

# 問題解決学としての統計入門

新村秀一  
成蹊大学 経済学部

## 1. 私の信念(人と異なるかも?)

- 理数系の学問は、「**問題解決の実学**」であるべき.
- 理数系の学問は、「**専門家教育**」から、高度な問題解決能力を涵養する「**ユーザー教育**」を独立させるべき.
- **統計, 数理計画法, 数学等の学問**は、数学が共通言語であり、その研究成果は**ソフトウェアに凝縮**されている.
- 学生から専門家までが**使いやすく**、専門家が**要求する機能を備えたソフトウェア**を使えば、効率的なユーザー教育が可能である.

## 2. ユーザー教育の光と影1

- 統計はかつて専門家教育一色であり, 文科系では一部の分かり易い入門部分のつまみ食いであった.
  - 社会に出て, 実際に役立てた人は少ない.
- 統計ソフトを用いて実際のデータを分析する「データ解析」, すなわち「統計のユーザー教育」が今日広く定着した.
- これによって, 統計の専門家と, それをささえる裾野の広い実務家が育った.

## 2. ユーザー教育の光と影2

- 一方, OR(経営科学)をみると, 過去3000人以上の学会員がいたが, 現在は2000人を切っている.
- 日本全国の大学には, 統計と異なり, 経営学部, 数理工学科, 経営工学科などの学部学科がある.
- 米国のInforms(代表的なOR学会の一つ)は, 41セッションが5日間開催され, 7割が社会人の発表(シカゴ大学 L. Schrage教授の調査)である.
- 日米のこの違いはなぜか?

### 3. 私の特異な経歴

- 昭和46年(1971):住商情報システムに1期生で入社
- 昭和46年から4年間, 大阪成人病センターで心電図の自動解析プロジェクトに参加. (32個の心電図所見を正常と判別が研究の原点)
- 昭和52年(1977):SASを計算センターに導入. 日本最強の統計開発
- 昭和59年(1984):LINDO社(シカゴ大学ビジネススクール)代理店
- 昭和60年(1985):SASミニコン版代理店(32製薬メーカ, 東洋信託)
- 昭和61年(1986):AHP代理店
- 平成2年(1990):Speakeasy代理店(Mathematica, Matlabの源流)
- 平成8年(1996):成蹊大学経済学部教授
- 平成15年(2003):ウイーンのIIASAで在外研究.『JMP活用 統計学とおき勉強法(講談社,JMP付き)』,『JMPを用いた統計およびデータ分析入門(Sallら著, 監修校正)』
- 平成22年(2010):統計入門(1年次必修)担当
- 平成23年(2011):経営科学(2年次準必修)担当

### 4. 人と違った考え

- 書籍文化とソフト文化の違い
  - 統計の専門書を勉強しても, 達成する目標が不明確.
  - ソフトウエアは, 出力結果が, 正しく理解できれば良い.
- ソフトの効用
  - 統計学全体をシステムの, 体系的に理解できる.
    - 推測統計学の恩恵を受けるのは, 平均, 標準偏差, 歪み度, 尖り度, 比率, 相関係数, 独立性の検定, 回帰分析, 等
    - 標準誤差や95%信頼区間の解釈が分かればよい
  - 判別分析の問題点
    - 推測統計学と無縁,
    - 回帰分析と異なり, 現実のデータを無視した理論
    - 種々の問題点
    - 『最適線形判別関数(日科技連, 2010)』ですべて解決
    - これまで, 12冊の統計書を出してきたが判別分析の紹介は上記の本が初めて(問題点を解決できなかったため)

## 5. ユーザー教育に必要なもの

- 新しいテキスト
  - 一つのデータをストーリー性をもって分析し小説のように解説
  - 統計手法を体系的に重要なものを全て教える
- 教育に適したデータ
  - 用いる変数の意味が分かり易い
  - 入門は40-50件(学生データ:元青山学院政経学部 高森寛)
  - Fisherのアイリスデータ, Anscombe, SPSSの銀行データ, 万博データ, Swiss銀行紙幣データ
- 無償の評価版の活用
  - 書籍に添付(Statistica, JMP, Speakeasy, LINGO, What'sBest!)
  - JMPは大学で契約すれば, 学生や教職員が自宅でも利用可能な契約形態に貢献.
- 統計入門で新しい教育法の確立
  - 2011年からは, 経営科学で新しいOR教育(問題解決学)
  - 非常勤講師, 集中講義引き受けます

## 6. ユーザー教育の試み

- 社会人向け(企業人の時代)
  - 東電, 日銀, 不動産経済研究所, 東ガス等の社内講師
- 非常勤講師(成蹊大学の教員になる1年前)
  - 東洋大学1年生の入門科目(渡辺先生の手の上で)
    - 「優しく実践データ解析の進め方(共立, 1989)」: 入門科目に不適
    - 大講堂で1年生全員に講義
    - SASの宿題で, 情報センター大混乱
  - 立教大学3年生のゼミ(山口ゼミ, 25人, 社会学部の代表)
    - 「パソコンによるデータ解析(性・人種差別の裁判)」の輪読
    - 学生に相互評価させ, 評価に利用. ただし学生にフィードバックしていないことが問題.
- 成蹊大学
  - 情報科学 II (15年間SAS, SPSS, JMPと変遷)
    - データを手で入力から, HP(総理府統計局)からダウンロード.
    - CUIからGUIへ. そして2005年以降学生は自宅でJMP利用可
    - しかし, 負担が少ないのに, ここ5年は外部公表できるレポートがない
  - 2010年度より統計入門(1年必修, 128人の受講生)
    - 工夫が必要

## 7. 統計入門の工夫

- テキスト『JMP活用 統計学とおき勉強法(講談社, 2004)』
  - 僅か4件のデータで, 基本統計量, 相関と単回帰, **分割表と独立性の検定(直接確率でp値が分かる)**, 一元配置の分散分析を計算させる. **検定も教えることが重要.**
  - 「学生データ」をJMPで分析し, その統計量を読み解く
- 入門科目では, 手計算の後で, Anscombeのデータを用い, 1)基礎統計量と2)相関・回帰をExcelで理解しているか否かを宿題で確認. 3)Speakeasyで直接確率の計算

## 8. 統計入門の拾いもの

- R.Fisher(1936)が, 「Fisherの仮説」に基づき, LDFを提案し, 判別分析という応用上重要な統計手法が提案された.
  - その後, 2次判別関数, ロジスティック回帰等豊饒な判別手法が開発.
  - 数理計画法でも研究が行われ, SVMに収斂.
  - しかし, 判別分析の数々の問題点を解決していない.
- MNM基準で**最適線形判別関数**を整数計画法で定式化し,
  - 判別分析の全ての問題点を解決し,
  - しかも判別成績は既存の手法より格段に良かった.
  - 統計的判別手法は, 線形分離可能なデータを識別できず, より多くの説明変数のモデルを選ぶ(Swiss銀行紙幣データ).
- 実は, 統計入門で10択100問のマークセンス試験を中間と期末で行ったが, 合否判定は良質な線形分離可能なデータである.
  - 統計的判別関数で, 種々の問題が分かった.
  - 100問中7問で合否判定可能.
  - センター試験などの試験の難易度の比較に利用できるのでは?

## 9. ユーザー教育への専門家の無理解

- 「素人が統計ソフトを使うと誤用する」という警句。
  - 単なる批判は、評論家の言であり、教育者の意見として不適切
- **グラフ**で、p値の解釈の逆転などを防止できる。
- **レポートを書かせ、間違いを添削し、返却する。**
  - 学部1年生の例、学部3年生の例(新村秀一のHP)
  - 2010年度大学院生の例(分割表を連続尺度に適用)
- **介護保険の誤用**
  - 数億円かけた1分間タイムSturdyのデータを決定木分析の後、システム化
  - 開発者土肥徳秀医師の後悔(順序尺度を名義尺度で分析、母集団の違い、開原しげこと)

## 10. OR教育の再編(数理計画法による問題解決学)

Excelで定義した**31個の種々の問題を、LINGOの汎用モデルで分析**し結果をExcelに出力。学生が社会に

出ても実問題を分析できる**問題解決学**の提案。

- 無償の評価版を授業に用い、学生は自宅で利用し学習できる。
- 日科技連出版社、6月刊行予定

- 1 魔法の学問による問題解決学(単体法を教えるのではなく、1次から3次の多項式の最大/最小問題から始め**領域の最大最小問題**がLPのアルゴリズムそのものであることを紹介)
- 2 非線形モデルと大域的最適解(微分との違い、目標計画法)
- 3 組み立て産業への応用(減少費用、双対価格)
- 4 配合計画(ある製鉄会社の配合問題、解がない場合の対応)
- 5 評価の科学DEA(良いところをほめる新しい評価法として普及させたい)
- 6 マーコウィッツのポートフォリオ分析と効率的フロンティア
- 7 時間をうまく管理する人生の達人(PERT、専用のPERTソフトは高額である)
8. 回帰分析(**最小自乗法**、**LAV回帰**、**L1.5ノルム**)と判別分析(**SVMと最適線形判別関数**)を40件のデータで分析。統計モデルに関する広い視野を与える。
9. 巡回セールスマン問題
10. 種々のモデル(**教独**、**CT(コンピュータ断層撮影)**を連立方程式としてとらえる、輸送問題、要員計画、ナップザック問題、マルコフ過程、ローン計算、在庫管理、品質管理、待ち行列、**Black Scholesのオプション価格**の計算など)
11. Material Requirements Planning, サンプル・サイズの決定、予測問題、季節調整

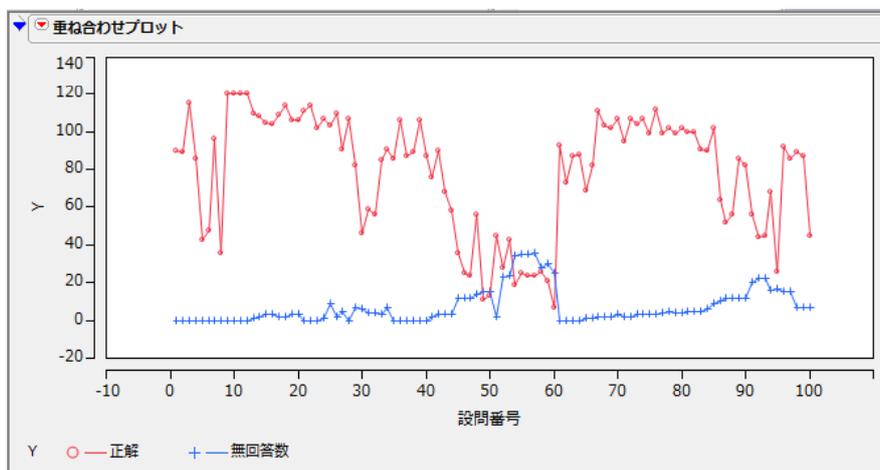
## 11. 終わりに

- 理数系の学問は、お茶や生け花の免許皆伝を授けることであってはいけない。
- 学力の低下が叫ばれているが、日本に体力のある今のうちに、
  - ソフトウェアでもって劣った能力を嵩上げし、
  - 正しく問題解決できる「**21世紀の一般教育**」を確立すべき。
- 理数系の学問は、それを正しく応用できれば著しい時間の節約になる。浮いた時間は、それ以外の複雑な問題解決に挑戦すべき。

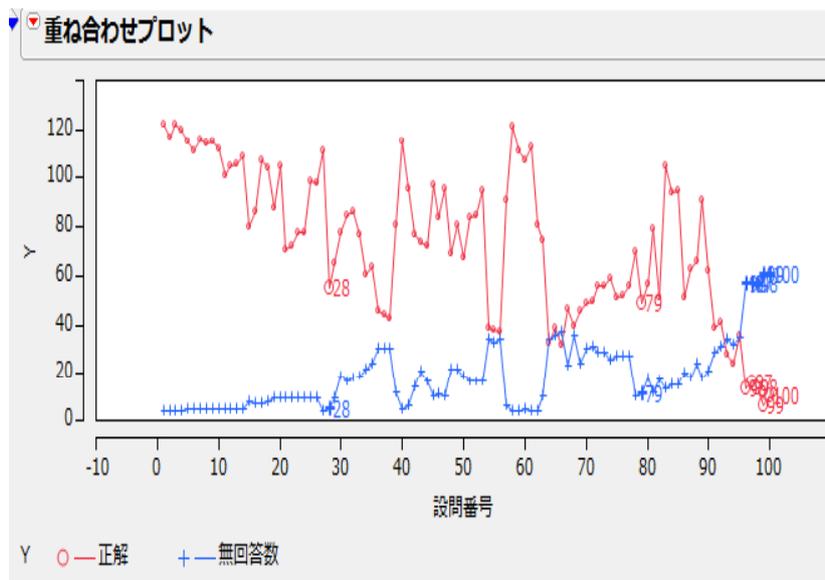
## 12. 配布資料の紹介

- 統計教育
  - 「マークシート試験によるFDの一提案」: 学生にこの論文の内容を公開して好評。統計が自分たちの試験の分析に利用できるという、統計の有用性を学生に示せた。
  - 「**試験の合否判定データの最適線形判別関数による分析**」: 統計入門の合否判定に関する論文は、今後数編出版予定。
  - 山口ゼミの報告: 「ゼミ学生による総合評価データを用いた解析手順の標準化」: 学生にフィードバックしなかったことが問題
  - **統計ソフトを使うと誤用**: 土肥徳秀, 筒井孝子(1995). 提供されたケアからみた要介護高齢者のタイプ分け. **第9回日本計算機統計学会シンポジウム**, 161-170.
- 「数学のできる問題解決学」: Speakeasyでローンの計算, GoogleのPageRank, Fisherの直接確率等。
- データマイニング・シンポジウム論文集(2000)。

### 図1 中間試験の正解と無回答数の比較



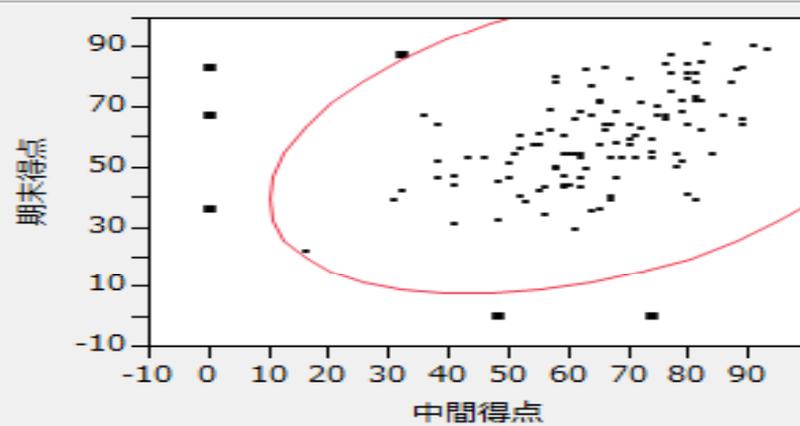
### 図2 期末試験の正解と無回答数の比較



## 図6 中間得点と期末の得点の散布図

(中間が30点で、期末が90点の学生をどう考えますか?)

### 中間得点と期末得点の二変量の関係



二変量正規楕円  $P=0.990$