CHAPTER 4

통계 소프트웨어

4.1. 사전 준비

조사 목적을 달성할 수 있는 설문지 (초안)이 만들어지면 사전 조사를 통해 설문지의 문제점이나 불완전한 부분을 보완하여 설문지를 완성한다. 조사원 교육 후 본 조사를 시행하고 조사원은 반드시 조사 보고서를 작성한다.

설문 조사가 완성되면 답변을 중단했거나 (완료하지 못한 설문은 응답을 신뢰할 수 없다) 성실하지 못하게 답변한 것 (각 문항 답변이 동일하게 계속 되거나 장난처럼 답변한 것 같은 설문지), 조사자가 설문지 회수 후 설문지 앞에 불성실 응답이라고 기록한 설문지를 제외하고 설문지에 일련 번호를 매긴다. 만약 서로 다른 집단을 조사하였으면 각 집단에 따로 일련 번호를 매긴다. 일련 번호를 부여하는 이유는 코딩 후 입력 오류가 발생하였을 때 쉽게 발견하기 위함이다.

첫 번째 설문지에는 통계 분석 시 문항에 부여하게 될 변수 이름을 적어 놓는다. 변수 이름 적는 방법은 4.1.2 절을 참고하기 바란다. 변수명을 적은 첫 번째 설문지는 분석이 완료되는 순간까지 항상 지니고 다닌다.

문항 번호를 부여할 때는 다음 원칙을 지킨다.(SAS 사용 위주)

(1)문항 하나에 변수 명을 배정하는 것이 원칙이며(Q1, Q2, …), 다중 문항이나 우선 순위 문항 경우에는 Q*_1, Q*_2, 식으로 변수명을 설정한다.

(2)문항 선택 보기에 번호가 없는 경우 분석자가 임의로 배정하되 1, 2, 3 식으로 부여한다.

(3)문항 보기가 10 개이면 1, 2, …, 10 이 아니라 1, 2, …,9, 0 식으로 배정한다. 이유는 일반적으로 설문 응답 결과를 입력할 때 빈 칸 없이 붙여 입력하므로 응답 결과를 한자리에 입력하는 것이 편리하기 때문이다.

4.1.1. 예제 설문

(1)조사 목적

○○대학교 중국경제통계학부 신입생들의 대학 시설 및 구성원에 대한 만족도와 전공 선택에 대한 의견을 수렴하고자 2001 학년 중국경제학부 신입생 전체를 대상으로 설문조사를 실시하였다.

(2)조사 대상

○○대학교 중국경제학부 01 학번 신입생

(3)데이터 수집 방법

설문지를 이용한 응답자 자기 기입식(self administrated survey) 면접 조사 방법을 사용하였으며 조사자는 각 반 “정보와 통계” 과목을 담당하는 교수이다. 그러므로 단체 면접 조사 방법이 사용되었다.

(4)표본 추출 방법 및 표본의 크기

모집단은 2001 년 3 월 26 일 현재 중국경제학부에 재학하는 01 학번 학생 140 명 중 조사 당일 결석한 학생을 제외한 모든 대상자 전수 조사(census) 실시하였다.

(5)설문지 설계

①서론

조사 목적, 응답자 협조의 글

②인구학적 변인(face item) (3 문항)

성별, 재수 유무, 출신 고등학교 소재 지역에 다른 본 항목 응답 결과의 차이를 보기 위하여 3 문항으로 구성하였다/

③본 항목

▪시설물 대한 만족도: 건물 공간, 강의실, 실습실, 화장실 (10 문항)

▪구성원에 대한 만족도: 교수, 조교 (4 문항),

▪교양 과목에 대한 만족도: 인성, 교양 세미나, 영어, 컴퓨터 과목 (4 문항)

▪한남대학교 전체에 대한 만족도: 한남대학에 대한 느낌 (4 문항)

▪전공 선택에 대한 의견: 전공 선택 기준 및 전공 (3 문항)

▪불만족 시설 선택 (1 문항)

(6)조사 일정 및 방법

학생이 공통으로 수강하는 “정보와 통계” 3 월 27 일 3 교시 수업 시간에 설문지를 나누어 준 후 학생들이 스스로 설문에 응답하는 방법으로 조사하였다. 과목 담당 교수가 설문지를 나누어 주고 응답 결과를 회수하였다. 응답 시간은 10 분을 초과하지 않게 실시하였다.

(7)사전 조사

동일 대상(99 학번 중국경제정보통계학부)에게 같은 설문 조사 목적으로 사용하였던 설문지를 이용하였으므로 사전 조사는 실시하지 않는다.

(8)본 조사 실시 결과 및 설문지에 번호 매기기

전혀 응답되지 않았거나 문항의 반 이상이 응답되지 않은 설문지는 제외하고 조사된 설문지는 총 130 부이다. 각 설문지에 1 부터 130 까지 일련 번호를 부여하였다. 각 문항에는 하나의 변수명을 부여하였고 다중 선택 문항에는 최대 선택 수만큼 지정해 주고 우선순위 문항에는 우선 순위 개수만큼 지정하였다.

(9)입력 사전 작업

①인구학적 변인에서 보기 번호가 없는 것은 임의로 번호를 부여한다.

▪성별: 여자➔1, 남자➔ 2

▪재수 여부: 재수 않음=1, 재수 이상=2

▪출신 고등학교: 대전➔1, 충남➔2, …, 기타 지역➔5

③주관식 문항은 분류표를 미리 만들어 그 번호를 입력한다.

영어 영문학 01, 경영학 02 … (학과가 10 개 이상으므로 코딩 할 때는 2 자리 입력한다. 물론 spreadsheet 에 데이터를 입력하면 01 대신 1 을 입력하면 된다. 코딩 데이터(CODING.txt)에는 이 문항을 입력하지 않았다. 자료 데이터는 저자 홈페이지에 올려져 있다. User id 는 SURVEY, Password 는 JAN2004 이다.)

4.1.2. 설문지

다음은 응답 된 첫 번째 설문지와 문항 번호를 부여한 결과이다.

본 설문 조사는 한남대학교 중국경제통계학부 신입생들의 대학 시설 및 구성원에 대한 만족도와 전공 선택에 대한 의견을 수렴하고자 실시하는 것입니다. 여러분의 응답이 보다 나은 학생 서비스 제공을 위한 학부 발전 계획 수립에 반영되므로 대학을 사랑하는 마음으로 성의 있게 응답해 주시기 바랍니다. 여러분의 협조에 심심한 감사를 드립니다.

중국경제정보통계학부

※여러분에게 해당하는 곳에 0표 하시오.

Q1□여자(1) 남자(2)

선택 보기 문항▶각 한 자리씩 부여

Q2□재수하지 않음(1) 재수 이상(2)

Q3□출신 고등학교 소재지? 대전(1) 충남(2) 서울(3) 경기(4) 그 외 지역(5)

각 문항에 미리 번호 부여

※다음은 대학 시설물에 대한 질문입니다. 여러분이 바라보는 수준에서 볼 때 다음 각 항목에 대해 자신의 의견을 잘 나타내는 숫자에 0표 하시오.

리커드 척도 문항

Q4①경상대학 건물 안의 공간은?

매우 쾌적하다 7 6 5 4 3 2 1 매우 답답하다

Q5②경상대학 건물 안팎의 휴식 공간은?

매우 충분하다 7 6 5 4 3 2 1 매우 부족하다

Q6③강의실 공간은 수업을 하는데 있어~

매우 여유 있다 7 6 5 4 3 2 1 매우 비좁다

Q7④강의실 안의 시설 및 비품은 수업을 하기에~

매우 잘 갖추어져 있다 7 6 5 4 3 2 1 매우 부족하다

Q8⑤강의시간에 보조기자재를 이용하는 것은?

매우 편리하다 7 6 5 4 3 2 1 매우 불편하다

Q9⑥경상대학 내에 외국어 공부를 하기 위한 시설은?

매우 적절하다 7 6 5 4 (3) 2 1 매우 부족하다

Q10⑦경상대학 내에 컴퓨터 실습을 위한 시설은?

매우 적절하다 7 6 5 4 3 (2) 1 매우 부족하다

Q11⑧경상대학 내에 도서관 시설은?

매우 적절하다 7 6 5 4 3 2 (1) 매우 부족하다

Q12⑨경상대학 화장실 시설은?

매우 청결하다 (7) 6 5 4 3 2 1 매우 불결하다

Q13⑩경상대학 시설에 대해 전체적으로 평가할 때~

매우 만족한다 7 6 5 4 3 2 (1) 매우 불만스럽다

※다음은 학내 각 구성원에 대해 자신이 동감하는 숫자에 0표 하시오.

Q14①나는 교수님들이 강의에 열성적이라고 생각한다.

매우 동의한다 7 6 5 4 (3) 2 1 전혀 동의하지 않는다

Q15②수업시간 중에 교수님과 질의응답은 충분히 이루어지고 있다.

매우 동의한다 7 6 5 4 (3) 2 1 전혀 동의하지 않는다

Q16③강의실 밖에서도 교수님과 상담할 기회는 적절히 제공되고 있다.

매우 동의한다 7 6 5 4 3 2 (1) 전혀 동의하지 않는다

Q17④조교 선생님은 학생들에게 협조적이다.

매우 동의한다 7 (6) 5 4 3 2 1 전혀 동의하지 않는다

※다음에 나열된 의견에 대해 자신이 동감하는 정도를 선택해 주십시오.

Q18①현재 개설 되어있는 교양 세미나 과목은 대체적으로 유익하다고 생각한다.

매우 동의한다 7 6 5 4 3 2 (1) 전혀 동의하지 않는다

Q19②현재 개설 되어있는 현대사회와 인성교육 과목은 대체적으로 유익하다고 생각한다.

매우 동의한다 7 6 5 4 3 2 (1) 전혀 동의하지 않는다

Q20③교양 영어(의사소통 영어) 수강은 영어 실력의 향상에 도움이 된다고 생각한다.

매우 동의한다 7 (6) 5 4 3 2 1 전혀 동의하지 않는다

Q21④교양 컴퓨터 과목은 컴퓨터 실력 향상에 도움이 된다고 생각한다.

매우 동의한다 7 6 (5) 4 3 2 1 전혀 동의하지 않는다

※다음은 한남대학교 학생으로서 느끼는 점에 대한 질문입니다.

Q22①○○대학교는 대전·충남 지역의 대학에 비해 그 수준이~

매우 우수하다 7 6 (5) 4 3 2 1 매우 떨어진다

Q23②나는 ○○대학교를 다니는 것에 대해~

매우 자부심을 갖는다 7 6 (5) 4 3 2 1 매우 부끄럽다

Q24③내가 ○○대학교에 입학한 것은~

매우 잘한 일이다 7 (6) 5 4 3 2 1 매우 잘못한 일이다

Q25④○○대학교에 대해 전체적으로~

매우 만족스럽다 7 6 5 (4) 3 2 1 매우 불만스럽다

※여러분이 선택할 전공에 관한 질문입니다. **우선 순위 문항**

①전공을 선택할 때 중요하게 생각하는 순서대로 번호(1~5)를 적으시오.

Q26_1▷취업 전망(1) Q26_2▷학문적 우월성(3) Q26_3▷나의 적성(2)

Q26_4▷전공 교수의 질(4) Q26_5▷선후배 관계(5)

Q27②오늘 전공을 선택한다면 어느 전공을 선택할 것입니까?

▷중국 전공( ) ▷경제 전공( ) ▷정보통계 전공( )

Q28③지난 대학 원서 접수하던 때로 돌아가 봅시다. 위에서 선택한 전공이 여러분이 대학에서 전공하기 원했던 전공입니까?(중국, 경제, 정보통계 전공 이외에 다른 모든 전공도 포함해서)

▷예( ) ▷아니오( ) ► 아니라면 원했던 전공은?( )

아래 문항은 실제 조사에서 실시되지 않았으나 다중 선택 문항 분석 방법을 보여 주기 위하여 문항을 삽입하였고 응답 결과도 임의로 입력한 것이다.

다중 선택 문항

※다음 시설 중 가장 불만족한 시설을 선택하시오.(다중 선택 가능, 최대 3개 까지)

①화장실( ) ②체육 시설( ) ③휴식 공간( )

④도서관( ) ⑤어학실( ) ⑥실습실( )

▶ Q29_1 ~ Q29_3

▶설문지 끝

4.2. 통계 소프트웨어

통계 소프트웨어(statistical software)는 수집된 데이터에 적절한 통계 분석을 적용하여 원하는 정보를 얻는데 도움을 주는 컴퓨터 소프트웨어이다. 통계 소프트웨어의 발달은 분석을 위한 계산 시간 절약 및 계산의 정확성 제고는 물론 통계 비전문가라도 손쉽게 통계 수치와 관련 그래프를 얻을 수 있게 하였다.

통계 소프트웨어는 입력 데이터를 인식할 때 그저 하나의 숫자로 인식하므로 명령의 오류가 없으면 데이터 분석이 적절하지 않더라도 분석 결과를 출력한다. 이로 인하여 통계의 오용과 남용을 불러 일으킨다. 통계적 안목이 없거나 적절하지 못한 분석 방법을 사용하여 얻는 결과는 이제 더 이상 정보가 아닐 뿐 아니라 여러(잘못된 의사 결정으로 인한 손실, 다른 연구에 불이익, 그릇된 발표로 인한 사회적 손실) 문제를 일으킨다. 통계 비전문가들은 분석에 대한 확신이 없을 때는 반드시 통계 전문가에게 통계 상담을 받기를 권한다.

통계 소프트웨어의 종류는 다양하다. 가장 많이 사용되는 통계 소프트웨어인 SAS (Strategic Application System 통합 응용 시스템, http://www.sas.com/offices/asiapacific/korea/index.html), 사회 과학 분야에서 주로 사용되는 SPSS (Statistical Package Software: 사회 과학), 경영과

학 QC(6-sigma)에서의 Minitab (경영과학, http://www.spss.co.kr/), 그래픽 툴이 강한 SYSSTAT, STATGRAPHICS, 시뮬레이션과 그래프의 리더 S-plus, 수학적 배경이 강한 Mat Lab, 경제학에서 주로 사용되는 RATS, RVIEW 등이 있다. 스프레드시트용 소프트웨어 EXCEL 에도 기초적인 데이터 분석 기능이 포함되어 있다.

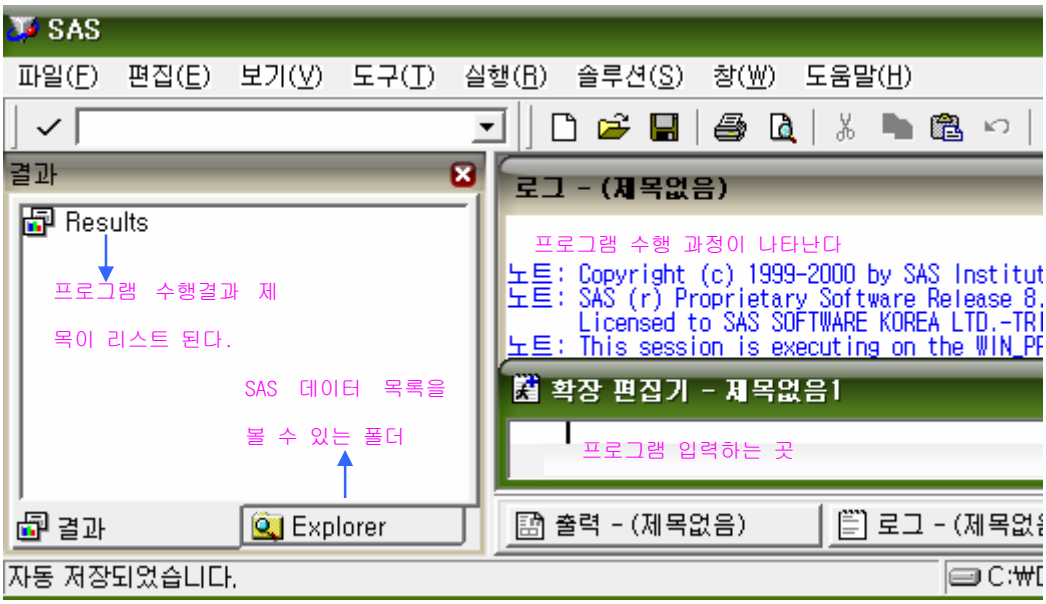
4.2.1. SAS

<SAS 의 주요기능 및 제품명>

주요기능	제품명
데이터 추출/변형	SAS/ACCESS, BASE SAS
데이터 전송	SAS/CONNECT, SAS/SHARE(SHARE*NET)
다차원 OLAP용 데이터 서버	SAS/MDDB Server
대용량 DSS 서버	Scalable Performance, Data Server
QUERY & REPORTING	Enterprise Reporter
데이터 조회 및 관리	SAS/FSP
데이터 추출 및 메타데이터관리	SAS/Warehouse Administrator
데이터 마이닝	Enterprise Miner
분석 지원 모듈	SAS/STAT, SAS/IML, SAS/OR, SAS/ETS, SAS/INSIGHT, SAS/LAB, SAS/CALC
시각화	SAS/GRAPH, SAS/SPECTRAVIEW
대화형 어플리케이션 개발 툴	SAS/AF
웹 어플리케이션 개발	SAS/Internet
IT 서비스 통합 평가 툴	IT SERVICE VISION
통합 재무제표 관리	CFO VISION
지리정보시스템 구축	SAS/GIS
편리한 사용자 인터페이스 지원	SAS/ASSIST
통계적 공정 관리	SAS/QC
품질관리를 위한 실험계획 도구	JMP

SAS *The Power to Know®*. 의 역사는 미국 North Carolina 주 Raleigh 에 있는 NCSU(North Carolina State University) 통계학과 대학원 과정 학생들이 주축이 되어 Statistical Analysis System 을 완성하던 1966 년으로 거슬러 올라간다. 1972 년 SAS72 가 각 대학에 Shareware 버전으로 제공되어 사용되다가, 1976 년 Cary (NCSU 에서 15 분 거리의 도시)에 SAS Institute 를 설립하면서 SAS 제품을 판매되기 시작했다. 초기에는 데이터를 검색하고 통계 분석 및 해석을 위한 소프트웨어였으나 제품이 개발되면서 통합 응용패키지(SAS 약어 바꿈: Strategic Application System)로 발전하였다. 현재 SAS 는 전세계 118 개국 (미국, 57 개국 지사), 40,000 업체(기업, 정부, 연구소, 학교)에서 사용되고 있으며, Fortune 500 기업 중 90%가 SAS 를 사용하고 있다. SAS version 8 이 사용 (version 9 시험판 출시)

<SAS 의 초기화면>



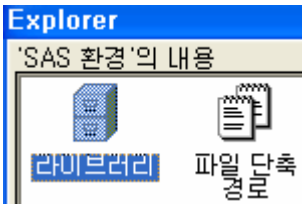
SAS 가 시작되면 초기 화면이 열리고 커서(cursor)는 PROGRAM(확장 편집기) 윈도우에 있다. SAS 의 초기 화면은 4 개의 창으로 구성되어 있는데, 프로그램 결과 창/SAS 데이터를 보여주는 Explorer 창이 함께 있는 결과/Explorer 창과 DMS(Display Manager System: 화면 관리 체계)의 기본적인 윈도우인 확장 편집기(Program editor), 로그 창(Log window), 결과 창(Output window)가 나타난다. 출력 창은 뒤에 숨어 있다.

확장 편집기는 프로그램을 작성하고 편집하는 기능과 작성된 프로그램을 실행하는 기능을 가지고 있다. 확장 편집기에서 실행된 프로그램 결과는 출력 창에 나타난다. 프로그램을 실행하였음에도 불구하고 출력 창에 결과가 출력되지 않으면 프로그램에 에러가 발생한 경우

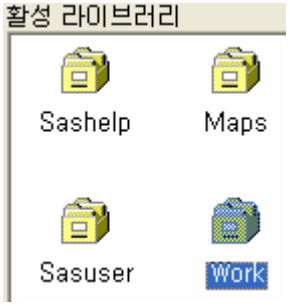
이다. 이때는 로그 창에서 에러 메시지를 확인한다. 에러 메시지 또는 경고 메시지를 참고하여 프로그램을 수정하고 재실행하면 원하는 결과를 출력 창에서 얻을 것이다.

로그 창에는 실행한 프로그램에 대한 과정이 출력한다. 실행된 프로그램은 물론, 프로그램에 대한 일반적인 정보(NOTE 문)와 프로그램에 문제가 있을 때 에러 메시지(ERROR 문)를 출력한다. 또한 에러는 아니지만 프로그램 수행 시 주의해야 할 사항에 대한 경고 메시지(WARNING 문)도 출력한다. 로그 창에는 프로그램의 실행결과에 대한 여러 가지 정보가 제공되므로 비록 결과 창에 결과가 출력되더라도 프로그램을 실행하면 반드시 로그 창을 참고하기를 권한다.

▶ Explorer 창을 살펴 보자.



라이브러리에는 SAS data 정보가 저장되어 있다.



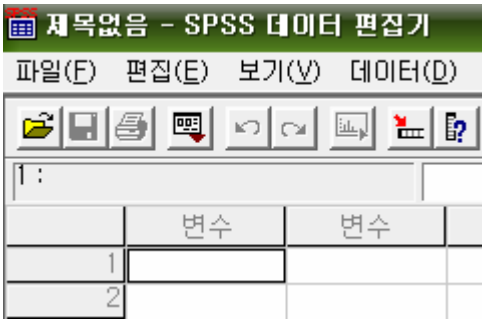
라이브러리에는 SAS 에 포함되어 있는 데이터나 여러 설정 Catalog 들이 있는 SASHELP, MAPS, SASUSER 라이브러리와 SAS 시작과 함께 만들어지는 WORK 라이브러리가 있다. 사용자들이 데이터를 만들면 이곳에 임시 저장되었다가 SAS 를 종료하면 없애 버린다. DATA ONE; 을 만들면 실제 이름은 WORK.ONE 이므로 WORK 라이브러리에 ONE 이라는 이름으로 저장된다.

SAS 확장자를 보면 프로그램은 *.sas, SAS 데이터는 *.sas7bdat, 출력 결과는 *.lst 이다.

4.2.2. SPSS

SPSS (Statistical Package for Social Science)는 1968 원시 데이터(raw data)로부터 기업의 의사결정에 이용되는 정보를 얻기 위한 통계 분석을 위해 개발되어 현재는 SPSS version 10 이 출시되어 사용되고 있다. 다음은 SPSS 제품의 주요 기능을 요약한 것이다.

- ▪데이터 접근 SPSS Data Entry
- ▪데이터 준비 Sample Power /SPSS Missing Value Analysis
- ▪데이터 분석 SPSS / Amos / Delta Graph
- ▪Data Mining Clementine / Answer Tree / Neural Connection
- ▪시계열, 예측 Decision Time & What If / SPSS Trends SPSS 초기 화면



데이터 입력 방법은 엑셀과 동일하다. SPSS 확장자를 보면 데이터는 *.sav, 출력 결과는 *.spo 가 사용된다.

4.3. 데이터 코딩

4.3.1. 데이터 행렬

통계학은 데이터에 관한 학문(Statistics is about data)이다. 데이터는 정보를 가진 숫자의 모임으로 통계학에서 데이터라 함은 관심이 있는 집단으로부터 (모집단: population) 추출한 표본(sample) 개체의 (예: 사람, 동물, 기업, 년도) 특성치에(변수, 예: IQ, 체중, 바이러스 수, 불량 여부, 도매 물가 지수) 대한 측정, 수집, 혹은 관측 값의 모임을 말한다. 데이터를 정리하거나 코딩 할 때는 열은 변수, 행은 개체(observation unit)로 하여 행렬의 형태로 하게 되는데 이를 데이터 행렬 (data matrix)이라 한다.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$ ➔ 데이터 행렬의 행은 개체, 열은 변수를 나타낸다.

다음은 임의로 선택한 3 명에 대해 성별, 키, 몸무게, 학력, 결혼 여부를 조사하여 정리 결과이다. 관측치 기호에 의하며 x_{11} = 남자, x_{21} = 180(cm) , …, x_{34} = no(미혼)에 해당한다.

<SPSS 데이터 예제>

	gender	height	weight	marriage
1	남자	180.00	80.00	yes
2	여자	165.00	52.00	no
3	남자	175.00	74.00	no

4.3.2. 변수(variable)의 종류

조사자가 관심을 갖는 집단의 특성에 대한 관측치를 변수라 하는데 설문 조사에서는 설문 항목이 하나의 변수가 된다. 변수 종류에 따라 적절한 데이터 분석 방법이 결정되므로 변수 종류를 아는 것은 적절한 분석 선택에 매우 중요하다. 지역 별 수능 성적의 차이가 있는지 알아 보고자 한다. 지역 변수는 분류형, 수능 성적은 측정형이고 지역이 수능에 영향을 미치는지 인과 관계에 (물론 지역별 차이로 보는 것이 더 적절한 표현이지만) 대한 분석이므로 분산 분석을 실시하면 된다. IQ 가 수능 성적에 영향을 미치는지 알아보려면 IQ 가 측정형이므로 회귀 분석을 실시하면 된다.

(1)Metric (측정형 변수, measurable)

실험 개체의 측정 가능한 특성을 측정하였거나 셀 수 있는 특성을 조사한 경우 이를 측정형 변수라 하며 키, 몸무게, 평점, IQ, 교통량, 사망자 수가 그 예이다. 설문 조사에서 측정형 변수는 주관식 문항 중 측정 가능한 것을 묻는 문항이다. 소득 수준, 용돈, 나이에 대해 개방형으로 물은 문항이 이에 해당된다. 일반적으로 측정형 변수는 3-4 범주로 분류하여 분류형 변수처럼 사용한다.

측정형 변수에 대한 분석도 평균, 표준 편차를 구하게 된다. 예제 설문 4.1.2 절에서 모든 리커트 척도 문항이 측정형 변수로 간주 된다. 또한 서스톤 척도, 거트만 척도도 측정형 변수로 간주된다.

(2)Non-metric (분류형 변수, classified, 범주형, categorical)

개체를 분류하기 위해 측정된 변수를 의미하며 성별, 결혼여부 등이 그 예이다. 설문 조사에서는 일반 선택 문항, 다중 선택 문항, 우선 순위 문항이 범주형 변수이다. 범주형 변수는 명목형 변수와 순서형 변수로 나눈다.

①명목형(nominal): 개체를 분류하는데 사용되는 변수 ➔ 성별, 결혼여부, 직업

②순서형(ordinal): 분류 범주가 순서를 가질 때 ➔ 성적(A, B, ..) 소득수준(상, 중, 하), 5 점 척도

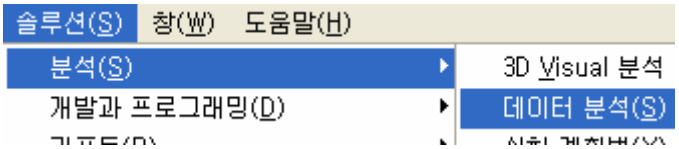
리커드 척도 문항은 순서형 변수로 사용할 수 있지만 각 범주를 수량화(quantify) 하여 측정형 변수처럼 사용한다. “①매우 불만족=1 점”, “②불만족=2 점”, “③보통=3 점”, “④만족=4 점”, “⑤매우 만족=5 점”으로 점수화 한다.

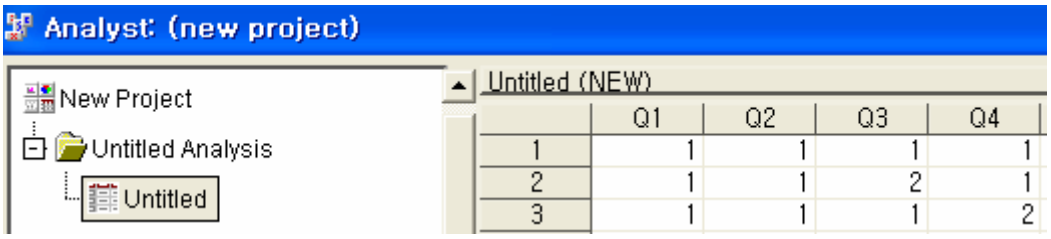
분류형 변수에 대한 분석으로는 각 범주 (보기)의 빈도나 비율을 계산한다. 예제 설문(4.1.2 절)에서 인구학적 문항(Q1~Q3), 우선 순위 문항(Q26_1~Q26_5), 다중 선택 문항 (Q29_1~Q29_3)이 분류형 변수에 해당한다.

4.3.3. SAS 에서 코딩

```
DATA SURVEY;
     INPUT Q1-Q25 Q26_1-Q26_5 Q27 Q28 Q29_1-Q29_3;
     DATALINES;
1 1 1 1 1 5 2 3 3 2 1 7 1 3 3 --(skipped)
1 1 2 1 1 1 1 1 3 1 . 3 3 3 2 --(skipped)
(skipped)
run;

PROC PRINT DATA=SURVEY;
RUN;
```

확장 편집기(프로그램 에디터)에서 직접 데이터를 입력할 수 있다. 그러나 응답자 수, 문항이 크므로 설문 조사 데이터에는 적절한 방법이 아니다.



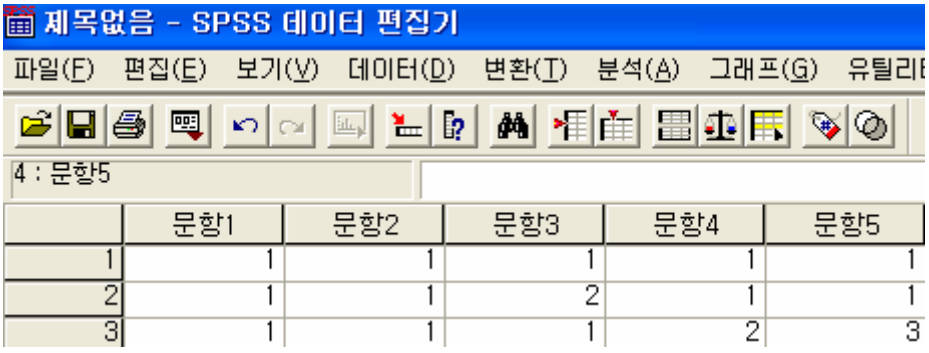


	Q1	Q2	Q3	Q4
1	1	1	1	1
2	1	1	2	1
3	1	1	1	2

엑셀과 같은 스프레드시트 입력 창이다. 관측치를 입력할 때마다 방향키를 사용해야 하므로 역시 좋은 방법이 아니다.

4.3.4. SPSS 에서 코딩

엑셀과 같은 스프레드시트 창에서 입력하면 된다.



	문항1	문항2	문항3	문항4	문항5
1	1	1	1	1	1
2	1	1	2	1	1
3	1	1	1	2	3

아무 조건 변화 없이 데이터를 입력하면 아래 창과 같다. 숫자 데이터는 소수점 2 자리로 입력되고 문자는 입력되지 않는다.

var00001	var00002	var00003	var00004	var00005
1.00	1.00	1.00	1.00	1.00

그러므로 창 아래 변수 보기에 가서 각 변수(열)의 속성을 바꾼 후 데이터를 입력해야 한다.

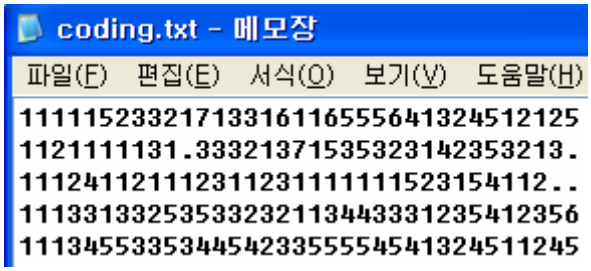
	이름	유형	자리수	소수점이하자리	설명
1	문항1	숫자	1	0	성별
2	문항2	숫자	1	0	재수여부
3	문항3	숫자	1	0	출신지역
4	문항4	숫자	1	0	건물
5	문항5	숫자	1	0	휴식공간

데이터 보기 변수 보기

4.3.5. 메모장에서 코딩

설문 데이터와 같이 개체(행: 응답자), 변수(열: 문항)가 많고 관측치가 한자리인 경우 스프레드시트 입력 방법은 시간이 많이 걸리고 (매번 방향키 사용해야 하므로) 입력 오류 발생 가능성이 높다. 그러므로 설문 조사 데이터 입력은 문항간 공백 없이 연속하여 입력하는 것이 편리하고 입력 시에는 키보드 오른쪽의 숫자 키 패드를 사용하면 편리하다.

반드시 한 행에는 한 사람의 응답 결과를 넣어야 하고 동일 문항의 데이터는 같은 열 위치에 맞도록 해야 한다. 다음은 메모장에서 예제 설문 데이터를 입력한 것의 일부이다.



입력이 끝나면 데이터 파일을 저장하면 된다. 이름을 coding.txtx(홈페이지에 올려져 있음. User id: “SUEVEY”, Password:”JAN2004”)라 하자.

(1)응답을 하지 않은 경우 (결측 문항)

응답하지 않은 문항에 대해서는 “.”을 입력한다. 0 과 결측치(.)는 다르다. 빈칸도 괜찮으나 .이 숫자 패드에 있어 훨씬 편리하다.

(2)문항 보기가 10 개 초과하는 경우

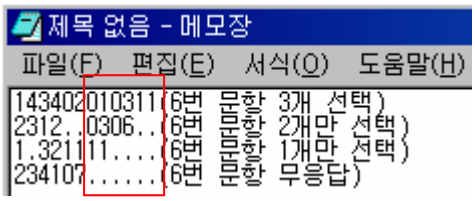
문항의 보기가 열 개인 경우는 10 번을 0 번으로 입력하면 그 문항은 한자리만 차지한다. 문항의 보기가 11 개 이상이면 동일 문항은 같은 열에 있어야 하므로 응답 결과는 반드시 두 자리를 입력해야 한다. 그러므로 1 번 응답은 01, 2 번 응답은 02, …로 입력한다. 두 자리를 입력하는 문항에서 결측치가 발생하면 “..”을 입력해야 한다.

(3)다중 선택 문항이 있는 경우

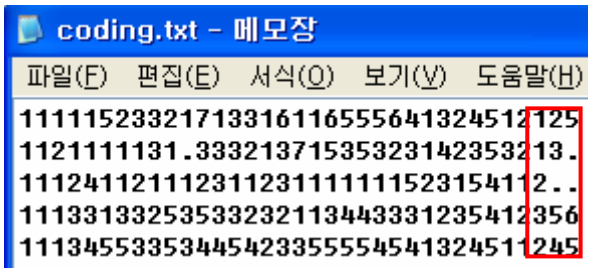
다중 선택 문항의 경우 응답자가 선택할 수 있는 최대 문항까지 자리 수를 잡아야 한다.

①응답 최대 개수를 언급하지 않은 경우: 코딩 전 조사된 설문지를 살펴 보았더니 그 문항에 대해 3 개까지 한 응답자가 있었다면 그 문항 입력을 위해서는 3 개 입력 공간이

있어야 한다. 다음은 보기가 12 개인 6 번 문항 응답 결과를 코딩한 예제이다. 첫번째 응답자는 6 번 문항에서 1 번, 3 번, 11 번을 택하였다.

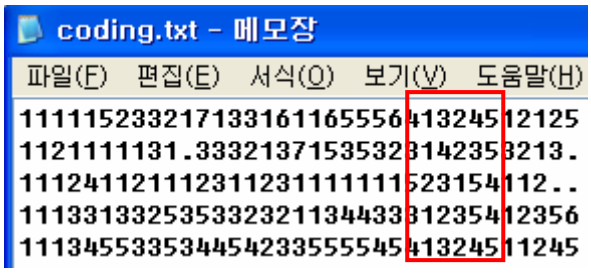


②**예제 설문 조사에서처럼 최대 개수를 지정한 경우:** 이 문항은 최대 3 개라고 명시하였고 보기가 6 개이므로 입력 공간이 3 열이면 된다. 첫번째 설문지 33 번 문항 응답 결과는 1, 2, 5 이고 두번째 설문지에서는 1 과 3 번 2 개만 선택되었다. 만약 4 개 선택한 응답자가 있다면 무응답으로 처리하면 된다. why? 적절한 3 개를 고를 방법이 없다.



(4)우선 순위 선택 문항

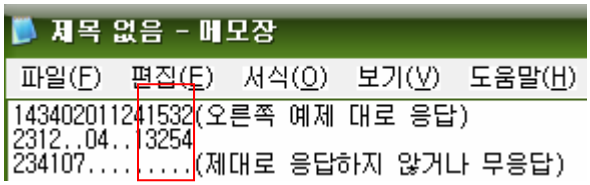
문항 평의 우선 순위 값을 차례로 입력하면 된다. 첫 번째 설문지의 응답 결과는 보기 차례로 1, 3, 2, 4, 5였다. (페이지 63)



우선 순위 문항이 다음과 같이 물어졌을 때 입력 방법을 살펴보자.

(문항 7)배우자를 선택할 때 고려되는 사항을 우선 순위로 부여하시오.

①건강 ②재력 ③직업 ④사랑 ⑤가족 사항

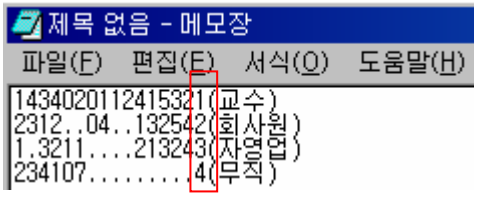


1 순위	4
2 순위	1
3 순위	5
4 순위	3
5 순위	2

(5)주관식 문항

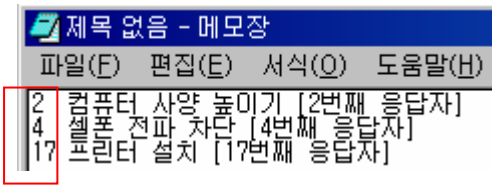
①분류가 가능한 경우

주관식 문항은 우선 몇 개의 범주로 분류한 후 각 범주에 번호를 부여하고 데이터 입력 시 함께 입력하면 된다. 예를 들어 귀하 아버지 직업은? _____ 분석자가 직업을 전문직(1), 비전문직(2), 자영업(3), 무직(4)으로 나누고 번호 부여 후 입력하면 된다. (아래 왼쪽)



②서술식 주관식

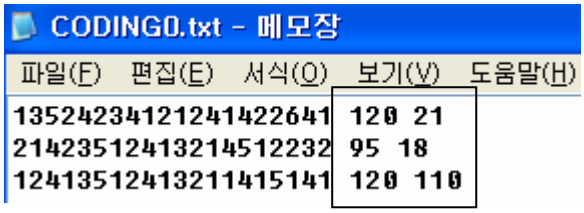
따로 입력하거나 정리하는 것이 편리하다. 만약 이 주관식 문항과 다른 문항과 관계 분석을 하는 경우라도 문항 번호, 응답 결과를 다른 파일에 저장하여 데이터를 합쳐 분석하는 것이 편리하다. 만약 다른 문항과 교차하여 분석하지 않을 것이면 설문지 번호를 (박스) 입력할 필요 없이 응답 내용만 정리하면 된다.



③숫자형 주관식

문항 번호에 관계없이 가장 뒤 부분에 입력한다. 마지막 열에 입력 하되 자릿수를 굳이 맞출 필요는 없다. 숫자형 주관식이 2 개 이상인 경우에는 각 문항마다 빈칸을 하나씩

입력한다. 다음은 IQ 와 용돈을 주관식으로 조사한 설문 자료 코딩 예이다. 첫 번째 응답자의 IQ=120, 용돈은 21(만원)인 경우이다.



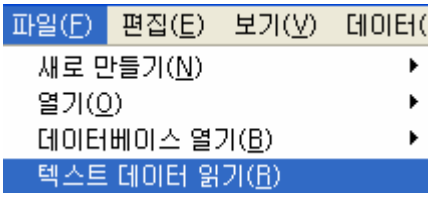
4.4. 입력 자료 불러 들이기

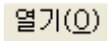
설문 조사 데이터를 메모장에서 입력한 경우(빈칸 없이) 통계 소프트웨어로 불러들이는 방법을 예제 자료를 (coding.txt) 이용하여 설명해 보자. 예제 자료를 C:\TEMP 폴더에 저장되어있다고 하자. 실습을 위하여 coding_error.txt 데이터에는 코딩 오류가 있으며 coding_a.txt 와 coding_b.txt 는 두 데이터로 분류한 것이다. coding_a.txt 는 A 집단으로부터 조사된 설문이고 coding_b.txt 는 B 집단으로부터 조사된 설문이라고 가정하자.

coding.txt	중국 경제학부 130명 응답 데이터, n=130
coding_error.txt	코딩 오류가 있는 응답 데이터, n=130
coding_a.txt	집단 A에서 설문 조사되었다고 가정된 응답 데이터, n=66
coding_b.txt	집단 A에서 설문 조사되었다고 가정된 응답 데이터, n=64

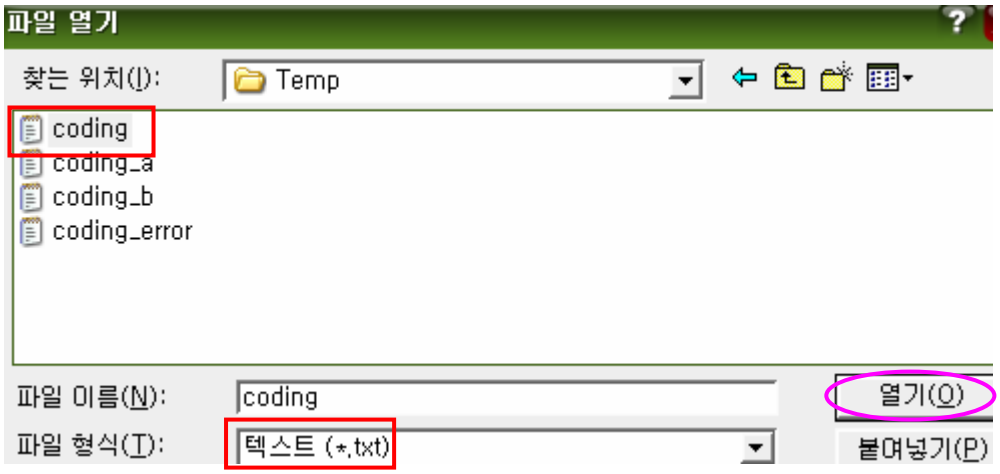
4.4.1. SPSS 에서 불러오기

(1)파일 메뉴에서 텍스트 데이터 (아스키 데이터) 읽기 메뉴를 선택하여 파일을 연다.

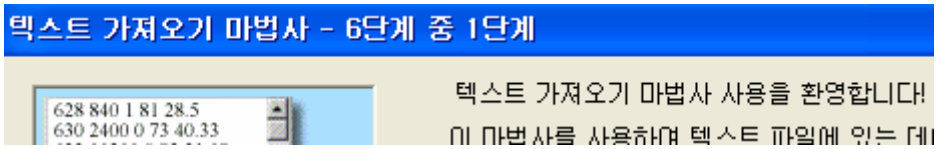


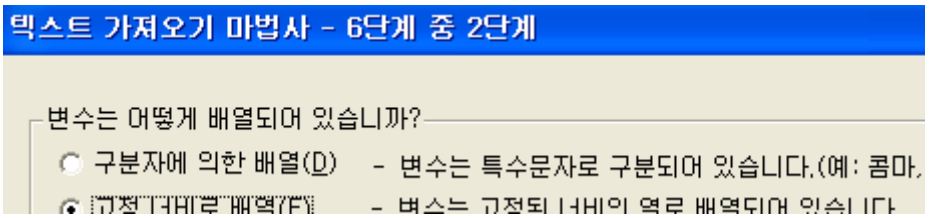
(2)텍스트 파일이 있는 폴더로 이동하여 데이터를 선택하고 

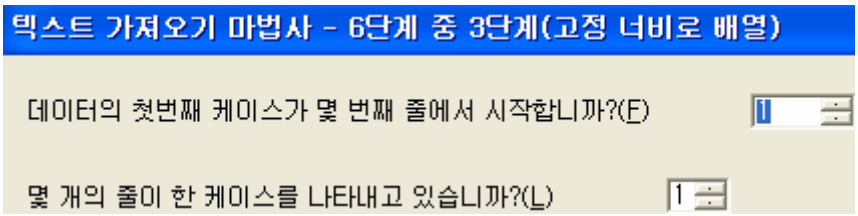
누른다.



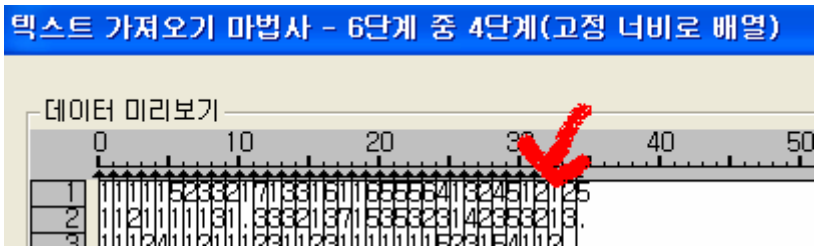
(3)아래와 같이 텍스트 가져오기 마법사를 설정한다. 1-3 단계까지는 아무 설정 없이 디폴트(default) 사용한다.



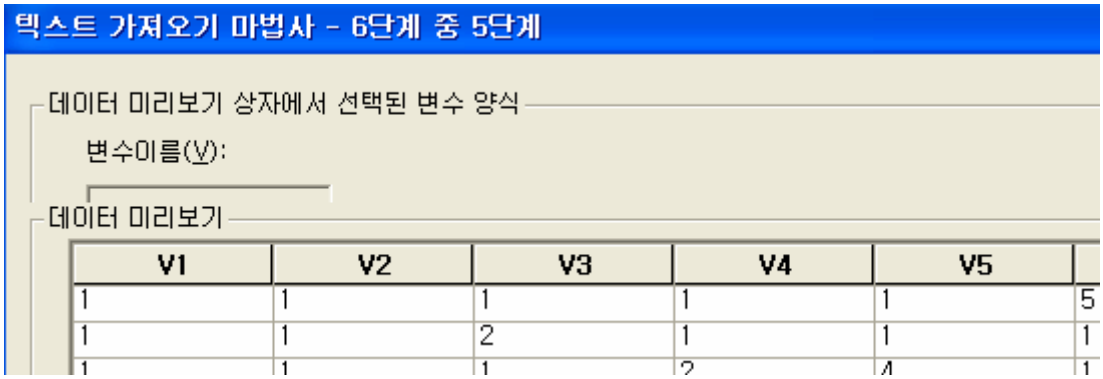


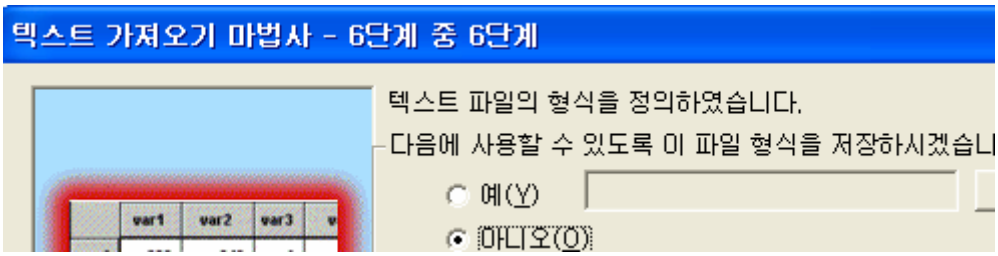


4 단계에서는 문항을 구분하는 구분선을 긋는다. 마우스를 원하는 위치에 놓고 누르면 화살표가 생긴다. 구분선은 데이터 열을 구분하는 역할을 한다. 만약 문항에 보기 수가 11 개 이상이라 두 칸(열)에 입력한 경우에는 두 열은 하나의 구분선으로 하면 된다. 화살표를 지우고 싶으면 화살표를 누른 후 마우스를 밖으로 이동하면 된다.



5 단계에서는 아래 데이터 미리 보기 창에서 데이터들이 제대로 읽혔는지 확인하고 문제가 없으면 다음(N) > 선택한다. 6 단계는 아무 설정 없이 디폴트로 내용을 그대로 사용한다.

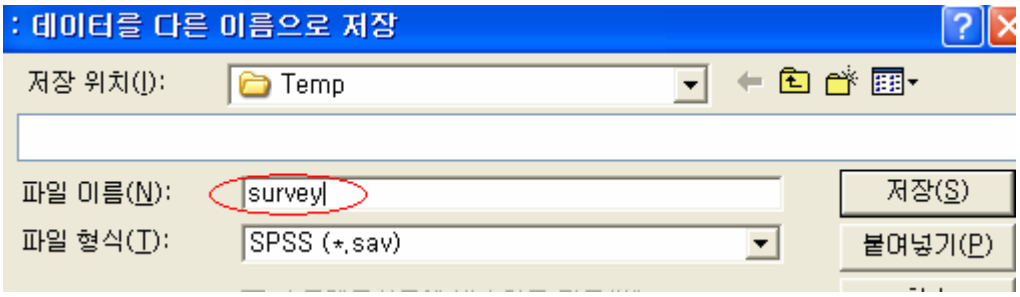




(4)다음은 성공적으로 읽어 오면 다음 화면이 나타난다.

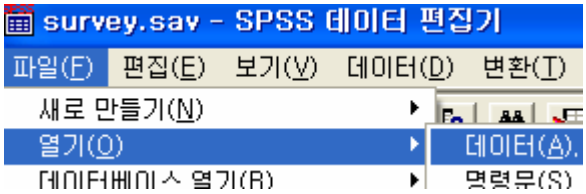
SPSS 데이터로 성공적으로 만들어졌는지 확인한다. 아래 변수 보기 옵션에 가서 변수명과 속성을 바꿀 수 있지만 그냥 v1, v2,... 사용해도 무방하다. 만약 SPSS 를 사용하려고 한다면 첫번째 설문지에 문항 번호를 매길 때 하나씩 배정한다. 즉 다중 선택 문항이나 우선 순위 문항이라도 Q26_1~ 이런 식이 아니라 각 문항에 하나의 번호를 배정한다. 그러므로 예제 설문(페이지 61-64)의 경우 SPSS 사용을 위해서는 v1-v35 로 배정하면 된다.

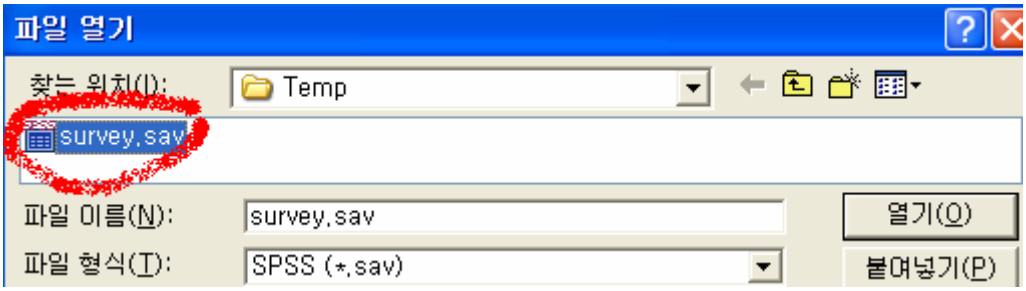
(5)SPSS 데이터 만들기에 성공하면 그 데이터를 다음에 다시 텍스트 데이터 가져오기로 실행하지 않으려면 일단 데이터를 SPSS 포맷으로 저장해야 한다. 저장 위치(I): 에서 저장할 폴더를 선택하고 파일 이름(N): 에 적절한 이름을 적고 저장(S) 누르면 된다.



C:\Temp 폴더 아래 survey.sav 라는 이름으로 저장된다. SPSS 데이터 확장자는 sav 이다.

(6)SPSS 종료 후 설문 데이터 분석이 필요한 경우 불러 오기를 이용해 저장해 둔 데이터를 불러 오면 된다. 메뉴에서 데이터 열기를 선택하고 불러오기 원하는 데이터를 선택하면 된다.





4.4.2. SAS 에서 불러오기

SAS 에는 외부 텍스트 파일을 불러오는 방법으로 파일(F) ➔ 데이터 가져오기(I)... 사용할 수 있으나 데이터를 빈칸 없이 입력하였으므로 이 방법보다는 확장 편집기에 다음과 같이 프로그램을 작성하여 데이터를 불러오는 것이 편리하다.

다음은 예제 설문 데이터를 SAS 데이터로 만드는 프로그램이다. (1.)의 의미는 한자리씩 읽어 들이라는 것이므로 보기가 11 개 이상이어서 한 문항에 두 자리를 배정한 경우에는 (2.)을 사용하면 된다.

확장 편집기 - 제목없음1 *

```
DATA SURVEY;
     INFILE 'C:\TEMP\CODING.TXT';
     INPUT (Q1-Q25) (1.) (Q26_1-Q26_5) (1.)
           (Q27-Q28) (1.) (Q29_1-Q29_3) (1.);
RUN;

PROC PRINT DATA=SURVEY;
RUN;
```

- SAS 는 자료를 만드는 DATA 단계와 만들어진 데이터를 이용하여 원하는 통계 분석 작업을 하는 Procedure 단계로 나누어져 있다. 각 단계는 RUN;에 의해 분리된다.
- INFILE: 텍스트 형식 데이터가 있는 폴더와 파일명을 지정하게 된다.
- INPUT: 변수명을 지정하는 공간이다.
- PROC PRINT 는 SAS 데이터를 출력하는 절차(procedure)이다.

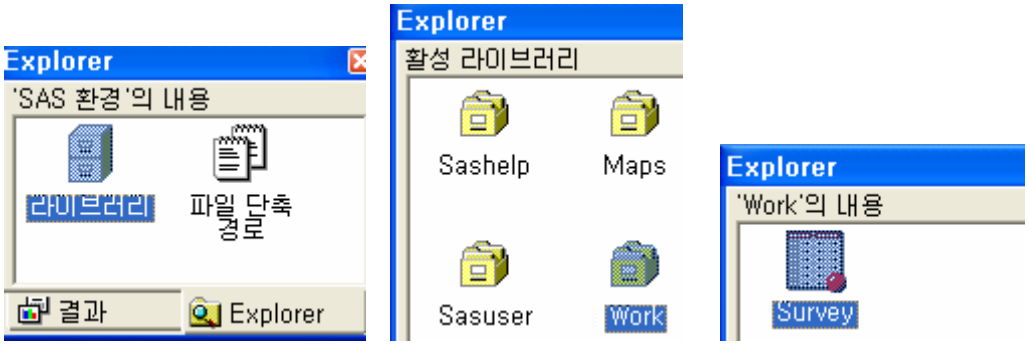
프로그램이 작성되면 F8-키(단축키)나 아이콘에서 을 눌러 프로그램을 실행한다. 프로그램 실행이 끝나면 로그 윈도우(프로그램 실행 과정 설명)를 살펴 오류 여부를 확인한다. 다음은 SURVEY 라는 SAS 데이터를 성공적으로 만들었으며 변수 35 개, 관측치 130 개가 있다는 것을 말해 주고 있다.

```
1    DATA SURVEY;
2        INFILE 'C:₩TEMP₩CODING.TXT';
3        INPUT (Q1-Q25) (1.) (Q26_1-Q26_5) (1.)
4              (Q27-Q28) (1.) (Q29_1-Q29_3) (1.);
5    RUN;

노트: The infile 'C:₩TEMP₩CODING.TXT' is:
      File Name=C:₩TEMP₩CODING.TXT,
      RECFM=V,LRECL=256

노트: 130 records were read from the infile 'C:₩TEMP₩CODING.TXT'.
      The minimum record length was 35.
      The maximum record length was 35.
노트: 데이터셋 'WORK.SURVEY'은(는) 130개 관측치, 35개 변수를 가지고 있습니다.
노트: DATA 문장 실행:
      실행 시간           1.00 초
      cpu 시간            0.06 초
```

로그 윈도우 확인 결과 오류가 없으면 출력 창에 가서 SAS 데이터가 성공적으로 만들어졌는지 확인한다. (변수명, 관측치 개수, 이상한 부분, 첫 번째 설문지와 Obs 1 을 대조하여 SAS 변수명과 첫 번째 설문지에 적힌 변수명이 일치하는지 확인한다.

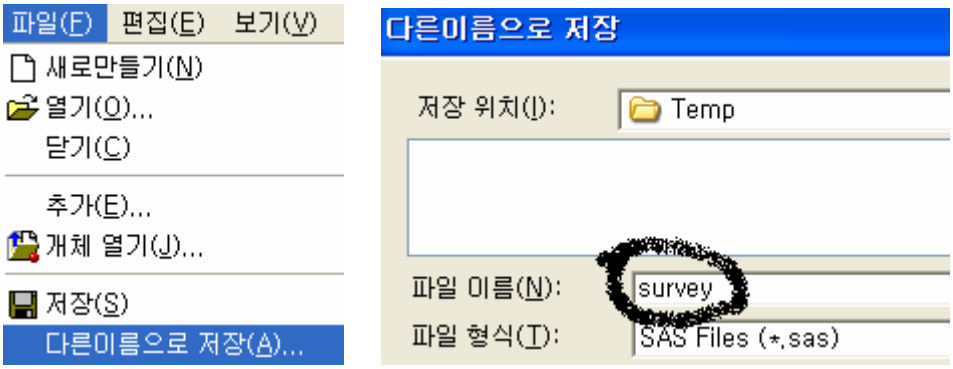
```
                                                                Q Q Q Q Q     Q Q Q
                                                                2 2 2 2 2     2 2 2
O                       Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q 6 6 6 6 6 Q Q 9 9 9
b  Q Q Q Q Q Q Q Q Q 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2           2 2
s  1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 1 2 3 4 5 7 8 1 2 3

 1 1 1 1 1 1 5 2 3 3 2 1 7 1 3 3 1 6 1 1 6 5 5 5 6 4 1 3 2 4 5 1 2 1 2 5
 2 1 1 2 1 1 1 1 1 3 1 . 3 3 3 2 1 3 7 1 5 3 5 3 2 3 1 4 2 3 5 5 3 2 1 3 .
 3 1 1 1 2 4 1 1 2 1 1 1 2 3 1 1 2 3 1 1 1 1 1 1 1 5 2 3 1 5 4 1 1 2 . .
 4 1 1 1 3 3 1 3 3 2 5 3 5 3 3 2 3 2 1 1 3 4 4 3 3 3 1 2 3 5 4 1 2 3 5 6
 5 1 1 1 3 4 5 5 3 3 5 3 4 4 5 4 2 3 3 5 5 5 5 4 5 4 1 3 2 4 5 1 1 2 4 5
```

SAS 데이터가 만들어지면 WORK 폴더에 SURVEY 라는 이름의 데이터가 생성되어 있다.



프로그램에서 DATA SURVEY 라고 되어 있는 것의 실제 이름은 WORK.SURVEY 이다. 이 SAS 데이터는 SAS 종료 시 사라져 버린다. 그럼 SAS 데이터를 저장해 놓아야 하나? 그럴

필요는 없다. 프로그램만 저장해 놓으면 된다. 먼저 확장 편집기를 선택한 후 (다른 윈도우가 선택되어 있으면 그 윈도우 내용이 저장된다) 다음 절차에 의해 프로그램을 저장한다.



프로그램은 C:\Temp 폴더 아래 SURVEY.sas 로 저장된다. SAS 프로그램 확장자는 sas 이다.

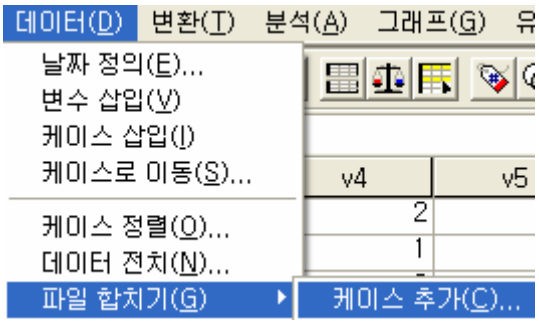
4.4.3. 두 개의 텍스트 데이터 합치기

설문 조사를 두 개의 집단으로 나누어 실시하고 각각 따로 저장하였다고 하자. CODING_A.TXT (집단 A), CODING_B.TXT (집단 B)로 텍스트 데이터를 저장하였다고 하자.

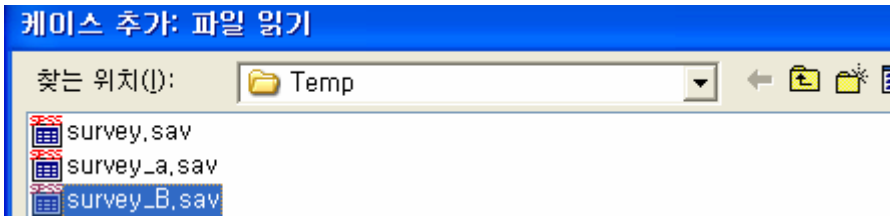
[SPSS 에서]

먼저 CODING_A.TXT 와 CODING_B.TXT 를 4.4.1 절의 방법대로 각각 SPSS 데이터로 만든다. SPSS 데이터 이름을 SURVEY_A, SURVEY_B 라고 하자.

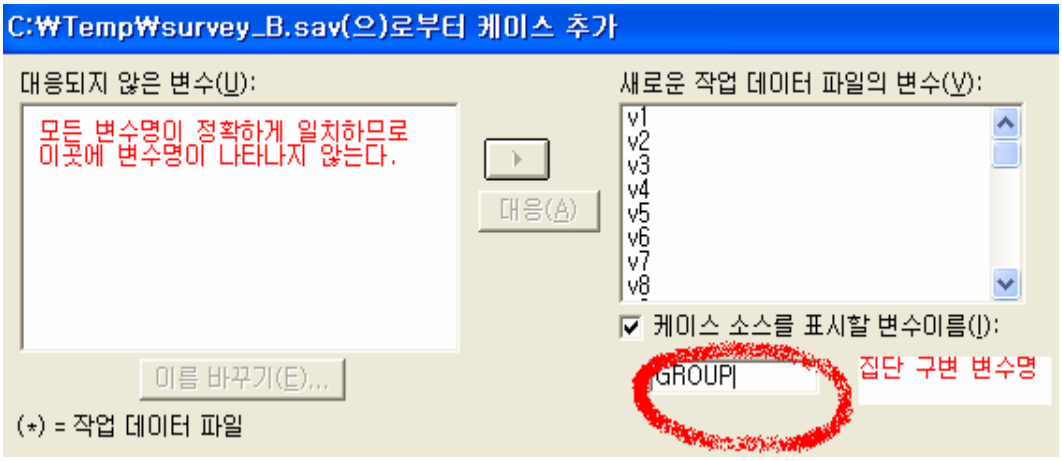
우선 SURVEY_A 를 불러온 후 데이터=>파일 합치기=>케이스 추가 메뉴를 선택한다. 케이스가 관측치가 된다.



뒤에 추가할 SPSS 데이터를 지정한다.



케이스 추가 창이 나타나면 다음과 같이 지정한다. 설문 자료의 경우에는 일치되지 않는 변수가 없으므로 케이스 소스를 표시할 변수이름(I): 부분을 체크하고 집단 구별하는 변수명을 적어 주면 된다. GROUP 변수 값은 0, 1 식으로 지정된다.



데이터가 제대로 읽혔는지 확인하고 데이터를 저장한다.

[SAS 에서]

두 집단을 GROUP 이라는 변수로 구별하였다. (집단 A: n=66, 집단 B: n=64)

```
DATA SURVEY1;
     INFILE 'C:\TEMP\CODING_A.TXT';
     GROUP='A';
     INPUT (Q1-Q25) (1.) (Q26_1-Q26_5) (1.)
           (Q27-Q28) (1.) (Q29_1-Q29_3) (1.);
RUN;

DATA SURVEY2;
     INFILE 'C:\TEMP\CODING_B.TXT';
     GROUP='B';
     INPUT (Q1-Q25) (1.) (Q26_1-Q26_5) (1.)
           (Q27-Q28) (1.) (Q29_1-Q29_3) (1.);
RUN;
```

```
DATA SURVEY;
    SET SURVEY1 SURVEY2;
RUN;

PROC PRINT DATA=SURVEY;
RUN;
```

SET 문은 두 개의 SAS 데이터를 아래로 합치는 명령문이다.

참고

숫자형 주관식 문항이 있는 경우(4.3.5 절의(5)) 이 문항을 마지막 열에 입력하는 것이 편리하다고 설명하였다. 이런 경우 SAS 에서 데이터를 읽는 프로그램은 다음과 같다.

```
DATA SURVEY0;
     INFILE 'C:\TEMP\CODINGO.TXT';
     INPUT (Q1-Q20) (1.) IQ MONEY;
RUN;
```

```
노트: 데이터셋 'WORK.SURVEY1'은(는) 66개 관측치, 36개 변수를 가지고 있습니다.
노트: DATA 문장 실행:
      실행 시간          2.12 초
      cpu 시간           0.18 초

노트: 데이터셋 'WORK.SURVEY2'은(는) 64개 관측치, 36개 변수를 가지고 있습니다.
노트: DATA 문장 실행:
      실행 시간          0.15 초
      cpu 시간           0.10 초
```

```
23   DATA SURVEY;
24       SET SURVEY1 SURVEY2;
25   RUN;

노트: 66개의 관측치를 데이터셋 WORK.SURVEY1. (으)로부터  읽었습니다.
노트: 64개의 관측치를 데이터셋 WORK.SURVEY2. (으)로부터  읽었습니다.
노트: 데이터셋 'WORK.SURVEY'은(는) 130개 관측치, 36개 변수를 가지고 있습니다.
노트: DATA 문장 실행:
      실행 시간           0.34 초
      cpu 시간            0.07 초
```

```
   G                                                       Q Q Q Q Q     Q Q Q
   R                                                       2 2 2 2 2     2 2 2
O  O                    Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q 6 6 6 6 6 Q Q 9 9 9
b  U Q Q Q Q Q Q Q Q Q 1 1 1 1 1 1 1 1 2 2 2 2 2 2           2 2
s  P 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 1 2 3 4 5 7 8 1 2 3

 1 A 1 1 1 1 1 5 2 3 3 2 1 7 1 3 3 1 6 1 1 6 5 5 5 6 4 1 3 2 4 5 1 2 1 2 5
 2 A 1 1 2 1 1 1 1 1 3 1 . 3 3 3 2 1 3 7 1 5 3 5 3 2 3 1 4 2 3 5 3 2 1 3 .
 3 A 1 1 1 2 4 1 1 2 1 1 1 2 3 1 1 2 3 1 1 1 1 1 1 1 5 2 3 1 5 4 1 1 2 . .
 4 A 1 1 1 3 3 1 3 3 2 5 3 5 3 3 2 3 2 1 1 3 4 4 3 3 3 1 2 3 5 4 1 2 3 5 6
```