

데이터분석가 실무자 교육 : SPSS 특강(1)

SPSS

- SPSS
 - Statistical Package for Social Sciences
 - 1968년 개발
 - 현재 IBM이 운영

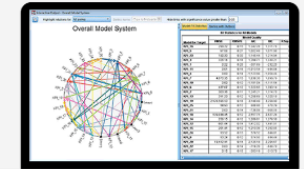
IBM SPSS Statistics

Propel research and analysis with a fast and powerful solution

IBM SPSS Statistics is the world's leading statistical software designed to solve business and research problems by means of ad hoc analysis, hypothesis testing, geospatial analysis and predictive analytics. Organizations use IBM SPSS Statistics to understand data, analyze trends, forecast and plan to validate assumptions and drive accurate conclusions.

→ [Learn more about IBM SPSS Statistics](#)

→ [Explore the student edition](#)



IBM SPSS Modeler

Gain insights quickly from all your data sources with powerful predictive analytics

IBM SPSS Modeler is a graphical data-science and predictive-analytics platform designed for users of all skill levels to deploy insights at-scale to improve their business. IBM SPSS Modeler supports the complete data-science cycle, from data understanding to deployment, with a wide range of algorithms and capabilities, such as text analytics, geospatial analysis and optimization.

→ [Learn more about IBM SPSS Modeler](#)

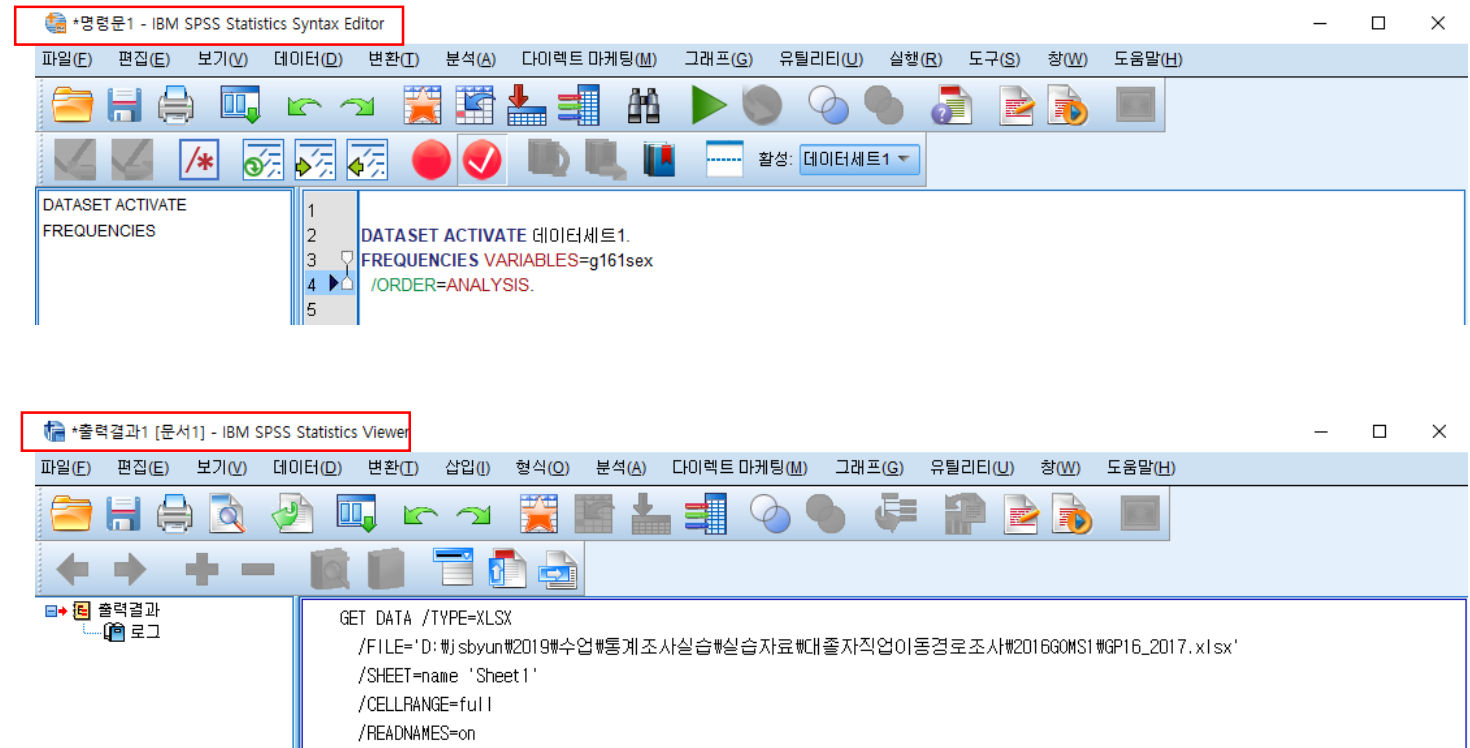
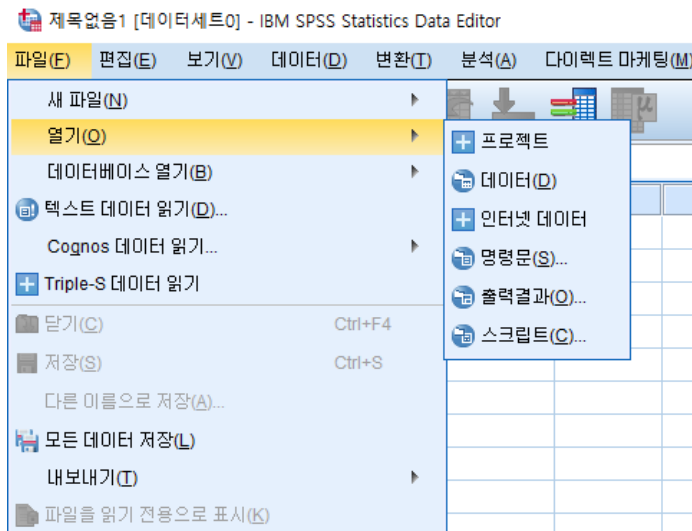


SPSS 분석 과정

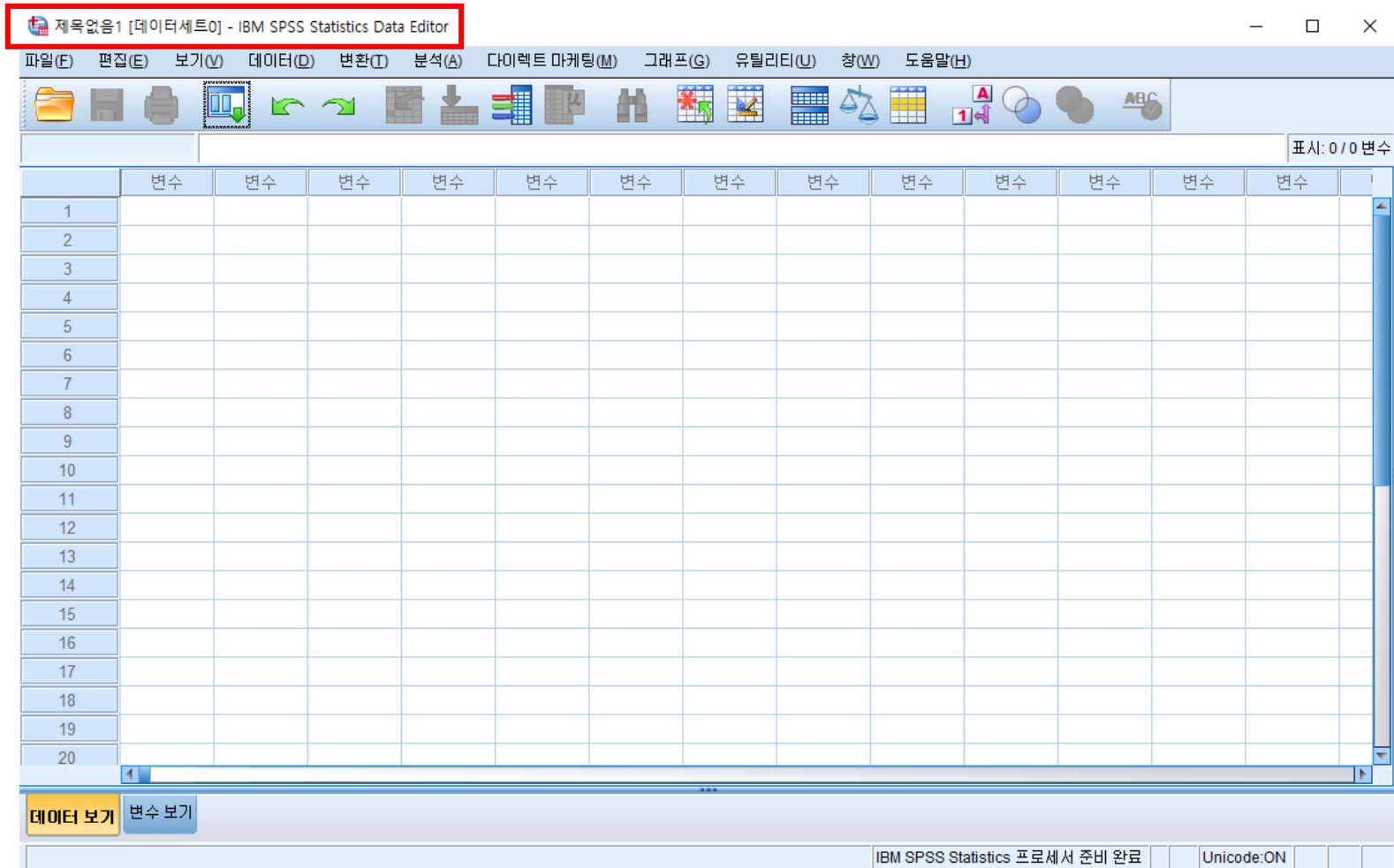
- Step 1 : 데이터 입력
- Step 2 : 분석 프로시저 선택
- Step 3 : 분석 변수 선정 및 옵션 선택
- Step 4 : 결과 검토
- Step 5 : 결과 분석 및 해석

SPSS 구조

• Windows of SPSS

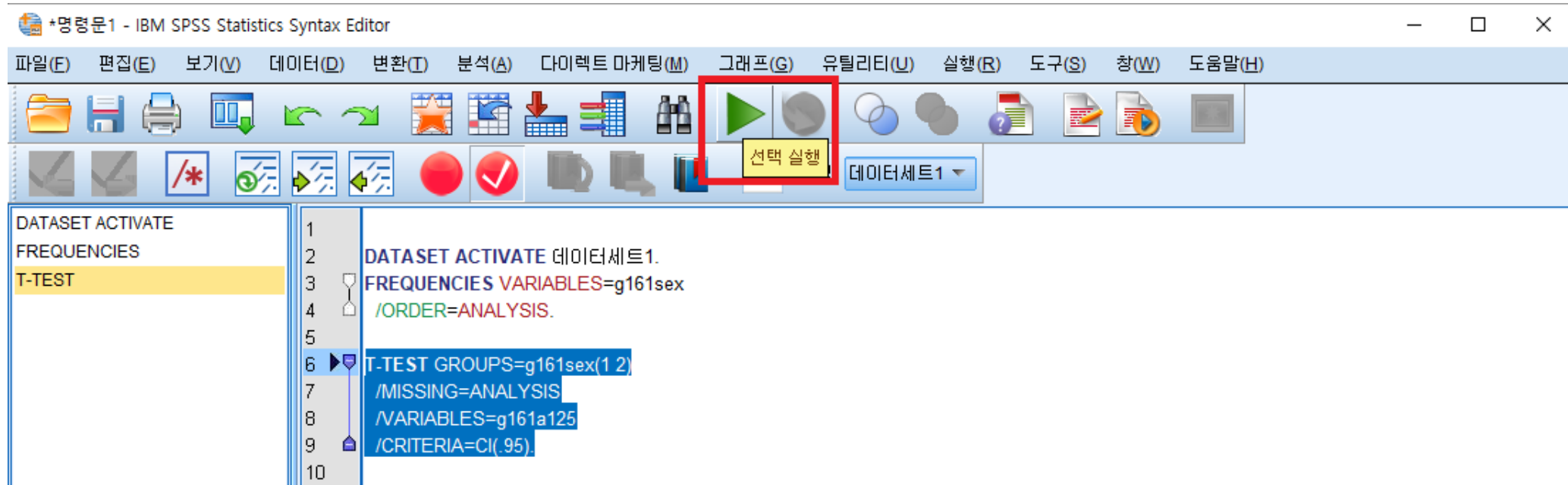


SPSS 데이터 window



SPSS 프로그램 window

- 프로그램(syntax) 작성과 실행



SPSS 출력 window

- 출력결과 내용
 - 프로그램
 - 분석프로시저
 - 분석 결과

IBM SPSS Statistics Viewer

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 삽입(I) 형식(O) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

2.0	8437	46.4	46.4	100.0
전체	18199	100.0	100.0	

T-TEST GROUPS=g161sex(1 2)
/MISSING=ANALYSIS
/VARIABLES=g161a125
/CRITERIA=CI(.95).

T 검정

집단통계량

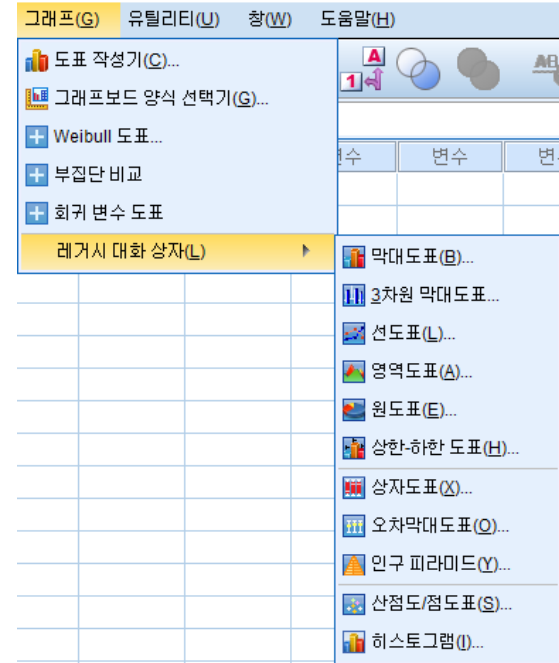
g161sex	N	평균	표준편차	평균의 표준오차
g161a125 1.0	7295	211.037	98.3248	1.1512
2.0	6065	169.201	73.3143	.9414

독립표본 검정

		Levene의 등분산 검정		평균의 동일성에 대한 T 검정						
		F	유의확률	t	자유도	유의확률 (양측)	평균차이	차이의 표준오차	차이의 95% 신뢰구간 하한	상한
g161a125	등분산을 가정함	233.583	.000	27.403	13358	.000	41.8363	1.5267	38.8437	44.9289
	등분산을 가정하지 않음			28.133	13207.086	.000	41.8363	1.4871	38.9214	44.9512

IBM SPSS Statistics 프로세서 준비 완료 | 케이스: 13,600 | Unicode:ON | H: 143, W: 909 pt.

SPSS 메뉴 : 주로 사용하는 메뉴 보기



SPSS 데이터 입력

- 데이터 입력
 - 간접 입력방식
 - TEXT : 메모장
 - Word : 입력방식(텍스트 방식 : .txt 혹은 .dat)
 - 엑셀
 - 기본 저장방식 : 엑셀 통합문서
 - SPSS 버전에 따라서는 엑셀 97-2003 통합문서 방식을 요구
 - 첫 행 : 변수명
 - 직접입력방식
 - SPSS
 - 다른 통계패키지 데이터
 - SAS

- SPSS에서 직접 입력하기

IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

4 : VAR00003 35.00 표시: 3 / 3 변수

	VAR00001	VAR00002	VAR00003	변수	변수	변수	변수	변수	변수	변수	변수	변수	변수
1	1.00 m		45.00										
2	2.00 f		43.00										
3	1.00 M		25.00										
4	3.00 f		35.00										
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON

• 변수 정보 입력 및 수정

IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도	역할
1	학년	숫자	8	0		없음	없음	8	오른쪽	명목형	입력
2	gender	문자	1	0		없음	없음	8	왼쪽	명목형	입력
3	VAR00003	숫자	8	2		없음	없음	8	오른쪽	척도	입력
4										척도	
5										순서형	
6										명목형	
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON

외부 입력 자료 불러오기

- spss, excel 등 가능

The screenshot displays the IBM SPSS Statistics Data Editor interface. The 'File' menu is open, showing options like 'New File', 'Open', 'Open Recent', 'Save', etc. The 'Open' option is highlighted, and its submenu is visible, showing 'Open', 'Open Recent', 'Open Recent', etc. The 'Open' option is highlighted in the submenu. The 'File' menu is also highlighted in the main menu bar.

The main window shows a data editor with columns labeled '변수' (Variable) and rows numbered 1 to 20. The status bar at the bottom indicates 'IBM SPSS Statistics 프로세서 준비' (IBM SPSS Statistics processor ready).

The 'File' menu options include:

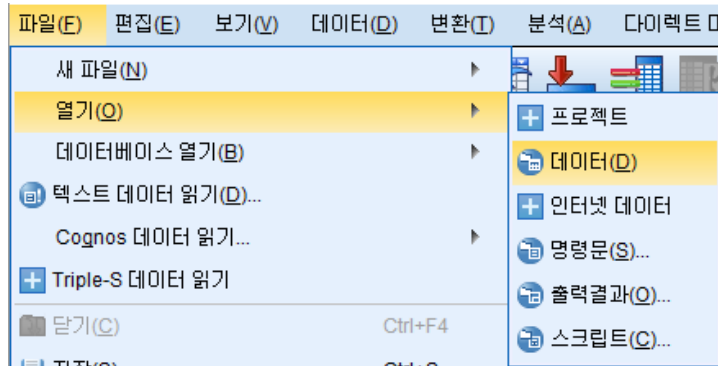
- 새 파일(N) (New File)
- 열기(O) (Open)
- 데이터베이스 열기(B) (Open Database)
- 텍스트 데이터 열기(O)... (Open Text Data...)
- Cognos 데이터 열기... (Open Cognos Data...)
- Triple-S 데이터 열기 (Open Triple-S Data)
- 닫기(C) (Close) - Ctrl+F4
- 저장(S) (Save) - Ctrl+S
- 다른 이름으로 저장(A)... (Save As...)
- 모든 데이터 저장(L) (Save All)
- 내보내기(T) (Export)
- 파일을 읽기 전용으로 표시(X) (Mark as Read-Only)
- 변수 정보 수집 (Collect Variable Information)
- 데이터 세트 이름 변경(M)... (Change Data Set Name...)
- 데이터 파일 정보 표시(I) (Show Data File Information)
- 데이터 세트 관리 (Manage Data Set)
- 데이터 캐쉬(H)... (Data Cache...)
- 프로세서 중단 (Stop Processor) - Ctrl+기간
- 뷰어 출력결과 옵션 설정(명령문)... (Set Viewer Output Options...)
- 서버 전환(W)... (Switch Server...)
- 리포지토리(R) (Repository)
- 인쇄 미리보기(V) (Print Preview)
- 인쇄(P)... (Print) - Ctrl+P
- 시작 대화 상자(W)... (Start Dialog Box...)
- 최근에 사용한 데이터(D) (Recently Used Data)
- 최근에 사용한 파일(F) (Recently Used File)
- 종료(X) (Quit)

The 'Open' submenu options include:

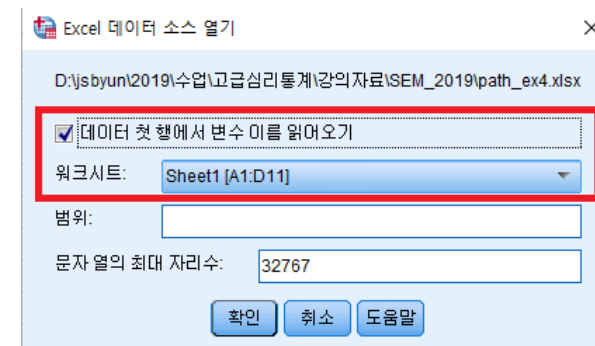
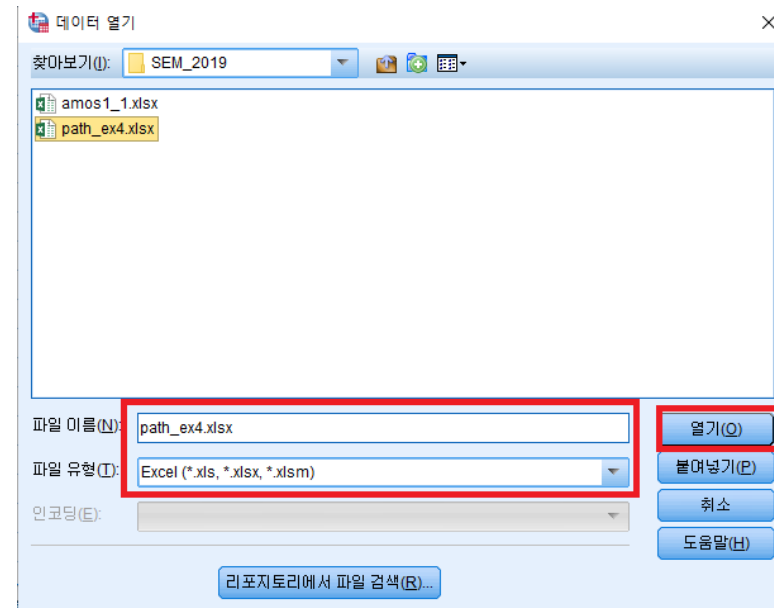
- 프로젝트 (Project)
- 데이터(D) (Data)
- 인터넷 데이터 (Internet Data)
- 명령문(S)... (Command...)
- 출력결과(O)... (Output...)
- 스크립트(C)... (Script...)

- 데이터 입력단계

- 데이터 입력 : 엑셀→SPSS
- 기본 메뉴 : 파일→열기→데이터

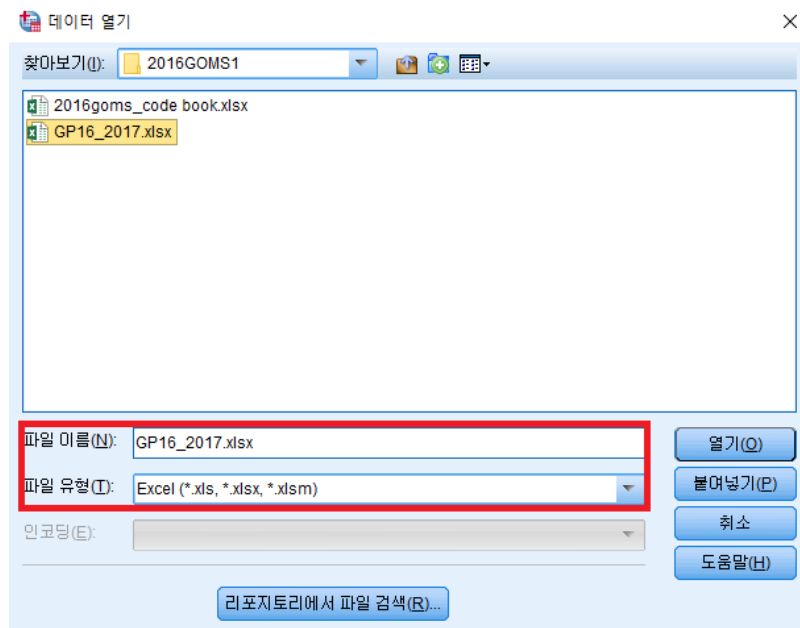


- SPSS dataset 저장 : 확장자 .sav

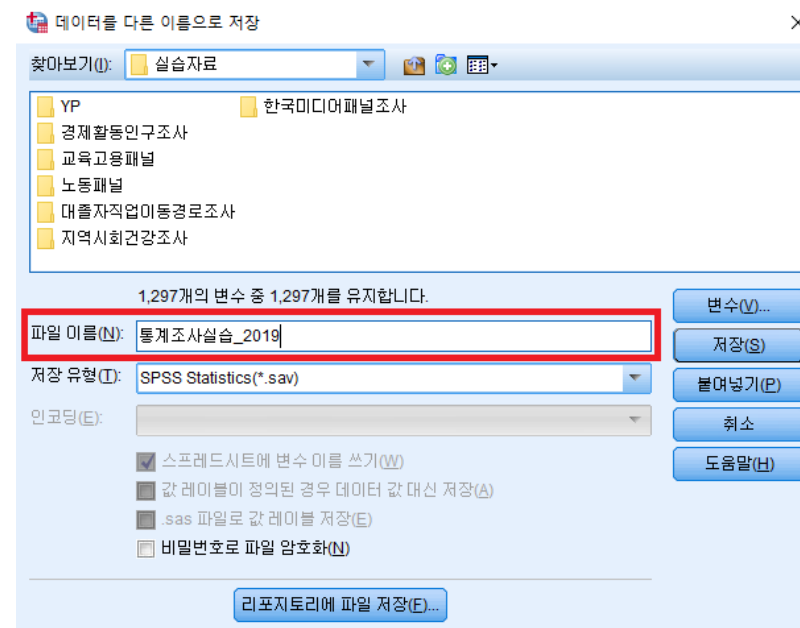


• 데이터 불러오기 예제

- 예제 : 대졸자 직업이동 경로조사(2016년) GP16_2017.xlsx
- 파일설계서 : 2016goms_code book.xlsx



저장



• 저장 결과 확인

통계조사실습_2019.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이어그램 마케팅(M) 그래픽(G) 유틸리티(U) 창(W) 도움말(H)

표시: 1297 / 1297 변수

	g161pid	g161branch	g161found	g161course	g161daynight	g161majorcat	g161dpmt	g161major	g161school
1	11.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100002	
2	31.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100002	
3	33.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100002	
4	39.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100002	
5	115.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100003	
6	125.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
7	130.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
8	141.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
9	161.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
10	177.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
11	192.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100005	
12	210.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100007	
13	213.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100007	
14	215.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100007	
15	217.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100007	
16	229.0	1.0	3.0	3.0	2.0	1.0	C010101	C01010100013	
17	237.0	1.0	3.0	3.0	2.0	1.0	C010101	C01010100013	
18	241.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100013	
19	243.0	1.0	3.0	3.0	1.0	1.0	C010101	C01010100013	
20	245.0	1.0	3.0	3.0	2.0	1.0	C010101	C01010100013	

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 케이스: 1,400 Unicode:ON

• 변수 확인 : 변수보기

통계조사실습_2019.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

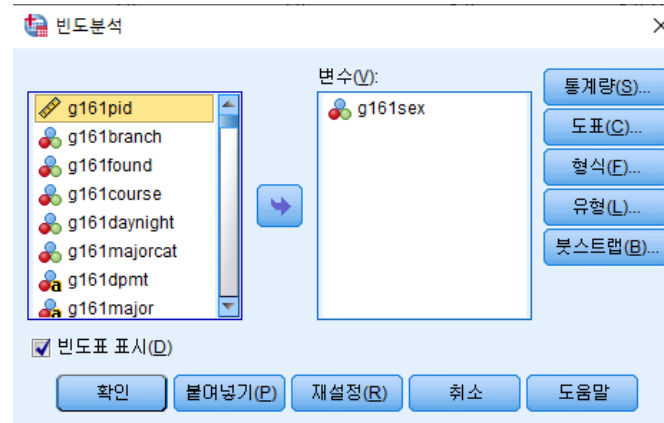
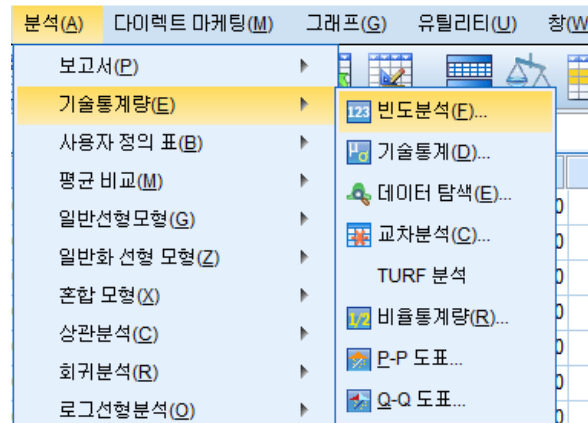
	이름	유형	너비	소수점미...	레이블	값	결측값	영	맞춤	측도	영향
1	g161pid	숫자	12	1		없음	없음	12	오른쪽	척도	입력
2	g161branch	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
3	g161found	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
4	g161course	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
5	g161daynight	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
6	g161majorcat	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
7	g161dpmt	문자	7	0		없음	없음	7	왼쪽	명목형	입력
8	g161major	문자	12	0		없음	없음	12	왼쪽	명목형	입력
9	g161school	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
10	g161province	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
11	g161area	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
12	g161sex	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
13	g161birthy	숫자	12	1		없음	없음	12	오른쪽	척도	입력
14	g161birthm	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
15	g161age	숫자	12	6		없음	없음	12	오른쪽	척도	입력
16	g161graduy	숫자	12	1		없음	없음	12	오른쪽	척도	입력
17	g161gradum	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
18	g161sq001	숫자	12	1		없음	없음	12	오른쪽	명목형	입력
19	g161sq002	숫자	12	1		없음	없음	12	오른쪽	척도	입력
20	g161sq003	숫자	12	1		없음	없음	12	오른쪽	척도	입력
21	g161sq004	숫자	12	1		없음	없음	12	오른쪽	명목형	입력

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 케이스: 1,400 Unicode:ON

SPSS 분석 기초

- 분석변수 : 성별

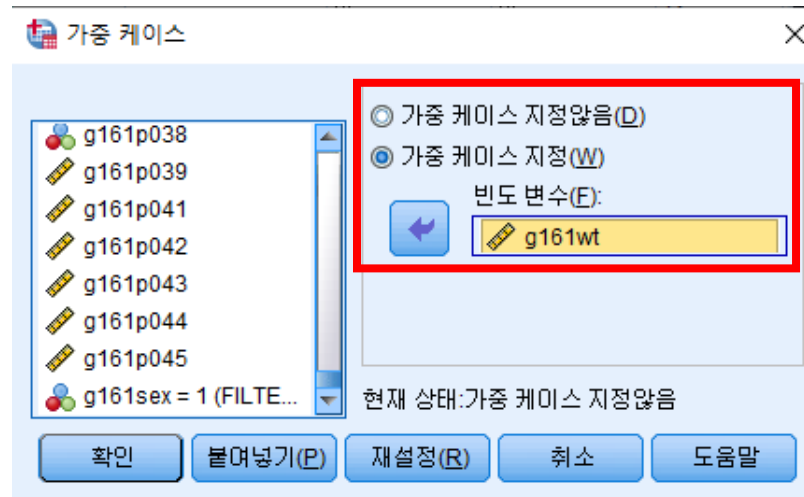
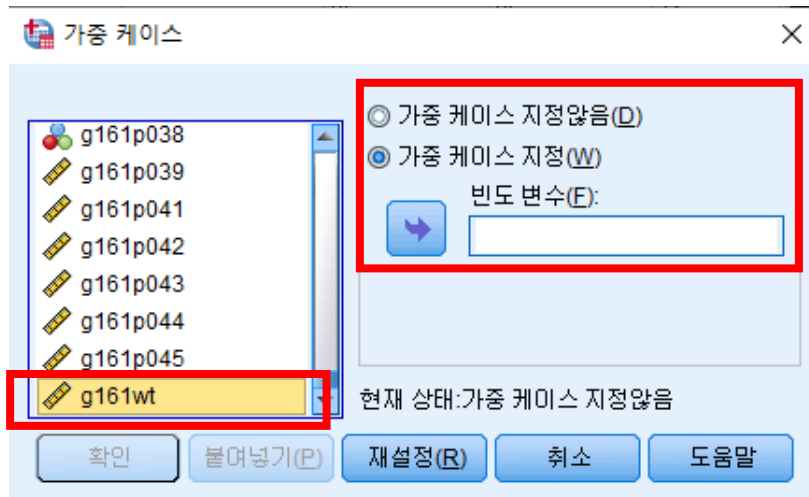


g161sex

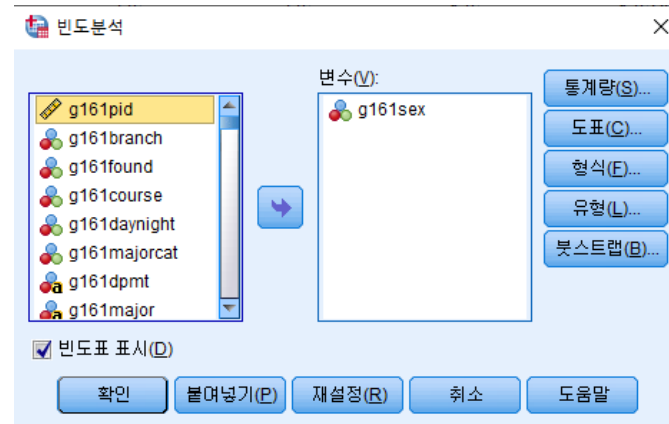
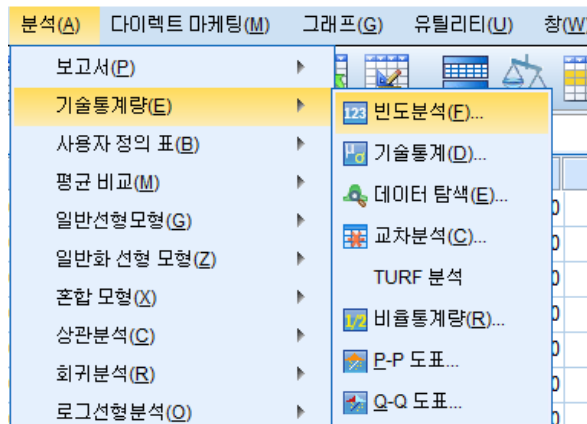
	빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효 1.0	9762	53.6	53.6	53.6
2.0	8437	46.4	46.4	100.0
전체	18199	100.0	100.0	

모수 추정 : 가중치를 부여한 분석

- 의미 : 표본단위의 가중값을 부여한 분석(모수 추정만 가능, 표준오차 추정은 편향 결과 제공)
- 가중값을 나타내는 변수 확인 : g161wt



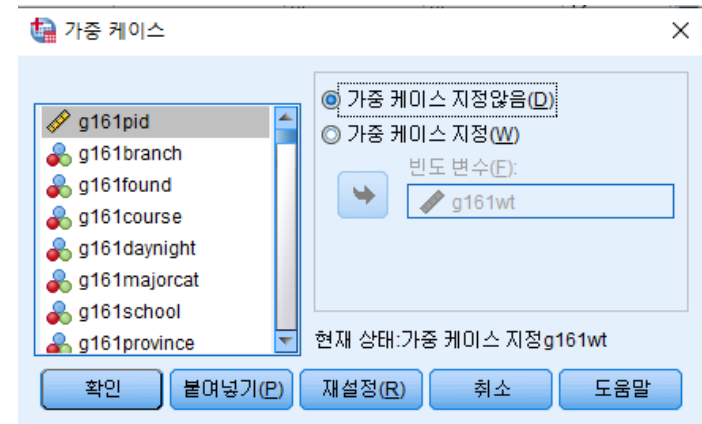
• 가중치를 부여한 분석



g161sex

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	1.0	250576	48.1	48.1	48.1
	2.0	269971	51.9	51.9	100.0
	전체	520547	100.0	100.0	

가중치 부여 해제

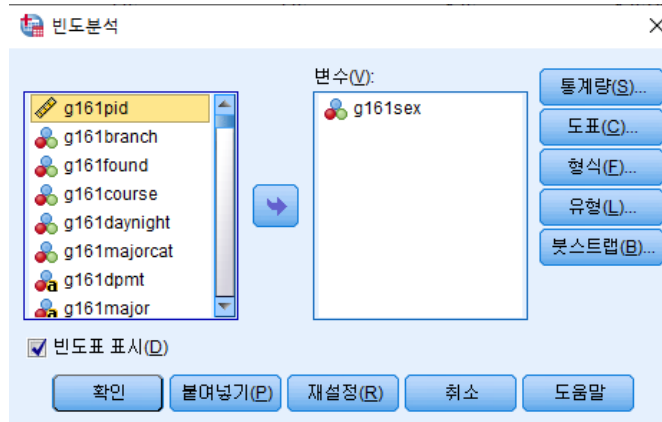
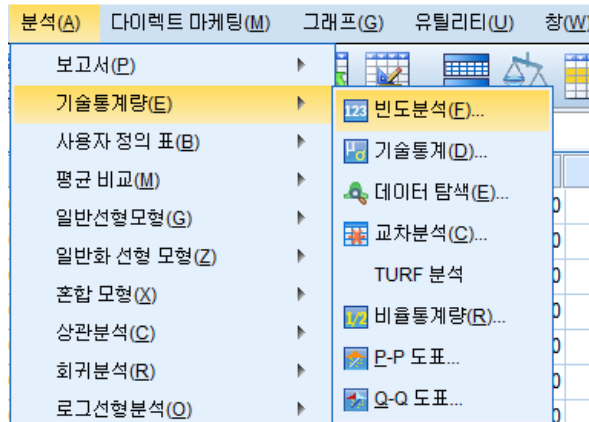


g161sex

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	1.0	9762	53.6	53.6	53.6
	2.0	8437	46.4	46.4	100.0
	전체	18199	100.0	100.0	

분석대상 선택

• 기본분석 : 성별



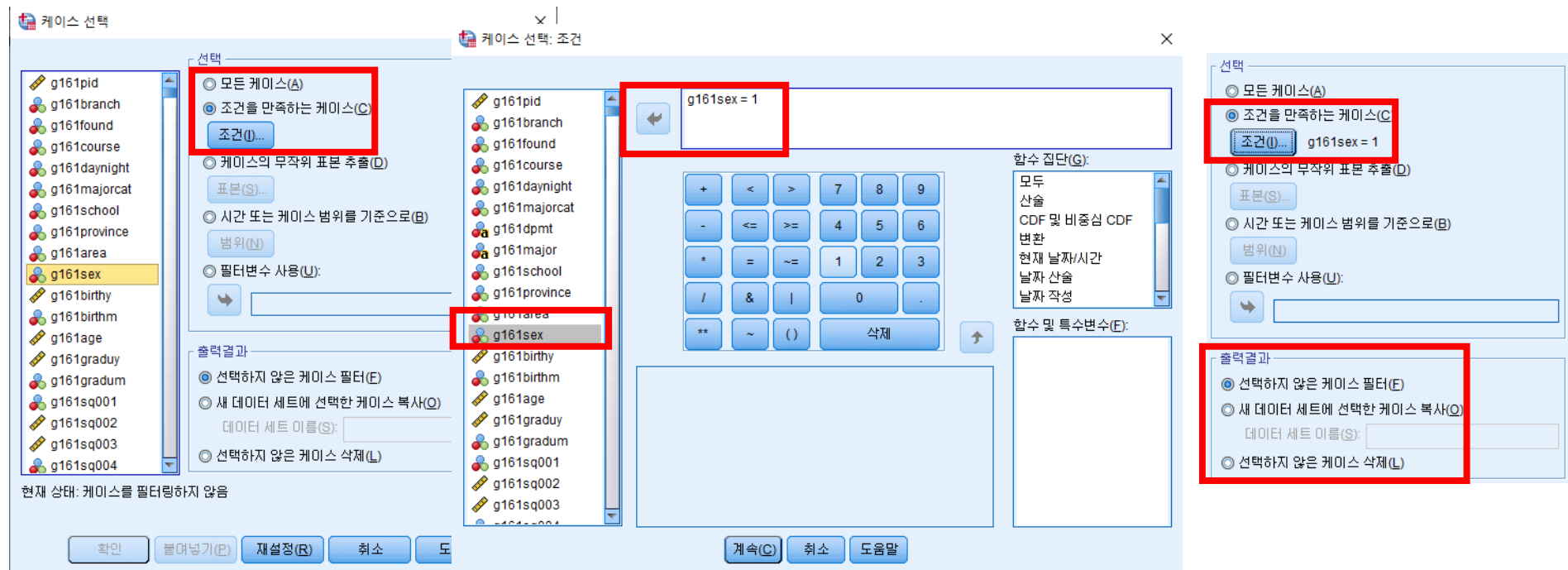
성별=1
선택하기

g161sex

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	1.0	9762	53.6	53.6	53.6
	2.0	8437	46.4	46.4	100.0
	전 체	18199	100.0	100.0	



• 성별=1



- 대상 선택한 데이터 및 분석 확인
- 성별=2는 선택하지 않음

*통계조사실습_2019.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(E) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

표시: 1298 / 1298 변수

	g161pid	g161branch	g161found	g161course	g161daynight	g161majorcat	g161dpmt	g161major	g161school
1	11.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100002		
2	31.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100002		
3	33.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100002		
4	39.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100002		
5	115.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100003		
6	125.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
7	130.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
8	141.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
9	161.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
10	177.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
11	192.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100005		
12	210.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100007		
13	213.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100007		
14	215.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100007		
15	217.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100007		
16	229.0	1.0	3.0	3.0	2.0	1.0 C010101	C01010100013		
17	237.0	1.0	3.0	3.0	2.0	1.0 C010101	C01010100013		
18	241.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100013		
19	243.0	1.0	3.0	3.0	1.0	1.0 C010101	C01010100013		
20	245.0	1.0	3.0	3.0	2.0	1.0 C010101	C01010100013		

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON 필터 설정

분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W)

보고서(P) 기술통계량(E) 사용자 정의 표(B) 평균 비교(M) 일반선형모형(G) 일반화 선형 모형(Z) 혼합 모형(X) 상관분석(C) 회귀분석(R) 로그선형분석(O)

123 빈도분석(E)... 기술통계(D)... 데이터 탐색(E)...

빈도분석

변수(V): g161pid g161branch g161found g161course g161daynight g161majorcat g161dpmt g161major

☒ 빈도표 표시(D)

통계량(S)... 도표(O)... 형식(F)... 유형(L)... 북스트랩(B)...

확인 붙여넣기(P) 재설정(R) 취소 도움말

g161sex

N	유효	9762
	결측	0

g161sex

	빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	1.0	9762	100.0	100.0

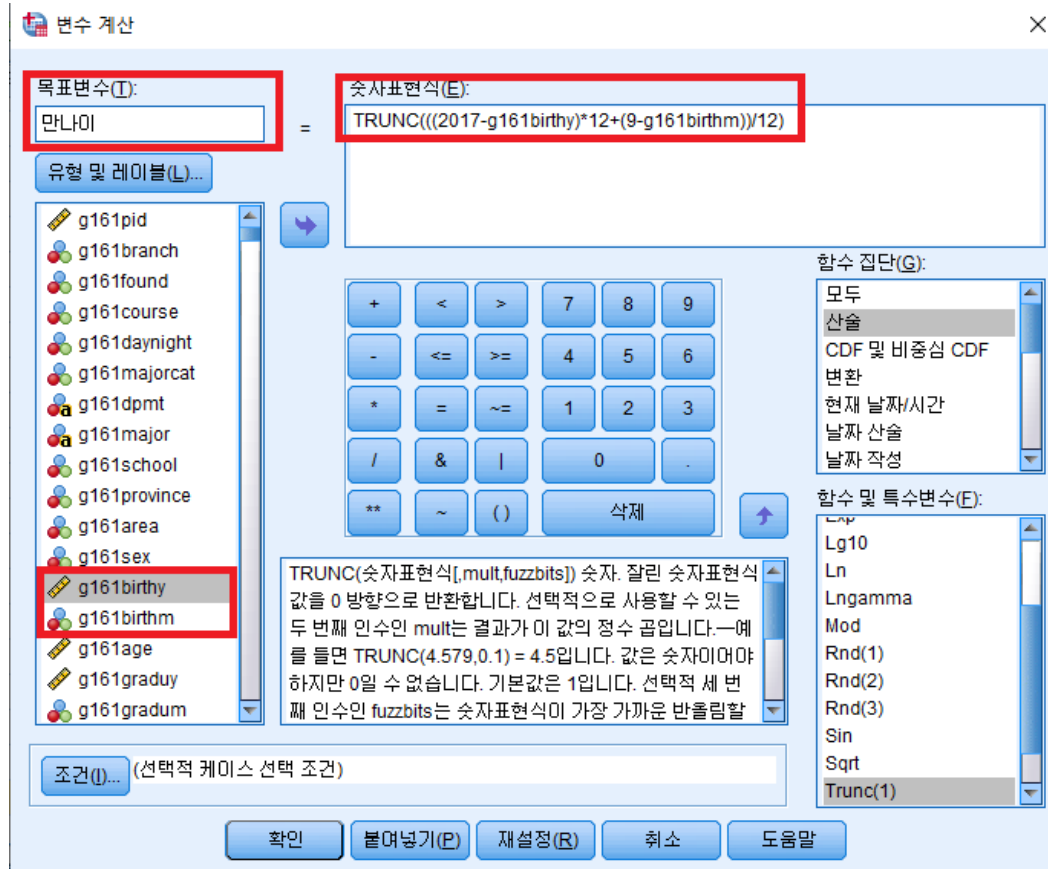
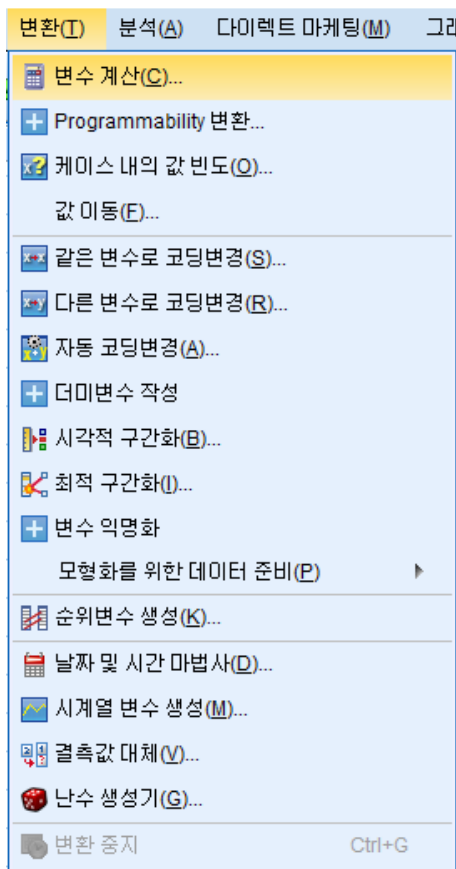
변수 계산

- 새로운 변수 계산

- 예: 만나이 계산

g161birthy	g161birthm
1991	9
1995	11

- 조사시점 : 2017년 9월 기준



- 여러 변수 계산 : syntax 이용

- 근속기간 계산 g161sq002 SQ1_1. 주된 활동 시작시점_년
 g161sq003 SQ1_1. 주된 활동 시작시점_월

- 만나이 계산

g161birthy	g161birthm
1991	9
1995	11

```
COMPUTE 만나이=TRUNC(((2017-g161birthy)*12+(9-g161birthm))/12).
COMPUTE 근속기간_m=(2017-g161sq002)*12+(9-g161sq003).
COMPUTE 근속기간_y=((2017-g161sq002)*12+(9-g161sq003))/12.
EXECUTE.
```

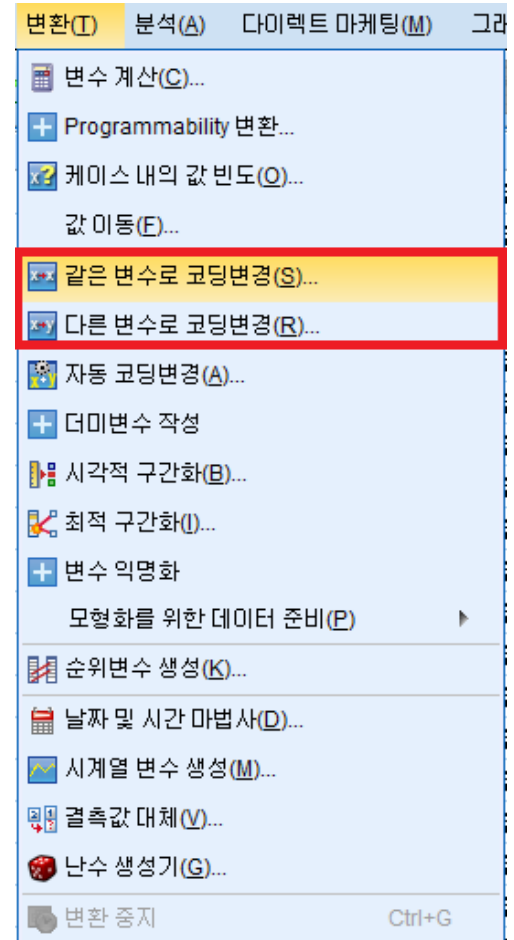

코딩변경 : 자료값 변경

- 의미 : 기존 자료 값을 다른 값 혹은 새로운 값을 부여하는 과정
- 방법
 - 다른 변수
 - 같은 변수
- 주의사항 : 동일 변수로 자료 값을 변경하면 기존 자료 값을 덮어버려 기존 자료값이 없어지므로 가능하면 다른 변수로 생성하는 것이 바람직함.
- 예1 : 성별의 현재 자료값 {1, 2}를 {1, 0}으로 변경

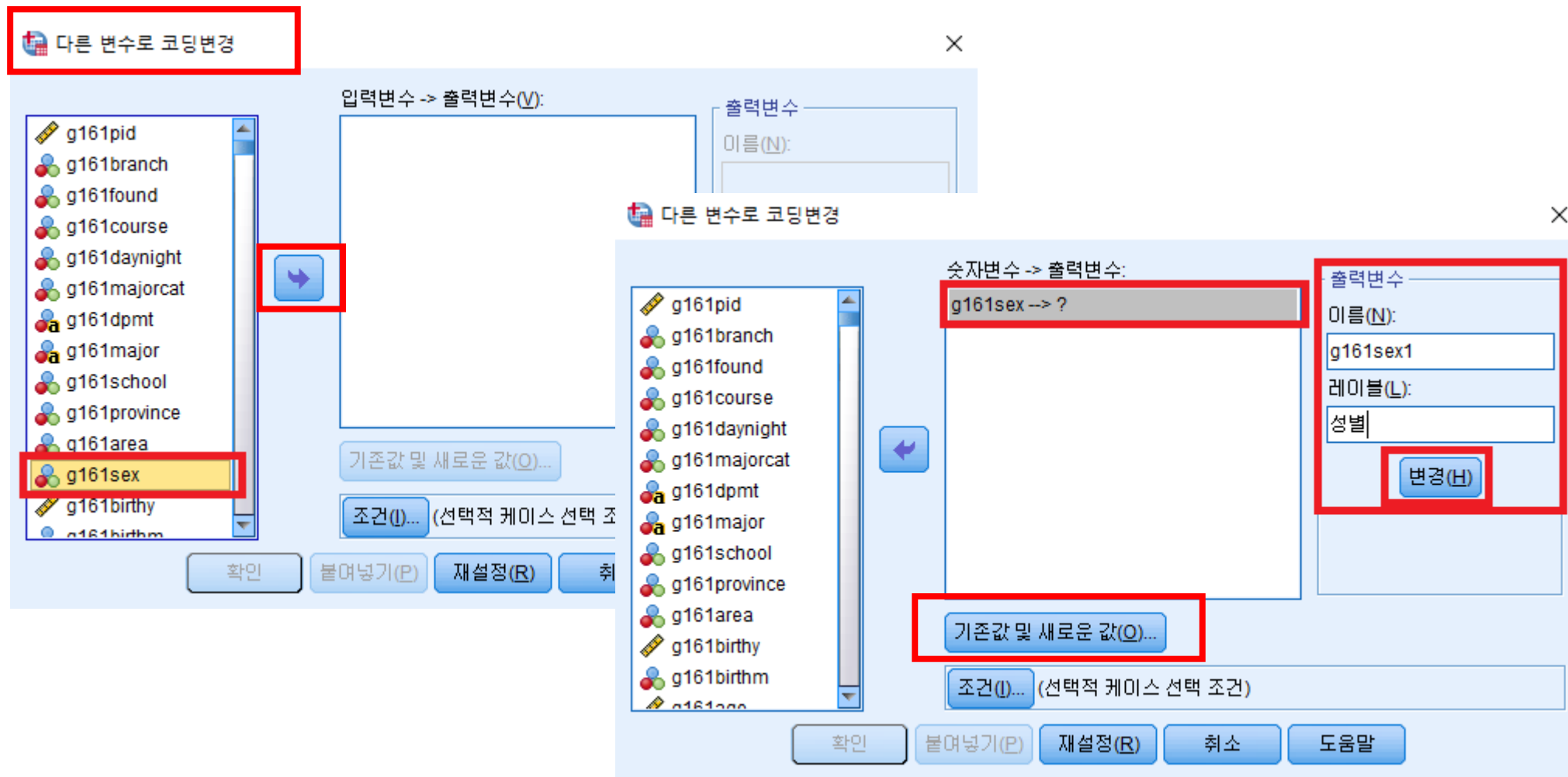
g161sex
1
2

- 예2 : 소득자료를 범주화하여 새로운 값을 부여하는 경우

g161a122	A31. <취업자 공통> 현 직장(일자리) 월평균 근로소득_만원
g161a125	A31_1. <취업자 공통> 현재 일자리 초임의 월평균 근로소득_만원



- 예1 : 다른 값으로 변경하는 경우



- 기존 값을 새로운 값으로 변경하는 과정

다른 변수로 코딩변경: 기존값 및 새로운 값

기존값

☒ 값(V):
1

☐ 시스템 결속값(S)

☐ 시스템 또는 사용자 결속값(U)

☐ 범위(N):
에서(I)

☐ 최저값에서 다음 값까지 범위(G):

☐ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☒ 기존값(L): 1

☐ 시스템 결속값(Y)

☐ 기존값 복사(P)

기존값 -> 새로운 값(D):
2 -> 0

☐ 출력변수가 문자열임(B) 너비(W):

☐ 숫자형 문자를 숫자로 변환('5' -> 5)(M)

다른 변수로 코딩변경: 기존값 및 새로운 값

기존값

☒ 값(V):

☐ 시스템 결속값(S)

☐ 시스템 또는 사용자 결속값(U)

☐ 범위(N):
에서(I)

☐ 최저값에서 다음 값까지 범위(G):

☐ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☒ 기존값(L):

☐ 시스템 결속값(Y)

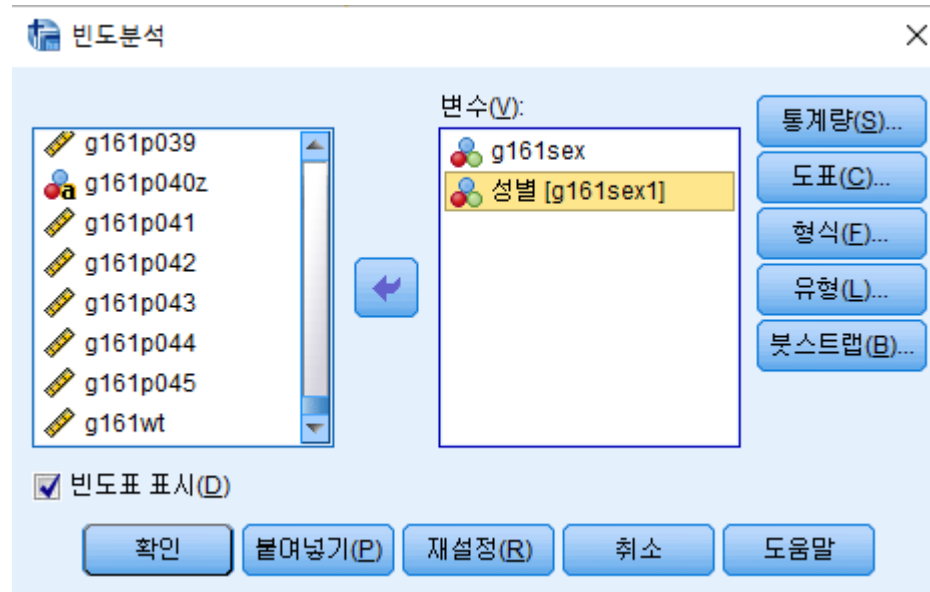
☐ 기존값 복사(P)

기존값 -> 새로운 값(D):
2 -> 0
1 -> 1

☐ 출력변수가 문자열임(B) 너비(W): 8

☐ 숫자형 문자를 숫자로 변환('5' -> 5)(M)

- 기존 값의 변경 결과 확인



g161sex

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	1.0	9762	53.6	53.6	53.6
	2.0	8437	46.4	46.4	100.0
	전체	18199	100.0	100.0	

성별

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	.00	8437	46.4	46.4	46.4
	1.00	9762	53.6	53.6	100.0
	전체	18199	100.0	100.0	

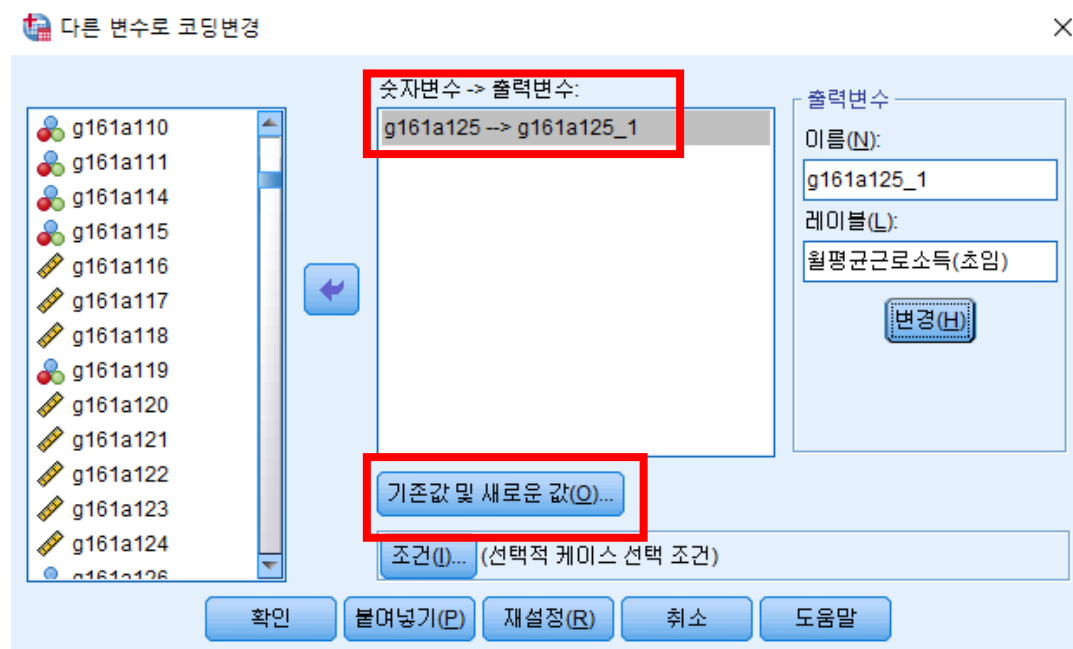
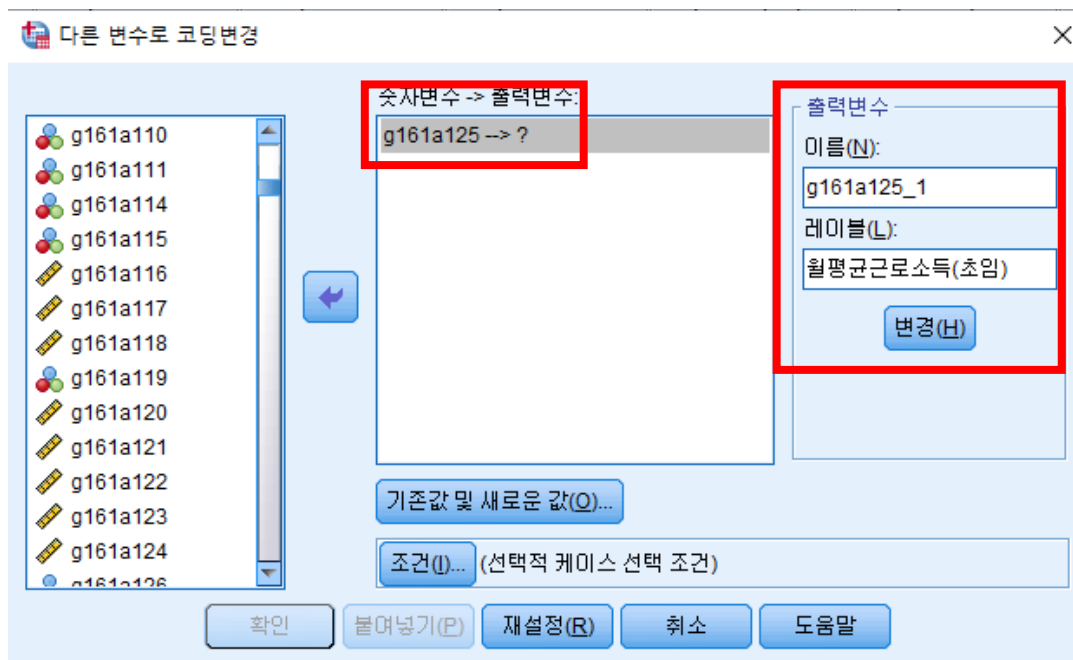
```
RECODE g161sex (2=0) (1=1) INTO g161sex1.
VARIABLE LABELS g161sex1 '성별'.
EXECUTE.
```

• 예2 : 소득자료를 범주화하는 경우

g161a122 A31. <취업자 공통> 현 직장(일자리) 월평균 근로소득_만원
 g161a125 A31_1. <취업자 공통> 현재 일자리 초임의 월평균 근로소득_만원

• 방법

- 분포를 고려한 범주화 : 소득자료의 분포 특성을 고려하여 범주화
- 관습적 범주화 : 일반적으로 사회에서 사용하는 범주화



• 소득자료의 범주화 값 부여 과정

다른 변수로 코딩변경: 기존값 및 새로운 값

기존값

☐ 값(V):

☒ 시스템 결측값(S)

☐ 시스템 또는 사용자 결측값(U)

☒ 범위(N):

에서(I)

☐ 최저값에서 다음 값까지 범위(G):

☐ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☐ 기존값(L):

☒ 시스템 결측값(Y)

☐ 기존값 복사(P)

기존값 --> 새로운 값(D):

Lowest thru 150 --> 1

151 thru 250 --> 2

251 thru 350 --> 3

351 thru 450 --> 4

451 thru Highest --> 5

추가(A)

변경(C)

제거(M)

☐ 출력변수가 문자열임(B) 너비(W): 8

☐ 숫자형 문자를 숫자로 변환('5'-'>5)(M)

계속(C) 취소 도움말

다른 변수로 코딩변경: 기존값 및 새로운 값

기존값

☐ 값(V):

☒ 시스템 결측값(S)

☐ 시스템 또는 사용자 결측값(U)

☐ 범위(N):

에서(I)

☐ 최저값에서 다음 값까지 범위(G):

☐ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☐ 기존값(L):

☒ 시스템 결측값(Y)

☐ 기존값 복사(P)

기존값 --> 새로운 값(D):

SYSMIS --> SYSMIS

Lowest thru 150 --> 1

151 thru 250 --> 2

251 thru 350 --> 3

351 thru 450 --> 4

451 thru Highest --> 5

추가(A)

변경(C)

제거(M)

☐ 출력변수가 문자열임(B) 너비(W): 8

☐ 숫자형 문자를 숫자로 변환('5'-'>5)(M)

계속(C) 취소 도움말

• 소득자료의 범주화 결과 확인(원자료 vs 추정)

원자료(표본)

기술통계량

	N	최소값	최대값	평균	표준편차	왜도	
	통계량	통계량	통계량	통계량	통계량	통계량	표준오차
g161a125	13360	-1.0	2000.0	192.045	90.2902	2.114	.021
유효 N(목록별)	13360						

월평균근로소득(초임)

	빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효 1.00	4391	24.1	32.9	32.9
2.00	6721	36.9	50.3	83.2
3.00	1760	9.7	13.2	96.3
4.00	328	1.8	2.5	98.8
5.00	160	.9	1.2	100.0
전체	13360	73.4	100.0	
결측 시스템	4839	26.6		
전체	18199	100.0		



추정

기술통계량

	N	최소값	최대값	평균	표준편차	왜도	
	통계량	통계량	통계량	통계량	통계량	통계량	표준오차
g161a125	388305	-1.0	2000.0	190.245	87.3374	1.806	.004
유효 N(목록별)	388305						

월평균근로소득(초임)

	빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효 1.00	130208	25.0	33.5	33.5
2.00	195415	37.5	50.3	83.9
3.00	49527	9.5	12.8	96.6
4.00	8879	1.7	2.3	98.9
5.00	4277	.8	1.1	100.0
전체	388305	74.6	100.0	
결측 시스템	132242	25.4		
전체	520547	100.0		

RECODE g161a125 (SYSMIS=SYSMIS) (Lowest thru 150=1) (151 thru 250=2) (251 thru 350=3) (351 thru 450=4) (451 thru Highest=5) INTO g161a125_1.
 VARIABLE LABELS g161a125_1 '월평균근로소득(초임)'.
 EXECUTE.

중복응답의 분석

- 중복 응답유형

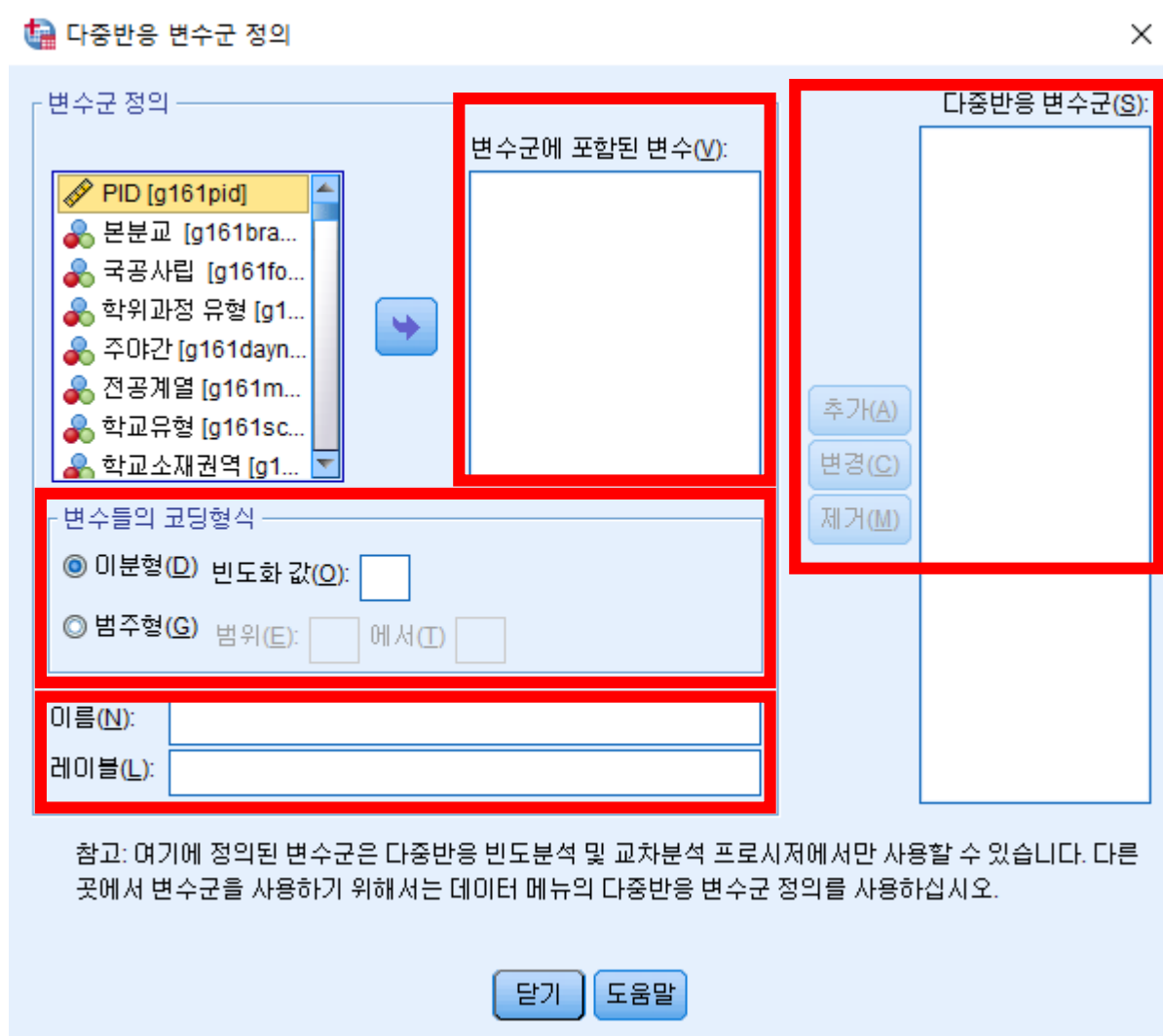
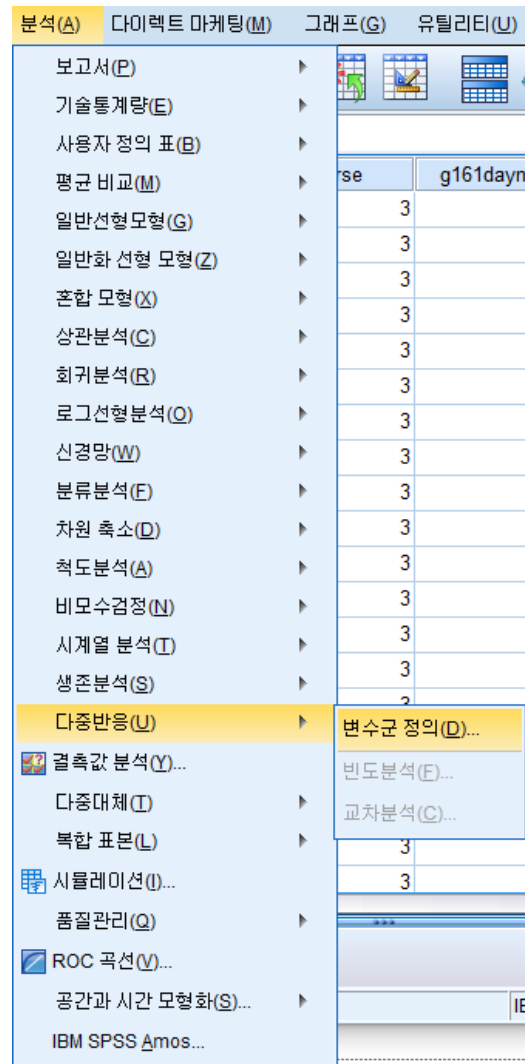
- 이분형 응답 : 예를 기준으로 통합 분석

g161d144 D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 1- 국민연금
g161d145 D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 2- 특수 직역 연금
g161d146 D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 3- 건강 보험
g161d147 D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 4- 고용보험
g161d148 D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 5- 산재 보험

- 중복응답 : 1순위분석, 통합 분석

g161d232 D42. <취업자 공통> 첫 일자리에 취업 시 중요 요소_1순위
g161d233 D42. <취업자 공통> 첫 일자리에 취업 시 중요 요소_2순위

중복 응답 분석



다중이분형 응답 분석

- 다중이분형 응답 선정 과정

다중반응 변수군 정의

변수군 정의

변수군에 포함된 변수(V):

다중반응 변수군(S):

추가(A) 변경(O) 제거(M)

변수이름 표시(D)
변수 레이블 표시(L)
문자순으로 정렬(A)
파일내 순서에 따라 정렬(F)
측정 수준별 정렬(M)
변수 정보(V)...

참고: 여기에 정의된 변수군은 다중반응 빈도분석 및 교차분석 프로시저에서만 사용할 수 있습니다. 다른 곳에서 변수군을 사용하기 위해서는 데이터 메뉴의 다중반응 변수군 정의를 사용하십시오.

닫기 도움말

다중반응 변수군 정의

변수군 정의

변수군에 포함된 변수(V):

다중반응 변수군(S):

추가(A) 변경(O) 제거(M)

변수들의 규정형식

이분형(D) 빈도화 값(O): 1

범수형(G) 범위(E): 에서(T)

이름(N): g161d37_t

레이블(L): 직장내 보형 가입여부

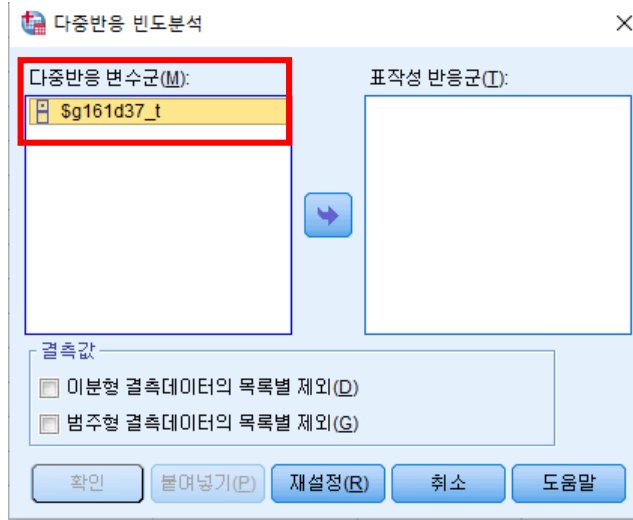
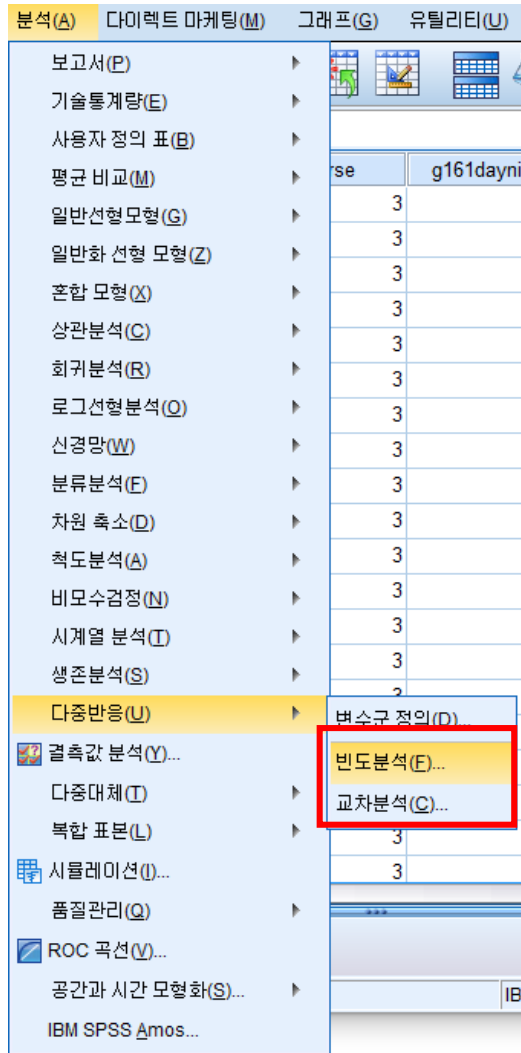
참고: 여기에 정의된 변수군은 다중반응 빈도분석 및 교차분석 프로시저에서만 사용할 수 있습니다. 다른 곳에서 변수군을 사용하기 위해서는 데이터 메뉴의 다중반응 변수군 정의를 사용하십시오.

닫기 도움말

다중반응 변수군(S):

\$g161d37_t

• 다중 이분형 분석



\$g161d37_t 빈도

		반응		케이스 중 %
		N	퍼센트	
직장내 보험 가입여부 ^a	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 1- 국민연금	4202	24.8%	94.0%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 2- 특수 지역 연금	60	0.4%	1.3%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 3- 건강 보험	4312	25.4%	96.5%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 4- 고용보험	4301	25.4%	96.2%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 5- 산재 보험	4074	24.0%	91.1%
전체		16949	100.0%	379.2%

a. 값 1을(를) 가지는 이분형 변수 집단입니다.

원자료(표본)

\$g161d37_t 빈도

		반응		케이스 중 %
		N	퍼센트	
직장내 보험 가입여부 ^a	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 1- 국민연금	127747	24.7%	93.7%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 2- 특수 지역 연금	2118	0.4%	1.6%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 3- 건강 보험	131568	25.5%	96.5%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 4- 고용보험	131431	25.5%	96.4%
	D37. <취업자 공통> 첫 직장(일자리) 직장 내 가입여부 5- 산재 보험	123522	23.9%	90.6%
전체		516385	100.0%	378.7%

a. 값 1을(를) 가지는 이분형 변수 집단입니다.

추정

중복형 응답 분석

다중반응 변수군 정의

변수군 정의

변수군에 포함된 변수(V):

다중반응 변수군(S): \$g161d37_t

추가(A) 변경(O) 제거(M)

변수들의 코딩형식

☐ 이분형(D) 빈도화 값(Q):

☒ 범주형(G) 범위(E): 에서(T)

이름(N):

레이블(L):

참고: 여기에 정의된 변수군은 다중반응 빈도분석 및 교차분석 프로시저에서만 사용할 수 있습니다. 다른 곳에서 변수군을 사용하기 위해서는 데이터 메뉴의 다중반응 변수군 정의를 사용하십시오.

닫기 도움말

다중반응 변수군 정의

변수군 정의

변수군에 포함된 변수(V):

다중반응 변수군(S): \$g161d37_t

추가(A) 변경(O) 제거(M)

변수들의 코딩형식

☐ 이분형(D) 빈도화 값(Q):

☒ 범주형(G) 범위(E): 1 에서(T) 99

이름(N): g161d43_t

레이블(L): 취업자공통순위

참고: 여기에 정의된 변수군은 다중반응 빈도분석 및 교차분석 프로시저에서만 사용할 수 있습니다. 다른 곳에서 변수군을 사용하기 위해서는 데이터 메뉴의 다중반응 변수군 정의를 사용하십시오.

닫기 도움말

• 중복응답 분석

원자료(표본)

추정

다중반응 빈도분석

다중반응 변수군(M):
☒ \$g161d43_t
☐ \$g161d37_t

표작성 반응군(I):

결측값
☐ 이분형 결측데이터의 목록별 제외(D)
☐ 범주형 결측데이터의 목록별 제외(G)

확인 붙여넣기(P) 재설정(R) 취소 도움말

\$g161d43_t 빈도

		반응		
		N	퍼센트	케이스 중 %
취업자공통순위 ^a	근로소득	2574	20.8%	41.7%
	근로시간	1927	15.6%	31.2%
	자신의 적성·흥미	1968	15.9%	31.9%
	전공분야와의 관련성	1414	11.4%	22.9%
	업무내용의 난이도	349	2.8%	5.6%
	업무량	270	2.2%	4.4%
	개인의 발전가능성	1090	8.8%	17.6%
	직업 자체의 미래 전망	449	3.6%	7.3%
	직장(고용) 안정성	484	3.9%	7.8%
	근무환경	563	4.6%	9.1%
	복리후생	204	1.7%	3.3%
	회사규모	156	1.3%	2.5%
	출퇴근거리	720	5.8%	11.7%
	일자리에 대한 사회적 평판	93	0.8%	1.5%
	하는 일에 대한 사회적 평판	28	0.2%	0.5%
	기타	62	0.5%	1.0%
전체		12351	100.0%	200.0%








a. 범주형 변수 집단



\$g161d43_t 빈도






		반응		
		N	퍼센트	케이스 중 %
취업자공통순위 ^a	근로소득	77216	20.9%	41.7%
	근로시간	57492	15.5%	31.1%
	자신의 적성·흥미	57824	15.6%	31.2%
	전공분야와의 관련성	41554	11.2%	22.5%
	업무내용의 난이도	10117	2.7%	5.5%
	업무량	7940	2.1%	4.3%
	개인의 발전가능성	31923	8.6%	17.3%
	직업 자체의 미래 전망	13625	3.7%	7.4%
	직장(고용) 안정성	14306	3.9%	7.7%
	근무환경	17408	4.7%	9.4%
	복리후생	6634	1.8%	3.6%
	회사규모	5075	1.4%	2.7%
	출퇴근거리	23476	6.3%	12.7%
	일자리에 대한 사회적 평판	2776	0.8%	1.5%
	하는 일에 대한 사회적 평판	785	0.2%	0.4%
	기타	1864	0.5%	1.0%
전체		370016	100.0%	200.0%

a. 범주형 변수 집단

기본분석

기술통계량(E)	▶	 빈도분석(F)...
사용자 정의 표(B)	▶	 기술통계(D)...
평균 비교(M)	▶	 데이터 탐색(E)...
일반선형모형(G)	▶	 교차분석(C)...
일반화 선형 모형(Z)	▶	TURF 분석
혼합 모형(X)	▶	 비율통계량(R)...
상관분석(C)	▶	 P-P 도표...
회귀분석(R)	▶	 Q-Q 도표...
로그선형분석(Q)	▶	

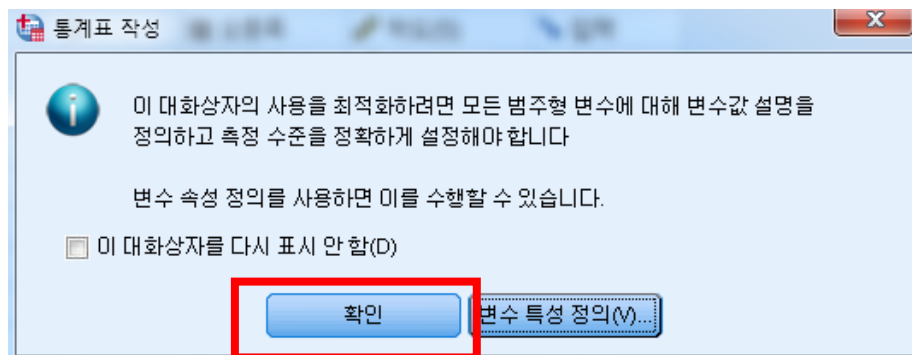
사용자 정의 표(B)	▶	 사용자 정의 표(C)...
평균 비교(M)	▶	 다중반응 변수군(S)...
일반선형모형(G)	▶	+ 다중 범주 세트 변환
일반화 선형 모형(Z)	▶	+ 범주순서 정의
혼합 모형(X)	▶	

평균 비교(M)	▶	 평균분석(M)...
일반선형모형(G)	▶	 일표본 T 검정(S)...
일반화 선형 모형(Z)	▶	 독립표본 T 검정...
혼합 모형(X)	▶	+ 요약 독립표본 T 검정
상관분석(C)	▶	 대응표본 T 검정(P)...
회귀분석(R)	▶	 일원배치 분산분석(Q)...
로그선형분석(Q)	▶	

SPSS : 표 작성

표 작성

- 표의 종류
 - 분할표 : 범주형의 2차원 이상 빈도표
 - 통계표 : 통계 요약값을 나타낸 표



- **분할표**

- 2차원 이상 범주형 자료의 범주별 빈도
- 행변수와 열변수 설정

[illegible]

• 행 합계 추가 : 행 변수 상황에서 가능

정의

N% 요약 통계량(S)...

범주 및 합계(C)...

요약 통계량

위치(I): 열(O) ☐ 감추기(H)

소스(U): 열 변수

범주 위치(E): 디폴트

범주별 합계

범주 및 합계

선택한 변수: SQ2) 성별

표시

값(U)

값	설명
1	남자
2	여자

제외(X):

보이기

☒ 전체(T)
설명(L): 전국

☒ 빈 범주(P)

☒ 데이터를 스캔했을 때 발견된 다른 값(F)

합계 및 소계 표시

☒ 적용할 상위 범주(V)

☐ 적용할 하위 범주(W)

소계 및 계산된 범주

소계 추가(S)... 범주 추가(C)... 편집(I)... 삭제(D)

모든 소계에서 범주가 생략되었습니다: 0

범주 정렬

기준(B): 변수값 순서(O): 오름차순

적용(A) 취소 도움말

전체 CHECK
설명 : 원하는 표현으로 수정

TOTAL 의 위치
- 상위 : 범주 이전에 표시

• 범주별 소계 추가

범주 및 합계

선택한 변수: SQ2) 성별

표시

값(U)

값	설명
2...1	소계
1	남자
2	여자

소계 및 계산된 범주

소계 추가(S)... 범주 추가(C)... 편집(I)... 삭제(D)

모든 소계에서 범주가 생략되었습니다: 0

범주 정렬

기준(B): 순서(O): 오름차순

제외(X):

보이기

전체(T)
설명(L): 전국

결측값(M)
☒ 빈 범주(P)
☒ 데이터를 스캔했을 때 발견된 다른 값(E)

합계 및 소계 표시

☒ 적용할 상위 범주(V)
☐ 적용할 하위 범주(W)

적용(A) 취소 도움말

• 열 변수의 합계와 요약 통계 수정

- 요약 통계량 추가 및 설명 수정
- 합계 추가

통계표 작성

표 제목 검정 통계량 옵션

변수(V):

보통(A) 최소(M) 레이어(L)

		문3) 이번 대통령 선거에 출마한 다음 후보 중 누구...	더불어 ...	자유한국당 ...	국민의당 ...	바른정 ...
		빈도	빈도	빈도	빈도	빈도
SQ2) 성별	전국	nnnn	nnnn	nnnn	nnnn	nnnn
	남자	nnnn	nnnn	nnnn	nnnn	nnnn
	여자	nnnn	nnnn	nnnn	nnnn	nnnn
연령	20대...	nnnn	nnnn	nnnn	nnnn	nnnn
	30 대	nnnn	nnnn	nnnn	nnnn	nnnn
	40 대	nnnn	nnnn	nnnn	nnnn	nnnn
	50 대	nnnn	nnnn	nnnn	nnnn	nnnn
	60 세 이 상	nnnn	nnnn	nnnn	nnnn	nnnn
권역	서울/인천/경...	nnnn	nnnn	nnnn	nnnn	nnnn
	대전 / 충청	nnnn	nnnn	nnnn	nnnn	nnnn
	광주 / 전라	nnnn	nnnn	nnnn	nnnn	nnnn
	대구 / 경북	nnnn	nnnn	nnnn	nnnn	nnnn
	부산/울산/경...	nnnn	nnnn	nnnn	nnnn	nnnn
	강원 / 제주	nnnn	nnnn	nnnn	nnnn	nnnn

정의

요약 통계량

N% 요약 통계량(S)...

범주 및 합계(C)...

위치(I): 열(O)

소스(U): 열 변수

감추기(H)

범주 위치(E): 디폴트

확인 붙여넣기(P) 재설정(R) 취소 도움말

- 범주형 자료의 요약통계치 추가 및 수정

오악 통계:

선택한 변수: 학력별

통계량(I):

가중되지 않은 빈도
행 N %
열 N %
표 N %
부표 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0

☐ 합계 및 소계에 대한 사용자 정의 요약 통계량(O)

통계량(I):

가중되지 않은 빈도
행 N %
열 N %
표 N %
부표 N %
레미어 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0

선택한 항목에 적용(S) 모든 항목에 적용(A) 닫기 도움말

요약 통계의 설명 수정

요약 통계: ×

선택한 변수: 문3) 이번 대통령 선거에 출마한 다음 후보 중 누구에게 투표하시겠어요? 후보는 무작위로 불러드리겠습니다.

통계량(I):

- 가중되지 않은 빈도
- 열 N %
- 표 N %
- 부표 N %
- 레이어 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0
행 N %	행 N %	nnnn.n%	1

☒ 합계 및 소계에 대한 사용자 정의 요약 통계량(C)

통계량(I):

- 가중되지 않은 빈도
- 행 N %
- 열 N %
- 표 N %
- 부표 N %
- 레이어 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0

선택한 항목에 적용(S) 모든 항목에 적용(A) 닫기 도움말

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0
행 N %	%	nnnn.n	1

- **열변수의 합계 추가**

범주 및 합계

선택한 변수: 문3) 이번 대통령 선거에 출마한 다음 후보 중 누구에게 투표하시겠습니까? 후보는 무작위로 불러드리겠습니다.

표시

값(U)

값	설명
1	더불어 민주당 문재인
2	자유한국당 홍준표
3	국민의당 안철수
4	바른정당 유승민
5	정의당 심상정
6	기타 후보

소계 및 계산된 범주

소계 추가(S)...

범주 추가(C)...

편집(I)...

삭제(D)

모든 소계에서 범주가 생략되었습니다: 0

범주 정렬

기준(B): 변수값

순서(O): 오름차순

제외(X):

보이기

☒ 전체(T)

설명(L): 전국

☐ 결측값(M)

☒ 빈 범주(P)

☒ 데이터를 스캔했을 때 발견된 다른 값(E)

합계 및 소계 표시

☒ 적용할 상위 범주(V)

☐ 적용할 하위 범주(W)

적용(A)

취소

도움말

- 합계에 대한 요약 통계량 추가

요약 통계: ×

선택한 변수: 문3) 이번 대통령 선거에 출마한 다음 후보 중 누구에게 투표하시겠어요? 후보는 무작위로 불러드리겠습니다.

통계량(I):

- 가중되지 않은 빈도
- 행 N %**
- 열 N %
- 표 N %
- 부표 N %
- 레이어 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0

☒ **합계 및 소계에 대한 사용자 정의 요약 통계량(C)**

통계량(I):

- 가중되지 않은 빈도
- 행 N %**
- 열 N %
- 표 N %
- 부표 N %
- 레이어 N %

표시(D):

통계량	설명	형식	소수점...
빈도	빈도	nnnn	0

선택한 항목에 적용(S)
모든 항목에 적용(A)
닫기
도움말

- SPSS 프로그램(syntax)

* 통계표 작성 : 전체 합계.

CTABLES

/VLABELS VARIABLES=SQ2 DM3 DM1 DM6 Q3 DISPLAY=LABEL

/TABLE SQ2 [C] + DM3 [C] + DM1 [C] + DM6 [C] BY Q3 [C][COUNT F40.0, ROWPCT.COUNT
'%' F40.1,

TOTALS[COUNT F40.0, ROWPCT.COUNT PCT40.1]]

/CATEGORIES VARIABLES=SQ2 Q3 ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
LABEL='전국' POSITION=BEFORE

/CATEGORIES VARIABLES=DM3 DM1 DM6 ORDER=A KEY=VALUE EMPTY=INCLUDE.

* 통계표 작성: 범주별 소계.

CTABLES

/VLABELS VARIABLES=SQ2 DM3 DM1 DM6 Q3 DISPLAY=LABEL

/TABLE SQ2 [C] + DM3 [C] + DM1 [C] + DM6 [C] BY Q3 [C][COUNT F40.0, ROWPCT.COUNT
'%' F40.1,

TOTALS[COUNT F40.0, ROWPCT.COUNT PCT40.1]]

/CATEGORIES VARIABLES=SQ2 [SUBTOTAL, 1, 2, OTHERNM] EMPTY=INCLUDE
POSITION=BEFORE

/CATEGORIES VARIABLES=DM3 DM1 DM6 ORDER=A KEY=VALUE EMPTY=INCLUDE

/CATEGORIES VARIABLES=Q3 [1, 2, 3, 4, 5, 6, 8, 9, OTHERNM] EMPTY=INCLUDE TOTAL=YES
LABEL='전국'

POSITION=BEFORE.

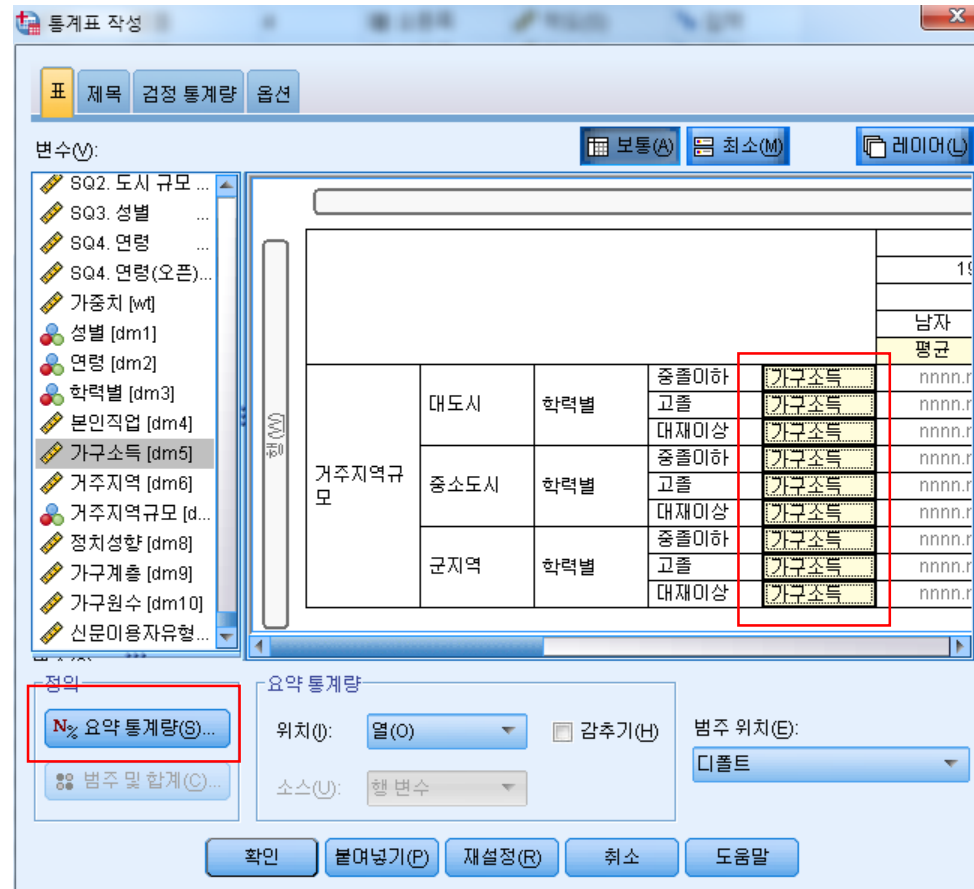
• SPSS 출력 결과

		문3) 이번 대통령 선거에 출마한 다음 후보 중 누구에게 투표하시겠습니까? 후보는 무작위로 불러드리겠습니다.																	
		전국		더불어민주당 문재인		자유 한국당 홍준표		국민의당 안철수		바른정당 유승민		정의당 심상정		기타 후보		없다		모르겠다	
		빈도	행 %	빈도	%	빈도	%	빈도	%	빈도	%	빈도	%	빈도	%	빈도	%	빈도	%
성별	전국	1000	1.0	358	35.8	84	8.4	302	30.2	28	2.8	28	2.8	19	1.9	106	10.6	75	7.5
	남자	495	1.0	173	34.9	44	8.9	159	32.0	14	2.8	18	3.7	11	2.2	51	10.4	25	5.1
	여자	505	1.0	185	36.6	40	7.9	144	28.5	14	2.8	9	1.8	8	1.6	55	10.9	50	9.8
연령별	20대	176	1.0	74	42.2	8	4.6	25	14.4	9	4.9	8	4.3	4	2.4	34	19.1	14	8.0
	30대	178	1.0	93	52.5	7	3.7	38	21.3	4	2.3	5	2.8	6	3.2	16	8.9	9	5.2
	40대	206	1.0	88	42.8	5	2.5	61	29.5	6	3.0	4	2.0	8	3.8	20	9.8	13	6.5
	50대	199	1.0	59	29.8	21	10.5	72	36.0	6	2.8	9	4.3	1	.4	17	8.7	15	7.4
	60세 이상	241	1.0	42	17.6	43	17.7	107	44.2	4	1.5	2	.9	1	.3	19	8.0	23	9.6

주) 20대는 19세 포함

- 통계표 :

- 수치자료의 요약 통계치를 표현
- 행변수 혹은 열변수에 수치형 자료를 설정



- 요약 통계치의 수정

