

# 検査・調査 Web システムにおける統計解析機能の導入と活用

情報社会システム研究室 I O 7 V M 0 3

工 藤 伸 之

## 1 はじめに

近年、インターネット普及率が向上し、それにより Web 上で多様なサービスの提供がなされるようになってきた。Web 上で行われる調査や検査のサービスもこれに含まれ、容易に調査や検査が行えることから、それらの数は増加の一途をたどってきた。一方では、調査・検査は Web 上で数多く行われているものの、さまざまな問題点もはらんでいる。

たとえば、調査においては、仮にデータの収集をしたとしても、そのデータからどのようなことが読み取れるのかを知る必要があり、また、検査においては、収集されたデータに基づく詳細な分析結果を提示する機能をもつことが望まれるが、そのような検査・調査に対する拡張機能が用いられる環境はほとんど見られない点である。

上記のような問題点に対し、収集したデータを基にして即座に分析結果として出力し、たとえ分析に関しての専門的知識をもたない人に対しても、統計手法を選択するだけで分析結果が即座に表示され、また、検査においても回答者に対して詳細な分析結果を提示する機能を搭載する。これができるれば、回答者・調査者双方に利点のある検査・調査 Web システムが実現できることになる。このようなシステムを提供することで、今よりも検査や調査がしやすい環境が整うことになり、Web 上で検査・調査が行われるケースが更に増えると考えられる。

そこで、検査・調査 Web システム上に統計解析システムを導入することで、検査や調査によって収集されたデータを調査者が容易に分析できるようにし、その分析結果を回答者にも与えることで、検査における判断材料とさせることを本研究の目的とする。

## 2 開発言語・Web アプリケーション

本システムを作成するにあたって使用する開発言語および Web アプリケーションは、以下のものである。

- (a) PHP...主にデータの受け渡し、調査票作成に用いる。
- (b) XML...調査票として受け渡されるデータの格納を行う。
- (c) MySQL...Web 上でデータ管理を行う。

## 3 統計手法と統計エンジン

### 3.1 統計手法

本研究では以下の 6 つの統計手法を実現した（小野寺・菱村，2005；渡辺，2005）。

#### (1) 基本統計量の算出

得られたデータの任意の変数の平均値や中央値などの代表値や、分散や標準偏差などの散布度を求める。

#### (2) ヒストグラムの作成

各データの項目ごとにヒストグラムを描く。

### （3）単回帰分析

データの中の一つの説明変数からもう一方の目的変数の値を予測する。

### （4）重回帰分析

データの中の複数の説明変数から一つの目的変数の値を予測する。

### （5）主成分分析

得られたデータの空間的配置を基にして、元の項目数よりも少ない次元でデータ間の関連性を分析する。

### （6）クラスター分析

空間内に配置されたデータの距離や類似性、非類似性を基にして、各データをいくつかのクラスターにまとめていく。

## 3.2 統計エンジン

上記の手法を検査・調査 Web システム上で実装するときに、データを収集する際の仕組みや分析する際の挙動については、それを制御するシステムを作成する必要がある。本研究では、統計解析ソフトウェアに「R」を利用して、その解析システムを作成する。

R は統計計算とグラフィックスのための言語・環境であり、多様な統計手法とグラフィックを提供しているので、広範な拡張が可能な統計解析環境である。

## 4 検査・調査 Web システムにおける統計解析機能

### 4.1 概要

本研究では、検査・調査 Web システム上で動く統計解析ツールの開発を行う。すなわち、調査者にとっては、作成された検査や調査を実施することでデータを収集し、蓄積されたデータを分析することで分析結果を得るもので、回答者にとっては、提供された分析結果および付加情報を得るものである。

### 4.2 システムの特徴

#### （1）データ収集

検査・調査 Web システムでは、調査者の作成した調査票を基にして、回答者が検査や調査を受けることができる。このとき、検査において、回答者に対して診断を与えるといった事例はこれまでもあるものの、Web 上の検査によってデータを収集する事例はほとんどない（包，2008）。本システムにおいて、検査においてもデータ収集を行うことで、より多くのデータ収集ができるようになっている。

#### （2）詳細な統計解析

データ収集によって蓄積された数多くのデータを調査者が有効活用できるようにするため、詳細な統計解析が可能となっている。また、詳細な統計解析機能はオンライン上で使用できるようになっているため、場所を選ばず、いつでもどこにいても分析結果を得ることができる。詳細な統計解析を行う際には、調査者は統計解析の用法を知る必要はない。それは、開発者によって事前に統計解析に関する処理が定められているからであり、調査者は実行したい分析を選択するだけで、収集されたデータを基にした分析結果を出すことができる。更には、汎用性の高いシステムを分析ごとに採用してあるため、調査者は、PHP や MySQL などの Web アプリケーションのソースに気を配らなくてもよい。

### (3) 回答者への情報提供

回答者に親切な診断結果を表示することが可能である。その結果に応じたアドバイスを提供したり、さまざまなインタラクティブ性をもつ部品を提供したりすることができる。これらを用いて、回答者はより価値のある情報を得ることができる。また、統計グラフなどの分析機能を利用して回答者は付加情報を得ることもできる。

### 4.3 機能

本システムがもつ機能は、次の5つである。

- (1) データ収集...回答データを調査者が利用するために、データベースに蓄積する。
- (2) データの一覧表示...調査者にデータの並び替えが行われたデータの一覧を表示する。
- (3) データの編集...データの書き換えができる。
- (4) ファイル出力...収集したデータをさまざまな形式に変換してファイル保存できる。
- (5) データ分析...収集したデータから特徴や傾向などを明らかにすることができる。

### 4.4 実際の動作

検査・調査 Web システム全体の構造は図 1 のとおりである。

回答者がさまざまな検査・調査に回答することで、回答データベースにデータが蓄積されていく。そのデータを<調査者>が確認するには、検査・調査 Web システムの[調査者入口]から調査者向けページに移動し、専用ページにログインする必要がある。正しい ID とパスワードを入力するとログインできる。この専用ページを使うことで調査者はデータの確認等を行うことができる。調査者専用トップページには、調査者が作成した検査・調査の一覧が表示される。

ここでは例として肥満診断の検査を行い、そのデータを収集する事例を示す。フレーム中のリンクをクリックすることで、[データの表示]や[データの編集]、[ファイル出力]などの処理ができる。また、収集したデータを基にした分析も、リンクをクリックするだけで即座に実行できる。それはさまざまな手法に対応しており、基本統計量の算出、ヒストグラム、単回帰分析（図 2）、重回帰分析、主成分分析（図 3）、クラスター分析（図 4）による分析が可能である。

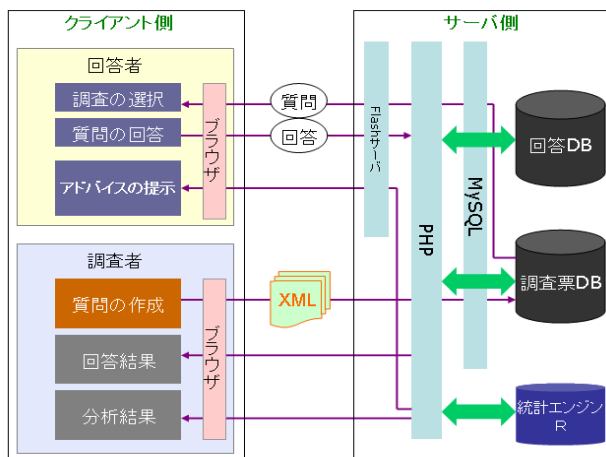


図 1：検査・調査 Web システムの構造

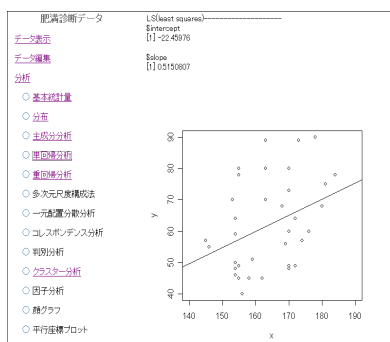


図 2：単回帰分析

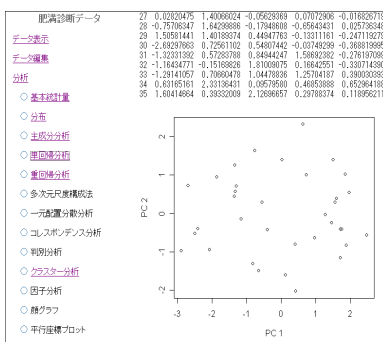


図 3：主成分分析

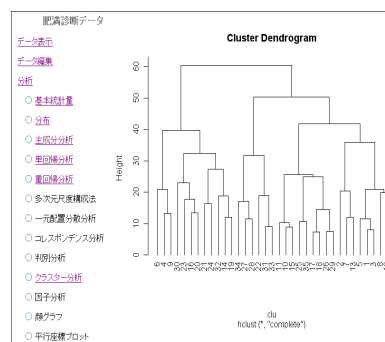


図 4：クラスター分析

<回答者>は検査・調査 Web システムの[回答者入口]から回答者向けページに移動すると、画面左下に調査票データベース中にある検査・調査の一覧が表示される。そこで実行したい検査・調査を選択するとサーバから調査票がダウンロードされ、画面左上にフォームが表示される。回答者はこのフォームに自分の情報を記入したり、与えられた質問に回答したりする。[診断]ボタンをクリックすると、即座に診断結果が回答者に与えられる（図 5）。診断結果、アドバイスと並んで、スライダーという部品も表示される。画面中の は本人の入力を基に求められた BMI 値の位置を表しており、この をドラッグし左右に動かすと、移動先の が指す BMI 値を得るための体重が計算され上部に表示される。回答者はこれを参考に、自分の目標とする体重を探ることができる。また、回答結果における統計解析を行い出力された分析結果も部品として同時に表示される。これには、回答者自身の情報を強調して表示し、全体中での自分の位置が一目でわかるようになっている。スライダーと同じように、診断結果の判断材料とすることができる。図 6 は単回帰分析で出力された図に回答者のデータ部分を強調したものである。なお、これらの部品はすべて調査票 XML ファイルの中で指定できる。

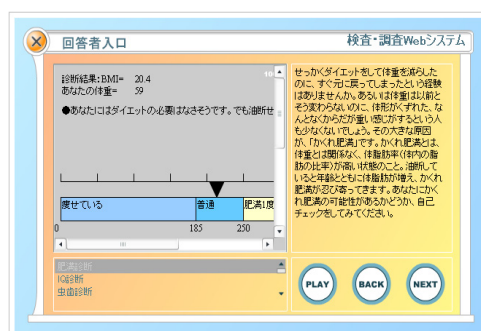


図 5：診断結果（部品：スライダー）

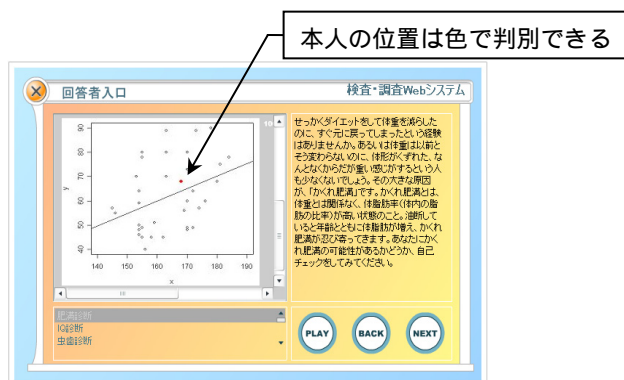


図 6：診断結果（部品：分析結果）

## 5 おわりに

検査・調査 Web システムにおいて統計解析機能を導入することで、調査者や回答者に対し有益な情報を提供することを目的に、実用的で汎用性の高いシステム構築をしてきた。

調査者や回答者は、特別な知識をもたずとも統計解析の結果を見ることができ、Web に特化したシステムであるので、時間や場所を問わず本システムを利用することができる。調査者に対して与えられる分析結果は複数用意されており、その分析結果を出力するための処理も容易で、表示されているリンクをクリックするだけでよい。そして回答者に対しても R によって統計解析された分析結果を診断結果として付与することで、検査・調査の結果を判断する際の新たな判断材料として提供することができるようになった。また、回答者に、分析結果やスライダー、アドバイスなどの部品を与えるとき、これらの情報は自動的に作成可能な調査票 XML にすべて含まれているため、部品の変更が容易にできる。

今後の課題としては統計手法の拡張が挙げられる。さまざまな統計手法を実現させることで、調査者にも回答者にもプラスとなって反映されることになる。

## 参考文献

- 小野寺孝義・菱村 豊(2005).文科系学生のための新統計学，ナカニシヤ出版.
- 包 一娜(2008).検査・調査 Web システムの開発，平成 19 年度岡山理科大学修士論文.
- 渡辺利夫(2005).フレッシュマンから大学院生までのデータ解析・R 言語，ナカニシヤ出版.