

データからのモードの推定

川合伸幸

初等統計学に残された問題の一つに、データからどのようにしてモードを推定したらよいかということがある。情報を集約する仕方としてモードをとり、モードを比較するのが適切な場合というのはかなりあるので、データから直接モードを推定する方式が確定すれば有用な道具になると考えられる。

モードは分布の中心を示す。そこで次のような素朴な推定量が求まる。各点 x_0 を中心とする長さ $a > 0$ の区間 $(x_0 - a, x_0 + a)$ を考え、この中に含まれるデータ点の数がいちばん多くなるような x_0 を中心あるいはモードと考える。これは長さ a に依存するので $Mo(a)$ と書くことにする。まず前提として、母集団の分布 $F(x)$ は連続な密度関数 $f(x)$ をもち、モードは一意に与えられるものとする。さらに数学的議論には関係ないが実際的な観点から密度関数は一山だと仮定する。そして $f(x)$ はモードのまわりで2階微分可能とする。 X_1, X_2, \dots, X_n を $F(x)$ からのランダム・サンプル、 X_1, X_2, \dots, X_n から得られる経験分布関数を $F_n(x)$ とする。漸近的には $f(x)$ が mode の a 近傍で左右対称でないと、 $Mo(a)$ はかたよりをもちることがわかる。そして $a \rightarrow 0$ とするとき $Mo(a)$ は母集団分布 $F(x)$ のモードに収束する。そこで、

- 1° a はできるかぎり 0 に近い
- 2° データから $Mo(a)$ は確率 1 で一意に定まる

という二つの条件を満たすように a を求める。それに対応するモード推定量 $Mo(a_0)$ を求めると、

$$\min_{1 \leq k \leq n-1} \{X(k+1) - X(k)\}$$

(ただし $X(k)$ は X_1, X_2, \dots, X_n の順序統計量) を与えるような k を k_0 とし、

$$Mo(a_0) = \{X(k_0+1) + X(k_0)\} / 2$$

となる。 $X(k_0+1) - X(k_0)$ は確率 1 で 0 に収束するから、 $Mo(a_0)$ は Chernoff (1964) より一貫性があることがわかる。一方、漸近分散は複雑な量になり、推定量の変動を評価するにはブートストラップ法を使うことが考えられる。

参考文献

Chernoff, H. (1964). Estimation of the mode, *Ann. Inst. Statist. Math.*, **16**, 31-41.

統計情報資源の統合化の試み

—— 統計メタ情報ナビゲーション・システム ——

大隅 昇

研究機関で収集・蓄積された統計情報資源（共同研究情報、図書登録情報、各種の出版刊行物、統計ソフトウェア、各種調査データなど）は、それぞれは自由に利用できても、保存形態やメディアの種類が多様で相互に関連づけて利用する支援環境が必ずしも十分とはいえない。これら統計情報資源の統合化を進め共有化を図ることは重要な課題である。その共通した特徴として、① 統計情報の形態が多様なメディアの上にある、② 統計情報が統合化されていない、③ 統計情報が一元化されず散在している、④ 各情報が断片的で相互の関連が必ずしも明らかでない、などがある。このことから、統計数理研究所を例として、ハイパーテキスト化やハイパーメディア環境の利用を前提とした統計情報資源活用のためのナビゲーション・システムの基本構想を検討した。この際、既存の統計情報取得の方法や利用環境を

保持しながら、段階的に新たなシステム開発を進めることが重要であるとの視点から、まず研究所組織の活動あるいは統計情報資源の全容を概観する機能（組織と統計情報の全体像のブラウジング機能）、次に、個々の情報資源の内容を詳細に参照し、相互の関連を照合するナビゲーション機能を中心に設計指針の試案を作成した。システムの設計方針としては、① ヒューマン・インターフェースが平易で操作が容易であり、利用者の興味を喚起するものであること、② 既存の統計情報資源の保存形態を正確に掌握して、これらをハイパーテキスト化する技術とそのアプリケーションの設計、つまり多様な形態の統計情報資源を共有化するための技法を確立すること、③ ハイパーテキストに基づく、リレーショナル・データベースの利用を前提に、ハイパーデータ化された統計情報の参照・検索を容易にするネットワーク・システム環境を構築すること、などの観点から設計試案の作成を進めた。とくに本年度は、情報資源として、実験的に共同研究情報と図書登録情報の一部を用いたリレーショナル・データベース化を試み、これを利用したナビゲーション・システムの構想の一部を試作システムとして具体化し、ハイパーテキスト化の技術的な問題、ハイパーメディア利用に適したインターフェースのあり方、ネットワーキング環境の問題点、マルチメディア環境下での統計情報のデータベース化の問題点などについて実験を通じて検討した。

生活習慣と健康事象に関する疫学研究

高木 廣文

我が国を含め先進国の現在の疾病構造は、感染症による死亡の減少と悪性新生物、心疾患、脳卒中などの非感染性慢性疾患の増加という点から特徴づけられる（AIDSの問題はあるが）。これらの、いわゆる「成人病」は長期に渡る日常的な各種の習慣の影響により、発症に到るものと考えられており、この点から「習慣病」とも呼ばれることもある。しかし、喫煙と肺癌のように、因果関係が比較的明らかなのは少なく、実際にはどのような習慣が特定の疾患と結びつくのかは、必ずしも明白ではない。また、実際問題として「生活習慣」をどのように測るのかも問題となる。我々は、1981年に初めて生活習慣調査のための質問紙を作成した（日野原 他（1982））。これは、食習慣、嗜好、価値観、性格特性、精神的健康度、身体的健康度などを測るための30尺度（255項目）からなるものであった。その後、簡便化を図り、1983年には23尺度（195項目）からなる調査票に改訂し（佐伯 他（1988））、1991年には22尺度（136項目）と飲酒、喫煙に関する項目を独立させた。生活習慣尺度の信頼性は θ 信頼性係数が比較的大きいこと、test-retest法での回答の一致率が大きいこと、単相関係数が十分大きいこと、などにより確認されている（Takagi et al. (1991)）。

上記のように開発した生活習慣尺度と健康事象との関係を検討するために、農村地域、漁業地域の主に40歳代、50歳代の主婦を中心とした調査を実施した。対象者数は1990年調査では112名、1991年調査では都会の主婦を加え、304名となり、生活習慣検査票、食事調査（3日間連続の国民栄養調査方式、および頻度調査）、24時間尿による塩分排泄量、血液・尿検査、などを調べた。興味ある点として、生活習慣尺度の高塩分尺度と塩分排泄量の単相関係数は0.410であったが、食事調査から算出される塩分摂取量と24時間尿からの塩分排泄量の単相関係は0.393しかなく、国民栄養調査などで従来からよく用いられている、食事調査からの塩分摂取量の算出には問題があるものと考えられた。今後、これらの点を含めて調査、解析を進める予定である。

参 考 文 献

- 日野原重明, 柳井晴夫, 高木廣文, 柏木恵子, 日野原緑 (1982). 循環器疾患予防のための生活習慣に関する研究 (第1報) 生活習慣の多変量解析による分析, 日本公衆衛生雑誌, **29**, 309-320.
佐伯圭一郎, 高木廣文, 日野原重明, 柳井晴夫, 道場信孝, 水口緑 (1988). LPC式生活習慣検査の作成, 行動計量学, **15**(2), 32-44.