最後は，第 2 次抽出単位である投票区の抽出である．抽出された市区町村について投票区別の有権者数を入力し，スタート番号を含む投票区を抽出する．市区町村でのスタート番号は投票区でのスタート番号へ引き継がれる．投票区別のデータを予め入力しておくのは 1 回限りのサンプリングでは不経済なので，市区町村を抽出した段階で入力している．

最終的な調査対象者の抽出は，各地点を担当した調査員が面接の直前に各市区町村の選挙管理委員会を訪れて行なう．指定された投票区を見つけ，その中のスタート番号にあたる有権者から 10 人あるいは 20 人おきに調査対象者を転記する．このあたりの現状についてはまた別のトピックになる．

以上述べたように，サンプリング計画はほぼ決まりきった手順で能率的に行なえるようになってきているが，社会調査は，回収率の低下に現れているように調査環境の悪化の中にある．このような状況にどのように対処していけばよいか，統計学がどのような役割を果たすことができるのか，取り組むべき課題は大きい．

パターン分類と e_{ij} 型数量化

林　　文

数量化の方法のうち，外的基準のない場合を扱うものに，数量化 III 類（パターン分類の数量化）と数量化 IV 類（e_{ij} 型数量化）がある．e_{ij} 型数量化は，項目同志の関係 —— 何等かの親近性を表す数値 —— のデータをもとに項目間の近さを再現する空間配置を求めるもので，データに何の制約も無いため，種々の分析の初期値等として適用対象が広い．一方，パターン分類は，個人が選択項目の中からいくつかを選択したときの反応パターンの似たものを集めて，人と項目を同時に分類するものである．このように二つの方法は出発点の考え方は異なるが，パターン分類の数量化も，解く数式は項目間のクロス集計による項目間関係を用いることになる．そこで，個人の項目選択のデータに対し，項目間の同時選択の頻度を用いて親近性として，ある適当な定義を与えて e_{ij} 型数量化を実行すると，パターン分類と同じ解が得られる場合がある．その一例として次のことが言える．

e_{ij} 型数量化の親近性として個人の回答データから

$$e_{ij}=\frac{1}{2N}\frac{d_{ij}-\dfrac{d_i d_j}{N}}{\sqrt{\dfrac{d_i}{N}}\sqrt{\dfrac{d_j}{N}}}$$

をとることとする．ここで，各個人の選択した項目数が一定という条件(条件 1)を仮定しておく．また，d_i：第 i 項目を選択した人数，d_{ij}：第 i 項目と第 j 項目を同時選択した人数とする．この e_{ij} は明らかに i, j 項目間の親近性の尺度となっている．このとき一致する例として，各項目を選択した人の数が一定（条件 2）のとき，また，条件 2 を満たす完全 scalable データのときにも一致する．条件 1 だけのとき，解くべき固有方程式の対角線上の要素は一致しないが，それ以外は一致し，解は scale を調整するとそれほど異ならない．条件 1，2 を共に満たさないデータに対しても，一般には解は異なるが，相互の位置関係が全く異なるということはない．

このほかにも解が同じになる定義やデータの条件は何か，また，解の異なりかたを調べることは，数量化の意味を考える上で必要であると考える．