

統計ソフトを活用した統計的データ分析の実習

数理工学科 高田佳和

1. プロジェクト実施の背景と目的

当学科の教育目的は、数学を応用し工学に現れる様々な問題を解決できる数理技術者を養成するところにあります。問題解決で大事なことは、実験、調査によって得られるデータを如何に解析し重要な情報を得るかである。そのための有用な方法が統計的データ解析法であります。当学科では、2年次前期に「確率統計」、3年次前期に「応用統計」の講義で、理論にもとづく解析手法の説明を行っている。それらの教育効果を上げるために、3年次後期の「計算数学第二」の講義では、学科の計算機室でパソコンを用いた統計的データ分析の実習を行っている。現在、学科の計算機室の各パソコンには、今回のものづくり経費で導入した統計解析専用ソフト、JUSE-StatWorks（日科技研）と JMP（SAS）がインストールされている。更に、数値解析のために Mathematica、表計算ソフトエクセルがインストールされている。統計解析専用ソフトを活用すれば大抵の統計的データ解析を行うことができる。しかし、学生が社会にでたとき、これらのソフトが自由に使える環境にはないと考えられる。そこで、この授業では多くのパソコンに標準的にインストールされているエクセルを用いて統計的データ分析の実習を行った。

又、大学院では「多変量データ解析特論」の講義を行っている。この科目は現在、全専攻共通科目であるため、例年、色々な専攻の学生が受講している。講義主体であるが、講義の理解を深めるため、今回は学科の計算機室で JUSE-StatWorks を用いた実習も行った。

2つの講義「計算数学第二」、「多変量データ解析特論」の中で、統計ソフトをいかに活用したかについて報告します。

2. 実施概要

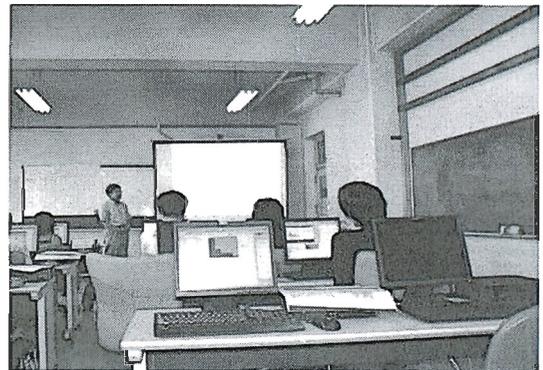
2-1 計算数学第二

品質管理は、企業や組織において、製品またはサービスの品質改善や問題を解決するために非常に重要です。この授業の目的は、品質管理において現れる様々な問題に対して統計的手法を駆逐して行う統計的品質管理を学生に身につけさせることにおいています。具体的には次の4つのテーマについて主にエクセルを用いて実習を行った（図1）。

- QC七つ道具

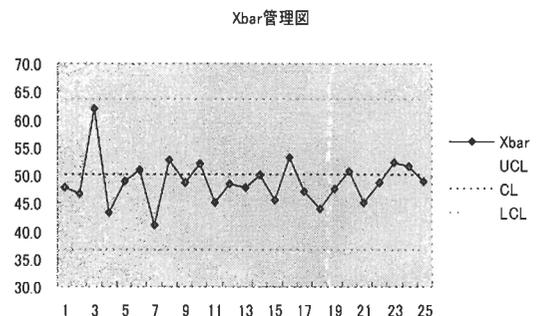
- 計量値、計数値の検定・推定
- 独立性の検定
- 多標本問題

図1 実習風景



QC七つ道具では、特に、問題発見に重要な管理図（図2）、ヒストグラム、パレート図の作成、データが正規分布に従っているかどうかを調べるための正規確率プロットの作成を行った。

図2 管理図（エクセルで作成）



計量値（正規分布）、計数値（二項分布、ポアソン分布）の検定・推定に関しては、エクセルにある統計分析のための関数（TTEST、FTEST 等）、統計分析ツールを利用した。これらを用いればかなり高度な統計分析が可能である。

独立性の検定では、ピボットテーブルを用いた分割表（クロス集計表）の作成と関数 CHITEST を用いた独立性の検定を行った。

最後に複数個の母集団を扱う多標本問題では、計量

値の場合、分析ツールを利用すれば簡単に分散分析表を作成することができる。計数値の場合も関数を利用すれば分析ができる。

講義の進め方は、資料をパワーポイントで作成し、それを学生に配布し、統計分析方法を説明した後、エクセルを用いて実習を行った。その資料の例、問題は、実際に統計的品質管理で現れる題材をもとにして作成した。

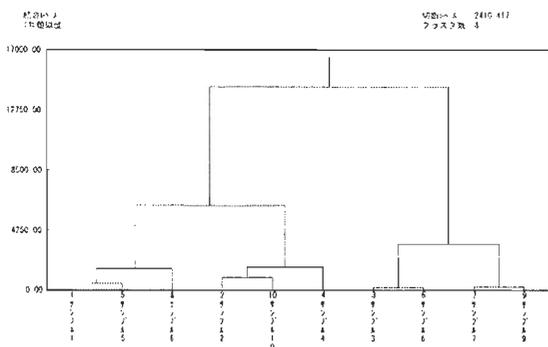
2-2 多変量データ解析特論

いくつかの個体が複数個の変数によって特徴づけられているとき、それらの変数間の相互関係を分析するための統計的手法の総称を多変量データ解析といわれる。その手法の中で特に重要な次の5つの分析方法について講義を行った。

- 重回帰分析
- 判別分析
- 主成分分析
- 数量化3類（対応分析）
- クラスタ分析

例えば、クラスタ分析は、異質なものの混ざり合っている個体を、それらの間の類似度にもとづいて似たもの同志を集めていくつかのクラスタ（集落）に分類する方法である（図3）。

図3 クラスタ分析（JUSE-StatWorksで作成）

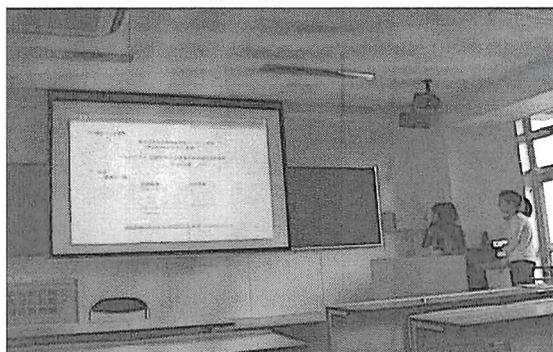


講義では各手法の概略と理論的背景について説明し、講義の終わりにその手法の使い方（JUSE-StatWorks）のデモを行った。講義は全てパワーポイントで行い、その内容は、学生の復習のためWebCTに載せた。最後の方の講義では、講義の復習を兼ねて、学科の計算機室で統計ソフトJUSE-StatWorksを用いて、これらの手法の実習を行った。

最終講義では、例年受講生をいくつかのグループに分け、これらの手法の一つを実データに適用し、その結果のプレゼンを行っている。昨年度の受講生（21名）は、情報電気電子専攻と建築学専攻の学生であった。所属する研究室ごとに7つのグループに分け、主

に研究室で扱っているデータに対して上記の手法の一つを適用し、その結果のプレゼンを行った（図4）。

図4 発表風景



各グループのプレゼンテーションのテーマは下記の通りであった。

- 2008年プロ野球セ・リーグの出場試合数上位選手の能力の分析（主成分分析）
- フォルマント周波数を用いた通常音声/障害者音声判別（判別分析）
- 出生率の要因分析（回帰分析）
- 車種判別（判別分析）
- ビブラートをかけたエレクトリックギター音の個人間の好みの差について（クラスタ分析）
- 熊本市中心市街来訪者の回遊行動特性に関する考察（回帰分析）
- 大学施設整備目標策定のための教員の重視する施設整備項目の整理について（クラスタ分析）

3. 今後の課題

統計的データ解析を単なる講義だけで、その使い方を含めて理解させることは困難である。この発表では、統計解析ソフトによる実習を講義と併用することで、その教育効果を上げることができた。

計算数学第二では、学生が社会に出たときの有用性を考え、エクセルを用いた統計的データ解析法の実習を行った。しかし、より高度なグラフを描いたり、少し複雑な解析を行うにはエクセルでは手間隙が掛かり不便である。統計分析専用ソフトとの併用について検討する必要がある。

多変量データ解析特論は、講義中心となり、実習の時間があまりとれなかった。しかし、学生は統計分析方法のユーザーであるので、理論より使い方に重点を置いた内容にした方が社会に出たとき役に立つと考えられる。その点から実習と講義のバランスをいかにとるかが今後の課題である。