

# 統計教育に関する授業内容の傾向と学生の授業評価

## —統計教育質保証に関する分析調査—

小野原 彩香・谷岡 健資・土山 玄・大森 崇

本報告は、同志社大学文化情報学部での統計教育の実態を把握するために行った調査、分析である。本報告は、統計教育における「質」を保証するための推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」の一環として行った調査分析を元にしており、文化情報学部の特徴をより明らかにする方向でまとめなおしたものである。本報告では、大きく2つの観点から統計教育に関する調査・分析を行った。1つ目は、統計教育大学間連携ネットワーク連携校8校のうち6校で行われている統計関連科目のシラバスを収集し、それらを総括するための統一基準を設定した上で、文化情報学部がそれらの大学の中でどのような位置づけになるかを明らかにしたものである。2つ目は、文化情報学部にて開講されているデータサイエンス科目を受講する学生に対し、授業評価アンケートの作成および調査、分析を行ったものである。本報告では特に「統計学への興味」という観点に基づいて単純集計および自由記述の分析を行った。これらの調査、分析により、文化情報学部における統計教育の内容とその受けとり手である学生の関係が明らかとなり、今後の文化情報学部のあり方についての示唆を得ることができた。

### 1. 本報告の経緯

同志社大学文化情報学部（以後、文化情報学部）は、様々な文化事象を「データサイエンス」の手法を中心に解明することを目的とした文理融合型学部である（宿久・村上, 2009）。この「データサイエンス」とは、統計学から発展した学問体系であるため、文化情報学部では、統計学の教育が徹底して行われるようなカリキュラムを編成している（林, 2001；宿久・村上, 2009）。統計学を学ぶ際に重要となるのが、「統計的なものの見方」（「データに基づく課題解決型人材育成に資する統計教育質保証」による用語）である。統計的なものの見方を養う過程で行われる情報の収集の仕方や集めた情報の適切な処理の仕方を学ぶことにより、「新たな課題を自ら発見し、データに基づく数量的な思考による課題解決の能力」の育成が期待される。2012年度から始まった文部科学省 大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」において、文化情報学部は「統計教育大学間連携ネット

ワーク」連携校となり、統計教育に関するいくつかの活動を行っている。当該事業では、課題発見と解決のための1つの重要なスキルである「統計的なものの見方と統計分析の能力」をもった人材を育成するための、標準的なカリキュラムのコンテンツと教授法を整備することを目標のひとつとして掲げている。これを受け、当該事業に設置された2012年度カリキュラム策定委員会の第2ワーキンググループ（ワーキンググループ2、以下、カリキュラム策定委員会WG2）では、現状のカリキュラムコンテンツと教授法の把握を目的として、連携校における統計学の学部授業実態調査を行った。この中で、著者らは、連携校における統計教育シラバスの現状把握を行うことと、同志社大学文化情報学部にて開講されたデータサイエンス科目の授業アンケート調査の実施と分析を行った。

これらの活動は、文化情報学部における教育の特徴を把握するよい機会となったと考えられる。そこで、本報告では、カリキュラム策定委員会WG2で行われた調査分析を、文化情報学部にお

ける統計教育内容の実態把握にクローズアップする形でまとめることにした。このことによって、文化情報学部のデータサイエンス科目における教育の特色の明確化と、現在抱えている課題を整理したい。

## 2. 連携校における統計教育シラバスの現状と文化情報学部の位置づけ

### 2.1 統一シラバス作成までの経緯と目的

カリキュラム策定委員会 WG2 として行った統一シラバスの作成では、各連携校で行われている統計学に関する授業のシラバスを収集し、統一的な基準を設定した上で、それらの基準に沿って、各シラバスの内容を記述し、各大学の統計関連科目を評価する作業を行った。

分析対象は、

<u>青山学院大学</u>	社会情報学部
<u>早稲田大学</u>	政治経済学部
<u>多摩大学</u>	経営情報学部
<u>大阪大学</u>	基礎工学部 共通教育 経済学部 人間科学部 医学部保健学科
<u>東京大学</u>	教養学部教養課程 教養学部専門課程 経済学部 情報理工学部
<u>同志社大学</u>	文化情報学部

(ただし、文化情報学部のデータサイエンス 1～2 は一般教養)

の 6 大学 13 学部合わせて 156 授業のシラバスである。

これらの大学のシラバス収集過程で明らかになったことや、統一項目設定段階で明らかになったことについては、小野原他 (2013a) として報告書の形でまとめた。

この中では、応用に重きを置いている授業と統計理論の習得に重きを置いている授業の対比構造、統計関連科目の開講数自体が大学ごとに大きくばらつくこと、同じ内容を扱っていたとしても、1つのトピックに割ける時間は、授業によって差があること、などをまとめている。

また、作業を進める当初の段階から、文化情報

学部における統計関連科目の授業数が他大学に比べてかなり多いことも報告している。

本報告の最終的な目的は、同志社大学文化情報学部の統計教育に関する特徴を把握することである。これを達成する為にシラバスに基づいて他大学との比較を行った。しかし、シラバスを収集した段階で、単純な比較が困難であることが明らかになった。なぜなら、各大学のシラバスは、フォーマットも書かれている内容にもばらつきがあったためである。このため、各大学のシラバスの内容を網羅的に比較できるような基準を設け、その基準に沿ってシラバスを記述し直した。以後これを統一シラバスと呼ぶことにする。こうして、作成した統一シラバスにより、現在、各連携校で行われている統計教育の全体像を捉える方針をとることとした。

なお、高等教育機関における統計教育には、大学学部での基礎的な教育の他に、大学院での高度な内容を含む教育があるが、ここでは学部での統計教育のみを取り上げている。

### 2.2 統一シラバス作成とその分析方法

統一シラバスの作成では、まず収集したシラバスを通読することによって、大半のシラバスに共通して出現する内容を統一シラバスの項目として設定した。

選定した項目は、「大学名」、「学部」、「授業名」、「授業題目」、「対象年次生」、「担当教員 (記号)」、「配布資料の有無」、「出席点の有無」、「課題・レポート提出の有無」、「試験の有無」、「演習系」、「定量的多変量データ分析」、「定性的多変量データ分析」、「手法」、「応用」、「心理学統計 (感性情報・行動科学含む)」、「社会学統計」、「経済統計」、「人文統計 (テキストマイニング含む)」、「データハンドリング」、「数理統計学の基本 (理論)」、「ベイズ統計」、「情報量・モデル選択」、「ノンパラメトリック」、「時系列分析」、「推定・検定の有無」、「事前履修科目等」、「教科書」、「指定図書」、「参考文献」である。これらの詳細は、小野原他 (2013a) に記載している。

次に、設定した項目に該当する内容を各シラバスから抜き出し、シラバスごとに各項目に該当する内容あり・なしを示す二値の表をまとめた。その後、学部と各項目の関係を把握するためにできあがった表を学部単位で集計し直し、多重対応分析を行った。この方法では、各学部間関係を把

握するのには適しているが、学部内の授業のばらつきを把握することは困難である。そこで、学部単位にまとめる前の各シラバスと各項目の関係を二値で入力したデータも用いて、多重対応分析を行い、同じ学部にも所属する授業のうちもっとも外側に位置する授業同士を凸包として結ぶことにより、学部の特徴を把握することも行った。

多重対応分析とは同一対象群に対して名義尺度によって観測された変数が複数得られた際に、対象間の近さ遠さ、各変量に属するカテゴリ間の近さ遠さ、対象と各カテゴリ間の近さ遠さの関係を低次元空間で同時に表現する視覚化法である（足立他, 2011）。

### 2.3 シラバスの項目分析による連携校と文化情報学部の位置づけ

統一シラバス作成の過程で設定した項目に該当する内容を各シラバスから抜き出し、これを学部ごとに集計したが、その結果のうち、授業数に関するものを図1に示す。図1から、文化情報学部における統計関連科目の授業数が他大学に比べてかなり多いことがわかる。図1中、大阪大学、東京大学での白抜きの棒は学部の合計を表しているが、それと比べても同志社大学文化情報学部の授業数が多いことがわかる。

図2は、統一項目と学部の関係について多重対応分析を行った結果を示している。ここでは、分析の都合から、度数の少ない「ベイズ統計」、「情報量・モデル選択」、「ノンパラメトリック」、「時系列分析」を除き、「人文統計（テキストマイニング含む）」と「社会学統計」の列を統合した行列を用いた。先行研究（小野原他, 2013b）によると、「社会学統計」は、シラバスに社会調査に関する言及がある場合であり、「人文統計（テキストマイニング含む）」は、計量文献学のみを対象とし、テキストマイニングも含む。これらの内容は、手法としては、ほぼ同じ手法を用いることから、「人文統計」と「社会学統計」統合しても良いと判断した。

なお、図の表現の都合から、以下の用語と学部名を矢印以下の通りに略記した。

「演習系」→演習

「定量的多変量データ分析」→定量

「定性的多変量データ分析」→定性

「心理学統計（感性情報・行動科学含む）」→心理

「経済統計」→経済

「人文統計」、「社会学統計」→人文社会

「データハンドリング」→データハン

「数理統計学の基本（理論）」→数理

「推定・検定の有無」→検定

青山学院大学社会情報学部→社会情報

早稲田大学政治経済学部→政治経済

多摩大学経営情報学部→経営情報

大阪大学基礎工学部→基礎工

大阪大学共通教育→共通教育

大阪大学経済学部→経済1

大阪大学人間科学部→人間科学

大阪大学医学部保健学科→医学部保健

東京大学教養学部教養課程→教養学部教養

東京大学教養学部専門課程→教養学部専門

東京大学経済学部→経済2

東京大学情報理工学部→情報理工

同志社大学文化情報学部→文化情報

図2では、第一軸の寄与率が、43.4%、第2軸が25.63%を示している。図中、赤字赤丸が分析対象とした学部を、青字青三角が分析対象とした項目を示す。この図からは、大阪大学人間科学部では、「心理学統計」、大阪大学医学部保健学科では「データハンドリング」、早稲田大学政治経済学部では「経済統計」、大阪大学情報理工学部、東京大学教養学部専門課程、大阪大学共通教育学部、東京大学教養学部教養課程、東京大学経済学部（経済2）、大阪大学経済学部（経済1）、大阪大学基礎工学部では、「数理統計学の基本（理論）」、「手法」、「検定・推定」といった項目に特徴がみられる傾向がある。また、「定量的手法」に重きを置いた授業は、どこの学部においても見られる。

図3は学部内の授業のばらつきを把握するために、各シラバスにおける各項目の該当の有無を二値データとして集計したデータに基づき、多重対応分析を行った結果である。第一軸の寄与率が、26.19%、第2軸が17.53%を示している。図3における凸包の面積が大きければ科目間の距離のばらつきも大きく、小さければばらつきも小さいと言える。

これを見ると、各学部同士の授業は、かなりの部分で共通していることが分かる。その一方で、もっとも授業の内容の散らばりが大きい、すなわち、もっとも多様な授業が行われているのは、文化情報学部であることが分かる。



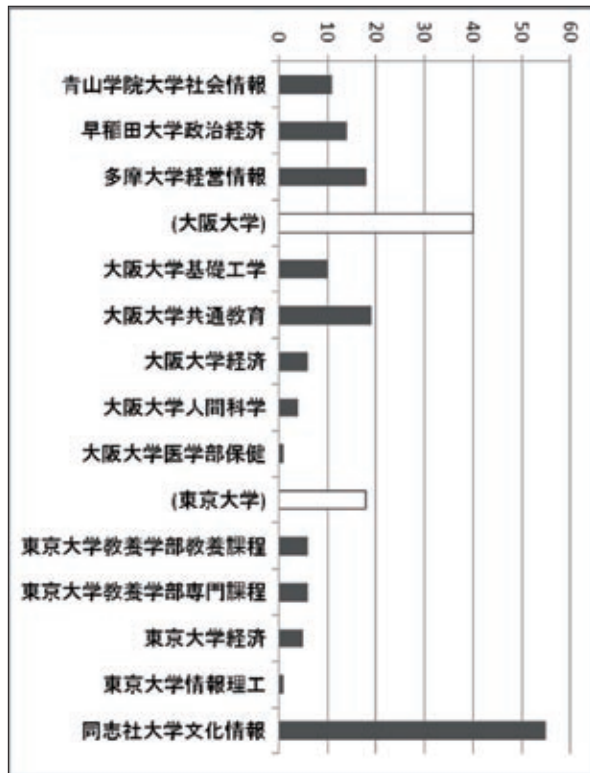


図1 各学部と項目の関係

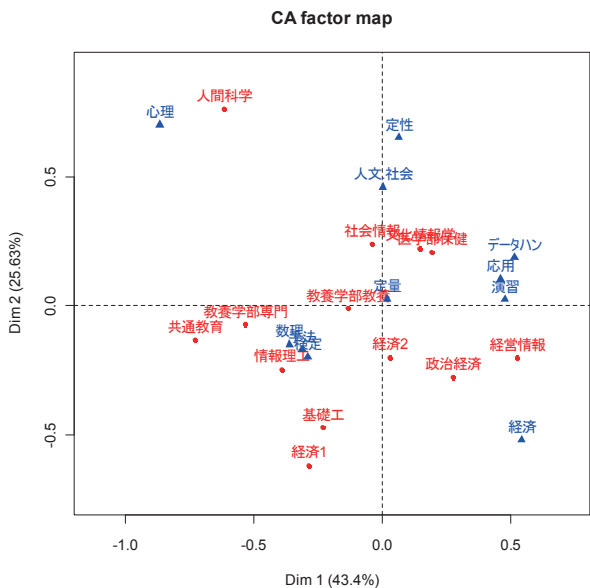


図2 各学部と項目の関係 (多重対応分析)

## 2.4 統一シラバス作成によりわかる同志社大学文化情報学部の統計教育の位置づけ

統一シラバスの項目の分析結果より、各大学、各学部における統計教育は、様々な形態と内容で行われていることが明らかとなった。また、小野原他 (2013a) では、作業の過程において、統計

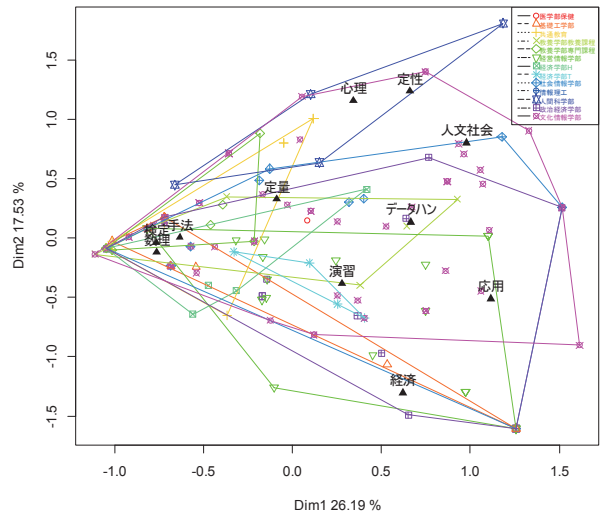


図3 多重対応分析：表1集計前の二値データより作成。経済H：大阪大学経済学部、経済T：東京大学経済学部

関連科目の開講数自体も大学ごとにばらつきがあるだけでなく、応用に重きを置いている授業と統計理論の習得に重きを置いている授業があることもわかってきた。また、同じ内容を扱っていたとしても、1つのトピックに割ける時間は、授業によって差があることも明らかとなった(小野原他, 2013a)。

作業を進める当初の段階から、文化情報学部における統計関連科目の授業数が他大学に比べてかなり多いことが直感的にわかっていたが、実際に授業数について集計したところ、図1はそのことを反映している。連携校のなかでは、学部ごとでも文化情報学部の授業数は圧倒的に多いが、大学ごとに集計しても、文化情報学部の授業数は、調査対象であった大阪大学や東京大学における対象学部同士の合計授業数を大きく上回る。多重対応分析の結果を見ると、図3から文化情報学部は13学部の中でもっとも散らばっていることがわかる。これらの結果からは、文化情報学部の授業が、どの項目にも偏りがなく、多様で体系的な教育がなされていることが明らかとなったといえるであろう。特に、図3における「人文」、「定性」の項目側にいくつかの授業が散らばることは、他学部には見られない特徴であるといえる。これは、文理融合学部としての文化情報学部が提供する授業に起因するものであろう。

以上より文化情報学部での統計教育の特徴をまとめると、統計学を体系的に学ぶことができ、また、授業数が多いため、個々の内容について時間

をかけて学ぶことができることが明らかとなったといえるであろう。しかしながら、こうした特徴を文化情報学部の学生が強く意識しているかどうかは疑問が残る。同志社大学では2013年現在14の学部が存在するが、これまでに学部の紹介などの場でも文化情報学部のキーワードとしてデータサイエンスを紹介することがあっても、統計教育における体系的な学びや授業数の多さが言及されることはほとんどなかった。この検討では、文化情報学部の教育の特徴的な一面を定量化することができたといえよう。この結果は、文化情報学部で学ぶ学生にとって、自らがどのような場で何を学んでいるのかを改めて見つめ直す機会を与えるだけでなく、他大学や他学部のカリキュラムと比較して何が異なっているのかについて自己アピールを行う際にも役に立つと考えられる。

統一項目として、どのような項目を設定すべきであったのかについては、議論の余地がある。項目の設定は集めたシラバスを見て大まかに決めたので主観的な側面があることは否めない。しかし、そうであったとしても上述した文化情報学部の特徴は、他の方法を用いたとしても同様の傾向を示す結果が得られるであろう。

### 3. 同志社大学文化情報学部「データサイエンス科目」のアンケート調査

#### 3.1 データサイエンス科目について

カリキュラム策定委員会WG2の中で、著者らが携わったもう1つの活動は、同志社大学文化情報学部の初年度に実施される科目である「データサイエンス入門」および「データサイエンス基礎」に関するアンケート調査である。調査を行った2012年度には両科目を受講した学生は約300人であった。元々、「データサイエンス入門」および「データサイエンス基礎」は、1人の担当者が大教室で行っていたが、学生の授業内容に対する理解の改善を行うために、現在は3名の担当者が同一の内容を行うこととし、グループワークや教材を導入している(大森, 2013)。このアンケート調査はこうした取り組みに対し、学生側がどのような印象を抱いているのかを把握するために行った試みの1つである。

これについて、「統計学への興味」という観点から、他大学での実施も視野に入れたアンケートの設計、調査、分析を行った。

実施したアンケートの詳細については小野原他(2013b)に譲ることにして、ここでは両授業の特徴であるグループワークと教材に焦点をあてて記すことにする。

#### 3.2 アンケート調査の方法と分析

アンケート調査の概要は以下の通りである。

##### 調査対象

データサイエンス入門、データサイエンス入門演習、データサイエンス基礎、データサイエンス演習を受講している学生303名。

##### 調査方法

集団調査法(演習教室3部屋にて)を用い、調査対象者が演習の時間にPC上で回答する自記式方法。

##### 調査場所

同志社大学京田辺校地 情報メディア館 3教室。

##### 調査日時

2012年度 12月21日 13時15分～終わり次第。

##### 質問項目(計38問)

統計学に対する興味について	2問
調査者個人の性質(付帯質問を含まない)	6問
グループ作業について	5問
4つのオリジナル教材について	8問
配布資料について	3問
先生の説明について	4問
授業で扱ったテーマについて	3問
付帯質問	7問

本報告では、グループワーク、教材とそれらの評価に関係する項目として、以下の項目を取り上げる。なお、各質問項目について5件法、6件法および自由記述により回答を得た。

質問16.「紙ヘリコプター」を使用する体験型学習に興味を持っていましたか。

質問19. あなたのグループは活発に意見交換できましたか。

質問21. グループで作業することで理解が促されたと感じましたか。

質問22. あなたはグループで作業することを楽しめましたか。

質問23. あなたは普段から積極的に話す性格ですか。

質問29. データサイエンスを学ぶことで統計学に

興味を抱きましたか。

質問 31 (自由記述). 本講義で、もっとこうして欲しかったと思うことを 200 字程度で記述してください。(250 字まで)

質問 37 (付帯質問). あなたは高校生のころ何系のクラスでしたか。

#### 理想的なグループの性質に関する分析

グループワークにおいて活発な意見交換が行われることが理想的である。しかしこうした意見交換が盛んに行われても、統計学への興味がわかなければ高い教育効果と言い難いであろう。そこで、理想的なグループとして、質問 29 の統計学に興味を抱いたか否かに関する回答項目、および質問 19 のグループ内で活発に意見交換を行えたか否かに関する回答項目を用いて、理想的なグループの特徴を把握することにした。まず質問 29 と質問 19 の回答を間隔尺度とみなし、各学生のグループ番号を用いて、グループに対する質問 29、19 ごとにそれぞれ平均を算出した。これを各グループの特徴量とした。そして、質問 29 の特徴量の中央値以上、かつ質問 19 の特徴量の中央値以上のグループを理想的なグループとして定義した。

このように定義した理想的なグループが、質問 21 のグループ活動によって理解が促されたか否かに関する質問および、質問 23 の普段から積極的に話すか否かに関する質問に対して、どのように関係するのかを層別散布図を用いて調べた。

#### 自由記述形式の計量的分析

質問 31 の自由記述形式の質問を形態素解析ソフトを用いて計量的に分析した。

まず、形態素解析の精度を向上させるために、得られた自由記述の回答における誤字脱字を修正し、不要な記号は削除した。テキストのクリーニングに次いで、同一概念が異なる語彙で表記されている場合、これを統一した。また、表記にゆれが認められる語彙に関しても同様に処理を行った。すなわち、回答者によって表記がゆれている語彙に関しては表記形式を統一した。

形態素解析の結果、連体詞・接続詞・助詞・助動詞を分析対象から除外し、本報告では名詞・動詞・形容詞・副詞の 4 品詞を対象とした。

語の頻度の単純集計、品詞別に隣接 2 語 (2 語連結) の組を意味する bigram を対象とした単純集計はすでに小野原他 (2013b) に記載している。

ここでは、文理融合学部である文化情報学部の特徴ともいえる質問 37 について理系出身と理系以外出身 (文系、スポーツ系、芸術系、その他) の学生間における特徴語に着目する。すなわち両グループの間の出現頻度の顕著な相違が認められる語について、検討を行った。

特徴語を抽出する際には、unigram と bigram の 2 つのタイプの分析を行った。unigram の分析は、単語 1 語の頻度を対象とした特徴語の抽出を指し、bigram は隣接する 2 語を 1 つの組とし、特徴語の抽出を行った。特徴語の抽出手法としてはランダムフォレスト (random forest) を用いた。

ランダムフォレストは、Breiman (2001) によって提案されたアンサンブル学習、あるいは機械学習と称される手法の 1 つであり、近年さまざまな領域で使用されはじめた手法である。ランダムフォレストは学習データから対象を判別・回帰するためのモデル、つまり分類器を構築し、未知の対象の判別や回帰を行う。

ランダムフォレストでは Breiman et al. (1984) の CART 法 (Classification and Regression Trees) による決定木を多数生成し、これを組み合わせ、1 つの分類器を構築する。ただし、決定木の生成においてはすべての対象を使用せず、ブートストラップ法により対象を選択する。これに加え、変数もランダムに選択し、1 本の決定木を生成する。つまり、ランダムフォレストは CART 法による決定木のアンサンブル学習と言える。本研究では 500 本の決定木を生成し、ランダムフォレストを行った。一般に、決定木は過剰適合 (overfitting) を避けるため、生成し得る最大の決定木に対して、クロスヴァリデーションを用いて木の剪定を行う (杉本・下川・後藤, 2005) が、ランダムフォレストでは未剪定の決定木を生成し組み合わせることで分類精度の向上を図る。なお、ランダムフォレストは過剰適合しないとされる (Breiman, 2001)。また、ランダムフォレストの主な特徴として、以下の点があげられる。(1) 分類精度が非常に高い (2) 大規模データに対しても効果的に動作 (3) 分類における変数の重要度の推定 (4) 欠損値に対して頑健 (5) 分析対象の近接行列の計算。このうち、とくに本研究と関連するのは (3) である。(3) の「分類における変数の重要度の推定」を特徴語の抽出に適用する。推定される変数の重要度は高ければ、より分類に寄与する変数であることを意味する。すなわち、重要度が高い変



数は、2つある外的基準のうち、どちらか一方の外的基準においては、値が大きく、他方においては小さい。本研究においては重要度の高い変数が特徴語であると考えられる。CART法による決定木は、すべての変数の不純度 (impurity) を計算し、不純度の高い変数から対象の分類に使用する。ランダムフォレストにおける変数の重要度の評価基準はこの不純度を用いる。変数の1つにノイズを与え、ノイズによる不純度と正しい不純度の差のすべての決定木による平均を正規化した値が重要度となる。ランダムフォレストをテキストデータに用いた先行研究である「ランダムフォレスト法による文章の書き手の同定」(金・村上, 2007)では、ランダムフォレストの分類精度を求める際に、ランダムフォレストを100回繰り返し、その平均を計算することで、分類精度を算出している。このことから、本研究におけるランダムフォレストの500回の繰り返しは妥当な回数であると考えられる。

### 3.3 文化情報学部「データサイエンス入門」および「データサイエンス基礎」のアンケート集計結果

分析の対象とした授業は、その目標としてアンケートの調査結果について主に「統計学に対する興味について」、「グループワーク」、「実習教材の活用」、「極力PPTには頼らず、配布資料を用いる」を掲げている(大森, 2013)が、それに関連すると思われる質問について単純集計を行った。

#### 統計学に対する興味、グループワーク、教材について

質問 29 データサイエンスを学ぶことで統計学に興味を抱きましたか。

回答項目	非常に興味をもった	興味を持った	興味をもてなかった	まったく興味をもてなかった	未回答
度数	41	177	48	13	24

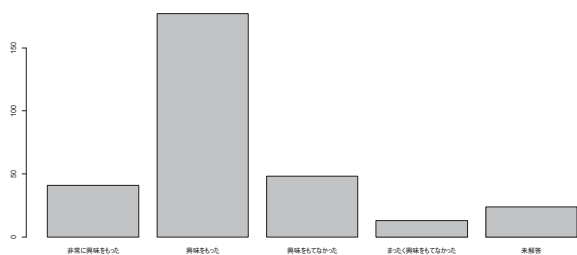


図 4 質問 29 の集計結果

図 4 質問 29 の集計結果より、受講した学生のなかで7割以上がデータサイエンスを学んで統計学に興味をもったと回答したことがわかる。

質問 19 あなたのグループは活発に意見交換できましたか。

回答項目	よくできた	できた	あまりできなかった	まったくできなかった	未回答
度数	63	149	67	1	23

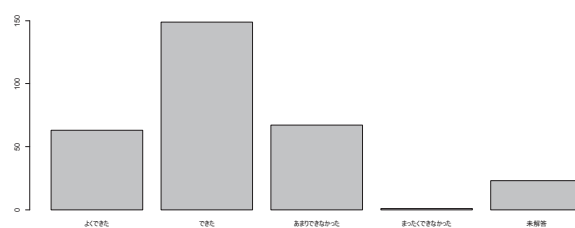


図 5 質問 19 の集計結果

質問 21 グループで作業することで理解が促されたと感じましたか。

回答項目	とても感じた	少し感じた	あまり感じなかった	まったく感じなかった	未回答
度数	95	141	31	12	24

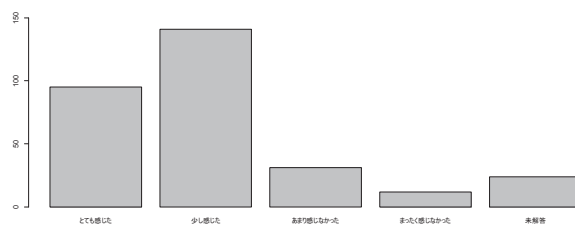


図 6 質問 21 の集計結果

質問 22 あなたはグループで作業する楽しめましたか。

回答項目	非常に楽しかった	楽しかった	つまらなかった	非常につまらなかった	未回答
度数	64	174	36	6	23

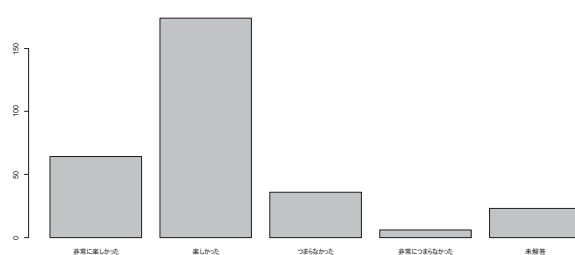


図 7 質問 22 の集計結果

図5から図7の結果から多くの学生がグループで作業することで理解が促されたと感じるとともに、グループ作業に満足していることがわかる。

質問16 「紙ヘリコプター」を使用する体験型学習に興味を持ってましたか。

回答項目	非常に興味をもった	興味をもった	興味をもてなかった	まったく興味をもてなかった	未回答	覚えていない
度数	48	167	54	11	23	0

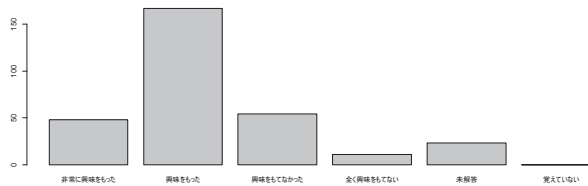


図8 質問16の集計結果

図8より約7割の学生が2群の検定問題を理解するための実習教材である「紙ヘリコプター」を使用した体験型学習に興味をもったと回答している。

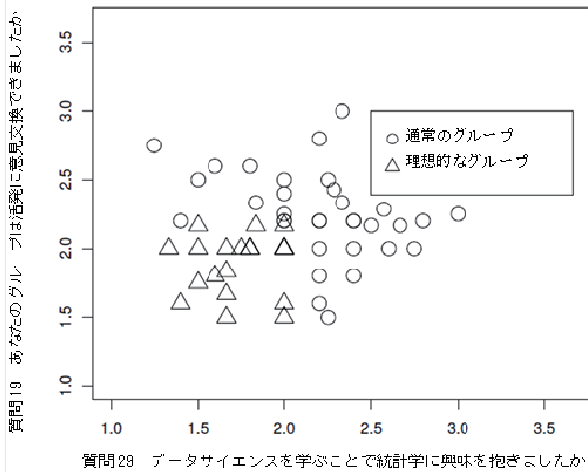


図9 理想的なグループの定義

理想的なグループの性質について

図9の横軸は質問29を表しており、縦軸は質問19を表している。値が小さいほどそれぞれ、興味をもっている、よくできた、という傾向にある。三角印に対応する箇所は理想的なグループということになる。このようにして決まった理想的なグループで、質問21と質問23に対する散布図を層別したものが図10である。縦軸横軸ともに値が小さいほどそれぞれ非常に感じた、積極的

という傾向を示している。

上記の結果より、グループ学習を行う際、グループ内で意見交換が活発に行われ、かつ統計学に興味を持った学生はグループ活動によって理解が促されたと感じる傾向にあることがわかる。また、普段から積極的に話す性格が多いグループだからといってグループ内で意見交換が活発に行われ、かつ統計学に興味を持つ傾向にある優良グループ

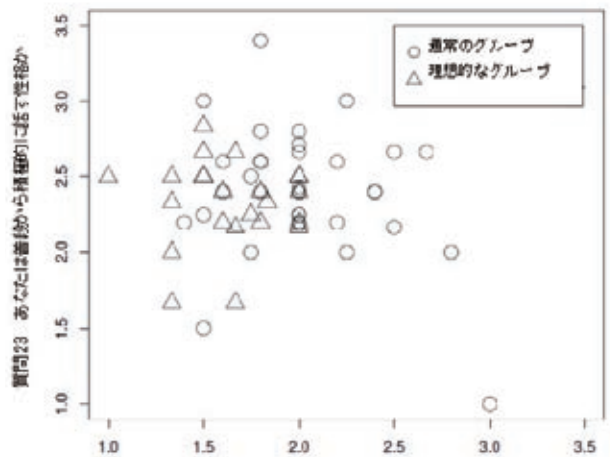


図10 理想的なグループと問21、問23との関係

になるとは限らない。

自由記述について

名詞の unigram についての特徴語抽出の結果、図11に示すように「データサイエンス」という語が理系出身の学生にもっとも特徴的に用いられていることが認められた。

「データサイエンス」は理系出身学生の自由記述において、「データサイエンスに興味を持つことができたかもしれない」や「データサイエンスを学んで、将来どうにかせるのか」、「データサイエンスを学ぶ意義について」などと記述されており、科目としてのデータサイエンス基礎やデータサイエンス演習などに対してではなく、データサイエンスそのものに興味を抱いていると言える記述が多い。一方、理系以外出身の学生は「データサイエンスの講義でもっと班の組み換えをしてほしかった」や「クラスによってあまりにも進度のズレや説明の分かりやすさに差があったのでこの状況が続くのはこれからデータサイエンスを学ぶ学生にとって全く良くありません」というように、班分けやクラス分けに対する記述が多いと言える。

「データサイエンス」という語以外では、図



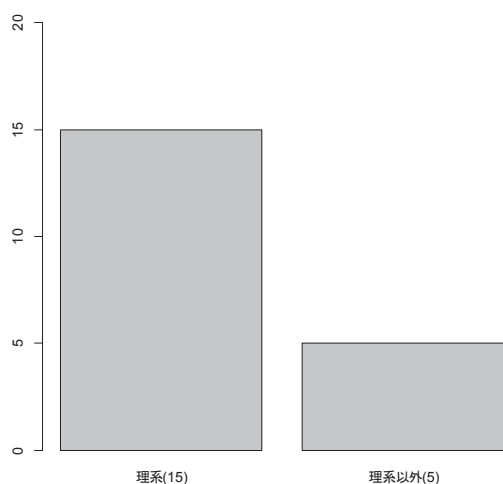


図 11 「データサイエンス」の出現頻度

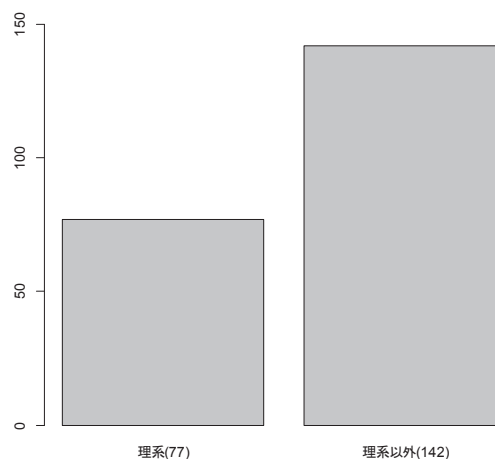


図 13 「説明」の出現頻度

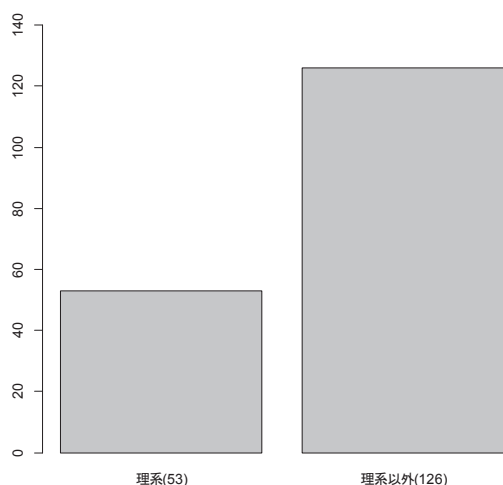


図 12 「理解」の出現頻度

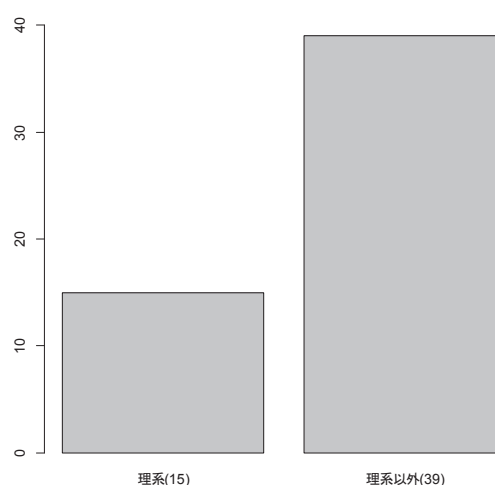


図 14 「理解」と「できる」という bigram の出現頻度

12、図 13 に示すように「理解」や「説明」という語の出現頻度が特徴的に理系以外出身の学生に多く現れる。これに関しては bigram による分析において考察する。

「理解」という語については、名詞と動詞のみを対象とし bigram を求め、ランダムフォレストを行った。分析の結果、図 14 に示すように名詞の「理解」と動詞の「できる」という bigram の出現頻度が理系以外出身の学生に特徴的に多く用いられていることが分かった。これらのうち、「理解できない」という意味をもつ回答が 30 件あり、その内訳は理系出身学生が 5 件、理系以外出身の学生が 25 件となる。ここで、理系以外出身の学生の自由記述に目を向けると「ほんやりとしか理解できなかったりして」や「ほとんど理解できていない人ほど質問しにくい雰囲気だった」などと

いう回答が認められる。

次いで、「説明」という語に関して、名詞と形容詞の bigram を求めランダムフォレストを行ったところ、図 15 に示すように「説明」と「ほしい」という bigram が特徴的に理系以外出身の学生に多く用いられていることが分かった。自由記述の回答では、「もう少し時間を取って詳しく説明してほしかった」、「もう少し細かく説明してほしいです」、「もう少しゆっくり説明してほしかったです」というような回答が理系以外出身の学生に多く認められる。

このようなことから、理系出身の学生と理系以外出身の学生の間では、データサイエンス科目の座学の授業および演習において、授業内容の理解、あるいは理解したという実感に差があると考えられ、それゆえに授業科目としてのデータサイエン

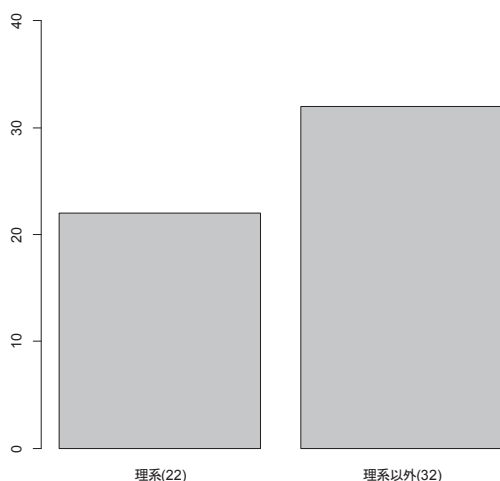


図 15 「説明」と「ほしい」という bigram の出現頻度

スではなく、学問としてのデータサイエンスに対する興味に影響が現れていることが推測される。

### 3.4 アンケート集計結果について

シラバスの記載により、他と比べた文化情報学部の教育の特徴は把握できるが、シラバスからは教育内容と学生の理解度や満足度を把握することができない。カリキュラム策定委員会 WG2 の活動では、文化情報学部が提供する統計科目の一部である「データサイエンス入門」および「データサイエンス基礎」という科目に関してアンケート調査を実施することができた。これらの科目については、学生側のやる気と理解を引き出すために、数年前からグループワークや教材の導入を行ってきた。

アンケートを通してわかったことは、授業以前は、「データサイエンス」がどのようなものであるかを知らなかった学生が 7 割程度いたにもかかわらず、授業を通して統計学に興味を持った学生の割合は 7 割を越え、全体の約 6 割 5 分の学生が、さらに発展した内容を取り扱った授業を受講したいと回答していた。また、将来的にも授業で学習したことが役に立つと回答した学生は、全体の 8 割にも及んだ（小野原他, 2013b）。比較対象がないため、大小関係を論ずることはできないが、これらの数字は少なくはない数字であると言えるのではないだろうか。

グループワークの効果については、約 7 割の学生がグループワークで積極的に意見交換できたと回答しており、グループワークにより、理解が促されたと回答している。また、グループワーク

を楽しめたという学生は約 8 割に及んだ。また、グループで活発に意見交換ができ、授業を通して統計学に興味を持った学生は、グループワークによって理解が促されたと感じる傾向にあることがわかった。このことは、グループワークを活性化させることで、統計学という学問への感心を引き出すことができる可能性を示しているといえよう。より質の高い統計教育を行うためには、学生の学習への意欲は最も考慮しなくてはならない要因の一つであることは間違いない。その要因としてグループワークの活性化が考えられるのであれば、グループワークを導入した統計教育では、統計学の知識やデータのおもしろさということとともにグループワークにおける受講者の充実をいかに行うかについて、教員が積極的に取り組みアイデアを試していく努力が必要であろう。

また、アンケート中にあった自由記述の回答からは、難易度より進度に関する要望が多くを占めていた（小野原他, 2013b）。本報告では、自由記述における理系出身とそれ以外出身の回答の差に注目した分析を行ったが、理系出身の学生が、「データサイエンス」という学問そのものに関心を抱いているのに対し、理系出身以外の出身の学生は、班分けや、クラス分けに対する意見が多く、理解や説明に対する不満が見られる結果となった。このことから、両者の間で、授業内容の理解、あるいは理解したという実感に差があると考えられる。

文理融合学部として、背景の異なる学生の多様性は文化情報学部の 1 つの魅力といえる。一方で、その多様性に起因して、何をどこまで前提とすればよいかという教育上の問題がしばしば話題になることがある。こうした観点からは、統計教育における理系出身とそれ以外の違いは、教育上の困難さの要因となっているかもしれない。自由記述の分析からは、理系出身とそれ以外出身で要望の質が異なる可能性が示唆されているといえよう。グループワーク型教育の 1 つの利点は互いに教え合うことにある。もしも学生間で授業内容を理解する上で大きく質の異なる層があるのであれば、これらの異なる層をグループ内にバランス良く配置し、互いに教え合うようなグループワークのやり方を構築できるのであれば、教育効果は上がる可能性がある。現在のグループワークのグループは学生の背景を考慮せずにグループを構成しているため、グループをどのように構成するかは今後

の課題となるであろう。

アンケート調査全体の課題としては、上記で述べた各要因が学生の統計学への興味に与える影響の検証、加えて学生の統計に関する知識も試験結果のデータを用いて複合的に検証する必要がある。また、それらの妥当性を比較・検証するためにも複数の授業に対してアンケート調査を行う必要があるかもしれない。

#### 4. 総括

文化情報学部における統計教育の実態把握という観点から、シラバスの特徴や特徴的な授業の内容、それに対する学生の反応を詳細に述べてきた。一連の調査分析を通じて、文化情報学部の統計教育は、他大学より多様かつ網羅性に優れ、内容的な部分については、テキストマイニングをはじめとした人文統計や定性的手法に重点のある授業が特徴的であることが明らかとなった。

また、グループワークや教育教材を導入した教育の取り組みに対し、学生側の反応も「統計学への興味」という観点において、良い結果を得られたということができ、さらに深い内容を学ぶ意欲を見せている点の特徴的であるといえる。しかしながら、グループワークを活性化させるさらなる取り組みが必要であろう。理系出身とそれ以外の出身の学生の授業内容に対する理解には、差が見られる結果が得られているため、学生の背景を考慮したグループ構成などが今後の課題であると思われる。

#### 文献

- 足立浩平, 村上隆 (2011). 『非計量多変量解析法 主成分分析から対応分析へ』, 東京, 朝倉書店
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Breiman, L., Freidman, J. H., Olshen, R. A., Stone, C. J. (1984). *Classification and Regression Trees*, Wadsworth International Group: Belmont, California, 43-49.
- 林知己夫 (2001). 『データの科学』, シリーズ〈データの科学〉1, 東京: 朝倉書店.
- 金明哲, 村上征勝 (2007). ランダムフォレスト法による文章の書き手の同定, *統計数理*, 55(2), 255-268.
- 大森崇 (2013). 4.1 同志社大学文化情報学部におけるデータサイエンス科目の初年次教育での取り組

み, 第4章 統計教育のフロンティア—連携校における特色ある授業—, 大学間連携共同推進事業連携大学(編)カリキュラム策定委員会 報告書 第2部 連携大学における統計学の学部授業実態調査, 文部科学省 平成24年度大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」統計教育大学間連携ネットワーク, 93-102.

小野原彩香, 谷岡健資, 土山玄, 大森崇 (2013a). 第1章 連携校学部シラバスの全容, 大学間連携共同推進事業連携大学(編)カリキュラム策定委員会 報告書 第2部 連携大学における統計学の学部授業実態調査, 文部科学省 平成24年度大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」統計教育大学間連携ネットワーク, 1-30.

小野原彩香, 谷岡健資, 土山玄, 大森崇 (2013b). 第2章 同志社大学文化情報学部「データサイエンス科目」のアンケート調査, 大学間連携共同推進事業連携大学(編)カリキュラム策定委員会 報告書 第2部 連携大学における統計学の学部授業実態調査, 文部科学省 平成24年度大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」統計教育大学間連携ネットワーク, 31-58.

宿久洋, 村上征勝 (2009). 文理融合型の統計教育～データサイエンスによる文化事象の解明～, 2009年度統計関連学会連合大会講演報告集, 統計関連学会.

杉本知之, 下川敏雄, 後藤昌司 (2005). 樹木構造接近法と最近の発展, *計算機統計学*, No.18, pp.123-164.

#### 参考 URL

- 文部科学省 大学間連携共同教育推進事業「データに基づく課題解決型人材育成に資する統計教育質保証」 <http://www.jinse.jp/>.