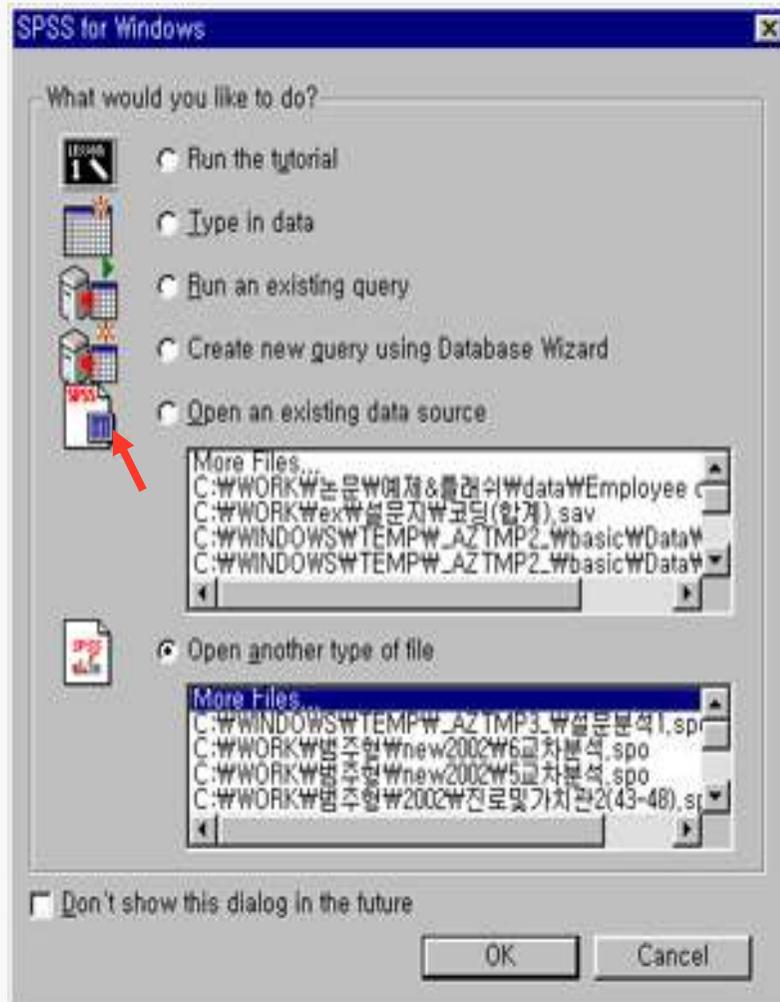


SPSS 사용법 특별강좌



- SPSS 아이콘을 더블 클릭 하거나, 시작 버튼을 누른 후 프로그램 메뉴에서 SPSS 12.0 for Windows를 선택하면 SPSS 는 작업선택 옆의 그림과 같은 대화상자(SPSS for Windows)를 연다.

- 이 대화상자는 지금까지 작업한 데이터를 보여줌으로써 기존작업을 계속하는데 편리하게 해준다.

- 'Cancel' 단추를 누르면 SPSS데이터 편집기 창이 열리면서 SPSS가 시작된다.

예제 1-1.sav - SPSS 데이터 편집기

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

4 : 마리수 141

	자극	반응	마리수	변수
1	1.00	1.00	25.00	
2	1.00	2.00	19.00	
3	2.00	1.00	31.00	
4	2.00	2.00	141.00	
5				
6				
7				

SPSS 프로세서

- ◇ 'Data View' : 자료를 화면에 보여줄 뿐 아니라 자료를 입력, 수정하는 데 사용된다.
- ◇ 해당 셀의 변수이름과 행 번호가 데이터 편집기 창의 왼쪽 위쪽에 나타난다.
- ◇ 변수이름은 'var00001'로 붙여진다.

예제 1-1.sav - SPSS 데이터 편집기

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

	이름	유형	자리수	소수점	설명	값	결측값
1	자극	숫자	8	2		{1.00, 반응유}	없음
2	반응	숫자	8	2		없음	없음
3	마리수	숫자	8	2		없음	없음
4							

프로세서 영역

SPSS 프로세서

- ◇ 자료 입력
- ◇ 변수 정의
  - ◆ 변수 이름, 유형, 설명
  - ◆ 변수값 설명
  - ◆ 결측치
- ◇ 파일 저장

2\_1표 - SPSS 데이터 편집기

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

19 : 보험금액 7

	일련번호	성별	직업	종교	교육수준	보험금액	월수입
1	1	m	1	1	3	7.0	110
2	2	m	2	1	4	12.0	135
3	3	f	2	3	5	8.5	127
4	4	f	3	3	5	5.0	150
5	5	m	1	3	3	4.5	113

2\_1표 - SPSS 데이터 편집기

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래프(G)

	이름	유형	자리수	소수점
1	일련번호	숫자	8	0
2	성별	문자열	8	0
3	직업	숫자	8	0
4	종교	숫자	8	0
5	교육수준	숫자	8	0
6	보험금액	숫자	8	1
7	월수입	숫자	8	0

- ◇ [파일]-[새로 만들기]-[데이터]
- ◇ 데이터 보기 sheet
  - ◆ Interactive 입력 방식
- ◇ 변수 보기 sheet
  - ◆ 변수 이름 지정
  - ◆ 유형 지정
  - ◆ 변수값 설명( [보기]메뉴)
  - ◆ 결측치 정의
  - ◆ 측도 정의
- ◇ 파일 저장

- ◇ 자료의 입력
  - ◆ 결측값 입력은 '9' 또는 '999', '.'
- ◇ 자료의 편집
  - ◆ [편집]-[복사]/[잘라내기]
  - ◆ [편집]-[붙여넣기]
- ◇ 새 변수 추가
  - ◆ [데이터]-[변수 삽입]
- ◇ 변수 삭제/ 케이스 삭제
  - ◆ 삭제대상 변수 또는 케이스 지정
  - ◆ [편집]-[지우기]
  - ◆ 오른쪽 마우스 버튼
- ◇ 자료 파일의 저장
  - ◆ [파일] -[저장]
  - ◆ [파일]-[다른이름저장]
- ◇ SPSS 파일 확장자
  - ◆ \*.SAV : Spss 전용파일
  - ◆ \*.dat : Notepad
  - ◆ \*.xls : MS Excel ( 변수 정보 저장 안됨)

**변수 유형** [?] [X]

숫자(N)  
 콤마(C)  
 점(D)  
 지수표기(S)  
 날짜(A)  
 달러(L)  
 사용자 통화(U)  
 문자열(B)

자리수(W):   
 소수점이하자리수(P):

**결측값** [?] [X]

결측값 없음(N)  
 미산형 결측값(D)  
 한개의 선택적 미산형 결측값을 더한 범위(B)

하한(L):   
 상한(H):   
 미산값(S):

**변수값 설명** [?] [X]

변수값 설명

변수값(U):   
 변수값 설명(E):

범	값	범
8	오른쪽	척도

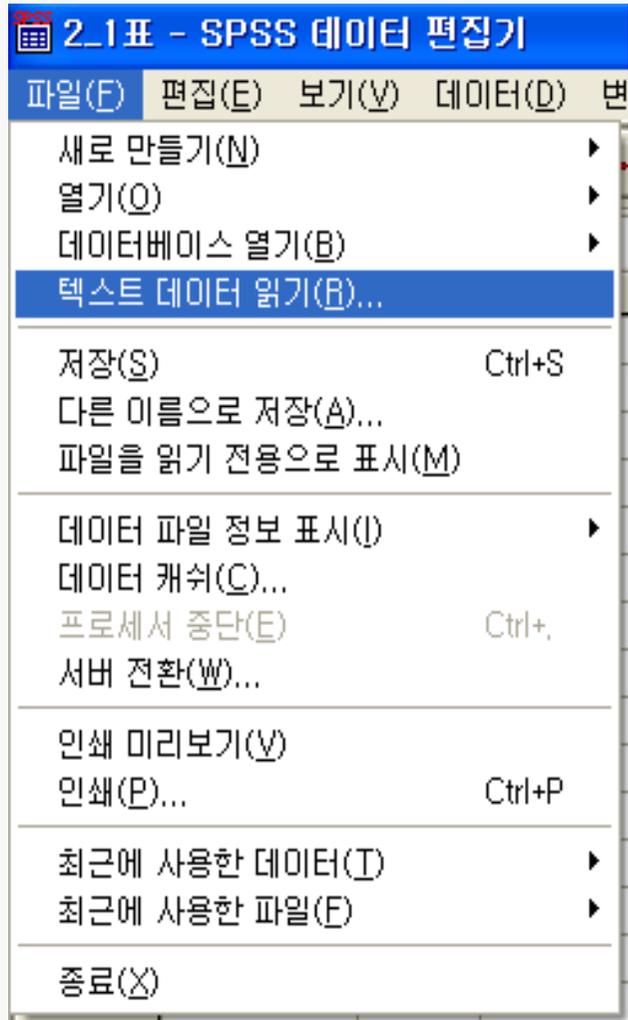
항 목	내 용	비 고
Name	이름 : 변수이름	
Type	유형 : 변수유형으로 숫자변수인지 문자변수인지를 지정	
Width	자리 수 : 숫자나 문자가 차지하는 총 자리 수	디폴트는 8
Decimals	소수점 이하 자리 : 숫자의 경우 소수점 이하의 자리 수	디폴트는 2
Label	설명 : 변수이름에 대한 설명(label)	
Values	값 : 변수값들에 대한 설명	
Missing	결측값 : 사용자가 정의하는 결측값	
Columns	열 : 열 자리수	디폴트는 8
Align	맞춤 : 오른쪽, 왼쪽, 가운데 맞춤	디폴트는 숫자는 오른쪽 문자는 왼쪽
Measure	척도 : 척도(구간/비율척도), 순서척도, 명목척도	

디폴트 확장자	
*.sav	데이터 편집기에 있는 파일
*.spo	출력항해사 창에 있는 결과 파일
*.sps	명령문 편집기 창에 있는 명령문 파일

종료할 때 **Save contents of Data editor to Untitled?** 라는 경고 메시지가 뜬다.

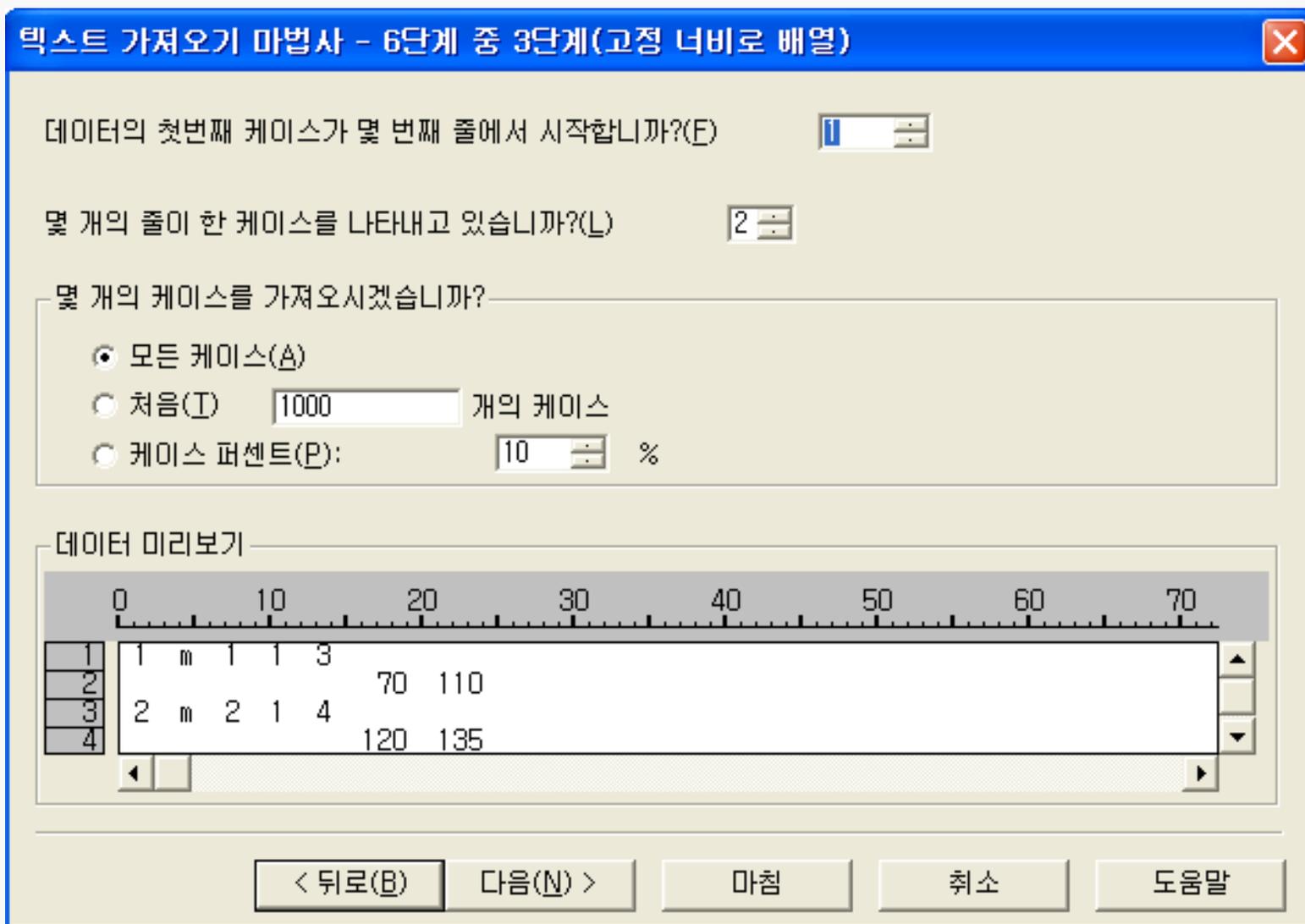
여기서 저장하려면 예를,

저장하지 않고 그냥 닫으려면 아니오를 누르면 된다.

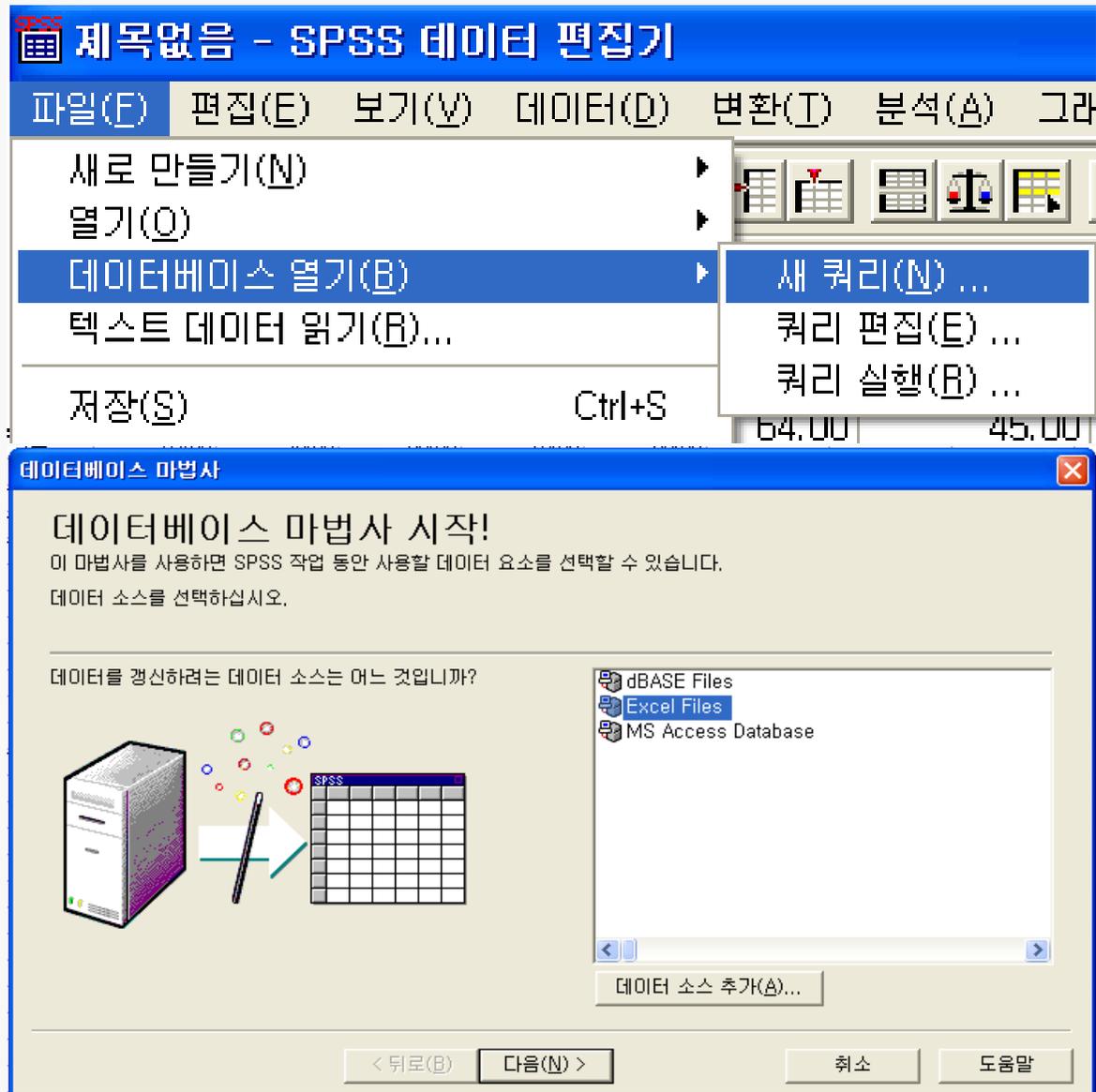


- ◇ 텍스트 파일 열기
  - ◆ [파일]-[텍스트 데이터 읽기]
  - ◆ 고정형식
    - ◆ 마법사(Wizard) 이용
    - ◆ 명령문
  - ◆ 자유형식
    - ◆ 마법사 이용
    - ◆ 명령문





# 파일 불러오기(EXCEL)



변환(T) 분석(A) 그래프(G) 유틸리티(U) 창(W)

- 변수 계산(C)...
- 코딩변경(B)**
  - 같은 변수로(S)...
  - 새로운 변수로(D)...
- 시각적 구간화(B)...
- 빈도변수 생성(O)...
- 순위변수 생성(K)...
- 자동 코딩변경(A)...

---

- 시계열변수 생성(M)...
- 결측값 바꾸기(V)...
- 난수 시작값(S)...

---

- 변환 중지(T)

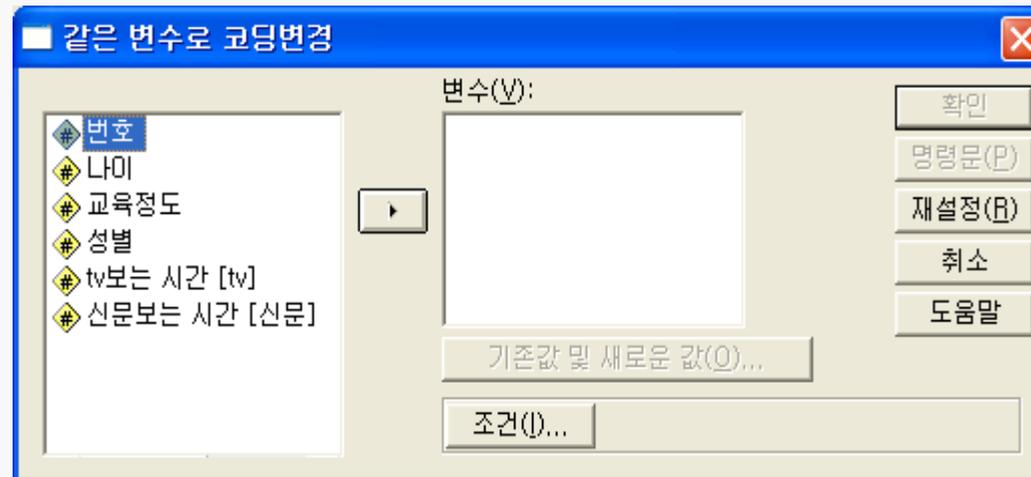
tv	신문
45.00	40.00
40.00	20.00
50.00	60.00
51.00	15.00
58.00	10.00
48.00	30.00
70.00	40.00

- 코딩변경 : 기존 변수내 값의 코드를 변경하거나 기존 변수 값의 코딩변경으로 새로운 변수를 만들 수 있다. 데이터 값의 코드를 변경하거나 수정, 몇 개 범주를 합치거나 조합할 때 사용

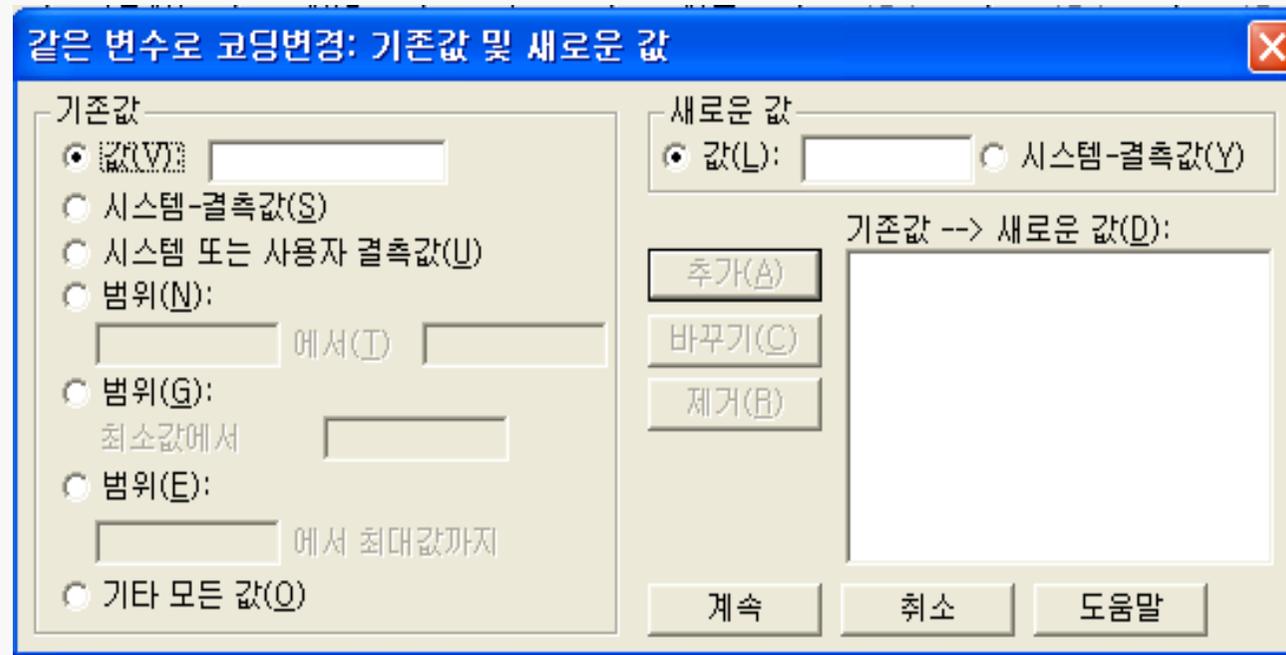
## (1) 같은 변수로 코딩 변경

메뉴에서 다음을 선택하면 [Recode Into Same Variables] 대화 상자가 나타난다

변환 → 코딩변경 → 같은 변수로



기존 변수의 값을 다시 할당하거나 기존 값 범위를 새 값으로 만들 수 있다.

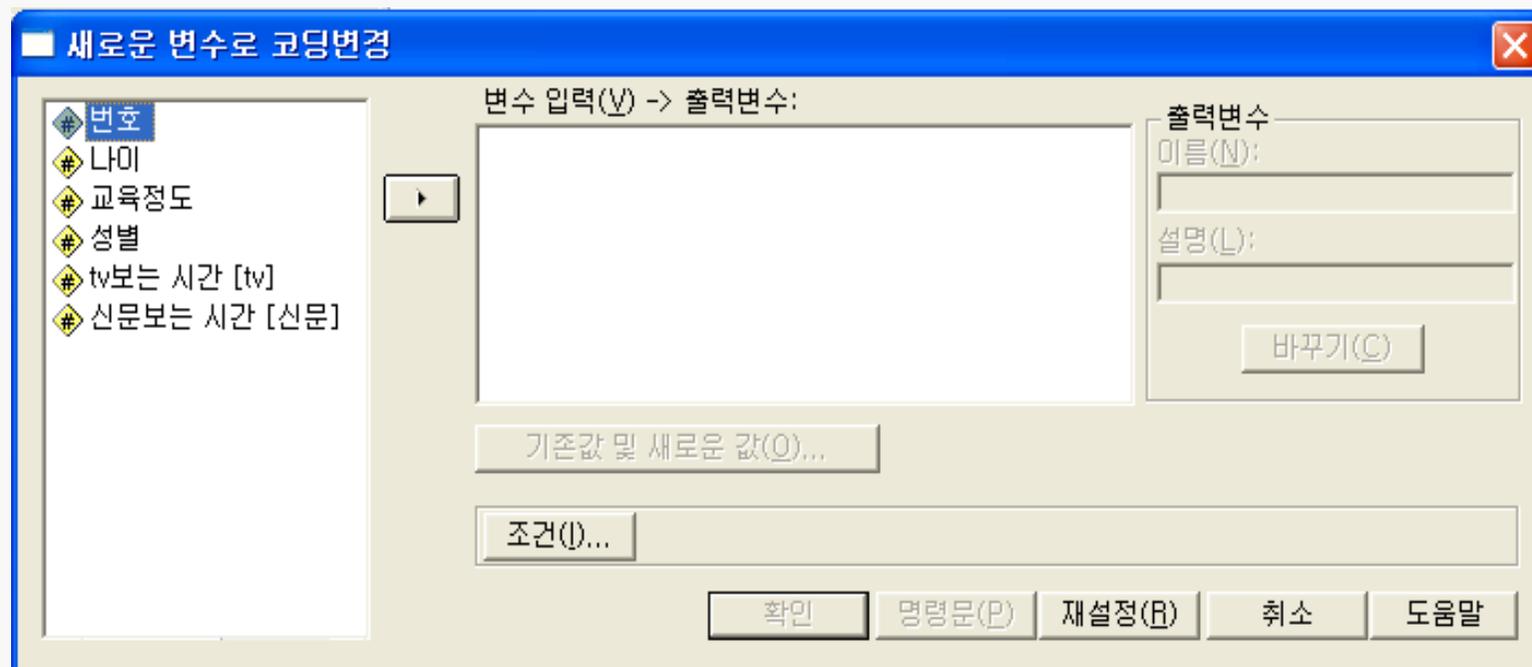


- ㄱ. 변수 목록으로부터 코드를 변경할 변수를 선택하고 [기존값 및 새로운값] 단추를 선택한다.
- ㄴ. 대화 상자에서 기존값과 새로운 값을 입력한 후 [추가(A)]를 선택한다.

## (2) 새로운 변수로 코딩변경

: 메뉴에서 다음을 선택하면 [Recode Into Different Variables] 대화 상자가 나타난다

변환 → 코딩변경 → 다른 변수로



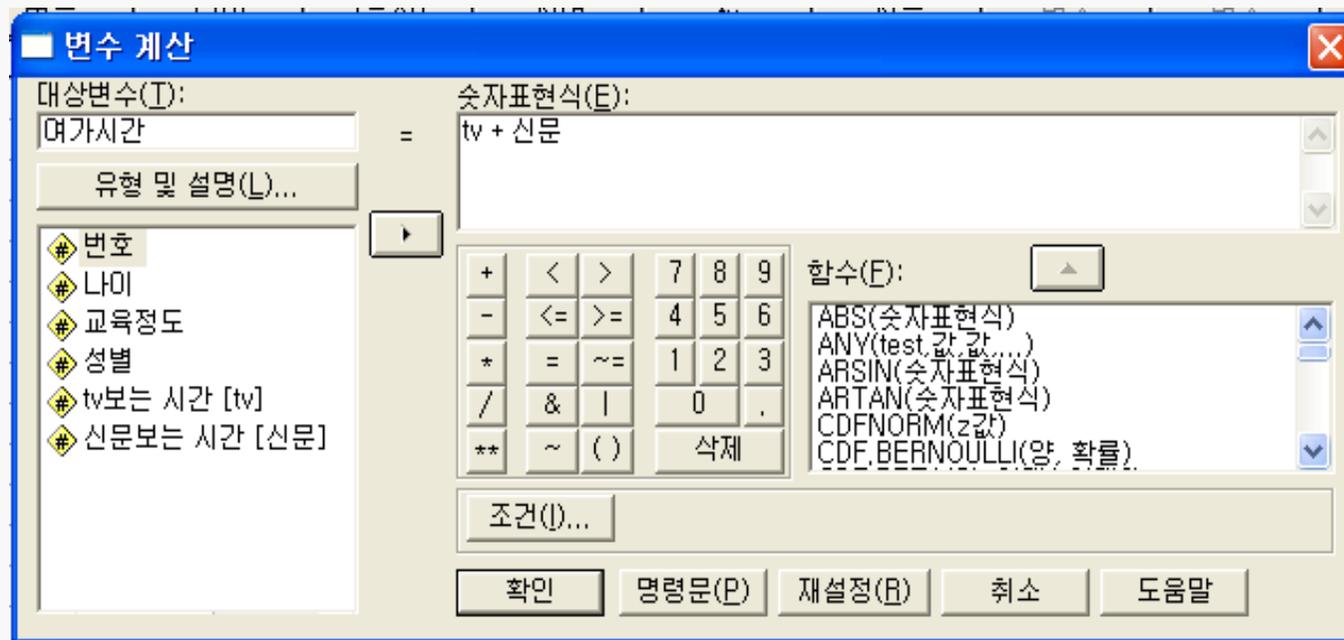
과정은 같은 변수로 코딩하는 것과 유사하다

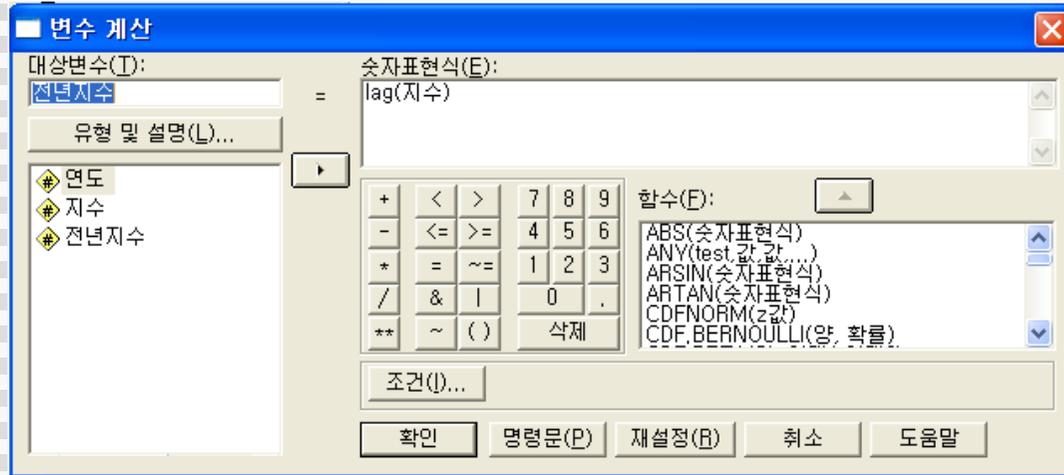
**변수계산** : SPSS데이터 파일을 용도에 맞게 변환하는 방법

기존의 변수를 이용하여 새로운 변수를 만들고자 할 때 사용하며 산술식 및 여러 가지 형태의 함수를 이용할 수 있다.

메뉴에서 다음을 선택하면 변수계산 대화상자가 나타난다.

변환 → 변수계산



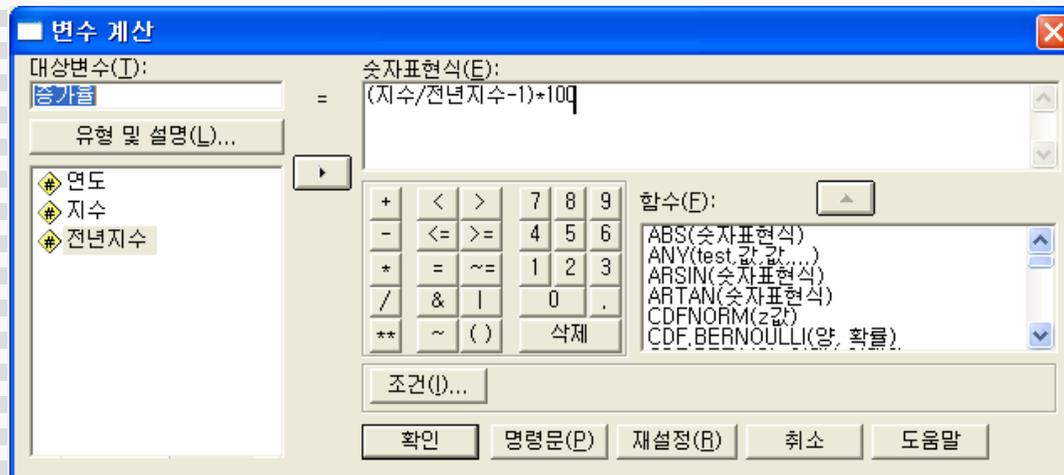


## [변환]-[변수계산]

- ◆ 여가시간 = TV + 신문
- ◆ 시청시간
  - ◆ 여성=> TV시청\*1.2
  - ◆ 남성 => TV시청\*1.5

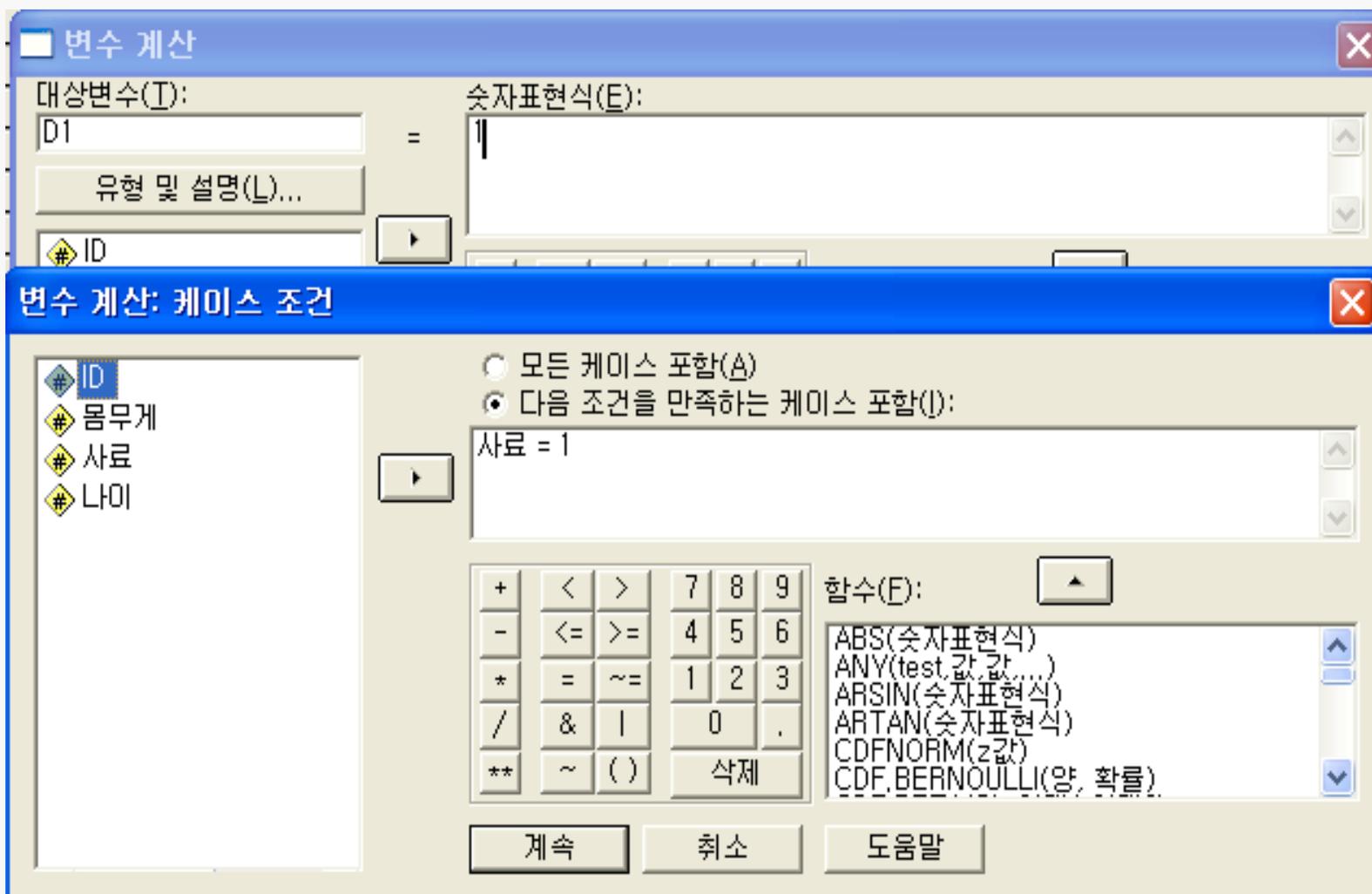
## 예제3.1(표3-2.sav)

- ◆ 전년지수=LAG(지수)
- ◆ 증감률=(지수/전년지수-1)\*100



## 예제3.2 가변수 생성하기

- ◆ 통계모형 적합시 유용



함 수	설 명
ABS(숫자표현식)	절대값
ARTAN(숫자표현식)	$\tan^{-1}$ 값
COS(숫자표현식)	cos 값
EXP(숫자표현식)	지수 $e^x$
LAG(변수)	변수의 바로 직전 케이스에 대한 값을 준다.
LG10(숫자표현식)	상용대수
LN(숫자표현식)	자연대수
MISSING(변수)	변수의 값이 결측값이면 1이 된다.
MOD(숫자표현식, modulus)	숫자 표현식 값을 modulus로 나눈 나머지
NORMAL(표준편차)	$N(0, \text{표준편차}^2)$ 분포를 갖는 확률 변수값을 준다.
RND(숫자표현식)	정수로 반올림한다.
SIN(숫자표현식)	sin 값
SQRT(숫자표현식)	제곱근
SYSMIS(숫자변수)	변수의 값이 시스템 결측값이면 1이 된다.
TRUNC(숫자표현식)	소수 이하를 버린다.
UNIFORM(최대값)	$U(0, \text{최대값})$ 분포를 갖는 확률변수값을 준다.
VALUE(변수)	변수에 대한 사용자 정의 결측값을 무시하고 계산에 포함
YRMODA(년, 월, 일)	1582년 10월 15일 이후의 총 날짜를 제공한다.

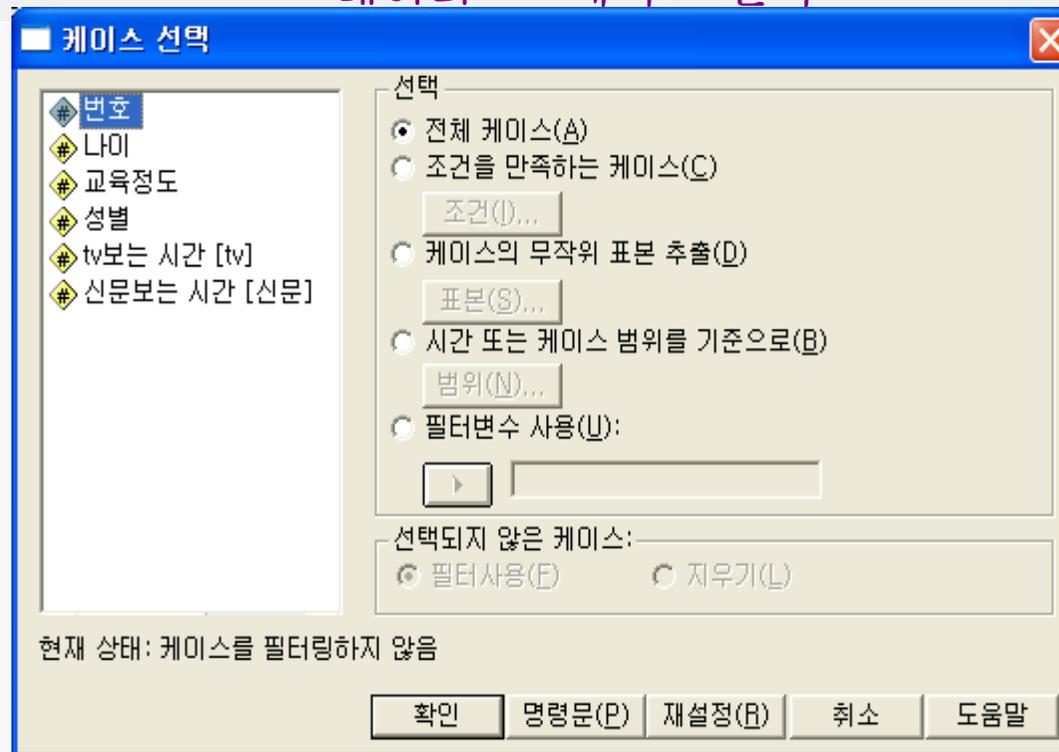
## 논리식에서 사용되는 연산자

연산자	종 류	
관계연산자 (relational operator)	= : 같다	~= : 같지 않다
	< : 작다	<= : 작거나 같다
	> : 크다	>= : 크거나 같다
논리연산자 (logical operator)	&(AND) : 참과 참을 연산할 때만 참이 되며 그렇지 않으면 거짓이 된다.	
	(OR) : 적어도 하나가 참이면 그 결과가 참이 된다.	
	~(NOT) : 참에 대해서는 거짓으로, 거짓에 대해서는 참으로 연산된다.	

**케이스 선택** : 분석에 필요한 케이스만을 선택한다.

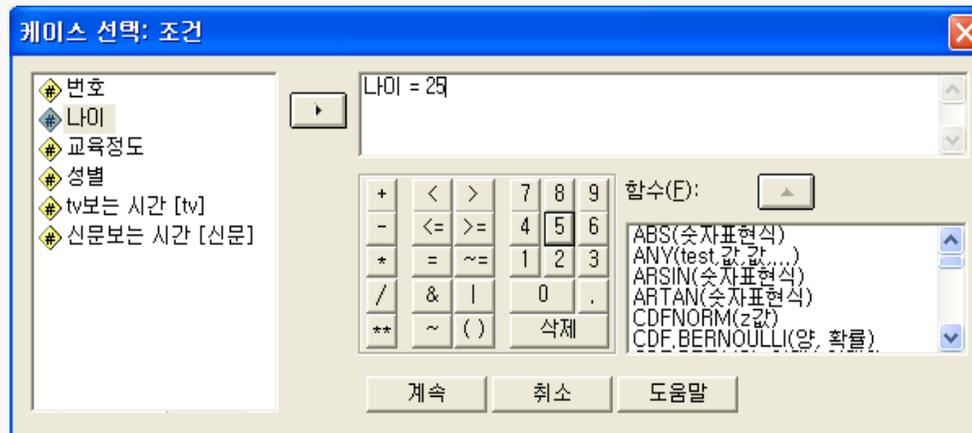
어떤 작업은 전체 케이스를 사용하지 않고 작업파일의 케이스에서 일부분만을 선택하여 사용한다.

데이터 → 케이스 선택

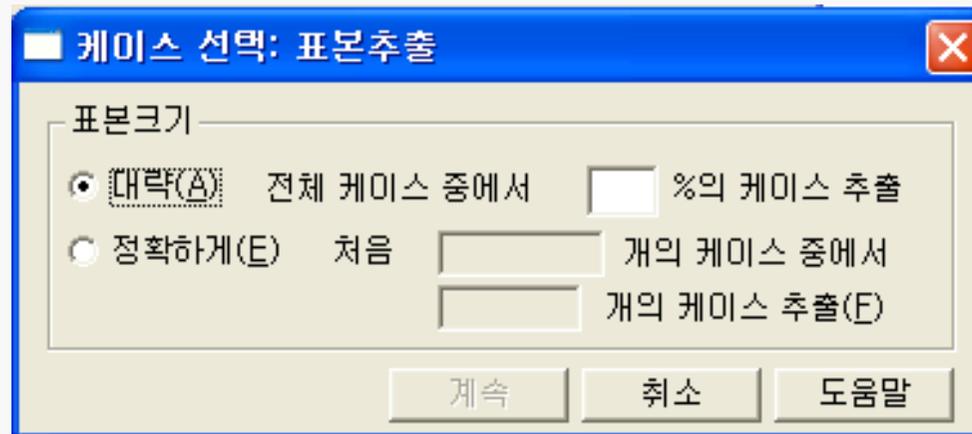


실습 : 전체 case중 25% 랜덤 추출 ,전체중 5개 랜덤 추출 , 3번째 case~6번째 case 선택, id=\$casenum

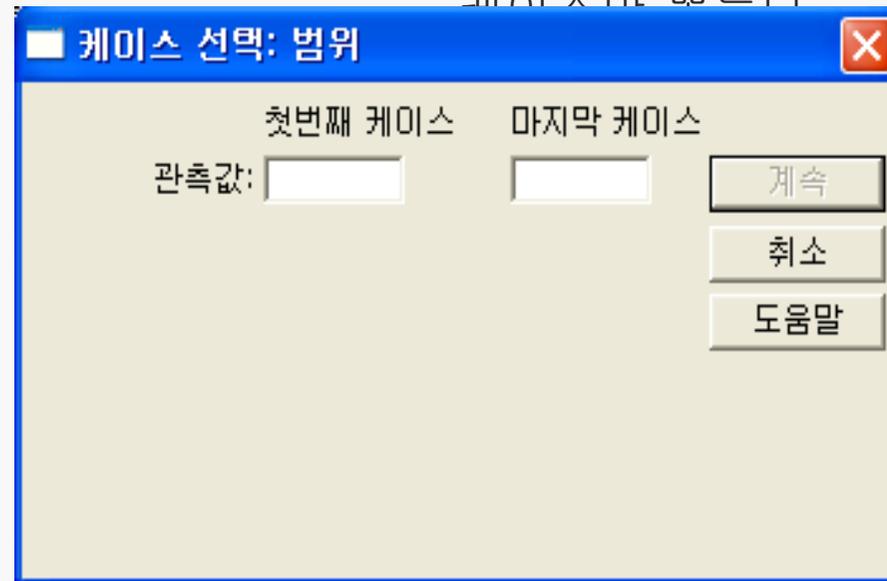
- **조건을 만족하는 케이스** : 케이스를 선택하는데 조건식을 사용한다.  
조건에 맞는 케이스만 선택된다.



- **케이스의 무작위 표본추출** : 케이스의 수나 비율을 지정하여 랜덤 표본을 뽑는다

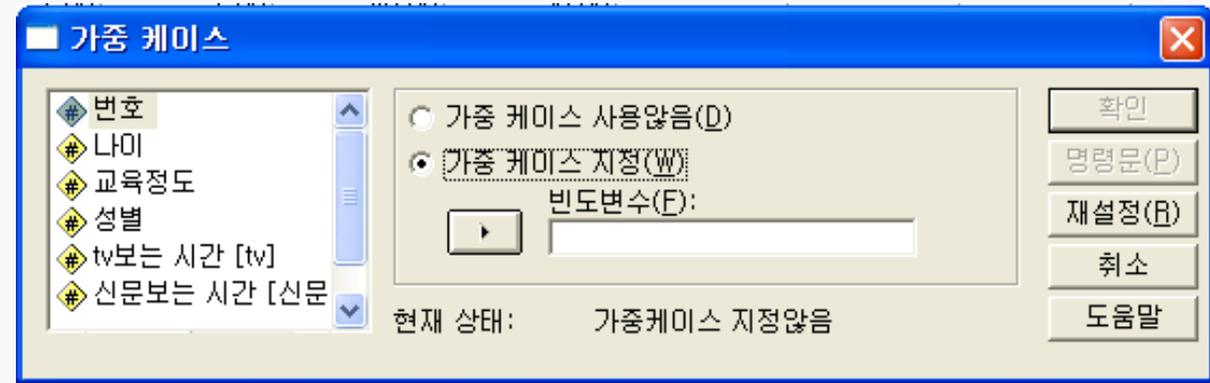


- 시간 또는 케이스 범위를 기준으로 : 시계열 데이터에서 시간의 범위에 해당되는 케이스만 뽑는다.



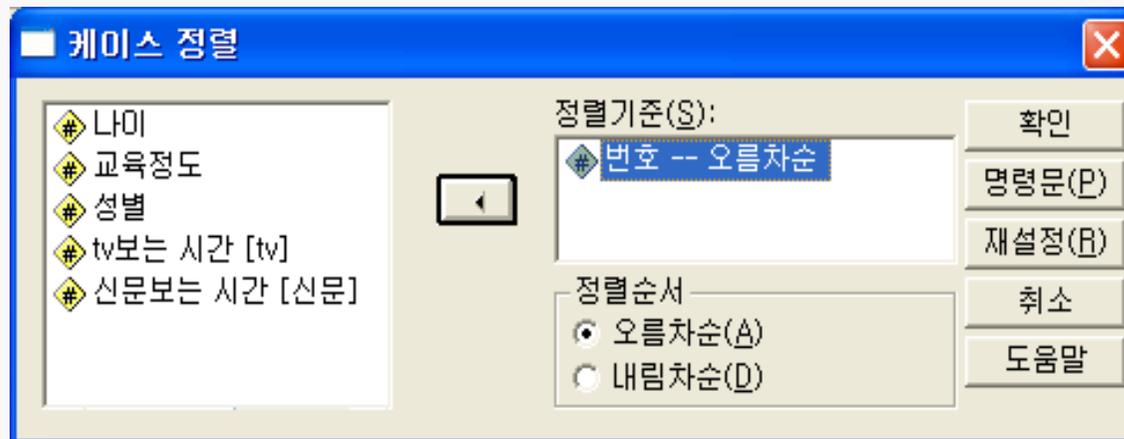
- 선택되지 않은 케이스의 처리 방법
- 현재 상태 : 현재 상태가 케이스 선택 상태인지 아닌지를 표시한다.

**가중케이스:** 가중변수 값을 통하여 케이스를 가중치를 부여한다.  
 이 명령을 사용하지 않으면 분석에서 모든 케이스는 똑같은 가중값 1을 가지게 된다. 그러나, 데이터의 한 변수를 가중 변수를 사용하여 케이스에 대한 가중값을 줄 수 있다.



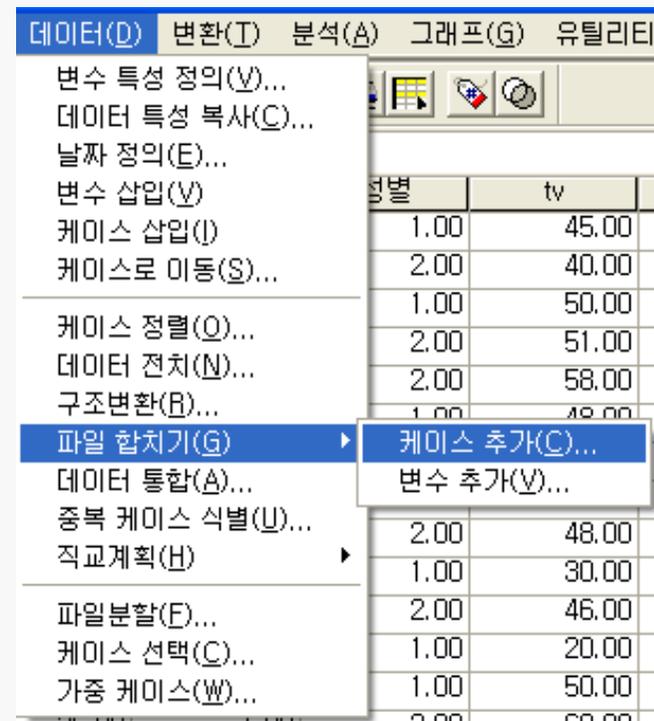
케이스 정렬(Sort Cases) : 하나이상의 변수 값을 기준으로 케이스를 정렬한다.

데이터 → 케이스 정렬



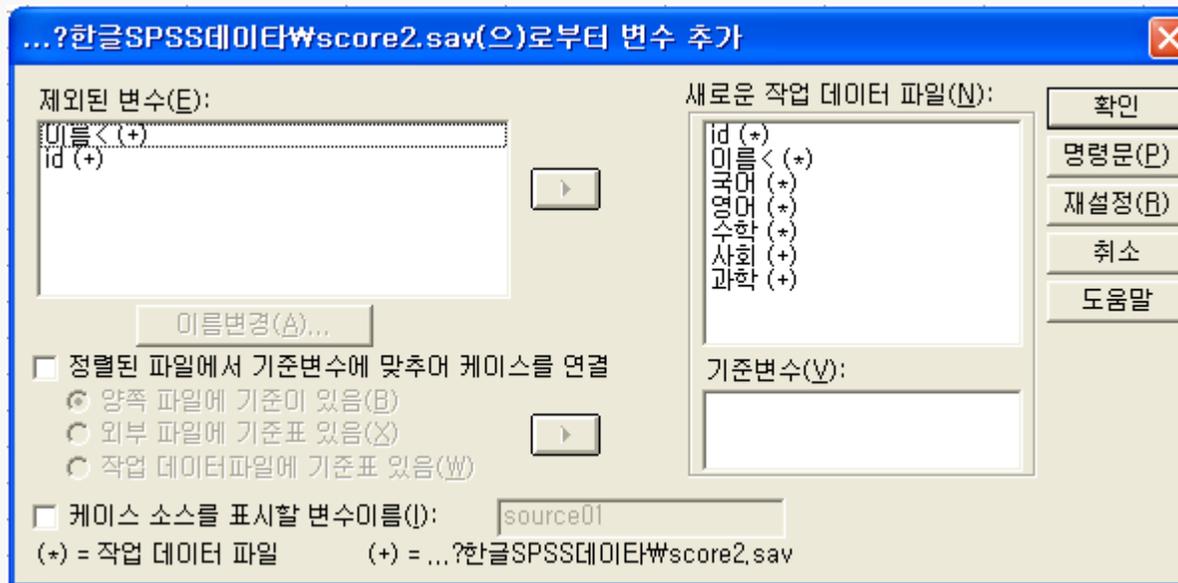
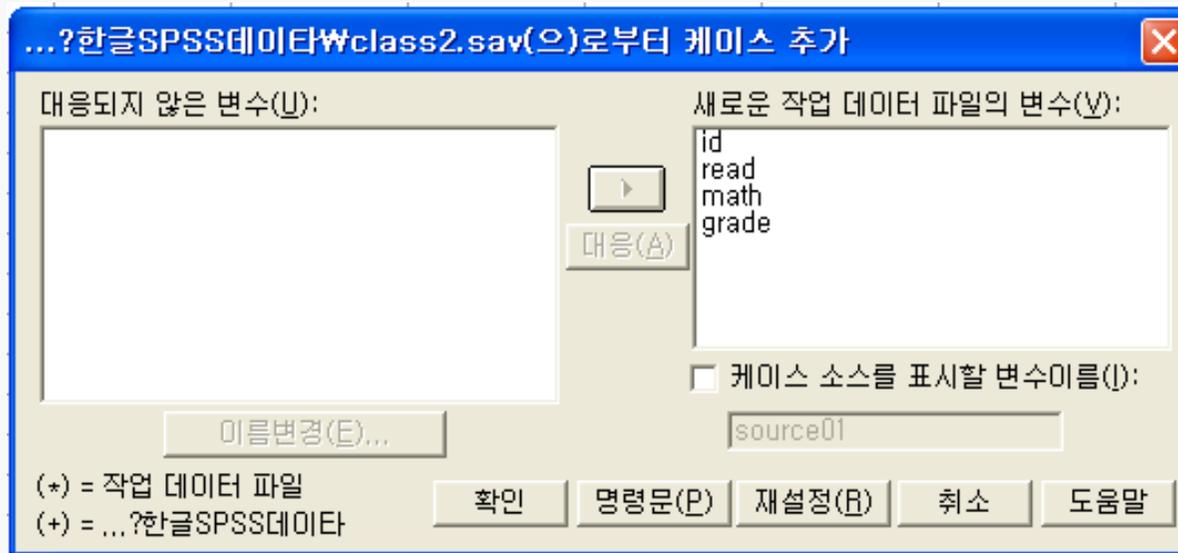
- ▶ 정렬 기준 : 목록에 나타나는 순서대로 분류한다.
- ▶ 정렬순서 : Ascending-오름차순 , Descending - 내림차순

데이터 파일 합치기 : 두개이상의 데이터 파일을 합친다



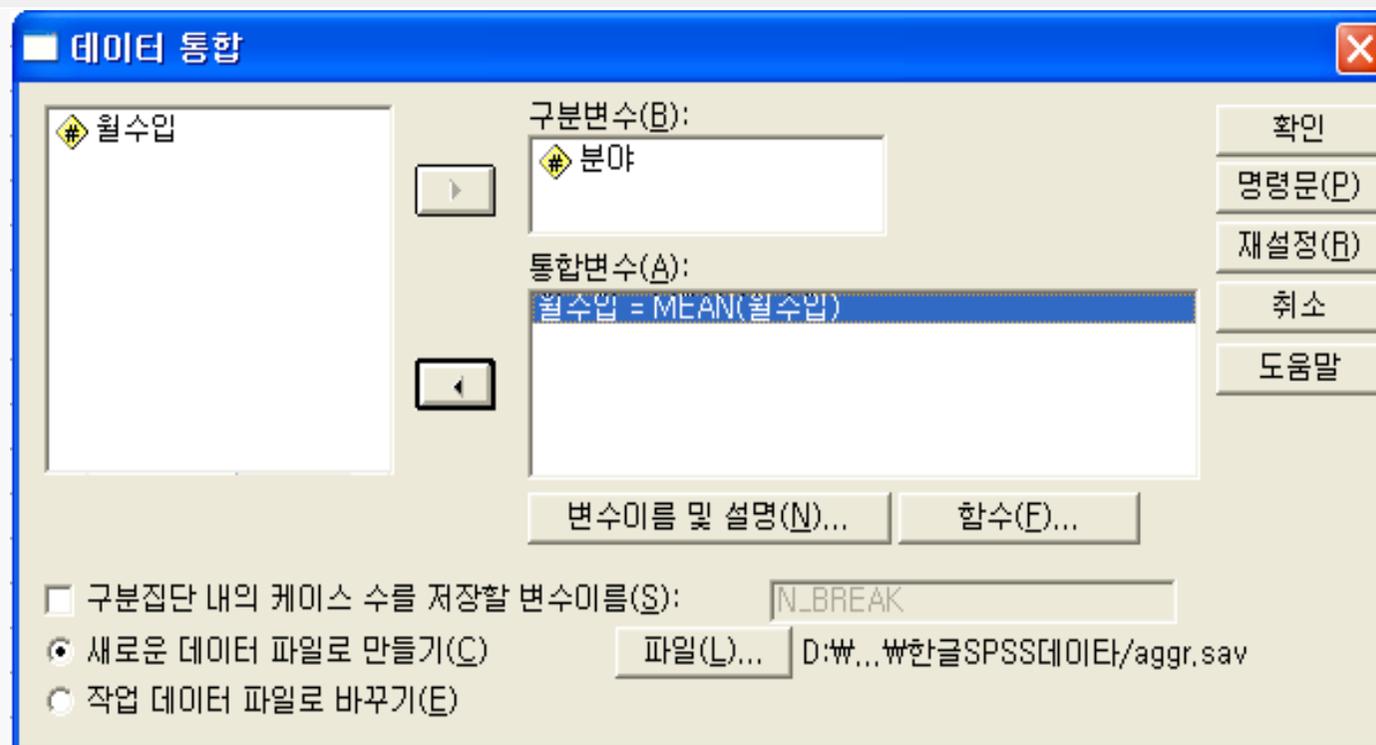
**케이스 추가** : 같은 변수들을 포함하고 다른 케이스들을 가지는 파일들의 합성

**변수 추가** : 같은 케이스들을 포함하고 다른 변수들을 가지는 파일들의 합성



**데이터 통합:** 집단 변수값을 가지고 집단에 대한 통계량을 데이터로 통합한다.  
 집단 특성을 나타내는 변수를 통합 파일로 작성한 후,  
 이 통합 파일의 변수를 집단별 케이스에 추가하여 합성할 수 있다.

### 데이터 → 데이터 통합(Aggregate)



## salary(4-5) - SPSS 데이터 편

파일(F) 편집(E) 보기(V) 데이터(D)



3 :

	분야	월수입
1	1	87
2	1	90
3	1	62
4	2	80
5	2	70
6	2	65
7	2	57
8	3	100
9	3	115
10	3	120
11	3	102



## 제목없음 - SPSS 데이터 편집기

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그라



7 :

	분야	월수입	월수입_1	N_BREAK
1	1	79.67	239.00	3
2	2	68.00	272.00	4
3	3	109.25	437.00	4

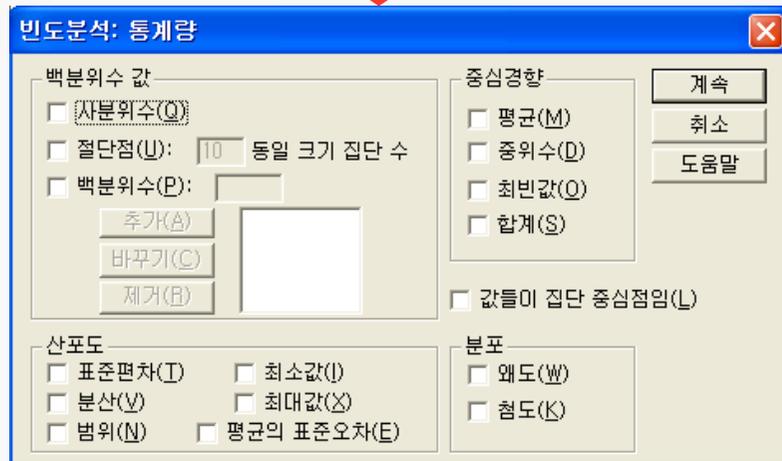
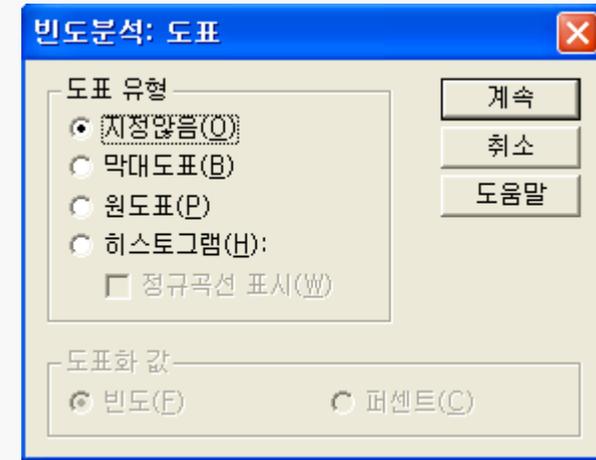
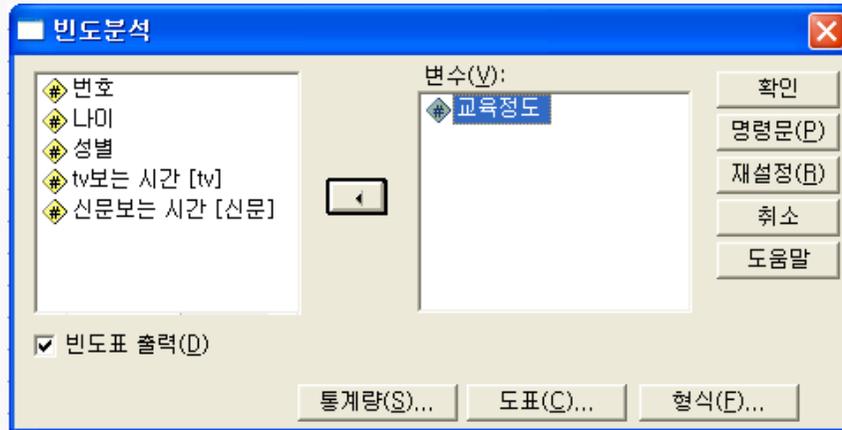
- ◆ 기술통계량(Summarize): 빈도수, 기술통계, 탐색적 자료분석, 분할표, 요약보고서
- ◆ 통계표작성(Custom tables) : 통계표를 작성하는 데 이용
- ◆ 평균 비교(Compare means) : 부그룹별 기술통계를 제공하는 집단별 평균분석 (MEANS) 절차, 일표본 T-검정, 두 집단간의 차이를 검정하는 T-검정, 일원배치 분산 분석
  - ◆ 일반선형모형(Anova models) : 분산분석, 다변량분산분석, 반복측정분석
  - ◆ 상관분석(Correlate) : 변수간의 상관분석, 편상관분석, 거리측도
  - ◆ 회귀분석(Regression)
    - 선형회귀, 곡선회귀, 로지스틱, 프로빗, 비선형, 가중회귀, 2-단계 최소제곱 등을 제공
  - ◆ 로그선형분석(Loglinear)
    - 일반로그선형모형, 로짓분석, 로그선형모형선택 등을 제공
  - ◆ 분류분석(Classify) : 판별분석, 군집분석 등을 제공
  - ◆ 데이터 축소(Data reduction): 인자분석, 대응분석, 최적화척도법 등
  - ◆ 척도화분석(Scale) : 신뢰도분석, 다차원척도법 제공
  - ◆ 비모수 검정(Nonparametric tests) : 카이제곱, 이항검정, 런검정 등의 비모수검정
  - ◆ 시계열 분석(Time series) : 지수평활, 자기회귀, ARIMA 모형 등의 시계열 분석
  - ◆ 생존분석(Survival) : 생명표, Kaplan-Meier 생존분석 등을 제공
  - ◆ 다중응답분석(Multiple response): 다중응답분석처리 대한 분석 제공
  - ◆ 결측값 분석(Missing value analysis) : 결측값 처리에 대한 분석

분석(A)	그래프(G)	유
보고서(P)		▶
기술통계량(E)		▶
표(T)		▶
평균 비교(M)		▶
일반선형모형(G)		▶
혼합 모형(X)		▶
상관분석(C)		▶
회귀분석(R)		▶
로그선형분석(O)		▶
분류분석(Y)		▶
데이터 축소(D)		▶
척도화분석(A)		▶
비모수 검정(N)		▶
시계열 분석(I)		▶
생존분석(S)		▶
다중응답(U)		▶
결측값 분석(V)...		▶
복합 표본(L)		▶

분석(A)	그래프(G)	유틸리티(U)	창(W)	도움
보고서(P)				
<b>기술통계량(E)</b>				
표(I)				
평균 비교(M)				
일반선형모형(G)				
혼합 모형(X)				
상관분석(C)		50.00	60.00	
회귀분석(R)		51.00	15.00	
로그선형분석(Q)		58.00	10.00	
분류분석(Y)		48.00	30.00	
데이터 축소(D)		70.00	40.00	
척도화분석(A)		25.00	40.00	
비모수 검정(N)		48.00	45.00	
시계열 분석(I)		30.00	55.00	
생존분석(S)		46.00	30.00	
다중응답(U)		20.00	70.00	
결측값 분석(V)...		50.00	60.00	
복합 표본(L)		50.00	.	

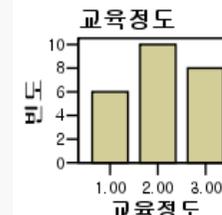
- ◇ 기술통계량
  - ◆ 중심위치의 측도
    - ◆ 평균
    - ◆ 중앙값
  - ◆ 산포도
    - ◆ 분산
    - ◆ 표준편차
    - ◆ 표준오차
    - ◆ 범위
    - ◆ 변동계수
  - ◆ 분포의 형태
    - ◆ 왜도
    - ◆ 첨도
  - ◆ 백분위수

# 빈도분석



교육정도

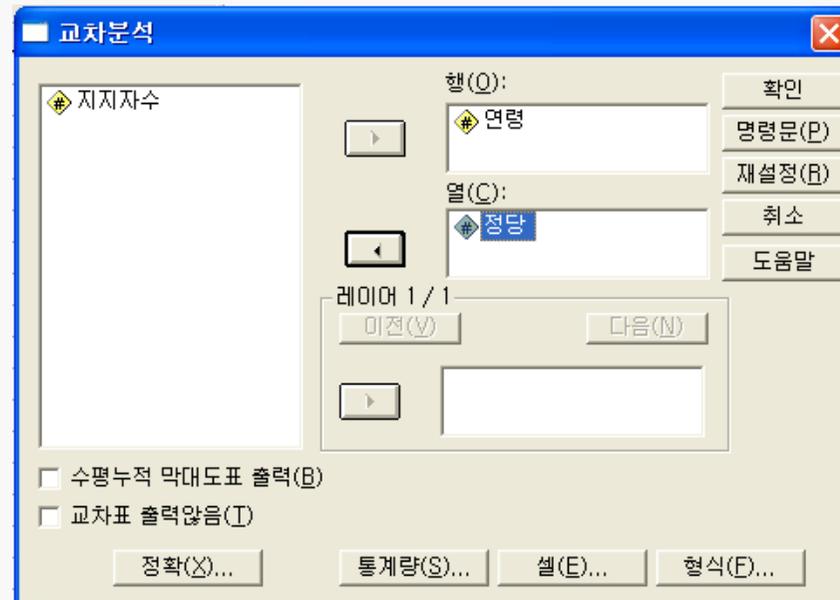
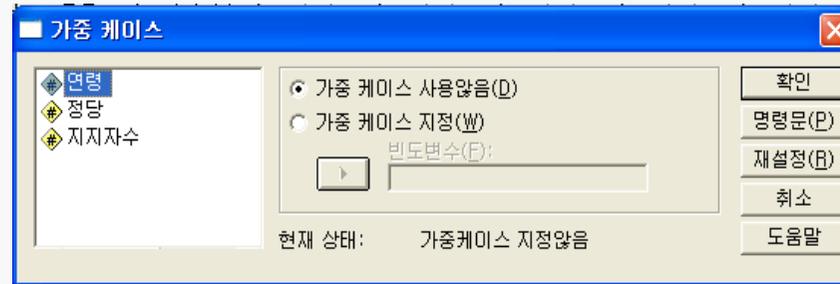
	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 1.00	6	24,0	25,0	25,0
2.00	10	40,0	41,7	66,7
3.00	8	32,0	33,3	100,0
합계	24	96,0	100,0	
결측 시스템 결측값	1	4,0		
합계	25	100,0		



# 교차분석(범주형)

## 교차분석:

두개 이상의 범주형 변수들 사이의 관련 정도를 알아보고자 하는 경우에 사용  
분석 → 기술통계량 → 교차분석



**교차분석: 통계량**

카이제곱(H)

명목 데이터

- 분할계수(O)
- 파이 및 Cram의 V(P)
- 람다(L)
- 불확실성 계수(U)

명목 대 등간 척도

- 에타(E)

순서 데이터

- 감마(G)
- Somers의 d(S)
- Kendall의 타우-b(B)
- Kendall의 타우-c(C)

상관관계(B)

- 카파(K)
- 위험도(I)
- McNemar(M)

Cochran 및 Mantel-Haenszel 통계량(A)

검정 공통승산비가 동일:

계속 취소 도움말

**교차분석: 셀 출력**

빈도

- 관측빈도(O)
- 기대빈도(E)

퍼센트

- 행(R)
- 열(C)
- 전체(T)

잔차

- 표준화하지 않음(U)
- 표준화(S)
- 수정된 표준화(A)

정수가 아닌 가중값

- 셀 수 반올림(N)
- 셀 수 절삭(L)
- 조정 없음(M)
- 케이스 가중값 반올림(W)
- 케이스 가중값 절삭(H)

계속 취소 도움말

	케이스					
	유효		결측		전체	
	N	퍼센트	N	퍼센트	N	퍼센트
연령 * 정당	940	100,0%	0	,0%	940	100,0%

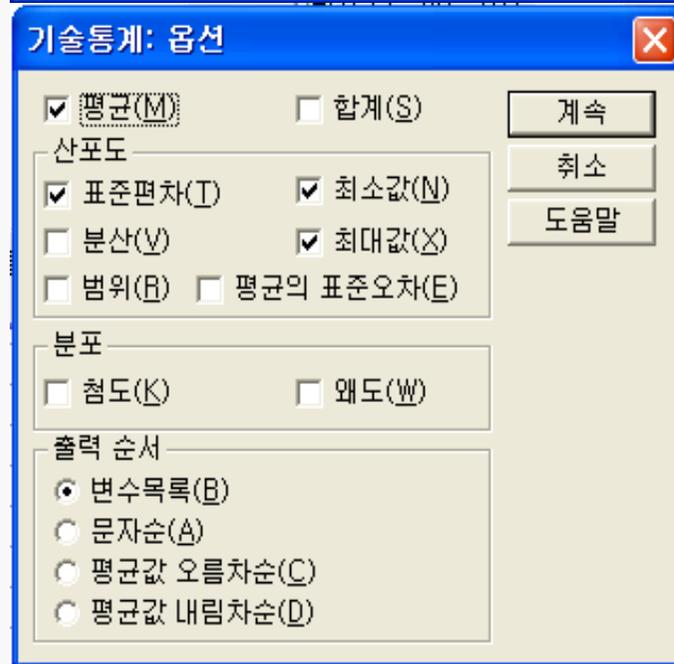
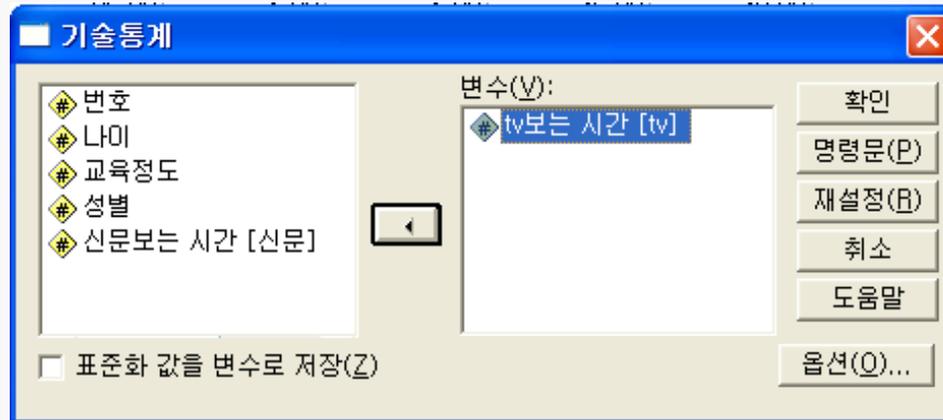
연령 \* 정당 교차표

			정당			전체
			1	2	3	
연령 1	빈도		158	53	62	273
	기대빈도		123,4	99,6	50,0	273,0
2	빈도		172	128	89	389
	기대빈도		173,2	139,8	70,1	383,0
3	빈도		95	162	27	284
	기대빈도		128,4	103,6	52,0	284,0
전체	빈도		425	343	172	940
	기대빈도		425,0	343,0	172,0	940,0

### 카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)
Pearson 카이제곱	91,341 <sup>a</sup>	4	,000
우도비	93,347	4	,000
선형 대 선형결합	3,056	1	,080
유효 케이스 수	940		

a. 0 셀 (.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 49,95입니다.

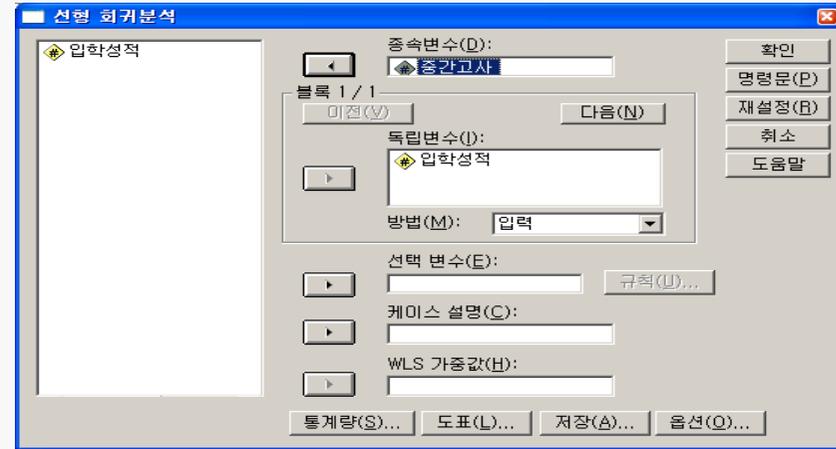
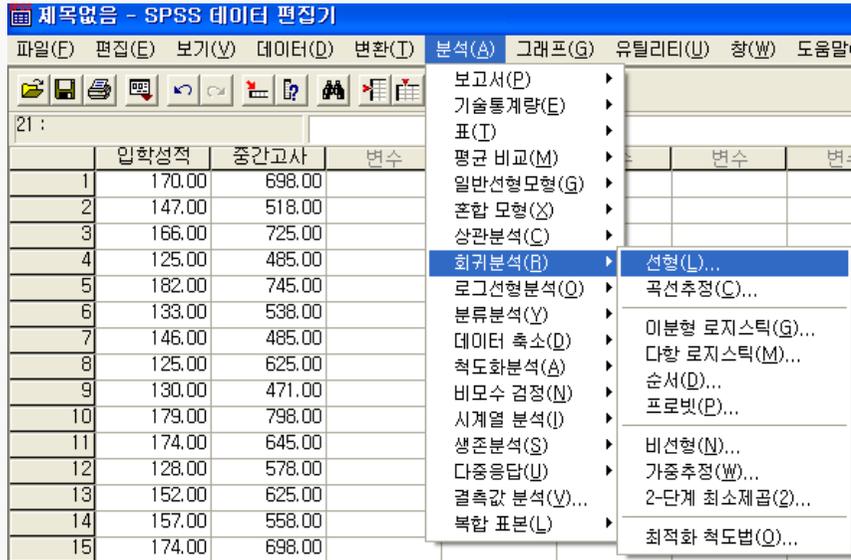


분석 → 기술통계량 → 기술통계

- ◆ 기술통계
  - 옵션으로 각종 통계량을 구할 수 있다.

# 회귀분석

SPSS

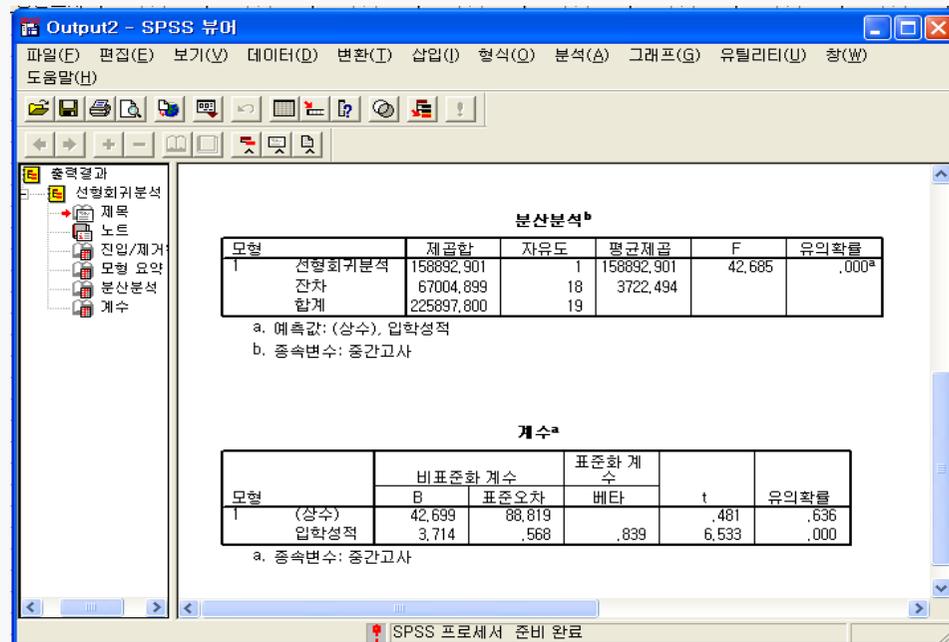


## [분석]

- [회귀분석]

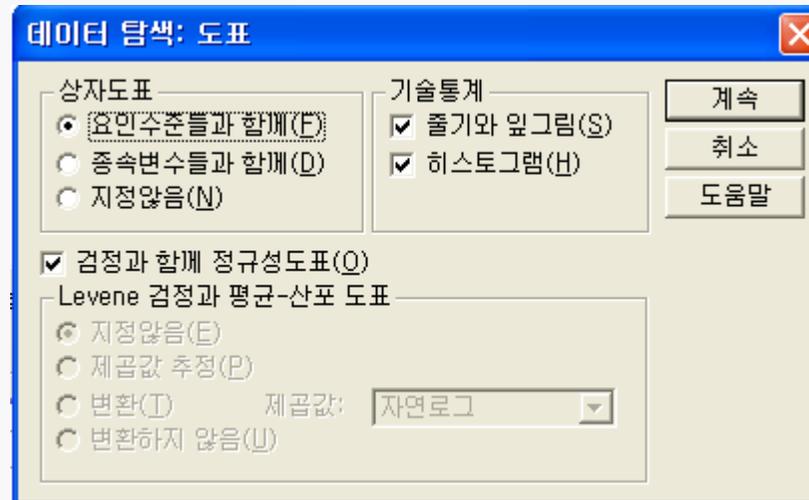
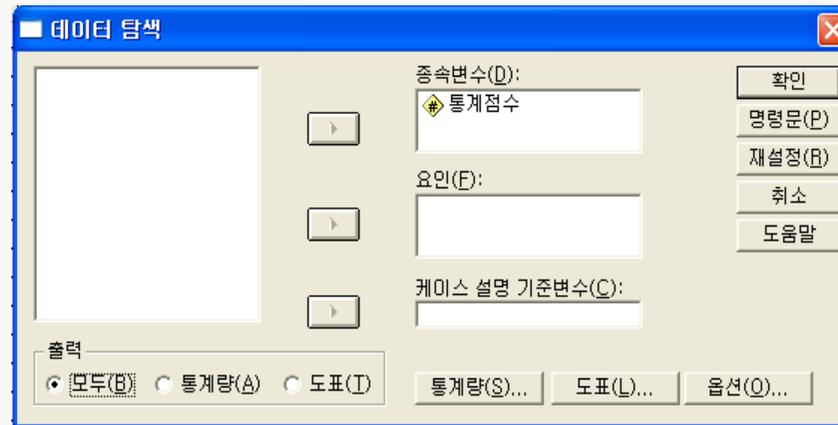
- [선형]

- ◆ 종속변수 : 중간고사
- ◆ 독립변수 : 입학성적



## 탐색적 자료분석:

자료의 대략적으로 분포의 형태를 알아볼 경우에 사용한다.

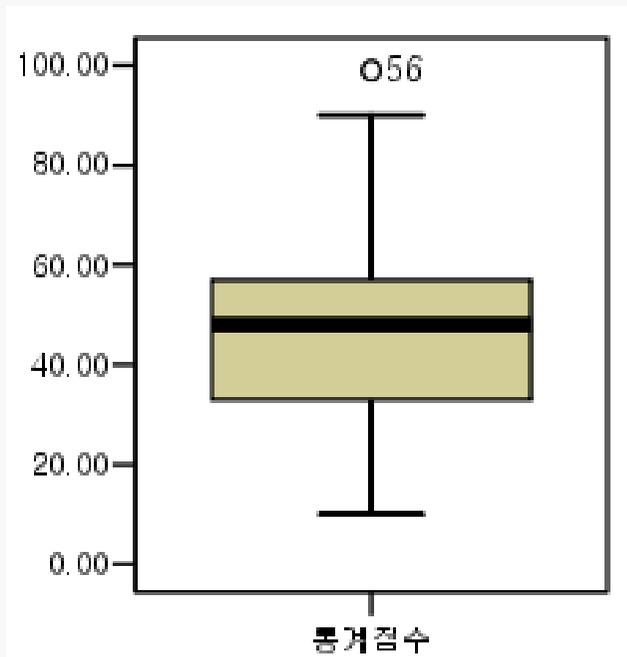


정규성 검정

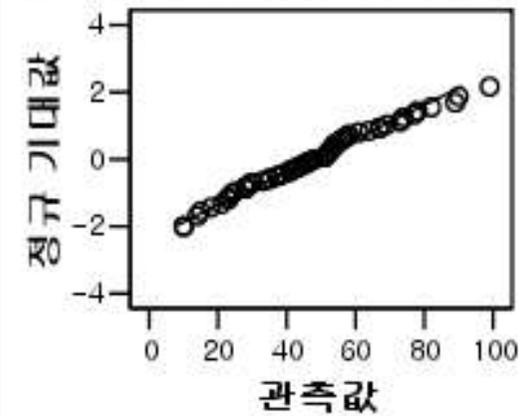
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
통계점수	,084	65	,200*	,982	65	,463

\*. 이것은 참인 유의확률의 하한값입니다.

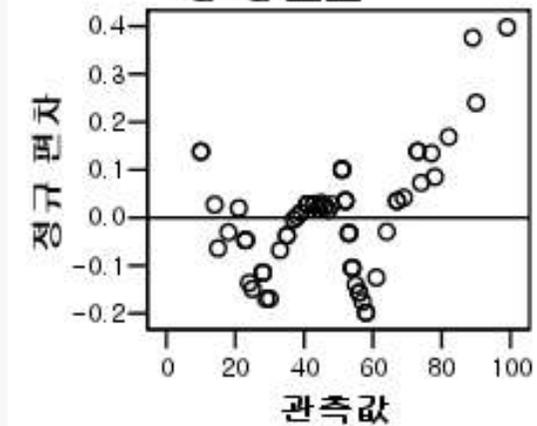
a. Lilliefors 유의확률 수정



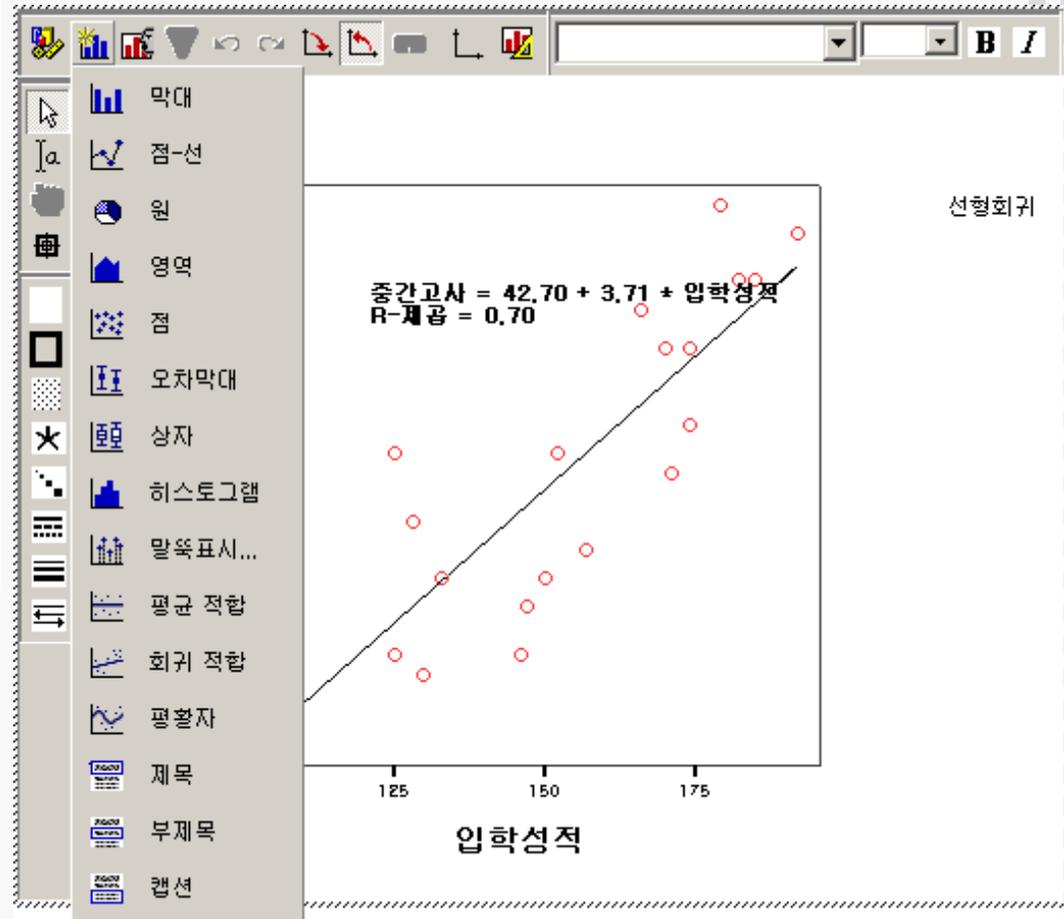
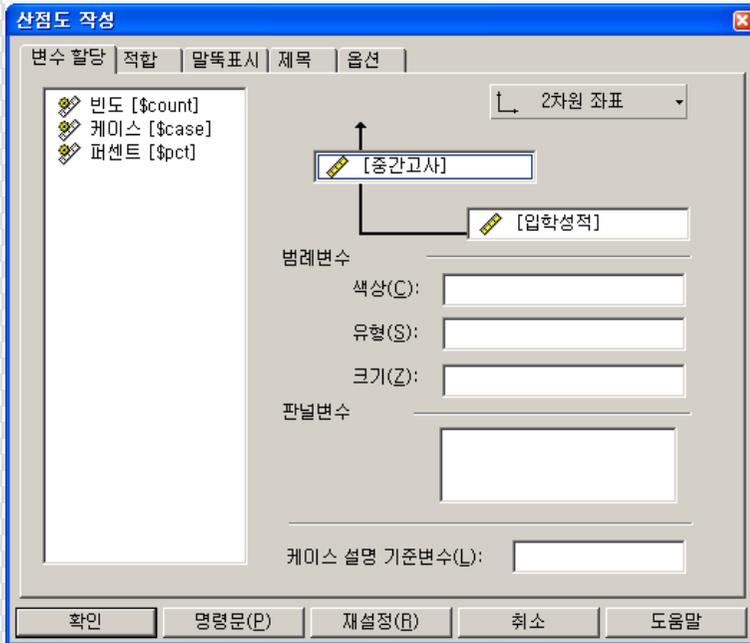
통계점수의 정규 Q-Q ...



통계점수의 기울기없는 정규 Q-Q 도표



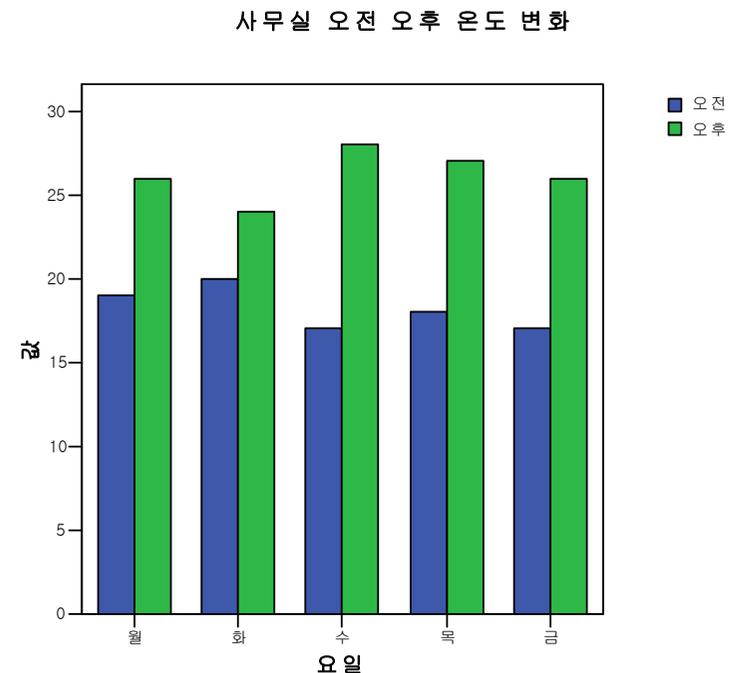
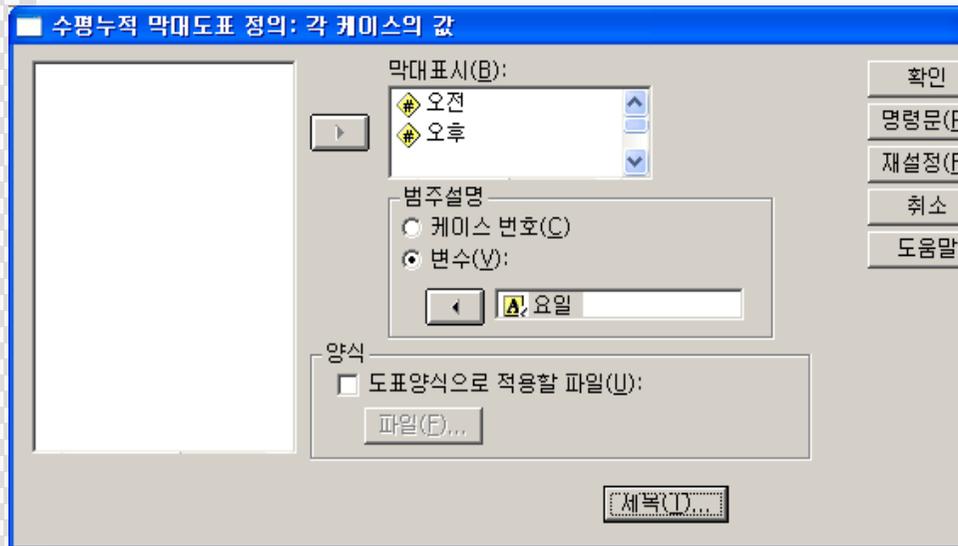
# 그래프(PLOT)



- ◆ [그래프]-[산점도]
- ◆ 단순
- ◆ 변수들간의 직선 관계 체크

# 그래프(Bar)

- ◇ [그래프]-[막대도표]
  - ◆ 단순 선택
  - ◆ 막대 표시 입력 : 오전,오후
  - ◆ 범주 설명 입력 : 요일
  - ◆ 제목 입력





감사합니다