

統計教育16年の総括

成蹊大学 経済学部

新村秀一

1. はじめに

- 統計などの理数系教育に対する考え
 - 専門家が必要とする機能を全て備えた, 使いやすいソフトウェアがあれば, それで高度な理数系のユーザー教育が行える.
 - 統計が先行し, 数理計画法, 数学, 各種設計,
 - 数理計画法などのOR分野は, 教員の理解が不足
 - 専門家教育と高度なユーザー教育(問題解決学)を, 分離すべき
 - 統計のユーザー教育は, 多くの統計家の努力で1970年代後半に始まり, 他の学問に先行した
 - しかし, 近年他の情報教育における競合科目が出てきて, 今後苦戦することが懸念される

表1 個人年表

年代	事項
1971-1974	卒業後, 大阪府立成人病センターで心電図の自動診断システムの開発. 心電図所見の多群判別を研究. 分散共分散行列による, FisherのLDFとQDFへの疑念で, 一生の研究.
1977(昭和52年)	SASを導入し, 統計の個人家庭教師
1984(昭和59年)	数理計画法ソフトLINDOの代理店
1985(昭和60年)	SASミニコン版代理店(製薬企業32社)
1996(平成8年)	成蹊大学経済学部教授
2010 (平成22年)	判別分析の全ての問題を解決し, 最小誤分類数が得られる最適線形判別関数の完成.
2012 (平成24年)	関数の最大/最少を扱う数理計画法による問題解決学の普及を最後の目標

統計の恩師(統計ソフト)

私にとって, 専門書で得られない成果が統計ソフトから得られた.
自分の勉強を兼ね, 日本にSASの紹介普及が出発点.

- SAS(1976年以降: CUI, 汎用機, システム開発)
 - J.Sall(新村訳, 1986). SASによる回帰分析の実践. 朝倉書店.
 - 新村秀一(1989). 易しく実践データ解析の進め方. 共立出版.
- SPSS(1993年以降: GUI, PC, Windows環境)
 - 新村秀一(1994). SPSS for Windows入門. 丸善.
 - 新村秀一(1995). パソコンによるデータ解析. 講談社.
- Statistica(2001年以降: グラフを重視したPC用統計ソフト)
 - 山口&渡辺監修のマニュアルが参考になる.
- JMP(2003年以降)
 - 新村秀一(2004). JMP活用統計学とおき勉強法. 講談社.
 - 新村秀一(2007a). JMPによる統計レポート作成法. 丸善.

恩師から学んだこと

- 分からないことは、小さなデータで分析し、手計算で確認.
- 壮麗な統計学を、現実問題に役に立つかたないかで優先順位を付け、体系的に考える癖がついた.
 - 例えば、推測統計学は、平均、標準偏差、歪み度、尖り度、比率、相関係数、独立性の検定、回帰係数などごく限られていることを認識し.
 - 標準誤差の意味と95%信頼区間は平均を例にして教えるだけでよいと達観.
 - また、p値の意味は2*2の分割表のFisherの直接確率を教えることで、理解が深まる.
- 配置行列の作成法を知れば、重回帰分析、分散分析、数量化理論、判別分析などを個別に勉強する必要がなく、負担が軽減(森村・牧野他, 1984). ダミー変数の理解も容易になる.
- 掃き出し演算子で、変数選択法の仕組みが理解できる(Sall,1986).
- 説明変数の全ての回帰モデルを出力することで、変数選択法の暗闇から解放された.

2. ユーザー教育16年の実施報告

- 1996年から2011年までの16年間の総括
- 2.0 立教と東洋大学の非常勤講師(1995年)
- 2.1 SASによる教育(1996年－2001年)
- 2.2 SPSS(2002年－2006年)
- 2.3 JMP(2007年－2011年)
- 2.4 LINGO(2012年以降)
 - 2012年度はJMPの契約を継続しないので、学部の2科目と大学院の2科目をやめ,
 - 数理計画法ソフトのLINGOで、主としてDEA法(効率性分析)で重回帰と判別手法の違いを教える.

授業の内容

表1と表2の重回帰と分散分析しか教えていない。判別分析が一生の研究テーマであるが、判別分析に問題があるとの認識。

表1 データを調べる。

	連続尺度。	名義尺度, 順序尺度。
1変数。	基本統計量とヒストグラム。	度数表(クロス集計)。
2変数。	散布図, 相関, 単回帰。	分割表(クロス集計)。
3変数以上。	クラスター分析, 主成分分析。	多重クロス集計。

表2 予測手法。

		目的変数。	
		連続尺度。	名義尺度, 順序尺度。
説明変数。	連続尺度。	重回帰。	判別分析, ロジスティック回帰。
	名義尺度, 順序尺度。	分散分析。	FUNCAT。

2.0 立教と東洋大学の非常勤講師

- 東洋大学経済学部の1年生の必修
 - SASが導入されていて、利用されていない
 - 500人以上の学生に、『易しく実践データ解析の進め方(共立出版)』で講義。入門科目の位置づけを考えていなかった。
 - SASの宿題を出し、情報センターが大混乱し、断念。
- 立教の山口ゼミ3年生(20名以上)で、『パソコンによるデータ解析(講談社)』を輪講。
 - シカゴのファーストイリノイ銀行の労使紛争の裁判データ。単純な統計分析で、性差別、人種差別があり倒産。
 - しかし、上手く職種で層別し、単回帰分析で昇給を分析すれば？

2.1 SASによる教育(成蹊大学)

- 2年次配当の情報科学Ⅱを前期・後期に開講
 - 定員は, PC室の制限で当初は60人(その後80人).
 - この時代は情報教育の黎明期で, 他に競合ソフトによる開講科目がなかったので, ほぼ定員を満たす
 - 定員制限のため受講に漏れて次年度に受講する者もいて3年生と4年生も若干名いた。(実は, 厳しいということ回避)
- テキスト:『統計処理エッセンシャル』.
 - 前半は, 高森教授の統計に関する解説,
 - 後半は「学生の成績データ(40人, 7変数)」を用いて, SASの操作法と出力結果の解釈を小説のように解説.

評価法と問題点

- 評価は, 授業で習ったことを参考に, 自分でデータを決めて, 20頁以上の統計レポートの提出.
- 授業の問題点.
 - この時代は多くの学生が自宅でPCを利用していなく, 入力時間に格差.
 - SAS言語(新村, 1994b)で, DATAを作成するDATAステップと, 統計処理のPROCステップのコマンドを次の構文で入力する必要.


```
DATA SASデータセット名;
  INPUT (X1-X19) (10*5. 3) ;
PROC MEANS; VAR X1-X10;
```
 - 「データが好きでも, データを入力するのでDATAですよ」とダジャレをいっても「DATE データセット名;」と入力しパニックになる学生もいて, サポートのため教室を駆け回って運動不足が解消できた.
 - 課題提出に際しては, この時代インターネット上に分析に耐えるデータがなかったため, 図書館に「DATAハンドブック」等を購入し紹介し, 手入力で作成させた.
 - 試験前にレポート作成が集中し, PC室の利用がネック.
 - 優秀な学生からは, 「他の授業の倍以上の学習時間を必要とするが, 終われば達成感で満足しています」などというコメントも得られた(統計実習の最良の時期).

(1) 高杉君:「日本・死者急増」

- 彼は、データを以下のように各種調査票から収集し、最終的に次の重回帰式を導く。

$$\bullet \text{ zikosuu} = 14.531054 + 0.012590 * \text{menkyo} + e$$

データ収集の出典の一例(彼のレポートからの抜粋):

- 運輸省自動車交通局技術安全部管理課編 平成6年(1994年)9月度車種別自動車保有車両数月報
- 財団法人自動車検査登録協力会編 平成6年(1994年)3月末現在市区町村別自動車保有車両数, 輸入車保有車両数

• (社)全国軽自動車協会連合会編市区町村軽自動車車両数による。

車種別自動車保有台数は、道路運送法、道路運送車両法により、運輸省が毎月、各都道府県の陸運事務所からの報告をもとにまとめている登録自動車台数の1994年9月30日現在における数値を掲げたもので現在の自動車普及台数とみてよい。

- 都道府県別保有台数は、乗用車(普通車, 小型車, 軽四輪車), トラック(普通車, 小型四・三輪車, 被けん引車, 軽四・三輪車), バス(普通車, 小型車), 特殊用途車両(消防車, 救急車, タンク自動車などの普通車, 小型四・三輪車, 大型特殊車)及び二輪車で構成される。この数値には駐留軍人, 軍属の私有車, 外国人の私有車などを含むが, 防衛庁関係の車両は含まれていない。

(2) 田中君:「プロ野球選手の成績についての分析」

- 彼は、私の授業が厳しいという風聞のため、4年生になって単位を全て取った上で受講してきた学生である。
- 余裕があるのか、20頁以上の統計レポートとPower Pointの発表資料も提出。
- また、自宅での勉強のため、SAS/ GRAPHが未契約のために『パソコン楽々統計学』に添付の評価版のStatiscaでグラフを作成するという使い分け。
- 参考文献に『易しく実践データ解析の進め方』を挙げているので、この時期はテキストをグレードアップしたようだ。
- そして、次の重回帰式を導いた。

$$\text{年俵} = -5405.51 + 3888.93 * \text{リーグ} + 77.11 * \text{打点} + 165.64 * \text{四球}$$

- その後、大学院に進学し、新聞社の懸賞論文の特選を受賞している。

データ: 2001年度シーズンのセリーグとパ・リーグの選手67人を対象とし、変数は選手氏名, チーム名, リーグ, 推定年俵, 打率, 試合, 打数, 得点, 安打, 2塁打, 3塁打, 本塁打, 打点, 盗塁, 四球, 死球, 三振の17変数である。推定年俵だけはホームページにはなく、筆者が授業で新村先生のご指摘を受け、別の資料[文献4]をもとに入力した(単位は万円)。

2.2 SPSS(～2006年ごろ)

- テキストは、「学生の成績データ」をSPSSで分析した『SPSS for Windows入門(丸善)』.
- SASのCUIから, SPSSのGUIに代わって学生の負担が軽減.
- それ以上に, 総理府統計局などのHP上に質の高いExcelデータが開示されているので, 学生の負担は圧倒的に少なくなった.
- この時期, 筆者のHP(<http://sun.econ.seikei.ac.jp/~shinmura/>)に、2名のレポートを掲載.
 - どのようなレポートを作成するか参考に供した.
- 学生にとって, 負担の少ない最良の時期

(3) 芦川さん:「日本映画産業についての分析」

- 授業ではめだたない学生のため, 25頁の秀逸なレポートの提出には驚いた.
- 時系列データは経済環境による年代区分が明確であることが, その後の他の学生の提出レポートで分かった.

(4) 前田さん:「日本の住宅事情に関するレポート」

- この学生は, 入学時のプレテスト(Excelを習得)に合格し, 2年次配当の「情報科学Ⅱ」にとび級で入ってきた1年生である.
- 2-3回目の授業後, 顔面蒼白な顔で「先生の話す日本語の意味が全く分かりませんので, ドロップアウトしたい」といつてきた.
- 聞くと1年生なので, もう少し辛抱すれば理解できるようになるし, 授業後分からないことを質問してきなさいと言って返した(授業後, 研究室に毎回質問).
- まったく統計の知識がないにもかかわらず25頁のレポートを提出した.
- 最近の学生は, 日本の都道府県を正しく把握していないものも多い.
 - そこで分かり易いことと, 47件のケース数で, 変数は10個以上が容易に手に入り, 変数の意味が理解しやすいので総理府統計局の都道府県データの利用を学生に進めている.
 - 面白さはないが, 授業内容を十分に理解したものである.
- **統計教育に用いるデータ: ケース数は印刷しA4に収まる50件程度, 変数は誰にも容易に内容が理解でき, 興味のある10変数以内.**

(5)高橋君:「プロ野球選手の推定年俵についての統計分析」

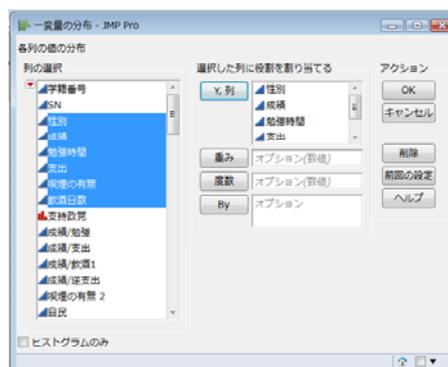
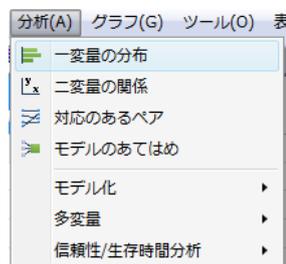
- 彼は、筆者の3年のゼミ生である.
- 過去の秀逸なレポートの中から田中君と同じテーマでレポートを書きたいと言ってきたので
- データを自分で集め直すこと, また最低2年以上のデータを集め, 初年度の重回帰式で次年度以降の予測と検定を行うことでの条件で認めた. 最終的に35頁のレポートを提出してきた.

2.3 JMP(2007-2011)

- 2004年4月から1年間ウイーンの郊外にある国連の研究機関のIIASAに在外留学.
 - 広大な公園内にあるハプスブルグ家の離宮がまるごと研究機関.
 - 日本人研究者も多く, 赴任2日目から日本語環境でPCが使えた.
- 出発前の3月にSASジャパン(株)のJMP事業部の責任者と打ち合わせ.
 - かねてより解説書の執筆を依頼されていたが, メニューとマニュアルの日本語化が前提.
 - ようやくそれが完成したので, ウイーンで執筆.
 - 評価版を本に添付することを了承してもらった.
 - **学生や社会人が自宅で自習できる.**
 - JMPの開発責任者のJohn Sall副社長の解説書(Sall et al.,2004)の校正と監修も引き受けた.
 - 帰り際に, 大学で普及する相談を受けた. 「大学の統計教育の一番の問題は学生が自宅で利用できないのが問題である. 」と回答.
- 2005年3月に帰国すると, 次のユニークな契約が可能
 - 大学や学部で契約すれば, **教員や職員や学生が自宅でも利用できる.**
 - 利用者が多ければ, 一人あたりの単価は非常に安くなる.
 - ようやく, 2007年にソフト購入の予算に余裕ができたので導入し授業に用いる.

JMPのユニークな点

- 操作法の革新性.
 - 従来の統計ソフトを授業に用いる場合、必要な統計手法と統計量のオプションをきめ細かく教える必要.
- 図1のJMPのメニューから分かるように、手法がわずか7個のカテゴリー(プラットフォーム)に整理.
- そして図2の画面で変数を選び、[Y列]をクリックし[OK]すれば、1変量のグラフから統計量すべてが出力.
- 授業では重要なものだけを説明し、不用な出力は無視することを宣言し、統計ソフトの操作法にさく時間が節約できる.
- 学生にとって、最高の教育環境が実現。しかし、ゆとり世代と、競合の情報科目の出現で、人気が薄.
- **多分、今後統計実習の受難時代になる。**



- 学生にとって、統計ソフトの操作法や分析データが容易に手に入るようになったが、以前のように秀逸なレポートが出なくなった.
 - ゆとり世代にとって、負担の大きな授業を避ける傾向と、
 - 他のPCを使った競合授業が増えてきた？

(7)黒岩:「合否判定によるFisherの判別分析の終焉」

- 黒岩さんは、本年度の3年のゼミ生である.
- 筆者は、1997年から2010年2月まででMNM(Minimum Number of Misclassifications)基準による最適線形判別関数の研究を完成した.
- 大学の統計教育で判別分析を取り上げる機会がないので、『最適線形判別関数』をテキストに輪読し、後期は私が行っている研究論文を参考にレポートを作成することをテーマとした.
- その中で黒岩さんのレポートの内容が一番早く提出されたので、今回成果報告させようと考えたが、就活に専念したいということで幻の計画に終わった.
- テキストは学部3年生には難しい内容であった.

(6) 坂本君:「2007年日本の優良企業分析」

- 坂本君は2008年に卒業した筆者の3年のゼミ生。
- 平均給与の高い優良企業に就職したいと考え、日経優良企業100社と自分でインターネットで調べて非優良企業30社を選定し、財務データなど23変数を集めた。
- このデータを用い彼のレポートの2章に見るように3つの作業仮説を検証した。

1 はじめに

2007年を振り返ってみると、思えばいろいろなニュースがあった。参議院選の自民党大敗から安倍総理の辞任、食品偽造問題、そしてアメリカ発で日本の株式市場を賑わせたサブプライムローン問題。ただ私の目下最大の関心事は就職活動だった。金利が上がろうが下がろうが、株式市場が低迷しようが、牛肉のほずが実は豚肉だろうが世の中で何が起ころうとも私は2009年には大学を卒業しなければならないわけで、従って就職活動をしなければならない。できるならば最初のキャリアは大企業でその中でも特に優良企業と呼ばれる会社に就職したいと思うのは多くの就職活動を向かえた大学生にとって共通の願いだろう。もちろん私もほかの就職活動を向かえた大学生と同じように考えていた。しかし、優良企業とはいったいどんな企業のことだろうか？SONYやTOYOTAのような誰もが名前を知っている企業だけが優良企業だろうか、もちろんこれらの企業は優良企業だろう。しかし、多くの人が名前を知らなくても堅実に利益を上げている優良企業もなかにはあるだろう。優良企業の定義は難しい。優良企業とは何かという問題に対する一つの答えとして日本経済新聞が発表している優良企業ランキングがある。実際よく出来たランキングであるが、結果だけを発表しているので細部でどのような評価が行われているのかよく分からない点もある。私は企業研究もかねて、日経新聞が発表している優良企業ランキングのリストをもとに、その財務上のデータを収集してそれを分析することによって優良企業の特徴を統計学的に把握したいと思う。

2 レポートの目的と作業仮説

レポートの目的は、優良企業の特徴を優良企業でない企業と比較することによって明らかにすることである。作業仮説は、①優良企業はそうでない企業よりも財務的に優れている。②従業員一人当たりの利益が大きい企業は平均給与が高い。③利益が高い企業は株価が高い、の3つを検証する。

(8) 井上さん:「油田規模分布の統計分析」

- 経済経営学研究科の院生は、GPAが低くなることを避けるのか、筆者のような分析系の授業を学部で取らない傾向。
- これまで学部と似た科目を開講してきたが、2007年入学の阿部君以外見るべきレポートはなかった。彼は成蹊高校から立命館大学に行き、大学院はUターン組。大学では、1年次にExcel教育で簡単な統計を習っただけということで、授業で教わったことを31頁のレポートにまとめた。
- 井上さんは、2010年に社会人入学で大学院生になり、昨年度は金融データを用いた分析を行った。今年は、彼自身の仕事である油田関連のデータを駆使した38頁の秀逸でレベルの高いレポートを提出した。
 - 私自身、グラフはデフォルト主義なので自分でマニュアルを調べてグラフを作成。
 - クラスタ分析ではWard法の説明は学生には難しいので、一般的には「群平均法」を使うことを指導している。
 - 彼はWard法を適用し、クラスタ分析で地域別に5個のクラスタ分けに成功している。

各クラスタを構成する主要な産油国は次のとおりである。

クラスタ1(赤): 中東巨大産油国(サウジアラビア、イラン、イラク、クウェイトなど)およびベネズエラ

クラスタ2(緑): ロシア(1国だけ)

クラスタ3(青): 伝統的大産油国(米国(アラスカ+沖合)、メキシコ、英国、ノルウェー、中国、インドネシア、アルジェリアなど)

クラスタ4(黄土): 新興産油国(カンボジア、マレーシア、ベトナム、スーダン、ウガンダ、シエラレオネ、ガボンなど)

クラスタ5(青緑): 中小産油国(ミャンマー、イエメン、タイ、ニュージーランド、チリ、日本、台湾など)

3. 大学の統計教育の問題点と反省

- 統計等の理数系教育を体系的に行うことは、ここ数年難しくなっている。
- SASの時代：CUI, データを自分で入力。筆者にとって至福の時期。
- SPSSの時代：GUI, データはHPからダウンロード。学生にとって至福の時。
- JMPの時代：
 - SPSSのGUIの利便性に加え、
 - 操作が簡単で、統計量の説明を中心にでき、
 - 自宅で利用可能。
 - 最良の環境を整えたのに、受講生が減少し、まともなレポートが出なくなった。

4. まとめ

- 1996年から2011年まで行ってきた、16年間の統計ソフトを使ったユーザー教育の総括を行った。
- 今後は、1年生の「統計入門」の講義科目で統計ユーザーの裾野を広げることに専念。
- 坂本君の「日経優良企業100社と不振企業30社」、田中君と高橋君の「野球選手の年俸」、岡野いつ佳さんの「日本車44車種の価格、販売台数と性能」等のデータを用い、重回帰分析の欠点をDEA法で補完。
 - 重回帰は、多入力多出力が扱えない、
 - 個々の評価対象に最適な評価が行えない、
 - 誰を手本に改善すれば良いかわからない。
 - DEAは、個々の評価対象に最適な重みを与え、出力/入力で効率性を定義。

参考文献

- Flury, B. & Rieduyll, H. (1988). Multivariate statistics : A Practical Approach, Cambridge University Press.
- Sall, J.P., Creighton, L. and Lehman, A. (2004). JMPを用いた統計およびデータ分析入門 (第3版), SAS Institute Japan 株. [新村秀一監修].
- Shinmura, S. (2011). Beyond Fisher's Linear Discriminant Analysis -New World of Discriminant Analysis -, ISI2011 Proceedings, 1-6.
- Shinmura, S. and Ishioka, T. (2012). Serious problems of Discriminant Analysis, 第4回日独分類シンポジウム・アブストラクト(発表予定).
- 森村・牧野他編(分担執筆, 1984). 統計・OR活用事典. 東京書籍.
- J.Sall (新村訳, 1986). SASによる回帰分析の実践. 朝倉書店.
- 高森寛・新村秀一 (1987). 統計処理エッセンシャル. 丸善.
- 新村秀一 (1989). 易しく実践データ解析の進め方. 共立出版.
- 新村秀一 (1994). SPSS for Windows入門. 丸善.
- 新村秀一 (1994). SAS言語入門. 丸善.
- 新村秀一 (1995). パソコンによるデータ解析. 講談社.
- 新村秀一 (1999). パソコンらくらく数学. 講談社.
- 新村秀一 (2004). JMP活用統計学とっておき勉強法. 講談社.
- 新村秀一 (2007a). JMPによる統計レポート作成法. 丸善.
- 新村秀一 (2007b). ExcelとLINGOで学ぶ数理計画法. 丸善.
- 新村秀一 (2007c). JMPによる統計レポート作成法. 丸善.
- 新村秀一 (2010). 最適線形判別関数, 日科技連出版社.
- 新村秀一 (2011a). 合否判定データによる判別分析の問題点, 応用統計学, 40/3, 157-172.
- 新村秀一 (2011b). 数理計画法による問題解決法, 日科技連出版社.
- 新村秀一 (2011c). 問題解決学としての統計入門. 第7回統計教育の方法論ワークショップ-問題解決力育成を旨とした統計教育の方法論-, 1-10.