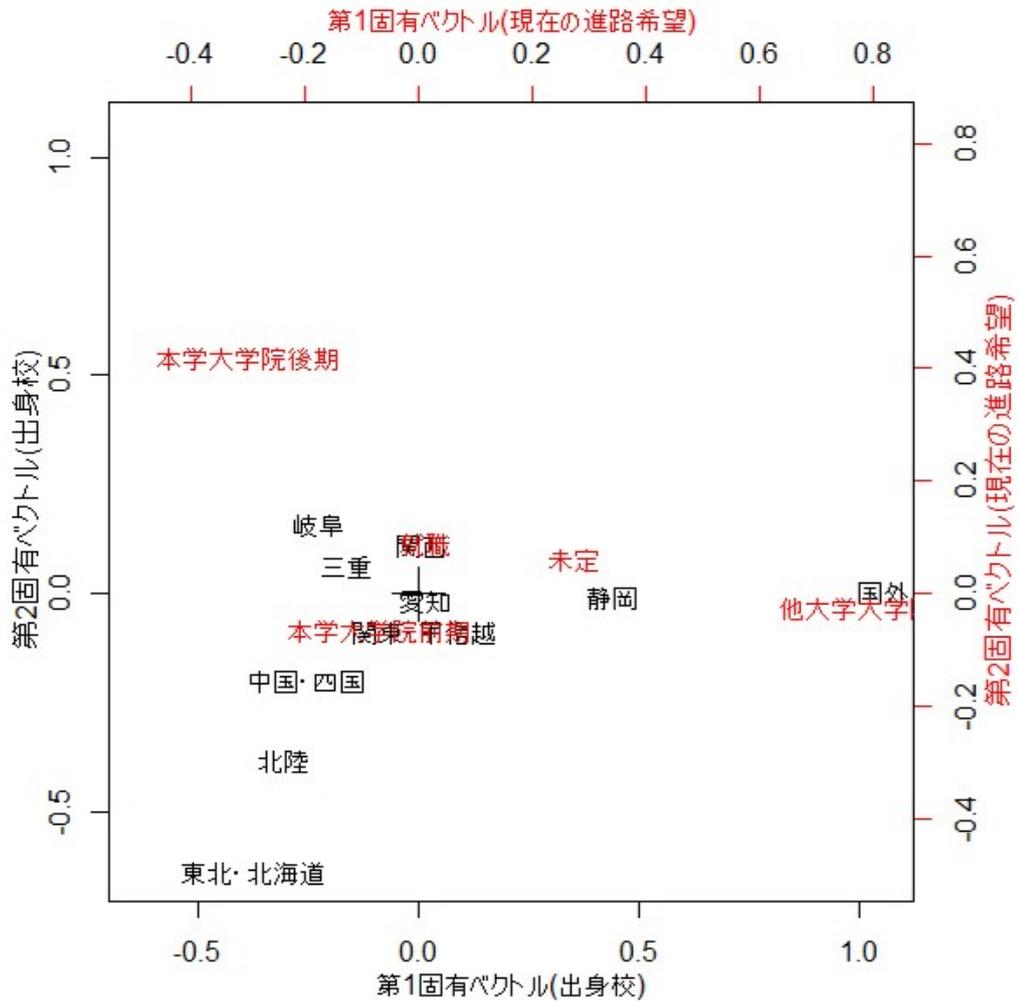


図 4.36: 出身校と現在の進路希望の対応分析

この図より、「国外」と「他大学大学院前期」について、原点からの方向が等しく、どちらの要素も原点からの距離が大きい。また、原点からの距離は大きくないが、「静岡」「未定」についても「国外」「他大学大学院前期」と原点からの方向が等しい。したがって、国外または静岡の出身の学生は他大学の大学院博士前期課程を志望する、もしくは進路がまだ定まっていない傾向があるといえる。

最後に、通学手段第2位と現在の進路希望についての分析結果を図4.37に示す。

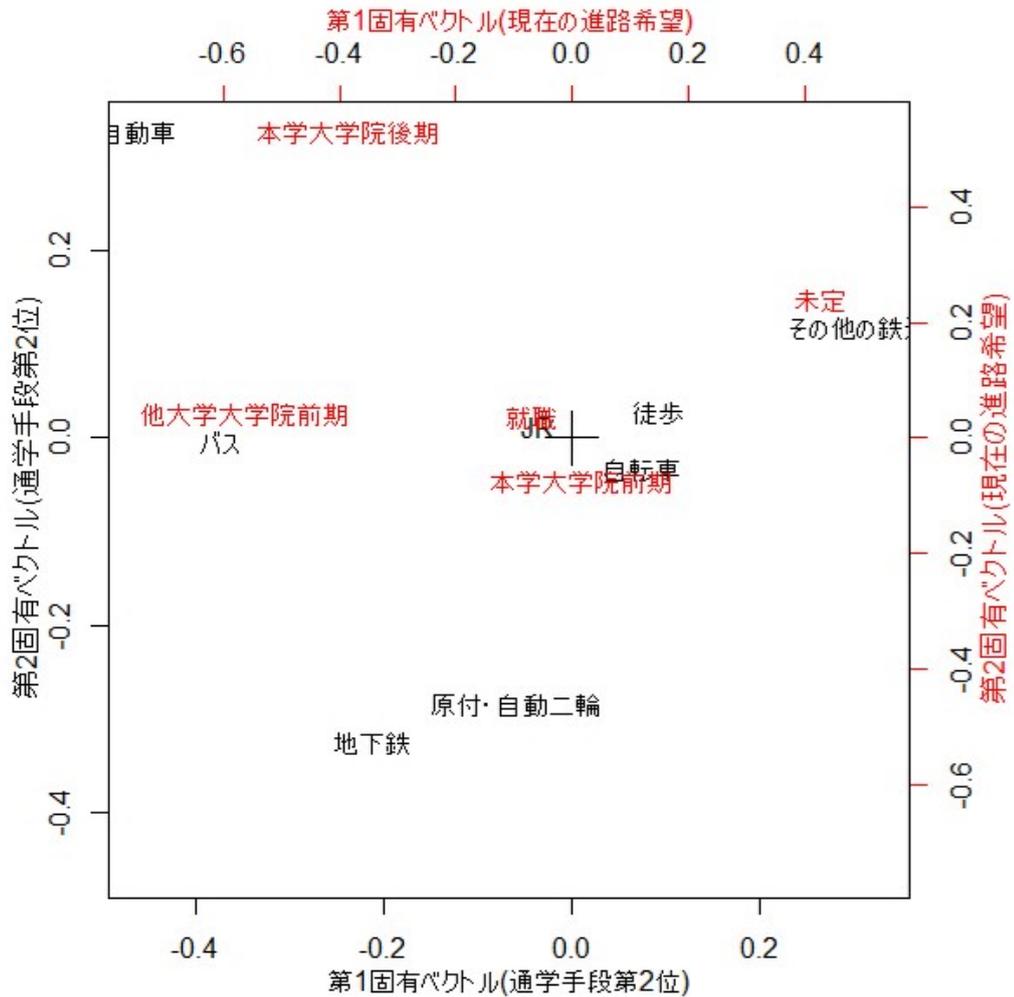


図 4.37: 通学手段第2位と現在の進路希望の対応分析

この図から、「バス」と「他大学大学院前期」、「その他の鉄道」と「未定」において、原点からの方向が等しく、どの要素も原点からの距離が大きい。したがって、第2の通学手段としてバスを用いている学生は他大学の大学院博士前期課程を希望する傾向がみられた。また、その他の鉄道を第2の通学手段として用いている学生に進路が未定である学生が多い傾向がみられた。

4.3.3 成績以外の関連性の分析のまとめ

通学状況においては、どの要素も距離に関連していることが再確認できた。また、名古屋工業大学の学生はJRとその他の鉄道や、公共交通機関と徒歩を組み合わせで通学している傾向にあることがわかった。また目的意識においては、優先順位と進路希望には関連性がほとんどないことがわかった。さらに通学状況と目的意識間においては、関西・関東出身の学生が友人関係を大切にしていることや、国外出身の学生が他大学の院を志望する傾向にあることなど、いくつかの小さな関連性が示唆された。

第5章 むすび

本研究では、名古屋工業大学の学生の通学状況及び目的意識について、成績との関連性をデータマイニング技術である相関分析、主成分分析、対応分析、クラスタリング分析、決定木分析をそれぞれ用いることで分析・考察した。その結果、通学状況に関しては、成績との関連性はほとんどみられなかった。よって、学生の通学距離や通学手段、同居人の有無に基づいた考え方で指導は避けるべきである。ただし、通学状況のなかで、通学手段において成績との微弱な関連性がみられた。また、目的意識に関しては、趣味などよりも学業を優先している学生は成績が良いという傾向がみられた。ただし、一定以下の成績の学生も学業を優先していると回答する傾向がみられたため、この点については注意が必要である。また、現時点における卒業後の進路として大学院の後期課程などのより専門的な教育課程への進学を希望している学生は、就職を目指している学生などよりも成績が良いという傾向がみられた。また、名古屋工業大学生の通学状況及び目的意識における特徴についても把握した。

今後の課題として、データの増量、特に低成績の学生や原付・自動二輪を利用している学生のデータを増やしての再分析、別の分析手法を用いての分析、さらに他大学の学生に対して同様のアンケートを実施しての分析などにより、今回得られた分析結果の検証や更なる特徴の抽出を目指す。また、実際の学生との面談による分析結果の確認や、実際の教育現場への活用などがあげられる。

謝辞

本研究を進めるにあたって、日頃から多大な御尽力を頂き、御指導を賜りました名古屋工業大学、舟橋健司 准教授、伊藤宏隆 助教に心から感謝致します。また、本研究に多大なご協力頂きました舟橋研究室諸氏に深く感謝致します。本研究の一部は JSPS 科研費 JP19K03028 の助成を受けたものです。

参考文献

- [1] 江谷典子：“創薬データマイニングにおける副作用予想モデルの提案”，情報処理学会研究報告，Vol.160, No.27, pp.1-6, 2014.
- [2] 小泉敬寛，小幡佳奈子，渡辺靖彦，近藤一晃，中村裕一：“映像対話型行動支援における頻出パターンに基づいたコミュニケーションの分析”，情報処理学会論文誌，Vol.56, No.3, pp.1068-1079, 2015.
- [3] 稲邑哲也：“ロボティクスにおけるベイジアンネットの応用”，人口知能学会誌，Vol.17, No.5, pp.546-552, 2002.
- [4] 坂本佳愛，岡田将吾，西田富明：“時系列マルチモーダルデータマイニングを用いたロボットの撮影行動則の獲得”，人口知能学会全国大会論文集，Vol.25, pp.1-4, 2011.
- [5] 原圭司，高橋健一，上田祐彰：“ベイジアンネットワークを用いた授業アンケートからの学生行動モデルの構築と考察”，情報処理学会論文誌，Vol.51, No.4, pp.1215-1226, 2010.
- [6] 伊藤圭佑，舟橋健司，伊藤宏隆：“データマイニングによる要注意学生の発見に関する研究”，平成 25 年度名古屋工業大学卒業論文，2013.
- [7] 松河秀哉，斎藤貴治：“データ・テキストマイニングを活用した授業評価アンケートフィードバックシステムの開発と評価”，日本教育工学会論文誌，Vol.35, No.3, pp.217-226, 2011.
- [8] 江本全志：“2018 年度の教学 IR における入試選抜区分による追跡調査”，秋草学園短期大学紀要，Vol.36, 2019.

- [9] 野津田雄太, 高橋健一, 稲葉通将: “大学生アンケートからの文系理系学生の特徴に関する分析”, 情報処理学会論文誌 教育とコンピュータ, Vol.1, No.4, pp.83-92, 2015.
- [10] 藤木美江: “大学における統計関連科目に対する意識調査と考察 - 担当科目における実験的試み - ”, 四條畷学園大学 リハビリテーション学部紀要, Vol.8, 2012.
- [11] 渡会純一: “ピアノ演奏指導に関する考察と展望 - 「表現技術I (音楽)」履修学生への意識調査から - ”, 教職研究, Vol.2017, pp.83-101, 2018.
- [12] 加藤利康: “授業支援システムにおける学習分析の展開”, 情報処理学会研究報告, Vol.124, No.23, pp.1-7, 2014.
- [13] 田中由香里: “相関分析・編相関分析”, <http://cogpsy.educ.kyoto-u.ac.jp/personal/Kusumi/datasem11/tanaka.pdf>, 2021-1-14 参照.
- [14] 樋口耕一: “計量テキスト分析における対応分析の活用 - 同時布置の仕組みと読み取り方を中心に - ”, コンピュータ&エデュケーション, Vol.47, pp.18-24, 2019.
- [15] 同志社大学: “R と対応分析”, https://www1.doshisha.ac.jp/~mjin/R/Chap_26/26.html, 2021-1-14 参照.
- [16] 株式会社アナリティクスデザインラボ: “決定木”, <http://www.analyticsdlab.co.jp/column/decisiontree.html>, 2021-1-18 参照.
- [17] 同志社大学: “R とクラスター分析 (1)”, https://www1.doshisha.ac.jp/~mjin/R/Chap_28/28.html, 2021-1-14 参照.
- [18] 同志社大学: “R と主成分分析”, <https://www.cis.doshisha.ac.jp/mjin/R/24/24.html>, 2021-1-14 参照.
- [19] “Microsoft Excel”, <https://www.microsoft.com/ja-jp/microsoft-365/excel>, 2021-1-14 参照.

- [20] “Weka 3 - Data Mining with Open Source Machine Learning Software in Java”, <https://www.cs.waikato.ac.nz/ml/weka/>, 2021-1-14 参照.
- [21] “The R Project for Statistical Computing”, <https://www.r-project.org/>, 2021-1-14 参照.
- [22] 小林稜介, 柴崎亮介: “職業属性を考慮した人流データの居住地及び通勤通学地推定”, 2019年度東京大学修士論文, 2019.

発表論文リスト

1. Tatsukaze Naganawa, Hirotaka Itoh, Kenji Funahashi: “Visualization and Analysis of Research Field in University Laboratories by Text Mining”, ICEMT 2019, pp.283-286, 2019.
2. Tatsukaze Naganawa, Hirotaka Itoh, Kenji Funahashi : “Analysis of the Relationship Between University Attendance and Grades by Data Mining”, IEEE-GCCE 2020, pp.199-202, 2020.

付録 分析ソフトについて

本研究では，データマイニングによる分析を行うため、フリーソフトウェアである R，および Weka を利用している。R は統計解析向けのプログラミング言語及びその開発実行環境であり，さまざまなデータマイニング手法を実行することができる。Weka はオープンソースの機械学習ソフトウェアであるが，同時に標準的なデータマイニングタスクも実行可能である。Weka はその性質から樹形図の描画能力に優れているため，本研究では決定木分析に利用している。本研究で用いた R は ver3.5.1 であり，Weka は ver3.8.4 である。また，R を効率よく利用するため R の統合開発環境である Rstudio を，データの管理のために Microsoft 社の Excel2016 をそれぞれ利用している。次節より各分析におけるコマンドや手順について述べる。

R での分析

本節では R による分析で用いたコマンドについて紹介する。

相関分析

相関分析を行うコマンドを以下に示す。

$$\text{cor.test}(\text{data1}, \text{data2}) \quad (5.1)$$

data1 と data2 における相関係数を計算する。このコマンドでは同時に t 検定により相関係数の有意性を判断できる。

主成分分析

主成分分析を行うコマンドを以下に示す。

$$\text{result} \leftarrow \text{prcomp}(\text{data}, \text{scale} = \text{T}) \quad (5.2)$$

`data` に対して主成分分析を実行し、その結果を `result` に格納する。 `result$x` に主成分得点、 `result$rotation` に主成分負荷量がそれぞれ格納される。また、 `summary(result)` コマンドにより寄与率及び累積寄与率を確認できる。

対応分析

対応分析を行うコマンドを以下に示す。

$$\text{result} \leftarrow \text{corresp}(\text{data}, \text{nf} = 2) \quad (5.3)$$

`data` に対して対応分析を実行し、その結果を `result` に格納する。その後、 `biplot(result)` コマンドにより結果をグラフに出力する。なお、このコマンドを実行するためには、ライブラリとして MASS 環境が必要である。

クラスタリング分析

クラスタリング分析を行うコマンドを以下に示す。

$$\text{result} \leftarrow \text{kmeans}(\text{data}, \text{n}) \quad (5.4)$$

`data` に対してクラスタリング分析を実行し、その結果を `result` に格納する。 `n` は分割数である。 `result$cluster` に各データの分類結果、 `result$centers` にクラスタの中心座標、 `result$size` に各クラスタに分類されたデータ数がそれぞれ格納される。なお、クラスタリング結果のプロットには `clusplot` コマンドを用いている。このコマンドの実行には `cluster` 環境が必要である。また、最適なクラスタ数を計算するコマンドを以下に示す。

$$\text{NbClust}(\text{data}, \text{method} = \text{"kmeans"}, \text{index} = \text{"alllong"}) \quad (5.5)$$

なお、このコマンドを実行するためには、ライブラリとして `NbClust`, `factoextra`, `Rmisc` 環境が必要である。

Weka での分析

本節では Weka による決定木分析の手順について紹介する。

1. あらかじめ分析用のデータセットを作成しておく。データ形式は csv 形式，もしくは arff 形式にする必要がある。なお arff 形式とは Weka 専用のデータ形式である。詳しくは Weka 公式 HP を参照されたい。また，データは英数字で統一する必要がある。数字のみのデータは数値データ，文字を含んだデータはカテゴリデータとして判断される。今回の分析においては，説明変数はカテゴリデータにする必要がある。
2. Weka を起動し，メニューの「Explorer」を選択する。その後「open file」から分析対象のデータを読み込む。
3. メニュー上部の「Classify」を選択し，「Choose → trees → J48」で，分析手法を決定する。なお，J48 は Weka における C5.0 手法である。
4. 「test options」中の「Cross-validation」を選択する。その後，「test options」の下部にあるボタンから説明変数に設定する変数を選択し，「start」により分析を実行する。
5. 「Classifier output」に結果が出力される。また，「Result list」に分析のリストが出力される。その中で行った分析にカーソルを合わせ，右クリックによりオプションが表示される。オプションの中から「Visualize tree」を選択することで，結果を図にプロットできる。

分析ファイル

各分析に用いたデータファイル名について表 5.1 にまとめる。

表 5.1: 分析に用いたファイル

分析	入力ファイル名	出力ファイル名
4.1.1	data_dist.csv	boxplot_ent3.png
4.1.2	data_dist_only.csv	prc_dist_12.png prc_dist_34.png
4.1.3	distance.csv	corresp_distance.png
	time.csv	corresp_time.png
	method1.csv	corresp_meth1.png
	method2.csv	corresp_meth2.png
	entry.csv	corresp_ent.png
4.1.3	alma.csv	corresp_alma.png
4.1.4	data_dist.csv	cluster_dist.png prc_cluster_dist.png boxplot_cluster_dist.png
4.1.5	data_dist_eng_2	-
	data_dist_eng_3	-
4.2.1	data_obj.csv	boxplot_pri1.png boxplot_now.png boxplot_level.png
4.2.2	data_obj_only.csv	prc_obj_12.png prc_obj_34.png
4.2.3	priority1.csv	corresp_pri1.png
	priority2.csv	corresp_pri2.png
	priority3.csv	corresp_pri3.png
	course_adm.csv	corresp_adm.png
	course_now.csv	corresp_now.png
4.2.3	priority1_2.csv	corresp_pri1x2_study.png
4.2.4	data_obj_eng_2.csv	tree_2.png tree_level_2.png
	data_obj_eng_3.csv	tree_3.png tree_level_3.png
4.2.5	data_obj.csv	cluster_obj.png prc_cluster_obj.png boxplot_cluster_obj.png
4.3.1	data_dist_only.csv	-
	data_obj_only.csv	-
4.3.2	method1x2.csv	corresp_meth1x2.png
	pri_now.csv	corresp_pri_now.png
	dist_pri.csv	corresp_dist_pri.png
	dist_now.csv	corresp_dist_now.png
	alma_pri.csv	corresp_alma_pri.png
	alma_now.csv	corresp_alma_now.png
4.3.2	meth2_now.csv	corresp_meth2_now.png