

SPSS의 利用方法

本研究院 資料處理課長 文 顯 相

[1] SPSS는 Statistical Package for the Social Science의 略字이다. 글자 그대로 社會科學에서 대상으로 하는 各種 社會調查資料를 統計的으로 分析하는 컴퓨터 프로그램이다.

이 프로그램은 미국의 Stanford 大學에서 처음 개발되어 IBM 7090型, IBM OS/360型 컴퓨터에서 작업이 가능하도록 되어 있었으나 最近에 와서 그 活用 범위가 확대되면서 CDC 6000型 컴퓨터에서도 사용할 수 있도록 프로그램이 개편되었다.

家族計劃研究院은 한국과학기술연구소가 가지고 있는 CDC 6000型 컴퓨터에 Terminal을 설치할 예정이므로 Terminal 利用者の 편의를 위하여 同 프로그램을 입수하여 보급함으로써 人口 및 家族計劃 分野뿐만 아니라 各種 社會調查 자료의 分析에 광범위하게 利用할 계획이다.

따라서 여기서는 SPSS의 구성과 그 利用方法을 간단히 소개하고자 한다.

[2] SPSS는 CONDESCRIPTIVE, CODEBOOK, MARGINALS, CROSSTABS, FASTABS, BREAKDOWN, PEARSON CORR, NONPAR CORR, PARTIAL CORR, REGRESSION, GUTTMAN SCALE, FACTOR로 Identify 되는 統計的 分析을 가능하도록 구성되어 있다.

CONDESCRIPTIVE 라는 Subprogram은 社會調查에서 취급되는 各 變數들의 MEAN, VARIANCE, RANGE, STANDARD ERROR, KURTOSIS, STANDARD DEVIATION, SKEWNESS, MINIMUM, MAXIMUM을 산출해 준다.

GODEBOOK은 各 變數들의 Category別 分布와 百分率을 계산한다. 예를 들어 어떤 調查에서 조사대상자의 혼인상태를 알아보기 위해서 既婚, 死別, 離婚, 別居, 未婚등으로 구분했다면 이 프로그램은 조사대상자의 이들 각각 Category에 대한 分布와 百分率을 계산한다.

MARGINALS는 CODEBOOK와 유사한 프로그램인데 Category수가 많은 경우 예를 들어 연령분포와 같이 0세~100세까지의 100개의 Category가 있을 경우 이에 대한 소위 Marginal Table을 빠른 속도로 낼 수 있는 프로그램이다.

CROSSTABS는 통상 社會調查 자료의 分析에서 쓰고 있는 Cross Tabulation을 가능하게 하는 프로그램이다. SPSS의 CROSSTABS는 表1과 같은 Printing form을 갖는다.

Cross Tabulation의 백분율 계산뿐만 아니라 이의 통계적인 有意性 검증까지도 산출한다.

FASTABS는 變數가 Numerically Code 되어 있을때 빠른속도로 CROSSTABS와 마찬가지로의 통계처리를 할 수 있다.

BREAKDOWN은 [圖1]과 같이 전체 조사대상을 특정한 GROUP로 나누었을 때 각 GROUP들의 MEAN, STANDARD DEVIATION, VARIANCE를 계산한다.

PEARSON CORR는 各 變數間의 相關係數를 산출하고 係數의 有意性도 검증한다. 統計學의 Pearson Correlation Coefficient를 계산한 프로그램이다.

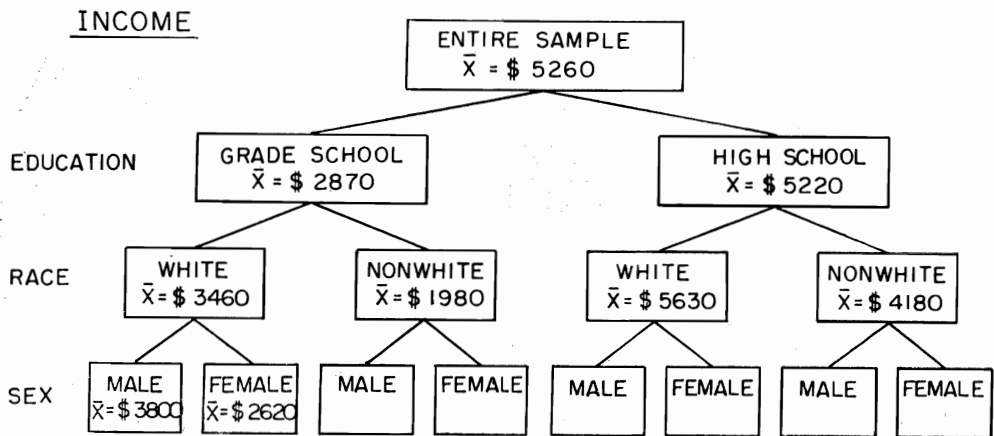
NONPAR CORR는 Spearman Correlation Coefficient, Kendall Correlation Coefficient, Rankorder Correlation Coefficient를 계산한다.

CROSSTABS RUN
 FILE FAMILY PLANNING
 SUBFILE SEOUL

CROSS TABULATION OF

EDUCATION	BY						NUMBER OF PREGNANCY	
	NUMBER OF PREGNANCY							
EDUCATION	NONE	ONE	TWO	THREE	FOUR	FIVE+	ROW TOTAL	
NONE								
1-8 YEARS								
9-11 YEARS								
12 YEARS AND OVER								
COLUMN TOTAL								

CHI SQUARE=
 CRAMER'S V=
 CONTINGENCY COEFFICIENT=
 KENDALL'S TAU B=
 GAMMA=
 SOMER'S D=



PARTIAL CORR 는 Partial Correlation Analysis 를 하는 프로그램으로 係數의 산출 뿐아
 니라 有意性 검증도 할 수 있다.

REGRESSION 은 Multiple-Regression Analysis 를 할 수 있는 프로그램이다. SPSS 의
 Standard Regression Coefficient 는 Path Analysis 도 가능하게 하고 특히 變數 변환기능은
 Harmonic, Polynomial regression 도 가능하게 한다.

GUTTMAN SCALE 는 Scalogram analysis 를 가능하게 하는 프로그램이다. FACTOR 는
 FACTOR Analysis 를 하는 프로그램이다. SPSS 의 FACTOR 는 correlation matrix,
 communalities, eigenvalues, proportion of variance, unrotated factor matrix, final
 communalities, percent of common variance rotated orthogonal factor matrix, factor

estimate matrix 를 산출한다.

[3] SPSS 利用者는 그가 使用하려는 Statistical Subprogram 에 따라 Control Card 를 작성함으로써 컴퓨터에서의 작업이 가능하다. 간단한 例를 들어 Control Card 작성법을 소개코자 한다. 例-1은 SPSS 의 가장 간단한 例이다. 분석될 자료는 20명의 대학교수들의 정치의견에 대한 자료이다.

例 1	<u>教授名</u> (Card Col. 1~8)	<u>政治意見</u> (Col. 10)	<u>年齡</u>	<u>性別</u>	<u>宗教</u>
	JONES	C	43	M	1
	SMITH	N	26	M	1
	OSGOODE	L	35	F	2
	MCFEE	R	50	M	1
	ALLISON	N	61	F	2
	SCHOLTZ	L	31	F	3
	KLINE	L	45	F	3
	BENSON	C	56	M	1
	RICHARDS	S	37	F	2
	AITKEN	N	45	M	4
	MACKAY	S	30	M	2
	SCHWARTZ	S	44	M	1
	YATES	N	0	F	2
	MIMFR	C	38	M	1
	ALVEREZ	L	0	F	2
	SAWCHOK	L	35	M	1
	FLYNN	C	42	M	2
	BATES	N	36	F	4
	NEMDY	L	29	M	4
	HALLER	L	39	F	3

즉 資料는 Fixed-column format 으로 작성되었다.

SPSS 의 利用者는 資料의 形態에 따라 Data-Definition Card 를 작성하여야 한다.

Data-Definition Card 의 첫 Card 는 FILE NAME card 이다. 이 카드는 Data File 의 이름이다. 例-1에서는

Card

Colum

1

16

FILE NAME FACSTUDY, SURVEY OF FACOLTY PARTY PREFERENCES

로 작성하였다. 즉 이용자는 그가 원하는대로 研究資料의 이름을 붙일 수 있다.

Data-Definition Card 의 두번째 카드는 VARIABLE LIST 카드이다.

이 카드는 研究資料에 포함되어 있는 變數들의 이름을 작성하는 카드이다.

例-1에서는

Card

Column

1

16

VARIABLE LIST

PROF, PARTYPRE, AGE, SEX, RELIGION

즉 교수 이름을 PROF 로, 정치의견을 PARTYPRE 로, 연령을 AGE 로, 성별을 SEX 로, 종교를 RELIGION 으로 이름 붙였다.

Data-Definition Card 의 세번째 카드는 INPUT FORMAT 카드이다.

이 카드는 分析될 資料가 Card 에 어떤 위치에 Punch 되어 있는지를 알려주는 카드이다.

例-1에서는

Card

Column

1

16

INPUT FORMAT

FIXED (A8, 1X, A1, 1X, F2.0, 1X, A1, 1X, F1.0)

資料의 구성이 Fix 되어 있느냐 아니면 Fix 되어 있지 않느냐에 따라, 資料가 數字로 되어 있느냐 혹은 文字로 되어 있느냐에 따라 이를 Define 해야되기 때문에 例-2의 FIXED 는 한개의 Data Record 가 아래와 같이 되어 있을 때 사용된다.

Case 1	001	36.2	4	9	17.3	Yes---
Case 2	002	45.4	7	8	25.9	No---
Case 3	003	71.6	4	3	96.2	Yes---

(Case 마다 한 카드)

Case 1	0011	D	19.01	4	Yes
	0012	F	37.38	7	Yes
Case 2	0021	A	14.73	3	No
	0022	R	29.24	7	No
Case 3	0031	R	43.27	7	No
	0032	A	82.91	8	No

(Case 마다 두 카드)

Data 의 Card 上의 위치 및 種數는 아래와 같이 작성된다.

Data 의 종류 文字 : A FORMAT

 數字 : F FORMAT

Data 의 위치 Skipping: X FORMAT

 Skipping Cards: 1

따라서 例-1에서와 같이 카드의 1 column~8 column까지는 文字로 기록된 변수이고 10 column이 역시 文字로 기록된 변수이고 12 column~13 column이 數字로 된 變數이고 15 column이 文字變數이고 17 column이 數字變數인 경우 control card의

Card

Column

1

16

INPUT FORMAT

FIXED(A8,1X, A1, 1X, F2.0, 1X, A1, 1X, F1.0)

에서 A8은 文字로 된 8 column 넓이의 變數라는 것을 말하고 1X,는 한 column 을 skip 하는 것이고, A1은 文字로 된 1 column 넓이의 變數이고, F2.0는 數字變數인데 그 column 넓이이고 소수점 이하가 없다는 것을 뜻한다. 참고로 F FORMAT 의 작용을 소개하면 아래와 같다.

<u>Format element</u>	<u>Value appearing on data card</u>	<u>Value read and stored</u>
F3.0	100	100
F3.1	100	10.0
F3.1	.10	0.10
F3.1	010	01.0
F3.2	1.0	1.00

control card 의 네번째 카드는 NO. OF CASES Card 이다. Data File 의 Record 數를 말한다. 例-1에서는

card column	
1	16
NO. OF CASES	20

20개의 record 로 작성된 file 이기 때문에 20으로 기록된다.

control card 의 다섯번째 카드는 INPUT MEDIUM 카드이다. 例-1에서는 資料가 card 를 통해 入力되기 때문에

card column	
1	16
INPUT MEDIUM	CARD 로 작성되었다.

control card 의 여섯번째 카드는 VAR LABELS 카드이다. 이 카드는 變數들의 이름을 설명해주는 카드로서 例-1에서는

card column	
1	16
VAR LABELS	PROF, FACULTY MEMBER'S NAME/PARTYPRE, POLITICAL PARTY PREFERENCE/

等으로 表示된다.

control card 의 7번째 카드는 VALUE LABELS 카드이다. 이 카드는 통상 調查資料의 CODEBOOK 에서 작성된다.

Card Column	
1	16
VALUE LABELS	PARTYPRE ("C") Cord SERVATIVE ("L") LIBERAL ("S") SOCIAL CREDIT ("N") NEW DEMOCRAT ("R") NOT GI- VEN/ SEX ("M") MALE ("F") FEMALE/ RELIGION (1) PROTESTANT (2) CATHOLIC (3) JEWISH (4) OTHER

以上이 Data-Definition Card 의 例이다. Data Definition Card 다음에는 Task Definition Card 가 온다. Task Definition Card 는 (A) CONDESCRIPTIVE 일 경우는

CONDESCRIPTIVE	VAR1 TO VARIO
STATIATICS	ALL

하면 VAR 1에서 VAR 10까지 즉 變數 10개에 대한 CONDESCRIPTIVE Subprogram 에 의한 통계적 처리가 진행된다.

(B) CODEBOOK 일 경우는

CODEBOOK	VAR 1 TO VAR 10
----------	-----------------

하면 10개 變數의 CODEBOOK Subprogram 에 의한 처리가 가능하다.

(C) MARGINALS 일 경우는

MARGINALS
STATISTICS

AGE, INCOME

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11

하면 AGE, INCOME 이라는 變數의 marginal tabulation 과 1~11까지의 통계치가 계산된다. 1에서 11까지의 통계치의 내용은 아래와 같다.

- | | |
|-----------------------|-------------|
| 1. MEAN | 7. KURTOSIS |
| 2. STANDARD ERROR | 8. SKEWNESS |
| 3. MEDIAN | 9. RANGE |
| 4. MODE | 10. MINIMUM |
| 5. STANDARD DEVIATION | 11. MAXIMUM |
| 6. VARIANCE | |

(D) CROSS TABS 일 경우는

CROSSTABS RACE, SEX, INCOME BY EDUCATION
CROSSTABS RACE TO OCCUPATION BY INCOME
CROSSTABS RACE TO INCOME BY EDUCATION BY CITY

등으로 Cross Tabulation 의 목적에 따라 2-way 부터 n-way cross tabulation 이 가능하다.

(E) BREAKDOWN 일 경우는

BREAKDOWN INCOME TO VAR 1 BY RACE TO AGE INCOME 변수 부터 VAR 1 변수 까지의 모든 변수들을 RACE와 AGE 順으로 Grouping 하여 BREAKDOWN 프로그램을 실행 한다.

(F) PEARSON CORR 일 경우는

PEARSON CORR VAR 1 WITH VAR 5 하면 VAR 1이 VAR 5의 상관계수를 산출한다.

以上에서 Task Definition Card 의 예를 보듯이 SPSS 는 매우 간단한 방법으로 SPSS PROGRAM FILE 에 들어 있는 각종 Subprogram 들을 자유로이 活用할 수 있다.

여기서 SPSS 사용을 위한 Deck Set up 을 소개하면

	카드	
	Column	Column
	1	16
	RUN NAME	EXAMPLE
Data Definition Card	—FILE Name	FAMILY PLANNING-1
	VARIABLE LIST	VAR 1, VAR 2, VAR3, ...
	INPUT FORMAT	FIXED (A1, A2, A3, F1.0, ...)
	⋮	
	—MISSING VALUES	VAR1 ("R")
Task Definition Card	—CROSSTABS	VAR1 by VAR 2
	—OPTIONS	3,5
	—STATISTICS	1,3
	READ INPUT DATA	
Data Record	
	
	
	FINISH	