第 2 章

はじめに

本調査における回答率は、第 1 章で示したように、必ずしも高いとはいえない。その結果、全体の約 30%の範囲しか説明できていない。一方で、回答者の傾向については、成績特に GPA との関連性が示されてきた。GPA が高い学生は回答し、逆に低い学生は回答しないという傾向である。これら背景から、GPA 等の変数を用いて欠損値の補完を行い、補完前後でどのような違いが見られるかを検証した。またそのための手法が妥当な範囲で運用できているか、その評価を行った。

対象のデータおよび手法

欠損値の補完に用いる説明変数の探索

解析の対象は、1 年生から 4 年生までの学生に共通する Q1 から Q4 までの計 52 項目とした。そして 52 項目の回答結果を目的変数とし、それに強い影響を持つ説明変数を探索するため、ランダムフォレスト分類を行った。説明変数には入学年度、所属学部、所属学科、性別区分、入試区分、高校評定平均、高校コード、出身都道府県、高等学校ランク等の属性に関する 9 つの変数と、GPA、成績区分（秀・優・良・可・不可、認定、合格、保留、履修取消）の数、単位数の合計、得点の合計、総修得単位の 31 の変数を用い、合計で 40 の変数を設定した。

高等学校のランクについては、高校生の進学事情に詳しい A 社の作成する高校ランクを使用し、ランクの設定がされていないものには、便宜的に特定の数値を代入した。また、高校の評定平均値が欠損しているデータに対しては、学年と学科をグループとし、その平均値を欠損値に代入した。

カテゴリカルデータに関しては、ラベルエンコーディングを施した。高校コードと出身都道府県は One-Hot エンコーディングにより bool 型に変換を行った。ランダムフォレスト分類の計算には、Python の scikit-learn ライブラリ（バージョン 1.4.1）を使用し、パラメータの設定については、木の数 (n_estimator) は 1000 とし、その他はデフォルトとした。

52 項目の質問に対する変数重要度を計算し、重要度の高い上位 10 件の変数を抽出した。これらの変数を、目的変数と強い関連性を持つと判断し、欠損データの予測精度を高めるために選出した。

抽出した説明変数
GPA，高校ランク，入学年度，学年コード，成績区分 1:SA(秀)の単位数合計，成績区分 1:SA(秀)の得点合計，成績区分 1:SA(秀)の数，成績区分 2:A(優)の単位数合計，成績区分 2:A(優)の得点合計，成績区分 2:A(優)の数，成績区分 3:B(良)の単位数合計，成績区分 3:B(良)の得点合計，成績区分 3:B(良)の数，成績区分 4:C(可)の単位数合計，成績区分 4:C(可)の得点合計，成績区分 4:C(可)の数，成績区分 5:D(不可)の単位数合計，成績区分 5:D(不可)の得点合計，成績区分 5:D(不可)の数，所属 2 コード，所属 3 コード，総修得単位数，高校評定平均

多重代入法の計算と欠損値補完

上記で抽出した変数を多重代入法による計算に組み込み、欠損値の補完を実施した。多重代入法は欠損データを扱う手法で、欠損値を複数の異なる値で補完し、それぞれの補完されたデータセットに対して解析を行い、その結果を統合する手法である。欠損がランダムに発生している場合に効果的に機能するとされているが、目的変数に影響を与える説明変数をモデルに組み込むことでバイアスを軽減し、信頼性の高い推論を行える。

今回、多重代入法の計算モデルには、ベイジアンリッジ回帰を使用した。アンケートの回答は4件もしくは5件法による順序尺度で作られているため、補完された値は四捨五入を行い、整数値に置き換えた。

結果

図1から4までは、欠損値を補完した前後の分布を可視化したもので、値は回答を単純平均したものである。図1では、元のデータセットと似た分布を持っており、極端な誤差は含んでいないものと考えられる。もっとも大きく変動しているのは、「履修登録時には、自分の学科の4年間のカリキュラムをよく確認した」で、10パーセント程度の開きが見られる。

図2では、図1と同様に元のデータセットに沿った分布を維持しているが、多くの項目で補完後に高い値が示されている。図3は、項目ごとで、もとの分布を維持しているものと、大きくかわったものとで分かれた。Q03_IT03, IT07からIT09までは、補完後の値は小さくなっている。

図4では、ほとんどの項目で補完後の値は小さくなっており、もとのデータセットの分布から大きく形を変えている。

考察および妥当性の検証

図1から図4を視覚的に評価すると、学業に強く関連する質問項目においては、欠損値補完の前後で値は変化するが、補完前の分布を概ね保持していることが認められた。一方で、学業に関係のない質問項目については、大きく変動が認められた。

これらの結果をうけ、妥当な範囲で適用されているかどうか確認するため、5分割クロスバリデーションを実施し、RMSE (Root Mean Square Error) により評価を行うこととした。RMSEは値が小さければ、モデル誤差が小さいことを示している。

この結果、RMSEの値はQ01で約0.42、Q02で約0.49、Q03で約0.63、Q04で約0.70であった。適切な評価はスケールを踏まえて個々の質問項目ごとに考える必要があるが、およそQ03、Q04では大きな誤差が見られたと評価できる。モデルに投入した変数はGPA等の成績であるため、環境や学生生活全般を扱う質問項目に対し、予測精度が低くなるのは妥当だと考えられる。

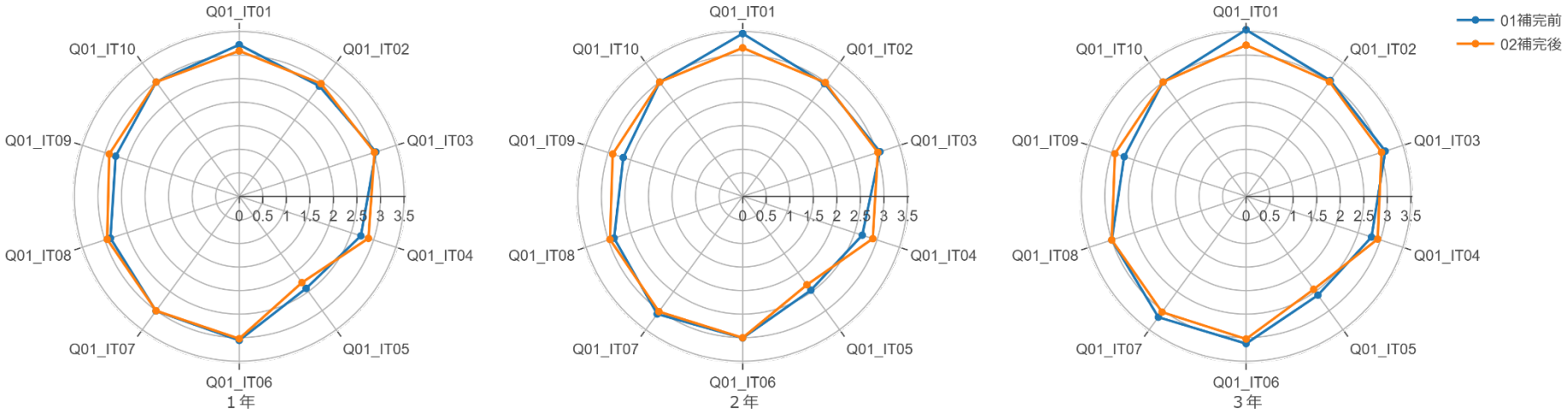
結論

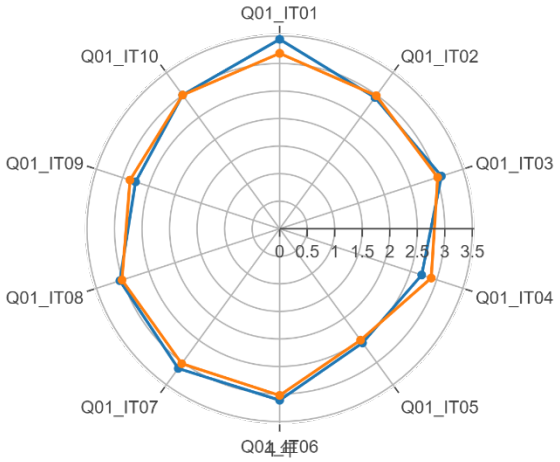
GPA等の変数を使用した多重代入法モデルは、学業に関する質問項目への適用時には有効に認められた。また適用の結果、数パーセントから最大で10パーセント程度の範囲で分布の変化が認められた。しかし、環境や学生生活全般に関する質問項目に対しては、モデルは適切に機能せず、

顕著な誤差が認められた。これらの質問項目に対して欠損値の補完を検討する場合、適用する変数を見直し、モデルの改良を行う必要がある。

図 1「Q01　あなたは、この 1 年間、どのような学び方をしてきましたか」の欠損値補完前後の分布

回答：“全くあてはまらない”から“とてもあてはまる”の 4 段階の順序尺度）

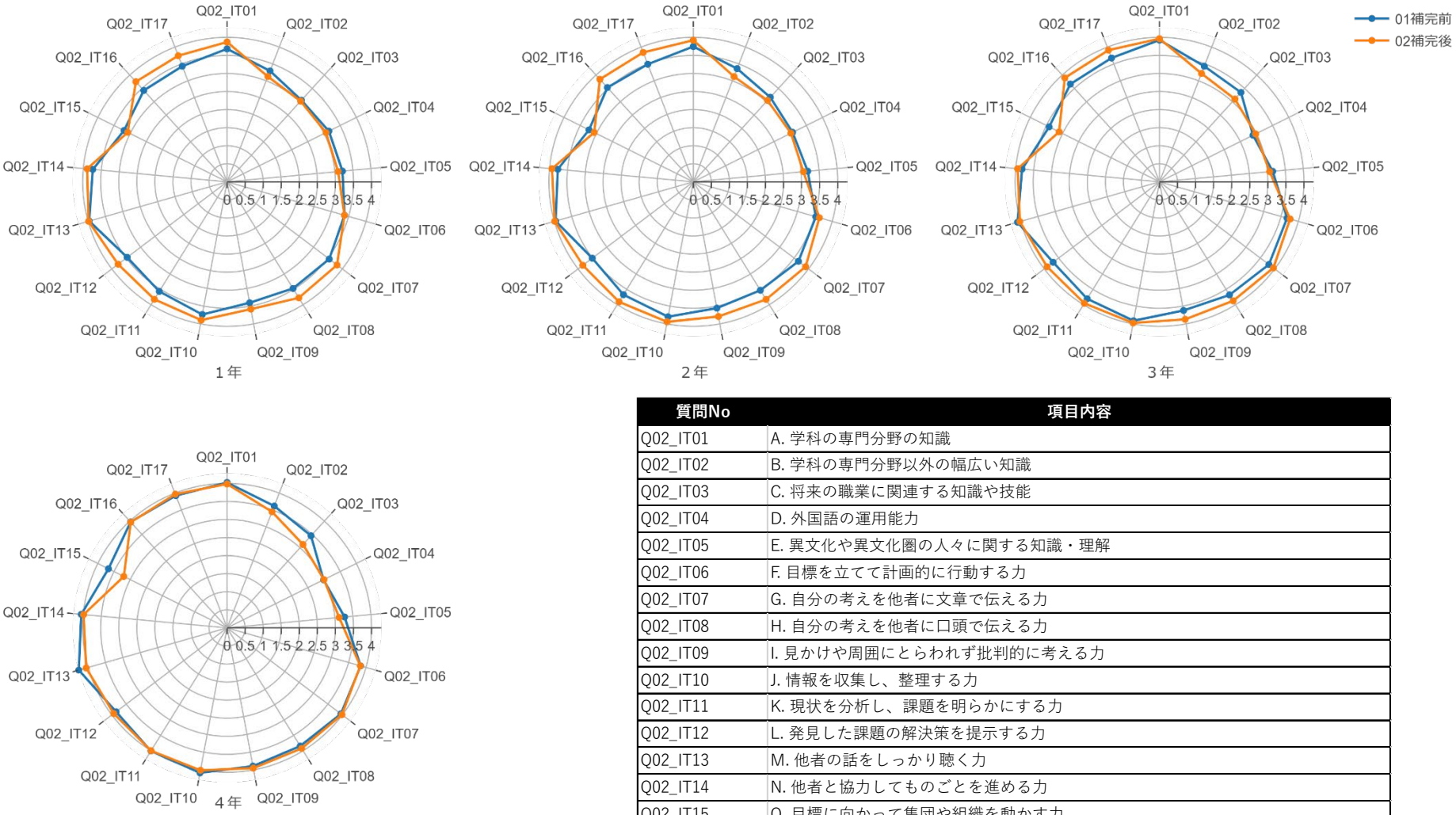




質問No	項目内容
Q01_IT01	01. 履修登録時には、自分の学科の4年間のカリキュラムをよく確認した
Q01_IT02	02. 授業は、その科目での到達目標を意識しながら受講した
Q01_IT03	03. 授業の受け方(ノートの取り方など)に、自分なりの工夫をした
Q01_IT04	04. グループ学習の機会があるときは、よく発言するほうだった
Q01_IT05	05. わからないことがあるとき、授業後などに先生に質問した
Q01_IT06	06. 授業時間外で、勉強の内容に関する調べ物をした
Q01_IT07	07. 課題や試験勉強には計画的に取り組んだ
Q01_IT08	08. 発表を行うときは、質疑応答などに備え、広く情報収集をした
Q01_IT09	09. 授業時間外で、授業等で学んだことをもとに友人と意見交換や議論をした
Q01_IT10	10. どちらかというと、学業以外の活動に注力していた

図２「Q02　あなたは大学入学から現在までに、以下の知識・能力をどのくらい身に付けることができたと実感していますか」の欠損値補完前後の分布

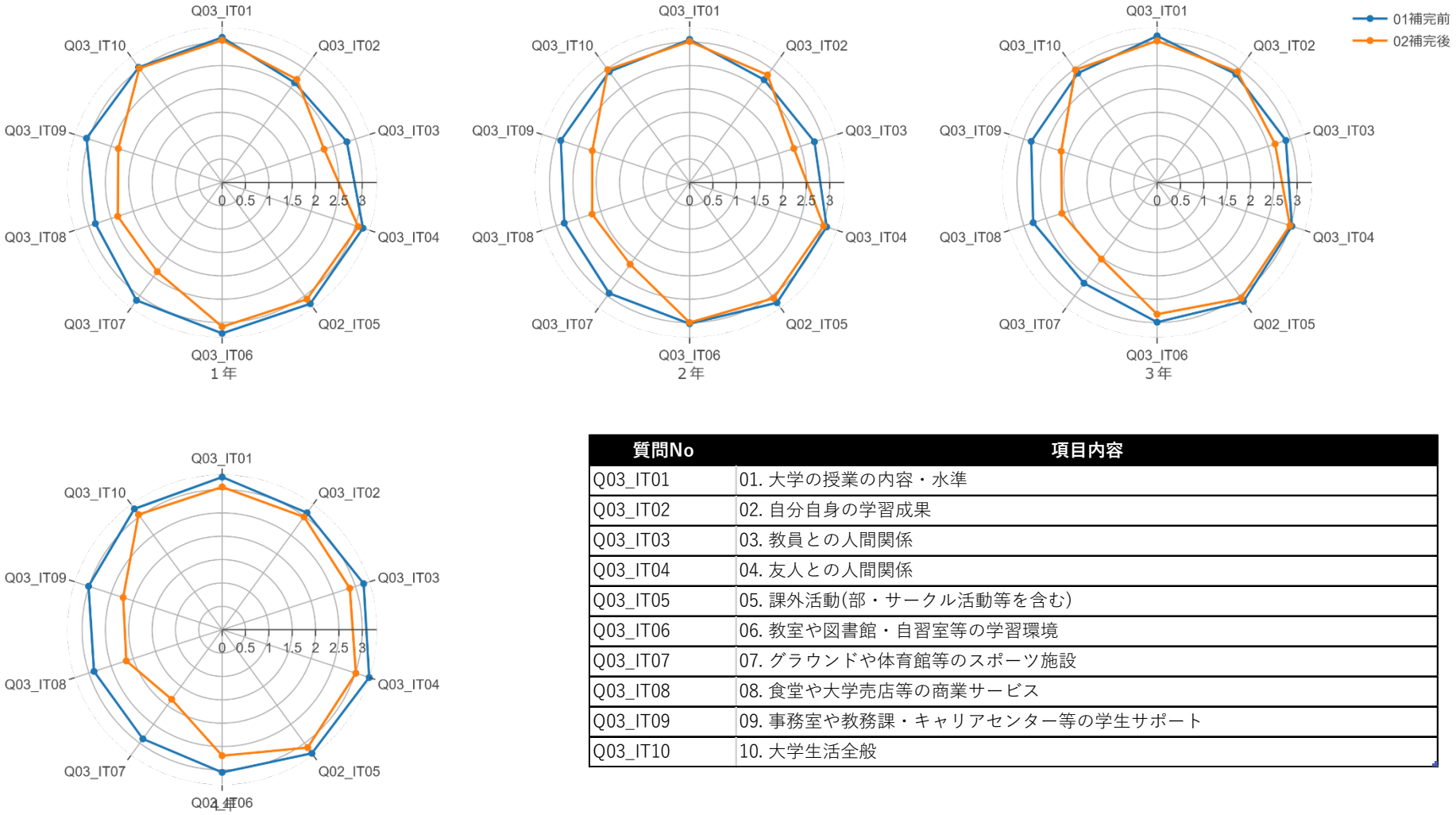
回答：回答は“全く身につかなかった”から“しっかり身についた”の５段階の順序尺度



質問No	項目内容
Q02_IT01	A. 学科の専門分野の知識
Q02_IT02	B. 学科の専門分野以外の幅広い知識
Q02_IT03	C. 将来の職業に関連する知識や技能
Q02_IT04	D. 外国語の運用能力
Q02_IT05	E. 異文化や異文化圏の人々に関する知識・理解
Q02_IT06	F. 目標を立てて計画的に行動する力
Q02_IT07	G. 自分の考えを他者に文章で伝える力
Q02_IT08	H. 自分の考えを他者に口頭で伝える力
Q02_IT09	I. 見かけや周囲にとらわれず批判的に考える力
Q02_IT10	J. 情報を収集し、整理する力
Q02_IT11	K. 現状を分析し、課題を明らかにする力
Q02_IT12	L. 発見した課題の解決策を提示する力
Q02_IT13	M. 他者の話をしっかり聴く力
Q02_IT14	N. 他者と協力してものごとを進める力
Q02_IT15	O. 目標に向かって集団や組織を動かす力
Q02_IT16	P. 自分の適性や能力を把握する力
Q02_IT17	Q. 広い視野から人間を探究する力

図 3「Q03　あなたは、この 1 年間の環境や学生生活にどの程度満足していますか」の欠損値補完前後の分布

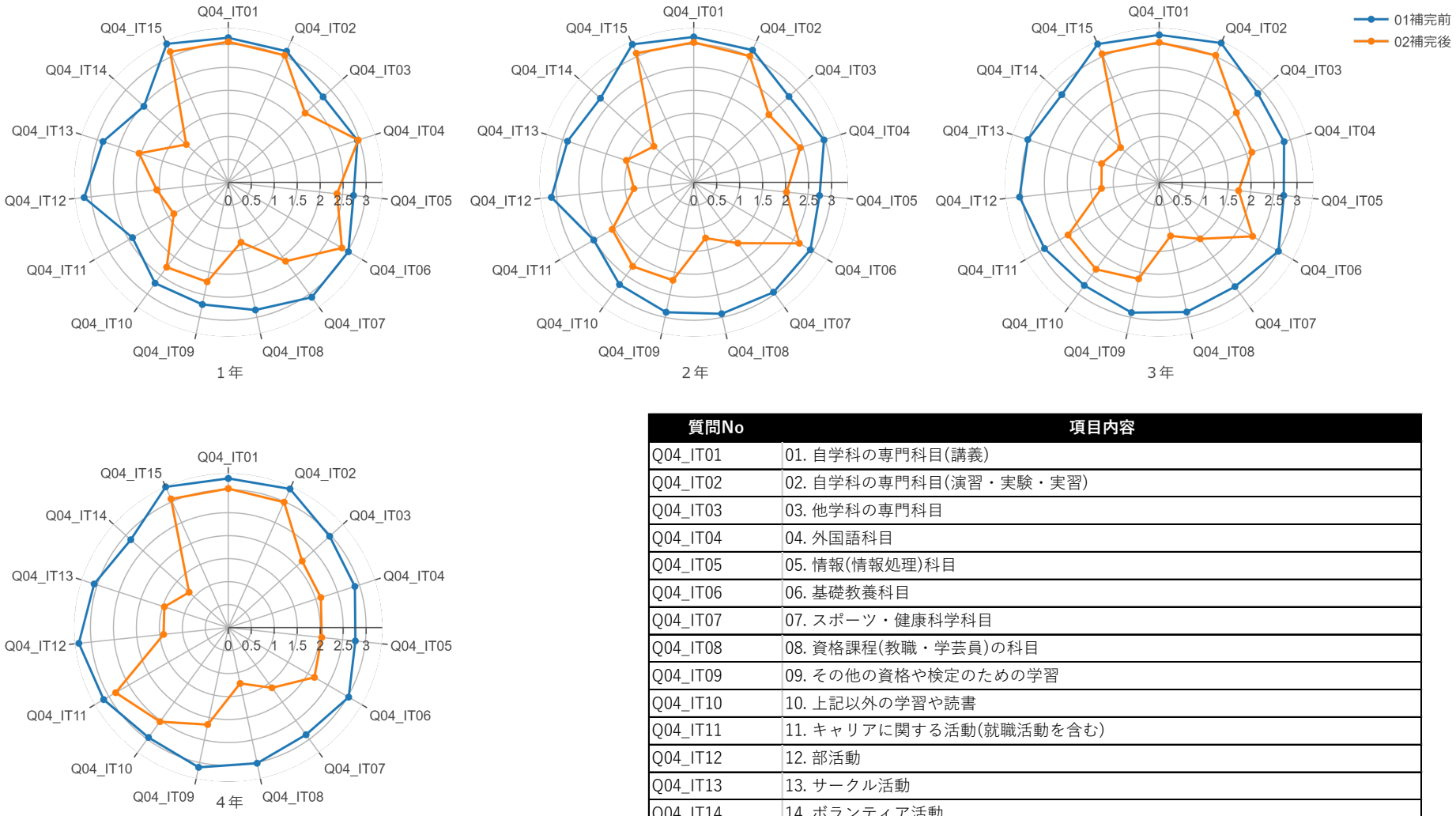
回答：“全く満足していない”から“とても満足している”の 4 段階の順序尺度



質問No	項目内容
Q03_IT01	01. 大学の授業の内容・水準
Q03_IT02	02. 自分自身の学習成果
Q03_IT03	03. 教員との人間関係
Q03_IT04	04. 友人との人間関係
Q03_IT05	05. 課外活動(部・サークル活動等を含む)
Q03_IT06	06. 教室や図書館・自習室等の学習環境
Q03_IT07	07. グラウンドや体育館等のスポーツ施設
Q03_IT08	08. 食堂や大学売店等の商業サービス
Q03_IT09	09. 事務室や教務課・キャリアセンター等の学生サポート
Q03_IT10	10. 大学生活全般

図４「Q04　あなたは、今年度の授業期間中、大学の授業やその他の学習などにどのくらい意欲的により組みましたか」の欠損値補完前後の分布

回答：“全く意欲的でなかった”から“とても意欲的だった”の 4 段階の順序尺度



質問No	項目内容
Q04_IT01	01. 自学科の専門科目(講義)
Q04_IT02	02. 自学科の専門科目(演習・実験・実習)
Q04_IT03	03. 他学科の専門科目
Q04_IT04	04. 外国語科目
Q04_IT05	05. 情報(情報処理)科目
Q04_IT06	06. 基礎教養科目
Q04_IT07	07. スポーツ・健康科学科目
Q04_IT08	08. 資格課程(教職・学芸員)の科目
Q04_IT09	09. その他の資格や検定のための学習
Q04_IT10	10. 上記以外の学習や読書
Q04_IT11	11. キャリアに関する活動(就職活動を含む)
Q04_IT12	12. 部活動
Q04_IT13	13. サークル活動
Q04_IT14	14. ボランティア活動
Q04_IT15	15. アルバイト