

## Comparative Study on Statistical Packages for using Multivariate Q-technique<sup>1)</sup>

Yong-Seok Choi<sup>2)</sup>, Hee-jung Moon<sup>3)</sup>

### Abstract

In this study, we provide a comparison of multivariate Q-techniques in the up-to-date versions of SAS, SPSS, Minitab and S-plus well known to those who study statistics. We can analyze data through the direct input method(command) in SAS and use of menu method in SPSS, Minitab and S-plus. The analysis performance method is chosen by the high frequency of use. Widely we compare with each Q-techniques form according to input data, input option, statistical chart and statistical output.

*Keywords:* Multivariate analysis, Q-techniques, SAS, SPSS, Minitab, S-plus.

### 1. 서 론

다변량 자료분석기법에는 크게 R-기법과 Q-기법으로 나눌 수 있다. R-기법은 개체공간(R-space)에서 변수들 간에 상관관계를 통하여 유사한 변수를 찾아내는데 관심을 두고 있고 Q-기법은 변수공간(Q-space)에서 개체간의 거리를 보아 서로 가까우면 유사한 집단으로 분류하는데 관심을 두고 있다(Chatfield and Collins, 1980, Chapter 1; Mardia et al., 1979, p. 16). 본 소고에서는 현재 많이 이용되고 있고 국내에서 비교적 쉽게 구할 수 있는 개인 컴퓨터용 통계 팩키지인 SAS, SPSS, Minitab, S-plus를 대상으로, 각 팩키지에서 어떤 Q-기법 다변량 분석이 가능한지를 파악하고, 어느 한계까지 입력조건과 출력조건을 수용하는지 비교하려 한다. 이러한 비교는 사용자들에게 가장 적합한 팩키지를 선택하여 Q-기법 다변량 자료분석을 할 수 있도록 도움을 주는데 목적이 있다. Q-기법은 변수공간(Q-space)에서 개체간의 거리를 보아 서로 가까우면 유사한 집단으로 분류하는데 관심을 두는데, 다변량 자료분석 중 대응분석(correspondence analysis), 군

1) This work was supported by Pusan National University Research Grant.

2) Professor, Department of Statistics, Pusan National University, Pusan 609-735, Korea.  
E-mail: yschoi@pusan.ac.kr

3) Graduate Student, Department of Statistics, Pusan National University, Pusan, 609-735, Korea

집분석(clustering), 판별분석(discriminant analysis) 그리고 다차원 척도법(multidimensional scaling)이 포함된다.

SAS, SPSS, Minitab, S-plus 모두 명령어를 통한 직접 입력 방식과 메뉴 선택 방식으로 분석을 수행 할 수 있는데, 본 소고에서는 사용자들이 각 팩키지를 사용할 때 많이 활용하는 분석 수행의 방법을 선택하였다. 그 결과 SAS는 명령어를 통한 직접 입력 방식, SPSS, Minitab, S-plus는 메뉴 선택 방식이 선택되었다. 2절에서는 다변량 Q-기법의 비교에 앞서 각 기법의 실행과정과 입력 자료 형태를 살펴보았다. 3절에서는 비교되는 Q-기법의 입력 옵션, 통계량 출력 항목 순서로 각 팩키지들을 비교하였다. 4절에서는 본 소고를 정리하고 요약하였다.

## 2. 다변량 Q-기법 실행경로와 각 기법의 입력 자료 형태

<표 2.1>은 SAS, SPSS, Mintab, S-plus에서 다변량 Q-기법을 수행하는 메뉴 경로와 명령어를 하나의 표로 나타낸 것이다. <표 2.1>에서 볼 수 있듯이 SAS는 직접 입력 방식의 명령어를, SPSS, Minitab, S-plus는 메뉴 선택 경로를 표기하였다(SAS Institute Inc, 1999; Minitab Inc 1999; 김명근·정강모, 2000). 그리고 SPSS, Minitab, S-plus에서 메뉴 선택 방법으로는 실행되지 않지만, 명령어를 통해 분석이 가능한 것은 명령어와 함께 command라는 표시를 해 두었다. 대응분석은 범주를 나타내는 변수가 둘 뿐인 이원 분할표를 위한 단순대응분석과 범주를 나타내는 변수가 세 이상인 다원 분할표를 위한 다중대응분석으로 구분하였고, 군집분석과 판별분석은 분석 방법에 따라 각각 계층적 군집분석과 비계층적 군집분석으로, Fisher의 판별분석과 정준 판별분석, 단계적 판별분석으로 구분하였다.

<표 2.1> Q-기법 다변량 분석의 실행경로

| 다변량 자료분석 |          | SAS                                       | SPSS                   | Minitab   | S-plus  |
|----------|----------|---|------------------------|---|---|
| 대응<br>분석 | 단순 대응    | PROC CORRESP                              | 분석> 데이터 축소><br>대응일치 분석 | Stat> Tables> Simple<br>Correspondence Analysis   | ×   |
|          | 다중 대응    |   | ×                      | Stat> Tables> Multiple<br>Correspondence Analysis                                       | ×   |
| 군집<br>분석 | 계층적      | PROC CLUSTER<br>PROC VARCLUS<br>PROC TREE | 분석> 분류분석><br>계층적 군집분석  | Stat> Multivariate><br>Cluster Observation,<br>Stat> Multivariate><br>Cluster Variables | Statistics> Cluster Analysis><br>Agglomerative<br>Hierarchical,<br>Statistics> Cluster Analysis><br>Divisive Hierarchical |
|          | 비계층적     | PROC FASTCLUS                             | 분석> 분류분석><br>K-평균 군집분석 | Stat> Multivariate><br>Cluster K-means  | Statistics> Cluster Analysis><br>K-Means  |
| 판별<br>분석 | Fisher's | PROC DISCRIM                              | 분석> 분류분석><br>판별분석      | Stat> Multivariate><br>Discriminant   | Statistics> Multivariate><br>Discriminant Analysis  |
|          | 정준       | PROC CANDISC                              |                        | ×   | ×   |
|          | 단계적      | PROC STEPDISC                             |                        | ×   | ×   |
| 다차원 척도법  |          | PROC MDS                                  | 분석> 척도화 분석><br>다차원 척도법 | ×   | CMDSCALE(command)   |

<표 2.2>는 SAS, SPSS, Mintab, S-plus에서 다변량 분석을 할 때 가능한 입력 자료의 형태를 비

교한 표이다(강병서·김계수, 1998; 우수명, 2001).

판별 분석의 입력 자료의 형태는 행은 개체, 열은 변수들로 표현된 원 자료인데, 개체 특성에 따라 구분되어진 범주형 변수가 자료에 포함되어야 한다. 다차원 척도법의 경우 SAS에는 이분형 자료나 빈도 자료가 입력 자료 형태가 되었을 때 그 자료를 유사성/비유사성 자료로 만드는 프로그램이 필요하다. 반면에 SPSS는 다차원 척도법에 포함되어 있는 옵션으로 유사성/비유사성 자료를 만들 수 있다.

<표 2.2> 다변량 Q-기법별 팩키지의 비교

| 다변량 자료분석         |       | SAS                                   | SPSS                 | Minitab                  | S-plus              |
|------------------|-------|---------------------------------------|----------------------|--------------------------|---------------------|
| 대<br>용<br>분<br>석 | 단순 대응 | 분할표, 분류변수 자료, 교차표                     | 분류변수 자료, 분할표         | 교차표, 분류 변수 자료            | ×                   |
|                  | 다중 대응 | 분류변수 자료, 버트 표, indicator table        | ×                    | 분류변수 자료, Indicator table | ×                   |
| 군집 분석            |       | 비유사성 행렬-거리, 원자료                       | 비유사성 행렬-거리 /이진수, 원자료 | 비유사성 행 -거리, 원자료          | 비유사성 행렬-거리/이진수, 원자료 |
| 판별 분석            |       | 원 자료(개체의 특성에 따라 구분되어진 범주형 변수가 자료에 포함) |                      |                          |                     |
| 다차원 척도법          |       | 유사성/비유사성 자료                           | 원자료, 유사성/비유사성 자료     | ×                        | ×                   |

### 3. 팩키지의 다변량 Q-기법 비교

#### 3.1 대응분석

<표 3.1>은 대응분석의 입력 옵션을 각 팩키지 별로 비교한 결과이다. 다중 대응분석에서 새로운 정보를 가진 범주를 추가하려고 할 때 SAS에서는 옵션 SUPPLEMENTARY를 사용하고 Minitab에서는 메뉴 Supp Data를 선택하면 되지만 SPSS에서는 이 기능의 옵션이 없다.

<표 3.1> 대응분석의 입력 옵션

| 팩키지     | 입력 옵션         |   |                        |
|---------|---------------|---|------------------------|
|         | 범주의 추가        | 좌표점 선택  | 차원지정                   |
| SAS     | SUPPLEMENTARY | ROW=A,AD,DA,DAD,DAD1/2,DAID1/2<br>COLUMN=B,BD,DB,DBD,DBD1/2,DBID1/2<br>PROFILE=BOTH,COLUMN,NONE,ROW | DIMENS=2               |
| SPSS    | ×             | 대칭적, 행 주 좌표,<br>열 주 좌표  | 해법차원: 2                |
| Minitab | Supp Data     | 대칭적, 행 주 좌표,<br>열 주 좌표  | Number of components:2 |

볼드체: 디폴트로 지정 ×: 입력 옵션이 없음

<표 3.2>는 대응분석의 출력항목을 각 팩키지 별로 비교한 것이다. SPSS는  $\chi^2$ 값의 교차표와  $\chi^2$ 값을 구하는 과정의 기대도수 교차표, 관찰도수-기대도수의 교차표는 출력되지 않았다. SAS에서 옵션 OBSERVED를 사용하면 단순 대응분석에서는 관측도수의 교차표를 출력하고, 다중 대응분석에서는 베트표를 출력할 수 있다. 그리고 옵션 RP, CP를 사용하면 행·열 프로파일을 구할 수 있고 옵션 CELLCHI2를 사용하면  $\chi^2$ 을 구할 수 있다. SPSS에서는  $\chi^2$ 값의 교차표는 출력되지 않고,  $\chi^2$ 의 전체 합만을 디폴트로 보여주고 있다(최용석, 2001).

&lt;표 3.2&gt; 대응분석의 통계량 출력 항목

| 출력 항목             | 팩키지 |      |         |
|-------------------|-----|------|---------|
|                   | SAS | SPSS | Minitab |
| 관찰도수의 교차표         | ○   | ◎    | ○       |
| $\chi^2$ 값 교차표    | ○   | ×    | ○       |
| 전체 $\chi^2$ 합     | ◎   | ◎    | ○       |
| 차원축소에 대한 설명(주결여성) | ◎   | ◎    | ◎       |
| 기대도수의 교차표         | ○   | ×    | ○       |
| 관찰도수-기대도수의 교차표    | ○   | ×    | ○       |
| 설명비율              | ◎   | ◎    | ◎       |
| mass              | ◎   | ◎    | ◎       |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

### 3.2 군집분석

<표 3.3>은 계층적 군집분석에서 입력옵션을 각 팩키지 별로 비교한 결과이다. SPSS에서 군집계산 방법으로 언급되어진 용어들은 실제 옵션으로 제공되어지고 있는 단어들이다. 집단간 연결은 군집 간 평균 연결법을, 집단내 연결은 군집 내 평균연결을 뜻한다. SAS에서 군집 추출 방법이 Average, Centroid, Median, Ward의 경우에는 개체간의 거리측정 방법의 디폴트가 제곱 유클리디안 거리이다. 유클리드 거리를 사용하려면 옵션 NOSQUARE를 쓴다. 군집화 방법이 Two-stage density나 Density인 경우에 거리를 표현하기 위해서 DIM옵션이나, TRIM옵션이 사용되기도 한다(김충련, 1996, 13장).

&lt;표 3.3&gt; 계층적 군집분석의 입력 옵션

| 파키지     | 계층적 군집의 입력 옵션   |   |
|---------|---|---|
|         | 군집 추출 방법  | 개체간의 거리측정 방법  |
| SAS     | average, centroid, complete, density, maximum-likelihood, flexible, MaQuitty's similarity, median, single, two-stage density, Ward's minimum-variance | <b>Squared Euclidean</b> , Euclidean,                                     |
| SPSS    | 집단간 연결, 집단내 연결, 가장 가까운 항목, 가장 먼 항목, 중심점 군집화, 중위수 군집화, Ward의 방법  | 제곱 유클리디안 거리, 코싸인, Pearson 상관, Chebychev, 블록, Minkowski                    |
| Minitab | Average, Centroid, Complete, McQuitty, Median, Single, Ward   | <b>Euclidean</b> , Manhattan, Pearson, Squared Euclidean, Squared Pearson |
| S-plus  | <b>Average</b> , Complete, Single, Ward, Weight   | <b>Euclidean</b> , Manhattan  |

볼드체: 디폴트로 지정

<표 3.4>는 계층적 군집분석에서의 통계량 출력 항목을 각 파키지 별로 비교하고 있다. 원 자료가 입력되었을 경우 SPSS와 Minitab에서 상이성 행렬을 볼 수 있는데 이는 개체들 사이의 유클리디안 거리들을 보이고 있다. SAS에서 CCC는 군집의 개수를 결정하는데 사용되는 판정 기준이다. 옵션 CCC을 이용하면  $R^2$ 의 값도 함께 구할 수 있다. 이외에 Pseudo F, Pseudo  $t^2$  이 군집을 결정하는데 사용된다. SPSS에서는 군집화 일정표에서 계수라는 명칭으로, Minitab에서는 Amalgamation Steps에서 Distance Level이라는 명칭으로 단계별 군집간의 거리를 나타내고 있는데 급속히 증가하는 단계에서 군집의 개수를 결정하도록 한다(허명희 · 양경숙, 2001, 6장; 최용석 · 정광모, 2001).

<표 3.5>는 K-means 군집분석의 통계량 출력항목을 나타내고 있다. SAS와 SPSS에서는 비교하는 모든 통계량이 출력되었다. 각 변수에 대한 초기의 군집 중심 값은 개체 할당을 위한 임시 값이라고 할 수 있다. 이러한 초기 군집 중심 값을 기준으로 각 개체와 각 군집의 중심점과의 거리를 계산하여 가장 가까운 군집에 개체를 할당한다. 반복 계산의 정보에는 반복계산에 따른 군집 중심의 변화량이 나타나 있다.

SAS에서는 옵션 MAXITER, LIST, DISTANCE를 이용하면 각각 반복계산 정보, 군집 중심과 개체간의 거리, 군집 중심간의 거리를 구할 수 있다. SPSS, Minitab, S-plus에서는 개체의 소속군집이 새로운 변수로 자료시트에 출력되어진다.

&lt;표 3.4&gt; 계층적 군집분석의 통계량 출력 항목

| 출력 항목            | 팩키지                                     |                   |                   |        |
|------------------|---|-------------------|-------------------|--------|
|                  | SAS                                     | SPSS              | Minitab           | S-plus |
| 군집의 형성단계         | ◎                                       | ◎                 | ◎                 | ×      |
| 상이성 행렬           | ×                                       | ○                 | ○                 | ×      |
| 개체의 소속<br>군집 표기  | ○                                       | ○                 | ○                 | ○      |
| 군집중심간의 거리        | ×                                       | ×                 | ○                 | ×      |
| 군집중심과<br>개체간의 거리 | ×                                       | ×                 | ○                 | ×      |
| 군집의 수치<br>판단 기준값 | CCC, $R^2$<br>Pseudo F,<br>Pseudo $t^2$ | 단계별<br>군집간의<br>거리 | 단계별<br>군집간의<br>거리 | ×      |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

&lt;표 3.5&gt; K-means 군집분석의 통계량 출력 항목

| 출력 항목            | 팩키지 |      |         |        |
|------------------|-----|------|---------|--------|
|                  | SAS | SPSS | Minitab | S-plus |
| 초기 군집 중심값        | ◎   | ◎    | ×       | ×      |
| 반복 계산 정보         | ○   | ◎    | ×       | ×      |
| 최종 군집 중심값        | ◎   | ◎    | ◎       | ◎      |
| 군집중심과<br>개체간의 거리 | ○   | ○    | ○       | ×      |
| 군집 중심간의 거리       | ○   | ○    | ○       | ×      |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

### 3.3 판별분석

<표 3.6>는 판별분석의 입력옵션을 팩키지 별로 비교한 것이다. 판별분석의 입력 옵션 중 사전 확률과 교차 확인법을 각 팩키지 별로 비교하였다. 한 집단이 다른 집단에 비해 규모가 훨씬 크기 때문에 표본의 출현 가능성이 더 많은 경우가 있는데, 이런 경우에는 사전 확률을 분류규칙에 반영하는 것이 바람직하다. S-plus의 경우 교차 확인법에 대한 옵션은 없지만 디폴트로 교차 확인법에 의한 분류표를 제시한다.

&lt;표 3.6&gt; 판별분석의 입력 옵션

| 팩 키지    | 입력 옵션                        |                      |
|---------|------------------------------|----------------------|
|         | 사전 확률                        | 교차 확인법               |
| SAS     | PRIORS PROP                  | CROSSVALIDATE        |
| SPSS    | 분류 > 사전 확률 > 집단 표본 크기로 계산    | 분류 > 출력 > 순차제거 복원 분류 |
| Minitab | Options > Prior probabilites | Use cross validation |
| S-plus  | Model > Group > proportional | 디폴트                  |

<표 3.7>과 <표 3.8>는 피셔의 판별분석, 정준 판별분석, 단계적 판별분석에서 공통으로 출력되는 결과들이다. 선형 판별함수에 관한 분류 규칙에는 다변량 정규분포와 공통 공분산 행렬의 가정을 필요로 하기 때문에 선형 판별함수를 적용할 때는 이에 대한 검토가 선행되어야 한다. SAS에서는 옵션 MANOVA를 사용하면 집단간 평균의 동질성에 대한 결과를 얻을 수 있고, 옵션 POOL=TEST를 이용하면 집단간 공분산 행렬이 동등한지 검정할 수 있다(SAS Institute Inc, 1999, 23장).

&lt;표 3.7&gt; 판별분석의 공통 통계량 출력 항목-1

| 팩 키지    | 판별분석 공통 출력                         |  |
|---------|------------------------------------|--|
|         | 집단간 분산의 동일성에 대한 Test               | 집단간 평균의 동일성에 대한 Test   |
| SAS     | Bartlett의 수정 최우법<br>검정( $\chi^2$ ) | Wilks Lamda, Pillai Trace, F<br>Hoteling Lawley Trace, Roy Greatest Root |
| Spss    | Box.M                              | Wilks Lamda, F   |
| Minitab | ×                                  | ×  |
| S.plus  | Box.M<br>adj.M                     | Wilks Lamda, Pillai Trace, F<br>Hoteling Lawley Trace, Roy Greatest Root |

×: 출력되지 않음

<표 3.8>은 SAS, Minitab, S-plus에서는 집단간 거리로 마할라노비스 거리가 디폴트로 출력됨을 보여 주고 있다. SAS에서 옵션 LISTERR을 사용하면 오 분류된 사례 목록을 구할 수 있고, CROSSLISTERR을 사용하면 교차 확인법에 의한 오 분류된 사례 목록을 구할 수 있다(김기영·전명식, 1994). SPSS는 따로 오 분류된 사례 목록을 보이지 않고 전체 개체 소속 집단에 별표로써 오 분류 된 목록을 표기하고 있다.

&lt;표 3.8&gt; 판별분석의 공통 통계량 출력 항목-2

| 출력 항목                  | 파키지 |      |         |        |
|------------------------|-----|------|---------|--------|
|                        | SAS | SPSS | Minitab | S-plus |
| 판별함수에 근거한 소속 분류표       | ◎   | ○    | ◎       | ◎      |
| 교차 확인법에 의한 소속 분류표      | ○   | ○    | ○       | ◎      |
| 교차 확인법에 의한 개체 예측 소속 집단 | ○   | ×    | ○       | ○      |
| 집단간 거리                 | ◎   | ×    | ◎       | ◎      |
| 개체 판별점수                | ○   | ○    | ×       | ×      |
| 오분류된 사례목록              | ○   | ×    | ◎       | ×      |
| 사후 확률 오차율              | ○   | ×    | ×       | ×      |
| 통합 상관행렬                | ○   | ○    | ×       | ×      |
| 개별 집단<br>공분산 행렬        | ○   | ○    | ○       | ×      |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

<표 3.9>는 정준 판별분석을 실행했을 때 보여지는 출력결과이다. 정준 상관 관계는 판별점수와 집단간의 관련정도를 나타내는 것이고, 고유값은 집단내의 분산을 집단간 분산으로 나눈 값이다. 고유치가 클수록 우수한 판별함수라고 할 수 있다.

&lt;표 3.9&gt; 정준 판별분석의 통계량 출력 항목

| 출력 항목              | 파키지 |      |        |
|--------------------|-----|------|--------|
|                    | SAS | SPSS | S-plus |
| 고유값                | ○   | ◎    | ○      |
| 정준 상관 관계           | ○   | ◎    | ○      |
| 정준 상관계수의 유의성 Test  | ○   | ○    | ○      |
| 변수의 판별<br>함수 기여도   | ○   | ○    | ×      |
| 표준화 정준<br>판별함수의 계수 | ○   | ◎    | ×      |
| 함수의 집단<br>중심점      | ○   | ○    | ×      |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

<표 3.10>은 단계별 판별분석을 실행했을 때 보여지는 출력결과이다. SAS에서는 속성 변수들 중에서 각 변수들이 진입했을 때  $R^2$ 의 증가량과 변수의 유의도를  $F$ 값으로 표시하고, 이 변수가 진입했을 다중공선성을 검정하는 통계량으로서 Tolerance가 제시되어 있다. SPSS에서는 단계별 변수에 대한 다중공선성 Test는 최소 공차 한계의 값으로 확인할 수 있다.

&lt;표 3.10&gt; 단계별 판별분석의 통계량 출력 항목

| 출력 항목                       | 팩키지 |      |
|-----------------------------|-----|------|
|                             | SAS | SPSS |
| 단계별 변수에 대한 결정계수             | ○   | ×    |
| 단계별 변수에 대한 유의도 Test( $F$ 값) | ○   | ○    |
| 단계별 변수에 대한 다중공선성 Test       | ○   | ○    |
| 단계별 판별식의 유의도                | ○   | ○    |

○: 입력 옵션에 의한 출력 ×: 출력되지 않음

### 3.4 다차원 척도법

<표 3.11>과 <표 3.12>는 다차원 척도법에서 각 팩키지 별로 입력 옵션의 형태를 비교한 결과이다. <표 3.11>은 다차원 척도법의 입력 자료의 조건을 지정하는 옵션과, 이원과 삼원 다차원을 지정하는 입력 옵션과 비유사성 측정 척도를 비교한 결과이다. SAS에서 입력 자료의 조건을 지정하는 옵션으로 CONDITION과 SHAPE가 있다. 그리고 옵션 COEF는 이원과 삼원을 지정하는 옵션이다. 유클리드 거리를 이용하는 이원 다차원척도법의 경우 IDENTITY로 이는 디폴트이다.

&lt;표 3.11&gt; 다차원 척도법의 입력 옵션-1

| 팩키지  | 입력 옵션   |                                 |   |
|------|---|---------------------------------|---|
|      | 입력 자료의 조건을 지정   | 이원과 삼원 다차원 척도법 지정               | 비유사성의 측정 척도                                 |
| SAS  | CONDITION=UN/MATRIX /SUBJECT/ROW SHAPE=TRIANGULAR /SQUARE | COEF=IDENTITY /DIAGONAL         | ABSOLUTE/INTERVAL/RATIO/ORDINAL/LOGINTERVAL |
| SPSS | 모형> 조건> 행렬/행/조건 없음  | 다차원척도법(PROXSCAL)>단일행렬소스/ 다중행렬소스 | 순서 척도/구간 척도/비율척도                            |

&lt;표 3.12&gt; 다차원 척도법의 입력 옵션-2

| 팩키지  | 입력 옵션                       |           |                          |
|------|-----------------------------|-----------|--------------------------|
|      | 유사성 행렬에서<br>비유사성 행렬로의<br>변환 | 차원 수 지정   | 최적 변환 지정                 |
| SAS  | SIMILAR                     | DIMENSION | FIT=DISTANCE/SQUARED/LOG |
| SPSS | ×                           | 모형 > 차원   | ×                        |

×: 입력 옵션이 없음

<표 3.12>에 대한 설명이다. SAS에서 옵션 SIMILAR는 입력 자료가 유사성임을 나타내고 이 유사성 행렬을 비유사성으로 바꾸라는 옵션이다. SAS에서 최적변환 지정의 디폴트는 FIT=DISTANCE로 Kruskal의 스트레스 1과 2를 이용한다. FIT=SQUARED는 Young의 S-스트레스 1과 2를 이용하게 한다(최용석·정광모, 2003, 7장). SPSS에는 최적변환을 지정하는 옵션은 없지만 출력에서 Kruskal과 Young의 스트레스를 표시한다.

&lt;표 3.13&gt; 다차원 척도법의 통계량 출력 항목

| 출력 항목            | 팩키지 |      |
|------------------|-----|------|
|                  | SAS | SPSS |
| 스트레스             | ○   | ○    |
| 다차원 그림을<br>위한 좌표 | ○   | ○    |
| 비유사성 행렬          | ×   | ○    |

◎: 디폴트로 출력 ○: 입력 옵션에 의한 출력 ×: 출력되지 않음

<표 3.13>는 다차원 척도법에서 각 팩키지 별로 통계량 출력 항목을 비교한 결과이다. 다차원분석을 할 때 입력 자료가 원 자료이면 SPSS에서는 비유사성 행렬을 출력하지만 SAS에서는 입력 자료가 비유사성 혹은 유사성 행렬이어야 된다. SAS에서는 입력 자료가 원 자료일 때 다차원분석을 하려면 원 자료를 유사성 혹은 비유사성 행렬로 바꿔야하는 절차를 거쳐야한다. 그런데 SAS의 다차원 척도법에서는 이 과정의 옵션이 없다.

#### 4. 결 론

통계 팩키지를 중 국내에서 비교적 접하기 쉬운 SAS, SPSS, Minitab, S-plus에서 다변량 분석 중 Q-기법을 비교한 결과 대응분석, 군집분석, 판별분석, 다차원 척도법이 가능하였다. SAS에서

는 이들 다변량 Q-기법이 모두 가능하였고, SPSS에서는 메뉴방식으로 다중 대응분석이 불가능하였다. Minitab에서는 단계적 판별분석과 정준 판별분석, 다차원 척도법이, S-plus에서는 메뉴 방식으로 대응분석과 단계적 판별분석, 다차원 척도법이 불가능하였다. SAS는 다변량 분석에서 많은 입력 옵션과 통계량 출력 결과를 세밀하게 제공하였다. SPSS와 Minitab은 사용자들이 쉽고 편리하게 다변량 분석을 할 수 있게 제공하고 있지만, 위에서 언급했듯이 SPSS는 다중 대응분석이, Minitab은 단계적 판별분석과 다차원 척도법이 불가능한 단점이 있다. 그리고 SPSS는 통계량 출력 항목이 표로 정리되어 사용자가 결과를 해석하는데 편리함을 준다. S-plus는 위에서 언급했듯이 대응분석과 단계적 판별분석, 다차원 척도법이 불가능하였다. 그리고 다른 팩키지 보다 통계량 출력 항목들을 많이 제공하지 못하였고, 통계량 출력의 형태 면에 있어서도 분석의 내용이 잘 정리되지 않은 느낌을 주었다.

### 참 고 문 현

- [1] 강병서 · 김계수 (1998). 「사회과학 통계분석」, SPSS 아카데미, 서울.
- [2] 김기영 · 전명식 (1994). 「SAS 판별 및 분류분석」, 자유 아카데미, 서울.
- [3] 김명근 · 정강모 (2000). 「S-plus를 이용한 다변량 자료분석」, 교우사, 서울.
- [4] 김정순 · 나종화 (2000). 「S-plus 사용법 및 프로그래밍」, 자유 아카데미, 서울.
- [5] 김충련 (1996). 「SAS라는 통계상자」, 데이터 리서치, 서울.
- [6] 우수명 (2001). 「마우스로 잡는 SPSS 10.0」, 인간과 복지, 서울.
- [7] 최용석 (2001). 「SAS 대응분석의 이해와 응용」, 자유 아카데미, 서울.
- [8] 최용석 · 정광모 (2001). 「실무자를 위한 MINITAB 다변량 분석」, 이레테크, 서울.
- [9] 최용석 · 정광모 (2003). 「SAS를 활용한 다변량 분석 기법과 응용」, 자유아카데미, 서울.
- [10] 허명희 · 양경숙 (2001). 「SPSS 다변량 자료분석」, SPSS아카데미, 서울.
- [11] Minitab Inc. (1999). *MINITAB User's Guide 2, Release 13*. Minitab Inc, USA.
- [12] SAS Institute Inc. (1999). *SAS/STAT User's Guide, Version 8*. SAS Institute Inc, Cary, NC.

[ 2003년 5월 접수, 2003년 7월 채택 ]