

令和 5（2023）年度 学習院大学 在学生調査

第 2 章

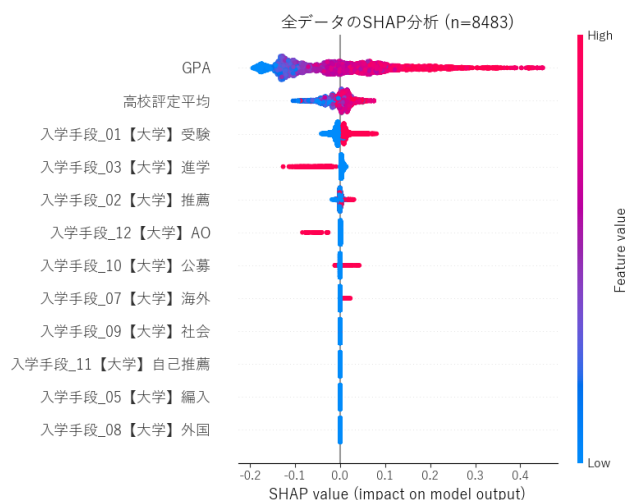
はじめに

高等教育機関における Institutional Research (IR) の実践において、学生を対象とした調査は、学修成果の把握および教育プログラムの改善の手段として広く採用されている。これらの調査分析では、回答と回答者属性情報を統合した分析が一般的に行われているが、得られたデータの代表性や潜在的なバイアスへの評価や検証が実施されているケースは多くない。

在学生調査においても、回答者の分布に関して学業成績上位層への有意な偏りが存在する可能性が示されている。これは「GPA が高い」学生が回答の中心となっていて、結果として調査に回答しない学生層の意見が欠落することで生じる偏り（無回答バイアス）が発生しているというものである。そこで、第 2 章では、回答の是非に寄与する変数を把握することおよび、その変数を用いて欠損値を補完し、前後の比較検討を行うことを目的に分析を進めた。

図 1 のバイオリンプロットは回答の有無（無を 0、有を 1 に置換）を目的変数にしたときの SHAP（SHapley Additive exPlanations）値を表したものである。SHAP 分析は、一般的に機械学習モデルを構築（あるいは解釈）する場合に、各変数の予測に対する寄与度を測るため使われるが、ここでは簡易的に変数重要度を測るため用いた。図中の赤色は正の影響を、青色は負の影響を示している。

図 1 SHAP 分析の結果



最も強い影響力持つのは GPA で、GPA が高いほど正の影響が強い。つまり回答確率が上がるという傾向が見られる。逆に GPA が低いと、負の影響があり、回答確率が下がるという傾向が示されている。また、GPA に次いで、高校評定平均に関しても、回答確率への影響が示唆されている。

この偏りは、学生の回答（自己評価）をもとにして、学修成果を測る場合、大きな誤りにつながりうる。このような背景を踏まえ、本分析では収集されたアンケートの回答データ、強い関連を示す変数を用いて、無回答すなわち欠損データの推定および補完を試みた。

ただし、本分析の主眼は、バイアスの緩和と、欠損データを補完し、データセットから洞察を得ることである。探索的解析の一環と捉えている点に留意されたい。

データならびに手法

対象のデータは、質問 1 から質問 4 までの回答、1 年生に限定したデータとした。比較的、回答率が高く（欠損の割合が低い）補完の精度が期待できるということ、データ量が多いと、計算時間が膨大になるという理由で対象を限定している。

欠損値の補完方法は、昨年度に引き続き多重代入法（Multivariate Imputation by Chained Equations : MICE）を用いた。MICE は欠損値を単一の値で埋めるのではなく、複数の可能な値のセットを生成することで、欠損値の不確実性を考慮する手法である。ランダムに欠損が発生している場合に効果的とされているが、目的変数に影響を与える変数を補助的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。回答の有無と GPA 等との関連性が示されていることから、ランダムでない欠損パターン（Missing not at random : MNAR）と解釈し、補助的な変数モデルへ組み込み、補完を行うこととした。

欠損値の補完には、R の Mice パッケージを活用した。Mice では順序プロビットモデル（ordered probit model）を使用した。昨年度は、Python と scikit-learn で補完を行ったが、順序尺度を直接扱えるライブラリがないこと、計算コスト（時間）の制約があることから、便宜的に連続値として扱った後に四捨五入し、整数値に置換するプロセスを挟んでいた。データ型に対して、適切なモデルを使うという意図で Python から R に変更している。

欠損値補完のための補助的な変数として、GPA、評定平均のデータを使用した。ただし外国人学生等を対象とする一部入試区分においては、評定平均がそもそも存在しないため、便宜的に学科と学年でグループ化した上で中央値による補完を行った。

Mice パッケージを用いて 20 個の補完済みデータセットを生成し、それらの平均値を最終的な補完結果として使用した。複数のデータセットの平均を取り、補完の前後を比較した。

結果

昨年度の分析と同様に、質問 1 から質問 4 までの回答についてレーダーチャートを作成し、補完前と補完後の回答分布を比較した（図 2 から図 5）。補完前後での回答分布の変化から、一部の質問項目で大きな変動が示された。

図 2 質問 1 における補完前後の比較

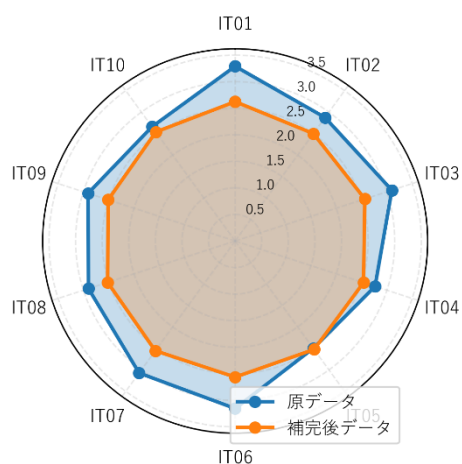


図 3 質問 2 における補完前後の比較

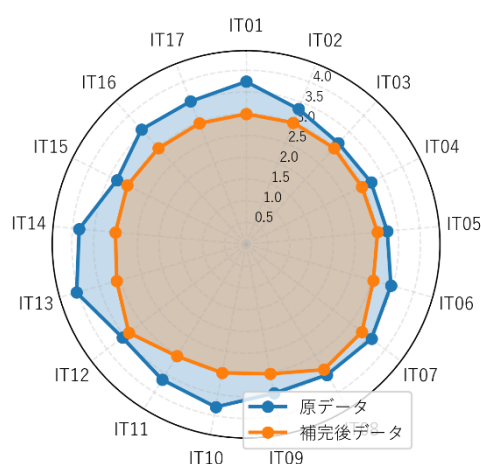


図 4 質問 3 における補完前後の比較

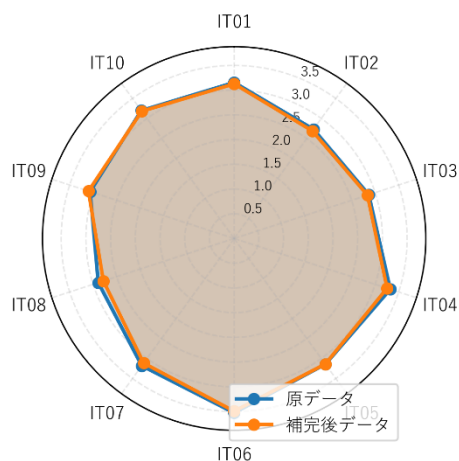
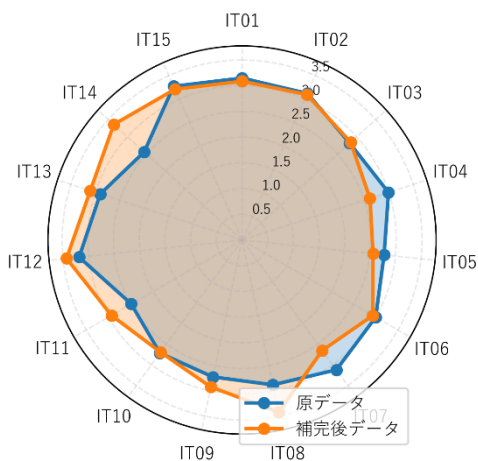


図 5 質問 4 における補完前後の比較



考察

欠損値補完の検証を行うため、20 個のデータセットに対して、多重代入法の提唱者である Rubin のルールに基づく統合結果の分析と感度分析をおこなった。

Rubin のルールに基づく多重代入法の統合結果を表 1 および表 2 に示す。グループ別の

分析結果から、全グループにおいて平均値の推定は比較的安定しており、標準誤差は 0.0266 から 0.0317 の範囲内に収まっている（表 1）。特に質問 1 では推定値の変動が最も小さく（Mean = 2.54 ± 0.0266）、質問 2 では最も高い平均値を示した（Mean = 3.08 ± 0.0317）。

Within 分散（実データのばらつきを表す分散）は全グループで 0.000480 から 0.000800 の範囲を示し、Between 分散（複数回の欠損値補完によって生じる推定値間のばらつきを表す分散）は 0.000150 から 0.000280 の範囲であった。Between 分散と Within 分散の比率（B/W Ratio：補完による不確実性がデータ本来のばらつきに対してどの程度の割合かを示す指標）はグループ間で大きく異なり、質問 2 で最も低い値（22.06%）、質問 3 で最も高い値（58.33%）を示した。これは質問 3 において欠損値の補完による不確実性が他グループと比較して相対的に大きいことを示唆している。

各グループ内の項目別分析（表 2）では、質問 3_IT06 が最も高い平均値（3.49 ± 0.0238）を示し、質問 1_IT09 が最も低い平均値（2.51 ± 0.0259）を示した。また、項目別の B/W 比率は 11.11%（質問 2_IT17）から 57.14%（質問 2_IT08）と広範囲に分布しており、欠損値補完の安定性が項目によって異なることが示された。

総じて、欠損値補完は比較的安定しており、特に質問 1 および質問 2 では補完による不確実性の影響が限定的であることが確認された。ただし、質問 3 においては欠損値補完による不確実性がやや高く、このグループの結果解釈には注意が必要である。

表 1 Rubin のルールに基づく統合結果：グループ別サマリー

Group	Mean ± SE	Within Var	Between Var	B/W Ratio (%)
Q01（質問 1）	2.54 ± 0.0266	0.000550	0.000150	27.27
Q02（質問 2）	3.08 ± 0.0317	0.000800	0.000176	22.06
Q03（質問 3）	3.06 ± 0.0275	0.000480	0.000280	58.33
Q04（質問 4）	2.99 ± 0.0303	0.000653	0.000267	40.82

表 2 Rubin のルールに基づく統合結果：特徴的な項目

質問項目	Mean ± SE	Within Var	Between Var	B/W Ratio (%)
Q01_IT01	2.62 ± 0.0286	0.0007	0.0001	14.29
Q01_IT09	2.51 ± 0.0259	0.0005	0.0001	20.00
Q02_IT08	3.39 ± 0.0342	0.0007	0.0004	57.14
Q02_IT17	2.98 ± 0.0326	0.0009	0.0001	11.11
Q03_IT06	3.49 ± 0.0238	0.0004	0.0002	50.00
Q03_IT02	2.69 ± 0.0275	0.0006	0.0001	16.67
Q04_IT12	3.43 ± 0.0306	0.0007	0.0003	42.86
Q04_IT05	2.56 ± 0.0290	0.0007	0.0002	28.57

感度分析は変動係数（Coefficient of Variation : CV）の比較を行った。CV は標準偏差を平均値で除したもので、異なるデータで変動の程度を比較するときに用いる。

なお全ての変数において CV_mean（変動係数の平均）の値は 0.002766 から 0.014834 の範囲に分布し、不確実性の程度に変数間で差が見られた。全ての変数において Violation（代入値における論理的な違反）は 0 であった。Violation とは、例えば「選択肢に存在しない値が代入される」など、変数の性質や定義域に反する不適切な代入が行われた回数を示す指標である。本分析では Violation が一切検出されなかったことから、代入プロセスは変数の基本的な特性や制約条件を適切に保持できていたと評価できる。

感度分析の結果、質問 1 および 2 については、いずれも CV は非常に低い値を示している（図 7 および図 8）。このことから、質問 1、質問 2 については、補完を行った後のデータ分布が統計的に安定しており、妥当性が確保されていることが示された。

質問 3 および質問 4 の項目については、質問 1、質問 2 と比較して CV が相対的に高い値を示している。特に質問 4 の複数項目（IT08: 0.014834、IT11: 0.011519、IT13: 0.011228、IT14: 0.011324）で CV が 0.01 を超えている（図 9 および図 10）。Violations は 0 であるものの、これらの項目では補完後のデータの変動が大きいことが示唆されている。

図 6 質問 1 から 4 までの変動係数（平均値）比較

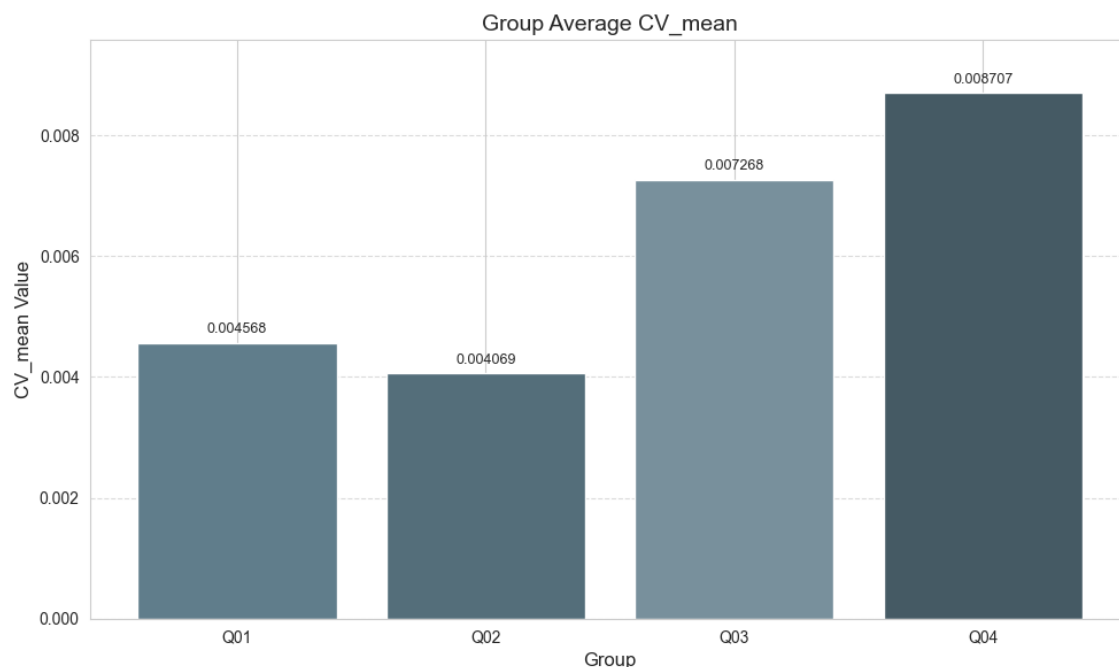


図 7 質問 1 の各項目の変動係数（平均値）

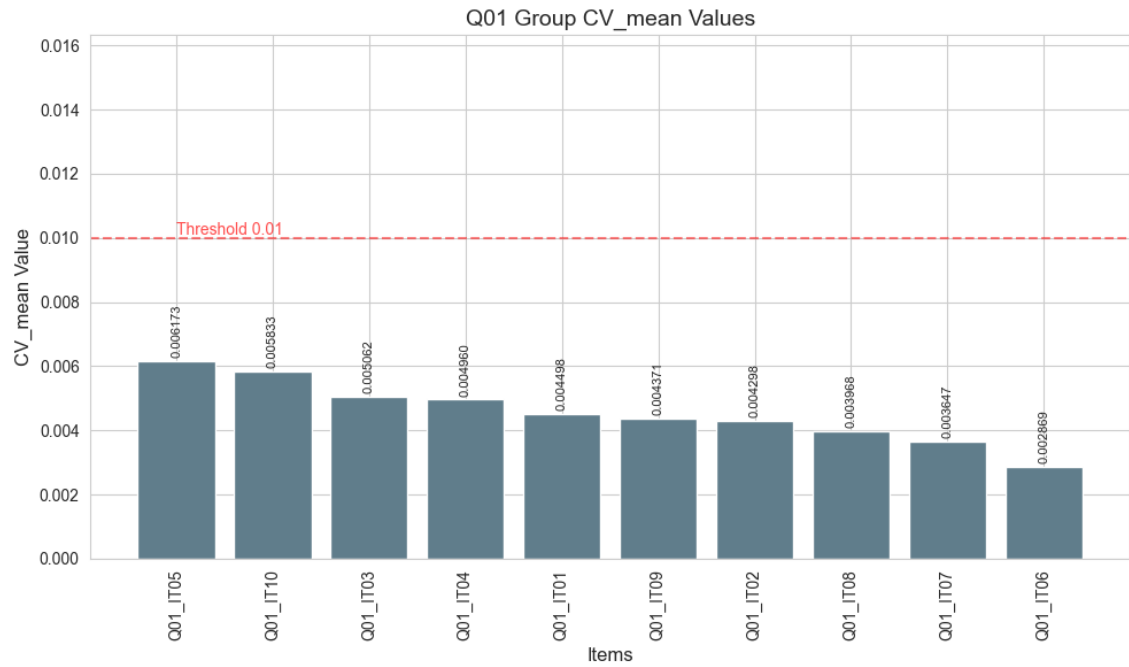


図 8 質問 2 の各項目の変動係数（平均値）

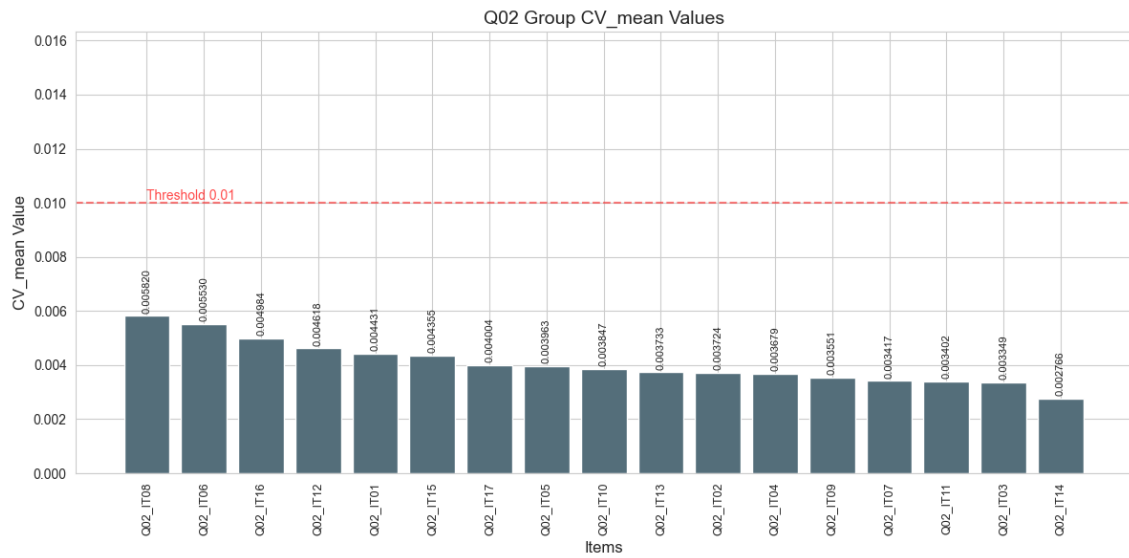


図 9 質問 3 の各項目の変動係数（平均値）

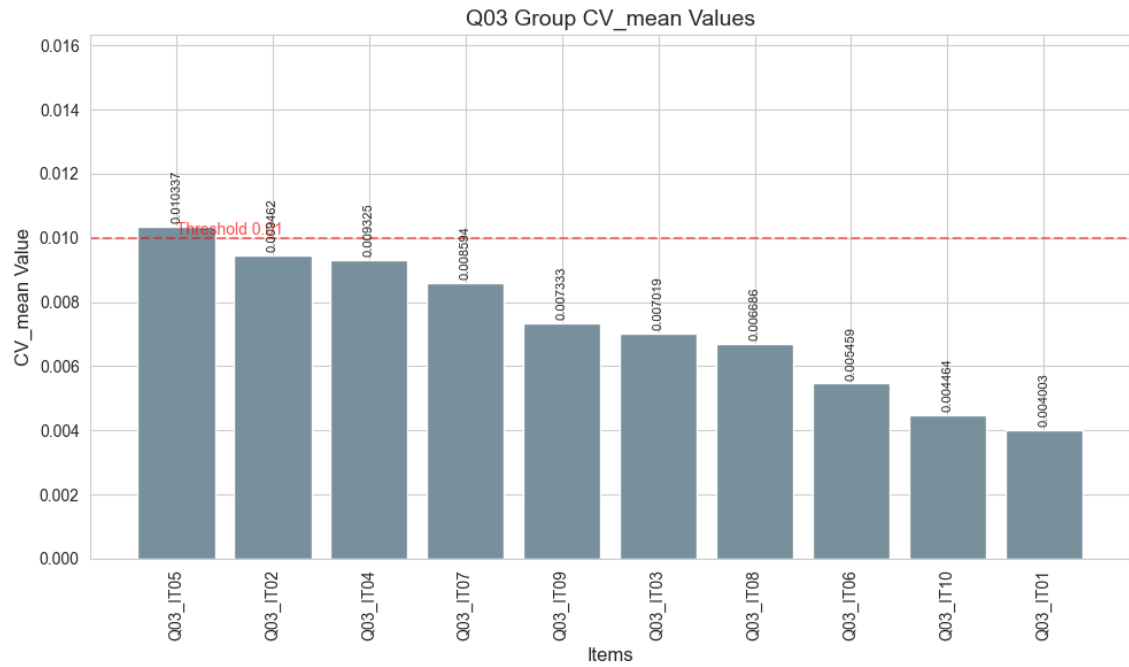
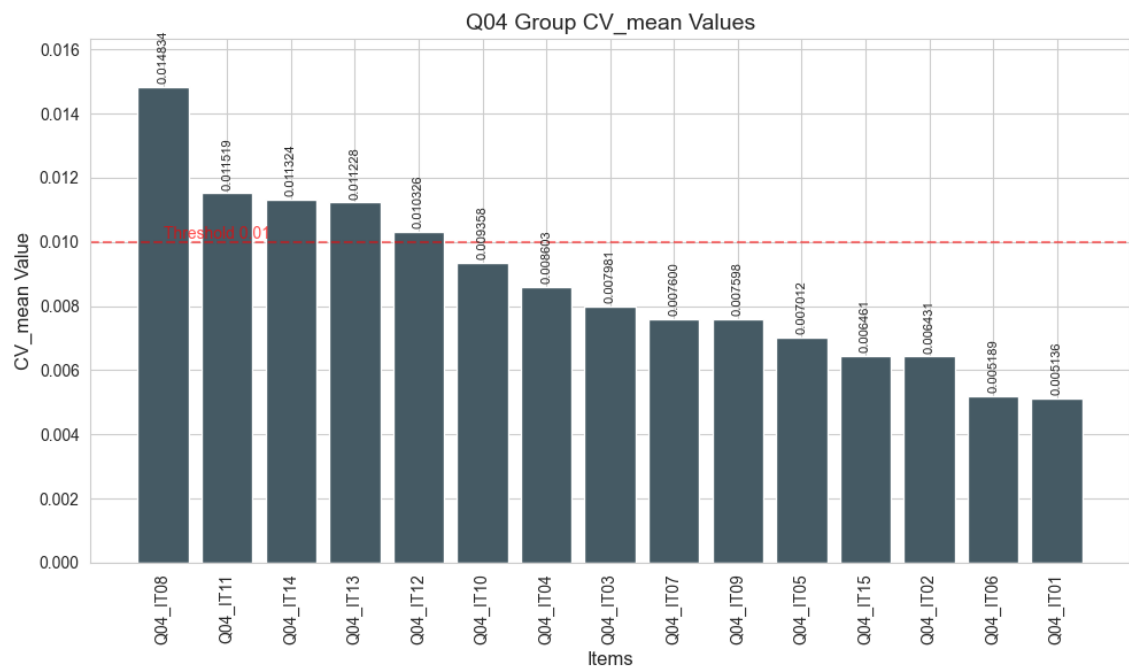


図 10 質問 4 の各項目の変動係数（平均値）



結論

多重代入法と感度分析の結果、GPA と評定平均値を基にした欠損値補完は、質問 1 および質問 2 の各項目において効果的に機能したと判断される。例えば質問 1_IT01（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）の平均値が補完前の 3.29 から補完後 2.62 へ変化したことは妥当であることが裏づけられた。質問 1 項目群への評価は、実測値と比較して最大約 20% の下方修正が生じると推定される。質問 2 項目群においては、質問 2_IT08（自分の考えを他者に口頭で伝える力）の評価は保留すべきだが、その他の項目は約 5%～25% 程度の下方修正が妥当と評価できる。

一方、質問 3 および質問 4 項目群については、B/W 比率と変動係数が相対的に高値を示しており、結果の解釈には慎重な検討が必要である。同時に、これらの項目群に対しては GPA などの成績データのみでは十分な欠損値補完が困難であることが示唆された。

課題や展望

本解析は探索的解析の一環として実施しており、欠損値の補完結果については、その精度や信頼性に関して最終的な結論を導く段階には至っていない点を強調したい。

Rubin のルールに基づいた統合結果が示す「精度の高さ」は、欠損値補完後の各データセット内や、補完結果同士のばらつきが非常に小さいこと、つまり統合推定値の不確実性が低いことを意味するが、真の値との一致については検証することが不可能である。

また、感度分析の結果が示すのはあくまで各疑似データセット間の統計的なばらつきであり、補完推定の精度を直接示すものではない点にも留意しなければならない。つまり「感度分析により得られた変動係数が低い」ということが統計的な安定性を示しているにすぎず、補完の正確性を担保するものではない。

さらに、本解析は初期段階の探索的検討であるため、補完手法の選択やパラメータの設定が結果に与える影響についても十分に検証する必要がある。特に計算コストの制約が強く、1 学年、20 個の推定データセットを作成するのに、約 15 時間を要している。作成する推定データセットを増やす、変数間の相互作用を想定したモデルを使う、予測に影響を持つ重要な変数を選定する、他の補完手法との比較検討を通して、妥当性や結果の安定性を確認する必要がある。