질병관리청

SPSS
복합표본설계 선형회귀분석

건강영양조사분석과

목 차

1. 복합표본설계 자료 분석을 위한 준비

2. 복합표본설계 기술 통계

3. 복합표본설계 선형회귀분석

01.
복합표본설계 자료 분석을 위한 준비

- 분석 변수 생성
- 복합표본설계 자료 분석 시 유의사항

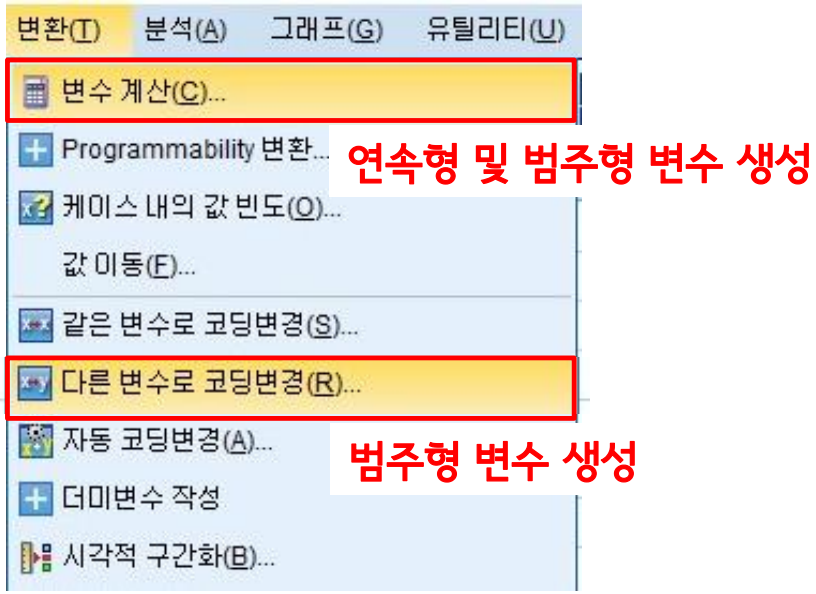
변수 생성_생성 방법

▪연구자의 필요에 따라 새로운 변수 생성 가능

- 새로운 연속형 변수 계산, 연속형 변수의 범주화, 범주형 변수의 코딩 변경 등

▪분석 변수 생성 방법

1. 대화상자 이용



2. 명령문 이용

변수 생성_예제 변수 설명

1. "변수 계산" 이용

변수명	설명	내용
BMI	체질량지수	연속형
sm_g	흡연 여부	1. 비흡연
		2. 과거흡연
		3. 현재흡연
HP	고혈압 유병 여부	1. 고혈압 유병 O (고혈압)
		0. 고혈압 유병 X (정상, 고혈압 전단계)
subg	부-모집단	1. 30세 이상 남자
		0. 그 외

2. "다른 변수로 코딩변경" 이용

변수	설명	내용
age30	30세 이상 여부	1. 30세 이상
		0. 30세 미만
age_g	10세 단위 연령 구분	3. 30대(30~39세)
		4. 40대(40~49세)
		5. 50대(50~59세)
		6. 60대(60~69세)
		7. 70세 이상(70세 이상)
		0. 그 외

변수 계산_연속형 변수 생성 예제

예제 연속형 변수 BMI(체질량지수) 생성하기

변수 정의 BMI(체질량지수)=체중(kg)/신장^2(m^2)

만성질환	주요지표	지표정의 및 프로그램	
비만	유병 여부 (체질량지수 기준)	지표정의	■BMI(체질량지수): 체중(kg)/신장^2(m^2) ■저체중(BMI<18.5), 정상(18.5≤ BMI< 25), 비만(25≤ BMI)
		프로그램	IF HE_dprg=. & HE_wt^=. & HE_ht^=. THEN HE_BMI = HE_wt / ((HE_ht*0.01)**2) ; IF 0< HE_BMI< 18.5 THEN HE_obe=1 ; /*저체중*/ ELSE IF 18.5<= HE_BMI< 25.0 THEN HE_obe=2 ; /*정상*/ ELSE IF 25.0<= HE_BMI THEN HE_obe=3 ; /*비만*/

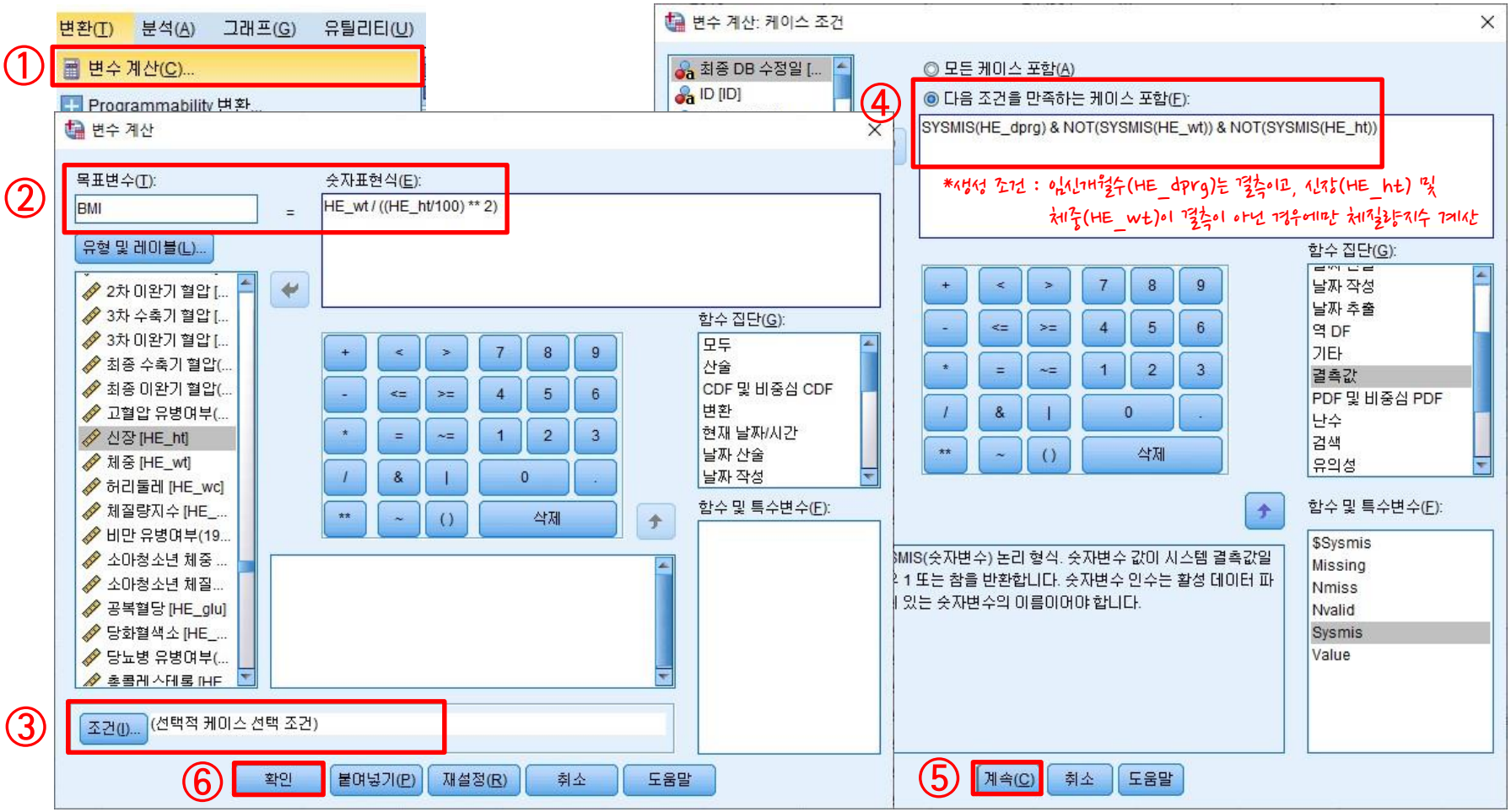
*생성 조건 : 임신개월수(HE_dprg)는 결측이고, 신장(HE_ht) 및 체중(HE_wt)이 결측이 아닌 경우에만 체질량지수 계산

생성 방법 [변환] – [변수 계산] 대화상자 이용

변수 계산_연속형 변수 생성 예제

예제 연속형 변수 BMI(체질량지수) 변수 생성하기

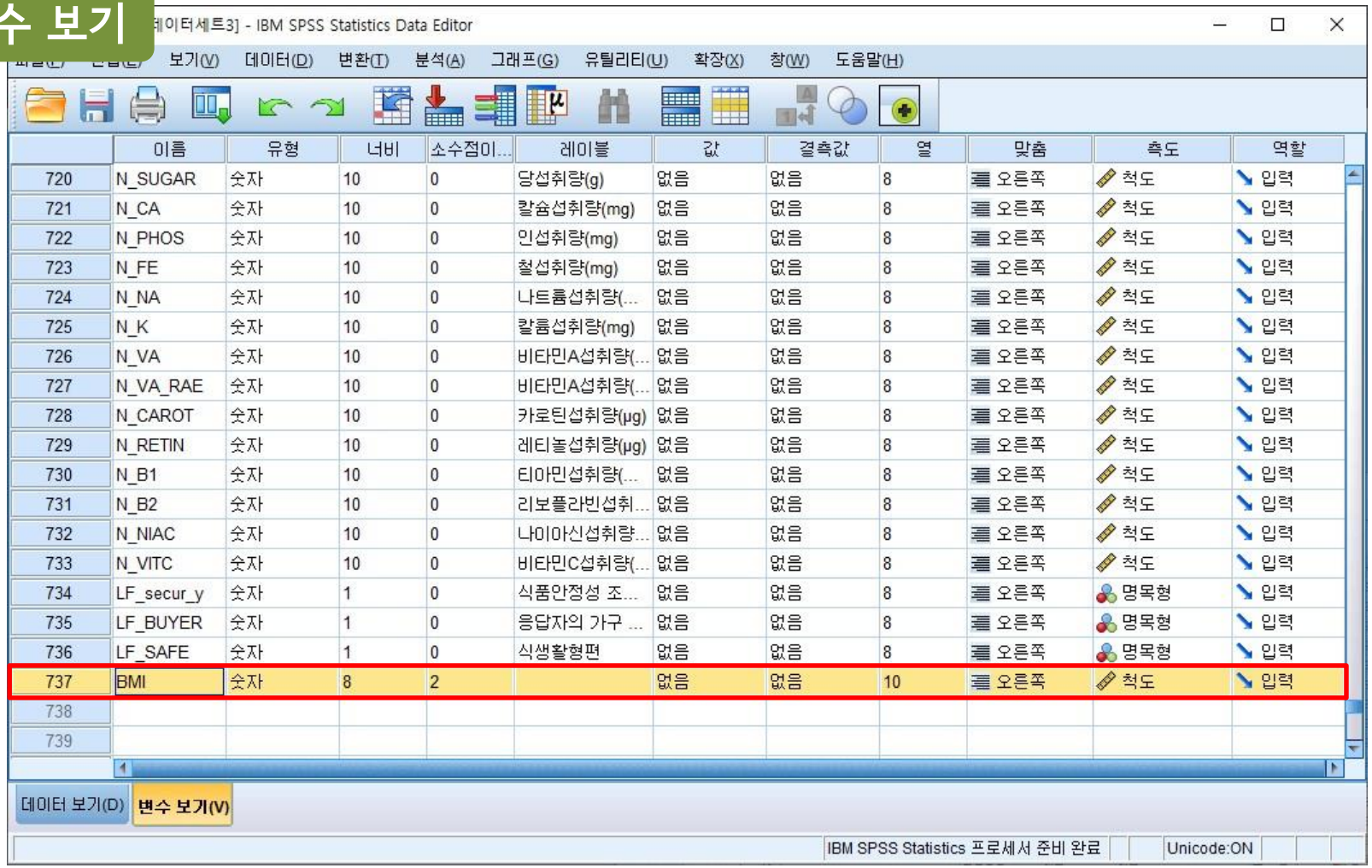
변환 ▶ 변수 계산 ▶ 목표변수(생성 변수명) 및 숫자표현식(변수계산식) 입력 ▶ 조건 지정



변수 계산_연속형 변수 생성 결과

예제 연속형 변수 BMI(체질량지수) 변수 생성하기

변수 보기

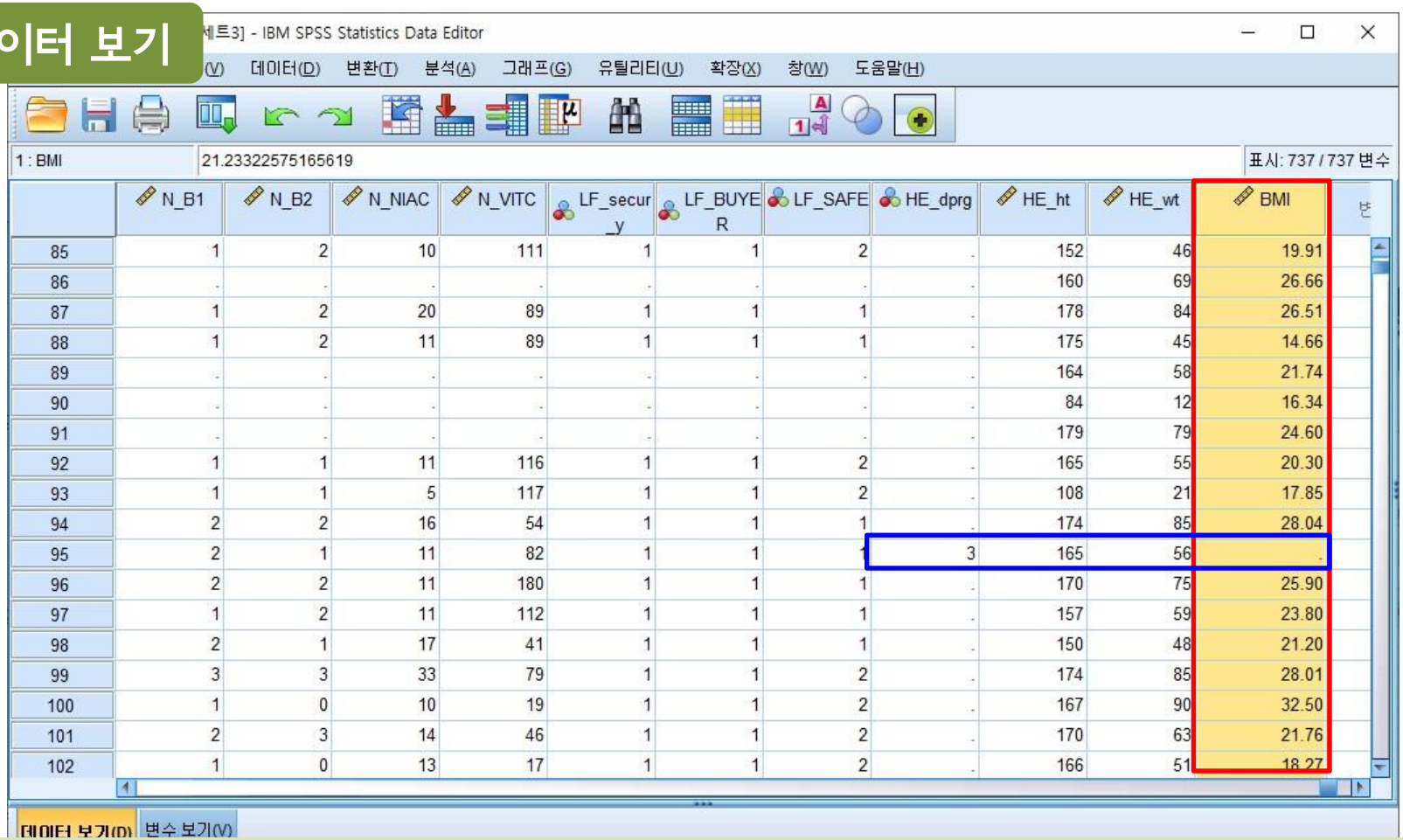


	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도	역할
720	N_SUGAR	숫자	10	0	당섭취량(g)	없음	없음	8	오른쪽	척도	입력
721	N_CA	숫자	10	0	칼슘섭취량(mg)	없음	없음	8	오른쪽	척도	입력
722	N_PHOS	숫자	10	0	인섭취량(mg)	없음	없음	8	오른쪽	척도	입력
723	N_FE	숫자	10	0	철섭취량(mg)	없음	없음	8	오른쪽	척도	입력
724	N_NA	숫자	10	0	나트륨섭취량(...	없음	없음	8	오른쪽	척도	입력
725	N_K	숫자	10	0	칼륨섭취량(mg)	없음	없음	8	오른쪽	척도	입력
726	N_VA	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
727	N_VA_RAE	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
728	N_CAROT	숫자	10	0	카로틴섭취량(μg)	없음	없음	8	오른쪽	척도	입력
729	N_RETIN	숫자	10	0	레티놀섭취량(μg)	없음	없음	8	오른쪽	척도	입력
730	N_B1	숫자	10	0	티아민섭취량(...	없음	없음	8	오른쪽	척도	입력
731	N_B2	숫자	10	0	리보플라빈섭취...	없음	없음	8	오른쪽	척도	입력
732	N_NIAC	숫자	10	0	나이아신섭취량...	없음	없음	8	오른쪽	척도	입력
733	N_VITC	숫자	10	0	비타민C섭취량(...	없음	없음	8	오른쪽	척도	입력
734	LF_secur_y	숫자	1	0	식품안정성 조...	없음	없음	8	오른쪽	명목형	입력
735	LF_BUYER	숫자	1	0	응답자의 가구 ...	없음	없음	8	오른쪽	명목형	입력
736	LF_SAFE	숫자	1	0	식생활형편	없음	없음	8	오른쪽	명목형	입력
737	BMI	숫자	8	2		없음	없음	10	오른쪽	척도	입력
738											
739											

변수 계산_연속형 변수 생성 결과

예제 연속형 변수 BMI(체질량지수) 변수 생성하기

데이터 보기



	N_B1	N_B2	N_NIAC	N_VITC	LF_secur_y	LF_BUYER	LF_SAFE	HE_dprg	HE_ht	HE_wt	BMI
85	1	2	10	111	1	1	2	.	152	46	19.91
86	.	.	.	.	.	.	.	.	160	69	26.66
87	1	2	20	89	1	1	1	.	178	84	26.51
88	1	2	11	89	1	1	1	.	175	45	14.66
89	.	.	.	.	.	.	.	.	164	58	21.74
90	.	.	.	.	.	.	.	.	84	12	16.34
91	.	.	.	.	.	.	.	.	179	79	24.60
92	1	1	11	116	1	1	2	.	165	55	20.30
93	1	1	5	117	1	1	2	.	108	21	17.85
94	2	2	16	54	1	1	1	.	174	85	28.04
95	2	1	11	82	1	1	1	3	165	56	.
96	2	2	11	180	1	1	1	.	170	75	25.90
97	1	2	11	112	1	1	1	.	157	59	23.80
98	2	1	17	41	1	1	1	.	150	48	21.20
99	3	3	33	79	1	1	2	.	174	85	28.01
100	1	0	10	19	1	1	2	.	167	90	32.50
101	2	3	14	46	1	1	2	.	170	63	21.76
102	1	0	13	17	1	1	2	.	166	51	18.27

임신개월수(HE_dprg)는 결측이고, 신장(HE_ht) 및 체중(HE_wt)이 결측이 아닌 경우에만 체질량지수 계산

변수 계산_범주형 변수 생성 예제

예제 부-모집단(subg) 변수 생성하기

변수 정의 만 30세 이상 남자

변수명	설명	내용
subg	부-모집단	1. 30세 이상 남자
		0. 그 외

생성 방법 [변환] – [변수 계산] 대화상자 이용

변수 계산_범주형 변수 생성 예제_1차

예제 부-모집단(subg) 변수 생성하기

1차 입력 : 변환 ▶ 변수 계산 ▶ 목표변수(생성 변수명) 및 숫자표현식(변수값) 입력 ▶ 조건 지정

변수 계산_범주형 변수 생성 결과_1차

예제 부-모집단(subg) 변수 생성하기

변수 보기 : 1차 입력

tatistics Data Editor

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

	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도	역할
721	N_CA	숫자	10	0	칼슘섭취량(mg)	없음	없음	8	오른쪽	척도	입력
722	N_PHOS	숫자	10	0	인섭취량(mg)	없음	없음	8	오른쪽	척도	입력
723	N_FE	숫자	10	0	철섭취량(mg)	없음	없음	8	오른쪽	척도	입력
724	N_NA	숫자	10	0	나트륨섭취량(...	없음	없음	8	오른쪽	척도	입력
725	N_K	숫자	10	0	칼륨섭취량(mg)	없음	없음	8	오른쪽	척도	입력
726	N_VA	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
727	N_VA_RAE	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
728	N_CAROT	숫자	10	0	카로틴섭취량(μg)	없음	없음	8	오른쪽	척도	입력
729	N_RETIN	숫자	10	0	레티놀섭취량(μg)	없음	없음	8	오른쪽	척도	입력
730	N_B1	숫자	10	0	티아민섭취량(...	없음	없음	8	오른쪽	척도	입력
731	N_B2	숫자	10	0	리보플라빈섭취...	없음	없음	8	오른쪽	척도	입력
732	N_NIAC	숫자	10	0	나이아신섭취량...	없음	없음	8	오른쪽	척도	입력
733	N_VITC	숫자	10	0	비타민C섭취량(...	없음	없음	8	오른쪽	척도	입력
734	LF_secur_y	숫자	1	0	식품안정성 조...	없음	없음	8	오른쪽	명목형	입력
735	LF_BUYER	숫자	1	0	응답자의 가구 ...	없음	없음	8	오른쪽	명목형	입력
736	LF_SAFE	숫자	1	0	식생활형편	없음	없음	8	오른쪽	명목형	입력
737	BMI	숫자	8	2		없음	없음	10	오른쪽	척도	입력
738	subg	숫자	8	2		없음	없음	10	오른쪽	명목형	입력
739											
740											

데이터 보기(D) 변수 보기(V)

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON

변수 계산_범주형 변수 생성 결과_1차

예제 부-모집단(subg) 변수 생성하기

데이터 보기 : 1차 입력

cs Data Editor

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

1 : subg 1.00 표시: 738 / 738 변수

	N_B1	N_B2	N_NIAC	N_VITC	LF_secur_y	LF_BUYER	LF_SAFE	BMI	sex	age	subg
1	1	1	5	6	1	2	1	21.23	1	80	1.00
2	1	1	17	109	1	1	1	27.17	1	78	1.00
3	1	1	5	97	1	1	1	21.39	2	75	.
4	.	.	.	.	.	.	.	23.46	2	43	.
5	1	0	6	11	1	1	1	26.98	2	80	.
6	1	1	7	37	1	1	1	21.53	2	71	.
7	1	1	12	54	1	1	1	26.42	1	44	1.00
8	0	0	2	6	1	1	1	19.02	2	42	.
9	1	2	5	10	1	1	1	18.61	2	13	.
10	1	1	6	82	1	1	1	15.78	1	11	.
11	.	.	.	.	.	.	.	23.16	1	36	1.00
12	1	1	10	12	1	1	2	23.09	2	30	.
13	1	1	7	71	1	1	2	14.76	1	2	.
14	2	2	20	85	1	1	1	30.00	1	57	1.00
15	1	2	12	30	1	1	1	21.73	2	54	.
16	2	2	21	68	1	1	1	35.70	1	27	.
17	1	1	7	33	1	1	1	27.02	1	26	.
18	.	.	.	.	.	.	.	24.41	1	68	1.00

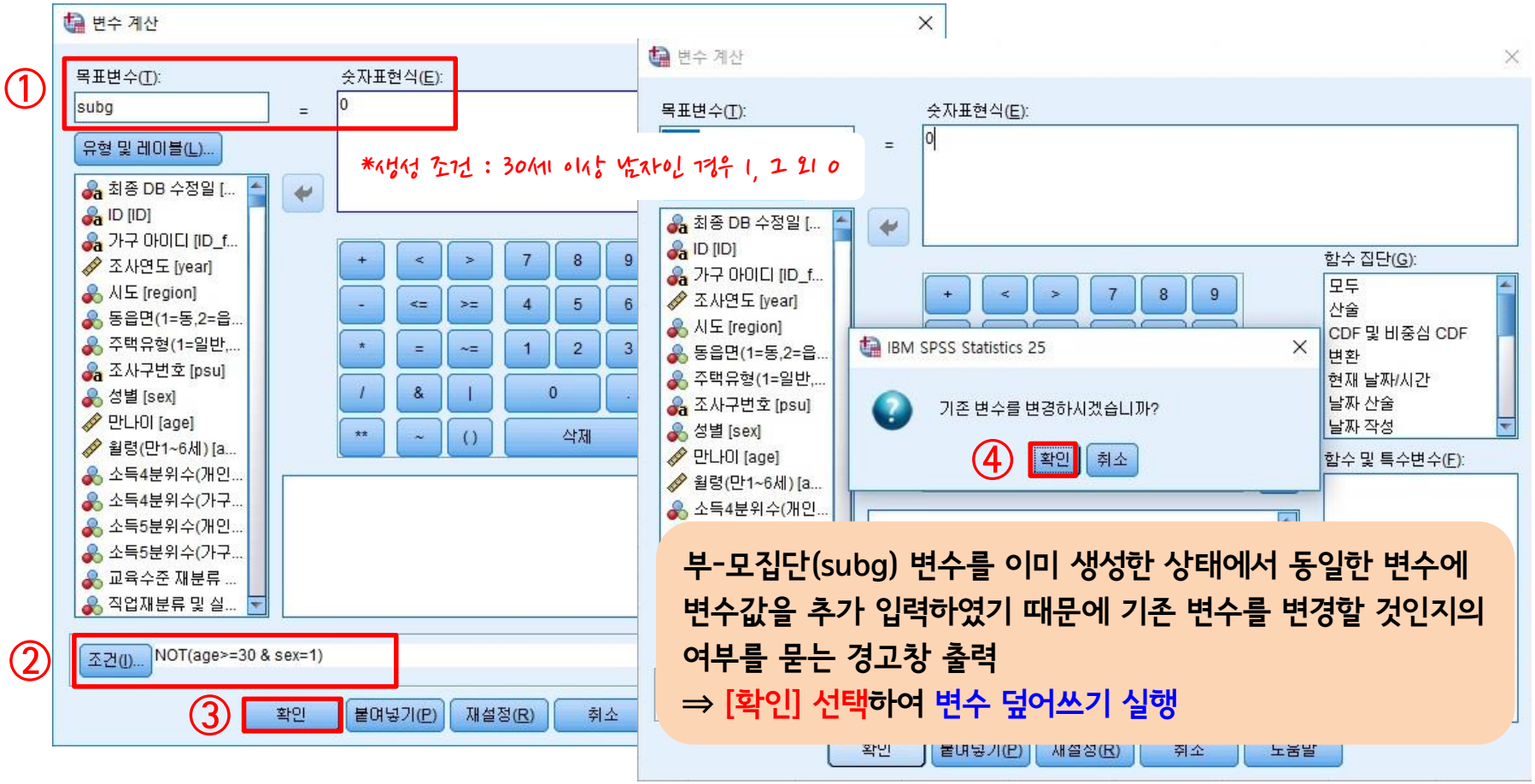
데이터 보기(D) 변수 보기(V)

30세 이상 남자가 아닌 경우 결측(.) 처리되어 변수값 0으로 입력 필요 ⇒ [변수계산] 재실행

변수 계산_범주형 변수 생성 예제_2차

예제 부-모집단(subg) 변수 생성하기

2차 입력 : 변환 ▶ 변수 계산 ▶ 목표변수(생성 변수명) 및 숫자표현식(변수값) 재입력 ▶ 조건 재지정 ▶ 기존 변수 변경 여부 [확인] 선택



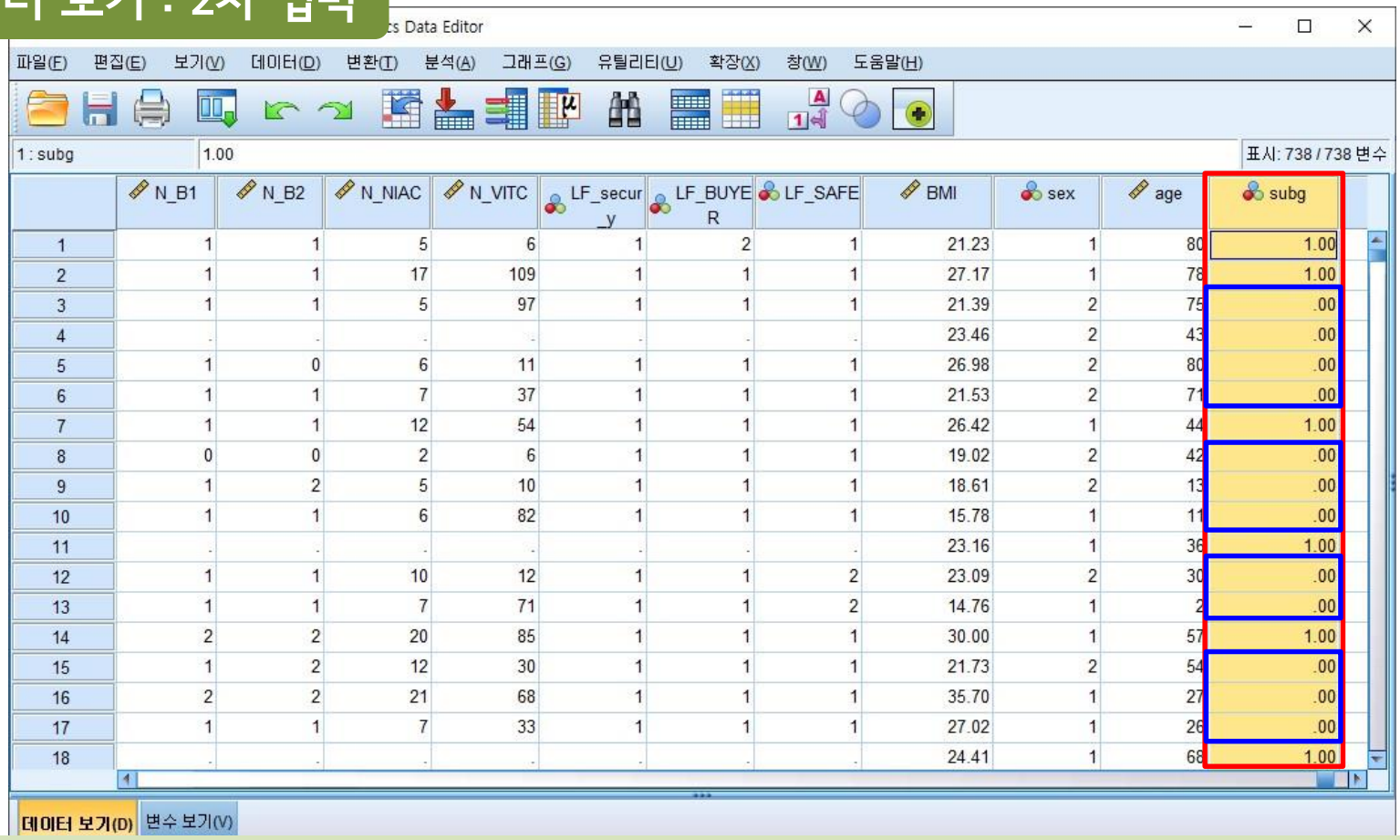
부-모집단(subg) 변수를 이미 생성한 상태에서 동일한 변수에 변수값을 추가 입력하였기 때문에 기존 변수를 변경할 것인지의 여부를 묻는 경고창 출력

⇒ [확인] 선택하여 변수 덮어쓰기 실행

변수 계산_범주형 변수 생성 결과_2차

예제 부-모집단(subg) 변수 생성하기

데이터 보기 : 2차 입력



	N_B1	N_B2	N_NIAC	N_VITC	LF_secur_y	LF_BUYER	LF_SAFE	BMI	sex	age	subg
1	1	1	5	6	1	2	1	21.23	1	80	1.00
2	1	1	17	109	1	1	1	27.17	1	78	1.00
3	1	1	5	97	1	1	1	21.39	2	75	.00
4	.	.	.	.	.	.	.	23.46	2	43	.00
5	1	0	6	11	1	1	1	26.98	2	80	.00
6	1	1	7	37	1	1	1	21.53	2	71	.00
7	1	1	12	54	1	1	1	26.42	1	44	1.00
8	0	0	2	6	1	1	1	19.02	2	42	.00
9	1	2	5	10	1	1	1	18.61	2	13	.00
10	1	1	6	82	1	1	1	15.78	1	11	.00
11	.	.	.	.	.	.	.	23.16	1	36	1.00
12	1	1	10	12	1	1	2	23.09	2	30	.00
13	1	1	7	71	1	1	2	14.76	1	2	.00
14	2	2	20	85	1	1	1	30.00	1	57	1.00
15	1	2	12	30	1	1	1	21.73	2	54	.00
16	2	2	21	68	1	1	1	35.70	1	27	.00
17	1	1	7	33	1	1	1	27.02	1	26	.00
18	.	.	.	.	.	.	.	24.41	1	68	1.00

30세 이상 남자가 아닌 경우 결측(.) 이 아닌 0으로 입력 완료

다른 변수로 코딩변경_범주형 변수 생성 예제

예제 연령그룹(age_g) 변수 생성하기

변수 정의 만 30세 이상 10세 단위 연령 그룹

	age(기존변수)
30대	30, 31, ··· , 39
40대	40, 41, ··· , 49
50대	50, 51, ··· , 59
60대	60, 61, ··· , 69
70세 이상	70, 71, ··· , 80
그 외	1, 2, ··· , 29
결측값	.

⟹

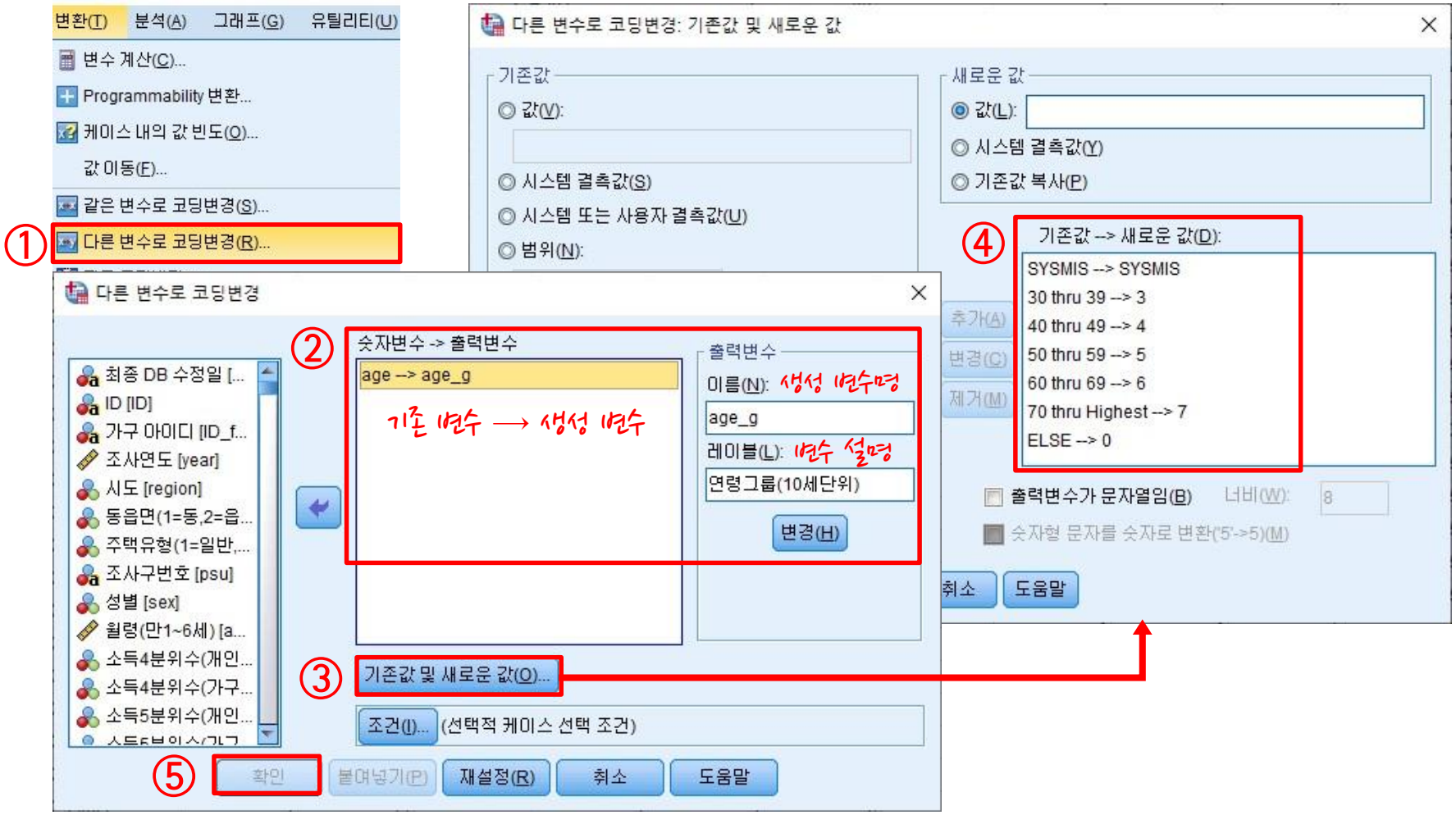
age_g(생성변수)
3
4
5
6
7
0
.

생성 방법 [변환] – [다른 변수로 코딩변경] 대화상자 이용

다른 변수로 코딩변경_범주형 변수 생성 예제

예제 연령그룹(age_g) 변수 생성하기

변환 ▶ 다른 변수로 코딩변경 ▶ 변수 선택 ▶ 출력변수 지정 ▶ 기존값 및 새로운 값 지정



다른 변수로 코딩변경_범주형 변수 생성 결과

예제 연령그룹(age_g) 변수 생성하기

변수 보기

데이터세트1] - IBM SPSS Statistics Data Editor

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

	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도	역할
722	N_PHOS	숫자	10	0	인섭취량(mg)	없음	없음	8	오른쪽	척도	입력
723	N_FE	숫자	10	0	철섭취량(mg)	없음	없음	8	오른쪽	척도	입력
724	N_NA	숫자	10	0	나트륨섭취량(...	없음	없음	8	오른쪽	척도	입력
725	N_K	숫자	10	0	칼륨섭취량(mg)	없음	없음	8	오른쪽	척도	입력
726	N_VA	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
727	N_VA_RAE	숫자	10	0	비타민A섭취량(...	없음	없음	8	오른쪽	척도	입력
728	N_CAROT	숫자	10	0	카로틴섭취량(μg)	없음	없음	8	오른쪽	척도	입력
729	N_RETIN	숫자	10	0	레티놀섭취량(μg)	없음	없음	8	오른쪽	척도	입력
730	N_B1	숫자	10	0	티아민섭취량(...	없음	없음	8	오른쪽	척도	입력
731	N_B2	숫자	10	0	리보플라빈섭취...	없음	없음	8	오른쪽	척도	입력
732	N_NIAC	숫자	10	0	나이아신섭취량...	없음	없음	8	오른쪽	척도	입력
733	N_VITC	숫자	10	0	비타민C섭취량(...	없음	없음	8	오른쪽	척도	입력
734	LF_secur_y	숫자	1	0	식품안정성 조...	없음	없음	8	오른쪽	명목형	입력
735	LF_BUYER	숫자	1	0	응답자의 가구 ...	없음	없음	8	오른쪽	명목형	입력
736	LF_SAFE	숫자	1	0	식생활형편	없음	없음	8	오른쪽	명목형	입력
737	BMI	숫자	8	2		없음	없음	10	오른쪽	척도	입력
738	subg	숫자	8	2		없음	없음	10	오른쪽	명목형	입력
739	age_g	숫자	8	2	연령그룹(10세...	없음	없음	10	오른쪽	명목형	입력
740											
741											

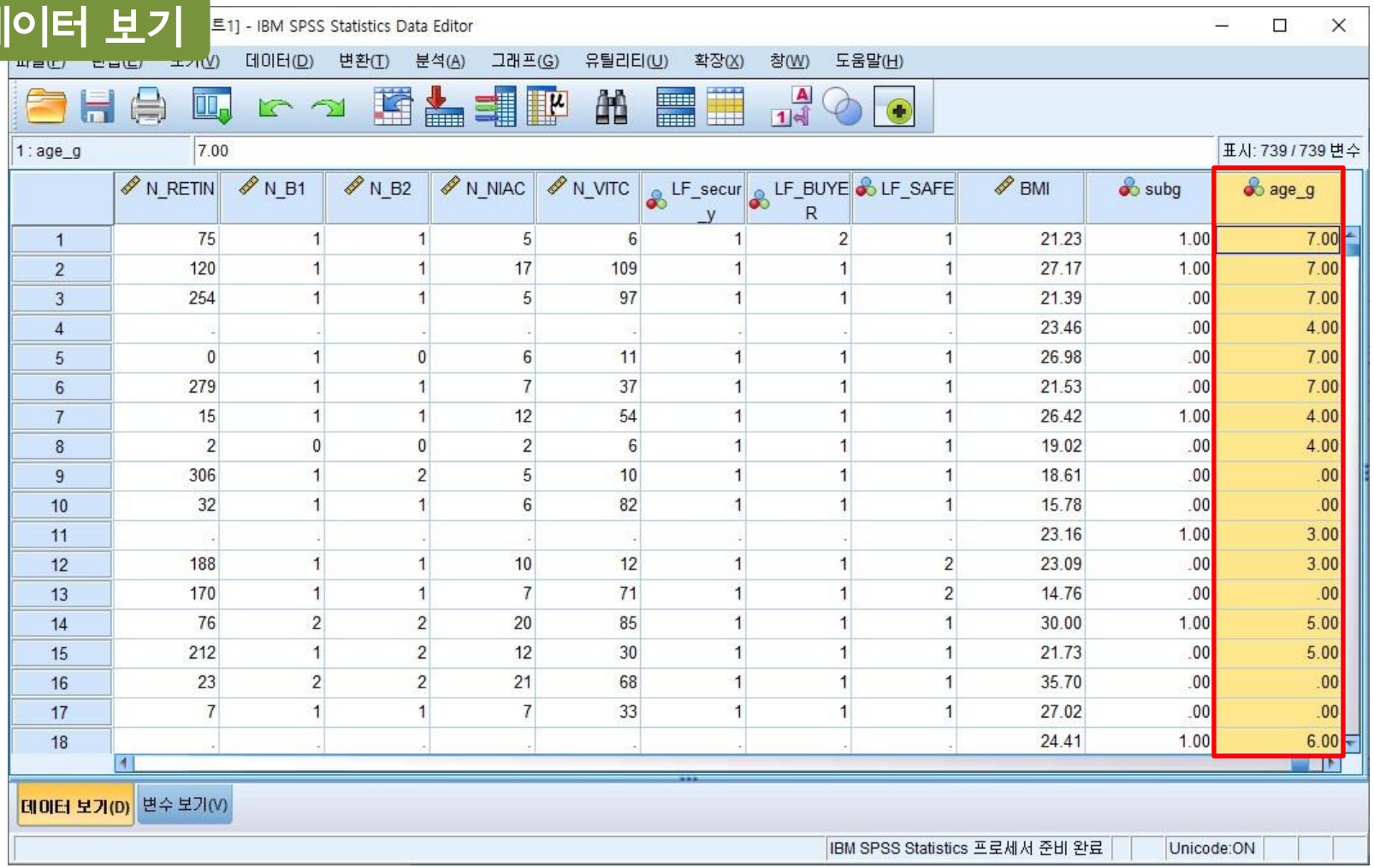
데이터 보기(D) 변수 보기(V)

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON

다른 변수로 코딩변경_범주형 변수 생성 결과

예제 연령그룹(age_g) 변수 생성하기

데이터 보기



	N_RETIN	N_B1	N_B2	N_NIAC	N_VITC	LF_secur_y	LF_BUYER	LF_SAFE	BMI	subg	age_g
1	75	1	1	5	6	1	2	1	21.23	1.00	7.00
2	120	1	1	17	109	1	1	1	27.17	1.00	7.00
3	254	1	1	5	97	1	1	1	21.39	.00	7.00
4	.	.	.	.	.	.	.	.	23.46	.00	4.00
5	0	1	0	6	11	1	1	1	26.98	.00	7.00
6	279	1	1	7	37	1	1	1	21.53	.00	7.00
7	15	1	1	12	54	1	1	1	26.42	1.00	4.00
8	2	0	0	2	6	1	1	1	19.02	.00	4.00
9	306	1	2	5	10	1	1	1	18.61	.00	.00
10	32	1	1	6	82	1	1	1	15.78	.00	.00
11	.	.	.	.	.	.	.	.	23.16	1.00	3.00
12	188	1	1	10	12	1	1	2	23.09	.00	3.00
13	170	1	1	7	71	1	1	2	14.76	.00	.00
14	76	2	2	20	85	1	1	1	30.00	1.00	5.00
15	212	1	2	12	30	1	1	1	21.73	.00	5.00
16	23	2	2	21	68	1	1	1	35.70	.00	.00
17	7	1	1	7	33	1	1	1	27.02	.00	.00
18	.	.	.	.	.	.	.	.	24.41	1.00	6.00

변수 생성 후 확인사항

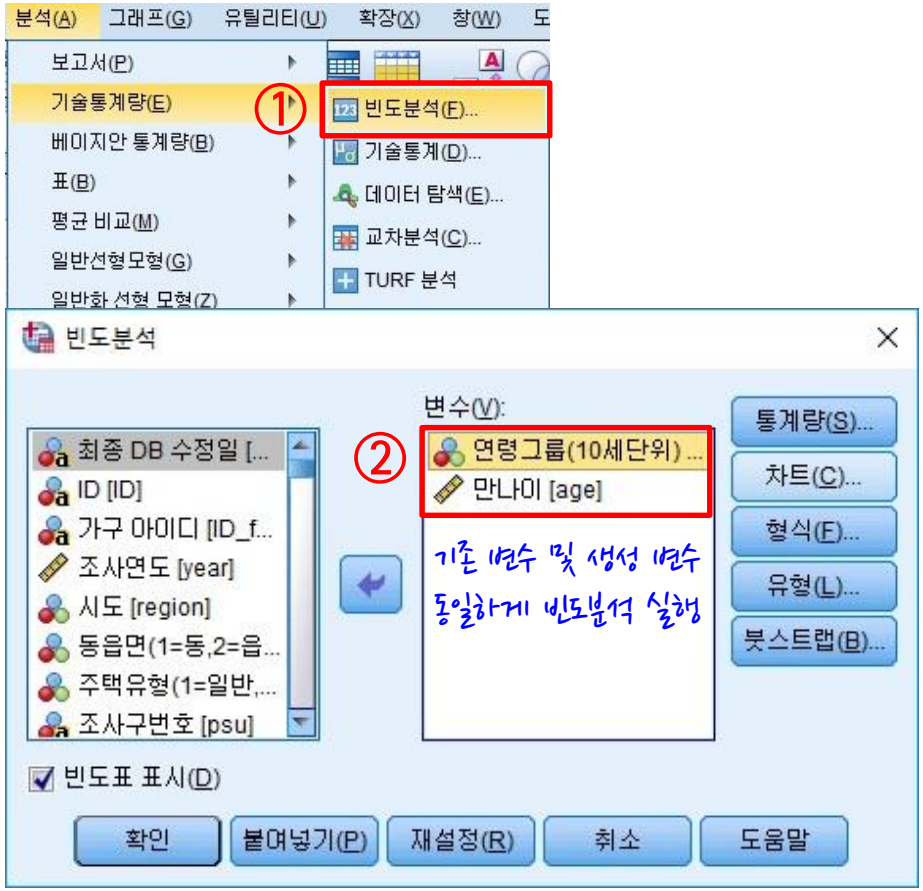
변수 생성 후 확인사항

1. 빈도분석 실시하여 생성하고자 하는 변수가 적절히 생성되었는지 범주 및 빈도 확인 필요

2. 기존 변수를 다른 변수로 코딩 변경한 경우에는 기존 변수와 생성 변수의 범주 및 빈도 비교 필요

생성 변수 범주 및 빈도 확인 예제

예제 연령그룹(age_g) 변수와 연령(age) 변수 빈도분석



<생성변수 age_g 빈도분석 결과표>

통계량

연령그룹(10세단위)

N	유효	7992
	결측	0

연령그룹(10세단위)

		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유효	.00	2317	29.0	29.0	29.0
	3.00	962	12.0	12.0	41.0
	4.00	1191	14.9	14.9	55.9
	5.00	1234	15.4	15.4	71.4
	6.00	1140	14.3	14.3	85.6
	7.00	1148	14.4	14.4	100.0
	전체	7992	100.0	100.0	

<기존변수 age 빈도분석 결과표 일부>

통계량

만나이

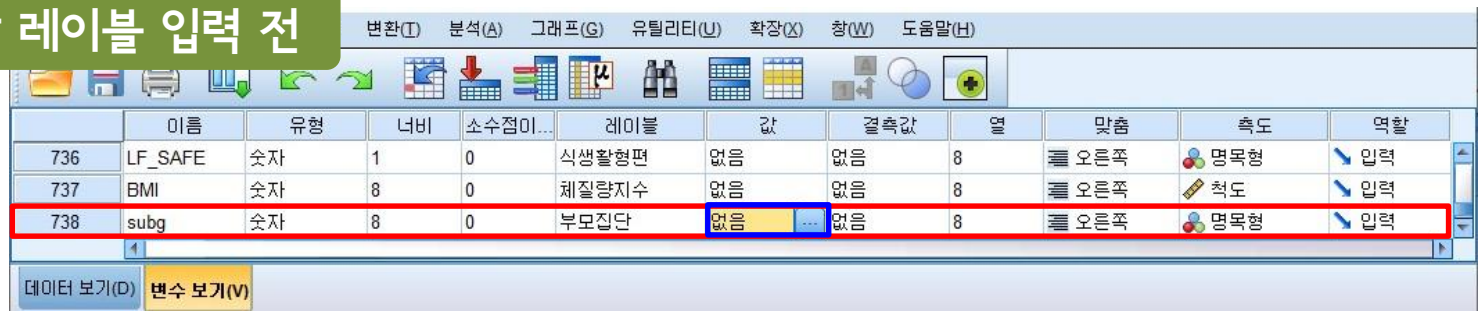
N	유효	7992
	결측	0

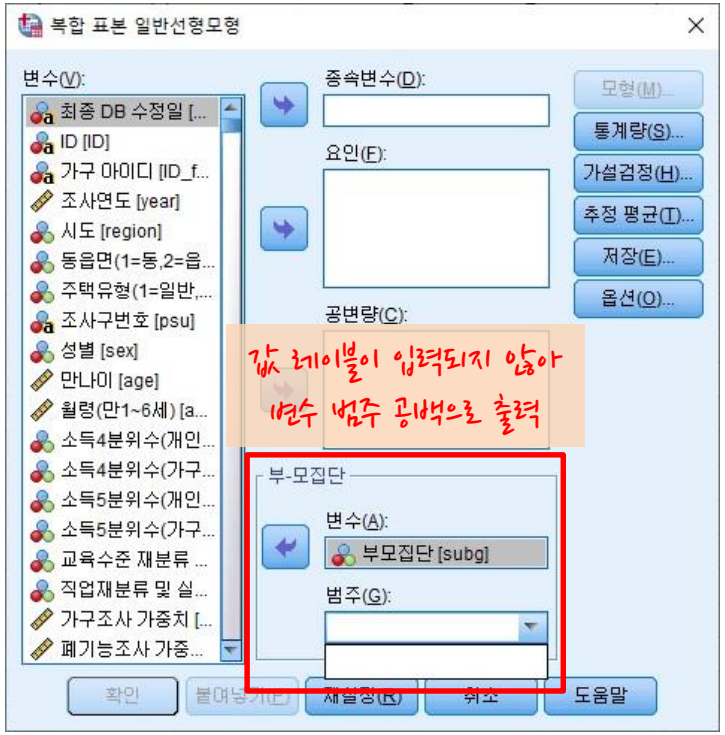
30	61	.8	.8	29.8
31	67	.8	.8	30.6
32	82	1.0	1.0	31.6
33	74	.9	.9	32.5
34	104	1.3	1.3	33.8
35	109	1.4	1.4	35.2
36	132	1.7	1.7	36.9
37	102	1.3	1.3	38.1
38	112	1.4	1.4	39.5
39	119	1.5	1.5	41.0
합계	962	12.0		

변수 생성 시 참고사항

- 변수 생성 후 [변수 보기] 창에서 **값 레이블을 입력하면 분석 시 편리**

값 레이블 입력 전





일변량 통계량

		B 추정값	표준오차
평균	최종 수축기 혈압(2,3차 평균)	116.81	.344

부-모집단 기술통계

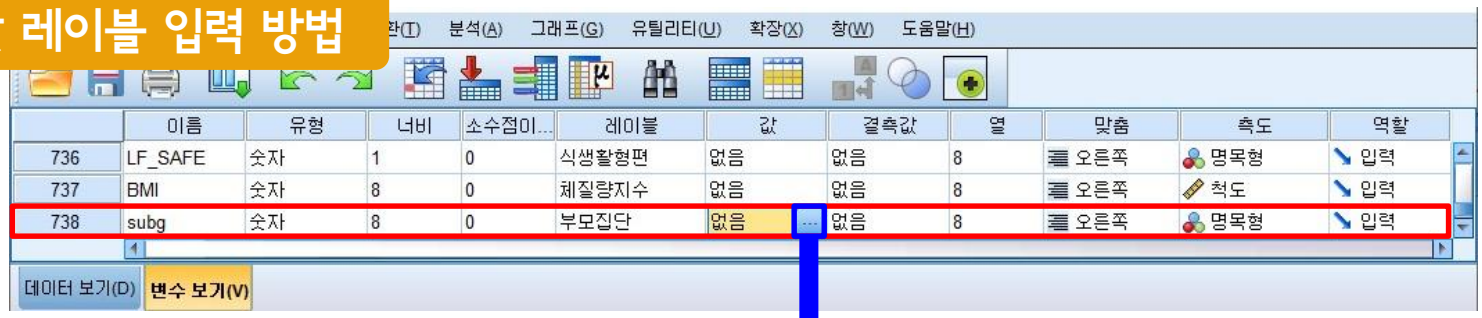
일변량 통계량

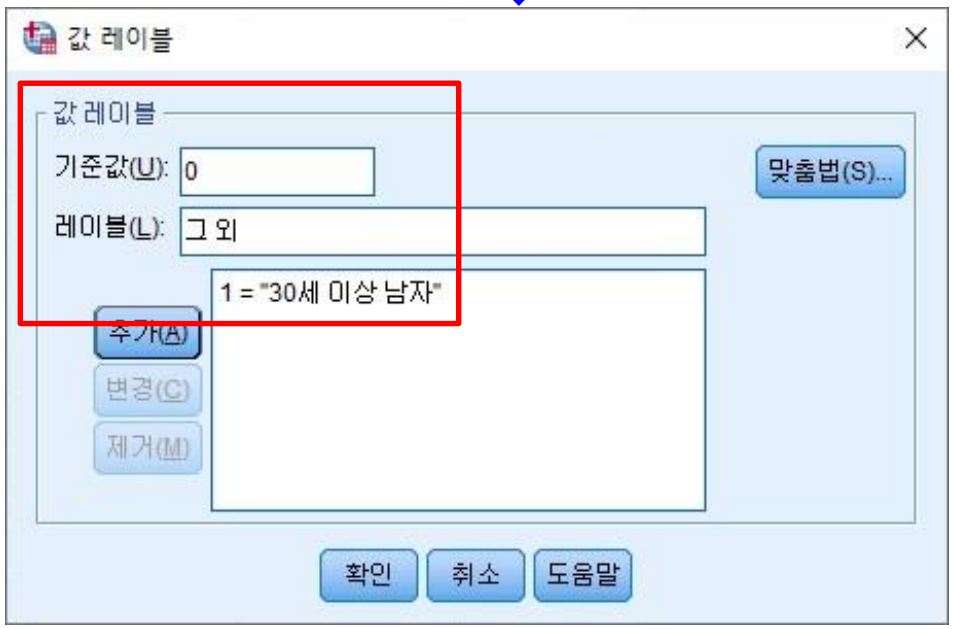
subg			B 추정값	표준오차
.00	평균	최종 수축기 혈압(2,3차 평균)	114.51	.373
1.00	평균	최종 수축기 혈압(2,3차 평균)	120.79	.470

변수 생성 시 참고사항

- 변수 생성 후 [변수 보기] 창에서 **값 레이블을 입력하면 분석 시 편리**

값 레이블 입력 방법

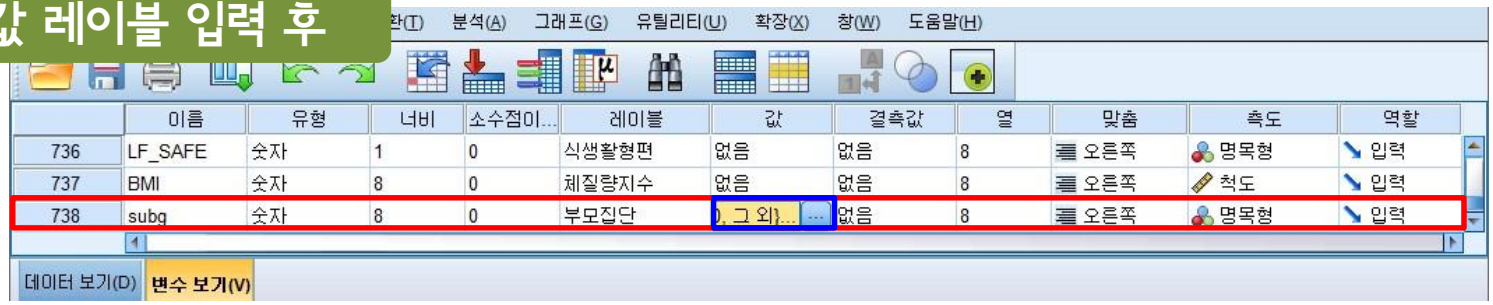


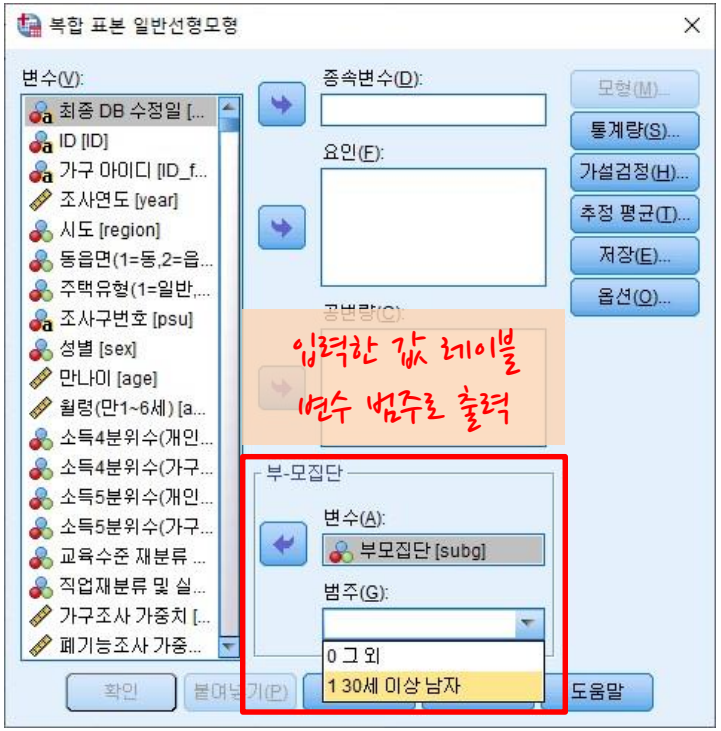


변수 생성 시 참고사항

- 변수 생성 후 [변수 보기] 창에서 **값 레이블을 입력하면 분석 시 편리**

값 레이블 입력 후





복합 표본: 기술통계

일변량 통계량

		B 추정값	표준오차
평균	최종 수축기 혈압(2,3차 평균)	116.81	.344

부-모집단 기술통계

일변량 통계량

subg			B 추정값	표준오차
그 외	평균	최종 수축기 혈압(2,3차 평균)	114.51	.373
30세 이상 남자	평균	최종 수축기 혈압(2,3차 평균)	120.79	.470

복합표본설계 자료 분석 시 유의사항(1)

- **관심 집단에 대한 분석 시에도 분석 DB는 항상 전체 자료 사용**
  - 복합표본설계 요소인 층, 집락, 가중치 정보는 원시자료 DB 전체에 포함되어 있음
  - 전체 자료가 아닌 **일부 자료 선택(또는 삭제)하여 생성한 DB로 분석하는 경우**, 삭제된 자료에 포함된 복합표본설계 정보가 누락되어 **추정치의 표준오차 편향 발생**

관심 집단 분석 방법

전체 자료를 분석 DB로 사용

↓

관심 집단을 구분하는 변수 생성

(예 : 관심 집단에 속하는 경우 1, 그 외 0)

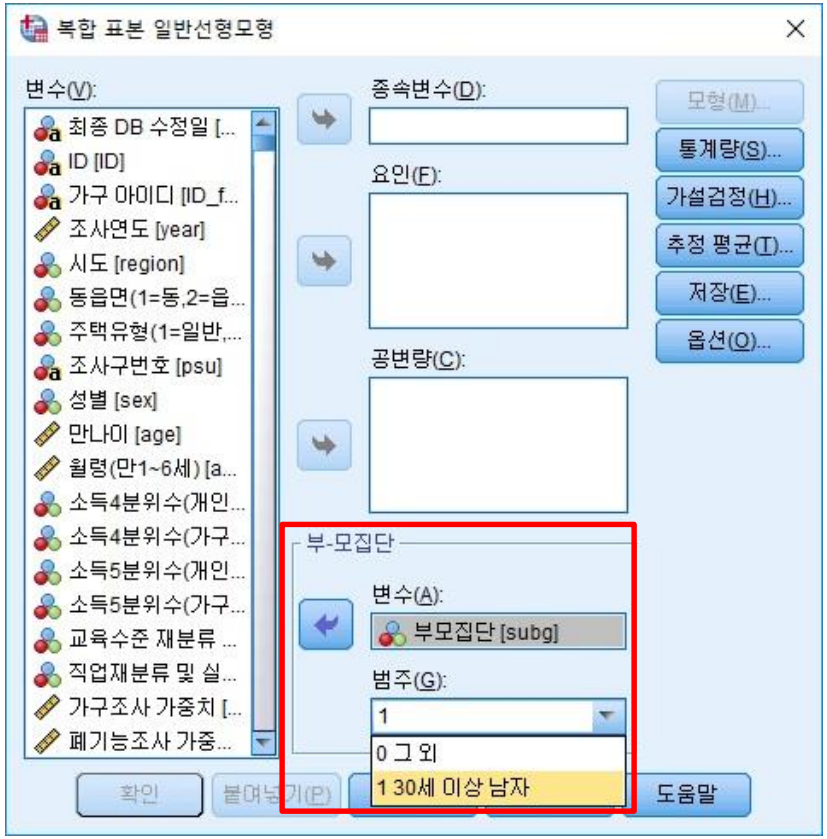
↓

생성된 관심 집단 변수를 "부-모집단"으로 지정

복합표본설계 자료 분석 시 유의사항(1)

- 관심 집단에 대한 분석 시에도 분석 DB는 항상 전체 자료 사용

세부 집단
분석 예시



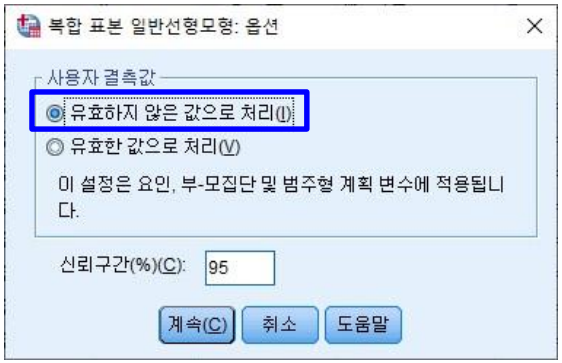
복합표본 자료 분석 시 반드시 적용!

복합표본설계 자료 분석 시 유의사항(2)

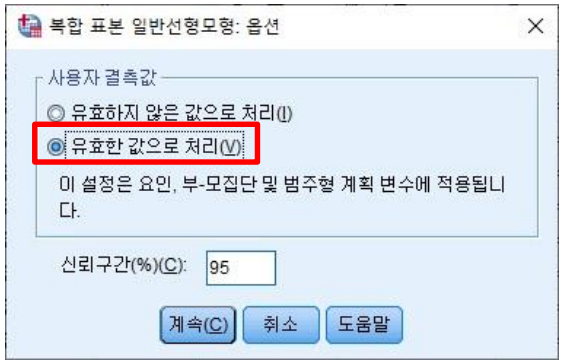
- **결측자료는 "유효한 값으로 처리" 하여 반드시 분석에 포함**
  - 분석 대화상자 옵션에서 사용자 결측값은 "유효하지 않은 값으로 처리" 되도록 기본 설정되어 있기 때문에 변경하지 않고 그대로 분석하는 경우 결측 케이스 삭제 후 분산 추정
  - 결측 케이스 삭제 시 삭제된 결측 자료에만 포함된 표본설계 정보가 있는 경우 분산 추정량에 편향 발생 가능

⇒ 따라서, 사용자 결측값을 "유효한 값으로 처리" 로 변경하여 분석변수에 결측이 있는 자료 모두 포함하여 분석

〈기본 설정〉



〈변경〉



! 복합표본 자료 분석 시 반드시 적용!

복합표본설계 자료 분석 시 유의사항_FAQ

Q 30세 이상 성인에 대한 결과를 산출하고자 하는 경우, 30세 미만에 해당하는 자료는 삭제해도 없나요 ?

A 아니요, 올바른 방법이 아닙니다.
30세 이상의 자료만 선택하여 분석하는 경우, 삭제된 30세 미만에 해당하는 자료에 포함된 복합표본설계 정보가 누락되어 추정치 표준오차의 편향이 발생합니다.

<기존DB 일부> HN18_all

			ID_fam	year	region	town_t	apt_t	psu	sex	age	age_month	incm
1	2020.03.05...	A90102050...	A9010205 ...	2018	1	1	2	A901 ...	1	80	.	1
2	2020.03.05...	A90102130...	A9010213 ...	2018	1	1	2	A901 ...	1	78	.	4
3	2020.03.05...	A90102130...	A9010213 ...	2018	1	1	2	A901 ...	2	75	.	4
4	2020.03.05...	A90102260...	A9010226 ...	2018	1	1	2	A901 ...	2	43	.	4
5	2020.03.05...	A90102330...	A9010233 ...	2018	1	1	2	A901 ...	2	80	.	2
6	2020.03.05...	A90102330...	A9010233 ...	2018	1	1	2	A901 ...	2	71	.	1
7	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	1	44	.	3
8	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	2	42	.	3
9	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	2	13	.	3
10	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	1	11	.	3

<생성DB 일부> 30세 이상만 선택

				year	region	town_t	apt_t	psu	sex	age	age_month	incm
1	2020.03.05...	A90102050...	A9010205 ...	2018	1	1	2	A901 ...	1	80	.	1
2	2020.03.05...	A90102130...	A9010213 ...	2018	1	1	2	A901 ...	1	78	.	4
3	2020.03.05...	A90102130...	A9010213 ...	2018	1	1	2	A901 ...	2	75	.	4
4	2020.03.05...	A90102260...	A9010226 ...	2018	1	1	2	A901 ...	2	43	.	4
5	2020.03.05...	A90102330...	A9010233 ...	2018	1	1	2	A901 ...	2	80	.	2
6	2020.03.05...	A90102330...	A9010233 ...	2018	1	1	2	A901 ...	2	71	.	1
7	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	1	44	.	3
8	2020.03.05...	A90102350...	A9010235 ...	2018	1	1	2	A901 ...	2	42	.	3
9	2020.03.05...	A90102390...	A9010239 ...	2018	1	1	2	A901 ...	1	36	.	3
10	2020.03.05...	A90102390...	A9010239 ...	2018	1	1	2	A901 ...	2	30	.	3

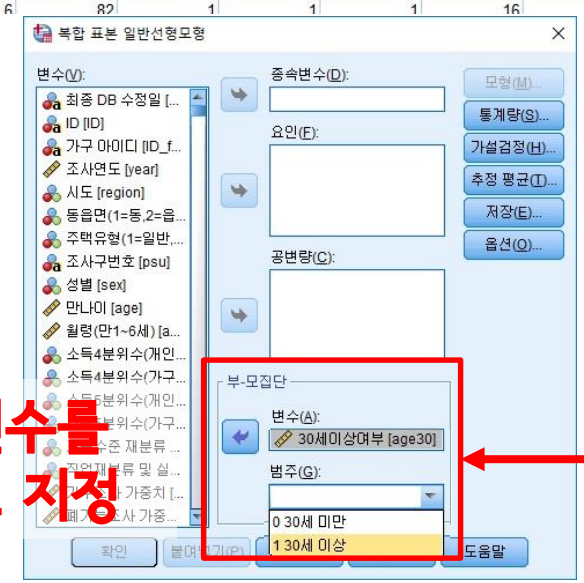
복합표본설계 자료 분석 시 유의사항_FAQ

Q 그렇다면 30세 이상에 30세 이상 성인에 대한 결과를 산출하고자 하는 경우 어떻게 분석해야 하나요?

A 이러한 경우에는 age30과 같이 30세 이상 여부를 구분하는 부-모집단 변수를 새로 생성한 후에 분석 시 해당 변수를 부-모집단 변수로 지정하여 분석하시면 됩니다.

	N_B1	N_B2	N_NIAC	N_VITC	LF_secur_y	LF_BUYER	LF_SAFE	BMI	subg	age30	age_g	sm_g
1	1	1	5	6	1	2	1	21	1	1	7	2
2	1	1	17	109	1	1	1	27	1	1	4	1
3	1	1	5	97	1	1	1	21	0	1	4	1
4	.	.	.	.	.	.	.	23	0	1	4	1
5	1	0	6	11	1	1	1	27	0	1	4	1
6	1	1	7	37	1	1	1	22	0	1	1	1
7	1	1	12	54	1	1	1	26	1	1	1	1
8	0	0	2	6	1	1	1	19	0	1	4	1
9	1	2	5	10	1	1	1	19	0	0	0	.
10	1	1	6	82	1	1	1	16	0	0	0	

관심집단 구분 변수 생성



관심집단 구분 변수를 부-모집단 변수로 지정

복합표본설계 자료 분석 시 유의사항_FAQ

Q 제공 원시자료인 ALL DB에 포함된 변수가 너무 많은데 분석에 필요한 변수만 선택하여 새로운 DB를 생성해도 되나요?

A 네, 가능합니다.
단, 표본설계 정보에 해당하는 집락(psu), 층화(kstrata), 가중치(wt_itvex 등, 조사부문별/영역별 선택 필요) 변수는 꼭 DB에 포함시켜주셔야 합니다.

<생성DB 일부> 일부 변수 선택

집락 가중치 층

	ID	ID_fam	region	town_t	apt_t	psu	sex	age	wt_itvex	kstrata	HE_sbp	HE_dt
1	A901020501	A9010205	1	1	2	A901	1	80	11065	801	150	
2	A901021301	A9010213	1	1	2	A901	1	78	3688	801	132	
3	A901021302	A9010213	1	1	2	A901	2	75	5244	801	156	
4	A901022601	A9010226	1	1	2	A901	2	43	6814	801	102	
5	A901023303	A9010233	1	1	2	A901	2	80	5244	801	149	
6	A901023304	A9010233	1	1	2	A901	2	71	3316	801	121	
7	A901023501	A9010235	1	1	2	A901	1	44	8459	801	109	
8	A901023502	A9010235	1	1	2	A901	2	42	6814	801	104	
9	A901023503	A9010235	1	1	2	A901	2	13	8233	801	110	
10	A901023504	A9010235	1	1	2	A901	1	11	5536	801	106	

표본설계 정보는 반드시 DB에 포함!

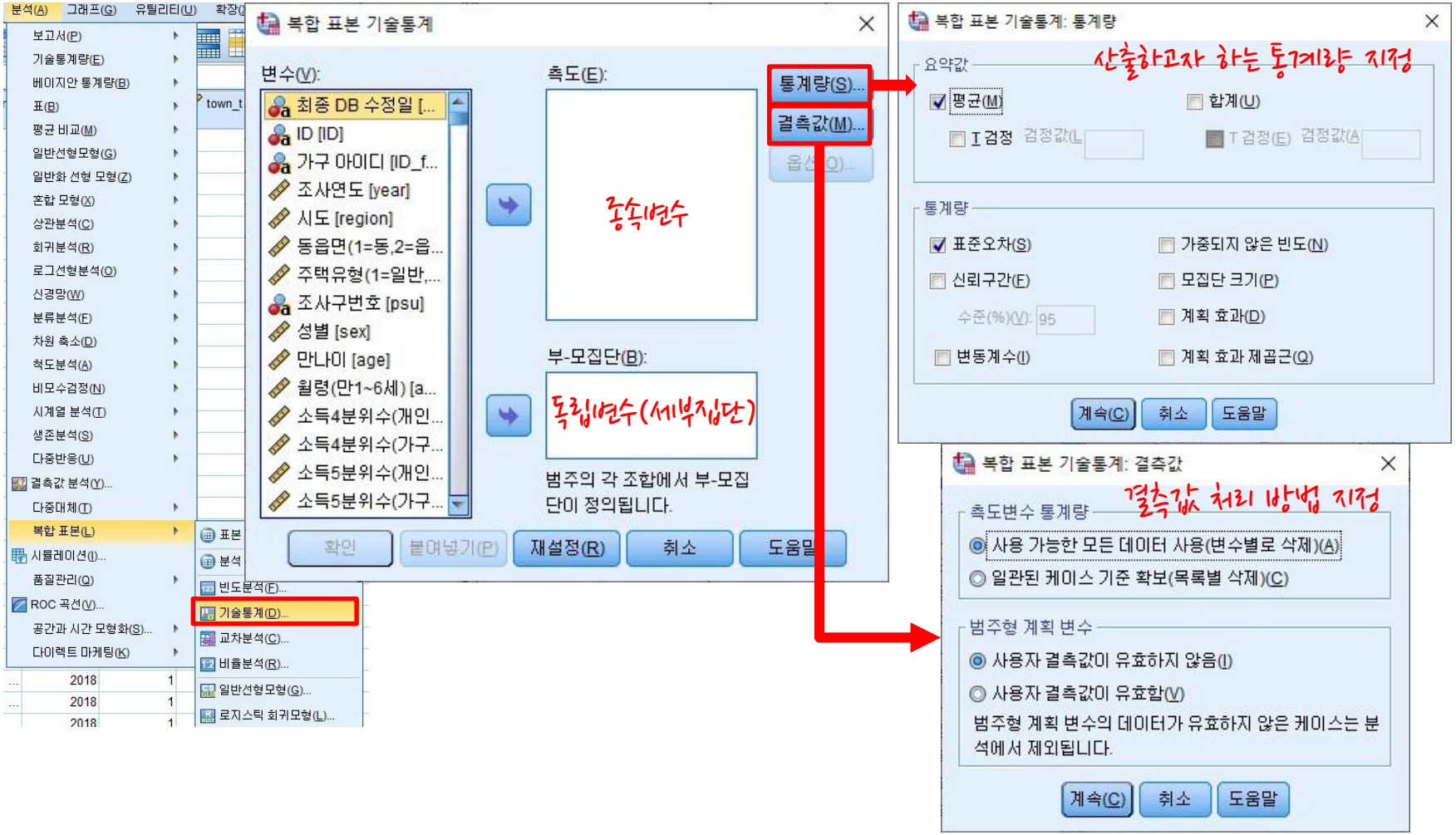
02.
복합표본설계 기술 통계

- 평균 추정
- 비율 추정

복합표본 기술 통계_대화상자

▪복합표본 기술통계

- 복합표본 자료에서 모집단의 **평균, 비율, 표준오차 등 추정치 산출**



복합표본 기술 통계_평균 추정 예제

예제 성별 연령그룹별 수축기혈압 평균

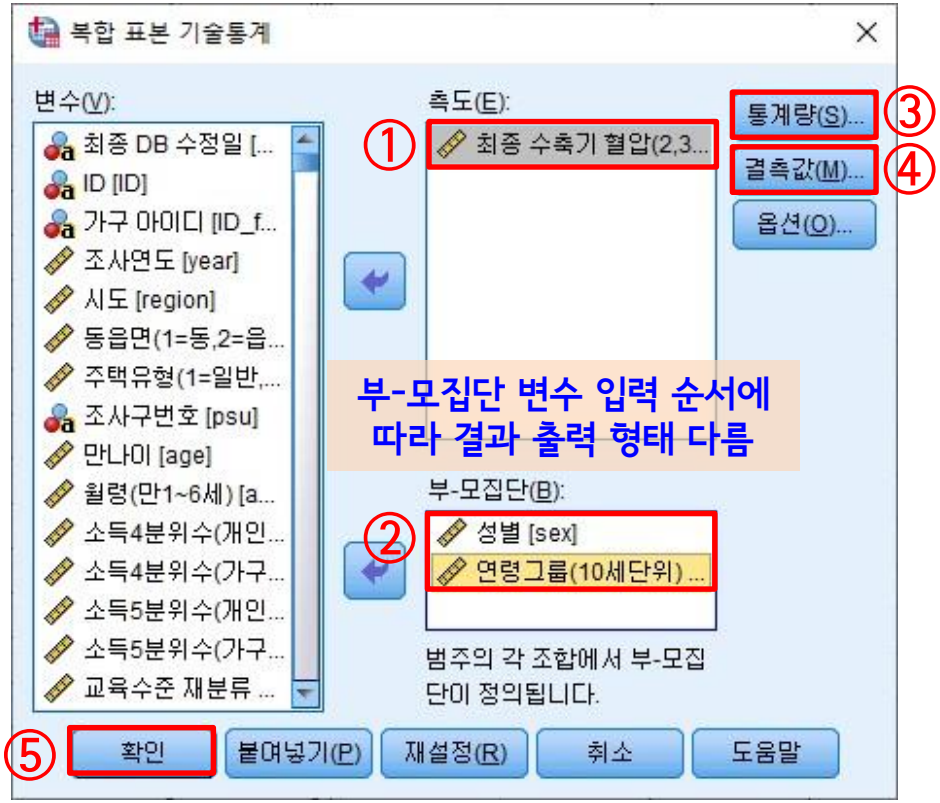
변수 설명

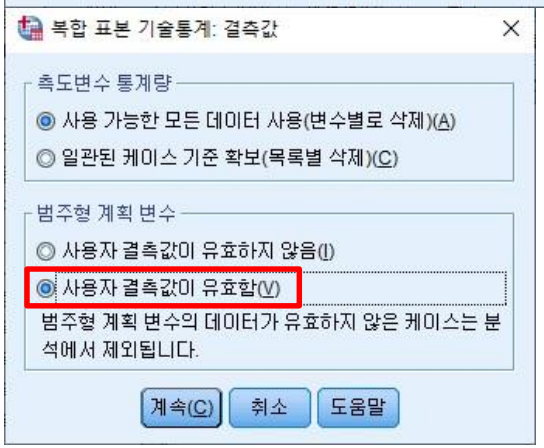
구분	변수	변수내용
종속변수(측도)	HE_sbp	수축기혈압(연속형)
독립변수(부-모집단)	sex	1. 남자
		2. 여자
	age_g	3. 30대
		4. 40대
		5. 50대
		6. 60대
		7. 70세 이상
		0. 그 외

복합표본 기술 통계_평균 추정 예제

예제 성별 연령그룹별 수축기혈압 평균

분석 ▶ 복합 표본 ▶ 기술통계





복합표본 기술 통계_평균 추정 결과

예제 성별 연령그룹별 수축기혈압 평균

일변량 통계량

		B 추정값	표준오차	가중되지 않은 빈도
평균	최종 수축기 혈압(2,3차 평균)	116.81	.344	6847

▶ 전체 및 부-모집단 통계량 출력

부-모집단 기술통계

일변량 통계량

성별	연령그룹(10세단위)			B 추정값	표준오차	가중되지 않은 빈도
1	그 외	평균	최종 수축기 혈압(2,3차 평균)	114.32	.561	702
	30대	평균	최종 수축기 혈압(2,3차 평균)	116.26	.757	414
	40대	평균	최종 수축기 혈압(2,3차 평균)	117.81	.813	476
	50대	평균	최종 수축기 혈압(2,3차 평균)	121.74	.796	505
	60대	평균	최종 수축기 혈압(2,3차 평균)	125.17	.903	500
	70세 이상	평균	최종 수축기 혈압(2,3차 평균)	127.71	1.117	468
2	그 외	평균	최종 수축기 혈압(2,3차 평균)	105.61	.406	699
	30대	평균	최종 수축기 혈압(2,3차 평균)	106.05	.624	494
	40대	평균	최종 수축기 혈압(2,3차 평균)	110.77	.747	652
	50대	평균	최종 수축기 혈압(2,3차 평균)	118.39	.737	689
	60대	평균	최종 수축기 혈압(2,3차 평균)	124.92	.937	602
	70세 이상	평균	최종 수축기 혈압(2,3차 평균)	131.34	.865	646

복합표본 기술 통계_비율 추정 예제

- 평균 추정과 동일한 방법 : 복합표본 기술 통계 대화상자 이용
- 비율 추정 개념

ID(10명)	A	B	C	D	E	F	G	H	I	J	합계	평균
SBP(mmHg)	118	114	138	117	104	109	152	103	120	95	1170	1170/10=117
DBP(mmHg)	69	77	92	67	71	76	86	74	75	65	752	752/10=75.2
약복용 여부	1	0	0	0	0	0	0	0	0	0	1	-
고혈압 유병여부 (140≤SBP or 90≤DBP or 약복용 O)	1	0	1	0	0	0	1	0	0	0	3	3/10=0.3 = 고혈압 유병 비율

예제 30세 이상의 성별 고혈압 유병률

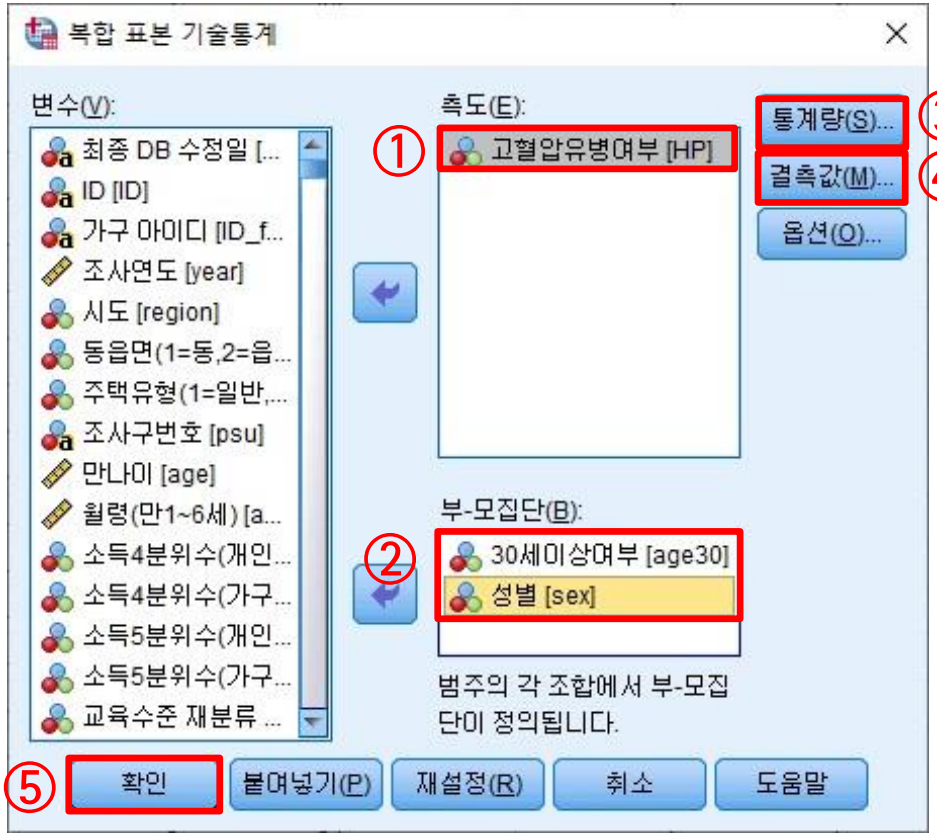
변수 설명

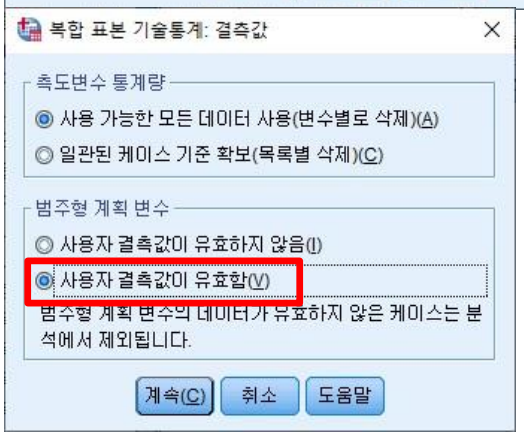
구분	변수	변수내용
종속변수(측도변수)	HP	1. 고혈압 유병 O
		0. 고혈압 유병 X
독립변수(부-모집단)	sex	1. 남자
		2. 여자
	age30	1. 30세 이상
		0. 30세 미만

복합표본 기술 통계_비율 추정 예제

예제 30세 이상의 성별 고혈압 유병률

분석 ▶ 복합 표본 ▶ 기술통계





복합표본 기술 통계_비율 추정 결과

예제 30세 이상의 성별 고혈압 유병률

일변량 통계량

		B 추정값	표준오차	가중되지 않은 빈도
평균	고혈압유병여부	.2585	.0084	6843

부-모집단 기술통계

일변량 통계량

30세이상여부	성별			B 추정값	표준오차	가중되지 않은 빈도
30세 미만	1	평균	고혈압유병여부	.0635	.0122	702
	2	평균	고혈압유병여부	.0114	.0042	698
30세 이상	1	평균	고혈압유병여부	.3638	.0120	2361
	2	평균	고혈압유병여부	.3041	.0124	3082

▶ 고혈압 유병률 = β 추정값 × 100 / 표준오차 = 표준오차 추정값 × 100

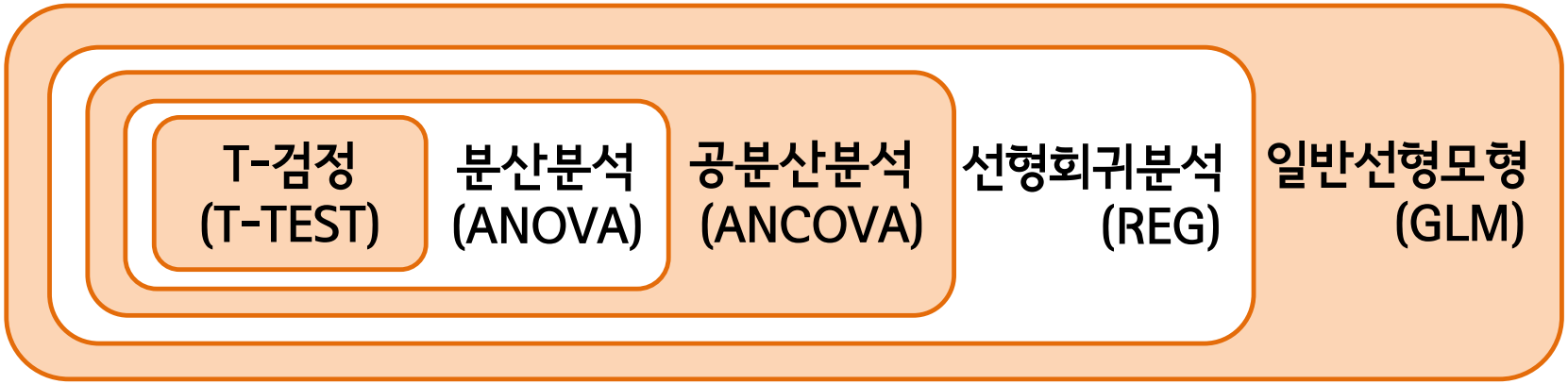
⇒ 30세 이상 남자 기준, 고혈압 유병률 36.4%(=0.3638 × 100) 표준오차 1.2(=0.012 × 100)

03.
복합표본설계 선형회귀분석

- 단순선형회귀
- T-검정
- 분산분석
- 공분산분석
- 보정평균
- 기준집단 변경

복합표본설계 선형회귀분석 개요

- 선형회귀분석 : 연속형 종속변수와 독립변수(연속형, 범주형)의 관련성 분석
- 회귀모형의 유형 및 분석



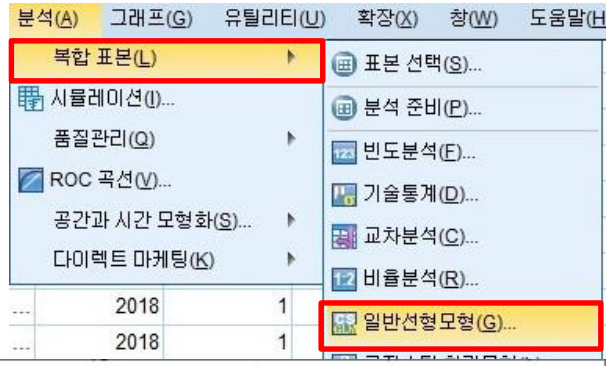
※일반선형모형 : 오차와 평균에 대한 선형회귀모형의 가정을 확장한 개념

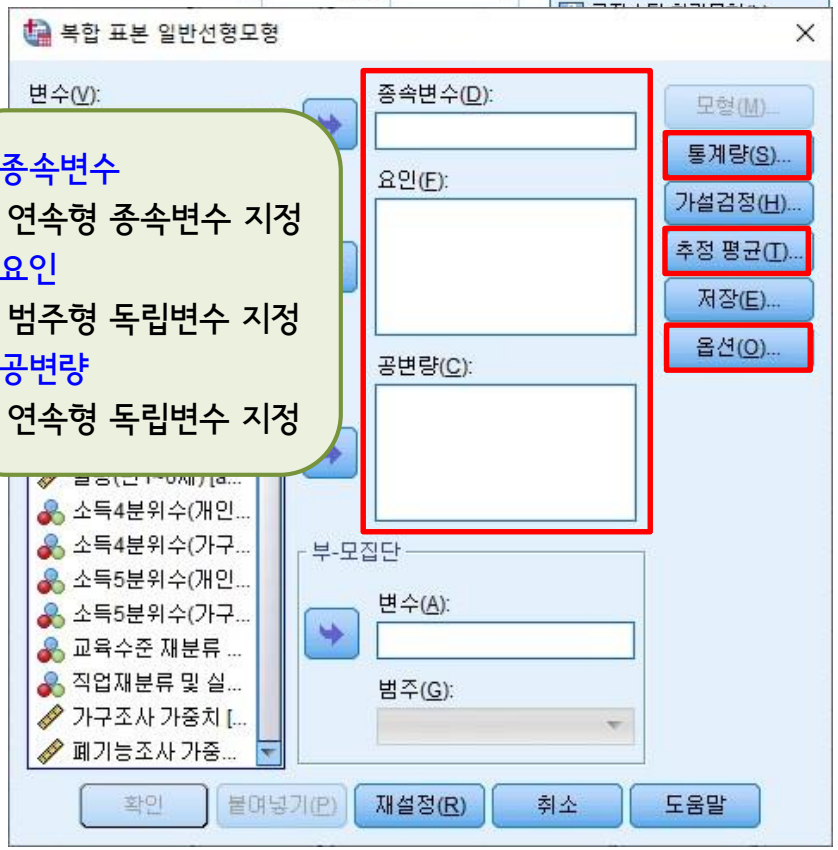
종속변수	독립변수	분석방법	검정방법	SPSS 복합표본설계 프로시저
연속형	범주형(2그룹)	T검정	T-TEST	복합표본 일반선형모형 (독립변수의 유형에 상관없이 모두 동일)
	범주형(3그룹 이상)	분산분석	ANOVA	
	연속형, 범주형	선형회귀분석	REG	
	연속형, 범주형	일반선형모형	GLM (T-TEST, ANOVA, ANCOVA, REG)	

복합표본 일반선형모형 대화상자

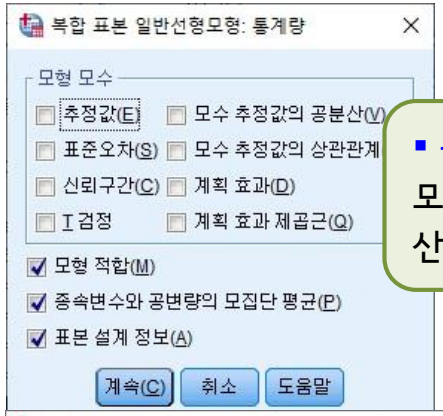
절차

분석
↓
복합 표본
↓
일반선형모형

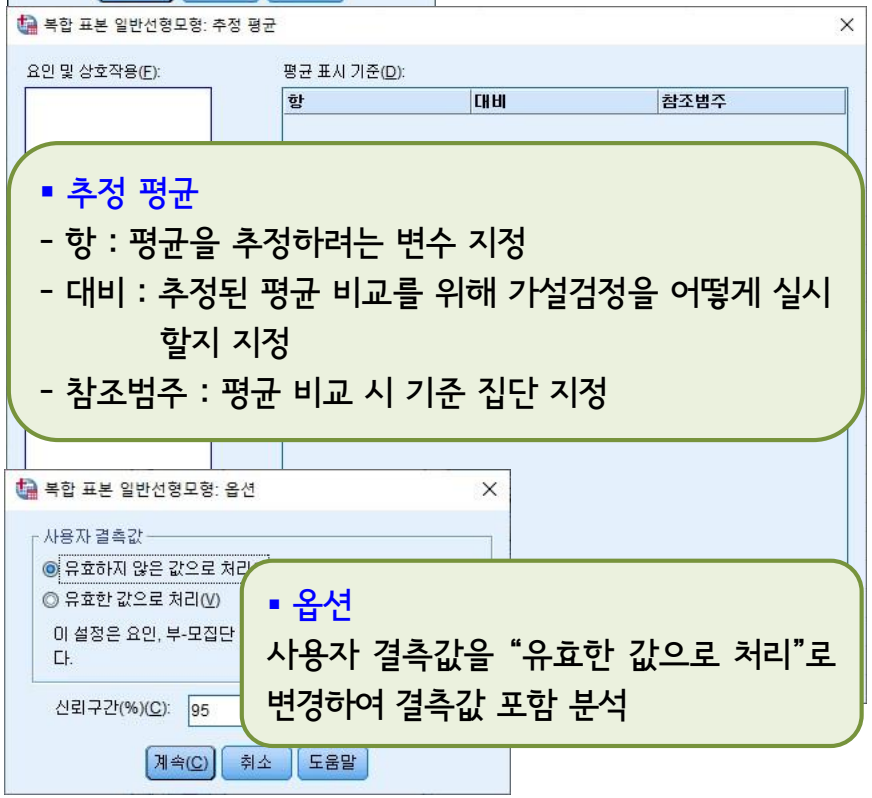




- **종속변수**
  - 연속형 종속변수 지정
- **요인**
  - 범주형 독립변수 지정
- **공변량**
  - 연속형 독립변수 지정



- **통계량**
  모형의 적합성 및 추정값 등 산출하고자 하는 통계량 지정



- **추정 평균**
  - 항 : 평균을 추정하려는 변수 지정
  - 대비 : 추정된 평균 비교를 위해 가설검정을 어떻게 실시할지 지정
  - 참조범주 : 평균 비교 시 기준 집단 지정

- **옵션**
  사용자 결측값을 "유효한 값으로 처리"로 변경하여 결측값 포함 분석

복합표본 단순선형회귀 예제

- 연속형 종속변수와 연속형 독립변수 간의 연관성 검정

예제 30세 이상에서 체질량지수와 수축기혈압의 연관성

변수 설명

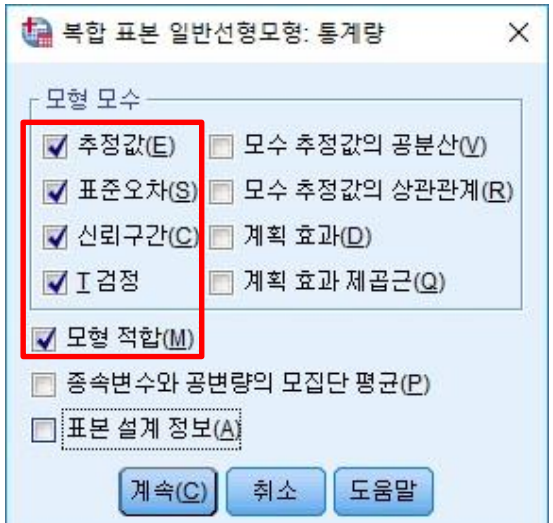
구분	변수	변수내용
종속변수	HE_sbp	수축기혈압(연속형)
독립변수	BMI	체질량지수
부-모집단	age30	1. 30세 이상
		0. 30세 미만

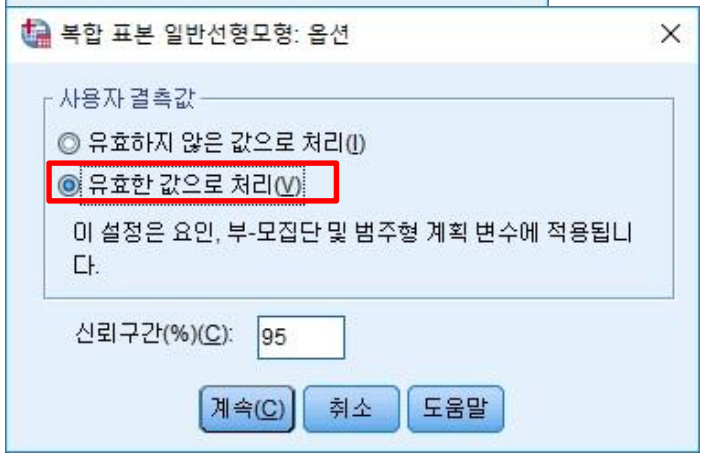
모형 $HE_sbp = \beta_0 + \beta_1 * BMI + \varepsilon$

복합표본 단순선형회귀 예제

예제 30세 이상에서 체질량지수와 수축기혈압의 연관성

분석 ▶ 복합 표본 ▶ 일반선형모형





복합표본 단순선형회귀 결과

분석 결과

모형 요약[a]

R 제곱	.052

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + BMI

▶ 결정계수(R^2)=0.052로 체질량지수가 수축기 혈압 변동 중 약 5.2% 설명 가능

모형 효과 검정[a]

	소스	자유도1	자유도2	Wald F	유의확률
모형	(수정된 모형)	1.000	165.000	213.008	.000
절편	(수정된 모형)	1.000	165.000	2451.662	.000
	BMI	1.000	165.000	213.008	.000

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + BMI

▶모형 효과 검정에서 p<0.05이므로 모형의 설명 유의

모수 추정값[a]

모수1	B 추정값	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한	가설검정 t	가설검정 자유도	가설검정 유의확률
(수정된 모형)	92.464	1.867	88.777	96.152	49.514	165.000	.000
BMI	1.103	.076	.953	1.252	14.595	165.000	.000

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + BMI

• 추정 모형 : E(HE_sbp)=92.46+1.10*BMI

▶ 결과해석

- 모수 추정값에서 p<0.05이므로 30세 이상에서 BMI가 1kg/m^2 증가하는 경우 수축기혈압은 약 1.1mmHg 증가

복합표본 T-검정

▪ T-검정 : 연속형 종속변수와 범주형(2그룹) 독립변수 간의 연관성 검정

예제 30세 이상에서 성별에 따른 수축기혈압 평균 차이 유무

변수 설명

구분	변수명	변수 내용
종속변수	HE_sbp	수축기혈압
독립변수	sex	1. 남자
		2. 여자(기준집단)
부-모집단	age30	1. 30세 이상
		0. 30세 미만

모형

$$HE_sbp = \beta_0 + \beta_1 * sex + \varepsilon$$ ∨ 범주형 변수

* β_0 : 기준 집단의 평균 / β_1 : 기준 집단과의 평균 차이

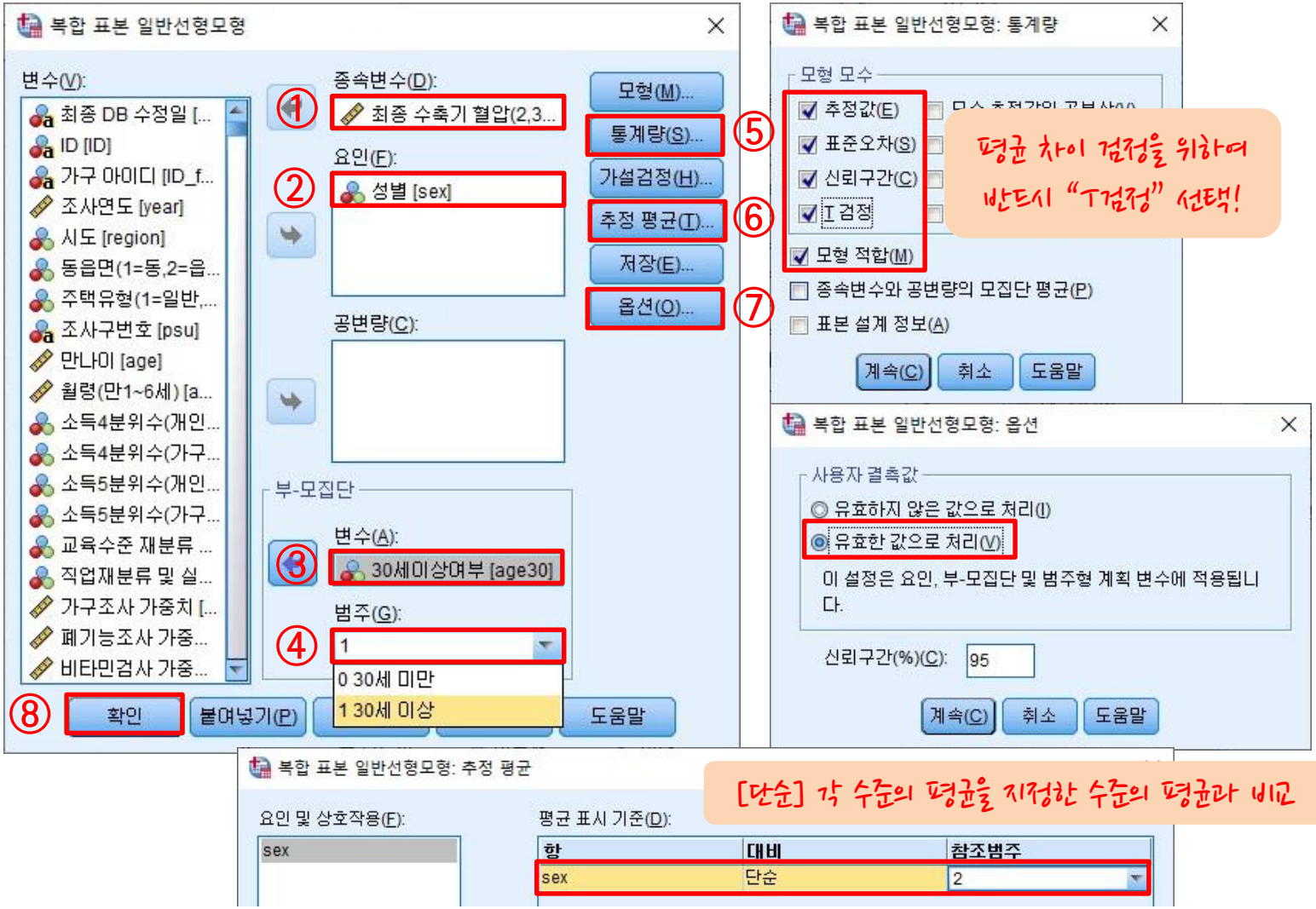
$$= \mu_2 + (\mu_1 - \mu_2) * sex_1 + \varepsilon$$ ∨ 가변수

* μ_1 : 남자의 평균 / μ_2 : 여자의 평균

- 30세 이상 남자(sex_1=1)의 평균 : $E(HE_sbp) = \mu_2 + (\mu_1 - \mu_2) * 1 = \mu_1$
- 30세 이상 여자(sex_1=0)의 평균 : $E(HE_sbp) = \mu_2 + (\mu_1 - \mu_2) * 0 = \mu_2$

복합표본 T-검정 예제

예제 30세 이상에서 성별에 따른 수축기혈압 평균 차이 유무



복합표본 T-검정 결과

분석 결과

모형 요약[a]

R 제곱	.010

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + sex

▶ 결정계수(R^2)=0.010로 성별이 수축기혈압 변동 중 약 1.0% 설명 가능

모형 효과 검정[a]

소스	자유도1	자유도2	Wald F	유의확률
(수정된 모형)	1.000	165.000	41.079	.000
(수정된 모형)	1.000	165.000	93258.114	.000
sex	1.000	165.000	41.079	.000

모형

절편

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + sex

▶모형 효과 검정에서 p<0.05이므로 모형의 설명 유의

모수 추정값[a]

모수1	B 추정값	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한	가설검정 t	가설검정 자유도	가설검정 유의확률
(수정된 모형)	117.379	.474	116.443	118.315	247.624	165.000	.000
[sex=1]	3.409	.532	2.359	4.460	6.409	165.000	.000
[sex=2]	.000[b]	.	.	.	.	.	.

(기준집단)

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + sex

b. 중복된 모수이므로 0으로 설정됩니다.

• 추정 모형 : $E(HE_sbp)=117.38+3.41*sex_1$

▶ 결과해석

- 모수 추정값에서 p<0.05이므로 30세 이상에서 성별에 따른 수축기혈압 평균 차이는 유의
- 여자 대비 남자가 약 3.41mmHg 높음

▶ 모수 추정값을 이용한 수축기혈압 평균 산출

- 30세 이상 여자 117.38mmHg
- 30세 이상 남자 120.79mmHg (=117.38+3.41)

복합표본 T-검정 결과

분석 결과

※[추정 평균] 옵션으로 평균 산출

추정 주변 평균: 성별

추정값

성별	평균	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한
1	120.79	.470	119.86	121.72
2	117.38	.474	116.44	118.31

부-모집단: 30세이상여부 = 30세 이상

▶ 성별 평균 추정값 표 출력

개별 검정 결과

성별 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화 오차	자유도1	자유도2	Wald F	유의확률
수준 1 및 수준 2	3.409	.000	3.409	.532	1.000	165.000	41.079	.000

부-모집단: 30세이상여부 = 30세 이상

a. 참조범주 = 2 (기준집단)

▶ 기준집단과의 평균 차이(대비 추정값) 및 유의확률 출력

복합표본 분산분석

▪ 분산분석 : 연속형 종속변수와 범주형(3그룹 이상) 독립변수 간의 연관성 분석

예제 30세 이상에서 연령그룹에 따른 수축기혈압 평균 차이 유무

변수 설명

구분	변수	변수내용
종속변수	HE_sbp	수축기혈압(연속형)
독립변수	age_g	3. 30대
		4. 40대
		5. 50대
		6. 60대
		7. 70세 이상(기준집단)
부-모집단	age30	1. 30세 이상
		0. 30세 미만

모형

$$HE_sbp = \beta_0 + \beta_1 * age_g + \varepsilon$$

∨ 범주형 변수

* β_0 : 기준 집단의 평균 / β_1 : 기준 집단과의 평균 차이

$$= \mu_7 + (\mu_3 - \mu_7) * age_g_3 + (\mu_4 - \mu_7) * age_g_4 + (\mu_5 - \mu_7) * age_g_5 + (\mu_6 - \mu_7) * age_g_6 + \varepsilon$$

∨ 가변수

* $\mu_3, \mu_4, \mu_5, \mu_6, \mu_7$: 연령그룹별 평균

- 70세 이상 ($age_g_3 \cdot age_g_4 \cdot age_g_5 \cdot age_g_6 = 0$) :

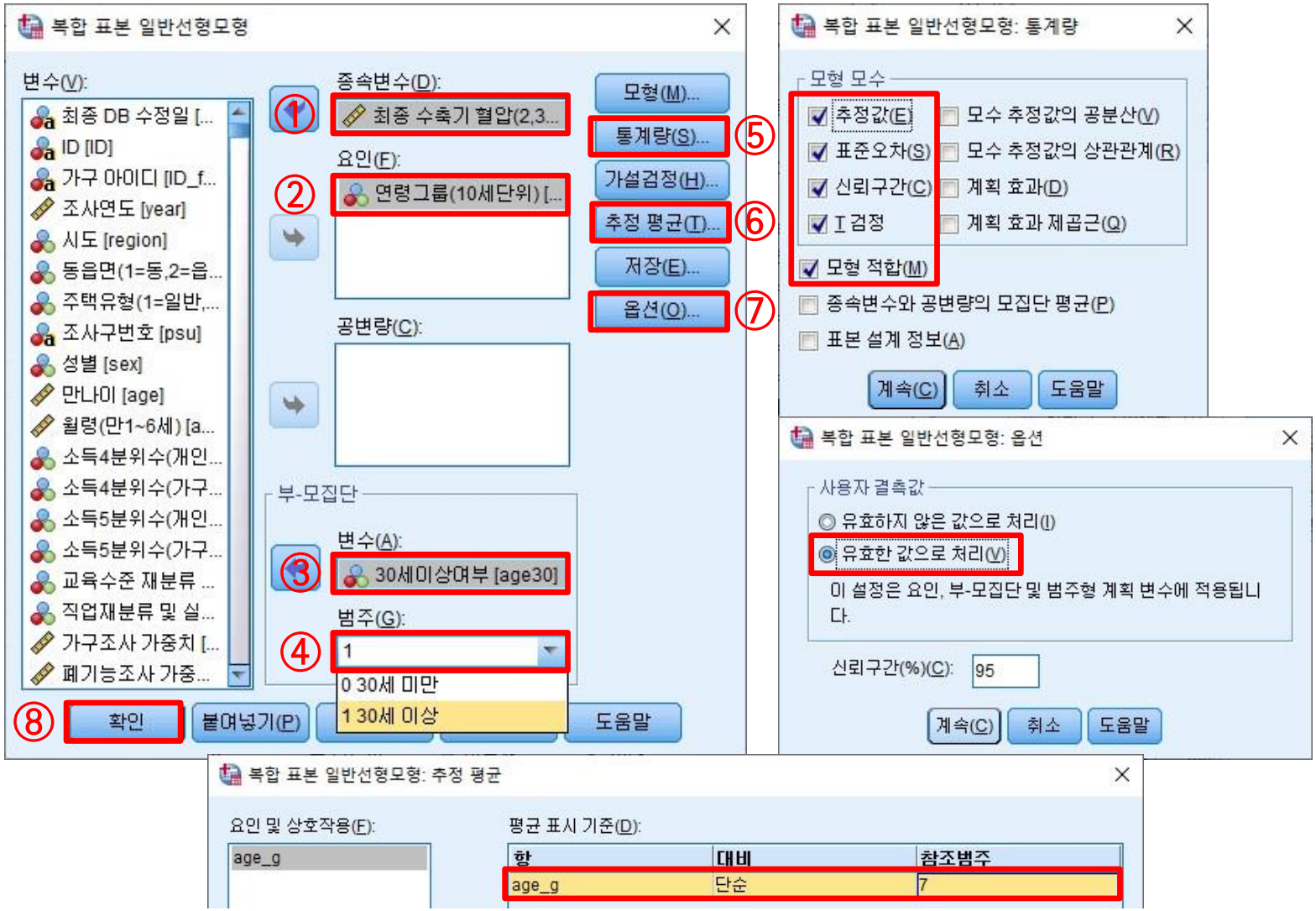
$$E(HE_sbp) = \mu_7 + (\mu_3 - \mu_7) * 0 + (\mu_4 - \mu_7) * 0 + (\mu_5 - \mu_7) * 0 + (\mu_6 - \mu_7) * 0 = \mu_7$$

- 30대($age_g_3 = 1$, $age_g_4 \cdot age_g_5 \cdot age_g_6 = 0$) :

$$E(HE_sbp) = \mu_7 + (\mu_3 - \mu_7) * 1 + (\mu_4 - \mu_7) * 0 + (\mu_5 - \mu_7) * 0 + (\mu_6 - \mu_7) * 0 = \mu_3$$

복합표본 분산분석 예제

예제 30세 이상에서 연령그룹에 따른 수축기혈압 평균 차이 유무



복합표본 분산분석 결과

분석 결과

모형 요약[a]

R 제곱	.146

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶ 결정계수(R^2)=0.146로 연령그룹이 수축기혈압 변동 중 약 14.6% 설명 가능

모형 효과 검정[a]

	소스	자유도1	자유도2	Wald F	유의확률
모형	(수정된 모형)	4.000	162.000	153.525	.000
절편	(수정된 모형)	1.000	165.000	117607.227	.000
	age_g	4.000	162.000	153.525	.000

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶모형 효과 검정에서 p<0.05이므로 모형의 설명 유의

모수 추정값[a]

			95% 신뢰구간		가설검정		
모수1	B 추정값	표준화 오차	하한	상한	t	자유도	유의확률
(수정된 모형)	129.876	.708	128.478	131.274	183.460	165.000	.000
[age_g=3]	-18.548	.860	-20.245	-16.850	-21.571	165.000	.000
[age_g=4]	-15.525	.836	-17.175	-13.875	-18.579	165.000	.000
[age_g=5]	-9.807	.817	-11.419	-8.194	-12.005	165.000	.000
[age_g=6]	-4.836	.882	-6.577	-3.095	-5.484	165.000	.000
(기준집단) [age_g=7]	.000[b]	.	.	.	.	.	.

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶ 추정 회귀 모형

$$E(HE_sbp) = 129.88 - 18.55 * age_g_3 - 15.53 * age_g_4 - 9.81 * age_g_5 - 4.84 * age_g_6$$

▶ 결과해석

- 모수 추정값에서 p<0.05이므로 70세 이상 기준 연령그룹에 따른 수축기혈압 평균 차이는 유의
- 70세 이상 기준 연령그룹 감소할수록 수축기혈압 평균 감소

▶ 모수 추정값을 이용한 수축기혈압 평균 산출

- 70세 이상 129.88mmHg
- 30대 111.33mmHg (=129.88-18.55)

복합표본 분산분석 결과

분석 결과

※[추정 평균] 옵션으로 평균 산출

추정 주변 평균: 연령그룹(10세단위)

추정값

연령그룹(10세단위)	평균	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한
30대	111.33	.566	110.21	112.45
40대	114.35	.555	113.26	115.45
50대	120.07	.583	118.92	121.22
60대	125.04	.679	123.70	126.38
70세 이상	129.88	.708	128.48	131.27

부-모집단: 30세이상여부 = 30세 이상

▶ 연령그룹별 평균 추정값 표 출력

개별 검정 결과

연령그룹(10세단위) 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화 오차	자유도1	자유도2	Wald F	유의확률
수준 30대 및 수준 70세 이상	-18.548	.000	-18.548	.860	1.000	165.000	465.317	.000
수준 40대 및 수준 70세 이상	-15.525	.000	-15.525	.836	1.000	165.000	345.196	.000
수준 50대 및 수준 70세 이상	-9.807	.000	-9.807	.817	1.000	165.000	144.113	.000
수준 60대 및 수준 70세 이상	-4.836	.000	-4.836	.882	1.000	165.000	30.070	.000

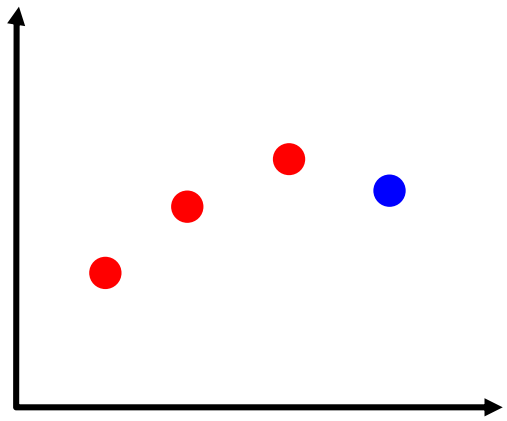
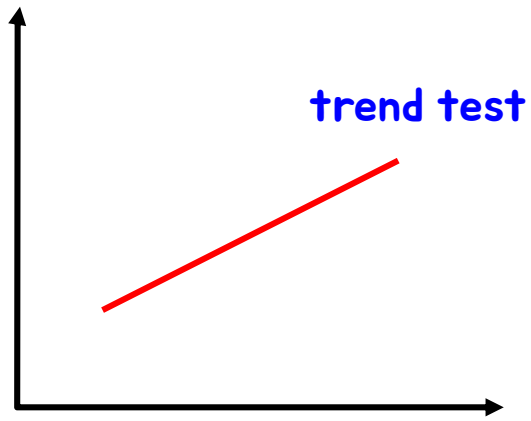
부-모집단: 30세이상여부 = 30세 이상

a. 참조범주 = 70세 이상

▶ 기준집단과의 평균 차이(대비 추정값) 및 유의확률 출력

복합표본 추이검정

- 독립 변수에 서열 척도인 범주형 변수를 요인이 아닌 공변량에 지정하여 분석할 경우 연속형 변수로 가정하여 분석

범주형 변수 요인으로 지정	범주형 변수 공변량으로 지정
	
기준집단 대비 나머지 집단에 대한 영향 추정 및 평균 비교	범주형 변수의 값이 증가함에 따른 효과 추정

복합표본 추이검정 예제

▪ 종속변수 : 연속형, 범주형 / 독립변수 : 연속형

- 범주형 변수를 공변량에 지정하여 분석할 경우 연속형 변수로 가정하여 분석

예제 30세 이상에서 연령그룹에 따른 수축기혈압 추이

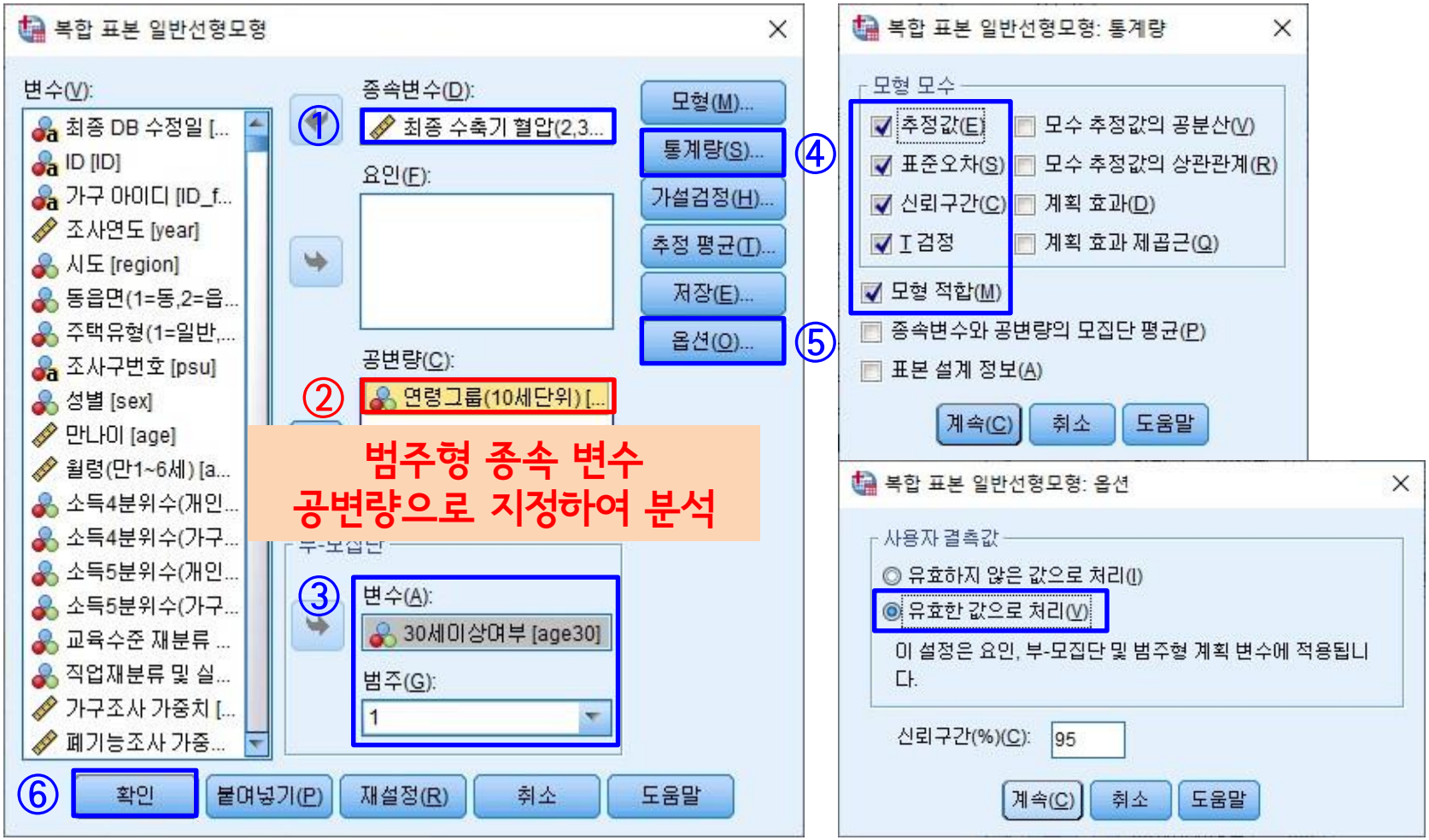
변수 설명

구분	변수	변수내용
종속변수	HE_sbp	수축기혈압(연속형)
독립변수	age_g	연령그룹(연속형)
	age30 (부-모집단)	1. 30세 이상
		0. 30세 미만

모형 $HE_sbp = \beta_0 + \beta_1 * age_g + \varepsilon$ ∨ 연속형 변수

복합표본 추이검정 예제

예제 30세 이상에서 연령그룹에 따른 수축기혈압 추이



복합표본 추이검정 결과

분석 결과

모형 요약[a]

R 제곱	.145

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶ 결정계수(R^2)=0.145로 연령그룹이 수축기 혈압 변동 중 약 14.5% 설명 가능

모형 효과 검정[a]

소스	자유도1	자유도2	Wald F	유의확률
(수정된 모형)	1.000	165.000	611.423	.000
(수정된 모형)	1.000	165.000	10551.136	.000
age_g	1.000	165.000	611.423	.000

모형

절편

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶모형 효과 검정에서 p<0.05이므로 모형의 설명 유의

모수 추정값[a]

모수1	B 추정값	표준화 오차	95% 신뢰구간		가설검정		
			하한	상한	t	자유도	유의확률
(수정된 모형)	96.222	.937	94.372	98.071	102.719	165.000	.000
age_g	4.772	.193	4.391	5.153	24.727	165.000	.000

부-모집단: 30세이상여부 = 30세 이상

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g

▶ 결과해석

- p<0.05이므로 30세 이상에서 연령그룹에 따른 수축기혈압 추이 유의
- 연령그룹이 증가함에 따라 수축기혈압 증가

복합표본 공분산분석

▪ 공분산분석

- 연속형 종속변수와 독립변수(범주형, 연속형) 간의 연관성 분석
- 종속변수에 영향을 줄 수 있는 변수(연속형)를 통제하여 집단 간 차이 분석하는 경우 사용

예제

연령그룹과 체질량지수 보정 시,
30세 이상 남자의 흡연 여부별 수축기혈압 평균 차이 유무

변수 설명

구분	변수	변수내용	구분	변수	변수내용
종속변수	HE_sbp	수축기혈압(연속형)	독립변수	sm_g	1. 비흡연
독립변수	age_g	3. 30대			2. 과거흡연
		4. 40대			3. 현재흡연
		5. 50대		BMI	체질량지수(연속형)
		6. 60대		subg (부-모집단)	1. 30세 이상 남자
		7. 70세이상			0. 그 외

* 공변량이란?
- 독립변수들이 하나의 종속변수에 대해 함께 공유하는 변량
- 독립변수들 외 종속변수에 영향을 주는 요인을 통제
- 연속형 변수

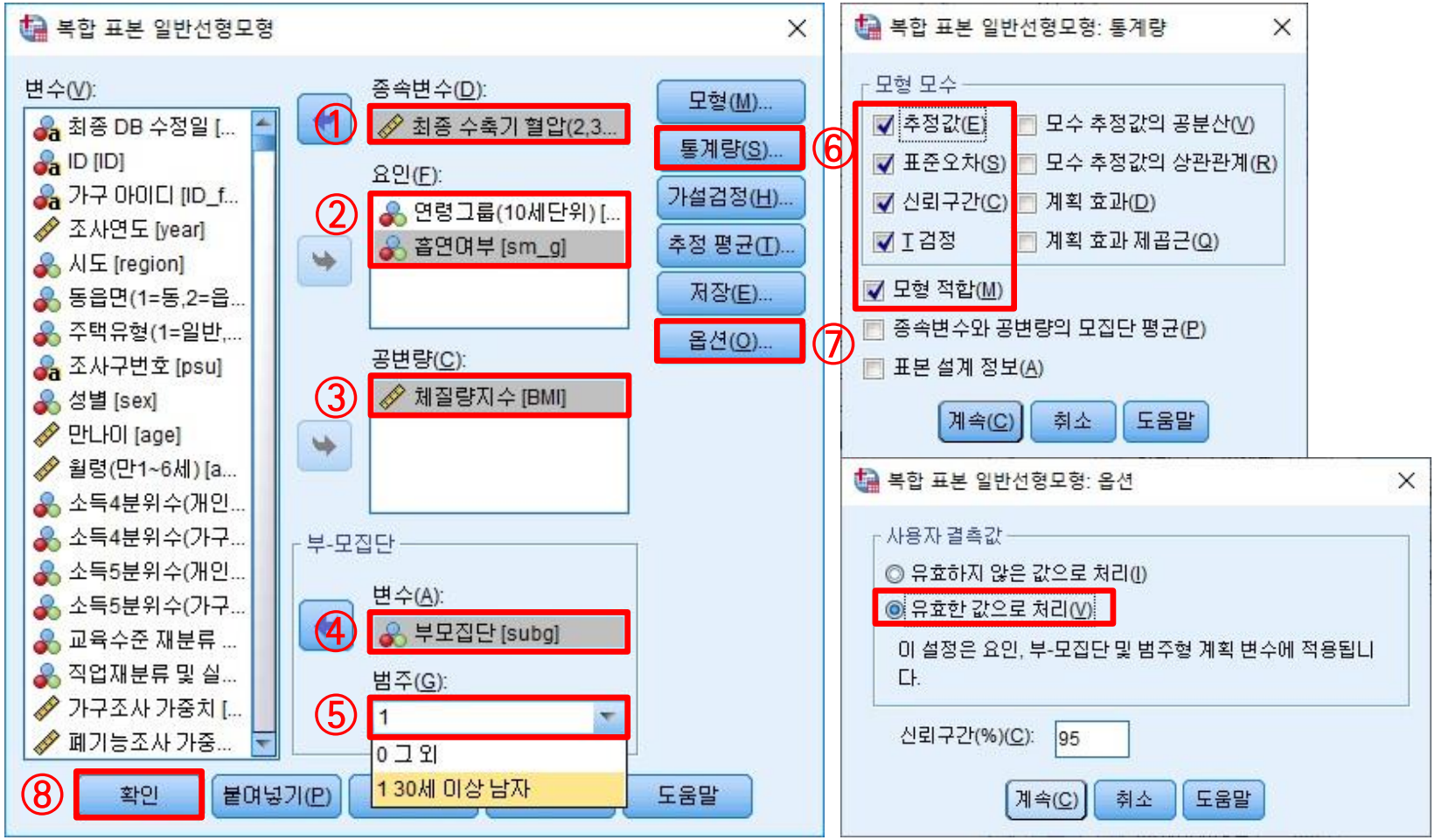
모형

$$HE_sbp = \beta_0 + \beta_1 * age_g + \beta_2 * sm_g + \beta_3 * BMI + \varepsilon$$

복합표본 공분산분석 예제

예제

연령그룹과 체질량지수 보정 시,
30세 이상 남자의 흡연 여부별 수축기혈압 평균 차이 유무



복합표본 공분산분석 결과

분석 결과

모형 요약[a]

R 제곱	.117

부-모집단: 부모집단 = 30세 이상 남자

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g + sm_g + BMI

▶ 결정계수(R^2)=0.117로 수축기혈압 변동 중 약 11.7% 설명 가능

모형 효과 검정[a]

소스	자유도1	자유도2	Wald F	유의확률
(수정된 모형)	7.000	159.000	46.330	.000
(수정된 모형)	1.000	165.000	1135.891	.000
age_g	4.000	162.000	50.675	.000
sm_g	2.000	164.000	.799	.451
BMI	1.000	165.000	85.438	.000

모형

절편

부-모집단: 부모집단 = 30세 이상 남자

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g + sm_g + BMI

▶모형 효과 검정에서 p<0.05이므로 모형의 설명 유의

모수 추정값[a]

모수1	B 추정값	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한	가설검정 t	가설검정 자유도	가설검정 유의확률
(수정된 모형)	103.264	3.006	97.330	109.199	34.357	165.000	.000
[age_g=3]	-13.300	1.223	-15.714	-10.886	-10.878	165.000	.000
[age_g=4]	-11.305	1.247	-13.768	-8.843	-9.066	165.000	.000
[age_g=5]	-6.951	1.316	-9.549	-4.353	-5.282	165.000	.000
[age_g=6]	-2.990	1.436	-5.826	-.154	-2.082	165.000	.039
[age_g=7]	.000[b]	.	.	.	.	.	.
[sm_g=1]	-.813	.964	-2.717	1.091	-.843	165.000	.401
[sm_g=2]	-.924	.760	-2.424	.576	-1.216	165.000	.226
[sm_g=3]	.000[b]	.	.	.	.	.	.
BMI	1.055	.114	.830	1.280	9.243	165.000	.000

(기준집단)

부-모집단: 부모집단 = 30세 이상 남자

a. 모형: 최종 수축기 혈압(2,3차 평균) = (수정된 모형) + age_g + sm_g + BMI

b. 중복된 모수이므로 0으로 설정됩니다.

▶ 결과해석

연령그룹별 체질량지수 보정 시, 30세 이상 남자의 흡연 여부별 수축기 혈압 평균 차이는 현재 흡연 기준 비흡연, 과거흡연 모두 유의하지 않음

• 추정 모형 :

$E(HE_sbp) = 103.26 - 13.30*age_g_3 - 11.31*age_g_4 - 6.95*age_g_5 - 2.99*age_g_6 - 0.81*sm_g_1 - 0.92*sm_g_2 + 1.1*BMI$

복합표본 보정 평균

▪ 보정평균 : 회귀모형에서 다른 독립변수를 보정하여 산출된 평균

예제

연령그룹, 체질량지수 보정 시,
30세 이상 남자의 흡연 여부별 수축기혈압 평균 차이 유무

변수 설명

구분	변수	변수내용	구분	변수	변수내용
종속변수	HE_sbp	수축기혈압(연속형)	독립변수	sm_g	1. 비흡연
독립변수	age_g	3. 30대			2. 과거흡연
		4. 40대			3. 현재흡연
		5. 50대		BMI	체질량지수(연속형)
		6. 60대		subg (부-모집단)	1. 30세 이상 남자
		7. 70세이상			0. 그 외

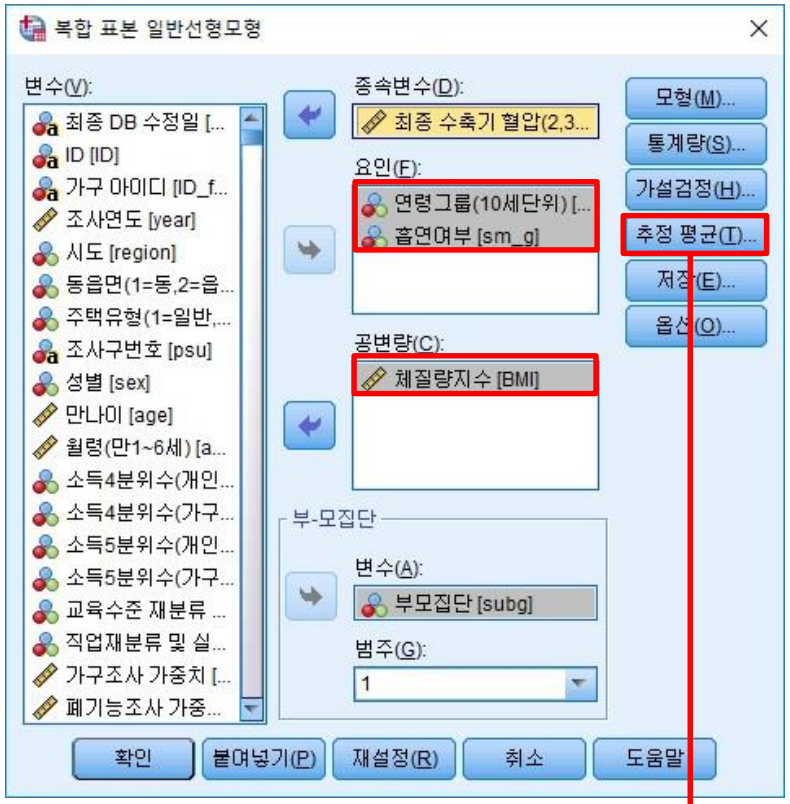
모형

$$HE_sbp = \beta_0 + \beta_1 * age_g + \beta_2 * sm_g + \beta_3 * BMI + \varepsilon$$

복합표본 보정 평균 예제

미보정

보정



복합표본 보정 평균 결과

보정 평균 산출 결과

추정값[a]

흡연여부	평균	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한
비흡연	121.59	.781	120.05	123.13
과거흡연	121.48	.589	120.31	122.64
현재흡연	122.40	.689	121.04	123.76

부-모집단: 부모집단 = 30세 이상 남자

a. 모형에 나타나는 공변량은 다음 값에서 고정됨: BMI=24.69

▶ 평균 추정값
연령그룹과 체질량지수 보정 후, 30세 이상 남자의 흡연 여부별 수축기혈압 평균

연속형 변수인 공변량(BMI)의 평균은 종속변수와 독립변수가 모두 존재하는 즉, 종속변수와 독립변수 모두 결측이 없는 자료만으로 자동으로 평균 산출

개별 검정 결과

흡연여부 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화 오차	자유도1	자유도2	Wald F	유의확률
수준 비흡연 및 수준 현재흡연	-.813	.000	-.813	.964	1.000	165.000	.710	.401
수준 과거흡연 및 수준 현재흡연	-.924	.000	-.924	.760	1.000	165.000	1.479	.226

부-모집단: 부모집단 = 30세 이상 남자

a. 참조범주 = 현재흡연

- **SPSS, SAS 간 보정평균 산출 방법 차이 존재 ⇒ 결과 산출 시 차이 발생**

복합표본 보정 평균_SPSS 및 SAS 비교

SPSS

Estimate Coefficients					
Effect	age_g	sm_g	Row1	Row2	Row3
Intercept			1	1	1
age_g	3		0.2	0.2	0.2
age_g	4		0.2	0.2	0.2
age_g	5		0.2	0.2	0.2
age_g	6		0.2	0.2	0.2
age_g	7		0.2	0.2	0.2
sm_g		1	1		
sm_g		2		1	
sm_g		3			1
BMI			24.685	24.685	24.685

≠

SAS

Estimate Coefficients					
Effect	age_g	sm_g	Row1	Row2	Row3
Intercept			1	1	1
age_g	3		0.2098	0.2098	0.2098
age_g	4		0.2438	0.2438	0.2438
age_g	5		0.2426	0.2426	0.2426
age_g	6		0.1652	0.1652	0.1652
age_g	7		0.1386	0.1386	0.1386
sm_g		1	1		
sm_g		2		1	
sm_g		3			1
BMI			24.686	24.686	24.686

- 연속형 변수 : 종속변수 및 독립변수의 값이 모두 결측이 없는 자료의 가중 평균
- 범주형 변수 : 1/K (K: number of level)

≠

- 연속형 변수 : 독립변수의 값이 결측이 없는 자료의 가중 평균
- 범주형 변수 : weighted proportion

복합표본 보정 평균_SPSS 및 SAS 비교

SPSS

추정값[a]

흡연여부	평균	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한
비흡연	121.59	.781	120.05	123.13
과거흡연	121.48	.589	120.31	122.64
현재흡연	122.40	.689	121.04	123.76

부-모집단: 부모집단 = 30세 이상 남자

a. 모형에 나타나는 공변량은 다음 값에서 고정됨: BMI=24.69

≠

SAS

sm_g Least Squares Means

sm_g	Margins	BMI	Estimate	Standard Error	DF	t Value
1	WORK.HN18	24.69	120.77	0.7704	165	156.77
2	WORK.HN18	24.69	120.66	0.5882	165	205.14
3	WORK.HN18	24.69	121.58	0.6727	165	180.73

개별 검정 결과

흡연여부 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화
수준 비흡연 및 수준 현재흡연	-.813	.000	-.813	
수준 과거흡연 및 수준 현재흡연	-.924	.000	-.924	.

부-모집단: 부모집단 = 30세 이상 남자

a. 참조범주 = 현재흡연

=

Differences of sm_g Le
Adjustment for Multiple

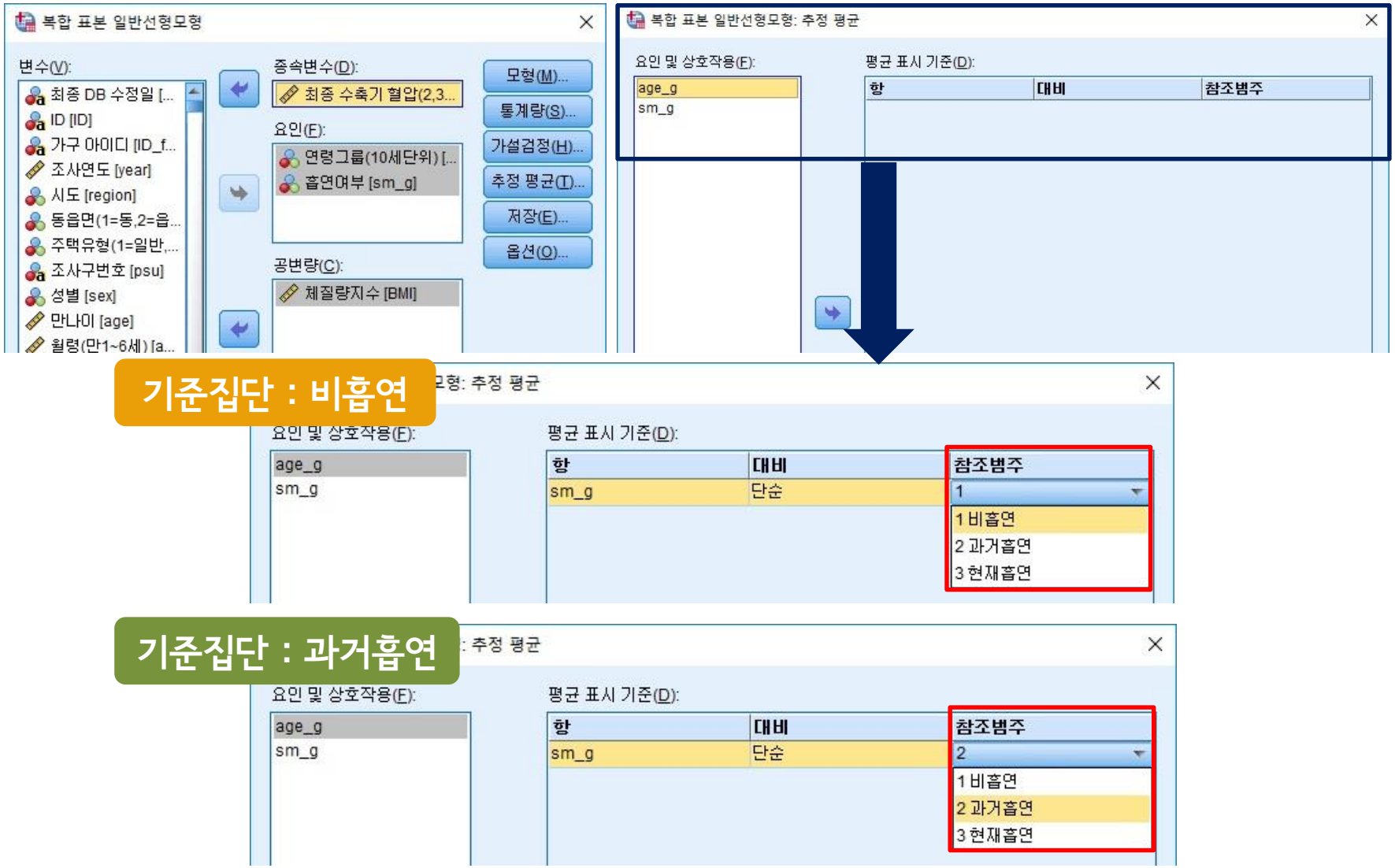
sm_g	_sm_g	Margins	BMI	Estimate	Standard Error	DF	t Va
1	2	WORK.HN18	24.69	0.1112	0.9407	165	
1	3	WORK.HN18	24.69	-0.8126	0.9643	165	-
2	3	WORK.HN18	24.69	-0.9238	0.7597	165	-

- 30세 이상 남자의 흡연 여부에 따른 수축기혈압 보정평균
  - SPSS : 비흡연 121.6mmHg, 과거흡연 121.5mmHg, 현재흡연 122.4mmHg
  - SAS : 비흡연 120.8mmHg, 과거흡연 120.7mmHg, 현재흡연 121.6mmHg

- 30세 이상 남자의 흡연 여부에 따른 수축기혈압 보정평균 추정값 결과는 분석 프로그램에 따라 다르게 산출되었으나, 보정평균 추정값의 차이는 모두 동일하게 산출

평균 비교 시 기준집단 변경 예제

▪ 기준집단 변경 : 추정평균 옵션 대화상자의 참조범주 이용



평균 비교 시 기준집단 변경 결과

추정값[a]

흡연여부	평균	표준화 오차	95% 신뢰구간 하한	95% 신뢰구간 상한
비흡연	121.59	.781	120.05	123.13
과거흡연	121.48	.589	120.31	122.64
현재흡연	122.40	.689	121.04	123.76

부-모집단: 부모집단 = 30세 이상 남자

a. 모형에 나타나는 공변량은 다음 값에서 고정됨: BMI=24.69

▶ [평균추정]의 결과로 30세 이상 남자의 흡연여부별 평균 추정값 표 출력

기준집단 : 비흡연

개별 검정 결과

흡연여부 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화 오차	자유도1	자유도2	Wald F	유의확률
수준 과거흡연 및 수준 비흡연	-.111	.000	-.111	.941	1.000	165.000	.014	.906
수준 현재흡연 및 수준 비흡연	.813	.000	.813	.964	1.000	165.000	.710	.401

부-모집단: 부모집단 = 30세 이상 남자

a. 참조범주 = 비흡연

▶ 기준집단과의 평균 차이 및 유의확률 출력

기준집단 : 과거흡연

개별 검정 결과

흡연여부 단순 대비[a]	대비 추정값	가설값	차이(추정값 - 가설값)	표준화 오차	자유도1	자유도2	Wald F	유의확률
수준 비흡연 및 수준 과거흡연	.111	.000	.111	.941	1.000	165.000	.014	.906
수준 현재흡연 및 수준 과거흡연	.924	.000	.924	.760	1.000	165.000	1.479	.226

부-모집단: 부모집단 = 30세 이상 남자

a. 참조범주 = 과거흡연

요약

- 자료 분석 시 복합표본설계를 기반으로 하는 **복합표본 분석방법 이용**
- 자료 분석 시 유의사항
  - 부-모집단 분석 : **관심집단과 그 외 집단을 구분하는 새로운 변수 생성** 후 생성된 관심 집단 변수를 부-모집단 변수에 지정하여 전체 자료 이용하여 분석
  - 결측 자료 처리 : **"유효한 값으로 처리"로 변경**하여 분석

- 복합표본설계 자료에서 종속변수가 연속형인 경우 독립변수 유형에 상관없이 SPSS 복합표본 일반선형모형 프로시저 사용

종속변수	독립변수	분석방법	검정방법	복합표본설계 SPSS 프로시저
연속형	범주형(2그룹)	T검정	T-TEST	복합표본 일반선형모형 (CSGLM)
	범주형(3그룹이상)	분산분석	ANOVA	
	연속형, 범주형	선형회귀분석	REG	
	연속형,범주형	일반선형모형	GLM (T-TEST, ANOVA, ANCOVA, REG)	

THANK YOU!