

職業調査において対象職業の就業者ではない回答者の  
混入を統計学的に検出する手法の検討  
——疑似混入シミュレーションによる3手法18条件の比較——

独立行政法人 労働政策研究・研修機構

研究員 鎌倉哲史

《要旨》

厚生労働省が2020年3月より公開している職業情報提供サイト(日本版O-NET)の職業横断的な数値情報はJILPTがWebモニター調査の結果に基づき作成しているが、自己申告形式のため対象職業の就業者ではない回答者が混入する懸念がある。これまでJILPTでは一定のスクリーニングは実施してきたが、より体系的な混入検出手法としてスキル、知識等の数値項目の回答傾向の利用が考えられる。ただし手法の妥当性の評価のためには職業真偽が判明しているデータセットが不可欠となる。

そこで本稿ではJILPTが2019年度に関連の省庁や団体に直接依頼にて数値情報のデータを収集した5職業をそれぞれ「真の回答者群」と設定し、意図的に「偽の回答者」1名分の回答を混入させる疑似混入シミュレーションにより判別手法の妥当性を評価した。混入者の設定は「無気力ランダム回答」と「異職業の就業者」の2種を置き、それぞれ100名分の混入者リストを作成し順次混入させた。判別手法の候補は「外れ値」出現回数に基づく手法A、Bと標準化ユークリッド距離に基づく手法Cの3種とし、それぞれ有意水準3パターン、ロバスト推定の適用有無2パターンで6条件を設定した(計18条件)。

シミュレーションの結果、無気力ランダム回答については、一部例外を除き高い判定精度が達成され各手法の適用の有効性が示された。異職業の就業者については、再現性は維持できるが無気力ランダム回答と比べ特異度は劣っていた。検証した18条件の中ではA3条件(手法Aの片側1%水準、非ロバスト条件)が最も再現性を高く保持するが、特異度も考慮するとC3条件(手法Cの片側1%水準、非ロバスト条件)も望ましい精度を持つことが示唆された。最後に本稿の課題と制約をまとめた。

---

(備考) 本論文は執筆者個人の責任で発表するものであり、独立行政法人 労働政策研究・研修機構としての見解を示すものではない。

## 目次

1. 問題意識 .....	1
2. 目的 .....	5
3. 方法 .....	5
3.1. 使用したデータセット.....	5
3.2. 使用した項目.....	6
3.3. シミュレーションの実行環境 .....	6
3.4. シミュレーションの枠組み .....	6
3.5. 比較対象とする 3 種の除外対象判別手法 .....	7
3.6. 各条件に対するロバスト推定の適用 .....	11
3.7. シミュレーション結果の評価指標 .....	12
4. 結果 .....	13
4.1. 各条件における「外れ値」出現回数、および標準化ユークリッド距離 ...	13
4.2. 無気力ランダム回答混入時の各条件における再現性と特異度 .....	16
4.3. 異職業の就業者混入時の各条件における再現性と特異度 .....	17
4.4. 手法 B 非ロバスト条件へのポアソン分布有意水準適用時の再現性と特異度 .....	18
5. 考察 .....	19
引用文献.....	23
付録 1. 使用した 110 項目の各群における平均値、標準偏差、尖度、歪度.....	25
付録 2. ランダム抽出 100 ケースの職業.....	28
付録 3. 手法 A、B のシミュレーションに使用した R スクリプト.....	29
付録 4. 手法 C のシミュレーションに使用した R スクリプト.....	34

## 1. 問題意識

2020年3月より厚生労働省が公開している職業情報提供サイト（日本版 O\*NET）では提供コンテンツの1つとして職業横断的な数値情報を掲載している。この数値情報は米国労働省が1998年より公開している Occupational Information Network（O\*NET）の調査項目を参考として、労働政策研究・研修機構（以下、「JILPT」という）が主に Web 調査会社のモニターを対象とした就業者調査（以下、「JILPT の調査」という）に基づき開発している（労働政策研究・研修機構, 2020）。しかし、当初開発にあたっては2年間で約500の職業の情報を速やかに開発する必要があったこともあり、その調査方法やデータ整備のプロセスには課題が多く残されている。代表的な課題としてはカバレッジ誤差の懸念、Web モニター特有のサンプリングバイアスの懸念<sup>1</sup>、対象職業の就業者ではない回答者が混入する懸念、スキルのレベル評価におけるアンカー設定の妥当性検証、過大な回答者負担の軽減等が挙げられている（詳細は Kamakura, Matsubara, & Matsumoto, 2020 を参照）。

これらはいずれも重要な課題であるが、本稿では特に対象職業の就業者ではない回答者が混入する懸念に焦点を当てる。米国 O\*NET の場合、原則として企業に調査協力を依頼し企業から各職業の就業者に調査票を配布してもらう二段階抽出を採っており（U. S. Department of Labor, 2018）、回答者の職業はある程度客観的に保証されている<sup>2</sup>。しかしこれは専属の組織（国立 O\*NET 開発センター）の指揮下で毎年約500万ドル（≒約5億円<sup>3</sup>）のプロジェクト予算を投じて初めて実現されているものであり（労働政策研究・研修機構, 2018）、我が国で同様の方法を採用することは難しい。そこで JILPT の調査では事前に Web 調査会社にモニター登録をしている人々に自身の職業を選択して回答してもらうことで調査費用を比較的 low コストに抑え数値情報を作成している。公的資金を原資とする事業という観点では low コストであることは大きなメリットであるが、反面、回答者の職業選択に裏付けはない。

もちろん我が国の他の職業調査でも自己申告形式は一般的であり、この意味で JILPT の調査方法が特段劣っているわけではない。たとえば国勢調査では職業に加えて本人の仕事内容を尋ねることで総合的観点から職業分類番号の割り当てを行っているが<sup>4</sup>、それでも各記入内容は自己申告に基づく。大規模な職業調査で個別に職業の裏付けを取る（e.g. 勤め先・取引先への問い合わせ、就業実績を示す資料の収集）

---

<sup>1</sup> たとえば Web モニター調査では回答者の学歴が高いこと、専門・技術職が多いこと等が先行研究から繰り返し指摘されている（e.g. 埴淵・村中・安藤, 2015）。

<sup>2</sup> なお米国 O\*NET の場合、就業者は郵送回答と Web 回答を自由に選択可能である。

<sup>3</sup> 1ドル100円換算の場合。

<sup>4</sup> 国勢調査2020の Web サイト上 FAQ「なぜ『本人の仕事の内容』の項目を調査するのか」について総務省は「就業者一人一人が勤め先の会社・工場・事務所などで実際にどのような仕事に携わっているかという、いわゆる『職業』を把握し、我が国人口の職業構造の実態を明らかにするためのものです。」と回答している。<https://www.kokusei2020.go.jp/about/faq.html>（2021年2月4日）

ということは現実的ではない。そこで、仮に一部に無回答、誤回答、無効回答、無気力回答等が含まれるとしても大半の回答者が誠実に回答するのであれば全体の傾向は把握可能であろうとの方針で調査が実施されることが一般的である。

ただし、JILPT の調査の場合には 3 つの観点で対象職業の就業者ではない回答者の混入の懸念が高い。第 1 に、Web モニター調査における「プロ回答者」の存在がある。本多（2006）によれば Web モニター調査では「謝金目当てで数多くの調査に回答しようとする”professional respondent”が参加しているために、短時間でいい加減に回答する者」がいるとの指摘がある（p.35）<sup>5</sup>。日本学術会議（2020）による Web 調査の学術目的利用に関する提言でも「satisfice 行動」<sup>6</sup>、すなわち「Web 調査の文脈では、質問を含む調査票の内容に十分な注意を払わず、素早く完遂すること（だけ）を目的として回答するような行動」を見せる回答者の存在に言及している（p.11）。

こうした「不真面目な回答」（日本学術会議，2020）を排除するため調査会社では様々な事後のスクリーニング手法を取り入れており、また当該ケースのモニターにはペナルティを与えることで回答品質の維持に努めている<sup>7</sup>。しかし、「可能な限り設問の読解や回答のための思考といった労力をカットしつつ、短時間回答にならないよう、かつランダムに回答する者」（以下、無気力ランダム回答者という）については調査会社、ならびに依頼元の調査者が事後に識別することは困難である。実際、Web モニター調査のバイアスについて多角的に検討した埴淵・村中・安藤（2015）でも、不良回答<sup>8</sup>の操作的定義の中で「ランダムに回答した場合」は判別できていないとしている（p.89）。こうした事態を想定して事前に調査計画の段階で設問文を読まない回答者を検出するための項目を挿入しておく IMC<sup>9</sup>、DQS<sup>10</sup>といった方法論も提唱されているが（尾崎・鈴木，2019）、三浦・小林（2016）によれば「正しく答えない」ことを求める設問の設定はモニターの信頼を損ねるとして調査会社が断る場合があり（p.2）、また項目数を増やすことで回答者負担の増加、調査コストの増大と

<sup>5</sup> Web モニターの募集広告等では空き時間に収入が得られる点を強調している場合もあり、こうした経済的合理性に従うモニターの存在は否定できない。

<sup>6</sup> Satisfice とは「satisfy と suffice の 2 単語の組み合わせに由来する造語で、『努力の最小化』などと訳される」（日本学術会議，2020，p.11）。

<sup>7</sup> 具体的には極端に回答時間が短いケースや自由記述内容が記号等無意味な内容であるケースを除外するほか、依頼元の調査者と協議しつつ設問内の回答がすべて同値といった単調回答のケース、複数の回答間で論理的な矛盾があるケース、選択式設問において規則的回答のケースを除外するという対策も選択し得る。なお、ここで例示している内容は埴淵・村中・安藤（2015）を参考としつつ、筆者自身が過去の Web モニター調査にて調査会社から聞いた対策内容もまとめている。

<sup>8</sup> 埴淵・村中・安藤（2015）では不良回答を「質問文をよく読まないで回答する」とこととしている（p.85）。

<sup>9</sup> IMC（Instructional Manipulation Check）とは「リッカート法や複数選択式の設問」に、「通常通りの回答をしないように求めるメッセージを教示に含めたもの」である（尾崎・鈴木，2019，p.39）。

<sup>10</sup> DQS（Directed Questions Scale）とは「リッカート法による項目群に『この項目では一番左（右）を選択してください』という内容の項目」を混ぜた設問を指す（尾崎・鈴木，2019，p.39）。

いった欠点も生じる（尾崎・鈴木，2019）。

第2に、「職業選択画面」における選択プロセスの難しさがある。高橋（2018）によれば、職業調査において選択式を採ると選択肢が多くなってしまいうため、自由記述で収集して事後に研究者がコードに変換する作業を行われる場合がある（e.g. 国勢調査）。しかし自由記述のコーディングは大規模調査の場合、「コードに膨大な負担がかかる」ほか、コードが分類体系を熟知していないための誤ったコード付与等のリスクがある（高橋，2018）。また Web モニター調査で自由記述のアフターコーディングを行う場合、職業別の目標回収数の管理が実現できない。そこで JILPT の調査では回答者に約 500 の調査対象職業リストの中から選択式で自分の職業を見つけてもらう方式を採っている<sup>11</sup>。このため大量のリストの中から自分の職業を探すというプロセスの性質上、本人が意図しない誤選択の懸念が発生する。たとえば 2018 年度調査では調査者が自由記述を目視チェックした結果、「調教師」の回答者に「調理師」が、「国際公務員」の回答者に「国家公務員」がそれぞれ混入していることが判明し、スクリーニングにて除外されている（労働政策研究・研修機構，2020，p.61）。こうした誤選択を軽減するため、2019 年度調査からは職業選択後に各職業の仕事の概要を表示させ、「この職業で間違いないか」を最終確認する画面遷移が採用された。また 2020 年度調査からは初めに職業のカテゴリを選択し、その後カテゴリ内で職業を選択するといった工夫が行われている<sup>12</sup>。それでも「職業選択画面」では常に意図しない誤選択の可能性や、自分の職業が見つからない時に謝礼目的で意図的に異なる職業、類似の職業を選択する可能性等が残されている。

第3に、サンプルサイズの小ささがある。JILPT の調査では対象職業数が多いため調査全体の規模は大きいものの、1 職業あたり目標回収数は 60 名とし、スクリーニング後の最終的なサンプルサイズが 20 名以上であることを Web サイトへの数値情報収録の基準としている（労働政策研究・研修機構，2020）<sup>13</sup>。これだけ小規模のサンプルサイズとなると本来その職業の就業者ではない人物が混入することで母集団推定において及ぼすリスクは大きくなる。

こうした対象職業の就業者ではない回答者の混入に関する懸念を可能な限り排除するため、JILPT では調査会社による短時間回答や単調回答等の除外に加えて、調査内容の1つである「タスク」群<sup>14</sup>において1つも実施タスクが無い、かつ「その

---

<sup>11</sup> なお、2018 年度と 2019 年度の当初開発ではそれぞれ半数となる約 250 職業ずつが調査対象とされた。2020 年度以降は調査項目数を大幅に減らし回答者負担を軽減しつつ、約 500 の職業を一括で調査対象としている。

<sup>12</sup> 詳細は今後公刊される見込みの 2020 年度開発報告書を参照されたい。

<sup>13</sup> 米国 O\*NET の場合、就業者調査において各職業で目標 20 件、最低 15 名としている（U. S. Department of Labor, 2018）。

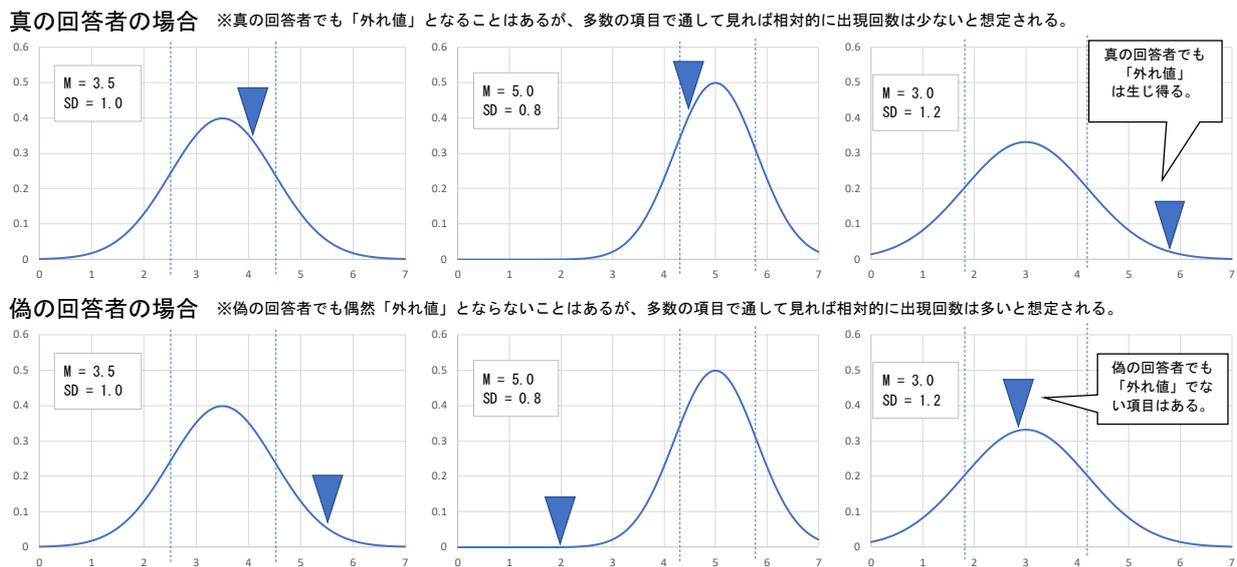
<sup>14</sup> 他の数値情報は全職業共通の内容だが、タスクのみ職業ごとに固有の内容について実施の有無を尋ねる形式となっている（労働政策研究・研修機構，2020）。

他のタスク」の自由記述もその職業の人物の記述として妥当とは言えないケースは、少なくとも調査者が想定しているその職業の人物ではないであろうと判断し調査後に除外している（労働政策研究・研修機構，2020）。ただし、タスクベースの方法には(1)タスクを毎回尋ねなければならないこと、(2)タスクの自由記述内容を人間の目でチェックする必要があること、(3)タスクリスト自体の妥当性もその都度検討が必要であること、といった課題がある。

そこで、調査後のより体系的な除外方法としてスキル、知識等の一定範囲で分布を持った数値項目の回答傾向を利用することが考えられる。たとえばある職業Aの回答者群の多くは実際にその職業の就業者（以下、「真の回答者」という）で、一部に対象職業の就業者ではない回答者（以下、「偽の回答者」という）が混じっていると仮定した場合、偽の回答者の回答傾向は群全体の傾向から乖離していることが予想される（図表 1）。もちろん、ある1つの変数だけを用いれば真の回答者が偶然乖離することもあれば、偽の回答者が偶然乖離しないこともあるため誤判定のリスクが大きいためであろう。しかし、50、100といった多数の項目の回答傾向を通して見れば、真の回答者の群全体からの乖離の程度や「外れ値」<sup>15</sup>の出現回数は偽の回答者より小さくなるものと期待される。この乖離の度合いを何らかの形で指標として抽出することで、事後に真の回答者と偽の回答者を判別できる可能性がある<sup>16</sup>。

**図表 1. 特定職業の回答者群における真の回答者と偽の回答者の回答傾向のイメージ**

(※グラフ、数値は全て架空のデータ)



<sup>15</sup> 「外れ値」とは「データのメインボディから外れている値」（藤澤，2017，p.1）、すなわちある集団の分布全体から見て大きく外れている値を指す。

<sup>16</sup> もちろん、偽の回答者が真の回答者と同じような回答傾向を示す場合にはこの方法では除外ができないが、偽の回答者が混入することの最大の懸念はそれによって本来推定されるべき平均値や比率が左右されることであり、真の回答者と同じような回答をする偽の回答者の混入は相対的に見て悪影響は小さいと言える。

この方法の利点は除外基準が就業者回答に応じて動的に変動すること、したがって偽の回答者が除外されないよう「正解」するためには、真の回答者と同じ程度に対象職業についてよく知っている必要があることである。もちろん真の回答者の中でも回答に大きなばらつきが出る項目はあるが、この場合はもとより分散が高くなるので「外れ値」の判定基準が厳しくなり合否判定にはあまり影響しない。逆に、たとえば外交官にとっての外国語スキル、医師にとっての医学知識のような真の回答者であれば小さい分散で固まるはずの項目が真偽の区別に大きく貢献することとなる。これは単調回答等の固定的な除外基準とは異なる、より柔軟な除外基準と言える。

しかし、実際には職業調査の事後に上記のような発想でスクリーニングを行っている先行研究は見られない。その背景には、仮にある指標に基づく判別手法を考案したとしても通常は調査者には最後まで回答者の職業真偽の「正解」が不明のままであるため、本当に除外したケースが偽の回答者で除外しなかったケースが真の回答者だったのかが分からず、その判別手法の妥当性を評価しようが無いことがある。埴淵・村中・安藤（2015）が指摘するように Web モニター調査において「不良回答」の懸念があるとしても「未知のバイアスを大きくしかねない回答除外には慎重であるべき」である（p.96）。ここで本稿では、JILPT が 2019 年度に関連の省庁や団体に直接依頼にて数値情報のデータを収集した 5 職業（後述）に注目する<sup>17</sup>。これらの職業は 2018 年度の Web 調査で十分なサンプルサイズが得られなかったため補足的に紙の郵送調査が実施されたもので（労働政策研究・研修機構, 2020）、米国 O\*NET の企業経由の調査と同じく、真の回答者で構成されていることがある程度客観的に保証されている。この 5 職業のデータの中に意図的に別の回答を混入させる疑似混入シミュレーションを行えば、判別手法の妥当性を評価することが可能となる。

## 2. 目的

前節で述べた問題意識から、本稿の目的は職業調査における対象職業の就業者ではない回答者を回答傾向に基づき事後に統計学的に除外する判別手法の妥当性を疑似混入シミュレーションによって評価することとする。

## 3. 方法

### 3.1. 使用したデータセット

JILPT が 2018 年度、2019 年度に Web 調査、および直接依頼調査にて収集した全

---

<sup>17</sup> 直接依頼の調査は 6 職業で実施されたが、直接依頼のみで最低 20 件を確保している 5 職業を本稿では対象とした。

ての職業の回答者 25,974 名のデータを使用した<sup>18, 19</sup>。

### 3.2. 使用した項目

職業情報提供サイト（日本版 O-NET）のために開発された数値情報には母集団の平均値を推定する平均値系と比率を推定する比率系があるが、本稿では平均値系の数値情報 110 項目（興味 6、価値観 10、スキル 39、知識 33、仕事の性質 22）を使用した<sup>20</sup>。後述の各群ごとに各項目の平均値、標準偏差、尖度、歪度を付録 1 に示す。折笠（2017）によれば歪度の絶対値が 1 以上のとき正規分布から「ひどく歪んでいる」、0.5 以上のとき「やや歪んでいる」と評価し得るが、この観点で本稿の使用項目は正規性が保たれていないと思われる項目が多い。こうした変数群を用いて正規性を仮定した各判別手法が判定精度を保てるかどうか、本稿の焦点の 1 つとなる。

### 3.3. シミュレーションの実行環境

R（version 4.0.3）にて、筆者が作成した 2 つのスクリプト（付録 3、4 参照）を用いて後述の検討対象条件に合わせてパラメータを書き換えつつシミュレーションを実行した。

### 3.4. シミュレーションの枠組み

2019 年度に関係省庁や関連団体への直接依頼により取得した 5 職業（麻薬取締官 28 名、入国警備官 25 名、検察官 20 名、家庭裁判所調査官 20 名、外交官 21 名）のデータを「真の回答者」だけで構成された回答者群と見なし、意図的に偽の回答者の回答を 1 名分混入させるという状況を各条件で 100 回繰り返し各判別手法の妥当性を評価した（評価指標は後述）。

その際、本稿では混入者の想定として「無気力ランダム回答者」と「異職業の就業者」の 2 パターンを置いた。無気力ランダム回答者については、各変数の取り得る範囲でランダムな数値を与えた 100 名分の架空のデータを作成し<sup>21</sup>、この混入者リストから順次混入させた（図表 2）。異職業の就業者については、全職業の回答者群の

---

<sup>18</sup> 短時間回答、単調回答、およびタスクに基づくスクリーニングが完了済みのケースである。

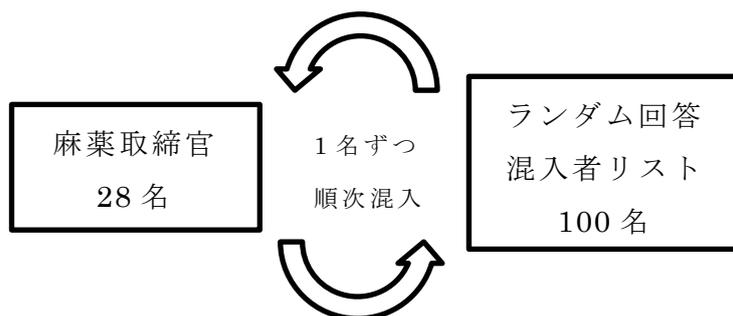
<sup>19</sup> なお米国 O\*NET では 1998 年の公開以来職業ごとの推定値のみを公開しており、細分化した属性別データや就業者単位のデータを外部に公開したことはない。その理由について米国労働省は、「回答者の暴露を防ぐためであり、また信頼できる推定を行うにはサンプルサイズが小さくなりすぎるためでもある。」としている（U.S. Department of Labor, 2018, p.27）。JILPT も同様の方針である。

<sup>20</sup> 仕事の性質は「スケジュールの規則性」のみ比率系のため除外している。各領域の項目の詳細は労働政策研究・研修機構（2020）、および職業情報提供サイト（日本版 O-NET）を参照されたい。

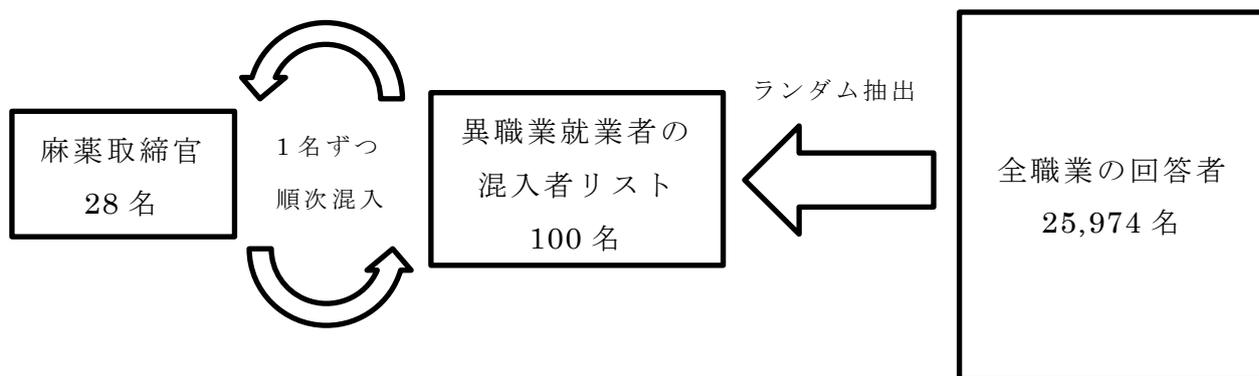
<sup>21</sup> 範囲内ランダム回答の作成には Microsoft Excel の RANDBETWEEN 関数を用いた。

中からランダムに 100 名を抽出した「混入者リスト」を作成し順次混入させた（図表 3）<sup>22</sup>。異職業就業者の混入者リストに含まれた 100 名の職業は付録 2 を参照されたい。混入者リストは 5 職業、すべての条件間で共通である。シミュレーション実施前に混入者リストを作成・固定した理由は、各条件を相対評価するにあたって混入者リストに含まれるケースの違いによる影響を排除したかったためである。

図表 2. 無気力ランダム回答者の疑似混入シミュレーションの枠組み(麻薬取締官の場合)



図表 3. 異職業の就業者の疑似混入シミュレーションの枠組み(麻薬取締官の場合)



### 3.5. 比較対象とする 3 種の除外対象判別手法

今回比較対象とする偽の回答者（＝除外対象）の判別手法は下記 3 種である<sup>23</sup>。

<sup>22</sup> 全候補ケースに通し番号を与え、Microsoft Excel の RANDBETWEEN 関数により抽選を行った。なお 1 回目の抽選時に混入先 5 職業の者が含まれていた場合は再抽選することとしたが実際には再抽選が必要なケースは発生しなかった。

<sup>23</sup> 当初、二項ロジスティック回帰分析による傾向スコアを用いた判別手法のシミュレーションも実施を検討した。しかし、(1)本稿で使用する真の回答者群には欠損値がありケースごと除外するか代替の値を与える等の処置が必要となる、(2)傾向スコア算出にあたっては真の回答者群と偽の回答者群を同時に分析に使用するため、他の手法とシミュレーションの枠組みが異なり比較が難しい、(3)通常は分析者が行う多重共線性や完全分離の評価等の投入変数の検討をすべてプログラミング上で自動化することは難しい、(4)仮に人力でパラメータを調整して良い結果が得られたとしても約 500 の職業に同手法で毎年スクリーニングを行うことはコストの観点から困難である、といった理由から見送ることとした。なお、多変量データに基づく群の判別は各種 AI の分類器も得意としており、実際に尾崎・鈴木（2019）では機械学習によって Web 調査における不適切な回答者の検出を試みている。しかし本稿の場合教師データにおける「該当」ケースが 14～28 名と少ないため十分な機械学

以下それぞれ手法 A、B、C と呼ぶ。

#### 判別手法 A：緩やかな「外れ値」判定に基づく、「外れ値」出現回数による判定

手法 A は、各設問（変数）の得点について個別に「緩やかな『外れ値』」を判定し、全項目を通しての出現回数（0～110 の範囲）について指定の有意水準（後述）の範囲を超えるものを偽の回答者の可能性が高いものとして除外対象とする手法である。ここで手法 A における「緩やかな『外れ値』」とは、具体的には  $M \pm 1SD$  の範囲を超えるものを指す。正規分布を仮定した場合この範囲には分布の 68.2% が含まれる。したがって仮にある人物がサンプルの想定する母集団に属する真の回答者で、かつ他のサンプルもすべて真の回答者であったとしても、各設問で 31.8% 程度の確率で「外れ値」と判定されることになる。このような緩やかな判定を採用した理由は、仮に統計学的な慣例に従い分布の 95% が含まれる範囲（ $M \pm 1.96SD$ ）等を採用した場合には同じ母集団の構成員とは考えづらい程度に大きく平均値から外れている回数を数える、言い換えれば 1 変数ごとに決定的なズレを判定してカウントすることになるが、今回のように 110 の変数で総合的に外れ値の出現回数を用いて判定しようとする場合には「少し外れている」ものも含めてカウントした方が得られる出現回数の分布の情報量が豊富となり、結果的により精度の高い判定ができる可能性があったためである。

上述の方法でカウントされた「外れ値」出現回数について、本稿ではさらに下記 3 つの有意水準により除外判定を行い後述の評価指標にて比較することとした。

##### 条件 A1：「緩やかな外れ値判定における厳しい合格判定」条件

$M \pm 1SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照らして片側 5% 水準（ $M + 1.64SD$ ）以下のときに「合格」、上回る場合に「不合格」（除外対象）とする。

##### 条件 A2：「緩やかな外れ値判定における中程度の合格判定」条件

$M \pm 1SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照らして片側 2.5% 水準（ $M + 1.96SD$ ）以下のときに「合格」、上回る場合に「不合格」（除外対象）とする。

##### 条件 A3：「緩やかな外れ値判定における甘い合格判定」条件

$M \pm 1SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照

---

習は困難と考えられた。

らして片側 1%水準 ( $M + 2.33SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象) とする。

除外判定の基準を正方向の片側のみとした理由は、基本的に真の回答者よりも偽の回答者のほうが外れ値出現回数は多いと予測され、外れ値が極端に多いケースのみを除外対象と想定すべきと考えたためである。仮に群の構成員が全員真の回答者であった場合、条件 A1 では 5%、条件 A2 では 2.5%、条件 A3 では 1%が理論上誤って除外されることになる。真の回答者の情報量の損失を避けるという意味では A3 条件が最も望ましいが、A1、A2 において偽の回答者の判定精度(特異度; 後述)が大きく向上する可能性も検証するためそれぞれシミュレーションを行うこととした。

なお、手法 A では欠損値は自動的に「外れ値」判定とカウントした。これは欠損値は分析者視点で望ましくない値であり、欠損値を生じさせる回答者はそうでない回答者よりもどちらかといえば厳しく判定した方が自然と考えたためである<sup>24</sup>。

#### 判別手法 B: 厳しい「外れ値」判定に基づく、「外れ値」出現回数による判定

手法 B は手法 A と構造的には同じ枠組みを採用するが、各設問における「外れ値」の判定基準を「 $M \pm 2SD$ 」(分布の 95.4%を含む範囲)の超過、とより厳しく設定する手法である。手法 A の説明における「緩やかな判定基準のほうが出現回数の分布の情報量が豊富となり、結果的により精度の高い判定ができる可能性」についての記述は分析者の推測に過ぎない。実際には項目ごとに決定的なズレのみカウントした方が判定精度が高い可能性もある。そこで手法 B についても、下記 3 条件にてそれぞれシミュレーションを実施することとした。

##### 条件 B1: 「厳しい外れ値判定における厳しい合格判定」条件

$M \pm 2SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照らして片側 5%水準 ( $M + 1.64SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象) とする。

##### 条件 B2: 「厳しい外れ値判定における中程度の合格判定」条件

$M \pm 2SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照らして片側 2.5%水準 ( $M + 1.96SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象) とする。

---

<sup>24</sup> ただしこの点については回答者の落ち度ではない無回答もあり得るため議論の余地がある。

### 条件 B3: 「厳しい外れ値判定における甘い合格判定」条件

$M \pm 2SD$  の条件下での「外れ値」出現回数が、群全体の出現回数の分布に照らして片側 1%水準 ( $M+2.33SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象) とする。

欠損値の取り扱いは手法 A と同様である。

### 判別手法 C: 標準化ユークリッド距離に基づく判定

手法 A、B では項目ごとに 0「外れ値ではない」か、1「外れ値である」かを結論付けた上でその回数を算出した。一方手法 C では、「距離」と呼ばれる幾何学的な指標に基づき群全体の平均値からの乖離の程度を総合的に評価し指標化する。項目単位では特に判定は行わず、総合的な「距離」を用いて群全体から著しく外れている回答を判別しようとする手法である。

群全体の回答傾向からの乖離の度合いを指標化する上で第 1 に検討されるべきはマハラノビス距離 (Mahalanobis, 1936) や、そのロバスト推定 (後述) としての拡張となる MSD 法 (e.g. 和田, 2010) を活用した多変量外れ値という考え方である。マハラノビス距離では多変量に対して分散・共分散行列を加味して群全体からの乖離の度合いを数値化することができる。変数間の相関がすべて 0 のときマハラノビス距離は標準化ユークリッド距離と一致することから (救仁郷, 2001)、ユークリッド距離のより一般化された指標と捉えることもできる。しかし、救仁郷 (2001) によればマハラノビス距離には数学的な制約があり、中でも「基準データ群のデータ数が変量数以下の場合には逆行列が存在しないため計算不能」という点で今回のシミュレーションの枠組みには適用できない<sup>25</sup>。

そこで次善の策として手法 C では標準化ユークリッド距離を活用することとした。標準化ユークリッド距離は変数間の相関 (共分散) は加味しないがデータ数の制約なく算出することが可能である。したがって数値項目をそれぞれ標準化してからユークリッド距離を回答者ごとに算出し (=標準化ユークリッド距離)、他の回答者と比べて極端に距離が離れている者を除外判定することとした。

なお手法 C においても、最終的に算出された標準化ユークリッド距離についてどの程度の有意水準で除外判定を行うかは手法 A、B と共通する下記 3 条件を置いた。

### 条件 C1: 「標準化ユークリッド距離に基づく厳しい合格判定」条件

標準化ユークリッド距離が、群全体の分布に照らして片側 5%水準

---

<sup>25</sup> 本稿の文脈では「基準データ群」は 15~29 名しかなく、「変量数」110 に満たない。

( $M+1.64SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象)とする。

**条件 C2: 「標準化ユークリッド距離に基づく中程度の合格判定」条件**

標準化ユークリッド距離が、群全体の分布に照らして片側 2.5%水準 ( $M+1.96SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象)とする。

**条件 C3: 「標準化ユークリッド距離に基づく甘い合格判定」条件**

標準化ユークリッド距離が、群全体の分布に照らして片側 1%水準 ( $M+2.33SD$ ) 以下のときに「合格」、上回る場合に「不合格」(除外対象)とする。

欠損値については、欠損値を除く他のケースについて当該項目の偏差の 2 乗値<sup>26</sup>を算出した後、その最大値と同じ値を挿入した。これは欠損値を含むケースには「どちらかといえば厳しい判定をすべき」との手法 A、B と同じ処置方針に基づき、少なくとも他のケースの中で最も平均値から外れているものと同等程度外れていると仮定した処置である。

### 3.6. 各条件に対するロバスト推定の適用

さて、実際の職業調査では職業の真偽が不明なまま各手法を適用することになるため、シミュレーションにあたっては偽の回答者を混入させる度に平均値と標準偏差を再計算し、これを用いて除外判定を行う必要がある。しかし、そもそも外れ値の除外を試みる理由は平均値が外れ値により「引っ張られる」ことを懸念しているためで、「引っ張られた」可能性があるサンプルの平均値と標準偏差を用いて外れ値を判定することには構造的な問題があると言える。具体的には、第 1 に外れ値に引っ張られた平均値と標準偏差は、外れ値を外れ値と見なさない方向に働く懸念がある。第 2 に外れ値に引っ張られた平均値と標準偏差は、平均値を挟んで反対側の外れ値でない真の回答を外れ値と見なす方向に働く懸念がある。

こうした事態において適用が検討されるのがロバスト推定である。ロバスト推定とは「外れ値が混入していたとしても、メインボディの中心やばらつきなどを妥当に推定する方法」を指す(藤澤, 2017, p.3)。そこで本稿の A1 から C3 までの 9 条件に

---

<sup>26</sup> 事前に変数はすべて標準化されているため、項目ごとの偏差の 2 乗値をすべて加算した値の平方根が標準化ユークリッド距離となる。

についても、ロバスト推定を適用しない非ロバスト条件に加えてロバスト推定を適用した条件 (A1'~C3') のシミュレーションを行った。

ロバスト推定条件における平均値については、10%刈り込み平均値 (トリム平均値) を適用した。平均値のロバスト推定の手法は様々あるが、藤澤 (2017) によれば刈り込み平均値は「外れ値の割合を事前に想定する必要がある」ことが唯一の問題点であるものの、「少々適当に想定しても、刈り込み平均は妥当な推定になりやすく、思いのほか推定量が小さくなることもあり、使いやすい推定法である」(p.8)。そこで、本稿では各設問において上位 10%と下位 10%の回答を除外した上で平均値を計算することで外れ値による影響を軽減することとした。

一方、標準偏差については四分位範囲 (IQR: interquartile range) の正規化によってロバスト推定を行った。四分位範囲とはある変数を得点順に並べ、下側から 25%点 (第 1 四分位数) と 75%点 (第 3 四分位数) を算出し、75%点から 25%点を引いた範囲を指す。藤澤 (2017) によれば「中央値が外れ値に強いと予想されることと同様に、25%点や 75%点も、外れ値に強いと考えられ」(p.12)、データが正規分布に従うと仮定した場合四分位範囲を 1.349 で割った値が標準偏差の不偏推定量となる。

なお、手法 C においては IQR を 1.349 で割った標準偏差のロバスト推定値によって各変数を標準化する必要があるが、IQR ではサンプルサイズにもよるが概ねデータの上位下位 25%のデータを使用しないため、本稿のシミュレーションにおいて範囲がゼロ (i.e. 25%点=75%点) となるケースがあり、標準化に使用できなかった。このため C1'、C2'、C3'に限り IQR がゼロとなる変数はそもそも距離の計算に使用しないこととした<sup>27</sup>。

### 3.7. シミュレーション結果の評価指標

以上の 3 手法、18 条件 (ロバスト推定条件含む) の評価指標として、本稿では再現性と特異度を用いた。これらは主に医学の臨床検査や機械学習の精度の評価などで用いられている指標である。まず再現性 (Recall) とは医学では感度 (Sensitivity) と呼ばれ (阿部, 2004)、「該当」と判定すべきケースを正しく「該当」と判定する確率を表す。本稿の文脈では真の回答者を正しく合格とした比率である<sup>28</sup>。一方、特異度 (Specificity) とは「非該当」と判定すべきケースを正しく「非該当」と判定する確率を表す。本稿の文脈で言えば偽の回答者を正しく不合格 (i.e. 除外対象) とし

---

<sup>27</sup> 厳密にはスクリプト上の処理としては当該の項目を全ケースで偏差の 2 乗値「0」と置くことで距離の算出への影響を取り除いた。

<sup>28</sup> なお臨床検査の場合「該当」すなわち「陽性」とは「問題があるケース」を指すため、再現性とは「望ましくないケースを検出する精度」である。しかし本稿では真の回答者であることを「該当 (合格)」としているため、再現性は「望ましいケースを検出する精度」である点に留意されたい。

た比率である。両指標はどちらも 100%であることが理想だが、原理的にトレードオフの関係にある。極端に緩い合格基準を置けば真の回答者は全員合格（再現性 100%）となるが偽の回答者も合格となり特異度は落ちる。逆に極端に厳しい合格基準を置けば特異度は高くなるが、真の回答者も切り捨ててしまうリスクが高まる。

本稿では両指標を総合的に評価するが、再現性を特に重視することとした。もとより限られたサイズのサンプルにおいて、真の回答者を誤って除外してしまう「副作用」は可能な限り低く抑えたかったためである。そこで、真の回答者をほぼ除外しないことを最低条件とした上で、どの程度特異度を高められるか（＝偽の回答者を除外できるか）、という観点で各条件の比較を行うこととした。

なお、特異度については 1 回の混入シミュレーションにおいて混入者の除外判定結果が 0（合格、除外しない）、1（不合格、除外する）で得られ、その 100 回平均となるためシンプルだが、再現性については同様の 0,1 データで真の回答者ごとに 100 回平均を求め、その後真の回答者群の中で「100 回平均の平均」を計算し、この値を 1 から引く（不合格率ではなく合格率を見る）という計算方法を採用した。

## 4. 結果

### 4.1. 各条件における「外れ値」出現回数、および標準化ユークリッド距離

手法 A ( $M \pm 1SD$ )、手法 B ( $M \pm 2SD$ ) の「外れ値」出現回数、および手法 C の標準化ユークリッド距離について真の回答者群と偽の回答者群の平均値と標準偏差、および群間の対応の無い  $t$  検定結果を図表 4、5 に示す<sup>29</sup>。偽の回答者群については自身が混入したときの出現回数/距離について 100 ケースの中で平均値と標準偏差を求めた。真の回答者群についてはまず 1 名ごとに 100 回の混入時の出現回数/距離の平均値を求めた上で、この平均値群に対して群内で平均値と標準偏差を求めた。計算方法のイメージを図表 6 に示す。

緩やかな判定基準による手法 A の出現回数は、厳しい判定基準による手法 B と比べて出現回数が多くなっている様子が窺える。ロバスト推定の適用については、手法 A では非ロバスト条件とさほど大きな違いは見られないものの、手法 B と手法 C では偽の回答者群だけでなく真の回答者群についても出現回数/距離が増加している。

こうした手法間の違いは見られるものの、無気力ランダム回答混入時（図表 4）、異職業の就業者混入時（図表 5）のすべての条件で偽の回答者群は真の回答者群よりも有意に出現回数が多い、または距離が大きかった。

---

<sup>29</sup> なお同一手法内では「外れ値」出現回数、ならびに標準化ユークリッド距離は共通である。これを活用して有意水準別 3 条件（e.g. A1、A2、A3）の条件分けが行われる。

**図表 4. 各手法における「外れ値」出現回数または標準化ユークリッド距離の平均値と標準偏差、および群間の対応の無い t 検定結果（無気ランダム回答混入の場合）**

	<i>n</i> (真群)		手法A 外れ値出現回数				手法B 外れ値出現回数				手法C 標準化ユークリッド距離			
			非ロバスト		ロバスト		非ロバスト		ロバスト		非ロバスト		ロバスト	
			真群	偽群	真群	偽群	真群	偽群	真群	偽群	真群	偽群	真群	偽群
麻薬取締官	28	<i>M</i>	31.0	66.1 ***	31.8	64.8 ***	4.3	29.3 ***	15.4	37.4 ***	9.9	18.2 ***	13.9	30.4 ***
		<i>SD</i>	(10.1)	(5.2)	(8.7)	(5.1)	(4.6)	(4.6)	(4.8)	(4.2)	(1.8)	(1.0)	(6.4)	(1.6)
入国警備官	25	<i>M</i>	27.1	68.4 ***	28.1	66.6 ***	3.6	37.3 ***	12.1	42.3 ***	9.4	21.8 ***	13.5	30.9 ***
		<i>SD</i>	(7.9)	(5.7)	(7.8)	(5.3)	(3.3)	(4.7)	(4.3)	(5.3)	(1.4)	(1.1)	(6.0)	(1.7)
検察官	20	<i>M</i>	27.7	72.5 ***	31.8	69.6 ***	4.3	41.4 ***	17.0	48.6 ***	9.4	21.8 ***	15.5	31.5 ***
		<i>SD</i>	(9.9)	(5.0)	(8.7)	(4.8)	(6.3)	(4.4)	(6.0)	(4.3)	(1.9)	(0.9)	(10.4)	(2.2)
家庭裁判所 調査官	20	<i>M</i>	25.7	74.9 ***	37.4	74.6 ***	3.0	46.7 ***	24.8	54.5 ***	9.1	22.9 ***	14.4	32.0 ***
		<i>SD</i>	(8.3)	(4.6)	(6.2)	(4.0)	(3.5)	(4.4)	(3.6)	(4.2)	(1.4)	(0.9)	(11.4)	(2.3)
外交官	21	<i>M</i>	31.4	67.4 ***	32.9	67.2 ***	3.6	30.7 ***	17.9	42.0 ***	9.7	18.1 ***	14.3	31.4 ***
		<i>SD</i>	(10.9)	(5.1)	(8.8)	(4.4)	(3.2)	(4.6)	(4.3)	(4.2)	(1.6)	(1.0)	(7.8)	(1.9)

真群は「真の回答者」群を、偽群は「偽の回答者」群を指す。図表5も同じ。

対応の無い t 検定で5%水準で有意のとき"\*"、1%水準で有意のとき "\*\*\*"、0.1%水準で有意のとき"\*\*\*\*"を付した。図表5も同じ。

**図表 5. 各手法における「外れ値」出現回数または標準化ユークリッド距離の平均値と標準偏差、および群間の対応の無い t 検定結果（異職業の就業者混入の場合）**

	<i>n</i> (真群)		手法A 外れ値出現回数				手法B 外れ値出現回数				手法C 標準化ユークリッド距離			
			非ロバスト		ロバスト		非ロバスト		ロバスト		非ロバスト		ロバスト	
			真群	偽群	真群	偽群	真群	偽群	真群	偽群	真群	偽群	真群	偽群
麻薬取締官	28	<i>M</i>	32.2	50.3 ***	32.7	49.9 ***	4.8	15.1 ***	16.1	27.0 ***	10.1	14.2 ***	13.8	24.1 ***
		<i>SD</i>	(10.2)	(11.1)	(8.6)	(8.5)	(4.8)	(7.7)	(4.7)	(6.7)	(1.9)	(2.2)	(6.4)	(5.8)
入国警備官	25	<i>M</i>	28.3	49.3 ***	27.7	47.5 ***	4.1	20.3 ***	11.7	26.6 ***	9.7	16.4 ***	13.6	24.7 ***
		<i>SD</i>	(8.2)	(9.8)	(7.8)	(9.9)	(3.4)	(8.8)	(4.4)	(8.1)	(1.4)	(3.3)	(5.9)	(6.0)
検察官	20	<i>M</i>	28.7	66.5 ***	30.7	60.1 ***	4.6	30.5 ***	16.6	38.5 ***	9.6	19.2 ***	15.1	23.8 **
		<i>SD</i>	(10.4)	(9.6)	(8.9)	(8.7)	(6.4)	(8.3)	(6.2)	(8.0)	(1.9)	(2.5)	(9.9)	(5.9)
家庭裁判所 調査官	20	<i>M</i>	27.4	62.6 ***	36.2	64.1 ***	3.7	33.8 ***	23.7	46.3 ***	9.4	19.6 ***	14.2	25.1 ***
		<i>SD</i>	(8.8)	(10.9)	(6.5)	(8.8)	(4.6)	(10.2)	(3.7)	(7.9)	(1.6)	(2.7)	(11.0)	(6.4)
外交官	21	<i>M</i>	32.0	52.2 ***	34.6	55.4 ***	4.2	20.1 ***	19.8	35.3 ***	9.9	15.1 ***	14.2	25.0 ***
		<i>SD</i>	(10.9)	(9.4)	(9.0)	(7.1)	(3.6)	(7.5)	(4.2)	(6.9)	(1.7)	(2.1)	(7.6)	(5.8)

図表 6. 真の回答者群、偽の回答者群ごとの外れ値発生回数と標準化ユークリッド距離の平均値・標準偏差の計算方法のイメージ

素点データのイメージ

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	...	v109	v110
case1	5	5	1	5	5	4	5	4	1	5	1		2	3
case2	5	5	3	4	4	2	4	4	4	3	1		2	3
case3	5	4	2	5	5	5	5	4	3	4	2	...	2	2
case4	5	5	5	5	5	5	4	3	2	2	2		0	0
case5	3	4	2	4	3	3	4	3	3	2	2		na	na
...														
case28	3	1	1	2	4	5	2	2	1	1	2	...	1	0
case29	1	5	5	1	2	1	4	2	4	2	5	...	0	4
M	3.8	3.9	2.3	4.4	3.7	3.5	4.3	3.9	3.2	3.2	2.3	...	1.1	1.6
SD	1.2	1.1	1.2	1.0	1.0	1.2	0.8	1.1	1.0	1.1	1.1	...	1.3	1.5

※欠損のあるケース  
※偽の回答者ケース  
※ロバスト条件では10%刈り込み平均、IQRによる推定標準偏差を用いる。



素点データに基づき、手法に応じて外れ値回数/距離を計算

【手法A、B】各素点の外れ値発生回数の計算方法イメージ  
(1:外れ値、0:非外れ値; 表は手法A適用時のイメージ)

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	...	v109	v110	各ケース 外れ値 発生回数
case1	1	0	1	0	1	0	0	0	1	1	1		0	0	31
case2	1	0	0	0	0	1	0	0	0	0	1		0	0	26
case3	1	0	0	0	1	1	0	0	0	0	0	...	0	0	32
case4	1	0	1	0	1	1	0	0	1	1	0		0	1	43
case5	0	0	0	0	0	0	0	0	0	1	0		1	1	17
...															
case28	0	1	1	1	0	1	1	1	1	1	0	...	0	1	30
case29	1	0	1	1	1	1	0	1	0	1	1	...	0	1	64

欠損値は自動的に「1」と判定

真の回答者群は同様の列が100回のシミュレーションで100列発生。図表4、5の「平均値」「標準偏差」とは、真の回答者1名ごとに100列の平均値を取って「個人単位の平均外れ値回数」を計算し、その後真の回答者群内（この表では28名）での平均値と標準偏差を計算したもの。

偽の回答者は100回のシミュレーションで数値が100個発生。図表4、5の「平均値」「標準偏差」はその100個内で計算したもの。

【手法C】標準化後の平均からの偏差2乗値とその合計値、標準化ユークリッド距離の計算方法イメージ

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	...	v109	v110	合計	標準化 ユークリッド 距離
case1	1.1	1.0	1.1	0.4	1.6	0.2	0.7	0.0	5.1	2.5	1.5		0.5	0.9	84.9	9.2
case2	1.1	1.0	0.4	0.2	0.1	1.6	0.1	0.0	0.7	0.0	1.5		0.5	0.9	80.6	9.0
case3	1.1	0.0	0.1	0.4	1.6	1.6	0.7	0.0	0.0	0.5	0.1	...	0.5	0.1	76.2	8.7
case4	1.1	1.0	5.2	0.4	1.6	1.6	0.1	0.6	1.5	1.2	0.1		0.7	1.2	129.9	11.4
case5	0.5	0.0	0.1	0.2	0.5	0.2	0.1	0.6	0.0	1.2	0.1		5.1	2.6	63.9	8.0
...																
case28	0.5	6.8	1.1	6.0	0.1	1.6	7.3	2.7	5.1	4.0	0.1	...	0.0	1.2	98.1	9.9
case29	5.6	1.0	5.2	12.1	3.0	4.5	0.1	2.7	0.7	1.2	5.7	...	0.7	2.6	337.4	18.4

欠損値は同列内の最大値を割り当て

## 4.2. 無気ランダム回答混入時の各条件における再現性と特異度

無気ランダム回答混入時の各条件下における再現性と特異度を図表 7 に示す。事前の評価方針に則り、再現性が 98.0%以上と特に高いセルを黄色くカラーリングした<sup>30</sup>。方法の節で述べた通り特異度は 100 回混入の単純平均なので小数点第 1 位は必ず 0 となる。また同一手法内の有意水準 3 条件の再現性は定義上必ず  $3 \geq 2 \geq 1$  となる (e.g.  $A3 \geq A2 \geq A1$ )。

まず全体を概観すると手法に関わらず高い再現性を保持したまま高い特異度を達成できている条件が多く見られた。中でも A2、A3、A3'、B3、C2、C3 の 6 条件は全職業で再現性が 98.0%以上であり、真の回答者を誤って除外してしまうリスクは非常に小さいことが示唆されている。そこで特異度の観点でこの 6 条件を比較すると B3、C2、C3 の 3 条件はすべての職業で偽の回答者を 100%除外できていた。

この他全体傾向としてロバスト条件は非ロバスト条件とほとんど判定精度に違いは無いこと、ただし手法 C のロバスト推定 3 条件 (C1'、C2'、C3') については再現

図表 7. 無気ランダム回答混入時の各条件の再現性と特異度

麻薬取締官				入国警備官				検察官									
非ロバスト		ロバスト		非ロバスト		ロバスト		非ロバスト		ロバスト							
再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度						
A1	95.9%	100.0%	A1'	98.0%	99.0%	A1	100.0%	100.0%	A1'	100.0%	100.0%	A1	100.0%	100.0%	A1'	95.1%	100.0%
A2	99.2%	99.0%	A2'	99.6%	97.0%	A2	100.0%	100.0%	A2'	100.0%	100.0%	A2	100.0%	100.0%	A2'	97.1%	100.0%
A3	100.0%	93.0%	A3'	100.0%	90.0%	A3	100.0%	100.0%	A3'	100.0%	97.0%	A3	100.0%	100.0%	A3'	99.9%	100.0%
B1	96.1%	100.0%	B1'	88.6%	100.0%	B1	98.9%	100.0%	B1'	93.3%	100.0%	B1	95.0%	100.0%	B1'	85.9%	100.0%
B2	97.3%	100.0%	B2'	90.8%	100.0%	B2	99.7%	100.0%	B2'	95.2%	100.0%	B2	95.4%	100.0%	B2'	88.4%	100.0%
B3	99.4%	100.0%	B3'	93.1%	100.0%	B3	100.0%	100.0%	B3'	96.0%	100.0%	B3	98.0%	100.0%	B3'	90.5%	100.0%
C1	95.3%	100.0%	C1'	89.4%	100.0%	C1	100.0%	100.0%	C1'	84.0%	100.0%	C1	97.5%	100.0%	C1'	97.5%	0.0%
C2	98.5%	100.0%	C2'	95.4%	100.0%	C2	100.0%	100.0%	C2'	86.5%	100.0%	C2	100.0%	100.0%	C2'	100.0%	0.0%
C3	99.9%	100.0%	C3'	98.9%	76.0%	C3	100.0%	100.0%	C3'	91.6%	100.0%	C3	100.0%	100.0%	C3'	100.0%	0.0%

家庭裁判所調査官				外交官							
非ロバスト		ロバスト		非ロバスト		ロバスト					
再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度				
A1	100.0%	100.0%	A1'	91.2%	100.0%	A1	98.7%	100.0%	A1'	96.7%	100.0%
A2	100.0%	100.0%	A2'	95.8%	100.0%	A2	100.0%	98.0%	A2'	99.6%	100.0%
A3	100.0%	100.0%	A3'	98.5%	100.0%	A3	100.0%	84.0%	A3'	100.0%	100.0%
B1	99.9%	100.0%	B1'	93.0%	100.0%	B1	99.8%	100.0%	B1'	95.1%	100.0%
B2	100.0%	100.0%	B2'	95.3%	100.0%	B2	100.0%	100.0%	B2'	97.0%	100.0%
B3	100.0%	100.0%	B3'	97.4%	100.0%	B3	100.0%	100.0%	B3'	98.8%	100.0%
C1	100.0%	100.0%	C1'	93.0%	13.0%	C1	99.9%	100.0%	C1'	90.8%	96.0%
C2	100.0%	100.0%	C2'	93.2%	11.0%	C2	100.0%	100.0%	C2'	96.2%	69.0%
C3	100.0%	100.0%	C3'	93.5%	10.0%	C3	100.0%	100.0%	C3'	100.0%	26.0%

再現性が98%以上のセルをカラーリングしている。図表7、8、9も同じ。

<sup>30</sup> 98.0%を目安とした理由は、各手法の第 3 条件 (A3, A3', B3, B3', C3, C3') において真の回答者のみだった場合に理論上期待される再現性が 99.0%であることを踏まえ、混入により 1.0%程度の低下は許容し得ると考えたためである。なお再現性 98.0%の場合、20 名の真の回答者のみサンプルの場合で誤除外人数の期待値は 0.4 名、40 名の場合で 0.8 名、60 名の場合で 1.2 名となり、JILPT の調査では各職業で概ね 0~1 名に抑えられるものと期待できる。

性は高い水準を維持するものの特異度が検察官ですべて 0.0%、家庭裁判所調査官で 10.0%～13.0%と際立って低かった。

#### 4.3. 異職業の就業者混入時の各条件における再現性と特異度

異職業の就業者混入時の各条件下における再現性と特異度を図表 8 に示す。図表 7 と同様、再現性が 98.0%以上と特に高いセルを黄色くカラーリングしている。全体として無回答ランダム回答混入時と比べて判定精度は低下しており、全職業で再現性 98.0%以上を達成していたのは A3 条件のみだった。ただし同条件における特異度は検察官や家庭裁判所調査官では 81.0%、77.0%と高いものの、その他の 3 職業では 50%を切っていた。

手法間で比較すると手法 B は手法 A と比べて総じて特異度は高いものの再現性は低かった。手法 C は麻薬取締官において再現性 98.0%以上を達成できていないものの、その他の職業では手法 A とほぼ同等の再現性、かつ手法 B 以上の特異度が見られた。

ロバスト推定については麻薬取締官の手法 B'、検察官の手法 B'、家庭裁判所調査官の手法 A'、外交官の手法 A'で非ロバスト条件よりも特異度が向上していた。特に麻薬取締官の B3→B3'では 30.0%から 56.0%に、外交官の A3→A3'では 16.0%から 47.0%に大きく精度が向上している。ただしこれらの条件では若干再現性は低下して

図表 8. 異職業の就業者混入時の各条件の再現性と特異度

麻薬取締官				入国警備官				検察官									
非ロバスト		ロバスト		非ロバスト		ロバスト		非ロバスト		ロバスト							
再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度						
A1	94.3%	52.0%	A1'	97.3%	44.0%	A1	97.8%	70.0%	A1'	99.9%	59.0%	A1	99.3%	96.0%	A1'	95.0%	91.0%
A2	97.0%	32.0%	A2'	99.4%	30.0%	A2	99.8%	63.0%	A2'	100.0%	48.0%	A2	100.0%	90.0%	A2'	97.2%	86.0%
A3	99.4%	21.0%	A3'	99.8%	16.0%	A3	100.0%	49.0%	A3'	100.0%	35.0%	A3	100.0%	81.0%	A3'	99.7%	74.0%
B1	93.5%	55.0%	B1'	88.9%	72.0%	B1	96.5%	86.0%	B1'	92.9%	83.0%	B1	95.0%	93.0%	B1'	87.8%	100.0%
B2	94.7%	43.0%	B2'	91.6%	64.0%	B2	97.0%	80.0%	B2'	95.0%	79.0%	B2	95.1%	89.0%	B2'	88.7%	98.0%
B3	96.9%	30.0%	B3'	93.4%	56.0%	B3	97.6%	76.0%	B3'	95.9%	72.0%	B3	95.8%	81.0%	B3'	90.4%	98.0%
C1	93.4%	61.0%	C1'	89.3%	45.0%	C1	98.7%	85.0%	C1'	83.8%	75.0%	C1	96.1%	99.0%	C1'	87.5%	3.0%
C2	95.0%	49.0%	C2'	94.4%	34.0%	C2	99.8%	78.0%	C2'	84.9%	64.0%	C2	98.0%	96.0%	C2'	94.9%	0.0%
C3	97.3%	35.0%	C3'	98.3%	22.0%	C3	100.0%	71.0%	C3'	90.0%	52.0%	C3	99.6%	92.0%	C3'	98.2%	0.0%

家庭裁判所調査官				外交官							
非ロバスト		ロバスト		非ロバスト		ロバスト					
再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度				
A1	97.8%	94.0%	A1'	89.1%	97.0%	A1	96.4%	50.0%	A1'	94.1%	75.0%
A2	99.3%	90.0%	A2'	93.3%	95.0%	A2	99.4%	38.0%	A2'	97.9%	65.0%
A3	100.0%	77.0%	A3'	97.1%	93.0%	A3	100.0%	16.0%	A3'	99.9%	47.0%
B1	96.8%	97.0%	B1'	91.3%	97.0%	B1	97.2%	85.0%	B1'	94.8%	85.0%
B2	97.2%	96.0%	B2'	93.5%	95.0%	B2	98.1%	81.0%	B2'	96.9%	80.0%
B3	97.9%	91.0%	B3'	96.7%	95.0%	B3	99.1%	75.0%	B3'	98.8%	74.0%
C1	97.4%	97.0%	C1'	79.9%	61.0%	C1	97.0%	82.0%	C1'	88.7%	40.0%
C2	98.2%	97.0%	C2'	80.4%	51.0%	C2	99.0%	72.0%	C2'	93.1%	30.0%
C3	99.4%	96.0%	C3'	81.4%	43.0%	C3	100.0%	56.0%	C3'	98.7%	20.0%

いることも読み取れる。

#### 4.4. 手法 B 非ロバスト条件へのポアソン分布有意水準適用時の再現性と特異度

さて、図表 7、8 からは手法 B の再現性が相対的に低いことが示唆されたが、図表 4、5 の「外れ値」出現回数を見るとそもそも正規分布を仮定した有意水準の設定が不適切だった可能性が考えられた。もともと「外れ値」出現回数は離散変数であるが、ラプラスの定理により  $np \geq 5$  かつ  $n(1-p) \geq 5$  のとき正規分布に近似できる。しかし手法 B の非ロバスト条件に限っては真の回答者群での平均「外れ値」出現回数が 3.0～4.8 となっており（図表 4、5）、1 名分の混入値によって期待値が引き上げられても上記の条件を満たしていない状況が発生していたと推測された。そこで該当する 3 条件（B1、B2、B3）について 100 回の混入シミュレーションでその都度ポアソン分布の適用是非を判定し、必要な場合のみポアソン分布を使用する「部分ポアソン条件」と、全ての混入シミュレーションで機械的にポアソン分布によって合否判定を行う「完全ポアソン条件」について再現性と特異度を確認した。

結果を図表 9、10 に示す。比較のため正規分布仮定時の結果も再掲する。検察官は混入値の「外れ値」出現回数が多かったため部分ポアソン条件で実際にポアソン分布が活用されることはなかった。その他の結果を概観すると、再現性は概ね正規条件  $\geq$  部分ポアソン条件  $\geq$  完全ポアソン条件となっていた。一方、特異度は無気力ランダム回答混入ではすべて 100.0%で差がつかなかったが、異職業の就業者混入では完全ポアソン条件  $\geq$  部分ポアソン条件  $\geq$  正規条件となった。したがって前述した手

**図表 9. ポアソン分布基準の有意水準を手法 B の非ロバスト 3 条件に適用した場合の再現性と特異度の比較（無気力ランダム回答混入の場合）**

麻薬取締官	部分ポアソン (19/100)						入国警備官	部分ポアソン (69/100)					
	正規		完全ポアソン		正規			完全ポアソン		正規		完全ポアソン	
	再現性	特異度	再現性	特異度	再現性	特異度		再現性	特異度	再現性	特異度	再現性	特異度
B1	96.1%	100.0%	94.0%	100.0%	84.9%	100.0%	B1	98.9%	100.0%	96.5%	100.0%	95.3%	100.0%
B2	97.3%	100.0%	95.5%	100.0%	88.8%	100.0%	B2	99.7%	100.0%	97.1%	100.0%	95.9%	100.0%
B3	99.4%	100.0%	98.0%	100.0%	92.4%	100.0%	B3	100.0%	100.0%	97.2%	100.0%	96.0%	100.0%
検察官	部分ポアソン (0/100)						家庭裁判所調査官	部分ポアソン (35/100)					
	正規		完全ポアソン		正規			完全ポアソン		正規		完全ポアソン	
	再現性	特異度	再現性	特異度	再現性	特異度		再現性	特異度	再現性	特異度	再現性	特異度
B1	95.0%	100.0%	—	—	94.5%	100.0%	B1	99.9%	100.0%	98.0%	100.0%	94.3%	100.0%
B2	95.4%	100.0%	—	—	94.9%	100.0%	B2	100.0%	100.0%	98.3%	100.0%	94.9%	100.0%
B3	98.0%	100.0%	—	—	95.0%	100.0%	B3	100.0%	100.0%	98.3%	100.0%	95.0%	100.0%
外交官	部分ポアソン (69/100)						部分ポアソン (69/100)						
	正規		完全ポアソン		正規		完全ポアソン		正規		完全ポアソン		
	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	再現性	特異度	
B1	99.8%	100.0%	93.1%	100.0%	89.8%	100.0%	B1	99.8%	100.0%	93.1%	100.0%	89.8%	100.0%
B2	100.0%	100.0%	95.0%	100.0%	92.8%	100.0%	B2	100.0%	100.0%	95.0%	100.0%	92.8%	100.0%
B3	100.0%	100.0%	97.0%	100.0%	95.7%	100.0%	B3	100.0%	100.0%	97.0%	100.0%	95.7%	100.0%

部分ポアソンの (X/100) は、100回の混入シミュレーションでX回ポアソン分布による有意水準が使用されたことを表す。図表9も同じ。

図表 10. ポアソン分布基準の有意水準を手法 B の非ロバスト 3 条件に適用した場合の再現性と特異度の比較 (異職業の就業者混入の場合)

麻薬取締官	部分ポアソン (24/100)						完全ポアソン	再現性	特異度	再現性	特異度	再現性	特異度
	正規		部分ポアソン		完全ポアソン								
	再現性	特異度	再現性	特異度	再現性	特異度							
B1	93.5%	55.0%	90.8%	63.0%	80.9%	78.0%	B1	96.5%	86.0%	92.7%	95.0%	92.4%	95.0%
B2	94.7%	43.0%	92.8%	53.0%	85.6%	77.0%	B2	97.0%	80.0%	94.9%	92.0%	94.7%	92.0%
B3	96.9%	30.0%	95.7%	41.0%	90.9%	72.0%	B3	97.6%	76.0%	96.3%	89.0%	96.0%	89.0%

検察官	部分ポアソン (0/100)						完全ポアソン	再現性	特異度	再現性	特異度	再現性	特異度
	正規		部分ポアソン		完全ポアソン								
	再現性	特異度	再現性	特異度	再現性	特異度							
B1	95.0%	93.0%	—	—	91.9%	100.0%	B1	96.8%	97.0%	95.0%	97.0%	93.1%	97.0%
B2	95.1%	89.0%	—	—	93.3%	100.0%	B2	97.2%	96.0%	95.5%	96.0%	94.3%	97.0%
B3	95.8%	81.0%	—	—	94.8%	99.0%	B3	97.9%	91.0%	96.4%	92.0%	94.8%	97.0%

外交官	部分ポアソン (63/100)						完全ポアソン	再現性	特異度	再現性	特異度	再現性	特異度
	正規		部分ポアソン		完全ポアソン								
	再現性	特異度	再現性	特異度	再現性	特異度							
B1	97.2%	85.0%	89.8%	92.0%	85.1%	93.0%	B1	96.8%	97.0%	95.0%	97.0%	93.1%	97.0%
B2	98.1%	81.0%	92.0%	87.0%	87.7%	89.0%	B2	97.2%	96.0%	95.5%	96.0%	94.3%	97.0%
B3	99.1%	75.0%	95.0%	85.0%	91.9%	88.0%	B3	97.9%	91.0%	96.4%	92.0%	94.8%	97.0%

法 B の特徴である「手法 A よりも再現性は低いの特異度は高い」という傾向が、ポアソン分布を用いることでさらに顕著になる可能性が示されたと言える。

## 5. 考察

本稿の疑似混入シミュレーションから得られた知見は主に 5 点ある。第 1 に、いずれの手法でも真の回答者群は偽の回答者群よりも統計学的に有意に「外れ値」が少なかった、もしくは距離が小さかった。方法の節でも述べた通り、本稿で使用した 110 の変数は正規性が担保されているとは言い難い (付録 1 参照)。こうしたデータであっても一定の正規性を仮定した各手法の指標において真の回答者群と偽の回答者群で一貫して有意差が確認されたことは少なくともこれらの指標を職業真偽の判別に活用しようとする発想自体には一定の妥当性があったことを示している。

第 2 に、無気力ランダム回答については C' の 3 条件を除けば非常に高い判定精度が達成され、各手法の適用の有効性が示された (図表 7)。先行研究ではランダムな回答は追加除外が難しいとされていることがあるが (e.g. 埴淵・村中・安藤, 2015)、本稿の手法を用いれば検出のための追加項目が無い、かつ機械学習等の適用が難しい少サンプルの職業調査において、人間の目で回答内容をチェックしたり特定の調査・職業に特有の除外方法を考案したりする必要なく、多変量データの振る舞いのみを根拠にランダム回答を除外できる可能性が示唆された。本稿では職業調査という文脈に限定しているが、原理的には回答者の大半が共通の母集団から正しく抽出さ

れ、一定の回答傾向を持つと仮定でき、かつそうした変数が一定数存在するのであれば、他の目的の調査において本稿の手法を応用できる可能性がある。

第 3 に、異職業の就業者を混入させた場合には無気力ランダム回答と比べ除外判定の精度は劣ることが分かった（図表 8）。再現性のみを見ると A3 条件が唯一 98.0% 以上を維持しているが、同条件では外交官や麻薬取締官の特異度が低くなっている（各 16.0%、21.0%）。異職業の就業者の検出が無気力ランダム回答よりも難しい理由は 2 点考えられる。第 1 に、現に母集団の回答傾向が類似する職業の存在がある。たとえば同じ事務の仕事に従事する 2 つの職業があったとき、両者の回答傾向は事務以外の仕事と比べて類似していると考えられる。この場合、一方の就業者がもう一方の回答者群に混入しても本稿の手法で除外することは相対的に難しいであろう。第 2 に、全職業である程度共通した回答傾向が示される項目の存在がある。たとえば付録 1 で 104 番目の「knowledge27：芸術」（範囲：0～5）を見ると、今回対象とした真の回答者 5 群はいずれも平均値 1.0 未満であり、異職業混入リスト 100 名の平均値も 0.9 と低くなっている。これは無気力ランダム回答 100 名分の平均値 2.4（範囲のほぼ中央）とは対照的である。つまり、無気力ランダム回答では「設問を読んでもいけば多くの職業で低くなるはずの設問」でランダムに「4」「5」といった回答が出るため外れ値出現回数、もしくは標準化ユークリッド距離が大きくなり、相対的に除外判定が容易であったと考えられる。

ただし、無気力ランダム回答と比べて除外判定が難しいとはいえ、異職業の就業者混入時にも適切な条件を設定すれば真の回答者を除外してしまうリスクを低く抑えられることも示唆された。仮に特異度が 50% に満たないとしても、従来は除外できなかった異職業の混入をほぼ「副作用」無しに一定程度除外できるとすれば、手法適用の有効性はある程度評価し得る。実際、A3 条件の外交官では再現性 100.0%・特異度 81.0%、同条件の家庭裁判所調査官では再現性 100.0%・特異度 77.0% となっており、職業によっては本稿の手法で異職業の就業者の混入を高い精度で除外できる場合もあることが示唆されている。

第 4 に、ロバスト推定に関しては手法 C' で特異度が極めて低いほか、手法 A、B では一貫した傾向は見られず、判定精度の向上に寄与しているとは言い難かった。まず手法 C' で特異度が低かった理由として、四分位範囲がゼロのとき項目得点の標準化が実行できないためその変数を距離計算から除外したことが原因と考えられる。四分位範囲がゼロの項目とは群のメインボディ（概ね、全体の中央付近 50%）が全員同値で回答している項目であり、本来は特に職業真偽の判定に役立つはずだった標準偏差の小さい項目である。これらの項目群が数学上の理由で取り除かれてしまったことで検察官や家庭裁判所調査官では極端に偽の回答者を検出する精度が落ちて

しまったものと考えられる。この意味でロバスト推定自体が判定精度を低くしたというより、本稿で採用したロバスト推定の方法が手法 C ではうまく機能しなかったと考えるべきである。

一方、手法 A、B のロバスト推定については結果が一貫しなかった。この理由として初めに考えられるのは、真の回答者の「刈り込みすぎ」の可能性である。本稿のロバスト推定では 10%刈り込み平均値を用いたが、これは 20 名の場合で各項目の上位 2 名、下位 2 名の得点を無視して平均値を求めることを意味する。今回混入は 1 名と設定しているため、仮に混入値が正しく刈り込まれたとしても同時に 3 名分真の回答者の情報量が減った推定を行ったことになる。このため、真の回答者群の分布状況によってはロバスト推定適用による恩恵よりも真の回答者の情報を刈り込んだ悪影響が大きくなっていった可能性がある。ただし、この可能性を検討するためいくつかの条件について 5%刈り込み平均値<sup>31</sup>で再度シミュレーションを実行したところ、再現性と特異度は 10%刈り込み時とほとんど変わらないか、むしろ悪化するケースも見られた。このことから、おそらく手法 A、B のロバスト推定で判定精度が落ちる場合がある理由は刈り込みすぎが原因ではなく、四分位範囲による標準偏差のロバスト推定にあったと考えられる。四分位範囲を 1.349 で割るという推定方法は、あくまで正規分布を仮定した場合のものである（藤澤，2017）。しかし付録 1 の歪度で確認されるように実際には回答項目は正規分布とは言い難いものが多かった。このため正規性の仮定が直接的に影響する推定方法では歪みが大きくなっていった可能性がある。

第 5 に、今回検証した 18 条件の中では A3 条件が最も再現性を高く保持するが、特異度も考慮すると C3 条件も望ましい精度を持つことが示唆された。まず A3 条件に関しては混入者の特性や職業によらず常に再現性 99.0%以上を維持しており、再現性を特に重視するという当初の評価方針から見て最も優れている。A3 条件を採用する限り真の回答者を誤って除外してしまう「副作用」のリスクは極めて小さいことが示唆されており、この意味で最も安心して適用できる手法と言える。一方、A3 条件は職業によっては特異度が低くなりがちであり、総合的な観点では C3 条件がバランスの良い判定精度を示した。C3 条件の場合麻薬取締官への異職業就業者混入時のみ再現性が 97.3%とやや低いですが、それ以外の職業、手法ではすべて 99.0%以上である。その上で C3 条件では異職業就業者混入時の特異度が入国警備官の 71.0%、外交官の 56.0%など、A3 条件を大きく上回っている（A3 条件ではそれぞれ 49.0%、16.0%）。したがって「副作用」のリスクをある程度許容するのであれば、C3 条件も優れた判別手法であると考えられる<sup>32</sup>。

<sup>31</sup> 5%刈り込み平均値の場合は 20 名の場合で得点上位 1 名、下位 1 名を刈り込むため、本稿のシミュレーション条件下では 10%刈り込み平均値よりも真の回答者の情報の減少量は抑えられる。

<sup>32</sup> なお、本稿のシミュレーションでは手法 B について主に再現性の観点で判定精度が比較的低かつ

以上の 5 つの知見から「職業調査における対象職業の就業者ではない回答者を回答傾向に基づき事後に統計学的に除外する判別手法の妥当性を疑似混入シミュレーションによって評価する」との本稿の目的は、一部ロバスト推定に関して明確な結果が得られなかったものの一定程度は達成されたものと考えられる。ただし本稿の方法と結果にはいくつかの課題と制約がある。この点について最後に 4 点述べる。

第 1 に、本稿で真の回答者群として採用した 5 職業はいずれも専門性を有する官職であり偏りがある。したがって得られた知見が他の多くの職業の回答者群にも一般化できるかは今後さらなる検証が必要である。JILPT の調査では今後民間職種についても、Web 調査でサンプルサイズが確保できないものについて直接依頼による調査を計画している。これらのデータを用いて今後はより多様な職業で同様のシミュレーションを実行し判定精度を評価することが求められる。

第 2 に、上記の第 1 の課題と関連して、本稿で使用した 5 職業のデータは単に職業が客観的に保証されているというだけでなく、紙の郵送調査で実施している、協力元の応答者が配布先や回答者を決定している、といった点で他の Web モニター調査データとはサンプリングの原理が質的に異なる。本来は「Web モニター調査で、かつ、職業真偽が分かっている」サンプルを「真の回答者群」と設定することが理想である。本稿では冒頭で述べた通りあえて偽の回答者混入の課題のみに着目したが、今後は他の交絡要因を統制したシミュレーションの実施も検討したい。

第 3 に、除外判定に必要な項目数についてもさらなる検証が必要である。本稿では 110 の項目を用いて回答者の判別を試みたが、通常、これほど多数の項目で分布のある回答を得られることは少ない。実際、JILPT の調査でも 2020 年度以降は回答者負担の軽減のため 1 ヶ年あたりの設問数は半数程度まで落として実施している。仮に除外判定に活用できる項目数が 50、40、30 と減っていった場合に、また各項目の尋ねている内容も変わったときに、本稿の提案手法が実用に耐える程度の判定精度を維持できるかどうかは検証していく必要がある。

第 4 に、本稿のシミュレーションのパラメータ設定の制約についても留意が必要である。ロバスト推定のさらなる妥当性検証も重要だが、特に「混入者 1 名」という本稿の仮定については議論の余地がある。本稿では「混入が生じるとしても全体の 5%程度(真の回答者 20 名に対して 1 名程度)であろう」という想定で「混入者 1 名」のシミュレーションを行ったが、もちろん実際の Web モニター調査で平均何名程度が混入するかは不明である。たとえば混入者が 2 名だった場合、3 名だった場合に各

---

たが、この理由として正規分布の仮定と実際の回答者群の分布形状の乖離が大きかったことが考えられる。同手法では決定的なズレを項目ごとに判定するため、手法 A と比べて「外れ値」出現回数が少なかった。このため正規分布を仮定した有意水準よりもポアソン分布を仮定した有意水準のほうが除外判定にあたって適切であった可能性がある。

条件の判定手法がどの程度の精度を示すのかについては、今後さらなる検証が求められる。

## 引用文献

- 阿部信一 (2004). 感度・特異度・精度 医学図書館, 51(4), 387-388.
- 尾崎幸謙・鈴木貴士 (2019). 機械学習による不適切回答者の予測 行動計量学, 46(2), 39-52.
- 折笠秀樹 (2017). 正規性の確認方法について 薬理と治療, 45(12), 1993-1995.
- 救仁郷 誠 (2001). マハラノビスの距離入門——MTS 法を理解するために—— 品質工学, 9(1), 13-21.
- 高橋和子 (2018). 機械学習を適用した自由回答収集時における有効情報追加システムの構想——職業コーディングを例として—— データ分析の理論と応用, 7(1), 21-42.
- 日本学術会議 (2020). Web 調査の有効な学術的活用を目指して <http://www.scj.go.jp/ja/info/kohyo/kohyo-24-t292-3-abstract.html> (2021年2月11日)
- 埴淵知哉・村中亮夫・安藤雅登 (2015). インターネット調査によるデータ収集の課題——不良回答、回答時間、および地理的特性に注目した分析—— *E-journal GEO*, 10(1), 81-98.
- 藤澤洋徳 (2017). ロバスト統計——外れ値への対処の仕方—— 近代科学社
- 本多則恵 (2006). インターネット調査・モニター調査の特質——モニター型インターネット調査を活用するための課題—— 日本労働研究雑誌, 551, 32-41.
- 三浦麻子・小林哲郎 (2016). オンライン調査における努力の最小限化 (Satisfice) を検出する技法：大学生サンプルを用いた検討 社会心理学研究, 32, 123-132.
- 労働政策研究・研修機構 (2018). 仕事の世界の見える化に向けて——職業情報提供サイト (日本版 O-NET) の基本構想に関する研究—— JILPT 資料シリーズ No.203
- 労働政策研究・研修機構 (2020). 職業情報提供サイト (日本版 O-NET) のインプットデータ開発に関する研究 JILPT 資料シリーズ No.227
- 和田かず美 (2010). 多変量外れ値の検出——MSD 法とその改良手法について—— 統計研究彙報, 67, 89-157.
- Kamakura, T., Matsubara, A., & Matsumoto, S. (2020). Development of the Input Data for the Occupational Information Network of Japan. *Japan Labor Issues*, 4(25), 13-21.

- Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*2(1), 49–55.
- U.S. Department of Labor. 2018. *O\*NET Data Collection Program: Office of Management and Budget clearance package supporting statement*. Retrieved from <https://www.onetcenter.org/reports/omb2018.html> (February 9, 2021)

付録 1. 使用した 110 項目の各群における平均値、標準偏差、尖度、歪度 1/3

No.	name	変数ラベル	麻薬取締官 (n=28)			入国警備官 (n=25)			検察官 (n=20)			家庭裁判所 調査官 (n=20)			外交官 (n=21)			異職業 混入者 (n=100)			ランダム 混入者 (n=100)									
			M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度								
1	interest1	現実的	3.9	1.1	0.7	-1.0	3.2	1.2	-0.3	-0.6	3.8	0.7	-0.8	0.4	3.0	1.0	0.1	-0.4	3.4	1.2	-0.5	-0.5	3.0	1.4	-1.3	0.0				
2	interest2	研究的	3.9	1.1	1.5	-1.3	3.8	1.1	1.6	-1.3	4.2	0.7	-0.5	-0.2	4.0	0.7	4.1	-1.6	3.1	1.3	-0.9	-0.2	2.7	1.3	-1.1	0.3				
3	interest3	芸術的	2.2	1.1	1.4	1.1	1.6	0.8	-0.9	2.7	0.6	-0.4	0.0	2.7	0.8	0.9	-0.9	3.1	0.9	1.0	2.6	1.2	-0.7	0.3	3.1	1.5	-1.4	0.0		
4	interest4	社会的	4.5	0.7	3.9	-1.9	4.2	0.8	0.6	-1.0	4.9	0.5	11.9	-3.4	4.9	0.3	7.0	-2.9	4.5	0.5	2.2	0.2	3.5	1.2	-0.7	-0.3	2.8	1.5	-1.4	0.2
5	interest5	企業的	3.8	1.0	1.4	-0.9	3.6	1.0	0.8	-0.5	3.9	0.9	-1.1	-0.1	3.8	0.7	1.3	-0.7	4.1	0.9	-0.2	-0.7	3.1	1.1	-0.6	-0.2	3.1	1.4	-1.4	0.0
6	interest6	慣習的	3.6	1.1	-0.4	-0.4	4.0	0.8	-0.2	-0.5	3.4	0.9	-0.6	0.3	3.1	0.9	-0.5	0.3	3.6	1.2	0.7	-0.8	3.1	1.1	-0.6	-0.2	3.0	1.3	-1.2	0.1
7	value1	達成感	4.3	0.9	1.9	-1.4	3.9	1.0	-0.4	-0.6	4.4	0.7	-0.4	0.8	4.0	0.7	2.2	-0.9	4.0	1.0	3.8	-1.7	3.5	1.1	-0.1	-0.6	3.0	1.4	-1.2	0.0
8	value2	自律性	3.9	1.1	0.6	-1.0	3.5	1.3	-1.3	-0.1	4.7	0.6	2.9	-1.8	4.2	0.7	-0.7	-0.3	3.1	1.1	-1.1	-0.1	3.5	1.0	0.1	-0.5	2.9	1.6	-1.5	0.1
9	value3	社会的認知・地位	3.1	1.0	0.5	-0.3	3.0	1.2	-0.6	-0.2	4.4	0.8	-0.6	-0.9	3.9	0.7	-0.9	0.2	3.4	1.1	0.9	-0.6	3.0	1.1	-0.4	-0.1	3.1	1.4	-1.3	-0.2
10	value4	良好な対人関係	3.3	1.1	-0.3	-0.3	3.4	1.1	0.0	-0.7	4.0	1.0	-0.6	-0.6	3.7	1.0	-1.3	0.3	3.8	0.6	-0.4	0.2	3.4	1.0	-0.2	-0.4	3.1	1.4	-1.3	0.0
11	value5	労働条件(雇用や報酬の安定性)	2.3	1.0	-1.1	0.2	3.8	1.2	1.0	-1.2	4.3	1.0	-0.8	-0.9	4.3	0.7	-0.8	-0.6	4.3	0.8	-1.1	-0.6	3.1	1.1	-0.5	-0.4	3.3	1.4	-1.2	-0.3
12	value6	労働安全衛生	1.8	0.8	0.3	0.8	2.6	1.2	-0.8	0.1	3.8	0.8	0.0	-0.2	4.3	0.7	-0.8	-0.6	3.4	0.9	-0.4	0.2	3.2	1.0	-0.3	-0.2	3.0	1.4	-1.3	-0.2
13	value7	組織的な支援体制	2.1	1.0	-1.2	0.3	3.1	1.2	-0.9	-0.2	4.2	1.0	-0.3	-1.0	3.9	0.6	-0.3	0.1	3.6	1.0	-0.9	0.0	3.0	0.9	0.0	0.0	3.3	1.3	-1.0	-0.2
14	value8	専門性	3.8	0.9	-0.7	-0.2	3.3	1.2	-1.1	0.1	4.9	0.3	7.0	-2.9	4.9	0.3	7.0	-2.9	4.2	0.8	1.2	-1.0	3.4	1.0	0.1	-0.5	3.3	1.3	-1.0	-0.3
15	value9	奉仕・社会貢献	3.9	1.0	1.3	-1.0	3.1	1.2	-0.9	-0.2	5.0	0.0	—	—	4.9	0.4	2.8	-2.1	3.4	0.9	-0.5	0.0	2.9	1.1	-0.5	0.0	3.1	1.4	-1.3	-0.1
16	value10	私生活との両立	1.8	0.8	0.2	0.9	2.9	1.3	-0.8	0.2	2.7	1.1	-0.3	0.3	4.1	0.8	-1.3	-0.2	2.6	1.0	-0.9	-0.4	3.1	1.1	-0.6	-0.2	2.8	1.5	-1.4	0.1
17	context1	他者とのかかわり	4.8	0.6	6.0	-2.6	4.9	0.4	15.3	-3.9	5.0	0.0	—	—	5.0	0.2	20.0	-4.5	4.8	0.6	5.2	-2.5	4.2	1.3	0.8	-1.5	3.1	1.4	-1.3	-0.1
18	context2	対面での議論	3.7	1.3	0.1	-1.0	4.0	1.2	0.0	-1.0	5.0	0.2	20.0	-4.5	4.5	0.6	-0.5	-0.6	4.4	0.8	2.5	-1.5	3.5	1.4	-0.9	-0.6	2.8	1.3	-1.1	0.3
19	context3	電話での会話	4.5	0.7	-0.2	-0.9	4.6	0.6	1.0	-1.4	5.0	0.0	—	—	4.8	0.4	-0.5	-1.3	4.9	0.5	12.6	-3.5	3.8	1.4	-0.4	-0.9	3.3	1.4	-1.1	-0.4
20	context4	ビジネス/スターやメモの作成	4.3	0.9	0.3	-1.0	4.2	1.3	1.2	-1.5	4.7	0.9	13.4	-3.5	4.7	0.6	1.6	-1.5	4.2	1.0	3.3	-1.7	3.3	1.3	-1.0	-0.4	3.2	1.4	-1.2	-0.2
21	context5	仕事上での他者との対立	3.1	1.2	-0.8	-0.3	3.7	1.3	-0.5	-0.6	4.9	0.3	7.0	-2.9	3.4	1.1	-0.4	-0.7	3.0	1.2	-0.8	-0.3	2.6	1.2	-1.0	0.2	2.9	1.3	-0.9	0.2
22	context6	時間的切迫	3.5	0.9	1.3	-0.5	3.8	1.1	0.2	-0.7	4.8	0.4	-0.5	-1.3	3.8	0.7	-0.8	0.4	3.7	1.1	0.7	-0.8	3.4	1.3	-0.7	-0.5	3.1	1.4	-1.3	0.0
23	context7	グループやチームでの仕事	4.5	0.7	3.9	-1.9	4.4	0.8	-0.7	-0.9	4.4	0.9	1.3	-1.3	4.5	0.7	0.1	-1.1	4.2	0.9	-0.3	-0.8	3.2	1.2	-1.2	0.0	3.1	1.5	-1.4	-0.1
24	context8	外部の顧客等との接触	3.9	1.1	-0.7	-0.7	3.8	1.0	-0.9	-0.2	4.4	0.8	-0.8	-0.9	4.7	0.5	-1.7	-0.7	3.2	1.2	-1.1	0.0	3.3	1.2	-0.9	-0.1	3.1	1.4	-1.4	0.0
25	context9	他者と調整し、リードする	4.0	0.8	-1.4	-0.1	4.0	1.0	-0.8	-0.5	4.5	0.7	0.1	-1.1	4.5	0.7	-0.2	-0.9	3.9	1.0	-1.0	-0.3	3.2	1.1	-0.8	-0.1	3.0	1.4	-1.3	-0.1
26	context10	厳密さ、正確さ	4.6	0.6	1.6	-1.6	4.5	0.7	-0.2	-1.0	4.9	0.3	7.0	-2.9	4.7	0.5	-1.2	-0.9	4.7	0.6	3.2	-1.9	3.7	1.1	-0.7	-0.4	3.1	1.4	-1.3	-0.1
27	context11	同一作業の反復	3.5	1.0	0.1	-0.5	3.4	1.2	-1.1	0.0	2.8	1.1	-0.3	0.8	2.5	0.8	-0.3	0.0	3.4	1.1	-0.5	-0.5	3.1	1.1	-0.5	0.1	3.2	1.5	-1.4	-0.2
28	context12	機器等の速度に応じた作業	3.1	1.2	-0.8	0.1	2.3	1.3	-1.1	0.4	1.8	1.1	3.0	1.7	1.8	1.0	-0.8	0.8	2.2	1.1	-1.2	0.4	2.7	1.2	-0.9	0.1	3.1	1.4	-1.4	0.0
29	context13	結果・成果への責任	3.8	1.0	-1.1	-0.3	3.4	1.2	-0.6	-0.2	4.3	1.0	-0.2	-1.1	3.9	1.1	0.5	-0.9	3.3	1.1	-1.1	0.2	3.4	1.1	-0.5	-0.1	3.1	1.5	-1.3	-0.1
30	context14	空調のきいた屋内作業	4.8	0.5	3.7	-2.0	4.9	0.3	9.6	-3.3	4.9	0.4	20.0	-4.5	4.8	0.5	7.4	-2.7	4.8	0.5	5.1	-2.3	4.0	1.4	0.2	-1.3	3.1	1.5	-1.4	-0.1
31	context15	空調のきいていない屋内作業	2.4	1.3	-0.6	0.5	2.6	1.4	-0.9	0.5	1.9	0.9	0.4	0.8	2.5	1.3	-1.8	0.0	1.6	0.9	1.0	1.4	2.8	1.6	-1.6	0.2	3.1	1.4	-1.4	-0.1
32	context16	屋外作業	3.2	1.1	-0.7	0.0	3.1	1.4	-1.3	-0.2	2.2	0.9	-0.9	0.1	3.0	1.1	-0.4	-0.8	1.2	0.4	-0.3	1.3	2.7	1.6	-1.6	0.2	2.8	1.5	-1.3	0.2
33	context17	盛り作業	4.0	0.7	1.1	-0.6	4.2	0.9	0.1	-0.8	4.5	0.5	-2.2	0.2	4.0	0.7	-0.9	0.0	4.7	0.7	8.7	-2.8	3.2	1.4	-1.3	-0.1	3.2	1.4	-1.3	-0.2
34	context18	立ち作業	2.1	0.6	0.4	0.0	2.0	0.8	0.8	0.8	1.9	0.4	2.8	-2.1	1.7	0.6	-0.4	0.0	1.4	0.5	-2.1	0.3	3.0	1.3	-1.2	0.1	3.0	1.4	-1.2	0.0
35	context19	反復作業	2.6	1.0	0.5	0.7	2.4	1.2	0.0	0.7	1.8	0.8	1.1	1.0	1.6	0.6	-0.6	0.4	1.9	0.8	1.6	0.9	2.9	1.3	-1.1	0.1	3.0	1.5	-1.5	-0.1
36	context20	ミスの影響度	4.3	0.8	0.2	-0.9	3.9	1.2	-1.4	-0.5	4.9	0.4	2.8	-2.1	4.4	0.9	0.5	-1.2	3.5	1.0	-0.8	0.2	3.2	1.2	-1.1	0.2	3.0	1.5	-1.4	0.0
37	context21	意思決定の自由	3.5	0.7	0.1	-0.7	3.4	0.6	-0.7	-0.1	4.3	0.6	-0.4	0.0	4.1	0.6	0.8	0.1	3.1	0.8	-1.2	0.2	3.3	1.0	0.0	-0.5	3.1	1.3	-1.1	-0.1
38	context22	仕事の構造化	3.4	0.7	-0.5	-0.7	3.5	0.7	-0.2	-0.9	4.1	0.3	7.0	-2.9	3.9	0.3	7.0	-2.9	3.2	0.9	0.3	-0.9	3.3	1.0	-0.4	-0.5	3.0	1.6	-1.6	0.0

付録1. 使用した110項目の各群における平均値、標準偏差、尖度、歪度 2/3

No.	name	麻薬取締官 (n=28)			入国警備官 (n=25)			検察官 (n=20)			家庭裁判所 調査官 (n=20)			外交官 (n=21)			異職業 混入者 (n=100)			ランダム 混入者 (n=100)											
		M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度	M	SD	尖度	歪度										
39	skill1	7	4.8	1.0	0.7	-0.7	4.1	1.1	1.5	0.7	6.6	0.8	5.9	-2.2	5.7	0.6	1.6	-1.5	5.5	1.1	-0.6	3.6	1.9	-0.7	-0.1	3.2	2.5	-1.4	0.2		
40	skill2	7	5.9	1.6	6.2	-2.3	5.5	1.3	-0.7	-0.6	6.9	0.4	2.8	-2.1	6.7	0.5	-1.2	-0.9	5.8	1.0	-0.6	-0.5	4.1	1.7	0.2	-0.3	3.2	2.2	-1.0	0.3	
41	skill3	7	4.8	0.9	-0.2	0.9	4.5	1.2	1.4	0.0	6.3	1.0	-0.3	-0.9	5.8	0.8	0.0	-0.2	5.6	1.1	-0.1	-0.7	3.6	1.9	-0.6	-0.3	3.3	2.1	-1.1	0.1	
42	skill4	7	4.8	0.8	0.2	0.9	4.2	0.9	3.2	0.5	6.2	1.0	-0.4	-0.8	5.7	0.7	-0.8	0.6	5.7	1.0	-0.6	-0.3	3.8	1.7	0.1	-0.4	3.4	2.3	-1.2	0.1	
43	skill5	7	3.3	1.4	0.9	0.6	2.8	1.2	0.3	-0.2	2.1	1.4	-1.2	-0.6	1.6	1.5	-0.3	0.7	5.6	1.3	2.6	-1.7	1.9	1.9	-0.5	0.7	3.0	2.3	-1.0	0.4	
44	skill6	7	2.9	1.6	0.7	0.3	2.6	1.0	2.0	0.4	1.5	1.4	-1.5	0.2	1.1	1.1	-1.3	0.4	5.5	1.3	1.4	-1.4	1.4	1.6	0.0	1.0	3.6	2.2	-1.1	0.0	
45	skill7	7	2.8	1.6	0.5	0.6	2.5	1.3	0.3	0.0	1.2	1.2	-1.2	-1.5	0.3	1.1	1.1	-1.3	0.4	4.9	1.7	2.7	-1.4	1.4	1.7	-0.3	1.0	3.1	2.5	-1.4	0.2
46	skill8	7	2.9	1.7	-0.1	0.3	2.6	1.2	1.0	-0.5	1.3	1.3	-1.7	0.2	1.2	1.1	-1.3	0.2	5.4	1.8	3.5	-1.9	1.3	1.7	1.4	3.2	2.1	-0.9	0.2		
47	skill9	7	1.6	1.3	-1.2	-0.2	1.0	1.1	-1.3	0.5	2.4	1.1	0.4	-1.1	1.8	1.2	-1.4	-0.5	2.6	1.2	0.4	-0.4	2.2	1.6	0.5	0.2	3.4	2.1	-1.1	0.0	
48	skill10	7	2.3	1.8	-1.7	-0.1	0.4	0.6	1.0	1.4	2.4	1.3	-0.8	-0.4	1.9	1.8	-1.9	0.2	1.9	1.7	-1.8	0.0	1.7	1.9	-0.6	0.8	3.5	2.4	-1.3	0.0	
49	skill11	7	4.3	1.5	3.3	-1.6	3.8	1.5	2.1	-1.4	6.6	0.6	0.6	-1.2	5.6	1.0	-0.9	0.0	5.3	1.0	1.1	-1.1	2.7	2.0	-0.9	0.1	3.8	2.2	-1.1	-0.2	
50	skill12	7	4.1	1.5	1.9	-1.0	3.0	1.3	1.0	-0.8	5.1	1.9	1.8	-1.4	5.0	1.5	-0.3	-1.0	5.1	0.8	0.7	0.6	3.0	1.9	-0.5	0.1	3.3	2.2	-1.1	0.0	
51	skill13	7	3.5	1.4	0.8	-1.2	3.0	1.3	0.3	-0.7	4.2	1.2	0.5	4.6	1.1	1.2	-0.3	3.6	1.8	-0.2	-0.5	2.8	1.6	-0.5	-0.2	3.6	2.4	-1.2	-0.2		
52	skill14	7	3.8	1.3	2.3	-1.4	3.4	1.4	1.2	-0.7	5.1	1.0	-0.2	-0.2	4.6	1.1	0.2	-0.7	4.5	1.4	4.2	-1.7	3.1	1.7	-0.4	0.0	3.4	2.3	-1.1	0.1	
53	skill15	7	4.7	1.2	6.8	-1.9	4.1	1.4	2.2	-1.3	6.4	0.8	2.7	-1.6	6.4	0.7	-0.5	-0.5	4.9	1.3	-1.0	0.2	3.4	1.6	-0.3	-0.1	3.4	2.2	-1.2	0.1	
54	skill16	7	4.1	1.5	1.3	-1.2	3.6	1.3	-0.2	0.0	5.8	1.2	-0.2	-0.8	5.2	1.2	1.5	-0.6	5.1	1.3	-0.1	-0.5	3.3	1.8	-0.2	0.0	3.5	2.5	-1.4	0.1	
55	skill17	7	4.9	1.6	2.4	-0.9	4.6	1.5	-0.2	-0.4	6.6	0.6	1.4	-1.4	5.7	1.0	-0.6	-0.4	5.1	1.4	-0.4	-0.6	3.2	1.6	-0.2	0.0	3.5	2.4	-1.2	0.0	
56	skill18	7	4.1	1.3	2.3	-0.9	3.7	1.1	-0.4	-0.6	5.6	1.3	-0.4	-0.8	5.1	1.2	-1.2	0.3	5.2	1.4	-0.3	-0.7	3.1	1.7	-0.4	0.0	3.4	2.4	-1.4	0.1	
57	skill19	7	4.2	1.2	6.5	-2.3	3.8	1.0	0.1	-0.1	5.0	1.0	-0.4	0.8	5.0	0.6	-0.3	0.0	4.2	0.8	2.0	-0.5	3.4	1.6	-0.4	-0.2	3.6	2.2	-1.1	0.1	
58	skill20	7	2.9	2.0	-1.2	-0.5	2.9	1.5	-0.1	-0.4	4.8	1.5	-0.2	-0.5	5.5	1.1	0.3	-0.6	3.4	2.0	-0.7	-0.6	2.9	1.8	-0.9	-0.2	3.5	2.4	-1.3	0.0	
59	skill21	7	4.0	0.9	0.0	0.1	3.3	1.3	0.2	-0.7	6.1	1.0	-0.4	-0.8	4.9	1.4	6.7	-2.0	4.7	1.7	1.4	-0.9	3.1	1.7	-0.8	-0.3	3.7	2.3	-1.2	0.0	
60	skill22	7	2.6	1.8	-1.0	-0.1	1.8	1.7	-1.1	0.4	2.6	2.6	-1.7	0.2	1.9	2.0	-1.0	0.6	2.7	2.5	-1.5	0.1	2.7	1.9	-0.7	0.0	3.6	2.4	-1.3	0.0	
61	skill23	7	1.8	1.8	-1.3	0.3	0.7	1.3	2.2	1.8	1.9	2.3	-1.0	0.7	1.8	2.1	-1.1	0.7	1.5	2.2	-0.5	1.1	2.5	2.0	-1.0	0.2	3.7	2.3	-1.2	-0.2	
62	skill24	7	3.1	1.8	-0.7	-0.7	1.8	1.6	-1.0	0.4	2.1	1.9	-1.7	0.1	1.4	1.9	-1.2	0.8	1.6	1.8	1.8	0.4	3.0	1.8	-0.6	0.1	3.3	2.3	-1.2	0.1	
63	skill25	7	2.6	2.1	-1.5	-0.2	1.3	1.3	-1.0	0.5	1.6	1.7	-1.0	0.6	0.9	1.5	0.4	1.4	1.4	1.7	-1.4	0.6	2.6	1.9	-0.9	0.1	3.6	2.1	-1.0	-0.1	
64	skill26	7	1.4	1.3	-1.0	0.3	0.9	1.1	-1.2	0.6	1.0	1.1	1.3	1.1	1.0	1.2	0.6	1.2	1.0	1.4	-0.2	1.1	1.5	1.7	0.3	1.0	3.2	2.3	-1.1	0.2	
65	skill27	7	1.5	1.6	-1.2	0.4	0.9	1.2	-1.4	0.7	0.8	1.4	0.9	1.4	0.4	1.0	10.5	3.2	0.6	1.2	2.2	1.8	1.8	1.9	-0.6	0.7	3.9	2.4	-1.2	-0.3	
66	skill28	7	1.8	1.5	-1.1	0.1	1.6	1.2	-1.0	-0.2	1.0	1.2	-1.5	0.6	1.0	1.3	-0.2	0.9	1.3	1.5	-1.2	0.6	2.3	2.0	-0.4	0.6	3.7	2.5	-1.3	-0.1	
67	skill29	7	1.9	1.7	-1.6	0.1	0.8	1.2	-0.6	1.0	0.3	0.7	2.4	2.0	0.4	0.9	3.5	2.1	0.6	1.2	2.2	1.8	2.1	2.1	-0.6	0.6	3.5	2.2	-1.1	-0.1	
68	skill30	7	1.5	1.6	-1.0	0.5	1.0	1.4	1.6	1.4	0.5	1.3	8.9	2.9	0.5	0.9	1.9	1.8	0.9	1.4	1.7	1.5	2.2	1.9	-0.7	0.5	3.6	2.4	-1.4	-0.1	
69	skill31	7	1.6	1.5	-0.8	0.4	0.6	1.0	-0.4	1.1	0.3	0.7	4.3	2.4	0.6	1.0	0.5	1.4	0.8	1.3	0.4	1.2	2.2	2.0	-0.8	0.5	3.9	2.3	-1.2	-0.2	
70	skill32	7	1.6	1.7	-1.4	0.4	1.2	1.3	-1.8	0.2	1.3	1.6	-1.4	0.6	2.5	2.0	-1.7	-0.2	1.3	1.6	-0.1	0.9	2.6	2.1	-0.7	0.4	3.4	2.3	-1.3	0.0	
71	skill33	7	2.9	1.6	-0.5	-0.7	2.3	1.5	-0.8	-0.3	4.9	1.7	-0.8	-0.5	3.7	1.7	1.3	-1.6	3.9	2.0	0.1	-0.8	2.6	1.8	-0.8	0.1	3.7	2.3	-1.2	-0.1	
72	skill34	7	2.4	1.6	-1.1	-0.6	2.4	1.4	0.0	0.0	4.2	1.9	0.6	-0.8	3.0	1.7	-0.6	-0.7	3.1	2.3	-1.2	-0.3	2.3	1.9	-0.5	0.5	3.4	2.3	-1.2	0.1	
73	skill35	7	1.9	1.6	-1.6	-0.1	2.2	1.5	-0.7	0.1	3.5	2.0	-0.5	-0.4	2.9	2.0	-1.3	-0.3	2.7	2.4	-1.6	0.0	2.4	1.9	-0.6	0.3	3.5	2.6	-1.5	0.0	
74	skill36	7	3.1	1.3	0.9	-0.5	3.2	1.1	-1.3	-0.2	5.7	1.2	-1.5	-0.1	3.6	1.1	0.1	0.8	4.1	1.6	0.7	-0.6	3.1	1.8	-0.4	0.2	3.3	2.3	-1.3	0.1	
75	skill37	7	1.6	1.9	-1.6	0.5	1.0	6.9	2.5	0.9	1.6	5.4	2.3	0.7	1.1	2.6	1.7	2.3	2.3	-1.6	0.2	2.1	1.9	-0.5	0.5	3.5	2.4	-1.4	0.0		
76	skill38	7	2.3	1.7	-0.8	0.0	1.2	1.3	-0.9	0.5	0.6	0.9	1.1	1.5	1.3	1.3	-0.5	0.5	1.8	1.8	-0.4	0.6	2.2	1.8	-0.6	0.5	3.1	2.3	-1.2	0.4	
77	skill39	7	3.0	2.0	-1.0	-0.4	3.1	1.3	0.6	-1.2	4.9	1.5	6.0	-1.7	4.5	0.9	-0.5	0.3	3.0	2.2	-1.5	-0.3	2.3	1.8	-0.9	0.4	3.1	2.2	-1.1	0.1	

付録1. 使用した110項目の各群における平均値、標準偏差、尖度、歪度 3/3

No.	name	変数ラベル		麻薬取締官 (n=28)			入国警備官 (n=25)			検察官 (n=20)			家庭裁判所 調査官 (n=20)			外交官 (n=21)			異職業 混入者 (n=100)			ランダム 混入者 (n=100)									
		範囲		M	SD	尖 度	歪 度	M	SD	尖 度	歪 度	M	SD	尖 度	歪 度	M	SD	尖 度	歪 度	M	SD	尖 度	歪 度								
		min	max																												
78	knowledge1	0	5	2.1	1.5	-0.9	-0.1	1.2	1.4	0.7	1.0	2.4	1.1	0.7	0.3	2.1	1.1	-0.9	0.0	1.8	1.6	-1.0	0.3	1.4	1.3	-0.9	0.4	2.6	1.7	-1.3	0.0
79	knowledge2	0	5	3.5	1.1	-0.3	-0.5	3.0	1.5	-0.4	-0.4	3.4	1.5	-1.7	0.1	3.6	1.1	0.1	-0.8	3.8	1.0	-1.2	-0.1	1.9	1.3	-0.7	0.1	2.5	1.7	-1.3	0.1
80	knowledge3	0	5	0.9	1.3	2.4	1.5	0.3	0.6	3.5	2.1	2.4	1.1	-0.1	-0.1	0.8	0.9	1.1	1.1	1.5	1.3	-1.1	0.2	1.2	1.2	-0.4	0.8	2.4	1.7	-1.3	0.1
81	knowledge4	0	5	0.6	1.3	5.2	2.3	0.1	0.3	9.6	3.3	1.1	0.9	-1.6	-0.2	0.5	0.7	0.5	1.3	0.6	0.9	1.9	1.6	1.5	1.4	-0.6	0.5	2.6	1.7	-1.3	0.0
82	knowledge5	0	5	0.9	1.4	1.8	1.5	0.6	0.8	-0.7	1.0	1.9	1.6	-1.1	0.3	1.5	1.6	-0.3	0.9	1.2	1.2	-1.5	0.3	2.0	1.5	-1.0	0.1	2.6	1.7	-1.3	-0.1
83	knowledge6	0	5	2.2	1.6	-1.0	-0.1	1.0	1.4	0.9	1.2	2.0	1.0	0.1	-0.3	2.2	1.3	-0.4	-0.7	1.7	1.5	-1.3	0.2	1.3	1.3	-0.9	0.5	2.4	1.8	-1.4	0.0
84	knowledge7	0	5	0.8	1.2	5.0	1.9	1.0	1.3	2.0	1.4	0.4	0.7	1.2	1.5	0.2	0.4	0.7	1.6	0.8	1.4	3.8	2.1	1.1	1.3	-0.3	0.9	2.3	1.6	-1.2	0.0
85	knowledge8	0	5	0.1	0.4	2.9	2.2	0.0	0.0	—	—	0.6	0.8	-0.6	0.9	0.1	0.3	7.0	2.9	0.2	0.6	5.2	2.5	1.2	1.4	-0.6	0.8	2.3	1.6	-1.0	0.2
86	knowledge9	0	5	0.3	1.0	18.7	4.2	0.0	0.0	—	—	0.7	0.7	-0.8	0.6	0.1	0.3	7.0	2.9	0.3	0.6	3.5	2.2	0.9	1.1	1.4	2.3	1.7	-1.3	0.1	
87	knowledge10	0	5	0.3	0.6	4.2	2.4	0.0	0.2	25.0	5.0	1.2	1.1	0.5	0.9	0.2	0.4	2.8	2.1	0.6	1.1	4.2	2.1	1.0	1.3	0.3	1.1	2.3	1.7	-1.1	0.3
88	knowledge11	0	5	1.7	1.5	-0.5	0.4	0.6	1.0	0.4	1.4	1.6	1.1	-0.5	0.3	0.7	0.8	-1.0	0.8	0.6	1.1	0.5	1.5	1.1	1.3	0.3	1.0	2.3	1.7	-1.2	0.1
89	knowledge12	0	5	0.5	1.1	4.3	2.3	0.0	0.2	25.0	5.0	1.0	0.8	-1.4	0.0	0.2	0.5	11.9	3.4	0.3	0.7	2.1	1.9	0.9	1.3	1.3	1.4	2.4	1.8	-1.4	0.1
90	knowledge13	0	5	0.1	0.4	17.4	4.1	0.1	0.3	9.6	3.3	1.1	0.9	-1.6	-0.2	0.1	0.3	7.0	2.9	0.4	1.0	7.1	2.7	1.0	1.4	0.1	1.1	2.3	1.8	-1.3	0.2
91	knowledge14	0	5	1.0	1.1	-1.1	0.6	0.3	0.6	2.5	1.9	1.1	0.9	-0.8	0.3	0.2	0.4	2.8	2.1	0.4	0.7	1.3	1.7	1.2	1.5	-0.3	0.9	2.8	1.7	-1.3	-0.2
92	knowledge15	0	5	0.6	0.9	0.6	1.3	0.4	0.7	0.5	1.2	1.4	0.9	-0.8	-0.4	1.0	0.8	-1.2	0.1	1.0	1.2	-1.3	0.6	1.4	1.3	-1.1	0.4	2.3	1.7	-1.3	0.1
93	knowledge16	0	5	0.5	1.0	2.4	1.9	0.0	0.0	—	—	1.3	1.0	-0.8	0.2	0.2	0.4	0.7	1.6	0.6	1.0	0.2	1.2	1.1	1.4	0.0	1.0	2.6	1.6	-1.1	-0.3
94	knowledge17	0	5	2.3	1.6	-0.9	0.1	0.0	0.0	—	—	1.3	1.0	-1.0	0.1	0.2	0.4	0.7	1.6	0.6	0.9	0.7	1.3	1.2	1.5	-0.1	1.0	2.6	1.6	-1.2	0.0
95	knowledge18	0	5	1.4	1.6	-0.1	1.0	0.1	0.4	25.0	5.0	1.5	1.1	-0.4	0.2	1.4	1.7	0.1	1.1	0.4	0.8	4.4	2.1	1.3	1.6	-0.2	1.0	2.1	1.6	-0.9	0.5
96	knowledge19	0	5	2.4	1.3	0.0	0.0	2.1	1.5	-1.4	-0.1	3.3	1.2	-0.9	0.1	5.0	0.2	20.0	-4.5	1.4	1.4	-1.4	0.3	1.5	1.4	-0.1	0.7	2.9	1.5	-1.1	-0.3
97	knowledge20	0	5	1.6	1.3	-1.4	-0.1	2.2	1.4	-1.1	-0.4	3.0	1.3	0.1	-0.2	4.7	0.6	1.6	-1.5	2.0	1.6	-1.0	0.1	1.3	1.5	0.2	1.0	2.5	1.7	-1.4	0.0
98	knowledge21	0	5	0.6	1.1	3.1	1.9	0.9	1.5	0.0	1.3	1.2	0.9	-1.1	0.1	0.7	0.9	1.3	1.3	1.9	1.8	-1.7	0.2	1.0	1.2	-0.5	0.9	2.4	1.7	-1.3	0.0
99	knowledge22	0	5	1.8	1.6	-0.7	0.5	1.3	1.5	0.5	1.0	3.3	1.2	-1.5	0.2	3.7	1.0	1.6	-1.0	0.4	0.7	1.3	1.7	1.5	1.7	-0.5	0.9	2.5	1.8	-1.3	0.0
100	knowledge23	0	5	2.0	1.4	-0.9	0.1	1.3	1.4	0.2	0.8	2.8	1.4	-0.6	0.1	4.6	0.7	1.2	-1.5	0.4	0.9	3.2	2.0	1.2	1.5	0.2	1.1	2.6	1.8	-1.4	0.0
101	knowledge24	0	5	1.8	1.4	-1.2	0.0	1.8	1.3	-1.0	0.1	2.1	0.8	-1.2	-0.1	3.7	0.8	0.2	-0.5	1.2	1.5	-0.5	0.9	1.4	1.4	-0.9	0.5	2.8	1.7	-1.2	-0.1
102	knowledge25	0	5	3.0	1.1	0.9	-0.6	3.3	1.2	0.8	-0.7	4.3	1.1	0.0	-1.2	4.3	0.6	-0.4	0.0	3.5	1.4	0.1	-0.6	1.5	1.5	-0.5	0.7	2.4	1.7	-1.2	0.1
103	knowledge26	0	5	2.3	0.9	-0.7	0.2	2.4	1.1	0.9	0.5	1.7	1.0	0.0	0.4	1.2	0.9	-0.6	0.2	4.2	1.1	-0.2	1.1	1.1	1.5	0.3	1.2	2.7	1.7	-1.2	-0.2
104	knowledge27	0	5	0.3	0.6	2.3	1.8	0.0	0.2	25.0	5.0	0.6	0.8	-0.6	0.9	0.9	1.1	-0.7	0.8	1.3	1.4	-1.4	0.3	0.9	1.4	0.8	1.4	2.4	1.7	-1.3	0.0
105	knowledge28	0	5	0.4	0.6	1.6	1.6	1.0	1.1	0.5	0.9	1.4	1.3	0.0	0.8	0.9	0.9	-1.9	0.3	3.3	1.5	0.3	-0.8	0.9	1.2	-0.2	1.0	2.5	1.9	-1.5	0.1
106	knowledge29	0	5	0.6	0.8	-0.7	1.0	1.7	1.3	0.1	0.4	1.6	1.4	0.5	0.9	1.8	1.1	0.1	-0.2	2.4	1.6	-0.9	-0.1	0.8	1.2	1.4	2.4	1.8	-1.3	0.1	
107	knowledge30	0	5	2.1	1.4	-1.1	-0.2	1.6	1.5	-0.8	0.4	2.1	1.3	1.1	1.1	1.6	1.6	-1.5	0.3	1.8	1.7	-0.9	0.4	1.5	1.6	-0.7	0.7	2.4	1.7	-1.2	0.1
108	knowledge31	0	5	4.1	1.1	1.8	-1.4	3.8	1.2	2.3	-1.2	4.9	0.7	20.0	-4.5	4.8	0.4	-0.5	-1.3	3.2	1.3	0.3	-0.5	1.3	1.4	-0.3	0.8	2.5	1.7	-1.2	0.0
109	knowledge32	0	5	1.2	1.3	-0.3	0.9	0.4	0.6	1.6	1.5	1.4	1.2	1.2	1.2	0.7	0.9	-1.2	0.8	1.4	1.4	-1.4	0.3	1.0	1.3	0.1	1.1	2.3	1.6	-1.1	0.1
110	knowledge33	0	5	1.6	1.4	-1.5	0.2	0.8	1.3	0.6	1.4	1.5	1.1	0.0	0.5	1.7	1.6	-0.8	0.4	1.8	1.5	-1.3	-0.1	1.5	1.5	-0.7	0.7	2.5	1.7	-1.3	-0.1

付録 2. ランダム抽出 100 ケースの職業（割当番号の小さい順に並べ替えたもの）

抽出番号	割当番号	回答者職業	抽出番号	割当番号	回答者職業
抽出 1	227	消防官	抽出 5 1	13736	広告ディレクター
抽出 2	590	水産ねり製品製造	抽出 5 2	13803	土木・建築工学研究者
抽出 3	626	作業療法士 (OT)	抽出 5 3	13823	会社経営者
抽出 4	900	しょうゆ製造	抽出 5 4	14058	電気技術者
抽出 5	1113	国家公務員 (行政事務)	抽出 5 5	14296	動画制作
抽出 6	1628	測量士	抽出 5 6	14552	電子機器技術者
抽出 7	2363	防水工	抽出 5 7	14669	織布工/織機オペレーター
抽出 8	2435	住宅・不動産営業	抽出 5 8	14767	玩具 (おもちゃ) 製作: 主に企画開発、個人制作
抽出 9	2620	電器店店員	抽出 5 9	15870	高分子化学技術者
抽出 1 0	2828	歯科衛生士	抽出 6 0	16234	グラフィックデザイナー
抽出 1 1	2931	スーパー店長	抽出 6 1	17022	製品包装作業員
抽出 1 2	2967	行政書士	抽出 6 2	17028	宅配便配達員
抽出 1 3	3974	ホームセンター店員	抽出 6 3	17344	ボイラーオペレーター
抽出 1 4	4251	専門学校教員	抽出 6 4	17360	製版オペレーター、DTPオペレーター
抽出 1 5	4522	ビル施設管理	抽出 6 5	17597	施設管理者 (介護施設)
抽出 1 6	4751	弁理士	抽出 6 6	18118	フロント (ホテル・旅館)
抽出 1 7	4753	水産技術者	抽出 6 7	18189	受付事務
抽出 1 8	5325	図書館司書	抽出 6 8	18415	宅配便配達員
抽出 1 9	5453	ピアノ調律師	抽出 6 9	18476	精神科医
抽出 2 0	5543	通関士	抽出 7 0	18548	福祉ソーシャルワーカー
抽出 2 1	5717	フラワーショップ店員	抽出 7 1	18752	障害者福祉施設指導専門員 (生活支援員、就労支援員等)
抽出 2 2	5803	歯科衛生士	抽出 7 2	19134	医療事務
抽出 2 3	5901	遊園地スタッフ	抽出 7 3	19254	産業廃棄物処理技術者
抽出 2 4	5997	レンタカー店舗スタッフ	抽出 7 4	19362	カウンセラー (医療福祉分野)
抽出 2 5	6182	臨床工学技士	抽出 7 5	19425	銀行・信用金庫渉外担当
抽出 2 6	6617	惣菜製造	抽出 7 6	19729	銀行・信用金庫渉外担当
抽出 2 7	6721	ネイリスト	抽出 7 7	20109	老人福祉施設生活相談員
抽出 2 8	6987	映像編集者	抽出 7 8	20330	フォークリフト運転作業員
抽出 2 9	7011	書店員	抽出 7 9	20731	計器組立
抽出 3 0	7014	臨床検査技師	抽出 8 0	20822	施設管理者 (介護施設)
抽出 3 1	7026	言語聴覚士	抽出 8 1	20927	動物看護
抽出 3 2	7253	歯科技工士	抽出 8 2	21287	医薬品販売/登録販売者
抽出 3 3	7383	デイスパッチャー (航空機運航管理者)	抽出 8 3	21335	児童指導員
抽出 3 4	7619	動物園飼育員	抽出 8 4	21724	ハウス野菜栽培者
抽出 3 5	7654	エレベーター据付	抽出 8 5	21757	貿易事務
抽出 3 6	8152	鉄道車両清掃	抽出 8 6	22201	精神科医
抽出 3 7	8442	コンビニエンスストア店員	抽出 8 7	22588	看板制作
抽出 3 8	8529	臨床工学技士	抽出 8 8	22859	ネット通販の運営
抽出 3 9	8738	鉄骨工	抽出 8 9	22999	証券外務員
抽出 4 0	9658	弁理士	抽出 9 0	23039	花火師
抽出 4 1	10184	駐車場管理	抽出 9 1	23207	新聞配達員
抽出 4 2	10663	日本料理調理人 (板前)	抽出 9 2	23878	知的財産サーチャージャー
抽出 4 3	10690	とび	抽出 9 3	24238	大学・短期大学教員 (教育学)
抽出 4 4	11367	臨床検査技師	抽出 9 4	24512	小児科医
抽出 4 5	11547	映像編集者	抽出 9 5	24786	積卸作業員
抽出 4 6	11559	客室乗務員	抽出 9 6	24896	治験コーディネーター
抽出 4 7	11890	カフェ店員	抽出 9 7	24963	アートディレクター
抽出 4 8	12021	鉄道線路管理	抽出 9 8	25336	施設警備員
抽出 4 9	12315	駅務員	抽出 9 9	25539	広報・PR担当
抽出 5 0	12388	アナウンサー	抽出 1 0 0	25688	損害保険事務

※サンプルサイズ不足のため職業情報提供サイト（日本版 O-NET）には掲載されていない職業の就業者が含まれる場合がある。また職業区分は考慮せずランダム抽選を行ったため、偶然同じ職業の就業者が複数回抽出される場合がある。

### 付録 3. 手法 A、B のシミュレーションに使用した R スクリプト

```
1 #Leveneの等分散検定用ライブラリ
2 library(lawstat)
3
4 #真の回答者群をdata1、偽の回答者プール100名をdata2として取り込む。
5 data1 <- read.table("matori_truth.csv", header=TRUE, sep=",")
6 data2 <- read.table("false_100.csv", header=TRUE, sep=",")
7
8 #シミュレーションに必要なパラメータを設定する。
9 outlier_level <- 2 #1ならば、±1 SDの範囲を超える場合に「外れ値」とする。2なら、2
SD。
10 amount_criterion <- 0.01 #外れ値の回数が、群から見てどの程度多いとき除外対象とする
か。片側5%なら0.05、片側2.5% (両側5%相当) なら0.025、片側1%なら0.01。
11 robust_outlier <- 0 #「外れ値」判定においてロバスト推定を行う (1)か、行わない (0)
か。
12 robust_amount <- 0 #「外れ値発生回数」に基づく除外対象判定時に、ロバスト推定を行う
(1)か、行わない(0)か。
13 trim_rate <- 0.1 #仮にロバスト推定を行う場合に、どの程度刈り込むのか。0.1ならば、上
位下位それぞれ10%を刈り込むことになる。
14
15 #外れ値発生回数を出力するための空行列を準備しておく。
16 final_outlier_count <- matrix(nr=nrow(data1)+1,nc=nrow(data2))
17
18 #除外判定結果を収納するための空行列を準備しておく。
19 final_judge <- matrix(nr=nrow(data1)+1,nc=nrow(data2))
20 final_judge_full_poisson <- matrix(nr=nrow(data1)+1,nc=nrow(data2))
21 final_judge_part_poisson <- matrix(nr=nrow(data1)+1,nc=nrow(data2))
22
23 #ポアソン分布適用チェックの初期化
24 check_poisson <- 0
25
26 #混入シミュレーション開始
27 for(i in 1:nrow(data2)){
28     #まず、真の回答者群にi番目の偽の回答者を1名加え、新たにdata_tempを作
る。
29     data_temp <- rbind(data1, data2[i,])
30
31     #外れ値判定結果を保存する、同じ大きさの空行列を作る。
32     data_temp_outlier <- matrix(nr=nrow(data_temp),nc=ncol(data_temp))
33
34     #1名ずつ、人数回 (真人数+1) のループ
35     for (j in 1:nrow(data_temp)){
36         #項目数回、ループ
37         for (k in 1:ncol(data_temp)){
38             #もし、j番目の人物のk番目の素点がNAな
らば、自動的に1 (外れ値) として判定。
39             if(is.na(data_temp[j,k])==TRUE){
40                 data_temp_outlier[j,k] <-
1
41
42             #NAではなく、かつロバスト推定を行わない場
合
43             }else if(robust_outlier == 0){
44                 #もし、j番目の人物のk番項
目の粗点が、群の平均±XSD
の範囲を超えていたら1 (外
れ値) と判定。
45                 if(data_temp[j,k] <
mean(data_temp[,k],
na.rm=TRUE) -
outlier_level*sd(data_temp
[,k], na.rm=TRUE) ||
data_temp[j,k] >
mean(data_temp[,k],
na.rm=TRUE) +
```

```

46         outlier_level*sd(data_temp
           [,k], na.rm=TRUE)){
           data_temp_outlier[j,k]
47             <- 1
           #範囲内ならば、0（外れ値
           ではない）と判定。
48         }else{
49             data_temp_outlier[j,k]
           <- 0
50         }
51     #NAではなく、かつロバスト推定を行う場合
52     }else if(robust_outlier == 1){
53         #もし、j番目の人物のk番項
           目の粗点が、群の刈り込み平
           均±X(IQR/1.349)の範囲を超
           えていたら1（外れ値）と判
           定。
54         if(data_temp[j,k] <
           mean(data_temp[,k],
           na.rm=TRUE,
           trim=trim_rate) -
           outlier_level*IQR(data_tem
           p[,k], na.rm=TRUE)/1.349
           || data_temp[j,k] >
           mean(data_temp[,k],
           na.rm=TRUE,
           trim=trim_rate) +
           outlier_level *
           IQR(data_temp[,k],
           na.rm=TRUE)/1.349){
55             data_temp_outlier[j,k]
           <- 1
56         }
57         #範囲内ならば、0（外れ値
           ではない）と判定。
58         }else{
59             data_temp_outlier[j,k]
           <- 0
60         }
61     }
62 }
63 }
64 }
65 #できあがった外れ値の判定行列を使って、人物ごとの外れ値発生回数をカウ
           ント。
66 data_temp_outlier_count <- matrix(nr=nrow(data_temp), nc=1)
67 for (j in 1:nrow(data_temp)){
68     data_temp_outlier_count[j,1] <-
           sum(data_temp_outlier[j,], na.rm=TRUE)
69 }
70
71 #後で出力できるよう、外れ値発生回数を保存しておく。
72 final_outlier_count[,i] <- data_temp_outlier_count[,1]
73
74 #人物ごとに、最終的な除外判定結果を記録。
75 for (j in 1:nrow(data_temp)){
76     #ロバスト推定を行わない場
           合
77     if(robust_amount == 0){
78         if(data_temp_outlier_count[j,1] >
           mean(data_temp_outlier_count[,1],
           na.rm=TRUE) + qnorm(amount_criterion,
           lower.tail=FALSE) *
           sd(data_temp_outlier_count[,1],

```

```

79         na.rm=TRUE)) {
80             final_judge[j, i] <- 1
81         } else {
82             final_judge[j, i] <- 0
83         }
84         #完全ポアソン条件
85         if (ppois(data_temp_outlier_count[j, 1], lambda =
            mean(data_temp_outlier_count[, 1], na.rm=TRUE)) > (1 - amount_criterion)) {
86             final_judge_full_poisson[j, i] <- 1
87         } else {
88             final_judge_full_poisson[j, i] <- 0
89         }
90
91         #部分ポアソン条件
92         if (mean(final_outlier_count[, i], na.rm=TRUE) < 5 && (1 -
            (mean(final_outlier_count[, i], na.rm=TRUE)) < 5)) {
93
94             check_poisson <-
95                 check_poisson + 1
96
97             if (ppois(data_temp_outlier_count[j, 1], lambda =
                mean(data_temp_outlier_count[, 1], na.rm=TRUE)) > (1 -
                amount_criterion)) {
98                 final_judge_part_poisson[j, i] <- 1
99             } else {
100                 final_judge_part_poisson[j, i] <- 0
101             }
102         } else {
103             final_judge_part_poisson[j, i] <- final_judge[j, i]
104         }
105         #ロバスト推定を行う場合
106         if (robust_amount == 1) {
107             if (data_temp_outlier_count[j, 1] >
                mean(data_temp_outlier_count[, 1], na.rm=TRUE, trim=trim_rate) +
                qnorm(amount_criterion, lower.tail=FALSE) *
                IQR(data_temp_outlier_count[, 1], na.rm=TRUE) / 1.349) {
108                 final_judge[j, i] <- 1
109             } else {
110                 final_judge[j, i] <- 0
111             }
112
113             #完全ポアソン条件
114             if (ppois(data_temp_outlier_count[j, 1], lambda =
                mean(data_temp_outlier_count[, 1], na.rm=TRUE, trim=trim_rate)) > (1 -
                amount_criterion)) {
115                 final_judge_full_poisson[j, i] <- 1
116             } else {
117                 final_judge_full_poisson[j

```

```

118                                     }, i] <- 0
119                                     }
120                                     #部分ポアソン条件
121                                     if(mean(final_outlier_count[, i],
122                                     na.rm=TRUE, trim=trim_rate) < 5 && (1 -
123                                     (mean(final_outlier_count[, i],
124                                     na.rm=TRUE, trim=trim_rate))) < 5 ){
125                                     check_poisson <-
126                                     check_poisson + 1
127
128                                     if(ppois(data_temp_outlier
129                                     _count[j, 1], lambda =
130                                     mean(data_temp_outlier_cou
131                                     nt[, 1], na.rm=TRUE,
132                                     trim=trim_rate)) > (1 -
133                                     amount_criterion)){
134                                     final_judge_
135                                     part_poisson
136                                     [j, i] <- 1
137
138                                     }else{
139                                     final_judge_
140                                     part_poisson
141                                     [j, i] <- 0
142
143                                     }
144                                     }else if(robust_amount == 1){
145                                     final_judge_part_poisson[j
146                                     , i] <- final_judge[j, i]
147
148                                     }
149                                     }
150                                     }
151                                     }
152
153                                     #真の回答者の外れ値出現回
154                                     数
155                                     truth_outlier_count = head(rowMeans(final_outlier_count), -1)
156                                     #平均
157                                     print(mean(truth_outlier_count))
158                                     #標準偏差
159                                     print(sd(truth_outlier_count))
160
161                                     #偽の回答者100名の外れ値出現回数
162                                     false_outlier_count = final_outlier_count[nrow(final_outlier_count), ]
163                                     #平均
164                                     print(mean(false_outlier_count))
165                                     #標準偏差
166                                     print(sd(false_outlier_count))
167
168                                     #Leveneの等分散検定
169                                     TF_outlier_count <- append(truth_outlier_count, false_outlier_count)
170                                     fx <- factor(rep(c("T", "F"), c(nrow(data1), nrow(data2))))
171                                     levene.test(y=TF_outlier_count, group=fx)
172
173                                     #等分散性を仮定した t 検定
174                                     t.test(truth_outlier_count, false_outlier_count, var.equal=T)
175
176                                     #等分散性を仮定しない t 検
177                                     定
178                                     t.test(truth_outlier_count, false_outlier_count, var.equal=F)
179
180                                     #真の回答者を正しく「合格」とする比率（再現性：Recall）
181                                     recall <- matrix(nr=1, nc=nrow(data2))
182                                     recall_full_poisson <- matrix(nr=1, nc=nrow(data2))
183                                     recall_part_poisson <- matrix(nr=1, nc=nrow(data2))
184
185                                     #計算のため、偽回答者の行を削除した行列を準備。

```

```

169 final_judge_temp <- final_judge[-nrow(final_judge),]
170 final_judge_full_poisson_temp <- final_judge_full_poisson[-
nrow(final_judge_full_poisson),]
171 final_judge_part_poisson_temp <- final_judge_part_poisson[-
nrow(final_judge_part_poisson),]
172
173 for (i in 1: nrow(data2)){
174     #i番目の混入者の時の除外判定率
175     recall[1, i] <- mean(final_judge_temp[, i], na.rm=TRUE)
176     recall_full_poisson[1, i] <- mean(final_judge_full_poisson_temp[, i],
na.rm=TRUE)
177     recall_part_poisson[1, i] <- mean(final_judge_part_poisson_temp[, i],
na.rm=TRUE)
178 }
179
180 #ファイル出力。
181 #write.table(recall, "recall.csv", sep=",")
182
183 #再現性の平均値（論文中で「再現性」としているもの）
184 print(1-mean(recall[1, ], na.rm=TRUE))
185 #偽の回答者を正しく「不合格」とする比率（特異度：Specificity）もコンソール上にのみ表
示。
186 print(mean(final_judge[nrow(final_judge), ], na.rm=TRUE))
187
188 #完全ポアソン条件
189 #再現性
190 print(1-mean(recall_full_poisson[1, ], na.rm=TRUE))
191 #特異度
192 print(mean(final_judge_full_poisson[nrow(final_judge_full_poisson), ], na.rm=TRUE))
193
194 #部分ポアソン条件
195 if(check_poisson > 0){
196     #100回中、何回ポアソン分布での判定が行われたか？
197     print(check_poisson / nrow(data_temp))
198     #再現性
199     print(1-mean(recall_part_poisson[1, ], na.rm=TRUE))
200     #特異度
201     print(mean(final_judge_part_poisson[nrow(final_judge_part_poisson), ],
na.rm=TRUE))
202 }

```

#### 付録 4. 手法 C のシミュレーションに使用した R スクリプト

```
1 #Levene の等分散検定用ライブラリ
2 library(lawstat)
3
4 #真の回答者群を data1、偽の回答者プール 100 名を data2 として取り込む。
5 data1 <- read.table("matori_truth.csv", header=TRUE, sep=",")
6 data2 <- read.table("false_100.csv", header=TRUE, sep=",")
7
8 #シミュレーションに必要なパラメータを設定する。
9 outlier_level <- 0.05 #どの程度、平均値から外れている時に「外れ値」とするか。片側
5%なら 0.05、片側 2.5% (両側 5%相当) なら 0.025、片側 1%なら 0.01。
10 robust_euclidean <- 0 #ユークリッド距離の計算時にロバスト推定を行う (1)か行わない
(0)か。
11 robust_outlier <- 0 #最終的な外れ値判定時にロバスト推定を行う (1)か行わない (0)
か。
12 trim_rate <- 0.1 #仮にロバスト推定を行う場合にどの程度刈り込むのか。0.1 ならば上位
下位それぞれ 10%を刈り込むことになる。
13
14 #ユークリッド距離を保存するための空行列を準備しておく。
15 data_euclidean <- matrix(nr=nrow(data1)+1, nc=nrow(data2))
16 final_euclidean <- matrix(nr=nrow(data1)+1, nc=nrow(data2))
17 #除外判定結果を収納するための空行列を準備しておく。
18 final_judge <- matrix(nr=nrow(data1)+1, nc=nrow(data2))
19
20 #混入シミュレーション開始
21 for(i in 1:nrow(data2)){
22 #まず、真の回答者群に i 番目の偽の回答者を 1 名加え、新たに data_temp を作る。
23 data_temp <- rbind(data1, data2[i,])
24 #ユークリッド距離計算時にロバスト推定を行わない場合
25 if(robust_euclidean == 0){
26 #Z 変換する。
27 data_temp <- scale(data_temp)
28 #ユークリッド距離計算時にロバスト推定を行う場合
29 }else if(robust_euclidean == 1){
30 for(j in 1:nrow(data_temp)){
31 for(k in 1:ncol(data_temp)){
32 #欠損値の場合はとりあえず飛ばす。
33 if(is.na(data_temp[j, k])==TRUE){
34 #トリム平均と IQR を用いて Z 変換する。
35 }else{
36 data_temp[j, k] <-
37 (data_temp[j, k] -
38 mean(data_temp[, k], na.rm=TRUE,
39 trim=trim_rate)) /
40 (IQR(data_temp[, k], na.rm=TRUE)/
41 1.349)
42 }
43 }
44 }
45 }
46 }
47 }
48 #偏差平方化する。標準化してあるので、単純に 2 乗すれば良い。
data_temp <- data_temp ^ 2
#na, inf の処理
for (j in 1:nrow(data_temp)){
for (k in 1:ncol(data_temp)){
#標準偏差、もしくはロバスト推定時の IQR がゼロのために
Z 値が nan (非数) もしくは inf(∞) になっているセルが
```

```

1つでもある列は、距離計算には使えないと考えゼロと置
く。
49   if(is.infinite(data_temp[j, k])==TRUE ||
50     is.nan(data_temp[j, k])==TRUE){
51       data_temp[j, k] <- 0
52     }
53   }
54   for (j in 1:nrow(data_temp)){
55     for (k in 1:ncol(data_temp)){
56
57       #nanを除外した後の単純なna欠損値については、同じ設
58       問列内の最高値に差し替える。
59       #なお、else ifで実行すると同じ設問列内がまだ0に置き
60       換わっていないInfとNaNのみで構成されている場合があ
61       るので、独立したif文とする。
62       if(is.na(data_temp[j, k])==TRUE){
63         data_temp[j, k] <- max(data_temp[, k],
64                               na.rm = TRUE)
65       }
66     }
67   }
68 }
69
70 #平方ユークリッド距離の平方根(=ユークリッド距離)にしておく。
71 data_euclidean <- sqrt(data_euclidean)
72
73 #除外判定の開始。
74 #1名ずつ、人数回(真人数+1)のループ
75 for (i in 1:nrow(data_euclidean)){
76   #項目数回、ループ
77   for (j in 1:ncol(data_euclidean)){
78     #ロバスト推定を行わない場合
79     if(robust_outlier == 0){
80       #i番目の人物のj人目混入時のユークリッド距離が、群の
81       平均+XSDの範囲を超えていたら1(外れ値)と判定。
82       if(data_euclidean[i, j] > mean(data_euclidean[, j],
83                                     na.rm=TRUE) + qnorm(outlier_level,
84                                     lower.tail=FALSE)*sd(data_euclidean[, j],
85                                     na.rm=TRUE)){
86         final_judge[i, j] <- 1
87       }
88       #範囲内ならば、0(外れ値ではない)と判定。
89       }else{
90         final_judge[i, j] <- 0
91       }
92     }
93     #ロバスト推定を行う場合
94     }else if(robust_outlier == 1){
95
96       #i番目の人物のj人目混入時のユークリッド距離が、群の
97       トリム平均+X(IQR/1.349)の範囲を超えていたら1(外れ
98       値)と判定。
99       if(data_euclidean[i, j] > mean(data_euclidean[, j],
100                                    na.rm=TRUE, trim=trim_rate) + qnorm(outlier_level,
101                                    lower.tail=FALSE) * IQR(data_euclidean[, j],
102                                    na.rm=TRUE) / 1.349){
103         final_judge[i, j] <- 1

```

```

94                                     #範囲内ならば、0（外れ値ではない）と判定。
95                                     }else{
96                                     final_judge[i, j] <- 0
97                                     }
98                                     }
99     }
100 }
101
102 #真の回答者の標準化ユークリッド距離
103 truth_euclidean = head(rowMeans(data_euclidean), -1)
104 #平均
105 print(mean(truth_euclidean))
106 #標準偏差
107 print(sd(truth_euclidean))
108
109 #偽の回答者 100名の標準化ユークリッド距離
110 false_euclidean = data_euclidean[nrow(data_euclidean),]
111 #平均
112 print(mean(false_euclidean))
113 #標準偏差
114 print(sd(false_euclidean))
115
116 #Leveneの等分散検定
117 TF_euclidean <- append(truth_euclidean, false_euclidean)
118 fx <- factor(rep(c("T", "F"), c(nrow(data1), nrow(data2))))
119 levene.test(y=TF_euclidean, group=fx)
120 #等分散性を仮定したt検定
121 t.test(truth_euclidean, false_euclidean, var.equal=T)
122 #等分散性を仮定しないt検定
123 t.test(truth_euclidean, false_euclidean, var.equal=F)
124
125 #真の回答者を正しく「合格」とする比率（再現性：Recall）
126 recall <- matrix(nr=1, nc=nrow(data2))
127 #計算のため、偽回答者の行を削除した行列を準備。
128 final_judge_temp <- final_judge[-nrow(final_judge),]
129 for (i in 1:nrow(data2)){
130     #i番目の混入者の時の除外判定率
131     recall[1, i] <- mean(final_judge_temp[, i], na.rm=TRUE)
132 }
133 #再現性の平均値（論文中で「再現性」としている値）
134 recall_average = mean(recall[1, ], na.rm=TRUE)
135 print(1-recall_average)
136 #偽の回答者を正しく「不合格」とする比率（特異度：Specificity）
137 specificity = mean(final_judge[nrow(final_judge), ], na.rm=TRUE)
138 print(specificity)

```