

국가승인통계
[승인번호 제11702호]

국민건강영양조사

원시자료 분석지침서(SPSS)

최초 공개 : 2013년 5월

이용자를 위하여

- ❶ 국민건강영양조사는 「국민건강증진법」 제16조에 근거하여 실시하는 법정조사로, 「통계법」 제17조에 근거한 정부 지정통계이며, 동법 제18조에 근거한 국가 승인통계입니다(승인번호 제11702호).

- ❷ 국민건강영양조사는 제1기(1998)부터 제3기(2005)까지 3년 주기로 실시하였고, 연중 조사체계로 개편 후 제4기(2007-2009)부터 매년 실시하고 있으며, 현재 제6기(2013-2015) 조사를 수행 중에 있습니다. 본 조사는 질병관리본부 연구윤리심의위원회의 승인을 받아 수행된 것이며(승인번호: 2007-02-CON-04-P, 2008-04EXP-01-C, 2009-01CON-03-2C, 2010-02CON-21-C, 2011-02CON-06-C), 조사결과의 공표는 조사익년 12월 보도자료 배포, 통계집 발간 및 원시자료 공개를 통해 이루어지고 있습니다.
※ 그 외 진행보고서(조사기획, 수행체계), 표본설계보고서, 질관리보고서, 기타 연구용역결과보고서는 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr/>) 자료실에서 다운로드 받으실 수 있습니다.

- ❸ 원시자료는 「개인정보 보호법」에 근거하여 조사대상자 개인을 추정할 수 있는 자료(조사일, 상세 행정구역명)를 제외하여 공개하고 있으며, 개인정보 보호법을 준수하는 범위 내에서 보건의료분야 연구 활성화를 위해 비공개자료를 분석할 수 있는 학술연구용 자료처리실을 운영하고 있습니다. 자료처리실 사용을 원하시는 경우, 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr>) 공지사항을 참조하시어 요청서 작성 후 담당자 이메일로 신청하여 주시기 바랍니다.

- ❹ 국민건강영양조사 원시자료 이용자는 「질병관리본부 원시자료 공개절차 등에 관한 규정」 및 「국민건강영양조사 원시자료 공개 및 활용 규정」을 준수해야 합니다. 특히 국민건강영양조사 원시자료를 활용하여 논문 및 학회발표, 포스터 게재 등이 이루어진 경우 해당 내용을 국민건강영양조사 홈페이지에 등록하여 주시기 바랍니다.

※원시자료 출처 표기 예

- (국문) • 국민건강영양조사 제5기 2차년도(2011), 질병관리본부
• 국민건강영양조사 제3기(2005), 질병관리본부
- (영문) • The Fifth Korea National Health and Nutrition Examination Survey (KNHANES V-2), 2011,
Korea Centers for Disease Control and Prevention
• The Third Korea National Health and Nutrition Examination Survey (KNHANES III), 2005,
Korea Centers for Disease Control and Prevention

- ⑤ 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr>) 공지사항 및 원시자료 요청시 등록 이메일을 통해 원시자료 공개(추가공개, 오류수정 포함), 조사결과 공표, 자료활용 워크숍 개최 등의 새소식을 공지하고 있사오니, 이메일 변동이 있을 경우 아래 연락처로 수정사항을 알려주시기 바랍니다.
- ⑥ 국민건강영양조사 원시자료는 복합표본분석방법을 권고하고 있으며, 이를 위해 SPSS13.0이상의 버전이 필요하오니 참고하시기 바랍니다.
- ⑦ 원시자료 문의사항은 건강영양조사과 통계팀(☎ 043-719-7465~8)으로 연락하시기 바랍니다.

목 차

01	분석 준비단계	1
02	복합표본설계 자료분석 개요	7
03	분석DB 준비단계	17
	• 자료결합: 자료의 가로결합	18
	• 자료통합분석	22
	-자료통합DB생성: 자료의 세로결합	22
	-자료통합DB생성: 통합가중치 산출	25
	• 식품섭취조사 개인단위 분석DB 생성	30
	• 기타	39
	-명령문 사용방법	39
	-변수보기 및 결측값 지정	42
	-변수 창 정렬	43
04	복합표본 자료분석	45
	• 개요	46
	• 분석계획 파일생성	47
	• 자료분석 시 유의사항	53
	• 복합표본 빈도분석	56
	• 복합표본 기술통계	63
	• 복합표본 교차분석	69
	• 복합표본 일반선형모형	82
	• 복합표본 로지스틱 회귀모형	103

01

분석 준비단계

□ 국민건강영양조사 원시자료 DB 구성

○ 원시자료 다운로드

- 제1기(1998)~제5기 2차년도(2011) 국민건강영양조사 원시자료는 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr/>)에서 SPSS, SAS DB형태로 다운로드 가능하다.

< 제1기(1998) ~ 제5기(2010-2012) 원시자료 구성 >

지침서	원시자료 DB (SAS, SPSS)
제1기(1998)	
<ul style="list-style-type: none"> • 원시자료 이용지침서 및 분석지침서 • 조사표 • 식품섭취조사 식품코드 및 음식코드 	<ul style="list-style-type: none"> - HN98_ALL : 건강설문, 검진, 영양 - HN98_IJMT : 손상 및 의료이용(건강설문) - HN98_DSRAW : 이환(건강설문) - HN98_24RC : 식품섭취조사(영양)
제2기(2001)	
<ul style="list-style-type: none"> • 원시자료 이용지침서 및 분석지침서 • 조사표 • 식품섭취조사 식품코드 및 음식코드 	<ul style="list-style-type: none"> - HN01_ALL : 건강설문, 검진, 영양 - HN01_IJMT : 손상 및 의료이용(건강설문) - HN01_DSRAW : 이환(건강설문) - HN01_24RC : 식품섭취조사(영양)
제3기(2005)	
<ul style="list-style-type: none"> • 원시자료 이용지침서 및 분석지침서 • 조사표 • 식품섭취조사 식품코드 및 음식코드 	<ul style="list-style-type: none"> - HN05_ALL : 건강설문, 검진, 영양 - HN05_IJMT : 손상 및 의료이용(건강설문) - HN05_24RC : 식품섭취조사(영양)
제4기(2007-2009)	
<ul style="list-style-type: none"> • 원시자료 이용지침서 및 분석지침서 • 조사표 • 식품섭취조사 식품코드 및 음식코드 	<ul style="list-style-type: none"> • 1차년도(2007) <ul style="list-style-type: none"> - HN07_ALL : 건강설문, 검진, 영양 - HN07_IJMT : 손상 및 의료이용(건강설문) - HN07_OE : 구강검사(검진) - HN07_24RC : 식품섭취조사(영양) • 2차년도(2008) 및 3차년도(2009) <ul style="list-style-type: none"> - HNYR_ALL : 건강설문, 검진, 영양 - HNYR_IJMT : 손상 및 의료이용(건강설문) - HNYR_OE : 구강검사(검진) - HNYR_DXA : 골밀도 및 체지방검사(검진) - HNYR_ENT : 이비인후검사(검진) - HNYR_EYE : 안검사(검진) - HNYR_24RC : 식품섭취조사(영양) <p>※ YR은 해당연도 두자리수 숫자 ※ 2008년 7월 안검사, 이비인후검사, 골밀도 및 체지방검사 신규도입</p>
제5기(2010-2012)	
<ul style="list-style-type: none"> • 원시자료 이용지침서 및 분석지침서 • 식품섭취조사 식품코드 및 음식코드 	<ul style="list-style-type: none"> • 1차년도(2010) 및 2차년도(2011) <ul style="list-style-type: none"> - HNYR_ALL : 건강설문, 검진, 영양 - HNYR_IJMT : 손상 및 의료이용(건강설문) - HNYR_OE : 구강검사(검진) - HNYR_DXA : 골밀도 및 체지방검사(검진) - HNYR_ENT : 이비인후검사(검진) - HNYR_EYE : 안검사(검진) - HNYR_24RC : 식품섭취조사(영양) <p>※ YR은 해당연도 두자리수 숫자</p>

○ 분석변수 및 범위 결정

- 「원시자료 이용지침서」 및 「조사표」 검토를 통해 조사항목, 응답보기, 대상연령, 준거기간 등을 연도별로 비교하여 분석 항목 및 범위 결정한다.

○ 분석변수의 상세내용 확인

- 「원시자료 이용지침서」에서 변수명, 해당변수 포함된 DB명, 코딩내용 확인한다.
- 대부분의 주요변수는 기본DB(HNYR_ALL, YR은 해당연도 2자리 숫자)에 포함되어 있으며, 손상 및 의료이용, 안검사, 이비인후검사, 골밀도 및 체지방검사, 구강검사, 식품섭취조사 자료는 별도 DB로 구분되어 있다.
- 가구 및 개인의 기본정보(성별, 연령, 가구원수, 소득 등)와 복합표본설계정보(가중치, 충화변수, 집락변수)는 모든 DB에 포함되어 있다.
- 「원시자료 이용지침서」에서 변수값, 결측치, 무응답, 비해당 유무를 확인하여 분석 시 고려하고, 필요시 분석목적에 적합한 새로운 변수를 생성하여 분석한다.

< 조사 부문별 DB 및 변수개수 >

(자료수 단위: 명, 건(식품섭취조사))

구분 (DB명) 자료수(변수개수)	제1기 (1998)	제2기 (2001)	제3기 (2005)	제4기		제5기	
				1차년도 (2007)	2차년도 (2008)	3차년도 (2009)	1차년도 (2010)
건강설문조사							
가구조사(HNYR_ALL)	39,060(47)	37,769(47)	34,145(57)	4,160(49)	9,744(60)	10,533(62)	9,688(40)
건강면접조사(HNYR_ALL)	39,060(78)	37,769(816)	34,145(355)	4,148(771)	9,281(771)	10,051(774)	8,388(443)
건강행태조사(HNYR_ALL)	10,808(111)	9,170(146)	14,165(162)				8,055(441)
손상및의료이용(HNYR_LJMT)	39,060(216)	37,769(548)	34,145(1,510)	4,148(830)	9,281(1,233)	10,051(874)	8,388(608)
이화카드(HNYR_DSRAW)	39,060(350)	37,769(414)	--	--	--	--	--
검진조사							
검진조사(HNYR_ALL)	9,771(70)	9,702(68)	7,597(130)	4,246(74)	9,307(82)	10,078(70)	8,473(166)
구강검사(HNYR_OE)	--	--	--	4,246(335)	9,307(346)	10,078(348)	8,473(341)
골밀도검사(HNYR_DXA)	--	--	--	--	3,583(156)	7,920(159)	7,043(140)
이비인후검사(HNYR_ENT)	--	--	--	--	4,592(132)	10,065(152)	8,313(142)
안검사(HNYR_EYE)	--	--	--	--	4,846(212)	9,760(217)	8,141(209)
영양조사							
영양조사(HNYR_ALL)	11,267(100)	10,000(123)	9,004(173)	4,099(205)	8,641(212)	9,397(215)	8,027(132)
식품섭취조사(HNYR_24RC)	454,841(50)	430,890(57)	420,794(60)	197,147(73)	418,628(87)	462,127(89)	441,890(82)

※DB명에서 YR은 해당연도 2자리 숫자 (예: 제5기 2차년도(2011) 이비인후검사 DB는 HN11_ENT)

※기본DB 부문별 변수개수 구분 : 공통변수(시도, 동읍면, 주택유형, 조사구, 소득4분위수, 성별, 만나이, 분산추정 층, 교육수준, 직업 분류, 가중치 등)를 가구조사 변수개수로 포함하여 계산

※손상 및 의료이용 DB(HNYR_LJMT) : 연도별 최대건수에 따라 변수개수 차이

	'07	'08'	'09'	'10'	'11'
- 손상 최대건수 :	4회	10회	5회	3회	3회
- 입원이용 최대건수 :	12회	6회	8회	14회	13회
- 외래이용 최대건수 :	13회	22회	15회	14회	20회
- 약국이용 최대건수 :	7회	12회	8회	10회	10회

〈 제5기(2010-2012) 조사항목 〉

조사부문	조사구분	해당연령	조사항목
건강설문	가구조사	만19세이상	성, 연령, 결혼상태, 가구원수, 세대유형, 가구소득, 건강보험 가입, 민간보험 가입
	교육	만1세이상	학력, 수학여부
	경제활동	만19세이상	경제활동여부, 미취업사유, 취업형태, 종사상지위, 근로시간형태, 직업, 최장 직업, 부모 학력
	이환	만1세이상	최근 2주간 이환, 천식증상 경험, 만성질환별(18개) 이환
	의료이용	만1세이상	외래이용, 입원이용, 약국이용
	건강검진	만19세이상	건강검진 수진, 암검진 수진
	예방접종	만1세이상	인플루엔자 예방접종
	활동제한	만1세이상	활동제한여부, 활동제한이유, 와병경험, 결근결석
	삶의 질	만19세이상	주관적 건강인지, EQ-5D(운동능력, 자기관리, 일상활동, 통증/불편, 불안/우울)
	손상	만1세이상	손상경험, 사고중독발생기전, 손상치료처, 손상으로 인한 와병, 손상으로 인한 결근결석
	흡연	만12세이상	평생흡연, 현재흡연, 과거흡연, 금연, 간접흡연
	음주	만12세이상	평생음주, 음주시작연령, 음주빈도, 음주량, 폭음빈도
	신체활동	만12세이상	격렬한신체활동, 중등도신체활동, 걷기실천, 근력운동, 유연성운동
	정신건강	만12세이상	수면시간, 스트레스인지, 우울증상, 자살, 정신문제상담
	안전의식	만1-11세	자동차보호장구 착용, 자전거헬멧 착용
		만12세이상	안전벨트 착용, 자전거헬멧 착용
	비만 및 체중조절	만6세이상	주관적 체형인지, 체중변화, 체중조절
	여성건강	만10세이상	현재월경여부, 초경연령
		만15세이상	임신경험, 출산경험
		만19세이상	모유수유 경험, 모유수유 자녀수 및 기간, 폐경연령, 경구피임약 복용경험 및 기간, 여성호르몬제 복용경험 및 기간
영양	식생활조사	만1세이상	결식, 외식 빈도, 가족동반식사여부, 식이보충제 복용경험, 식품지원프로그램 수혜, 식품안정성(가구조사)
		초등학생이상	영양표시 이용현황, 영양교육 및 상담 수혜
		만1~3세	출생정보, 수유 방법 및 기간, 이유식 및 영아기 식이보충제 섭취 정보
	식품섭취빈도조사	만12세이상	에너지 및 영양소 주요 급원식품(63항목)에 대한 섭취빈도
	식품섭취조사	조리자	조사1일전 가구에서 섭취한 음식에 대한 조리 내용
		만1세이상	조사1일전 섭취 음식(식이보충제 포함)의 종류 및 섭취량(24시간 회상법)
	신체계측	만1세이상	신장, 체중, 허리둘레
검진 ("10~'11)	혈압 및 맥박	만10세이상	수축기혈압, 이완기혈압, 맥박수
	혈액검사	만10세이상	총콜레스테롤, 중성지방, HDL콜레스테롤, LDL콜레스테롤, 혈중요소질소, 크레아티닌, 공복 혈당, 당화혈색소, 인슐린(10), B형간염표면항원, 지오티, 지피티, 감마지티피, 철, 철결합능, 혜모글로빈, 혜마토크릿, 페리틴, 비타민D, 알칼라인 포스포테이즈
		만10세이상	수은, 납, 카드뮴, 아연(10) (매년 2,400명)
		만50세이상	부갑상선 호르몬
	소변검사	만10세이상	요단백, 요당, 요점혈, 요비중, 요산도, 요크레아티닌, 요나트륨, 코티닌(매년 2,400명), 유로빌리노젠, 케톤, 빌리루빈, 아질산염, 미세알부민
	구강검사	만1세이상	치아상태, 보철물 상태, 치주조직 상태, 저작불편경험, 구강검진수진, 칫솔질 실천 등
	폐기능검사	만40세이상	FVC(Forced Vital Capacity), FEV ₁ (Forced Expiratory Volume)
	흉부X-선검사	만15세이상	결핵, 폐기종, 기흉
	안검사	만3세이상	시력 및 굴절이상, 사시, 안검하수, 세극등검사, 안압검사, 무산동안저촬영검사 ※ 필요시 산동후 안저촬영, 시야검사
	이비인후검사	만1세이상	안면신경마비, 전이개누공, 외이기형, 고막검사, 수축전비강검사, 수축후비강검사, 후두내시경, 어지럼증검사, 청력검사
	골밀도 및 체지방검사	만10세이상	골밀도, 체지방('08. 7월~'11. 5월)
	골관절염검사	만50세이상	무릎관절, 엉덩관절

□ 국민건강영양조사 보고서

- 분석을 위해 국민건강영양조사 결과 보고서, 진행 보고서, 질관리 보고서, 표본 설계 보고서를 참조할 수 있다.

< 국민건강영양조사 보고서 목록 >

구분	진행보고서 ¹⁾	결과보고서 ²⁾	표본설계보고서 ³⁾
제1기 (1998)	<ul style="list-style-type: none"> 1998년도 국민건강영양조사 진행 보고서 	<ul style="list-style-type: none"> 1998년도 국민건강영양조사 자료 처리 및 분석 – 총괄보고서 1998년도 국민건강영양조사 <ul style="list-style-type: none"> – 건강면접조사 – 건강검진조사 – 보건의식행태조사(20세이상 성인) – 여성건강실태 – 노인건강실태 – 청소년건강실태 1998년도 국민건강영양조사 결과 보고서(영양조사부문) 심층분석보고서 <ul style="list-style-type: none"> – 1998 국민건강영양조사 심층연계분석 – 1998년도 국민건강영양조사 심층 연계분석(영양조사부문) 	<ul style="list-style-type: none"> 1998년도 국민건강영양조사 자료 처리 및 분석 – 총괄보고서
제2기 (2001)	<ul style="list-style-type: none"> 2001년 국민건강영양조사(국민 건강부문) 진행보고서 2001년 국민건강영양조사 진행보고서(영양조사부문) 	<ul style="list-style-type: none"> 2001년 국민건강영양조사 <ul style="list-style-type: none"> – 총괄편 – 만성질병편 – 급성질병사고증독편 – 보건의식행태편 – 검진편 심층분석보고서 <ul style="list-style-type: none"> – 한국인의 주요 상병 및 건강행태 분석–2001년 국민건강영양조사 건강 부문 심층분석 결과 – 2001년도 국민건강영양조사 심층연계분석(영양조사부문) 	<ul style="list-style-type: none"> 2001년 국민건강영양조사 및 표본 설계
제3기 (2005)	<ul style="list-style-type: none"> 2005 국민건강영양조사 진행보고서 <ul style="list-style-type: none"> – 건강면접 및 보건의식행태부문 – 검진조사부문 2005 국민건강영양조사(영양조사부문) 	<ul style="list-style-type: none"> 국민건강영양조사 제3기(2005) <ul style="list-style-type: none"> – 총괄 – 성인이환 – 활동제한 및 삶의질 – 사고 및 증독 – 아동 및 청소년의 건강수준 및 보건 의식행태 – 성인보건의식행태 – 의료이용 – 영양조사(I), (II) 국민건강영양조사 제3기(2005)검진조사 심층분석보고서 <ul style="list-style-type: none"> – 국민건강영양조사 제3기 조사결과 심층분석 연구 : 건강면접 및 보건 의식부문 – 국민건강영양조사 제3기 조사결과 심층분석 연구보고서 : 검진부문 – 국민건강영양조사 제3기(2005) 심층 분석 : 영양부문 	<ul style="list-style-type: none"> 2004년도 국민건강영양조사 및 표본 설계

구분	진행보고서 ¹⁾	결과보고서 ²⁾	표본설계보고서 ³⁾
제4기 (2007– 2009)	<ul style="list-style-type: none"> 국민건강영양조사 진행보고서 1,II : 제4기 1차년도(2007) 	<ul style="list-style-type: none"> 2007 국민건강통계: 제4기 1차년도(2007) 2008 국민건강통계: 제4기 2차년도(2008) 2009 국민건강통계: 제4기 3차년도(2009) 	<ul style="list-style-type: none"> 제4기(2007–2009) 국민건강영양조사 표본설계
질관리보고서			
제4기 1차년도(2007)			
<ul style="list-style-type: none"> 국민건강영양조사 제4기 1차년도(2007) 임상검사사업 국민건강영양조사 제4기 1차년도(2007) 임상검사 정도관리 및 표준화사업 국민건강영양조사 제4기 1차년도(2007) 폐기능검사 인력훈련 및 정도관리사업 국민건강영양조사 제4기 1차년도(2007) 구강검사 인력훈련 및 정도관리사업 국민건강영양조사 제4기 1차년도(2007) 혈장수행 정도관리사업 			
제4기 2차년도(2008)			
<ul style="list-style-type: none"> 국민건강영양조사 제4기 2차년도(2008) 임상검사사업 국민건강영양조사 제4기 2차년도(2008) 진단의학검사(임상검사) 정도관리 실시 및 검사 표준지침서 개발 국민건강영양조사 제4기 2차년도(2008) 조사원 교육 및 정도관리사업: 결핵검사 국민건강영양조사 제4기 2차년도(2008) 조사원 교육 및 정도관리사업: 골밀도 검사 국민건강영양조사 제4기 2차년도(2008) 조사원 교육 및 정도관리사업: 안질환 검사 국민건강영양조사 제4기 2차년도(2008) 조사원 교육 및 정도관리사업: 이비인후질환 검사 국민건강영양조사 제4기 2차년도(2008) 조사원 교육 및 정도관리사업: 구강검사 국민건강영양조사 제4기 2차년도(2008) 혈장수행정도관리사업 국민건강영양조사 제4기 2차년도(2008) 영양조사 정도관리 및 조사결과 산출지원 			
제4기 3차년도(2009)			
<ul style="list-style-type: none"> 국민건강영양조사 제4기 3차년도(2009) 진단의학검사(임상검사) 사업 국민건강영양조사 제4기 3차년도(2009) 진단의학검사(임상검사) 수행기관 질관리 및 검사항목의 기준치 설정 국민건강영양조사 제4기 3차년도(2009) 호흡기계검사 조사원 교육 및 질관리사업 국민건강영양조사 제4기 3차년도(2009) 골밀도검사 조사원 교육 및 질관리사업 국민건강영양조사 제4기 2차년도(2009) 안검사 조사원교육 및 질관리 사업 국민건강영양조사 제4기 3차년도(2009) 이비인후검사 조사원 교육 및 질관리사업 국민건강영양조사 제4기 3차년도(2009) 구강검사 조사원 교육 및 질관리사업 국민건강영양조사 제4기 3차년도(2009) 골관절염 조사원 교육 및 질관리사업 			
제5기 (2010– 2012)		<ul style="list-style-type: none"> 2010 국민건강통계: 제5기 1차년도(2010) 2011 국민건강통계: 제5기 2차년도(2011) 	<ul style="list-style-type: none"> 제4기(2007–2009) 기중치 산출 및 제5기(2010–2012) 표본설계 국민건강영양조사 제6기 표본설계 및 제5기 기중치 산출
질관리보고서			
제5기 1차년도(2010)			
<ul style="list-style-type: none"> 국민건강영양조사 제5기 1차년도(2010) 골밀도검사 전문조사원(방사선사) 질관리 및 산출지표 표준화 국민건강영양조사 제5기 1차년도(2010) 구강검사 전문조사원(공중보건치과의) 교육 및 질관리 국민건강영양조사 제5기 1차년도(2010) 안(눈)검사 전문조사원(안과전공의) 질관리 및 산출지표 표준화 국민건강영양조사 제5기 1차년도(2010) 영양조사 전문조사원(영양사) 교육 및 질관리 국민건강영양조사 제5기 1차년도(2010) 이비인후(귀, 코, 목)검사 전문조사원(이비인후과전공의) 교육 및 질관리 국민건강영양조사 제5기 1차년도(2010) 진단의학검사 수행기관 질 관리 및 주요항목 측정 소급성(traceability) 조사 국민건강영양조사 제5기 1차년도(2010) 진단의학검사(임상검사) 국민건강영양조사 제5기 1차년도(2010) 호흡기계검사 전문조사원(간호사, 방사선사) 교육 및 질관리 			
제5기 2차년도(2011)			
<ul style="list-style-type: none"> 국민건강영양조사 제5기 2차년도(2011) 골관절염검사 전문조사원(방사선사) 교육 및 질관리 국민건강영양조사 제5기 2차년도(2011) 구강검사 전문조사원(공중보건치과의) 교육 및 질관리 국민건강영양조사 제5기 2차년도(2011) 안(눈)검사 전문조사원(안과전공의) 교육 및 질관리 국민건강영양조사 제5기 2차년도(2011) 영양조사 수행 및 자료 산출 질관리 국민건강영양조사 제5기 2차년도(2011) 이비인후(귀, 코, 목)검사 전문조사원(이비인후과전공의) 교육 및 질관리 국민건강영양조사 제5기 2차년도(2011) 진단의학검사 수행기관 질 관리 국민건강영양조사 제5기 2차년도(2011) 진단의학검사(임상검사) 국민건강영양조사 제5기 2차년도(2011) 혈입측정 질관리 국민건강영양조사 제5기 2차년도(2011) 호흡기계검사 전문조사원(간호사, 내과의사, 방사선사) 교육 및 질관리 			

1)조사 수행체계, 자료수집방법, 조사항목, 지침서 등; 2)결과표(당해년도 현황 및 연도별 추이), 결과 해석, 지표정의, 대상자 현황, 분석방법 등); 3)모집단, 표본추출틀, 표본추출방법, 모수추정, 사후관리 등

02

복합표본설계 자료분석 개요

□ 국민건강영양조사 원시자료 복합표본분석 개요

○ 표본설계

- 국민건강영양조사 표본 지역(조사구)은 표본의 대표성 및 추정의 정확성 향상을 위해 복합표본설계방법인 다단계총화집락학률추출법으로 추출하였다.
- 이는 단순임의표본설계와 대비되는 개념이며, 복합표본분석을 위해 복합표본설계의 3요소인 가중치(변수명: wt_itvex, wt_ntr 등), 층(변수명: kstrata), 집락(변수명: psu) 변수들을 원시자료 DB에 포함하여 공개하고 있다.

< 연도별 표본설계 >

구분	제1기(1998)	제2기(2001)
조사기간	1998년 11~12월 (2개월)	2001년 11~12월 (2개월)
추출틀	인구주택총조사(1995), 신축아파트목록(1997)	인구주택총조사(2000)
추출단위	조사구 → 가구	조사구 → 가구
총화변수	통합시(군), 동읍면	통합시도, 동읍면, 주택유형
표본크기	<ul style="list-style-type: none"> • 건강면접조사: 200조사구, 12,000가구 • 건강행태/검진/영양조사: 200조사구, 4,000가구 <p>※연간 조사구당 60가구(건강면접조사), 20가구 (건강행태/검진/영양조사) 조사</p>	<ul style="list-style-type: none"> • 건강면접조사: 600조사구, 12,000가구 (통합조사구 200, 비통합조사구 400) • 건강행태/검진/영양조사: 200통합조사구, 4,000가구 <p>※연간 조사구당 20가구 조사</p>
구분	제3기(2005)	제4기(2007~2009)
조사기간	2005년 4~6월 (3개월)	2007년 7~12월, 2008~2009년 1~12월
추출틀	인구주택총조사(2000), 신축아파트목록(2001)	인구주택총조사(2005)
추출단위	조사구 → 가구	동읍면 → 조사구 → 가구
총화변수	시도, 동읍면, 주택유형	시도, 동읍면, 주택유형 (내재적층: 권역, 연령)
표본크기	<ul style="list-style-type: none"> • 건강면접조사: 600조사구, 12,000가구 (통합조사구 200, 비통합조사구 400) • 건강행태/검진/영양조사: 200통합조사구, 4,000가구 <p>※연간 조사구당 20가구 조사</p>	<ul style="list-style-type: none"> 600조사구, 13,800가구 (연간 200조사구, 4,600가구) <p>※2007년은 반년조사로 100조사구, 2,300가구 완료 ※연간 조사구당 23가구 조사</p>
구분	제5기(2010~2012)	제6기(2013~2015)
조사기간	2010년 1월~2012년 12월	2013년 1월~2015년 12월
추출틀	주민등록인구(2009), 아파트목록(2008)	인구주택총조사(2010)
추출단위	조사구 → 가구	조사구 → 가구
총화변수	시도, 동읍면, 주택유형 (내재적층: 일반- 성별, 연령; 아파트- 평당가격, 평균평수 등)	시도, 동읍면, 주택유형 (내재적층: 성별, 연령, 주거면적, 가구주 학력)
표본크기	<p>총 576조사구, 11,520가구 (연간 192조사구, 3,840가구)</p> <p>※연간 조사구당 20가구 조사</p>	<p>총 576조사구, 11,520가구 (연간 192조사구, 3,840가구)</p> <p>※연간 조사구당 20가구 조사</p>

○ 분석방법

- 국민건강영양조사 자료를 분석하는 경우에는 복합표본설계 정보를 고려한 분석방법을 사용해야 하는데, 이를 고려하지 않는 경우 추정치(평균, 유병률, 오즈비 등) 및 이들의 분산(표준오차) 추정치에서 편향된 결과를 얻을 수 있다.
- 아래 표와 같이 복합표본분석에서 평균 계산 시 가중치가 반영된 가중평균으로 계산되는데, 단순임의표본분석에서는 단순평균으로 계산되고 또한 복합표본분석에서 분산 추정 시 가중치, 총, 집락(조사구)의 정보가 반영되어 추정되나, 단순임의표본분석에서는 이들 정보가 이용되지 않고 계산된다.
- 따라서 국민건강영양조사 자료분석 시에는 **복합표본분석방법을 사용하여야 하며, SPSS 복합표본분석은 「분석→복합표본」 메뉴를 선택하여 수행한다.**

〈 복합표본설계 추정식 〉

	단순임의설계	복합표본설계
평균 (비율) 추정	$\hat{Y} = \frac{\sum_{k=1}^n y_k}{n}$	$\hat{Y} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \times y_{hij}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}}$
분산 추정	$\hat{V}(\hat{Y}) = \frac{\sum_{k=1}^n (y_k - \hat{Y})^2}{n-1}$	$\hat{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi.} - \bar{e}_{h..})^2$ 여기서, $f_h = n_h/N_h$, $\bar{e}_{h..} = (\sum_{i=1}^{n_h} e_{hi.})/n_h$, $e_{hi.} = [\sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \hat{Y})]^2 / w_{..}$
기호	(h: 총, i: 조사구, j: 가구) n : 전체 완료자수 y_{hij} : h th 총, i th 조사구, j th 참여자의 응답값 w_{hij} : h th 총, i th 조사구, j th 참여자의 가중치 N_h : h th 총의 모집단 조사구수 n_h : h th 총의 표본 조사구수 m_{hi} : h th 총, i th 조사구 완료자수	

□ 가중치

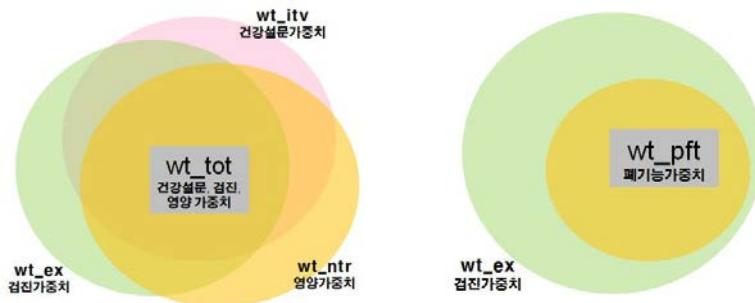
○ 가중치 개요

- 국민건강영양조사 가중치는 추정치가 우리나라 전체 모집단을 대표하도록 부여하는 확대승수로, 추출률, 응답률, 모집단 분포 등을 반영하여 산출한다.
- 가중치는 **조사부문별**로(건강설문조사, 검진조사, 영양조사) 부여하되, 조사항목별 대상자수 또는 참여자수 고려가 필요한 경우 **조사항목별** 가중치를 부여하였다.
 - **제1기(1998)~제3기(2005)** : 건강면접조사 대상자의 1/3에 대해서만 건강행태조사를 실시하여, 건강면접조사 가중치(wt_itv)와 건강행태조사 가중치(wt_bhv) 별도 부여
 - **제4기(2007-2009)** : 동일한 대상자에 대하여 건강면접조사와 건강행태조사를 실시하여 건강설문조사 가중치(wt_itv) 부여
 - **제5기(2010-2012)** : 건강설문조사 및 검진조사 완료자가 거의 일치하여 건강설문조사 또는 검진조사 참여자에 대해 건강설문-검진조사 가중치(wt_itvex) 부여
※동일한 조사부문 가중치라 하더라도 변수명 변경에 주의하여 분석
 - **중금속검사** : 표본 중 매년 2,000~2,400명을 선정하여 검사하므로 중금속검사 참여자에 대해 별도의 중금속검사 가중치(wt_hm) 부여
 - **폐기능검사** : 판독불능 등의 신뢰성이 낮은 검사결과 자료 제외로 인해 검진조사 자료수와 폐기능검사 자료수 차이가 발생하여 별도의 폐기능검사 가중치(wt_pft) 부여
 - **안검사, 이비인후검사, 흉부X선검사, 골밀도 및 체지방검사**는 2008년 7월 신규도입, 골관절염검사는 2009년 8월 신규도입, 골밀도 및 체지방검사는 2011년 5월 완료 등, 검진항목이 연중 신규도입 또는 완료되어 전체 검진조사 자료수와 차이가 발생하는 경우 별도 가중치 부여

기중치 종류	기중치 변수명							
	제1기 (1998)	제2기 (2001)	제3기 (2005)	제4기			제5기	
기구 가중치	—	—	wt_hs	wt_hs	wt_hs	wt_hs	wt_hs	wt_hs
건강설문조사								
건강면접조사			wt_itv			wt_itv		
건강행태조사			wt_bhv					
검진조사								
검진조사	wt_ex	wt_ex	wt_ex	wt_ex	wt_ex	wt_ex	wt_jtvex	wt_itvex
검진보충조사		wt_exhs	—	—	—	—		
안/이비인후/흉부X선검사	—	—	—	—	wt_ex1	wt_ex	wt_jtvex	wt_itvex
골밀도검사					wt_ex1	wt_ex	wt_jtvex	wt_ex1
체지방검사	—	—	—	—	wt_ex1	wt_dw	wt_jtvex	wt_ex1
골관절염검사	—	—	—	—	—	wt_ex1	wt_jtvex	wt_itvex
폐기능검사	—	—	—			wt_pft		
중금속검사	—	—	wt_hm	—			wt_hm	
영양조사								
식생활조사			wt_life					
영유아식생활조사			wt_ylife					
식품섭취빈도조사			wt_ffq					
24시간화장조사			wt_24rc				wt_ntr	

- 각 조사부문(또는 조사항목) 참여자에게 부여한 **기본가중치**와 2개 이상의 조사부문(또는 조사항목)에 모두 참여한 대상자에게 부여하는 **연관성 가중치**가 있다.
 - 제4기(2007-2009) : 건강설문조사 참여자는 건강설문조사 가중치(wt_itv)를, 검진조사 참여자는 검진조사 가중치(wt_ex)를, 영양조사 참여자는 영양조사 가중치(wt_ntr)를 부여하고, 아래 그림과 같이 건강설문조사, 검진조사, 영양조사 3개 부문 참여자는 건강설문-검진-영양조사 연관성분석 가중치(wt_tot)도 별도로 부여. 따라서 건강설문조사, 검진조사, 영양조사 3개 부문의 변수가 모두 분석에 포함되는 경우, 건강설문-검진-영양조사 연관성분석 가중치(wt_tot)를 사용
 - 제5기(2010-2012) : 검진조사와 건강설문조사의 참여자수가 거의 일치하여 두 조사의 가중치를 통합하였으므로, 두 조사부문 중 1개 조사부문의 변수가 분석에 포함되는 경우와 2개 조사부문의 변수가 분석에 포함되는 경우 모두 건강설문-검진조사 가중치(wt_itvex)를 사용

〈부문별 포함관계에 따른 연관성분석 가중치의 산출〉

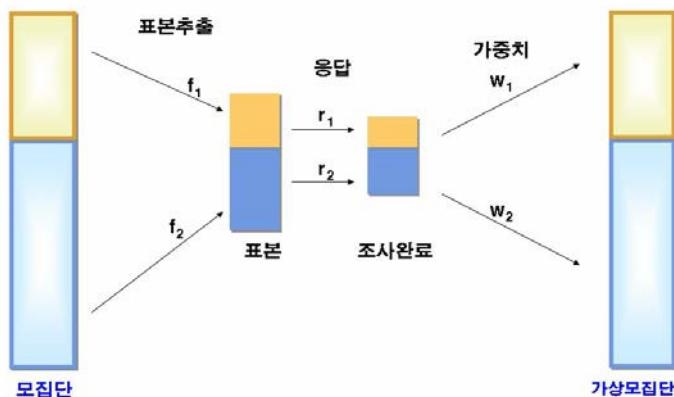


- 검진조사 가중치(제4기 wt_ex, 제5기 wt_itvex)를 사용하는 자료(신체계측, 혈압측정, 혈액검사, 소변검사)와 별도의 가중치가 부여된 검진조사 자료를 함께 분석하는 경우에는 별도의 가중치가 부여된 검진항목의 참여자가 전체 검진조사 참여자 안에 포함되므로, 별도 부여된 검진항목 가중치를 사용한다.
 - 기본 검진조사 가중치를 사용하는 자료와 폐기능검사 자료를 함께 분석하는 경우 : 폐기능검사 참여자와 전체 검진조사 참여자의 교집합은 폐기능검사 참여자이므로 폐기능검사 가중치(wt_pft)를 사용
 - 제5기 2차년도(2011) 자료에서 흡연(건강설문조사), 혈압(검진조사), 영양소(영양조사) 변수를 함께 분석하는 경우 : 건강설문-검진-영양 가중치(wt_tot)를, 흡연(건강설문조사), 혈압(검진조사), 폐기능(검진조사) 자료를 함께 분석하는 경우에는 폐기능검사 가중치(wt_pft)를 사용

〈 연관성분석 가중치 〉

기중치 종류	제3기 (2005)	기중치 변수명				
		제4기		제5기		
		1차년도 (2007)	2차년도 (2008)	3차년도 (2009)	1차년도 (2010)	2차년도 (2011)
부문별 연관성분석						
건강설문(건강행태)-검진	wt_bhvx			wt_jtvex		
건강설문(건강행태)-영양	wt_bhvnt	wt_jtvnt				
검진-영양	wt_exnt	wt_exnt			wt_tot	
건강설문-검진-영양	wt_tot	wt_tot				
폐기능	(건강설문/검진)	—			wt_pft	
	(건강설문/검진)-영양	—			wt_pfnt	
	중금속	—	—		wt_pfhm	
	중금속-영양	—	—		wt_pfhmnt	
중금속	(건강설문/검진)	wt_hm	—		wt_hm	
	(건강설문/검진)-영양	—	—		wt_hmnt	
	폐기능	—	—		wt_pfhm	
	폐기능-영양	—	—		wt_pfhmnt	
특정기간 조사항목	※당해년도에만 특정기간동안 조사하고, 이후부터 연간 조사			안/이비인후/ 골밀도/체지방/흉부 X선(08.7~12월)	골관절염검사 (09.8~12월)	-
	(건강설문/검진)	—	—	wt_ex1	wt_ex1	—
	(건강설문/검진)-폐기능	—	—	wt_ex1pf	wt_ex1pf	—
	(건강설문/검진)-중금속	—	—	wt_ex1hm	wt_ex1hm	—
	(건강설문/검진)-영양	—	—	wt_ex1nt	wt_ex1nt	—
	(건강설문/검진)-폐기능-중금속	—	—	wt_tot1	wt_tot1	—
	(건강설문/검진)-폐기능-영양	—	—	wt_ex1pfnt	wt_ex1pfnt	—
	(건강설문/검진)-중금속-영양	—	—	wt_ex1hmnt	wt_ex1hmnt	—
	(건강설문/검진)-폐기능-중금속-영양	—	—	wt_tot1nt	wt_tot1nt	—

〈 가중치의 구성 〉



○ 가중치 산출방법

- 국민건강영양조사 제3기(2005) 이후부터 가중치는 설계가중치 계산, 무응답률 조정, 사후충화, 극단가중치 처리를 거쳐 최종 산출한다.

1) 설계가중치 계산 : 표본조사구 추출률과 표본가구 추출률 곱의 역수로 표본추출율 반영

(예) 제5기 2차년도(2011) : 모집단 조사구중 총 192개 표본조사구 추출, 표본조사구내 적절가구중 20개 표본가구 추출, 표본가구에서 부적절가구원(만세미만, 타지역거주 등) 제외한 모든 가구원을 조사대상으로 선정

$$\text{설계가중치(추출률 역수)} = \frac{\text{모집단 조사구수}}{\text{표본 조사구수}} \times \frac{\text{조사구내 적절가구수}}{\text{표본 가구수}}$$

2) 응답률 조정 : 표본조사구내 가구 응답률과 표본가구내 개인응답률 곱의 역수로 응답률 반영

$$\text{무응답조정가중치(응답률 역수)} = \frac{\text{조사대상 가구수}}{\text{참여 가구수}} \times \frac{\text{조사대상 가구원수}}{\text{참여 가구원수}}$$

3) 기본가중치 계산 : 추출률과 응답률 역수의 곱으로 기본가중치 산출

$$\text{기본 가중치} = \text{추출률 역수} \times \text{응답률 역수}$$

4) 가중치 사후보정 : 우리나라 인구(가구)수에 근거하여 기본가중치 사후보정

※ 사후보정 결과 가구가중치는 조사 참여가구의 가중치 합이 우리나라 전체 가구수와 같아지도록, 부문별 개인가중치(건강설문검진조사, 영양조사 등)는 조사 참여자수의 가중치 합이 우리나라 전체 인구수와 같아짐

$$\text{사후보정가중치} = \text{기본 가중치} \times \frac{\text{시도} \cdot \text{성별} \cdot \text{연령별 추계인구수}}{\text{시도} \cdot \text{성별} \cdot \text{연령별 가중치 합}}$$

5) 극단가중치 처리 및 최종가중치 산출 : 사후보정가중치 값이 ($Q1 - 2 \times IQR$, $Q3 + 2 \times IQR$) 범위를 벗어나는 경우, 하한($Q1 - 2 \times IQR$)보다 작은 경우 하한값을, 상한($Q3 + 2 \times IQR$)보다 큰 경우 상한값을 부여. 이러한 극단가중치 절사 후 가중치 사후보정 과정을 거쳐 최종 가중치 산출

※ Q1: 1사분위수(25백분위수), Q3: 3사분위수(75백분위수), IQR: 사분위수범위($Q3 - Q1$)

□ **증화변수 : 제1기(1998) ~ 제5기(2010-2012)**

- 일부 표본설계층을 통합하여 생성한 분산추정층(변수명: kstrata)을 기본 증화변수로 사용한다.
- 단, 2008년 7월 신규도입된 안검사, 이비인후검사, 골밀도 및 체지방검사, 흉부X-선검사 자료분석 시에는 2008년 하반기 분산추정층 변수(kstrata1)를, 2009년 8월 신규도입된 골관절염검사 자료분석 시에는 2009년 하반기 분산추정층 변수(kstrata2)를 사용한다.

※kstrata1, kstrata2는 1개 조사구만 있는 일부 표본설계층을 통합한 분산추정용 층이므로, 신규도입 항목에 대해 도입연도에 kstrata를 사용하여 분석할 경우 분산 추정 편의 발생 가능

□ **집락변수 : 제1기(1998) ~ 제5기(2010-2012)**

- 집락변수는 1차 표본추출단위를 의미하므로 제1기(1998)~제5기(2010-2012) 조사구 변수(PSU)를 지정한다.
※제4기(2007-2009) 1차 추출단위 동읍면, 2차 추출단위 조사구이나, 1개 동읍면당 1개 조사구만을 추출하기 때문에 집락변수에 조사구 지정

〈 분산추정을 위한 층 〉

제2기(2001)				제3기(2005)				제4기(2007-2009)				제5기(2010-2012)							
구분		(신규) kstrata	(기존) kstrata0	구분		(신규) kstrata	(기준) kstrata0	구분		(신규) kstrata	(기준) kstrata1	(09골관) kstrata2	(기준) kstrata0	구분		(신규) kstrata	(기준) kstrata1	(기준) kstrata2	(09골관) kstrata0
1. 서울	동	아파트	201	201	1. 서울	동	아파트	301	301	1. 서울	412	410	48	417	1. 서울	512	510	508	517
	동	일반	202	202		동	일반	302	302	2. 부산	413	411	49	418	2. 부산	513	511	509	518
2. 부산	동	아파트	203	203	2. 부산	동	아파트	303	303	3. 대구	414	412	410	419	3. 대구	514	512	510	519
	동	일반	204	204		동	일반	304	304	4. 인천	415	413	411	420	4. 인천	515	513	511	520
3. 대구	동	아파트	205	205	3. 대구	동	아파트	305	305	5. 광주	416	414	412	421	5. 광주	516	514	512	521
	동	일반	206	206		동	일반	306	306	6. 대전	417	415	413	422	6. 대전	517	515	513	522
4. 인천	동	아파트	207	207	4. 인천	동	아파트	307	307	7. 울산	418	416	(414)	423	7. 울산	518	516	(514)	523
	동	일반	208	208		동	일반	308	308	8. 경기	419	417	415	424	8. 경기	519	517	515	524
5. 광주	동	아파트	209	209	5. 광주	동	아파트	309	309	9. 강원	420	418	416	425	9. 강원	520	518	516	525
	동	일반	210	210		동	일반	310	310	10. 충북	421	419	417	426	10. 충북	521	519	517	526
6. 대전	동	아파트	211	211	6. 대전	동	아파트	311	311	11. 충남	422	420	417	427	11. 충남	522	520	5517	527
	동	일반	212	212		동	일반	312	312	12. 전북	423	421	412	428	12. 전북	523	521	512	528
7. 울산	동	아파트	213	213	7. 울산	동	아파트	313	313	13. 전남	416	414	412	429	13. 전남	516	514	512	529
	동	일반	214	214		동	일반	314	314	14. 경북	424	422	416	430	14. 경북	524	522	516	530
8. 경기	읍면	일반	215		8. 경기	동	아파트	315	315	15. 경남	418	416	414	431	15. 경남	518	516	514	531
	동	아파트	216			동	일반	316	316	16. 제주	425	423	418	432	16. 제주	525	523	518	532
	동	일반	216			읍면	아파트	317	317										
	읍면	아파트	217	218		읍면	일반	318	318										
	읍면	일반	218	219															
	동	아파트	219	220															
9. 강원	동	일반	220	221															
	읍면	일반	221	222															
10.11.	동	아파트	222	223															
충북, 충남	동	일반	223	224	10.11. 충북, 충남	읍면	아파트	322	324	4. 인천	404	404	404	404	4. 인천	504	504	504	504
	읍면	아파트	224	225		동	일반	323	323	5. 광주	405	409	405	405	5. 광주	505	509	505	505
	읍면	일반	225	226		읍면	일반	324	325	6. 대전	406	405	406	406	6. 대전	506	505	506	506
	동	아파트	226	227						7. 울산	407	406	407	407	7. 울산	507	506	507	507
12.13.	동	일반	227	228	12.13. 전북, 전남	동	일반	326	327	8. 경기	408	407	404	408	8. 경기	508	507	504	508
전북, 전남	읍면	아파트	226	229		읍면	아파트	325	328	9. 강원	409	408	403	409	9. 강원	509	508	503	509
	읍면	일반	228	230		읍면	일반	327	329	10. 충북	10	405	406	410	10. 충북	510	505	506	510
16.제주	동	일반	235		16.제주	동	아파트	325	334	11. 충남	406	405	406	411	11. 충남	506	505	506	511
	읍면	일반	229	236		읍면	일반	328	335	12. 전북	411	409	405	412	12. 전북	511	509	505	512
14.15.	동	아파트	230	231	14.15. 경북, 경남	동	아파트	329	330	13. 전남	411	409	405	413	13. 전남	511	509	505	513
경북, 경남	동	일반	231	232		읍면	일반	330	331	14. 경북	409	408	403	414	14. 경북	509	508	503	514
	읍면	아파트	232	233		읍면	아파트	331	332	15. 경남	407	406	407	415	15. 경남	507	506	507	515
	읍면	일반	233	234		읍면	일반	332	333	16. 제주	411	409	(405)	416	16. 제주	511	509	(505)	516

※ 제1기(1998) 분산추정 총

시도, 읍면동 구분	kstrata
6대 시(서울, 부산, 대구, 인천, 광주, 대전)의 동	101
기타 시(경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주)의 동	102
시(경기, 강원, 충남, 전북, 전남, 경북, 경남)의 읍면	103
군(경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주)의 읍	104
군(경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남)의 면	105

03

분석DB 준비단계

- 자료결합: 자료의 가로결합
- 자료통합분석
 - 자료통합DB생성: 자료의 세로결합
 - 자료통합DB생성: 통합가중치 산출
- 식품섭취조사 개인단위 분석DB 생성
- 기타
 - 명령문 사용방법
 - 변수보기 및 결측값 지정
 - 변수창 정렬

□ 자료결합: 자료의 가로결합

- 국민건강영양조사 원시자료는 기본조사 DB 및 분야별 DB로 구분되어 있다. 자료분석시 분석변수가 여러 DB에 포함되어 있는 경우, 해당 DB들을 가로 결합하여 1개의 새로운 DB로 생성한다.
- 국민건강영양조사 원시자료에서 개인별 조사자료는 개인아이디(ID)로 구분되어 있고, DB별 개인아이디가 동일한 경우는 동일인에 대한 자료를 의미한다. 따라서 2개 이상의 DB를 1개의 DB로 결합하려면 개인아이디를 기준으로 정렬한 후 붙이면 된다.
- 두개의 DB를 가로결합하는 방법은 **1. 자료준비(자료열기), 2. 케이스 정렬, 3. 파일 합치기(변수추가)** 단계로 진행한다.

자료의 가로결합 예제

- 2011년 국민건강영양조사 기본조사DB(HN11_ALL)와 구강DB (HN11_OE) 자료를 결합하여 1개의 DB로 생성하는 과정은 다음과 같다.

자료 준비

케이스 정렬

파일 합치기(변수추가)

- 자료를 결합할 기본조사DB(HN11_ALL)와 구강조사DB(HN11_OE)를 활성화한다.

HN11_ALL.sav [데이터집합1] - PASW Statistics Data Editor

	ID	ID_fam	year	region	town_t
1	A951929601	A9519296	2011	1	1
2	A951929603	A9519296	2011	1	1
3	A951930002	A9519300	2011	1	1
4	A951930201	A9519302	2011	1	1
5	A951930202	A9519302	2011	1	1

HN11_OE.sav [데이터집합2] - PASW Statistics Data Editor

	ID	ID_fam	year	region	town_t
1	A951929601	A9519296	2011	1	1
2	A951930002	A9519300	2011	1	1
3	A951930201	A9519302	2011	1	1
4	A951930202	A9519302	2011	1	1
5	A951930203	A9519302	2011	1	1

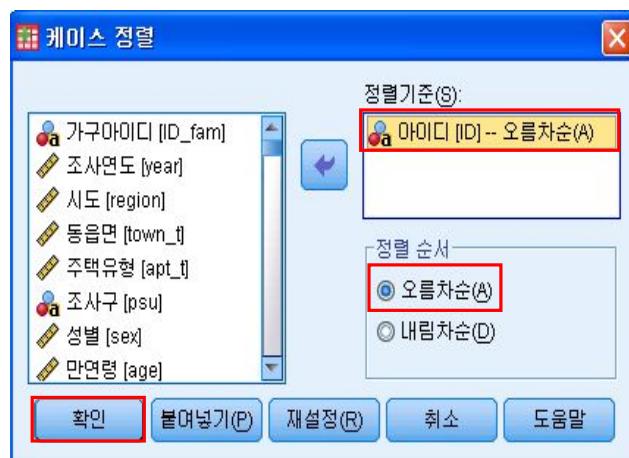


- DB 결합의 기준변수인 개인아이디(ID)로 오름차순 정렬을 실행한다.

1) 기본조사 DB(HN11_ALL)를 활성화 한 상태에서 **데이터** ► **케이스 정렬** 을 선택한다.



2) 정렬 기준은 **아이디(ID)**, 정렬순서는 **오름차순** 선택 후 **확인** 을 클릭한다.



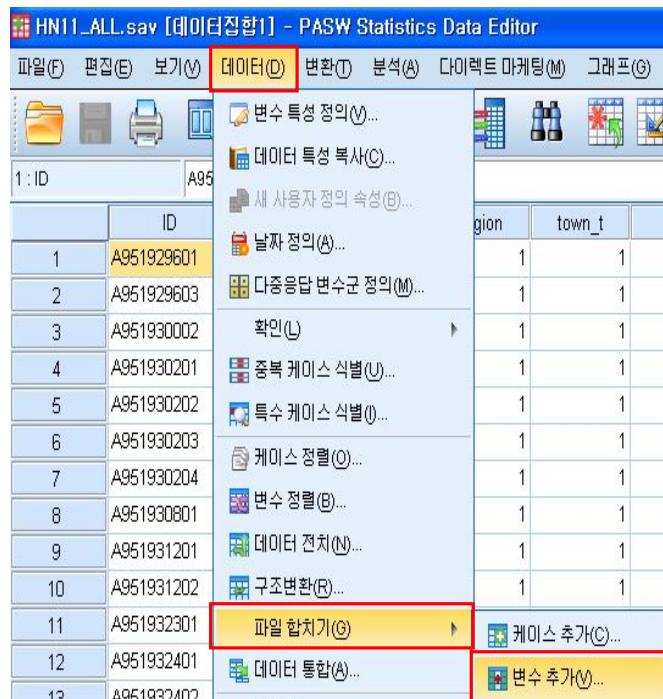
3) 구강조사 DB(HN11_OE)도 아이디(ID) 변수로 오름차순 케이스 정렬을 실행한다.

자료 준비하기 **케이스 정렬** **파일 합치기(변수추가)**

- 기본조사DB와 구강조사DB를 ID변수로 케이스 정렬 후, 기본조사DB를 기준 DB로 활성화하고 아래의 과정으로 진행한다.

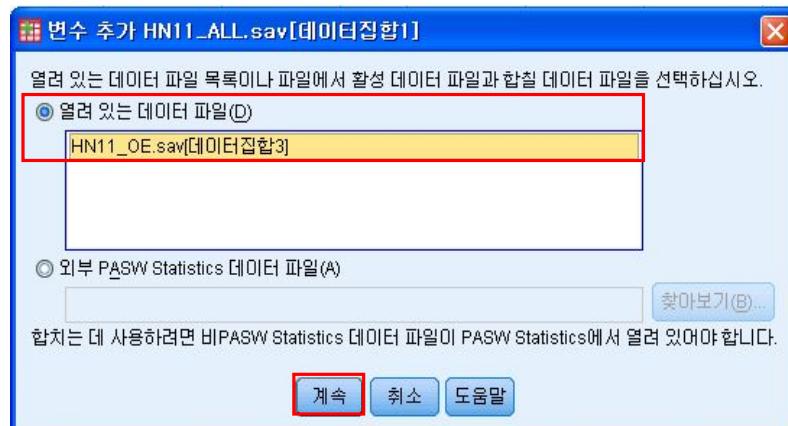
※ 기본조사DB를 기준DB로 하여 구강조사DB의 일부변수를 붙이는 경우는 기본조사DB를 활성화하고, 그 반대는 구강조사 DB를 활성화함

- 1) 기본조사DB(HN11_ALL) 활성화 상태에서 메뉴 창의 **데이터** ► **파일 합치기**
 ► **변수추가** 를 선택한다.



- 2) 변수 추가 창의 **열려 있는 데이터 파일**에서 구강조사DB(HN11_OE)를 선택 후 **계속** 을 클릭한다.

※ 구강조사DB(HN11_OE)를 케이스 정렬 후 파일을 닫으면 <열려있는 데이터 파일>에 나타나지 않으므로 케이스 정렬 후 파일은 닫지 않아야 한다.

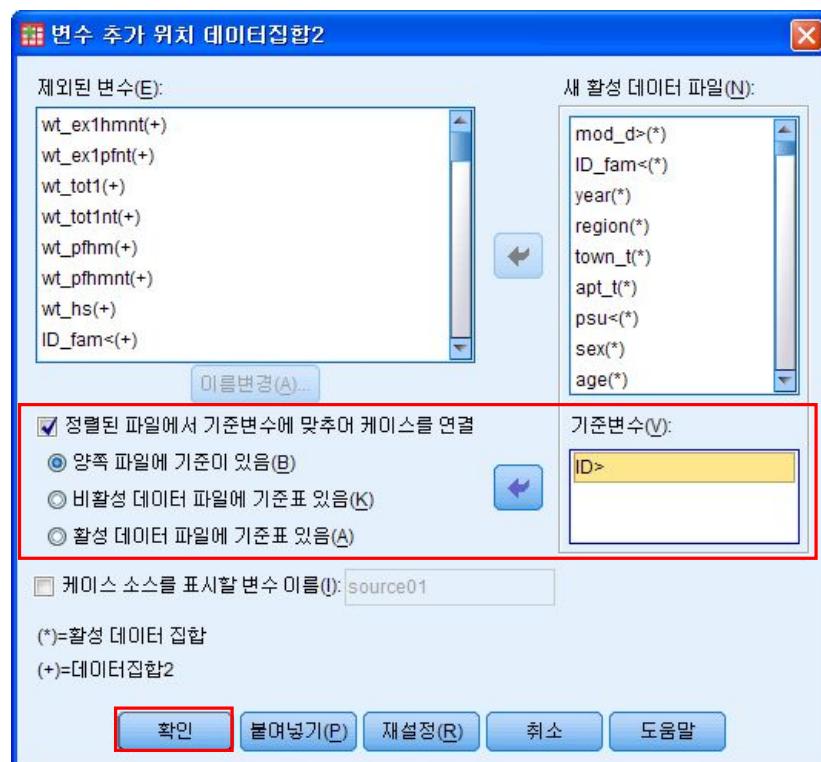


3) 가로결합 후 생성될 DB에 포함변수를 지정한다.

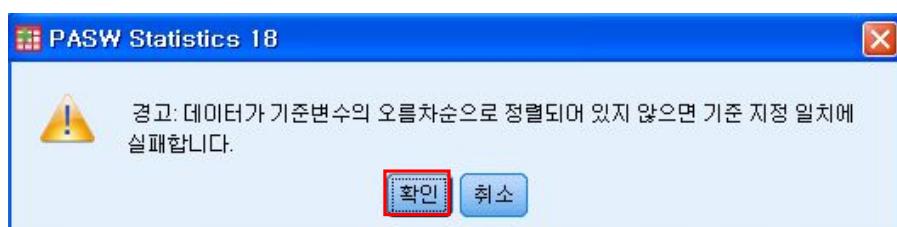
- 생성될 DB에 포함변수는 **새 활성 데이터 파일**에 나타나며, 파일에서 제외할 변수는 **제외된 변수**창으로 이동시킨다.

※ 기본값으로 기본조사DB와 구강조사DB의 모든변수가 **새 활성 데이터 파일** 안에 포함된다. 두 DB에 공통으로 존재하는 변수들은 기준DB가 아닌 불임DB(구강)의 변수들이 제외되며 **제외된 변수**창에 나타난다. 기준DB 변수명 뒤는 (*), 불임DB 변수명 뒤는(+)로 구분되며, 예제에서는 기본조사DB (*), 구강조사DB (+)로 표시

- **정렬된 파일에서 기준변수에 맞추어 케이스를 연결**에서 **양쪽 파일에 기준 있음** 선택 후 개인아이디(ID)를 **제외된 변수**창에서 찾아서 **기준변수**에 넣는다.



- 확인을 클릭하면 아래의 경고창이 나타나며, 아이디(ID)변수로 정렬한 경우는 **확인**을 클릭하면 된다.



□ 자료통합분석

- 세부집단별 분석 또는 낮은 유병률 자료 분석 시, 1개년도의 자료만으로 자료수가 충분하지 않은 경우 다른 조사년도와의 자료통합을 고려해 볼 수 있다.
- 자료통합 분석에서 통합DB를 사용함으로써 분석 자료수가 많아지는 장점이 있지만, 분석변수의 경향 또는 분석변수와 다른 변수간 관련성의 경향이 다른 경우 오히려 그 경향이 희석되는 경우도 있으므로 연도별 경향을 먼저 확인한다. 검토가 완료되면 자료통합 분석을 위해 **통합DB 및 통합 가중치**를 생성한다.

① 자료통합DB생성: 자료의 세로통합

- 다른년도의 자료를 통합하는 방법은 **1. 자료준비(자료열기), 2. 파일 합치기(케이스추가)** 단계로 진행한다.

자료통합DB 생성[세로결합] 예제

- 국민건강영양조사 2010년 기본조사DB(HN10_ALL)와 2011년 기본조사DB (HN11_ALL) 자료를 1개의 통합DB로 생성하는 과정은 다음과 같다.

자료 준비

파일 합치기(케이스추가)

- 통합할 자료인 2010년 기본조사DB(HN10_ALL)와 2011년 기본조사DB (HN11_ALL)를 활성화한다.

HN10_ALL.sav [데이터집합3] - PASW Statistics Data Editor

1:ID	A901963701	ID	ID_fam	year	region	town
1	A901963701	A9019637		2010	1	
2	A901963702	A9019637		2010	1	
3	A901963703	A9019637		2010	1	
4	A901963801	A9019638		2010	1	
5	A901963802	A9019638		2010	1	

HN11_ALL.sav [데이터집합1] - PASW Statistics Data Editor

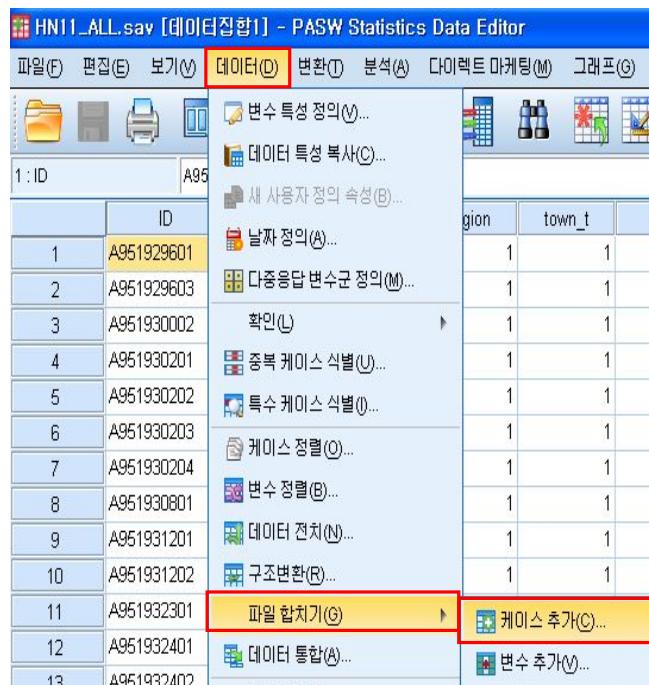
1:ID	A951929601	ID	ID_fam	year	region	town_t
1	A951929601	A9519296		2011	1	1
2	A951929603	A9519296		2011	1	1
3	A951930002	A9519300		2011	1	1
4	A951930201	A9519302		2011	1	1
5	A951930202	A9519302		2011	1	1

자료 준비

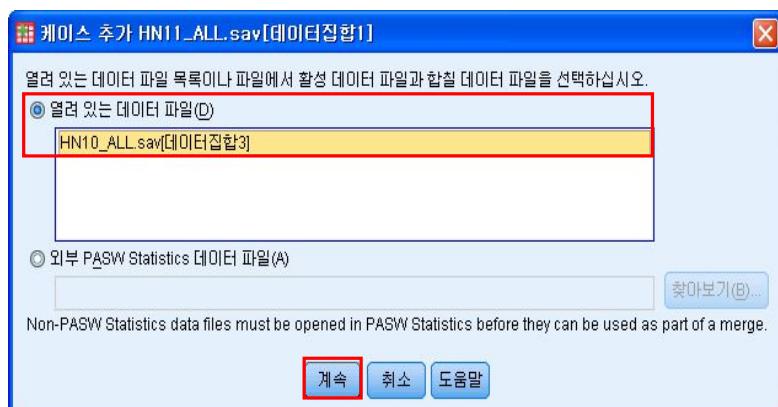
파일 합치기(케이스추가)

- 2011년 기본조사DB(HN11_ALL)을 활성화하고 아래의 과정으로 진행한다.

- 1) 기본조사DB(HN11_ALL) 활성화 상태에서 메뉴 창의 **데이터** ► **파일 합치기**
▶ **케이스 추가** 를 선택한다.



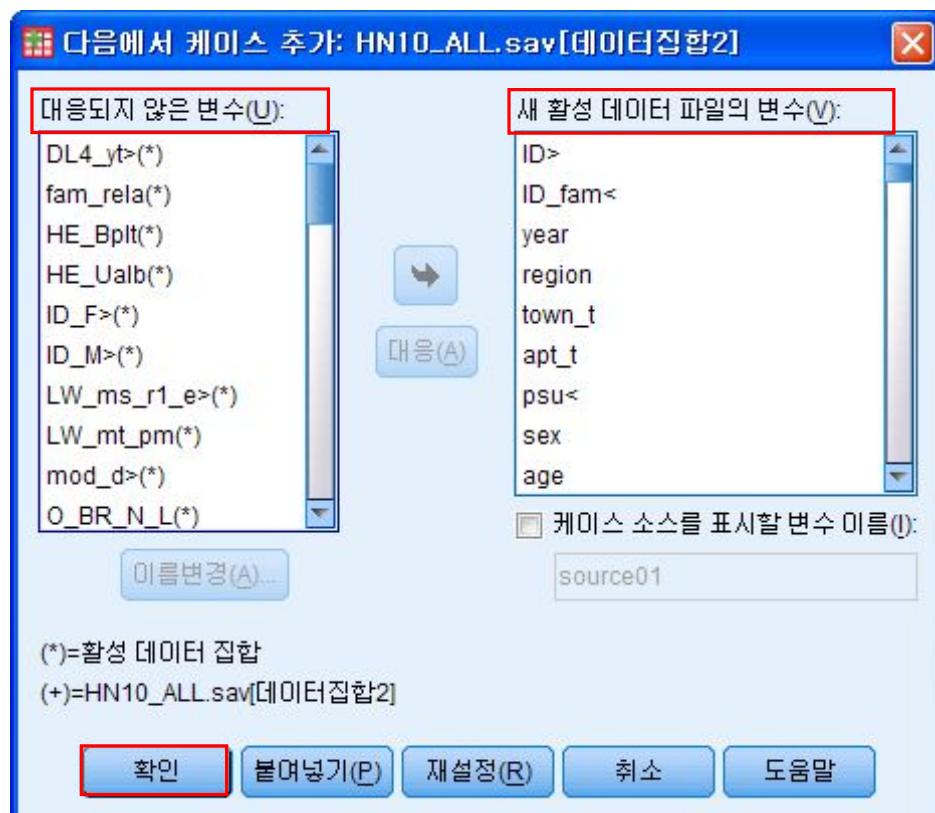
- 2) 변수 추가 창에서 **열려있는 데이터 파일** 을 선택하고 2010년 기본조사DB(HN10_ALL) 선택 후 **계속** 을 클릭한다.



* 통합자료인 2010년 기본조사DB(HN10_ALL)가 열려있는 경우에는 <열려있는 데이터 파일>에서 선택, 열려있지 않은 경우는 <외부 PASW Statistics 데이터 파일>에서 찾아서 지정하기

- 3) 다음에서 케이스 추가 창이 나타나면 확인 을 클릭한다. 생성DB에 포함변수를 변경할 경우는 아래의 방법으로 변수를 선택한 후 확인 을 클릭한다.

* 기본값은 기준DB(활성화DB)와 추가DB에 공통으로 있는 변수들이 새 활성 데이터 파일의 변수 창에 포함되며, 두 DB에 공통으로 없는 변수는 대응되지 않는 변수 창에 나타난다. 공통으로 없는 변수를 추가할 때는 대응되지 않는 변수 창의 변수를 새 활성 데이터 파일의 변수 창으로 이동시킨다.



2 자료통합DB 생성 : 통합가중치 산출

- 자료통합 분석 시 다음과 같이 통합가중치를 산출한 후 분석을 실시한다.

1) 기수내 자료통합

제4기(2007-2009), 제5기(2010-2012)의 기수내 자료를 통합하는 경우, 연도별 표본 조사구수 비례로 통합 가중치를 산출한다. 예를 들어, 제4기(2007-2009) 3개년도의 자료를 통합하는 경우, 조사구수가 2007년 100개, 2008년 200개, 2009년 200개이므로 통합 가중치는 다음 표와 같이 계산한다.

조사년도 통합년도	2007	2008	2009
조사구수	100	200	200
연도별 가중치	w_{07}	w_{08}	w_{09}
통합 가중치			
2007-2008	$w_{07-08} = w_{07} \times 1/3$	$w_{07-08} = w_{08} \times 2/3$	-
2008-2009	-	$w_{08-09} = w_{08} \times 1/2$	$w_{08-09} = w_{09} \times 1/2$
2007-2009	$w_{07-09} = w_{07} \times 1/5$	$w_{07-09} = w_{08} \times 2/5$	$w_{07-09} = w_{09} \times 2/5$

※ 「국민건강영양조사 제4기(2007-2009) 가중치 산출 및 제5기(2010-2012) 표본설계」(질병관리본부, 한국조사연구학회; 2009) 연구용역보고서 참조

2) 기수간 자료통합

제4기(2007-2009)와 제5기(2010-2012)의 기수간 자료를 통합하는 경우, 연도별 조사 구 자료가 그 연도를 대표한다고 간주하여 통합 가중치를 산출한다. 단, 2007년은 연중 반기(7~12월) 동안만 조사된 점을 고려하여 다른 연도가 1의 통합 비율을 갖는다면, 2007년이 1/2의 통합 비율을 갖도록 적용한다 예를 들어, 2007~2011년 자료 통합시 2007, 2008, 2009, 2010, 2011년의 통합 비율은 각각 1/2, 1, 1, 1, 1을 적용하여 통합 가중치를 계산한다.

조사년도 통합년도	제4기			제5기	
	2007	2008	2009	2010	2011
조사구수	100	200	200	192	192
기수간 통합비율	1	2	2	2	2
연도별 가중치	w_{07}	w_{08}	w_{09}	w_{10}	w_{11}
통합 가중치					
2007-2010	$w_{07-10} = w_{07} \times 1/7$	$w_{07-10} = w_{08} \times 2/7$	$w_{07-10} = w_{09} \times 2/7$	$w_{07-10} = w_{10} \times 2/7$	-
2007-2011	$w_{07-11} = w_{07} \times 1/9$	$w_{07-11} = w_{08} \times 2/9$	$w_{07-11} = w_{09} \times 2/9$	$w_{07-11} = w_{10} \times 2/9$	$w_{07-11} = w_{10} \times 2/9$
2008-2010	-	$w_{08-10} = w_{08} \times 1/3$	$w_{08-10} = w_{09} \times 1/3$	$w_{08-10} = w_{10} \times 1/3$	-
2008-2011	-	$w_{08-11} = w_{08} \times 1/4$	$w_{08-11} = w_{09} \times 1/4$	$w_{08-11} = w_{10} \times 1/4$	$w_{08-11} = w_{10} \times 1/4$

※조사구수 비례 및 연도별 비례 방법으로 산출된 통합 가중치를 적용한 추정치 비교 결과, 표준오차 차이가 거의 없으므로 기수간 자료 통합 분석시 연도별 비례 통합 방법 사용(「국민건강영양조사 및 청소년건강행태온라인조사 표본설계」(질병관리본부, 한국조사연구학회; 2012) 연구용역보고서 참조)

3) 안검사 또는 이비인후검사 자료 통합

안검사 또는 이비인후검사의 연도별 자료를 통합하는 경우, 기수내 또는 기수간 자료 통합 방법을 적용하되, 2008년 7월 도입된 점을 고려하여 2008년의 통합 비율은 1/2을, 다른 연도의 통합 비율은 1을 적용합니다.

통합자료 조사기간	제4기		제5기	
	2008.7-12	2009	2010	2011
기수내 통합비율 (조사구수)	108	200	192	192
기수간 통합비율	1	2	2	2
2008-2009	$w_{08-09} = w_{08} \times \frac{108}{308}$	$w_{08-09} = w_{09} \times \frac{200}{308}$	-	-
2008-2011	$w_{08-11} = w_{08} \times 1/7$	$w_{08-11} = w_{09} \times 2/7$	$w_{08-11} = w_{10} \times 2/7$	$w_{08-11} = w_{11} \times 2/7$
2009-2011	-	$w_{09-11} = w_{09} \times 2/6$	$w_{09-11} = w_{10} \times 2/6$	$w_{09-11} = w_{11} \times 2/6$

4) 골밀도검사 또는 체지방검사 자료 통합

골밀도검사 또는 체지방검사의 연도별 자료를 통합하는 경우, 기수내 또는 기수간 자료 통합방법을 적용하되, 2008년 7월 도입된 점을 고려하여 2008년의 통합 비율은 1/2을, 다른 연도의 통합 비율은 1을 적용한다. 단, 2011년 5월에 조사가 완료된 점을 고려하여 제5기 2차년도(2011) 자료가 포함된 기수간 자료 통합의 경우 예외적으로 조사구수 비례로 통합 가중치를 계산한다.

통합자료 조사기간	제4기		제5기	
	2008.7-12	2009	2010	2011.1-5
기수내 통합비율 (조사구수)	108	200 (체지방 199)	192	80
기수간 통합비율	1	2	2	-
2008-2009	$w_{08-09} = w_{08} \times 108/308$ (체지방: 108/307)	$w_{08-09} = w_{09} \times 200/308$ (체지방: 199/307)	-	-
2008-2010	$w_{08-10} = w_{08} \times 1/5$	$w_{08-09} = w_{09} \times 2/5$	$w_{08-10} = w_{10} \times 2/5$	-
2008-2011	$w_{08-11} = w_{08} \times 108/580$ (체지방: 108/579)	$w_{08-11} = w_{09} \times 200/580$ (체지방: 199/579)	$w_{08-11} = w_{10} \times 192/580$ (체지방: 192/579)	$w_{08-11} = w_{11} \times 80/580$ (체지방: 80/579)
2009-2010	-	$w_{09-10} = w_{09} \times 1/2$	$w_{09-10} = w_{10} \times 1/2$	-
2009-2011	-	$w_{09-11} = w_{09} \times 200/472$	$w_{09-11} = w_{10} \times 192/472$	$w_{09-11} = w_{11} \times 80/472$
2010-2011	-	-	$w_{10-11} = w_{10} \times 192/272$	$w_{10-11} = w_{11} \times 80/272$

통합가중치 산출 예제

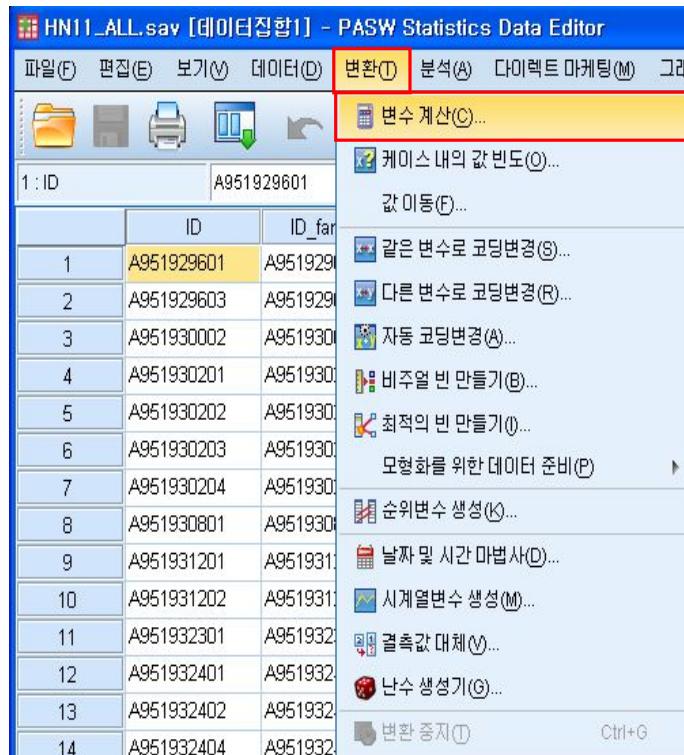
- 국민건강영양조사 2007년~2011년 기본조사DB(HN07_ALL~HN11_ALL)의 통합DB에서 건강설문-검진 통합가중치 산출하고자 한다.

※ 2007~2011년 5개년도 자료의 통합DB 생성과정은 「자료통합DB생성: 자료의 세로통합」 참조 단, 이 경우에는 케이스추가 과정을 4번 반복하여 진행한다. 즉, 2011년 DB 활성화+케이스추가(2010년 DB), 생성DB(2011년 DB+2010년 DB)+케이스추가(2009년 DB), …, 생성DB(2011년 DB + 2010년 DB + 2009년 DB + 2008년 DB) +케이스추가(2007년 DB)

- 1) 건강설문-검진 통합가중치 계산식을 확인한 후, 통합가중치 산출을 위해 메뉴창에서 **변환** ► **변수계산** 을 선택한다.

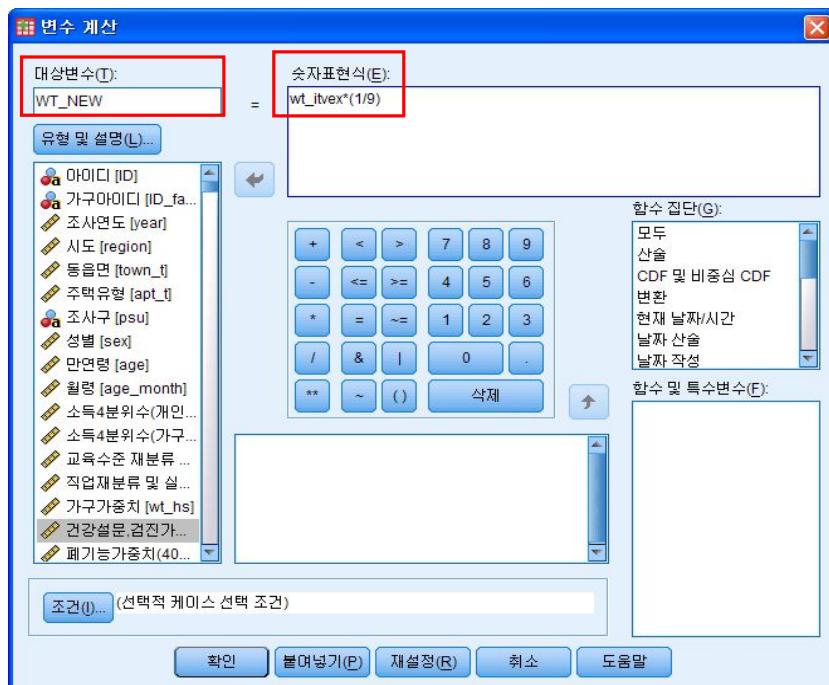
- 건강설문-검진 통합가중치 계산식

년도	2007	2008	2009	2010	2011
검진·건강설문 가중치	wt_itvex	wt_itvex	wt_itvex	wt_itvex	wt_itvex
통합 가중치	통합비율	1/2	1	1	1
	통합비율×2	1	2	2	2
	가중치× 통합비율 통합비율합	wt_itvex*1/9	wt_itvex*2/9	wt_itvex*2/9	wt_itvex*2/9

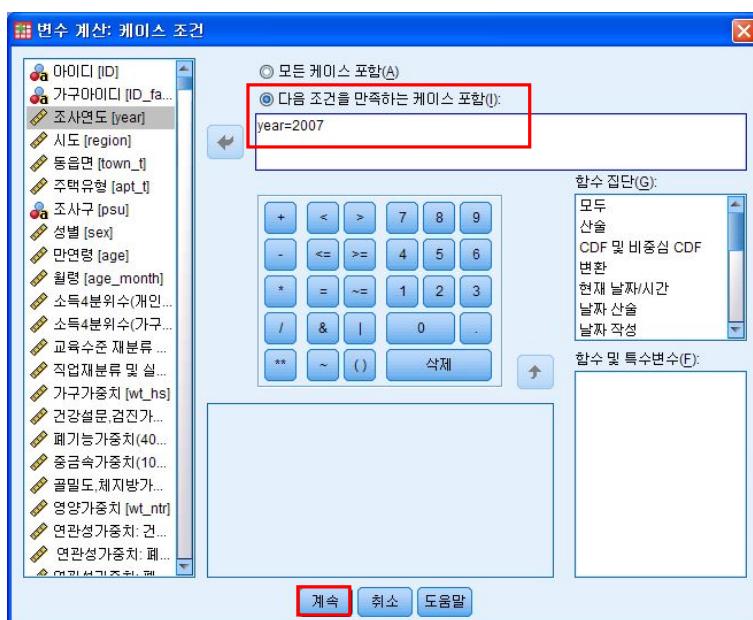


2) 위와 같은 통합비율로 2007년 건강설문-검진 통합가중치를 산출한다.

- ① **대상변수** : 통합가중치 변수명(WT_NEW)을 입력
- ② **숫자표현식** : $wt_itvex^*(1/9)$ 을 입력
- ③ **대상변수**, **숫자표현식** 을 지정한 후, **조건** 을 클릭한다.



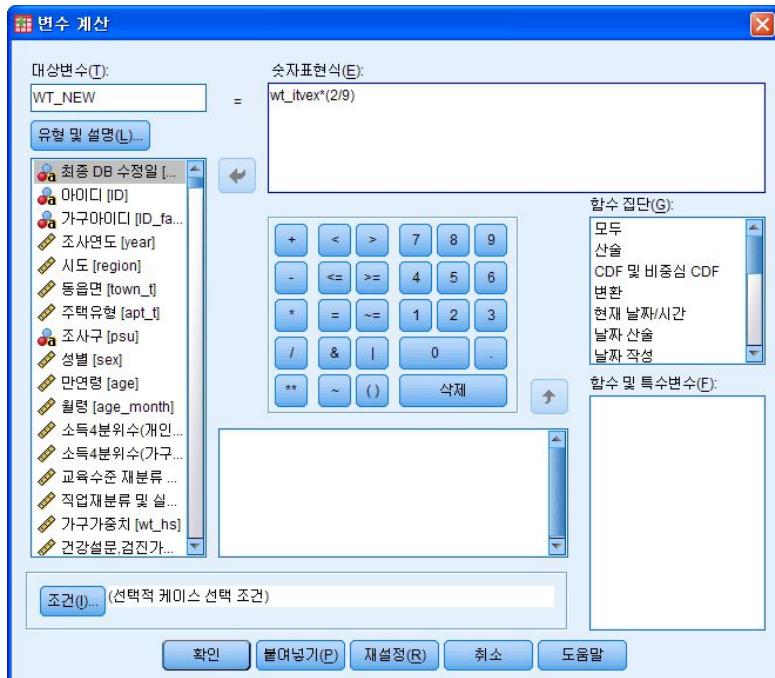
- ④ 다음 조건을 만족하는 케이스 포함에 year=2007 입력한 후, **계속** 을 클릭한다.



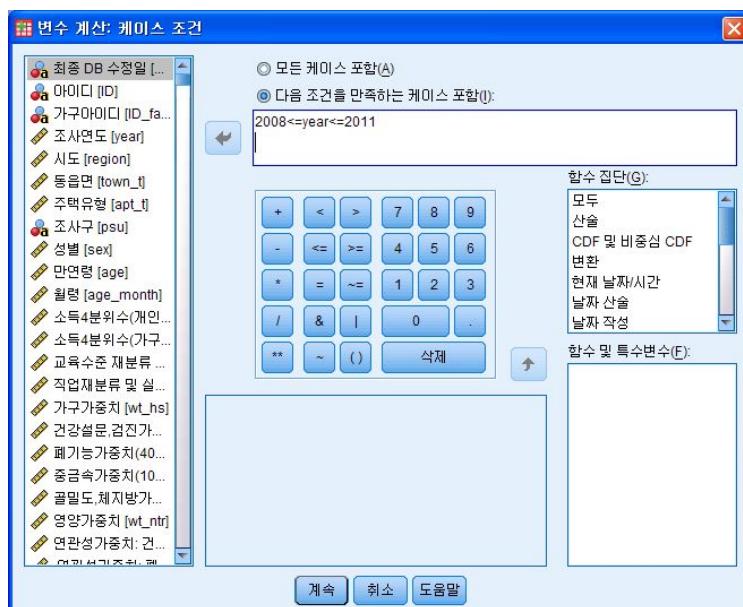
- ⑤ 지정한 값을 확인한 후, **확인** 을 클릭한다.

3) 2008~2011년 통합비율을 확인한 후 건강설문-검진 통합가중치를 산출한다.

- ① 메뉴창에서 **변환** ► **변수계산** 을 클릭한다.
- ② **대상변수** : 통합가중치 변수명(WT_NEW) 입력
- ③ **숫자표현식** : $wt_itvex^*(2/9)$ 을 입력
- ④ **대상변수**, **숫자표현식** 을 지정한 후, **조건** 을 클릭한다.



- ⑤ 다음 조건을 만족하는 케이스 포함에 2008<=year<=2011 입력한 후, **계속** 을 클릭한다.



- ⑥ 지정한 값을 확인한 후, **확인** 을 클릭한다.

□ 식품섭취조사 개인단위 분석DB 생성

- 식품섭취조사DB는 개인이 조사 1일전 하루 동안 섭취한 음식, 식품, 그에 따른 영양소 섭취량에 대한 자료이며, 아래와 같이 개인당 여러 개의 행으로 구성되어 있다.

<식품섭취조사DB(HNyr_24RC) 자료 예시>

ID	끼나구분	음식명	총부피	섭취부피	식품재료	섭취량	에너지	단백질	지방	탄수화물
A101020101	1	훈합잡곡밥	943	300	훈합잡곡	52.15	193	3.9	0.6	41.3
A101020101	1	훈합잡곡밥	943	300	쌀	78.23	291	5	0.4	63.8
A101020101	1	된장국, 냉이	1000	300	냉이	10.98	3	0.5	0.1	0.4
...										
A101020101	1	멸치볶음, 풋고추	300	50	깻소금	0.5	3	0.1	0.3	0
A101020101	1	멸치볶음, 풋고추	300	50	식용유	2	18	0	0	0
A101020101	1	멸치볶음, 풋고추	300	50	풋고추	21.71	4	0.3	0.1	0.8

- 식품섭취조사DB(HNyr_24RC)에서 개인별 섭취량, 에너지 및 영양소를 합산하여, 개인이 1일간 섭취한 섭취량은 기본조사DB(HNyr_ALL)에 포함되어 있다. 따라서 개인별 1일 섭취량, 에너지 및 영양소자료를 활용시에는 기본조사DB를 사용하며, 이외에 식품군별 분석, 끼니별 분석, 특정 식품 분석 등을 위해서는 식품섭취조사 DB를 사용하여야 하며, 이를 위해 개인당 여러 개의 자료를 1개의 자료로 변환하는 과정이 필요하다.

식품섭취조사 개인단위 분석DB생성 예제

- 2011년 식품섭취조사DB(HN11_24RC)에서 개인별 채소류 섭취량을 산출과정은 다음과 같다.

- 식품섭취조사DB에서 식품군분류(N_KINDG1)가 채소류(코드:06)인 섭취량(NF_INTK)을 합하여 개인별 채소류 섭취량 산출
 - 식품별 섭취량(NF_INTK)에서 채소류 섭취량 변수 생성
 - 채소류 식품섭취량을 합하여 개인별 총 채소류 섭취량 변수 생성

변수명	변수설명	내용
NF_INTK	식품섭취량	식품별 섭취량(g)
N_KINDG1	식품군분류	01. 곡류 02. 감자류 03. 당류 04. 두류 05. 견과류 06. 채소류 ...

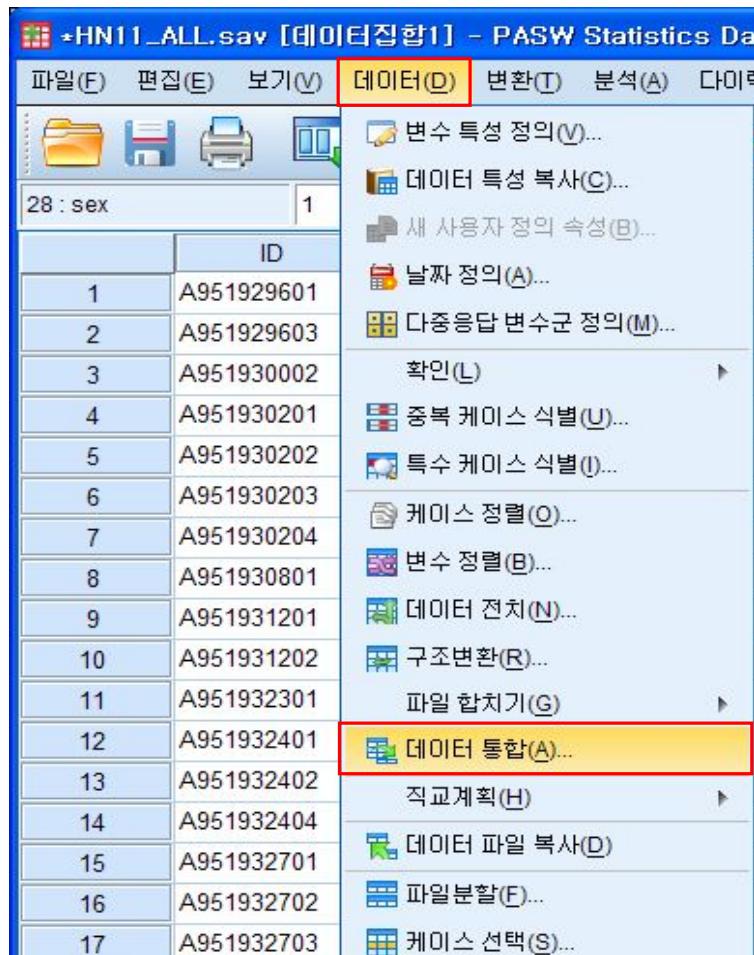
- 1) 식품별 섭취량(NF_INTK)에서 채소류 섭취량변수(Vegetable) 생성은 식품섭취량조사DB를 활성화한 상태에서 메뉴 창에서 [변환] ▶ [변수계산]에서 다음과 같이 지정한다(지정방법은 p90-91 참조).

※채소류 식품섭취량변수(Vegetable)에 식품군분류가 채소인 식품섭취량(NF_INTK)값만을 갖도록 함.



2) 채소류 식품섭취량변수(vegetable)을 개인별로 합한 개인별 총 채소류섭취량 변수(vegetable_sum) 생성은 다음과 같이 진행한다.

- ① 메뉴창에서 **데이터** ► **데이터 통합** 을 선택한다.

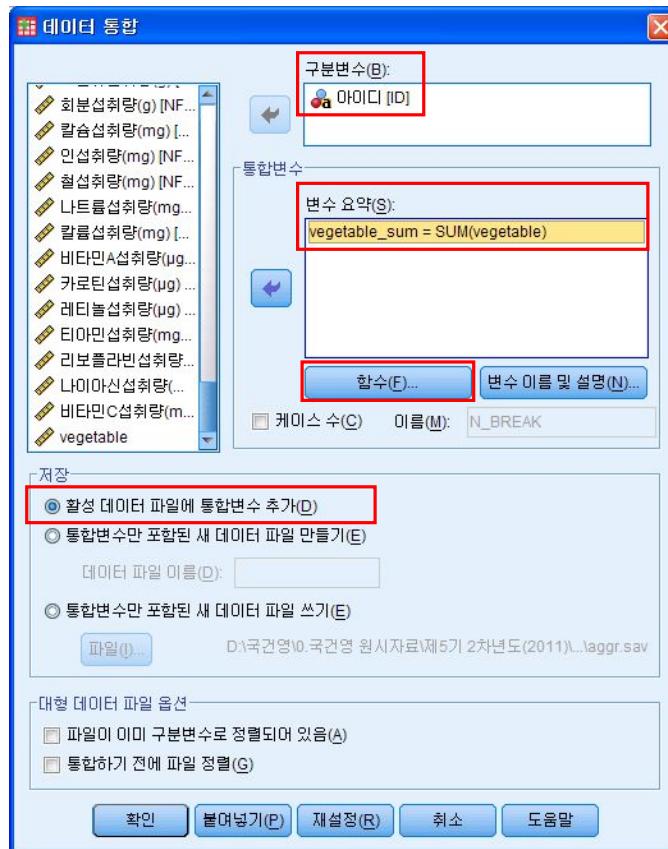


② 데이터 통합 창이 나타나면 다음과 같이 지정한다.

- **구분변수** : ID선택 (개인아이디(ID)별 자료합산을 위해)

- **변수요약** : 개인별 채소류섭취량변수(vegetable) 선택

※변수선택은 좌측변수창에서 해당변수를 찾아서 그 창으로 이동시킨다.



- **함수** 를 클릭하여 <요약통계량>에서 합계를 선택한다.

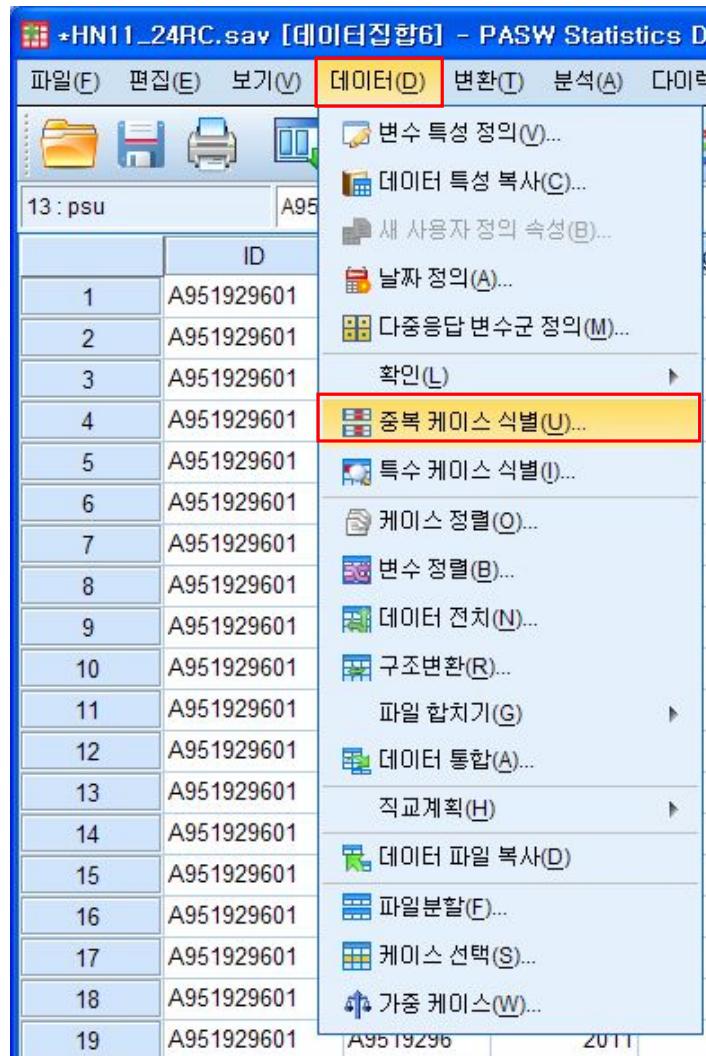
※기본값으로 설정된 평균함수를 합계함수로 변경한다.



- **저장** 옵션에서 <활성 데이터 파일에 통합변수 추가>를 선택한다.

- 3) 데이터통합을 통해 개인별 총 채소류섭취량 변수(vegetable_sum)를 생성하면, 식품 섭취조사DB(HN11_24RC)에 vegetable_sum 변수가 추가된다. 따라서 개인단위 분석이 가능하도록 생성DB를 개인별 자료DB로 만드는 과정이 필요하며, 그 과정은 다음과 같이 <중복 케이스 식별>과 <케이스 선택>으로 진행한다.

① **데이터** ► **중복 케이스 식별** 을 선택한다.



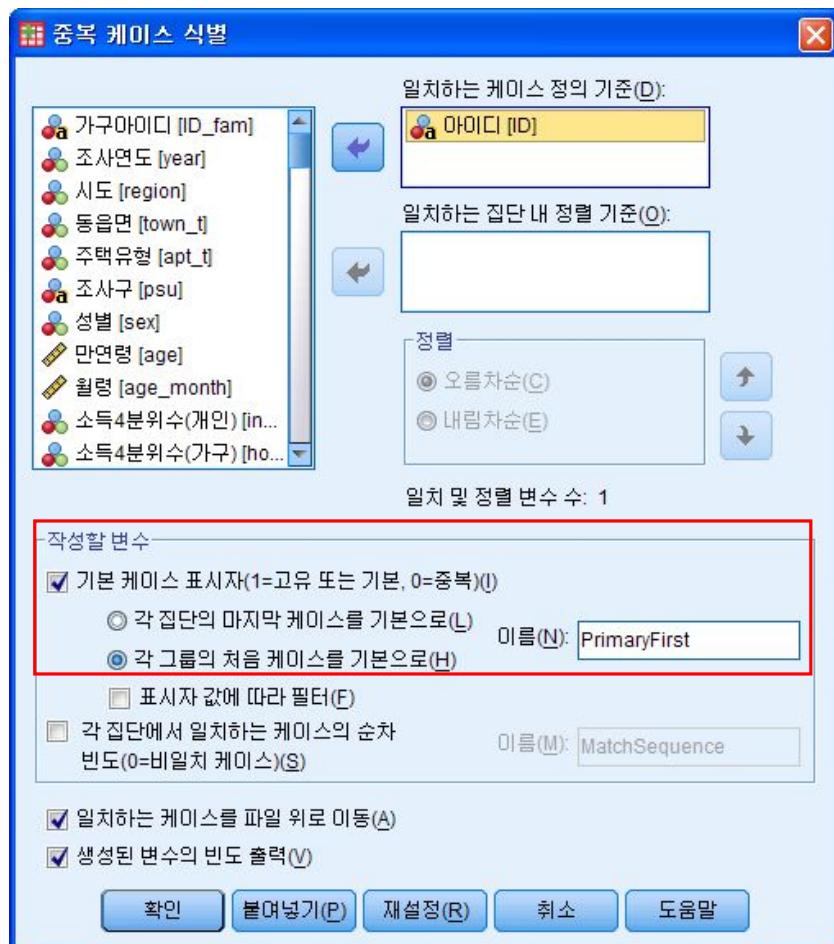
② 중복 케이스 식별 창이 나타나면 다음과 같이 지정하고 **확인** 을 클릭한다.

- **일치하는 케이스 정의 기준** : ID선택

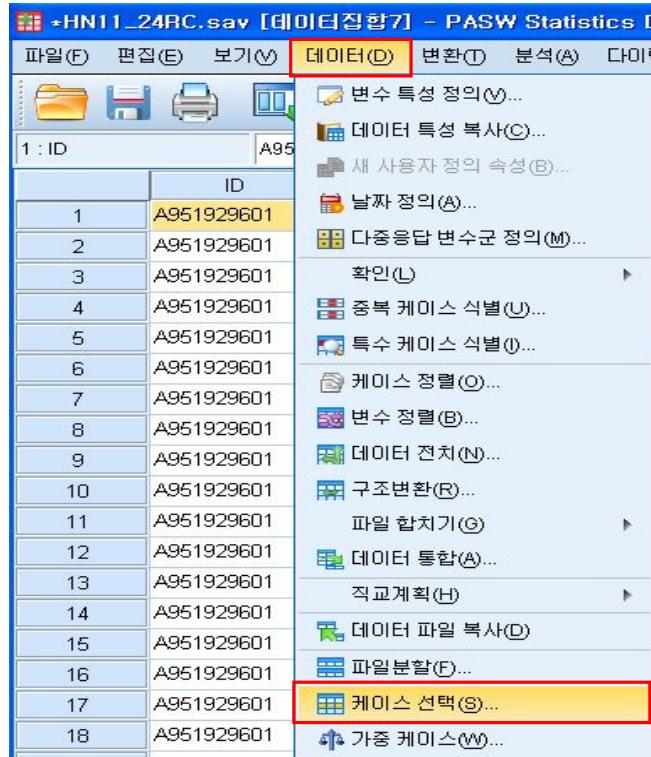
※변수선택은 좌측변수창에서 해당변수를 찾아서 그 창으로 이동시킨다.

- **작성할 변수** 에서 “기본 케이스 표시자” 클릭, “각 그룹의 처음 케이스를 기본으로” 선택

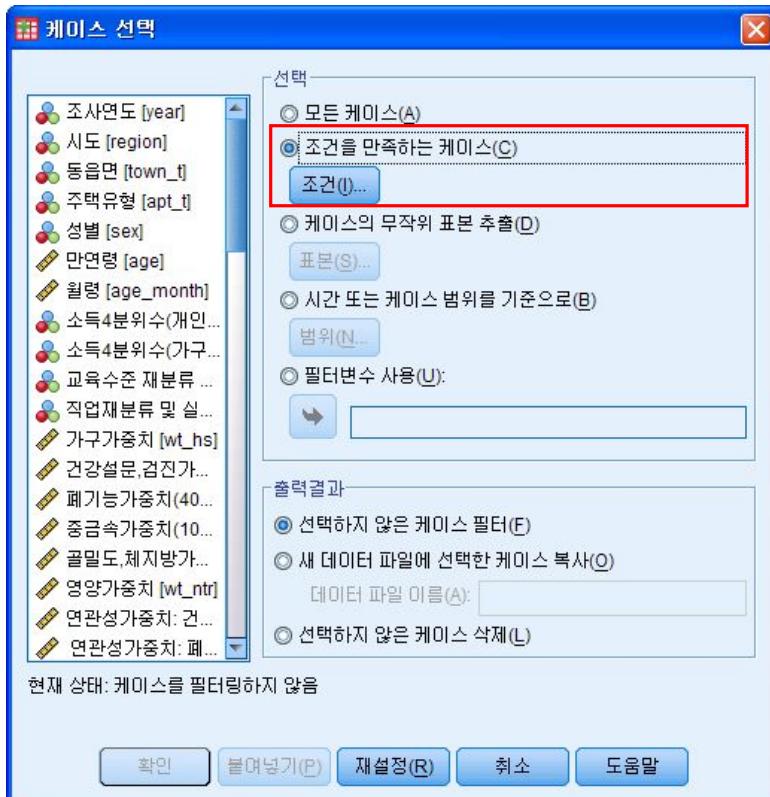
※본 예제에서는 **작성할 변수** 에 “각 그룹의 처음 케이스를 기본으로”로 지정하며, “집단의 마지막 케이스를 기본으로”를 지정한 경우에는 PrimaryLast라는 변수가 생성되며 이 후, PrimaryLast변수를 사용하면 된다.



③ 메뉴창에서 **데이터** ► **케이스 선택** 을 선택한다.



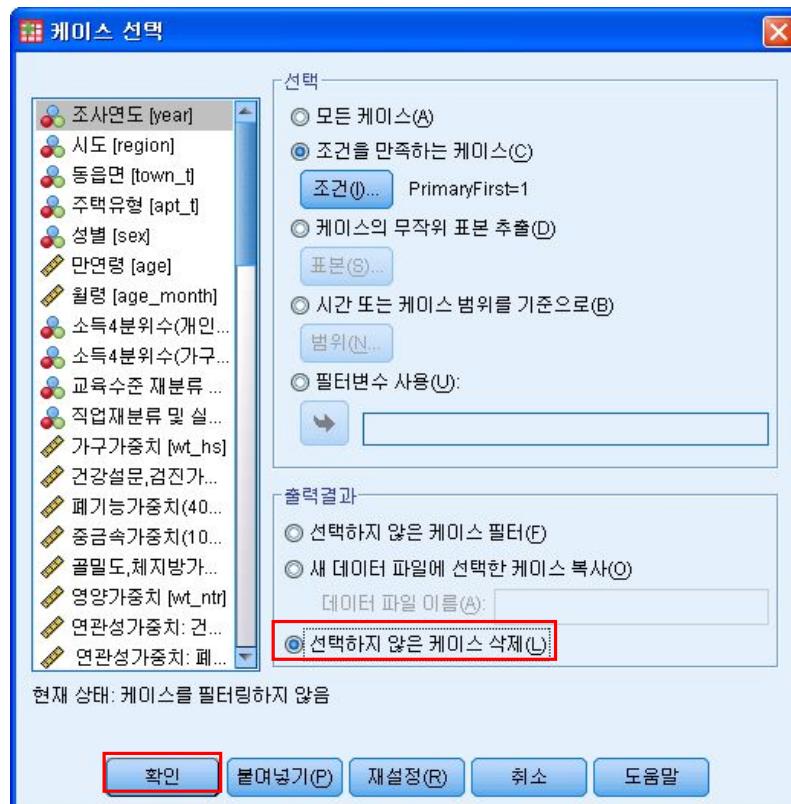
④ **조건을 만족하는 케이스** 선택하고, **조건** 을 클릭한다.



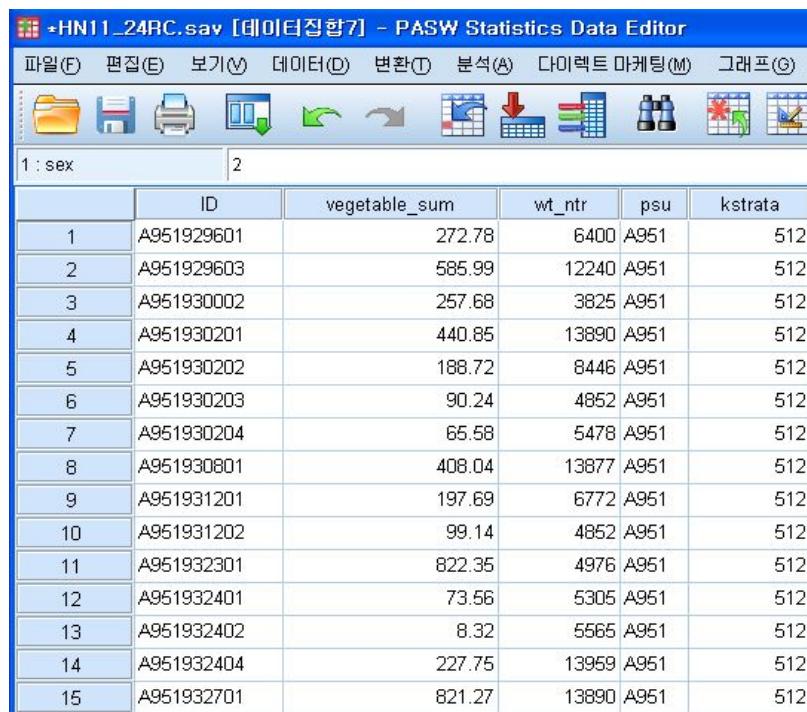
⑤ 케이스 선택: 조건 창이 나타나면, PrimaryFirst=1 을 입력 후 **계속** 을 클릭한다.



⑥ **출력결과**에서 “선택하지 않은 케이스 삭제”를 선택하고 **확인** 을 클릭한다.



⑦ 위의 과정을 완료하면, 다음과 같이 개인별 DB가 생성된다.



The screenshot shows the PASW Statistics Data Editor interface. The title bar reads '+HN11_24RC.sav [데이터집합7] - PASW Statistics Data Editor'. The menu bar includes '파일(F)', '편집(E)', '보기(V)', '데이터(D)', '변환(I)', '분석(A)', '다이렉트 마케팅(M)', and '그래프(G)'. Below the menu is a toolbar with various icons for file operations like opening, saving, printing, and filtering. The main window displays a table with 15 rows and 6 columns. The first column is labeled '1 : sex' and contains row numbers from 1 to 15. The second column is labeled 'ID' and contains values starting with A951929601. The third column is labeled 'vegetable_sum' and contains numerical values like 272.78, 585.99, etc. The fourth column is labeled 'wt_ntr' and contains values like 6400, 12240, etc. The fifth column is labeled 'psu' and contains values like A951, A951, etc. The sixth column is labeled 'kstrata' and contains values like 512, 512, etc. The table has a light blue header row and white data rows.

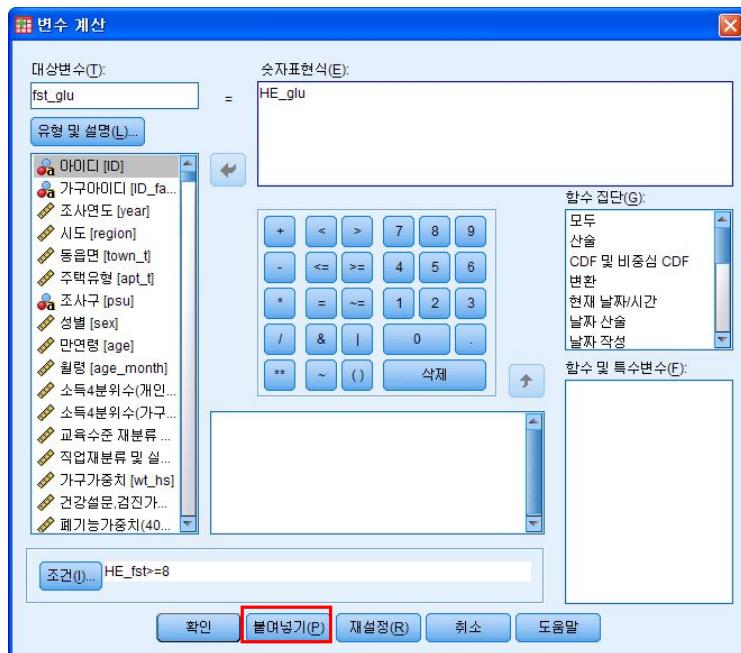
1 : sex	ID	vegetable_sum	wt_ntr	psu	kstrata
1	A951929601	272.78	6400	A951	512
2	A951929603	585.99	12240	A951	512
3	A951930002	257.68	3825	A951	512
4	A951930201	440.85	13890	A951	512
5	A951930202	188.72	8446	A951	512
6	A951930203	90.24	4852	A951	512
7	A951930204	65.58	5478	A951	512
8	A951930801	408.04	13877	A951	512
9	A951931201	197.69	6772	A951	512
10	A951931202	99.14	4852	A951	512
11	A951932301	822.35	4976	A951	512
12	A951932401	73.56	5305	A951	512
13	A951932402	8.32	5565	A951	512
14	A951932404	227.75	13959	A951	512
15	A951932701	821.27	13890	A951	512

□ 명령문 사용방법

- SPSS 실행방법은 GUI(Graphic User Interface)방식과 명령문(Syntax)방식으로 구성되어 있다. GUI방식은 앞에서 제시한 예제와 같이, 메뉴와 대화상자를 이용하여 데이터를 처리하고 분석하는 방법이고, 명령문 방식은 프로그램을 작성하여 데이터 입출력 및 분석 등 모든 절차를 수행하는 방법이다.
- 명령문 방식은 반복적인 작업 수행시 유용하게 사용할 수 있으나 명령문에 익숙하지 않은 경우에는 사용이 쉽지 않다. 명령문을 익숙치 않은 경우에는 먼저 GUI 방식으로 실행하고, 그 실행과정을 명령문으로 저장한 후 이후 분석시에 명령문에서 변경해야 하는 부분을 수정하고 명령문을 실행하면 유용하게 명령문을 이용할 수 있다.
- 명령문 사용 방법
 - 명령문을 새로 작성하거나 저장된 명령문을 불러와서 사용할 수 있다.
 - 메뉴창에서 **파일** ► **새파일** ► **명령문** 선택하면 다음과 같이 명령문 창이 활성화된다. 명령문 창에서 직접 명령문을 작성한 후 명령문을 실행한다.
 - 저장된 명령문이 있는 경우는 **파일** ► **열기** ► **명령문** 를 선택하고, 저장된 명령문 파일을 열어서 명령문을 실행한다.
 - GUI 방식으로 실행하고, 그 실행과정에 대한 명령문을 저장하기 위해 대화상자에서 <붙여넣기> 클릭하여, 실행과정 명령문이 나타나면 명령문을 실행한다.

명령문 사용방법 예제

- 1) GUI 방식의 실행과정을 명령문으로 저장하는 방법은 대화상자창에서 클릭 방식으로 입력완료 후, **붙여넣기** 를 클릭한다.



※예제에 대한 설명은 p90-91 참조.

- 2) **붙여넣기** 실행 결과, 위의 예제의 실행과정에 대한 명령문코드가 나타난다.

```

1 DATASET ACTIVATE 데이터집합1.
2 IF (HE_fst>=8) fst_glu=HE_glu.
3 EXECUTE.
4
5

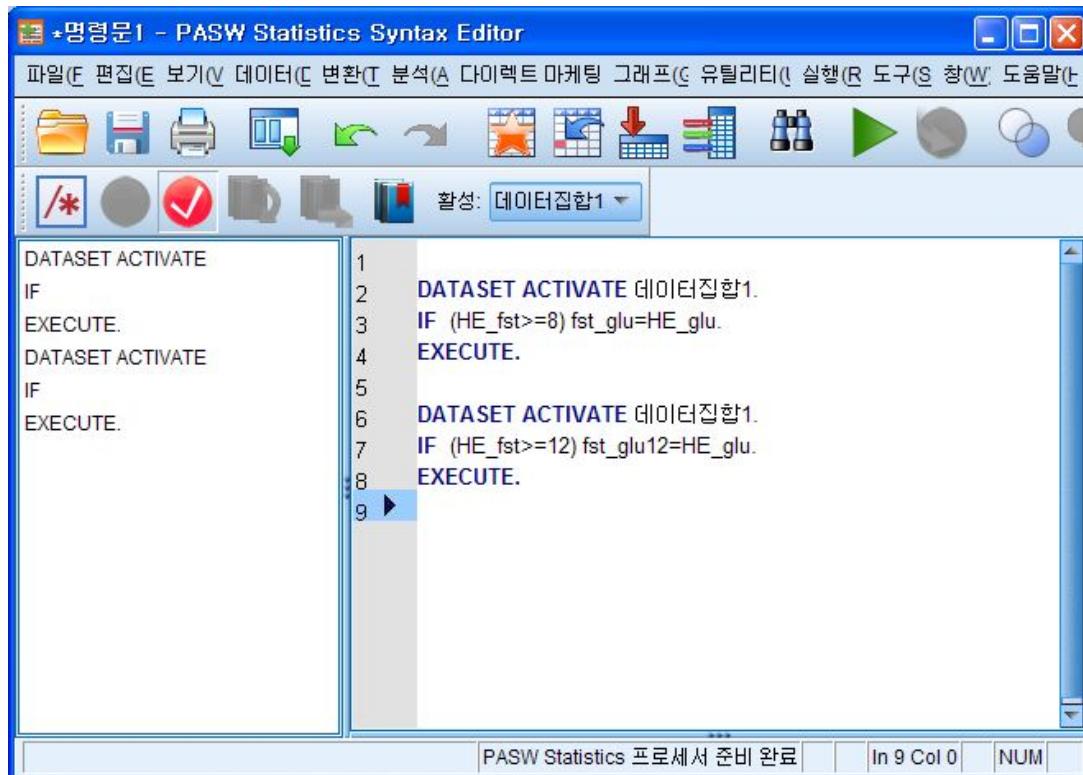
```

3) 명령문에서 명령문코드가 수정이 필요한 경우는 명령문 창에서 수정한다.

- 예를 들어, 공복시간 조건을 8시간에서 12시간으로 변경하고, 새로운 변수 (fst_glu12)로 저장하는 경우에는 다음과 같이 명령문코드를 수정한다.

기존명령문	수정된 명령문
<pre>DATASET ACTIVATE 데이터집합1. IF (HE_fst>=8) fst_glu=HE_glu. EXECUTE.</pre>	<pre>DATASET ACTIVATE 데이터집합1. IF (HE_fst>=12) fst_glu12=HE_glu. EXECUTE.</pre>

4) 수정된 명령문을 실행하면(▶ 클릭)하면, 변수계산 대화상자를 이용하지 않고 fst_glu12가 생성된다.



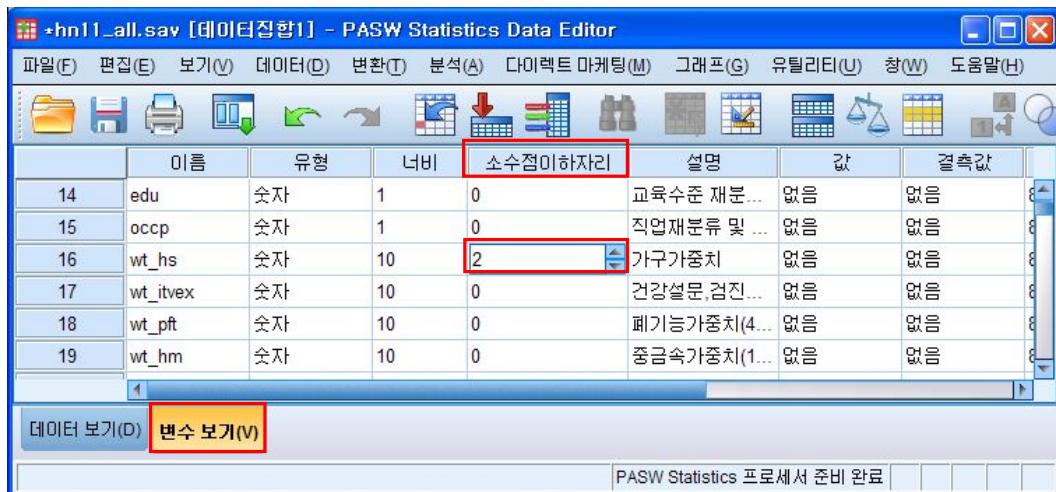
5) 자주 사용하는 명령문은 저장하여 사용하면 편리하다. 사용시에는 메뉴창에서

파일(F) ▶ 열기(O) ▶ 명령문(S) 을 선택하고, 명령문파일을 불러들여 실행한다.

□ 변수보기 및 결측값 지정

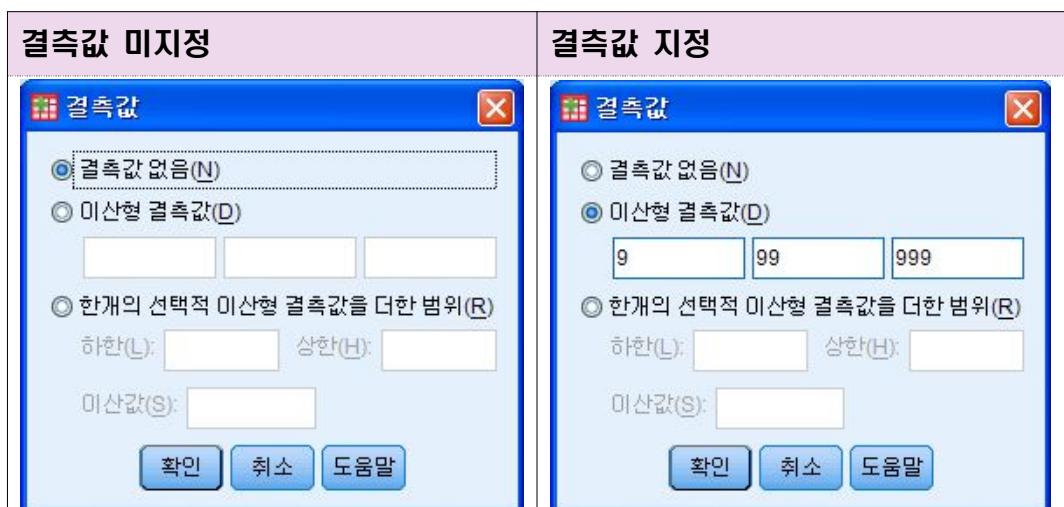
○ 소수점 이하 값 보기

- 국민건강영양조사 원시자료 SPSS DB의 변수값은 정수 형태로(소수점 없이) 보여진다. 소수점 이하 값이 있는 변수의 경우에는 다음과 같이 지정하면, 소수점 이하 값을 볼 수 있다. SPSS 창 하단의 ‘변수보기’ 탭 선택 후, 소수점 이하 값이 필요한 변수의 ‘소수점’ 셀에서 소수점 이하 자리수를 지정한다.



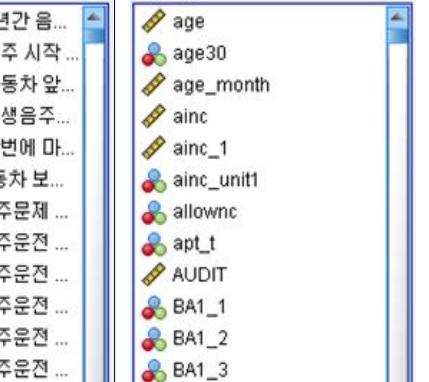
○ 결측값 지정

- 국민건강영양조사 원시자료에는 비해당(8/88/888/...), 무응답(9/99/999/...) 값이 존재한다. 분석 시 이러한 값들이 분석 결과값에 반영되지 않도록 해당값을 결측값으로 지정하여야 한다. ‘변수보기’ 탭에서 결측값을 클릭하면, 아래와 같이 결측값 창이 나타난다. 여기서 이산형 결측값에 해당 결측값을 입력하고, 확인을 클릭한다.



□ 변수 창 정렬

- SPSS 변수를 쉽게 찾기 위한 방법으로 변수 창 정렬을 이용 할 수 있다.
- 변수창 정렬은 모든 프로시저의 대화상자에서 가능하며, 마우스를 우클릭하여 다음과 같은 조건을 선택할 수 있다.

변수설명 표시 파일 내 순서에 따라 정렬	변수이름 표시 파일내 순서에 따라 정렬	변수설명 표시 알파벳 순으로 정렬	변수이름 표시 알파벳 순으로 정렬
<input type="radio"/> 변수이름 표시(V) <input checked="" type="radio"/> 변수설명 표시(L) <input type="radio"/> 알파벳 순으로 정렬(A) <input checked="" type="radio"/> 파일내 순서에 따라 정렬(F) <input type="radio"/> 측정 수준별 정렬(M) 변수 정보(V)...	<input checked="" type="radio"/> 변수이름 표시(V) <input type="radio"/> 변수설명 표시(L) <input type="radio"/> 알파벳 순으로 정렬(A) <input checked="" type="radio"/> 파일내 순서에 따라 정렬(F) <input type="radio"/> 측정 수준별 정렬(M)	<input type="radio"/> 변수이름 표시(V) <input checked="" type="radio"/> 변수설명 표시(L) <input checked="" type="radio"/> 알파벳 순으로 정렬(A) <input type="radio"/> 파일내 순서에 따라 정렬(F) <input type="radio"/> 측정 수준별 정렬(M)	<input checked="" type="radio"/> 변수이름 표시(V) <input type="radio"/> 변수설명 표시(L) <input checked="" type="radio"/> 알파벳 순으로 정렬(A) <input type="radio"/> 파일내 순서에 따라 정렬(F) <input type="radio"/> 측정 수준별 정렬(M)
변수(V): 	변수(V): 	변수(V): 	변수(V): 

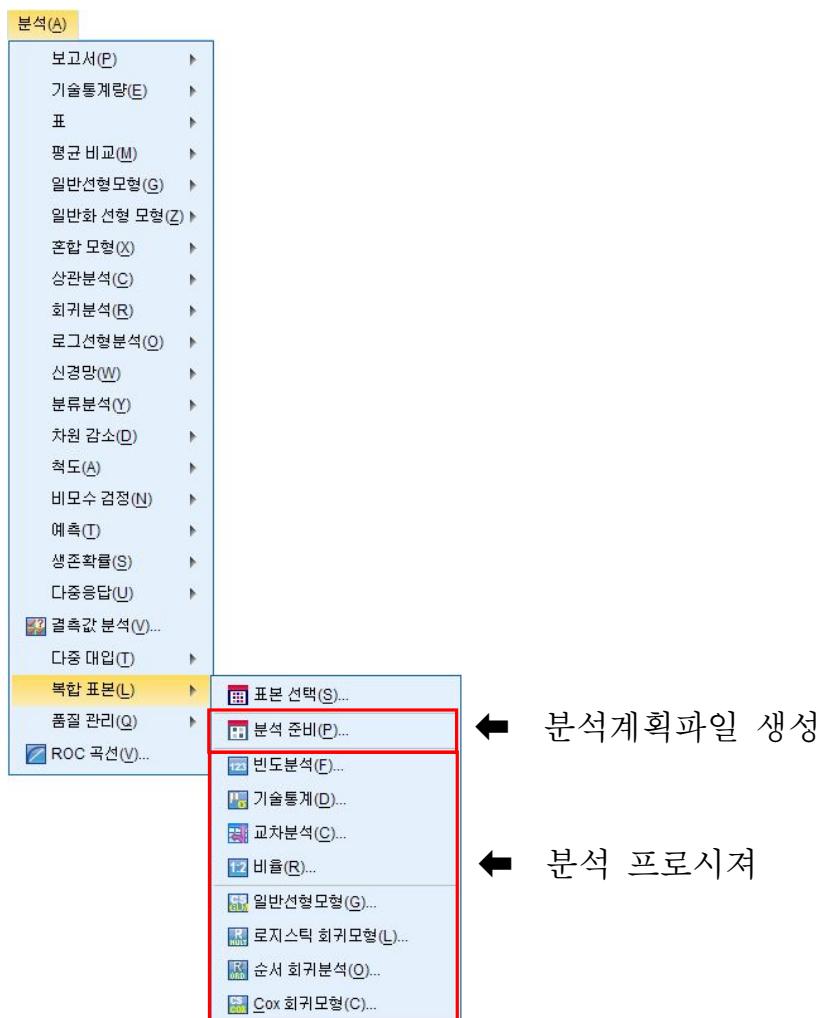
04

복합표본 자료분석

- 개요
- 분석계획 파일생성
- 자료분석 시 유의사항
- 복합표본 빈도분석
- 복합표본 기술통계
- 복합표본 교차분석
- 복합표본 일반선형모형
- 복합표본 로지스틱 회귀모형

□ 개요

- 국민건강영양조사는 복합표본설계되어 단순임의추출자료의 분석과는 다르게 복합표본설계 요소(총, 군집, 가중치)를 반영하여 자료분석을 하도록 권고하고 있다.
- SPSS에서의 복합표본설계 자료분석은 **분석** ▶ **복합표본** 을 사용하며, 복합표본설계 자료분석 전에 **복합표본** ▶ **분석준비** 를 통해 복합표본설계 요소를 반영한 분석계획파일을 생성한다.
 - ※ 분석계획파일은 복합표본자료를 SPSS로 분석하기 위해 표본설계요소(총, 군집, 가중치)를 지정하여 생성한 파일을 말하며, 모든 복합표본 분석 프로시저 첫단계에서 분석계획파일을 생성후 분석과정 진행



- 복합표본설계 자료분석은 **복합표본** 내의 분석 프로시저와 생성된 복합표본 설계 파일을 사용하여 진행한다.

□ 분석 계획파일 생성

- 분석DB에 복합표본설계 요소(층, 군집, 가중치)를 지정하여 분석계획파일을 생성한다.
- 분석계획파일 생성은 복합표본의 분석준비 마법사를 이용하여 표본 구조, 추정 방법, 계층, 군집, 표본 가중값 변수를 지정한다.

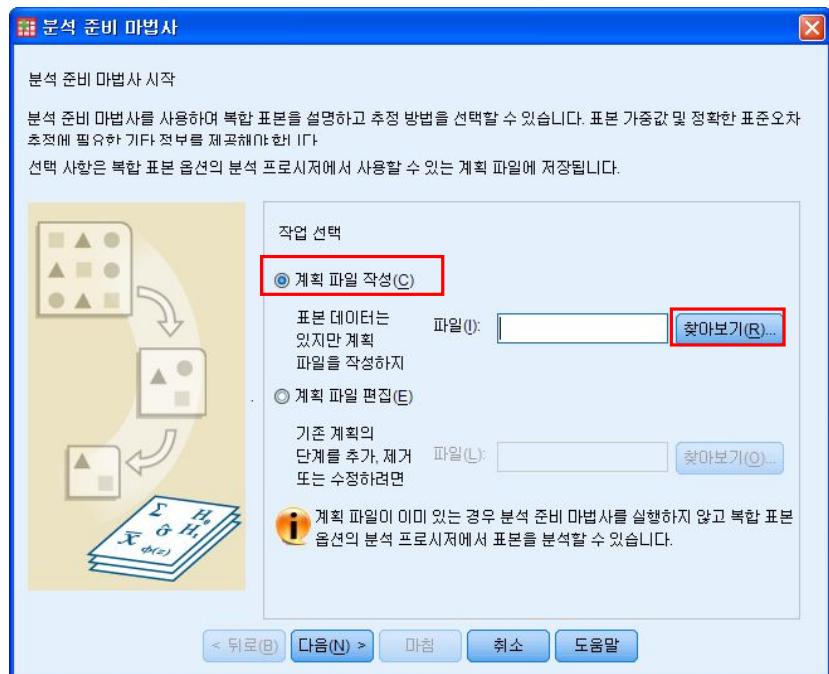
분석 준비

- **분석** ► **복합표본** ► **분석준비** 를 선택 한다.



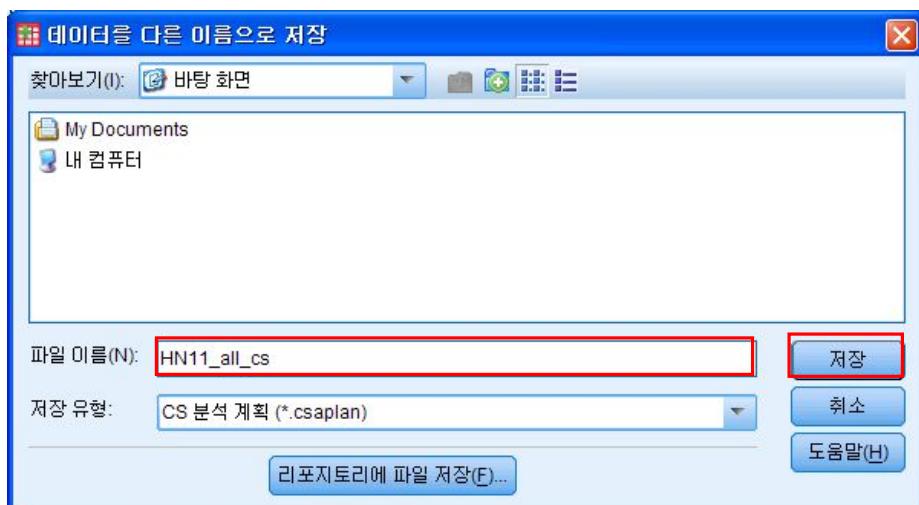
계획 파일 작성

- 계획 파일 작성 ► 찾아보기 를 클릭한다.



계획 파일 저장

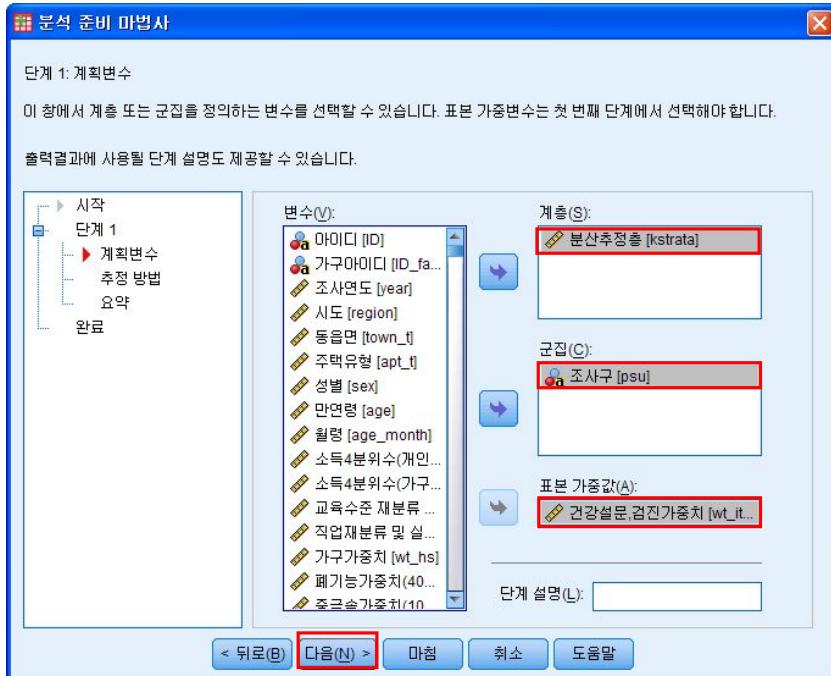
- 데이터를 다른 이름으로 저장 창이 나타나면 저장 경로 지정 후, 생성할 DB명 (예: hn11_all_cs)을 파일이름에 입력하고 다음 을 클릭한다.



- 계획 파일 작성 화면이 다시 나타나면 다음 을 클릭한다.

계획변수 설정

- 복합표본 설계요소 변수를 선택한다.
 - 1) 계층 : 충화변수인 **분산추정층(kstrata)** 지정
 - 2) 군집 : 집락변수인 **조사구(psu)** 지정
 - 3) 표본가중값 : 분석변수에 적합한 가중치 지정(예: 건강설문, 검진가중치(wt_itvex))



- 복합표본 설계요소 변수들을 지정한 후 **다음(N) >** 을 클릭한다.

추정방법 설정

- 표준오차 추정방법을 선택한다.

- 국민건강영양조사 표본은 비복원추출방법으로 설계되었으나, 1차 추출단위의 추출율이 낮으므로 **복원 표본추출과 무한 모집단수정(FPC)**을 선택한다.

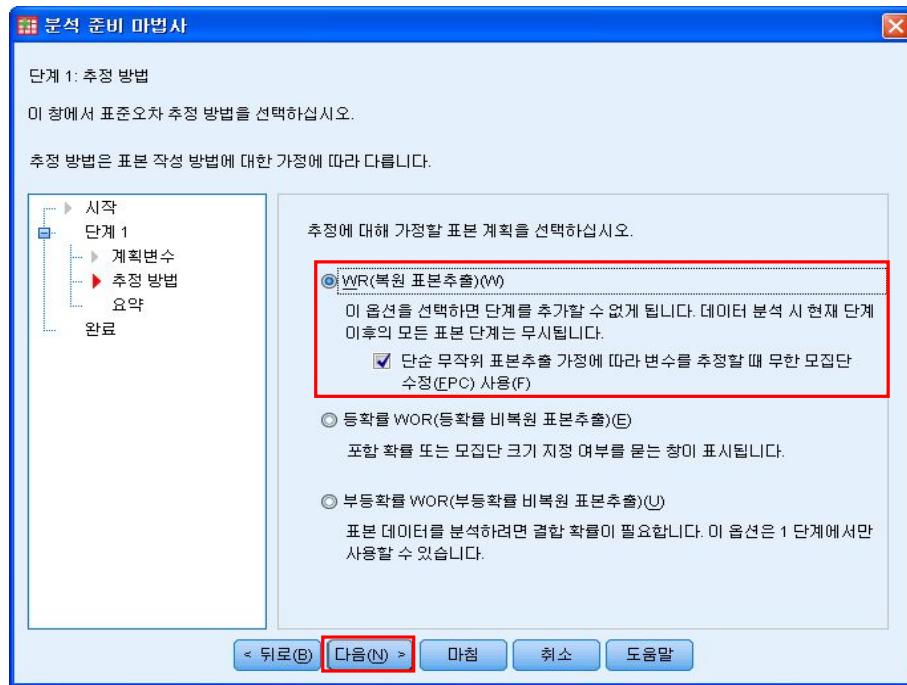
※ 복합표본설계 분산추정 방법

$$\sum_{h=1}^H \frac{n_h (1 - f_h)}{n_h - 1} \sum_{i=1}^{n_h} (e_{hi.} - \bar{e}_{h..})^2$$

$$f_h = n_h / N_h$$

$$e_{hi.} = \sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \hat{y}) / w_{...}$$

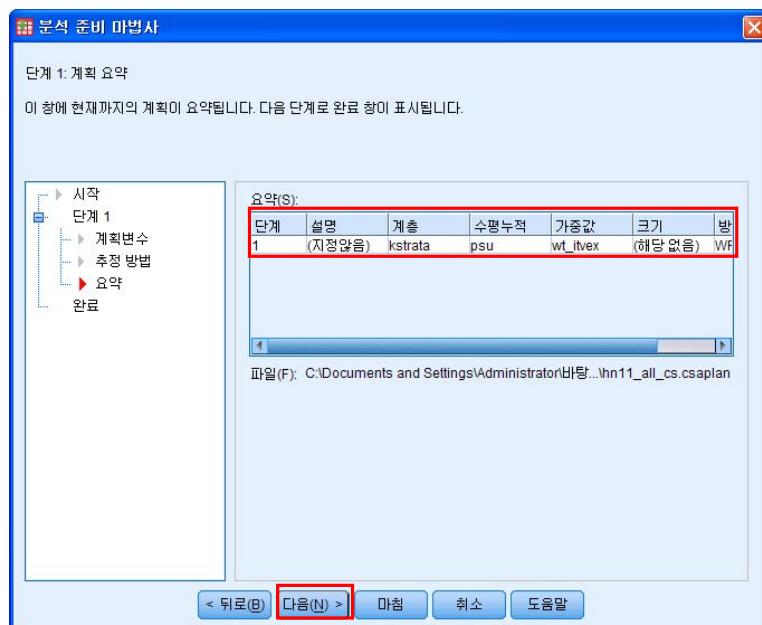
$$\bar{e}_{h..} = \sum_{i=1}^{n_h} e_{hi.} / n_h$$



- 표준오차 추정방법을 선택한 후 **다음(N)** 을 클릭한다.

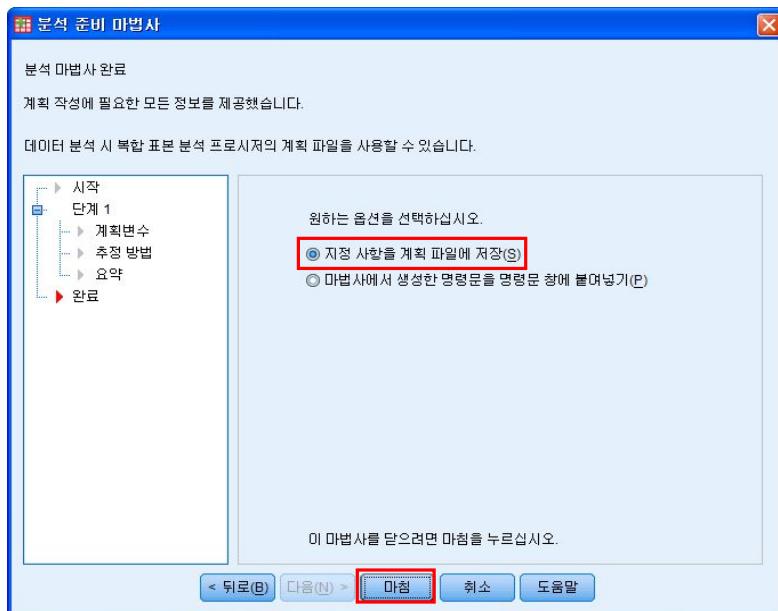
계획 요약

- 분석 계획파일에서 지정된 복합표본설계 요소를 확인 후 **다음(N)** 을 클릭한다.
 - 1) 각 단계에서 진행된 내용을 요약하여 보여준다.
 - 2) 계층, 수평누적, 가중값 등을 확인한다.

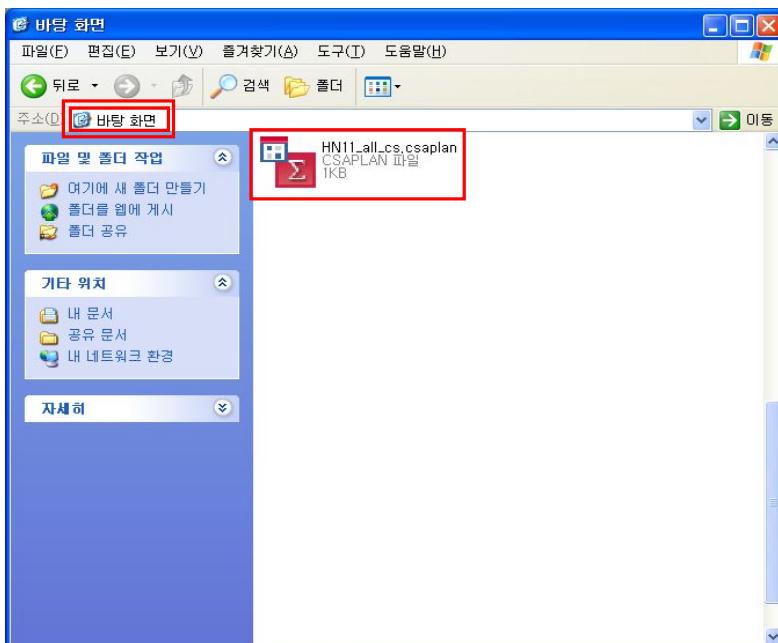


계획파일 생성 완료

- 분석 준비 마법사의 마지막 단계로 **지정 사항을 계획 파일에 저장** 을 선택 후 **마침** 을 클릭한다.

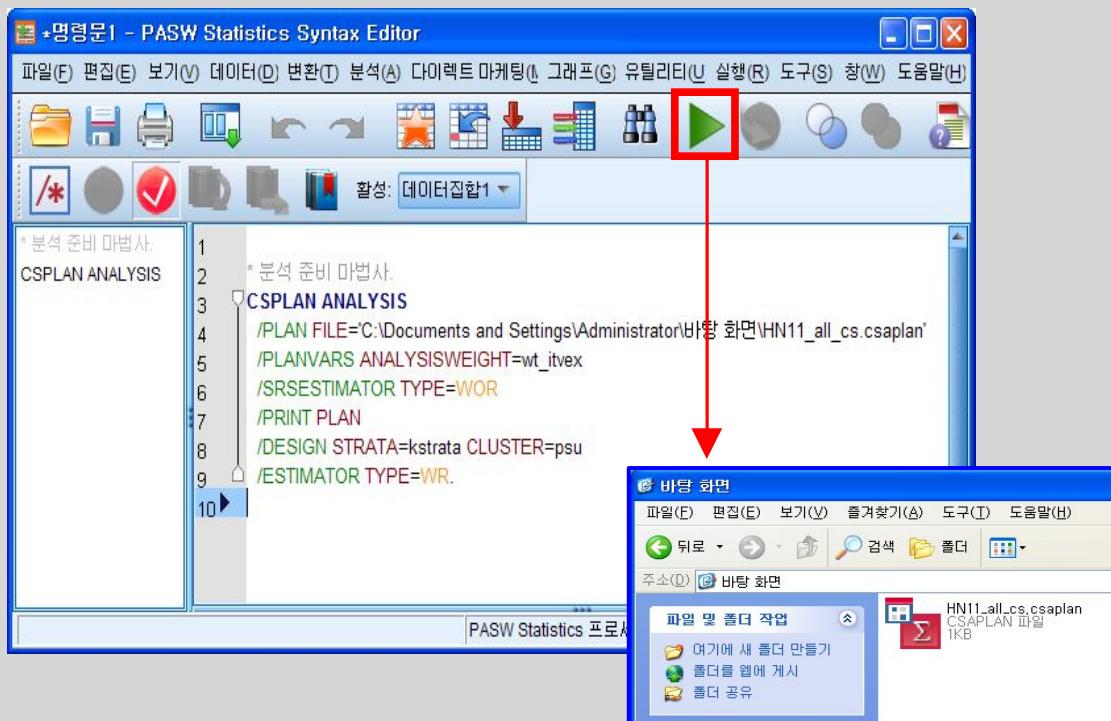


- 생성된 계획파일(HN11_all_cs.CSPLAN)이 계획파일 저장(p48) 단계에서 지정한 폴더(바탕화면)에 생성되었는지 확인한다.



※ 명령문을 이용한 분석계획파일 생성

- [분석 준비 마법사] 의 마지막 단계(p.51)에서 [마법사에서 생성한 명령문을 명령문 창에 붙여넣기]를 선택한 후 마침을 클릭한다.
- 명령문 창에는 각 단계에서 지정한 내용이 포함된 코드가 나타나고, 명령문 실행을 하면(▶클릭) 지정된 폴더에 분석계획파일(hn11_all_cs.csplan)이 생성된다. 이후 분석시, 분석계획파일을 다시 생성해야 할 때, 기존의 클릭방식으로 분석계획파일을 생성하거나 명령문에서 해당 부분을 수정한 후 명령문 실행을 통해서 생성할 수 있다.



※명령문 사용방법은 P39-41 참조.

□ 자료분석 시 유의사항

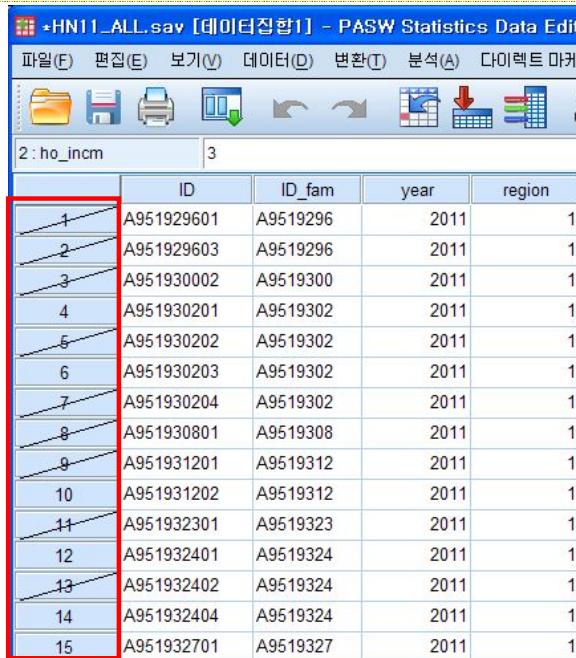
분석 유의사항 1 : 복합표본설계자료에서 케이스 삭제 및 선택 시 추정치의 표준오차 편향 발생 => 집단변수 생성 후 부모집단 분석

국민건강영양조사의 표본은 복합표본설계방법인 다단계총화집락화률추출방법에 의해 선정되었고, 복합표본설계 요소인 층, 집락, 가중치 정보가 원시자료DB 전체에 포함되어 있다. 따라서 자료분석 시 전체 자료가 아닌 일부자료만으로 분석DB를 생성할 경우(즉, 비해 당 자료를 분석DB에서 삭제하는 경우) 삭제된 자료에 포함된 복합표본설계 정보가 누락됨에 따라 추정치의 표준오차에 편향이 발생할 수 있다. 따라서 관심집단에 대한 분석 시, 전체 자료를 분석 DB로 하고, 관심집단과 그외 집단을 구분하는 변수(예: 관심집단 group=1, 그 외 group=0) 생성 후, 생성된 관심집단 변수를 부모집단으로 지정하여 분석한다. 단, 복합표본 빈도분석 및 교차분석에서는 집단변수를 부모집단으로 지정하기 위해서 분석계획파일 생성 시 집단변수를 계층 변수에 반드시 포함하여야 한다.

* 상세방법: “교차분석 예제(p72-74)” 참고

- 그림1과 같이 케이스선택을 이용하여 일부자료 사용 시, 제외된 자료의 복합표본설계정보가 누락되어, 추정치의 표준오차에 편향이 발생할 수 있으므로, 그림2와 같이 전체 자료를 사용하되, 관심집단 변수를 생성 후 복합표본설계 분석 내의 부모집단창 이용

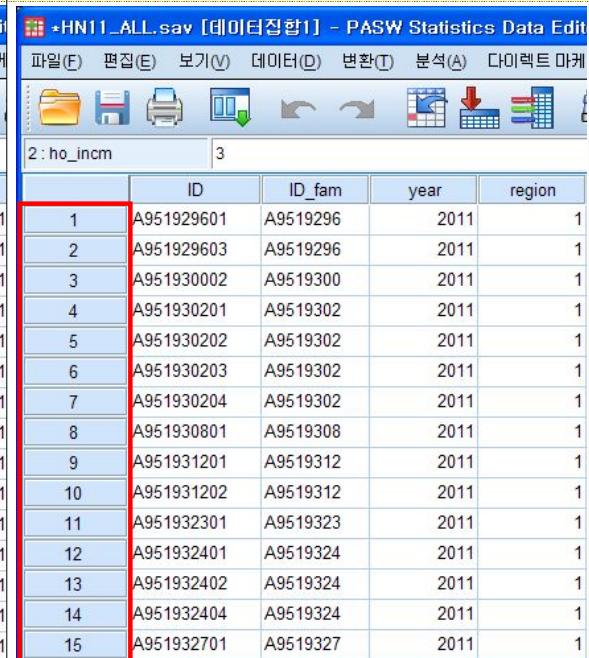
<그림1. 부분 데이터 사용(케이스 선택)>



The screenshot shows the PASW Statistics Data Editor interface. A red box highlights the first 15 rows of data, which correspond to cases 1 through 15. The columns are labeled ID, ID_fam, year, and region. All data points in these columns are identical across all 15 rows.

	ID	ID_fam	year	region
1	A951929601	A9519296	2011	1
2	A951929603	A9519296	2011	1
3	A951930002	A9519300	2011	1
4	A951930201	A9519302	2011	1
5	A951930202	A9519302	2011	1
6	A951930203	A9519302	2011	1
7	A951930204	A9519302	2011	1
8	A951930801	A9519308	2011	1
9	A951931201	A9519312	2011	1
10	A951931202	A9519312	2011	1
11	A951932301	A9519323	2011	1
12	A951932401	A9519324	2011	1
13	A951932402	A9519324	2011	1
14	A951932404	A9519324	2011	1
15	A951932701	A9519327	2011	1

<그림2. 전체자료 사용 및 관심변수 생성>



The screenshot shows the PASW Statistics Data Editor interface. A red box highlights the first 15 rows of data, which correspond to cases 1 through 15. The columns are labeled ID, ID_fam, year, and region. All data points in these columns are identical across all 15 rows.

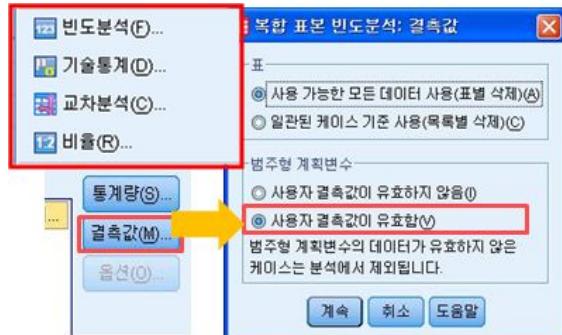
	ID	ID_fam	year	region
1	A951929601	A9519296	2011	1
2	A951929603	A9519296	2011	1
3	A951930002	A9519300	2011	1
4	A951930201	A9519302	2011	1
5	A951930202	A9519302	2011	1
6	A951930203	A9519302	2011	1
7	A951930204	A9519302	2011	1
8	A951930801	A9519308	2011	1
9	A951931201	A9519312	2011	1
10	A951931202	A9519312	2011	1
11	A951932301	A9519323	2011	1
12	A951932401	A9519324	2011	1
13	A951932402	A9519324	2011	1
14	A951932404	A9519324	2011	1
15	A951932701	A9519327	2011	1

분석 유의사항 2 : 복합표본설계자료에서 결측자료 처리

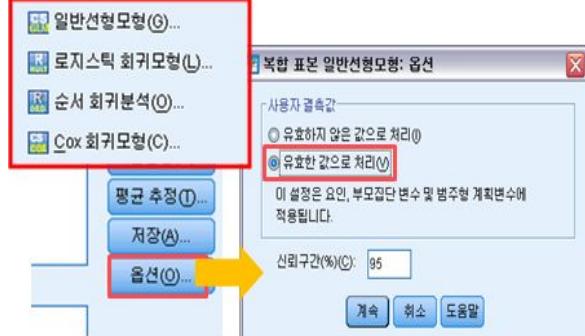
국민건강영양조사 대상자 중 조사에 참여하지 않은 경우(단위 무응답), 가중치에 응답률의 역수를 반영함으로써 응답률이 낮은 집단의 조사참여자가 더 큰 가중치를 부여받는 방법인 무응답 가중치 조정을 실시하였다. 조사참여자는 조사부문내 일부 조사항목에 대해서만 응답하고 일부 조사항목에 대해서는 무응답 하였더라도 해당 부문의 가중치를 부여받는다. 이와 같이 특정 조사항목에 대해 무응답한 경우(항목 무응답)에도, 자료분석 시에는 특정 조사항목 무응답자들의 표본설계정보도 누락되지 않고 반영되어야 한다. 이를 위해 복합표본 빈도분석, 복합표본 기술통계, 복합표본 교차분석 시에는 결측값 옵션 ▶ <범주형 계획변수>의 <사용자 결측값이 유효함>을 선택하고, 복합표본 일반선형모형, 복합표본 로지스틱회귀분석 시에는 옵션 ▶ <사용자 결측값>의 <유효한 값으로 처리>를 선택한다 (그림 참조). 이러한 옵션을 지정하지 않을 경우 추정치의 표준오차가 과소추정될 수 있으므로, 옵션지정 후 분석하기를 권고한다.

○ 결측값 처리 옵션

<복합표본 빈도분석 · 기술통계 · 교차분석>

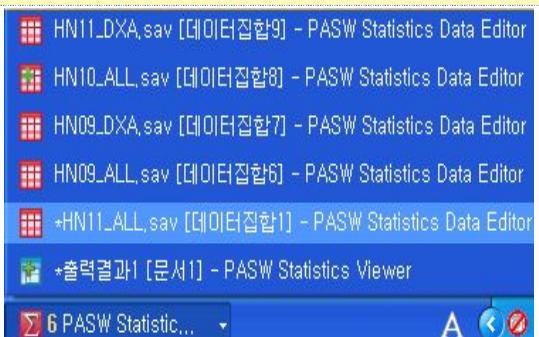


<복합표본 일반선형모형 · 로지스틱회귀분석>



분석 유의사항 3 : 분석연도DB만을 활성화하여 사용

그림1과 같이 여러연도의 DB를 열어놓은 경우, DB들의 형태가 유사하여, 분석연도DB가 아닌 다른년도DB를 현재 화면에 보이는 창으로 두고 분석을 실행할 경우, 다른년도의 결과를 얻을수 있으므로, 그림2와 같이, 가급적 분석연도 DB만을 활성화하여 사용한다.

<그림1. 여러연도의 DB가 열려있는 경우>	<그림2. 다른연도의 DB를 닫고 분석DB만 열어서 분석시>
	

복합표본 빈도분석

잠깐!! 분석 전에는 자료분석 유의사항(p53~55)을 반드시 확인해주시기 바랍니다.

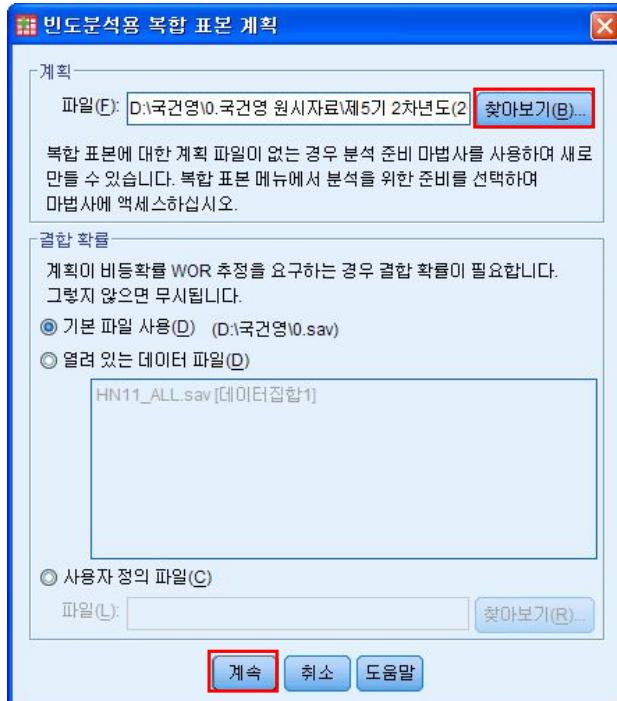
- 복합표본 빈도분석 프로시저에서는 범주형 변수의 빈도표와 일변량 통계량(모집단의 크기, 표 퍼센트, 표준오차, 신뢰구간, 변동계수, 가중되지 않은 빈도, 계획효과, 계획효과의 제곱근, 누적값)을 제공한다.

분석 준비

- 분석 ► 복합표본 ► 빈도분석 을 선택 한다.

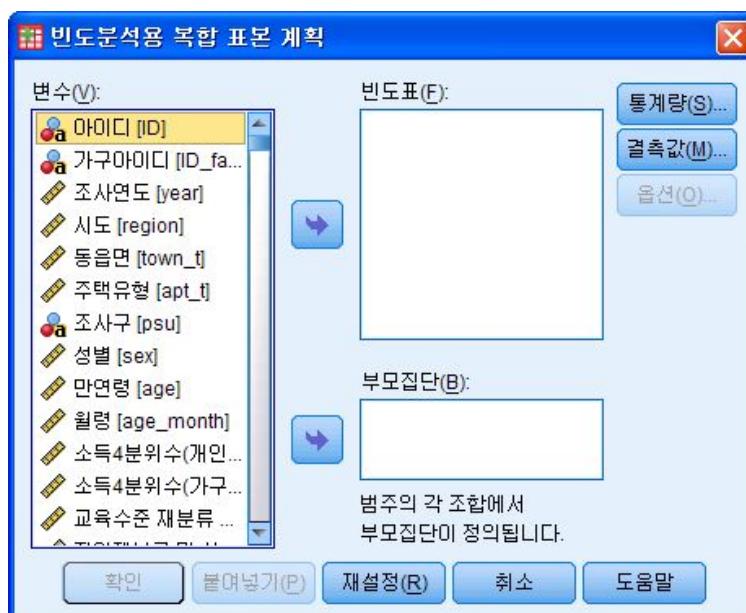


- 분석계획파일을 **찾아보기**에서 선택한 후, **계속**을 클릭한다.



빈도분석 대화상자

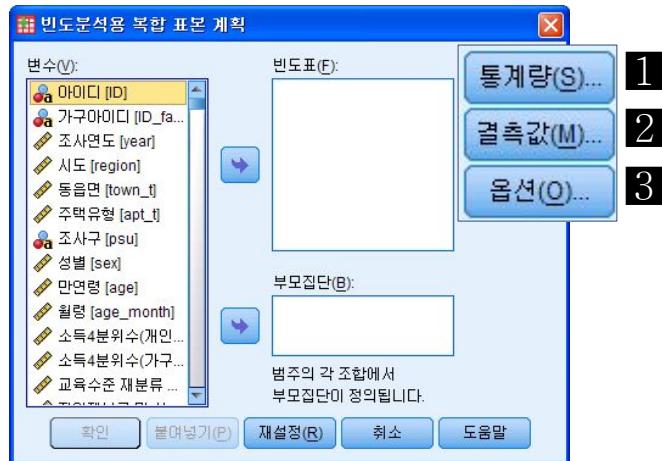
- 복합표본 빈도분석 대화상자가 나타나면 분석변수는 빈도표에, 관심집단은 부모집단에 넣는다.



- 빈도표 : 빈도표로 작성하고자 하는 **범주형** 변수를 선택한다.
- 부모집단 : 부모(세부)집단으로 사용할 변수를 선택한다. 부모집단으로 사용하고자 하는 변수가 없을 때는 지정하지 않는다.

빈도분석 옵션 대화상자

- 복합표본 빈도분석 대화상자에서 분석할 변수를 선택한 다음, 그 외 필요한 옵션들을 옵션 대화상을 통해 선택한다.



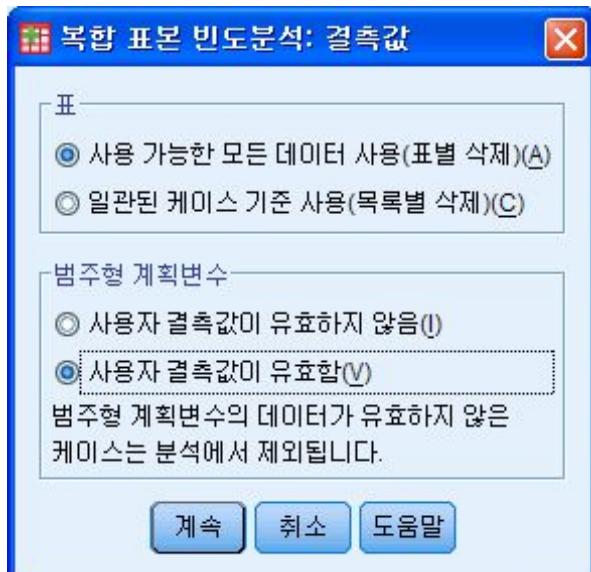
1 통계량

통계량 대화상자에서 산출 통계량을 선택할 수 있다.



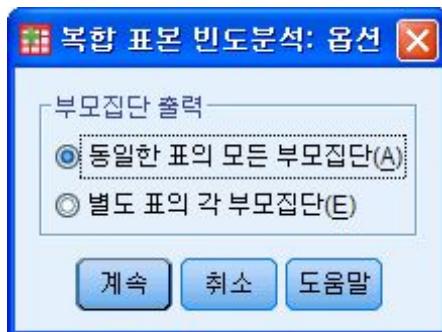
2 결측값

국민건강영양조사의 경우 사용 가능한 모든 데이터를 사용하므로 결측값을 유효하게 처리한다. 따라서 사용 가능한 모든 데이터 사용, 사용자 결측값이 유효함을 선택한다.



3 옵션

옵션 대화상자는 부모(세부)집단을 지정한 경우 활성화되며, 부모집단에 따라 동일한 표나 별도의 표로 출력하도록 한다. 원하는 출력형태를 선택 할 수 있다.

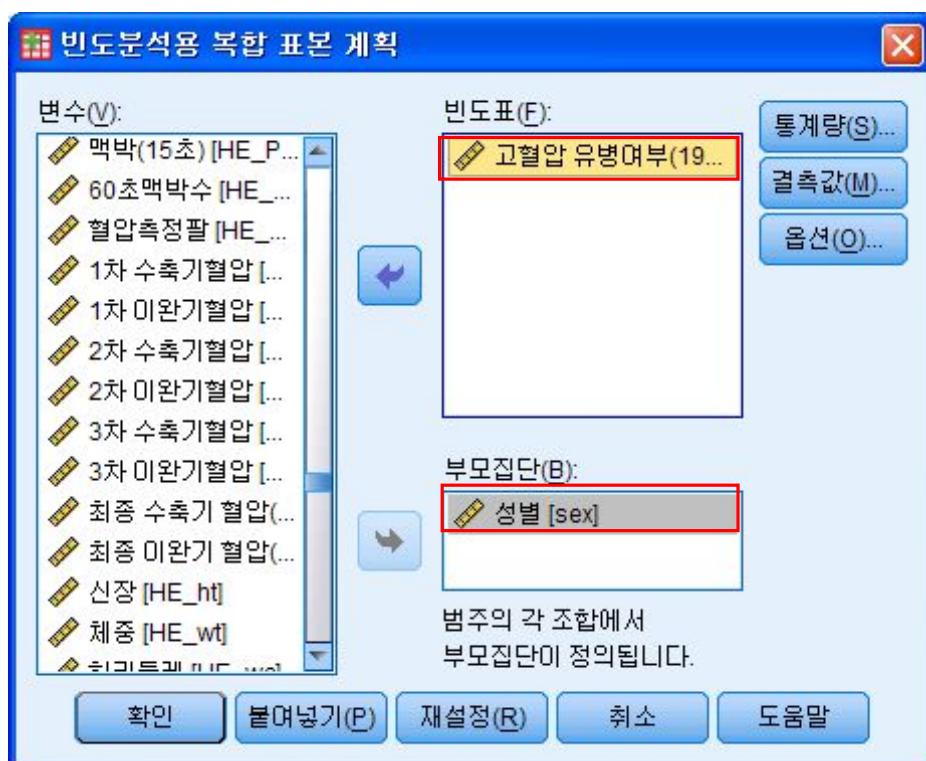


복합표본 빈도분석 예제

- 2011년 국민건강영양조사 자료를 이용하여 성별 고혈압유병여부의 빈도와 일변량 통계량을 산출하고자 한다.

자료 분석

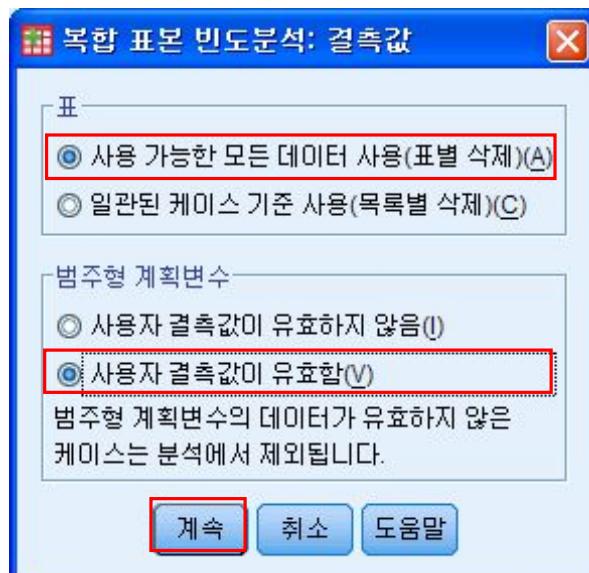
- **분석** ► **복합표본** ► **빈도분석** 을 선택한 후, 분석 계획 파일을 선택하면 다음과 같은 복합표본 빈도분석 대화상자가 나타난다.
※ 분석계획파일의 복합표본설계 요소(가중치:wt_itvex, 층:Kstrata, 군집:PSU) 지정
- 분석에 사용될 변수를 해당 창에 지정 한 후, 보고자 하는 옵션을 선택한다.
 - 1) 빈도표 : 고혈압유병여부(**HE_HP**)
 - 2) 부모집단 : 성별(**SEX**)



- 통계량으로 표 퍼센트, 가중되지 않은 빈도를 선택하고, 계속을 클릭한다.



- 결측값 대화상자에서는 사용 가능한 모든 데이터 사용, 사용자 결측값이 유효함을 선택하고, 계속을 클릭한다.



- 원하는 옵션을 선택 한 후, 최종적으로 빈도분석용 복합표본계획 창에서 확인을 클릭하면 다음과 같은 결과가 산출된다.

분석 결과

복합 표본: 표

고혈압 유병여부(19세이상)

	추정값	가중되지 않은 빈도
전체 중 % 1	47.8%	2622
2	26.0%	1464
3	26.2%	1932
합계	100.0%	6018

부모집단 표

고혈압 유병여부(19세이상)

성별		추정값	가중되지 않은 빈도
1	전체 중 % 1	37.7%	858
	2	32.8%	777
	3	29.5%	920
	합계	100.0%	2555
2	전체 중 % 1	57.6%	1764
	2	19.5%	687
	3	23.0%	1012
	합계	100.0%	3463

변수	코딩값	설명
SEX	1	남자
	2	여자
HE_HP	1	정상
	2	고혈압 전단계
	3	고혈압

결과해석

성별에 따른 고혈압 유병여부는 남자 29.5%, 여자 23.0%로 나타났다.

▶ 복합표본 빈도분석 예제 명령문

CSTABULATE

```
/PLAN FILE='D:\제5기 2차년도(2011)\SPSS\HN11_all_cs.csaplan'
/TABLES VARIABLES=HE_HP
/SUBPOP TABLE=sex DISPLAY=LAYERED
/CELLS TABLEPCT
/STATISTICS COUNT
/MISSING SCOPE=TABLE CLASSMISSING=INCLUDE.
```

*명령문 활용에 대한 상세내용은 p39 참조.

복합표본 기술통계

잠깐!! 분석 전에는 자료분석 유의사항(p53~55)을 반드시 확인해주시기 바랍니다.

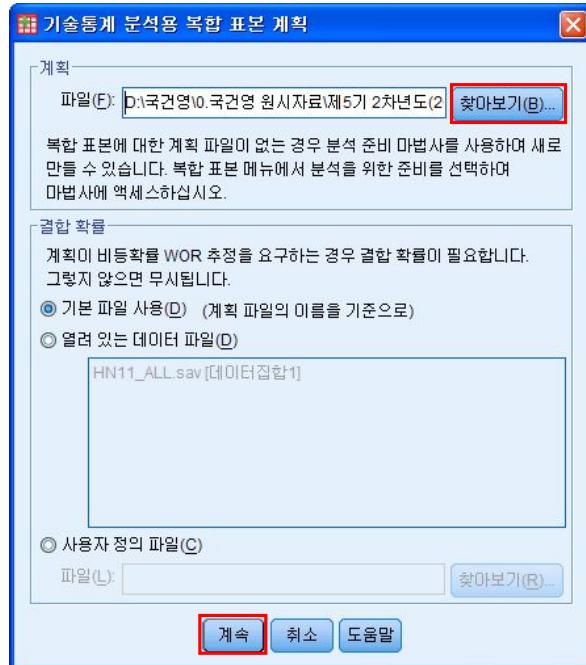
- 복합표본 기술통계 프로시저에서는 연속형 변수의 일변량 통계량(평균, 합계, T-test, 표준오차, 신뢰구간, 변동계수, 가중되지 않은 빈도, 모집단의 크기, 계획 효과, 계획 효과 제곱근)을 제공한다.

분석 준비

- 분석 ► 복합표본 ► 기술통계 를 선택한다.

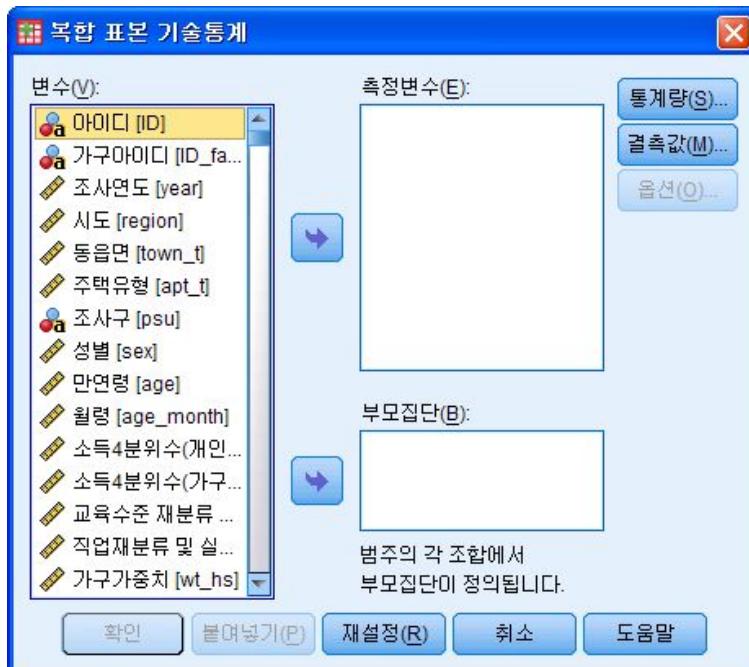


- 분석계획파일을 **찾아보기**에서 선택한 후, **계속**을 클릭한다.



기술통계 대화상자

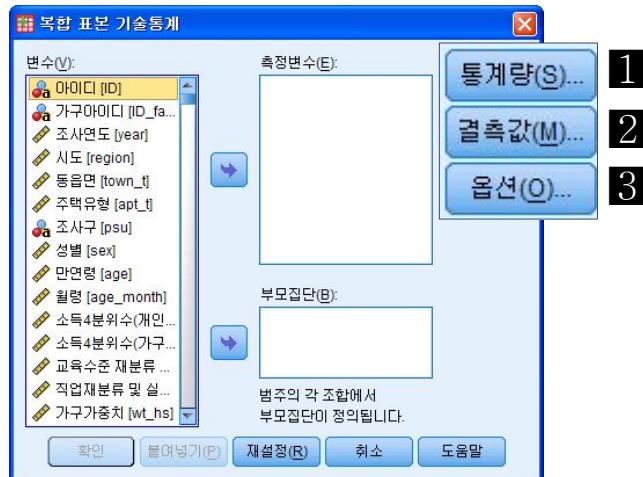
- 복합표본 기술통계 대화상자가 나타나면 분석변수는 측정변수에, 관심집단은 부모집단에 넣는다.



- 측정변수 : 연속형 변수를 선택한다.
- 부모집단 : 부모(세부)집단으로 사용할 변수를 선택한다. 부모집단으로 사용하고자 하는 변수가 없을 때는 지정하지 않는다.

기술통계 옵션 대화상자

- 복합표본 빈도분석 대화상자에서 분석변수를 선택한 다음, 그 외 필요한 옵션들은 옵션 대화상을 통해 선택한다.



1 통계량

통계량 대화상자에서 산출 통계량을 선택할 수 있다.



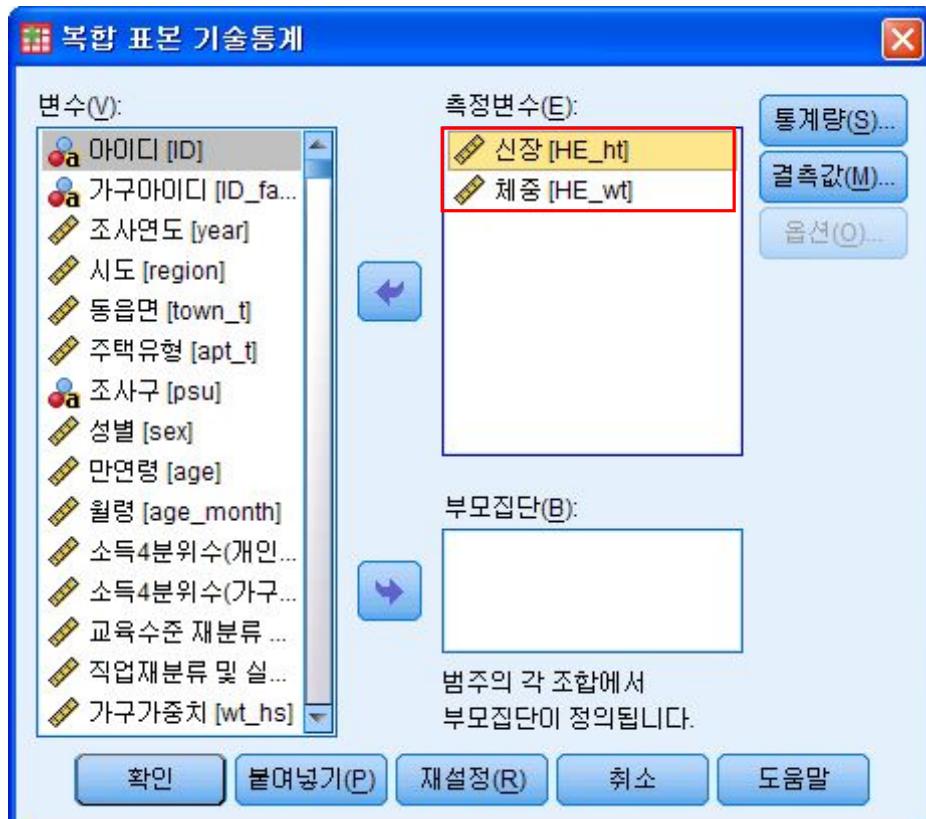
2 결측값, 3 옵션은 복합표본 빈도분석과 동일함(p59 참조).

복합표본 기술통계 예제

- 2011년 국민건강영양조사 자료를 이용하여 신장 및 체중의 평균, 신뢰구간, 가중되지 않은 빈도를 산출하고자 한다.

자료 분석

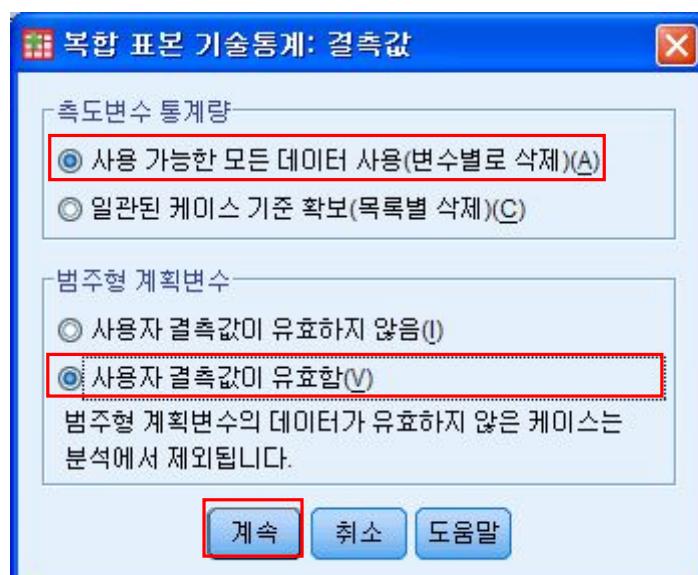
- **분석** ► **복합표본** ► **기술통계** 를 선택한 후, 분석 계획 파일을 선택하면 다음과 같은 복합표본 기술통계 대화상자가 나타난다.
※ 분석계획파일의 복합표본설계 요소(가중치:wt_itvex, 층화:Kstrata, 군집:PSU)
 - 분석에 사용될 변수를 해당 창에 지정 한 후, 산출할 옵션을 선택한다.
- 1) 측정변수 : 신장(HE_ht), 체중(HE_wt)**
- ※ 측정변수창에는 산출하고자 하는 여러개의 연속형 변수를 넣을 수 있다.
- 2) 부모집단 : 지정하지 않음**



- 통계량으로 **평균**, **표준오차**, **신뢰구간**, **가중되지 않는 빈도**를 선택하고, **계속**을 클릭한다.



- 결측값 대화상자에서는 **사용 가능한 모든 데이터 사용**, **사용자 결측값이 유효함**을 선택하고, **계속**을 클릭한다.



- 원하는 옵션을 선택 한 후, 최종적으로 복합표본 기술통계창에서 확인을 클릭하면 다음과 같은 결과가 산출된다.

분석 결과

복합 표본: 기술통계

일변량 통계량

	추정값	표준오차	95% 신뢰구간		가중되지 않은 빈도
			하한	상한	
평균 체중	59.04	.268	58.51	59.57	8024
신장	159.06	.313	158.44	159.67	8023

결과해석

평균 체중은 59.0kg, 평균 신장은 159.1cm로 나타났다.

▶ 복합표본 기술통계 예제 명령문

CSDESCRIPTIVES

```
/PLAN FILE='D:\제5기_2학년도(2011)\SPSS\HN11_all_cs.csaplan'
/SUMMARY VARIABLES=HE_wt HE_ht
/MEAN
/STATISTICS SE COUNT CIN(95)
/MISSING SCOPE=ANALYSIS CLASSMISSING=INCLUDE.
```

※ 명령문 활용에 대한 상세내용은 p39 참조.

복합표본 교차분석

잠깐!! 분석 전에는 자료분석 유의사항(p53~55)을 반드시 확인해주시기 바랍니다.

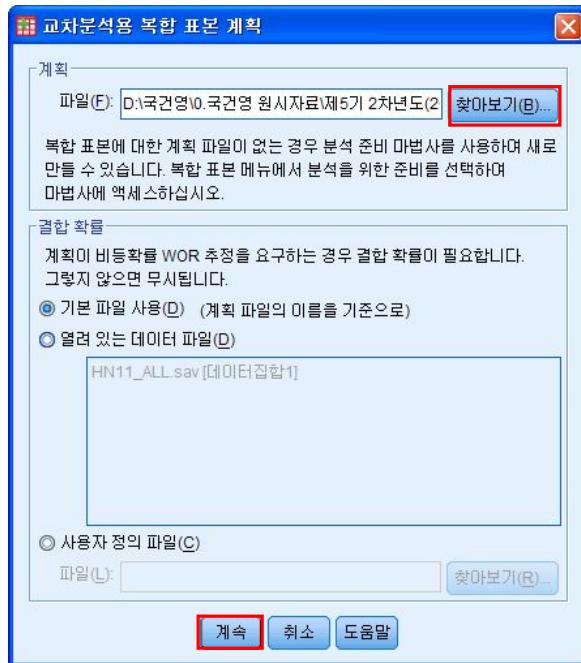
- 복합표본 교차분석 프로시저는 이원 교차표에 대한 검정 및 통계량을 제공한다.

분석 준비

- 분석 ► 복합표본 ► 교차분석 을 선택한다.



- 분석계획파일을 **찾아보기**에서 선택한 후, **계속**을 클릭한다.



교차분석 대화상자

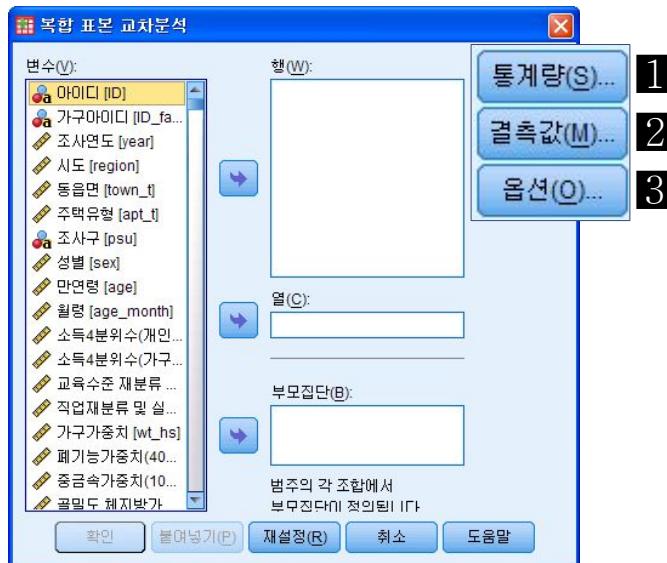
- 복합표본 교차분석 대화상자가 나타나면 분석변수는 행과 열에, 관심집단은 부모집단에 넣는다.



- 행, 열 : 범주형 변수를 선택한다.
- 부모집단 : 부모(세부)집단으로 사용할 변수를 선택한다. 부모집단으로 사용하고자 하는 변수가 없을 때는 지정하지 않는다.

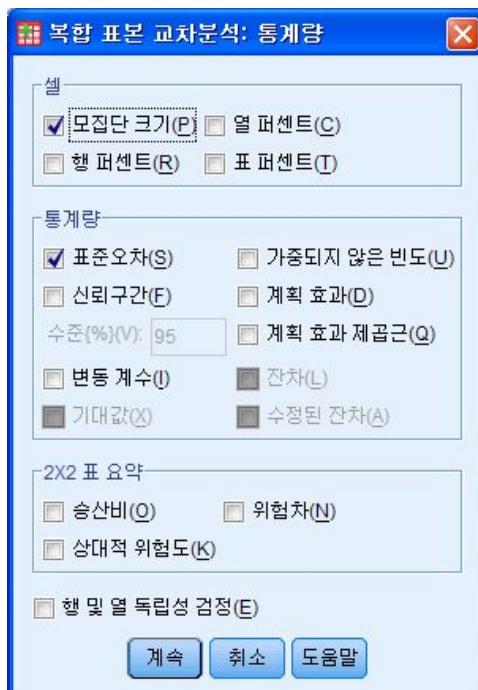
교차분석 옵션 대화상자

- 복합표본 교차분석 대화상자에서 분석할 변수를 선택한 다음, 그 외 필요한 옵션들을 옵션 대화상을 통해 선택한다.



1 통계량

통계량 대화상자에서 산출하고자 하는 하위 옵션을 선택 할 수 있다. 디폴트 값은 모집단 크기, 표준오차이다.



2 옵션, 3 결측값은 복합표본 빈도분석과 동일함(p59 참조).

복합표본 교차분석 예제

- 2011년 국민건강영양조사 자료를 이용하여 19세 이상 남자에서 연령과 비만과의 연관성 여부를 검정하고자 한다.
- 분석변수 요약
 - 기존 변수 : 성별(sex), 비만유병여부(HE_obe)
 - 생성 변수 : 연령대(age_group)

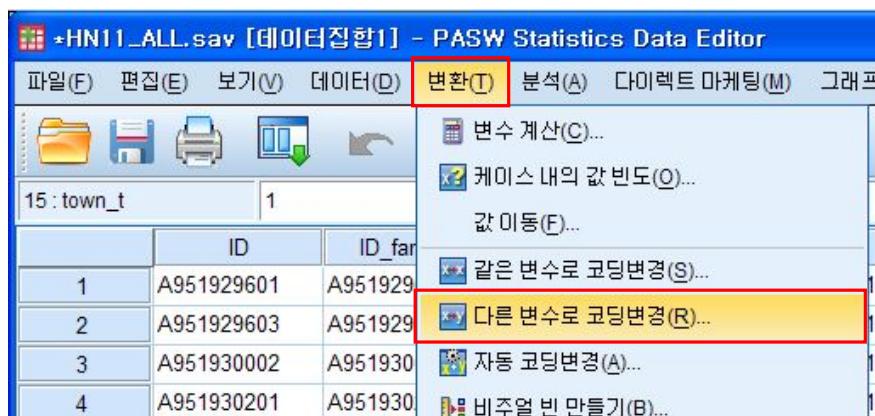
구분	변수명	변수설명	내용
분석대상	SEX	부모집단(성별)	1. 남자 2. 여자
종속변수	HE_BMI	비만유병여부	1. 저체중 2. 정상 3. 비만
독립변수	age_group	연령대	1. 19-44세 2. 45-64세 3. 65세 이상

- 분석에 앞서, 생성변수를 만든 후 복합표본설계 교차분석을 실시한다.

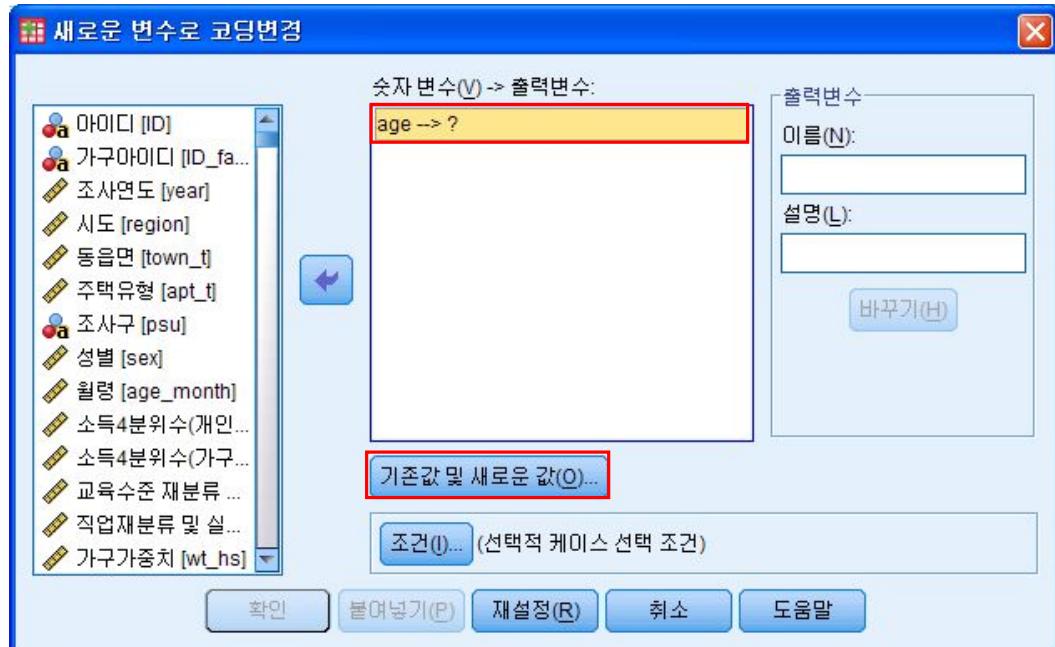
1 부모집단변수 생성

생성변수	설명	
age_group	지표설명	연령대
	변수유형	범주형
	조건식	RECODE age (MISSING=SYSMIS) (19thru44=1) (45thru64=2) (65thruHighest=3) INTO age_group. 사용변수 (age: 연령)

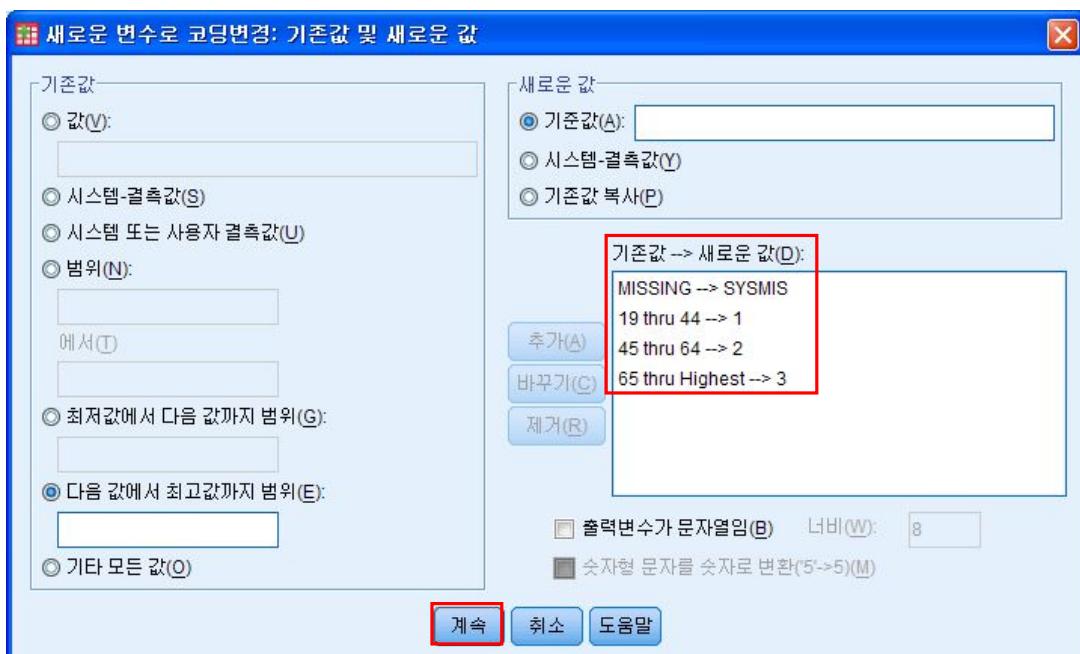
- 메뉴 창의 **변환** ▶ **다른 변수로 코딩변경** 을 선택한다.



- 생성변수를 만들기 위해 기존변수인 **age**를 선택하여 **숫자변수->출력변수** 창으로 옮겨놓은 후, **기준값 및 새로운 값** 을 클릭한다.

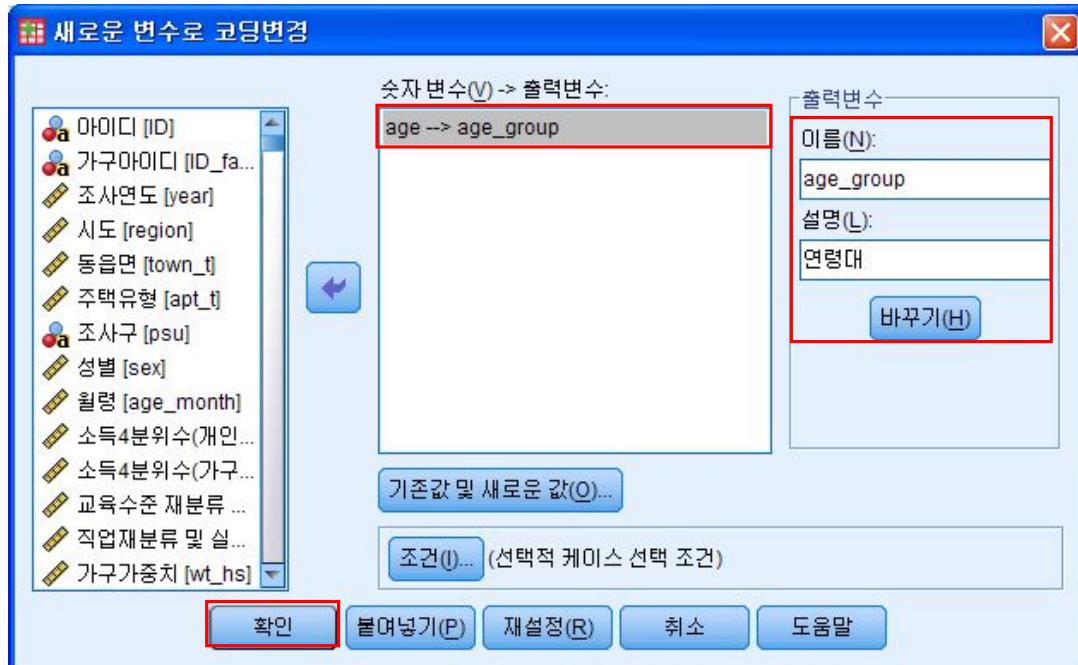


- 기준값(age변수)를 이용하여 새로운 값을 지정 한 후 **계속** 을 클릭한다.
※기준값이 시스템 또는 사용자 결측값 인 경우 새로운 값 또한 시스템-결측값 으로 지정해 준다.



- 출력변수

: 이름에 생성할 변수명(age_group)을 입력하고 **바꾸기**를 선택 한 후 **확인**을 클릭하면 새로운 변수가 생성된다. 설명은 연구자가 필요시에 작성한다.



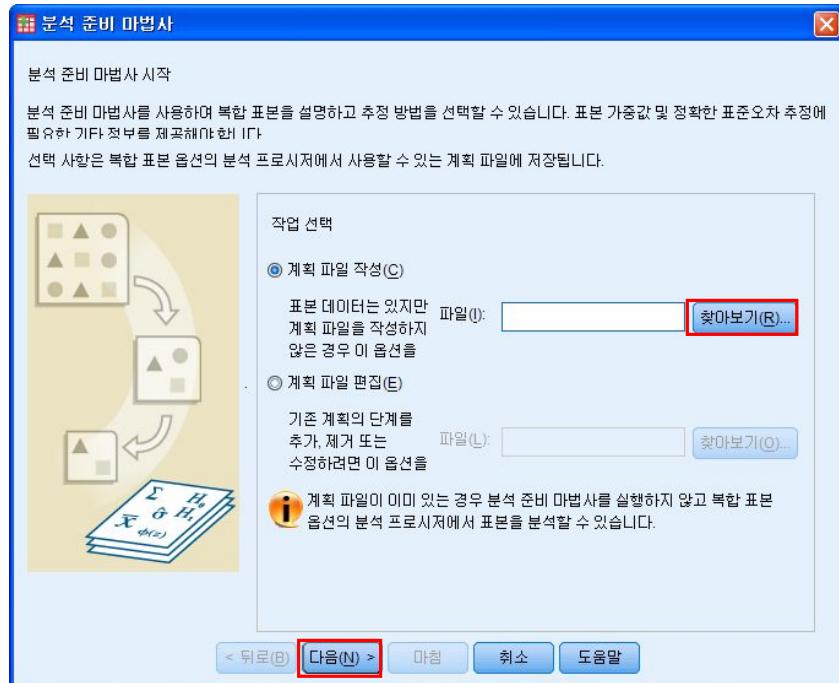
자료 분석

○ **분석** ► **복합표본** ► **분석준비** 를 선택 한다.

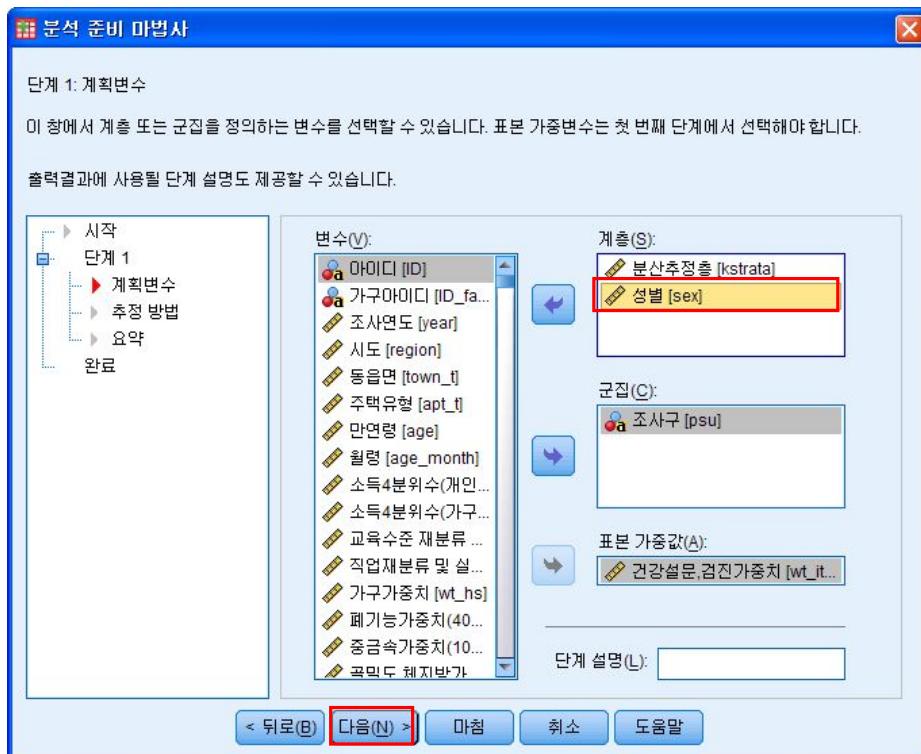
The screenshot shows the SPSS menu bar with '분석(A)' highlighted. Below it, the '복합 표본(L)' option under the '복합 표본' section is selected. A sub-menu '분석 준비(P)...' is displayed, with the '분석 준비(P)' option highlighted.

ID	ID_fam	own_t	apt_t	psu
A951929601	A9519296	1	1	A951
A951929603	A9519296	1	1	A951
A951930002	A9519300	1	1	A951
A951930201	A9519302	1	1	A951
A951930202	A9519302	1	1	A951
A951930203	A9519302	1	1	A951
A951930204	A9519302	1	1	A951
A951930801	A9519308	1	1	A951
A951931201	A9519312	1	1	A951
A951931202	A9519312	1	1	A951
A951932301	A9519323	1	1	A951
A951932401	A9519324	1	1	A951
A951932402	A9519324	1	1	A951
A951932404	A9519324	1	1	A951
A951932701	A9519327	1	1	A951
A951932702	A9519327	1	1	A951
A951932703	A9519327	1	1	A951
A951932704	A9519327	1	1	A951
A951932801	A9519328	1	1	A951
A951934101	A9519341	1	1	A951
A951934102	A9519341	1	1	A951
A951934902	A9519349	2011	1	
A951935201	A9519352	2011	1	
A951935202	A9519352	2011	1	
A951936001	A9519360	2011	1	
A951936002	A9519360	2011	1	
A951936003	A9519360	2011	1	
A951936004	A9519360	2011	1	
A951936801	A9519368	2011	1	
A951936802	A9519368	2011	1	

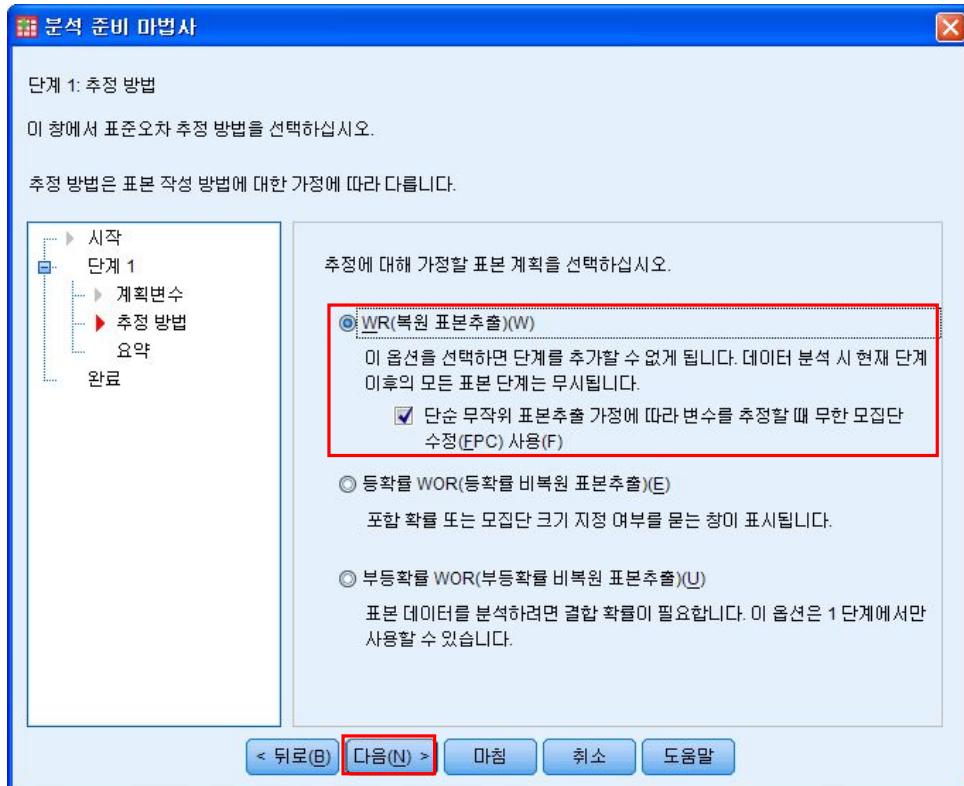
- 계획 파일 작성 ► 찾아보기 를 선택하여 저장할 경로를 지정하고 파일이름을 작성한 후 다음 을 선택한다.



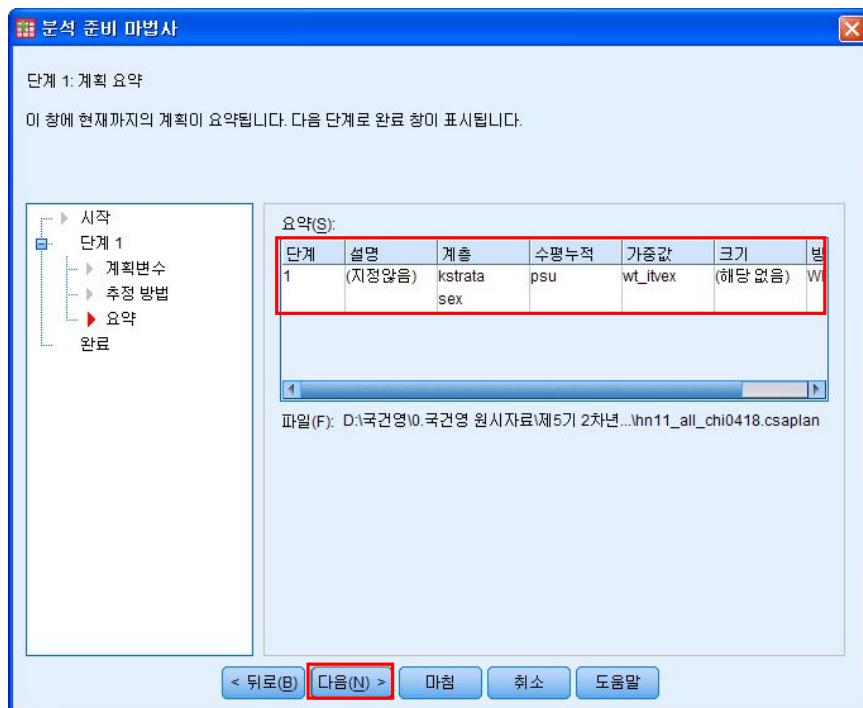
- 복합표본 설계요소 변수를 지정한다. 이때 부모집단으로 사용할 변수인 성별(SEX)을 계층변수에 추가로 선택 한다.



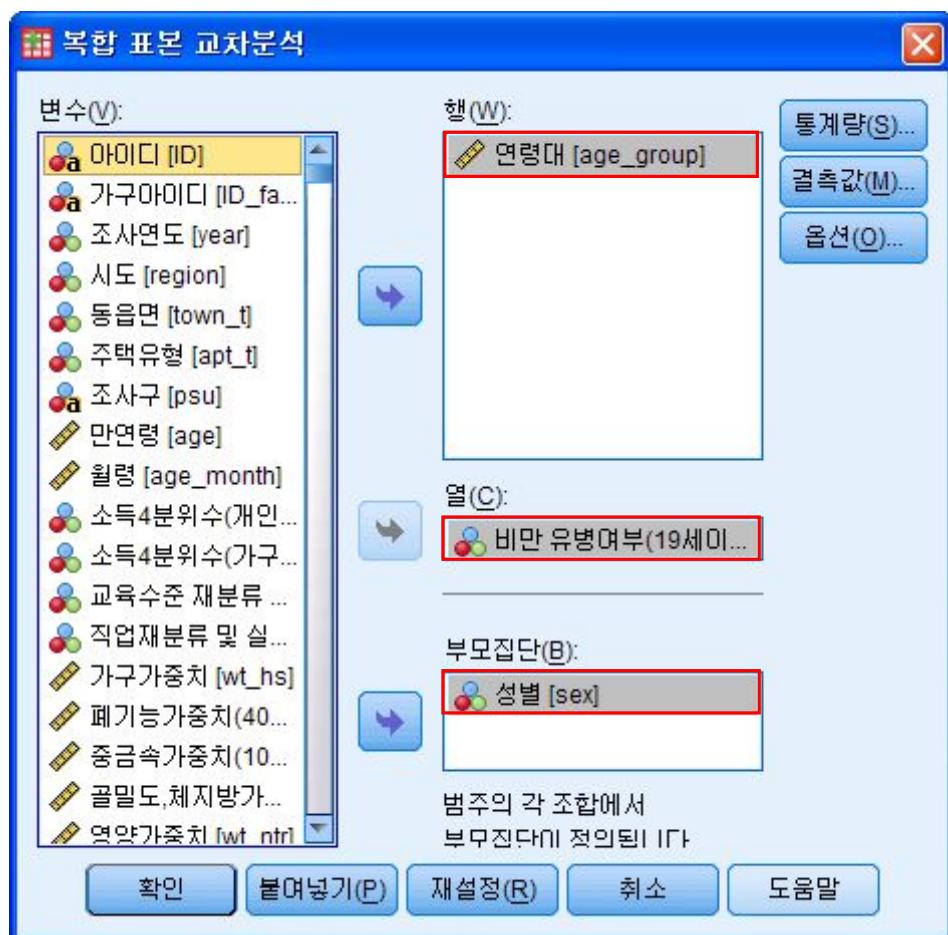
- 표준오차 추정방법을 선택한다.



- 분석 계획파일의 복합표본설계 변수가 지정된 것을 확인하고 분석계획 파일을 생성한다.



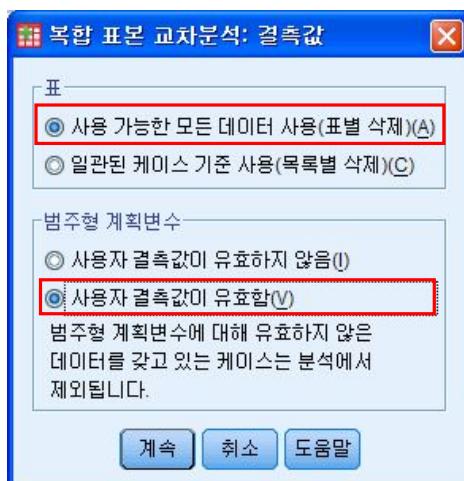
- 분석 ► 복합표본 ► 교차분석 을 선택한 후, 교차분석용 분석 계획 파일을 선택하면 다음과 같은 복합표본 교차분석 대화상자가 나타난다.
※ 분석계획파일의 복합표본설계 요소(가중치:wt_itvex, 층:Kstrata, 군집:PSU) 지정
- 분석에 사용될 변수를 해당 창에 지정 한 후, 보고자 하는 옵션을 선택한다.
 - 1) 행 : 연령대(age_group)
 - 2) 열 : 비만 유병여부(HE_obe)
 - 3) 부모집단 : 성별(SEX)



- 통계량은 **행 퍼센트**, **표준오차**, **가중되지 않는 빈도**, **행 및 열 독립성 검정** 을 선택하고, **계속** 을 클릭한다.



- 결측값은 **사용 가능한 모든 데이터 사용**, **사용자 결측값이 유효함** 을 선택하고, **계속** 을 클릭한다.



- 원하는 옵션을 선택 한 후, 최종적으로 복합표본 교차분석창에서 확인을 클릭하면 다음과 같은 결과가 산출된다.

분석 결과

성별 = 남자

연령대 * 비만 유병여부(19세이상)^a

연령대		비만 유병여부(19세이상)			
		1	2	3	합계
19-44 연령대 중 %	추정값	2.8%	60.0%	37.1%	100.0%
	표준오차	.6%	1.7%	1.8%	.0%
	가중되지 않은 빈도	27	595	376	998
45-64 연령대 중 %	추정값	2.5%	62.3%	35.2%	100.0%
	표준오차	.7%	1.8%	1.9%	.0%
	가중되지 않은 빈도	22	614	351	987
65+ 연령대 중 %	추정값	5.6%	67.8%	26.6%	100.0%
	표준오차	1.0%	2.3%	2.3%	.0%
	가중되지 않은 빈도	38	437	175	650
합계 연령대 중 %	추정값	3.1%	61.8%	35.2%	100.0%
	표준오차	.5%	1.2%	1.2%	.0%
	가중되지 않은 빈도	87	1646	902	2635

a. 성별 = 남자

독립성 검정^a

	카이 제곱	수정된 F	df1	df2	유의 확률
연령대 * 비만 유병여부 (19세이상)	Pearson	18.552	3.569	3.385	565.302 .011
	우도비	17.831	3.430	3.385	565.302 .013

수정된 F는 2차 Rao-Scott 수정된 카이제곱 통계량의 변형입니다. 유의도는 수정된 F 및 그 자유도를 기준으로 합니다.

a. 성별 = 남자

결과해석

19세 이상 남자에서 연령과 비만과의 연관성 검정 결과, 비만유병률은 19-44세 37.1%, 45-64세 32.5%, 65세이상에서 26.6%이며, p-value<0.05로 연관성이 유의적으로 있는 것으로 나타났다.

▶ 복합표본 교차분석 예제 명령문

* 변수생성 : 연령대.

```
RECODE age (MISSING=SYSMIS)(19thru44=1)(45thru64=2)(65thruHighest=3) INTO
    age_group.
VARIABLE LABELS age_group '연령대'.
EXECUTE.
```

* 복합 표본 교차분석.

CSTABULATE

```
/PLAN FILE='D:\제5기 2차년도(2011)\SPSS\HN11_all_chi.csaplan'
/TABLES VARIABLES=age_group BY HE_obe
/SUBPOP TABLE=sex DISPLAY=SEPARATE
/CELLS ROWPCT
/STATISTICS SE COUNT
/TEST INDEPENDENCE
/MISSING SCOPE=TABLE CLASSMISSING=INCLUDE.
```

※ 교차분석 부모집단 설정에 따라 csplan 파일이 달라지므로 주의.

※ 명령문 활용에 대한 상세내용은 p39 참조.

복합표본 일반선형모형

잠깐!! 분석 전에는 자료분석 유의사항(p53~55)을 반드시 확인해주시기 바랍니다.

- 복합표본 일반선형모형 프로시저에서는 복합표본설계자료에서 종속변수가 연속형일 때 종속변수와 독립변수간의 관련성을 분석한다.
- 일반선형모형에서는 T-검정, 분산분석, 회귀분석 등의 분석이 가능하며, 독립변수를 연속형으로 분석시에는 분석프로시저내의 공변량창에 지정, 범주형으로 분석시에는 요인분석창에 지정한다.

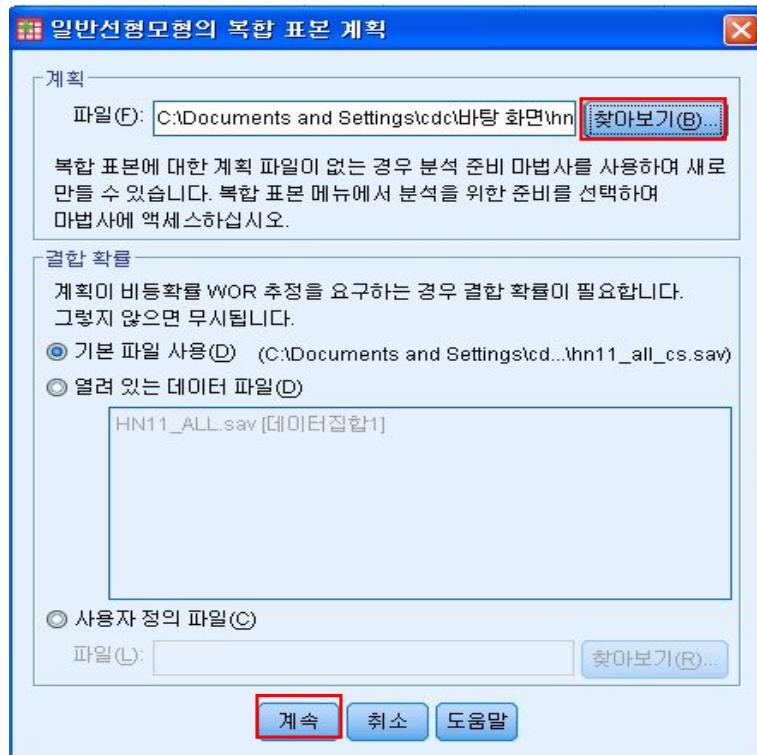
종속변수	독립변수	분석방법	SPSS 복합표본 프로시저
연속형	범주형(2그룹)	T검정	복합표본 일반선형모형
	범주형(3그룹이상)	분산분석	
	연속형	회귀분석	
	연속형,범주형	일반선형모형	

분석 준비

- 분석 ► 복합표본 ► 일반선형모형 을 선택한다.



- 분석계획파일을 **찾아보기**에서 선택한 후, **계속**을 클릭한다.



일반선형모형 대화상자

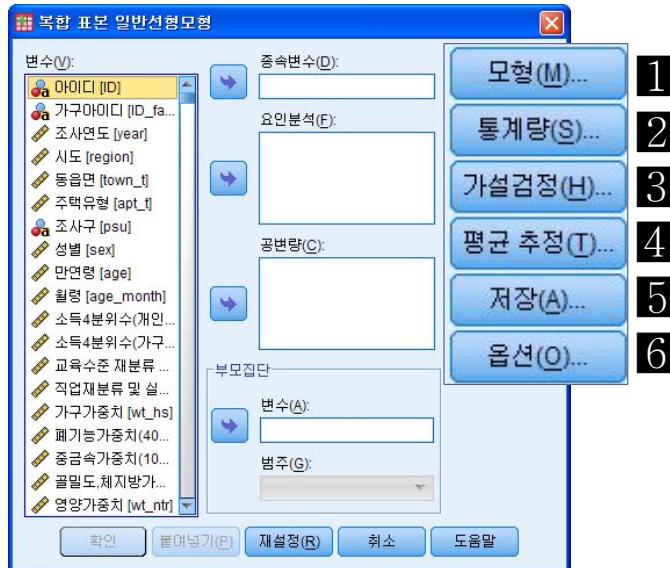
- 계획파일 선택을 완료하면 다음과 같은 복합표본 일반선형모형 대화상자가 나타난다.

종속변수	독립변수의 영향을 받는 변수(연속형)	
요인분석	범주형 독립변수	
공변량	연속형 독립변수	
부모집단	변수	분석하고자 하는 집단을 구분하는 변수
	범주	분석하고자 하는 집단을 구분하는 변수 중 해당집단의 범주 값



일반선형모형 옵션 대화상자

- 복합표본 일반선형모형 대화상자에서 분석을 원하는 변수를 선택하고, 그 외 필요한 옵션은 옵션 대화상을 통해 선택한다.



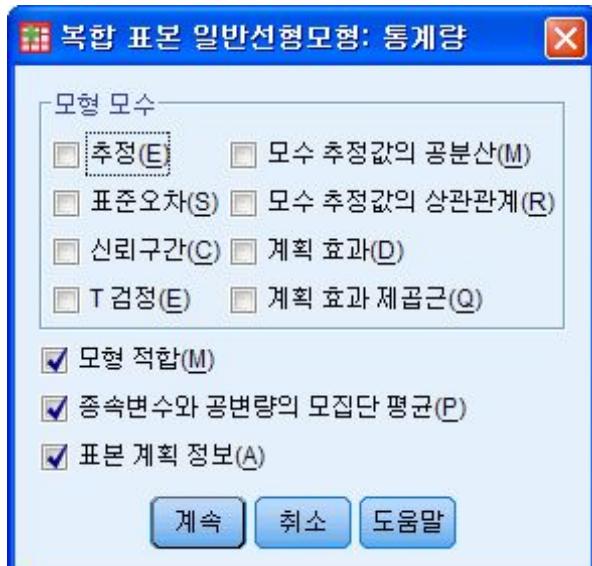
1 모형

사용자정의를 이용하여 연구자가 모형을 지정할 수 있으며, 기본값으로 일반선형모형 대화상자에서 지정한 요인변수와 공변량변수를 이용하여 주효과 모형을 제시한다.



2 통계량

산출하고자 하는 통계량을 선택 할 수 있다. 기본값은 모형적합, 종속변수와 공변량의 모집단 평균, 표본계획정보이다.



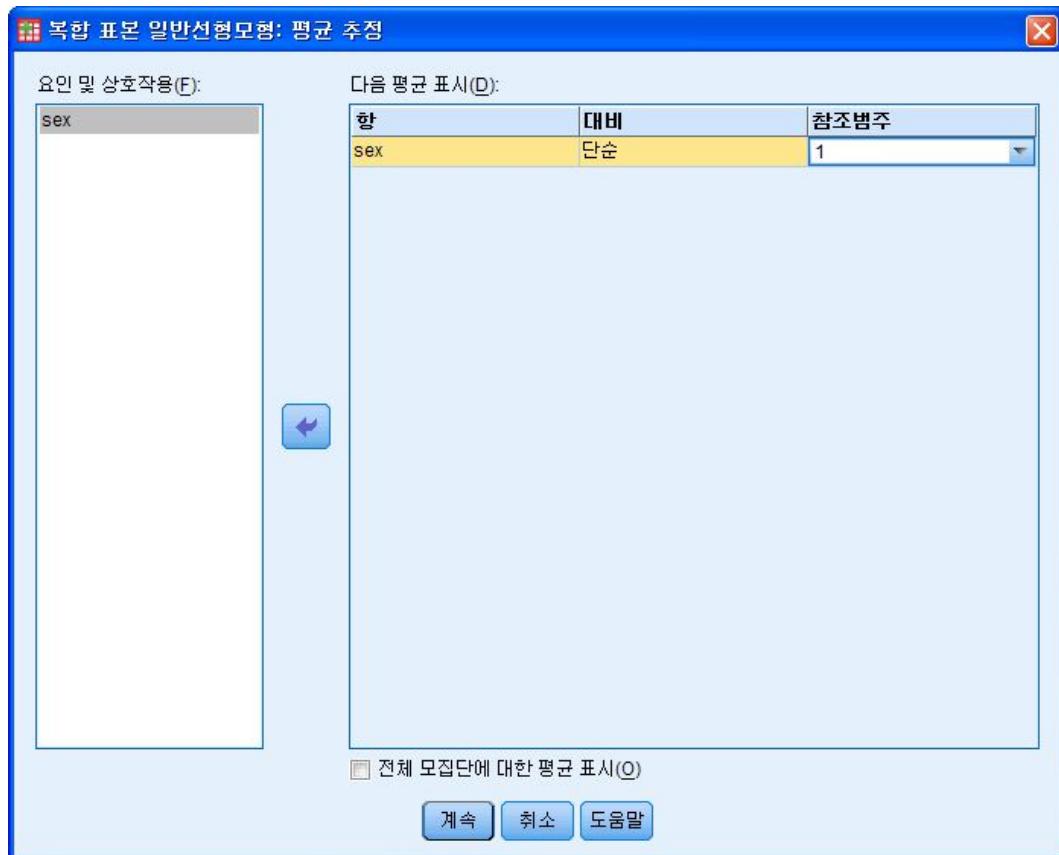
3 가설검정

가설을 검정하는데 사용되는 검정통계량, 다중비교에 대한 조정, 자유도 표본추출방법을 지정할 수 있다. 검정에 사용되는 옵션을 선택 할 수 있다.



4 평균추정

요인 및 상호작용 항에 있는 범주형 변수의 추정평균을 산출할 수 있다. 연구자가 평균을 추정할 경우, 다음 평균 표시로 해당 변수와 기준이 될 범주값을 지정 할 수 있다.



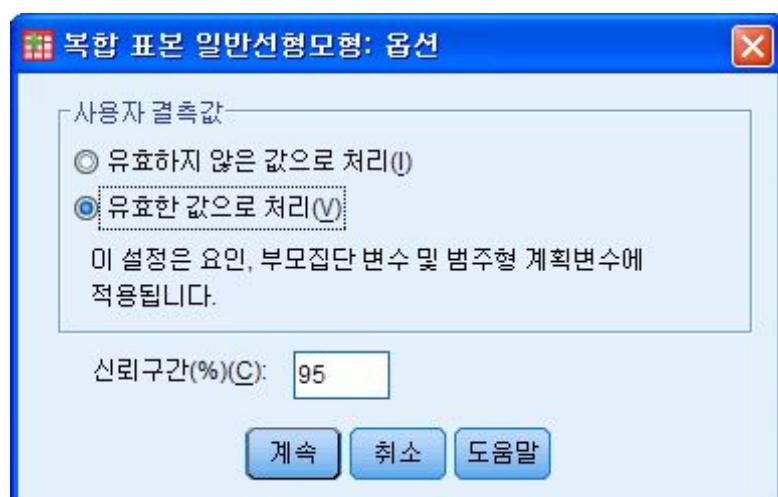
5 저장

분석 DB에 예측값과 잔차를 변수로 저장하기와 모형 내보내기를 할 수 있다. 예측값을 이용하여 다른 분석을 하거나 잔차분석을 할 경우와 데이터 및 모형을 내보낼 경우 해당부분을 선택 할 수 있다. 기본값으로 지정되어 있지 않으면 연구자가 지정해야 한다.



6 옵션

분석에 사용되는 변수의 결측값의 사용 여부를 선택할 수 있다. 국민건강영양조사 자료의 경우, 사용자결측값은 **유효한 값으로 처리**를 선택한다.



복합표본 일반선형모형 예제

- 2011년 국민건강영양조사 자료를 이용하여 공복시간 8시간 이상을 만족하는 사람에서 성별, 연령, 흡연여부, **BMI** 그룹에 따른 공복혈당의 평균 차이 검정을 복합표본 일반선형모형으로 실시한다.

- 분석변수 요약

- 기존 변수 : 성별(sex), 연령(age)
- 생성 변수 : 부모집단(GRP), 공복혈당(fst_glu), 흡연여부(SM), BMI그룹 (BMI_group1)

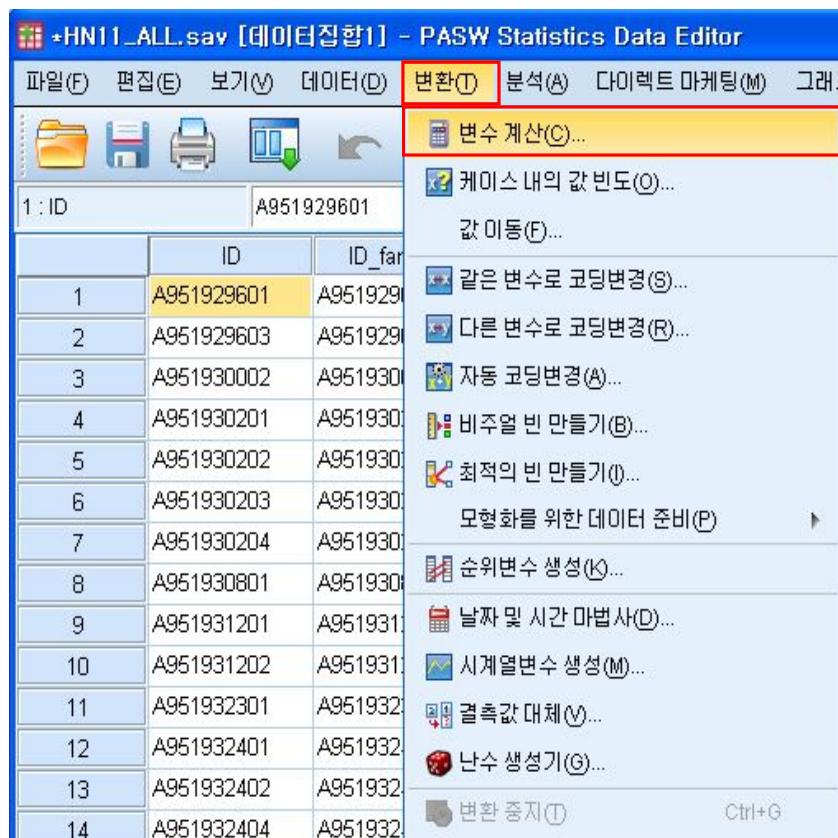
구분	변수명	변수설명	내용
분석대상	GRP	부모집단 (30세이상, 흡연여부, BMI그룹,공복혈당이 결측이 아닌 대상자)	1. 관심집단
종속변수	fst_glu	공복시간이 8시간이상인 경우의 공복혈당	연속형
독립변수	SEX	성별	1. 남자 2. 여자
	SM	흡연여부	1. 흡연 2. 과거흡연,비흡연
	BMI_group1	BMI그룹	1. 저체중 2. 비만 3. 정상
	age	연령	연속형

○ 분석에 앞서, 생성변수를 만든 후 복합표본설계 일반선형모형 분석을 실시한다.

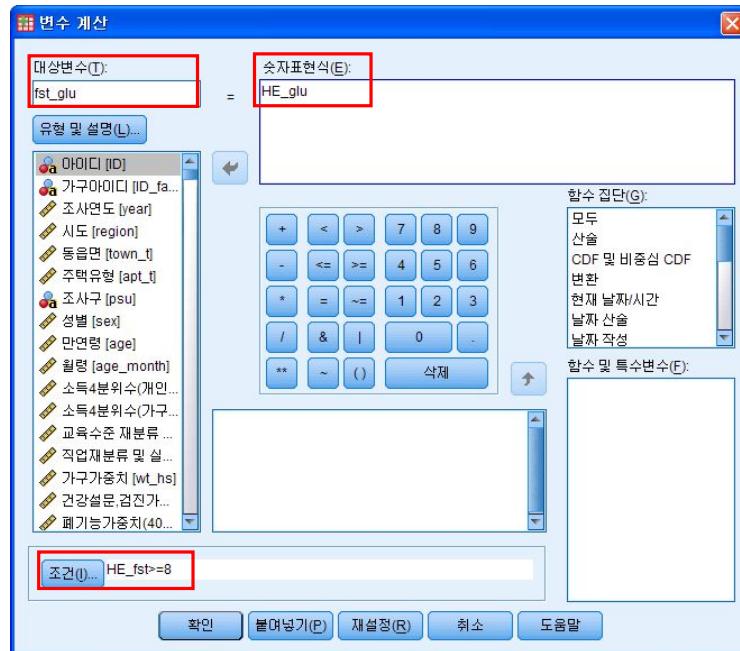
1 공복시간 8시간이상 공복혈당 변수 생성

생성변수	설명	
fst_glu	지표설명	공복시간 8시간 이상인 경우의 공복혈당
	변수유형	연속형
	조건식	IF (HE_fst>=8) fst_glu=HE_glu. 사용변수 (he_fst: 공복시간, he_glu: 공복혈당)

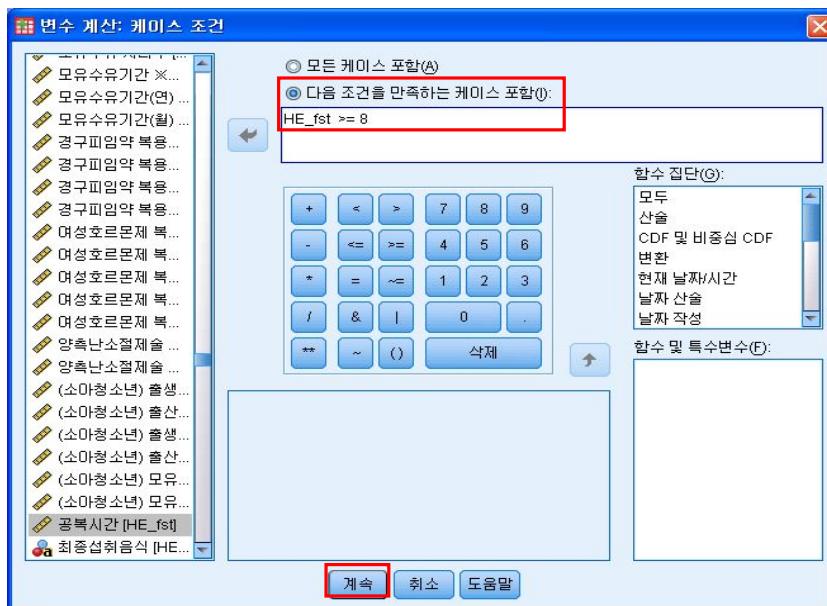
- 메뉴 창의 **변환** ▶ **변수계산** 을 이용하여 공복혈당 변수 생성



- 대상변수 : 신규 생성될 변수명(fst_glu) 입력
- 숫자표현식 : 신규 생성될 변수에 입력되는 값(HE_glu) 지정



- 조건 : 신규 생성될 변수값이 조건에 의해 선택될 때 변수 계산 창 하단의 **조건** 을 클릭하면 아래의 화면이 나타난다. 변수생성 조건($he_fst \geq 8$)을 입력한 후 **계속** 을 클릭하면, 윗 화면의 조건 옆에 입력된 조건이 나타난다.

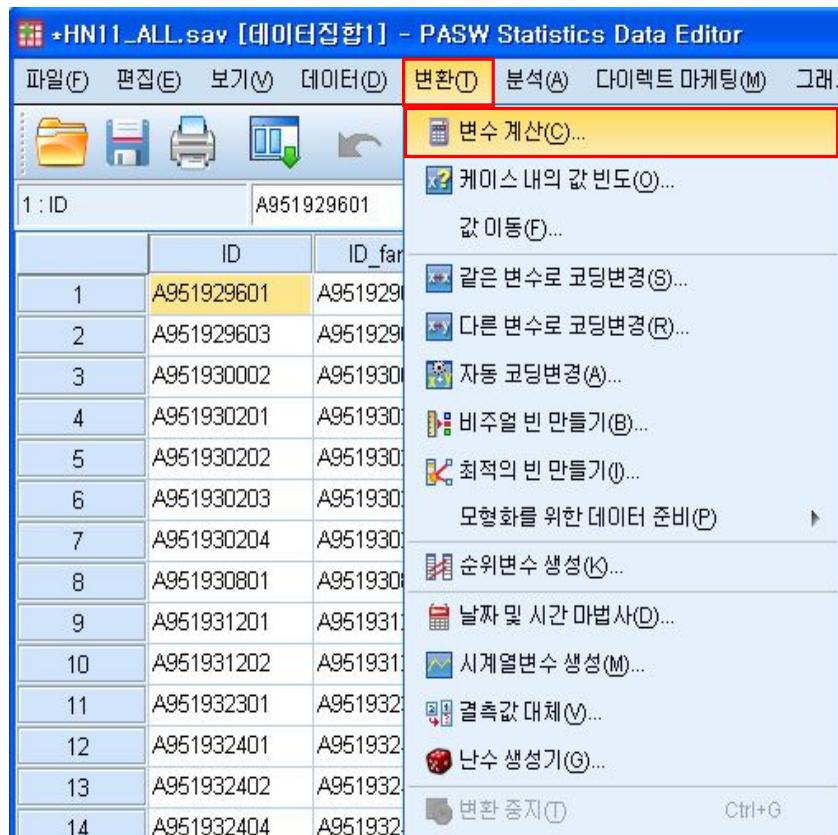


- 대상변수, 숫자표현식, 조건을 지정한 후, 확인을 클릭하면 공복혈당 변수가 생성된다.

2 흡연여부 변수 생성

생성변수	설명	
SM	지표설명	<현재흡연> 평생 흡연량이 5갑(100개비) 이상이며 현재 피우거나 가끔 피움 <과거흡연, 비흡연> 평생 흡연량이 5갑(100개비) 이상인 경우 과거에는 피웠으나 현재 피우지 않거나, 5갑(100개비)미만인 경우
	변수유형	범주형 1.현재흡연 2.과거흡연,비흡연
	조건식	IF ((BS1_1=2) & (BS3_1=1 or BS3_1=2)) SM=1. IF (BS1_1=1 or (BS1_1=2 & BS3_1=3)) SM=2. 사용변수 (BS1_1:평생흡연여부 , BS3_1:현재흡연여부)

- 메뉴 창의 **변환** → **변수계산** 을 이용하여 흡연여부 변수 생성



- 아래와 같이 대상변수, 숫자표현식, 조건을 입력 후 확인을 선택한다.

※ 입력과정은 공복혈당 변수 생성 참조(p90~91).

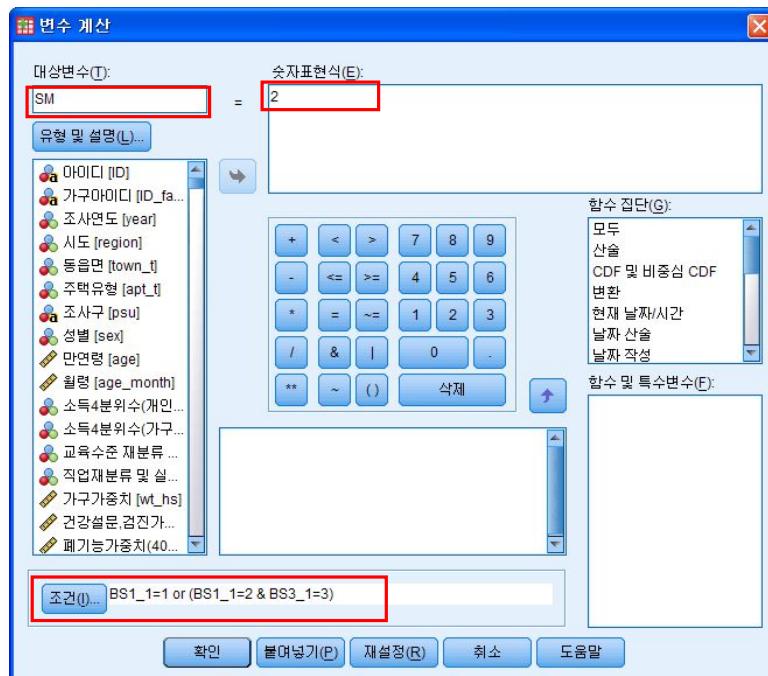
- 현재흡연(SM=1)인 경우

대상변수에 sm, **숫자표현식**에 1, **조건**에 $(BS1_1=2) \& (BS3_1=1 \text{ or } BS3_1=2)$ 입력.



- 과거흡연,비흡연(SM=2)인 경우

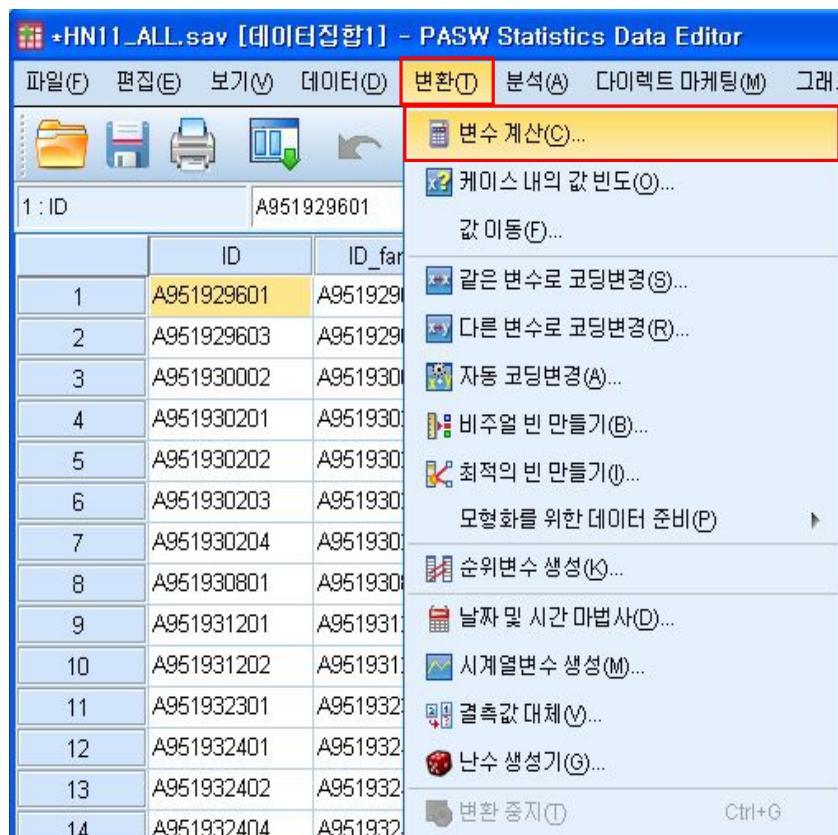
대상변수에 sm, **숫자표현식**에 2, **조건**에 $BS1_1=1 \text{ or } (BS1_1=2 \& BS3_1=3)$ 입력.



3 체질량지수(BMI) 그룹 변수 생성

생성변수	설명
BMI_group1	<p>지표설명</p> <p><저체중> 체질량지수(BMI)가 18.5미만</p> <p><정상> 체질량지수(BMI)가 18.5이상 25미만</p> <p><비만> 체질량지수(BMI)가 25이상</p>
	<p>변수유형</p> <p>범주형</p> <p>1. 저체중 2. 비만 3. 정상</p>
	<p>조건식</p> <p>IF (HE_BMI > 0 and HE_BMI < 18.5) BMI_group1=1. IF (HE_BMI >= 25) BMI_group1=2. IF (HE_BMI >= 18.5 and HE_BMI < 25) BMI_group1=3.</p> <p>사용변수 (HE_BMI:체질량지수)</p>

- 메뉴 창의 **변환** ▶ **변수계산** 을 이용하여 BMI_group1 변수 생성



- 아래와 같이 대상변수, 숫자표현식, 조건을 입력 후 확인을 선택한다.
- ※ 입력과정은 공복혈당 변수 생성 참조(p90~91).
- ※ BMI그룹(BMI_group1) 변수는 조건식을 변경하여 저체중(BMI_group1=1), 비만(BMI_group1=2), 정상(BMI_group1=3)인 경우를 반복하여 수행한다.
- BMI그룹(BMI_group1=1) : 저체중

대상변수에 BMI_group1, **숫자표현식**에 1, **조건**에 HE_BMI>0 and HE_BMI<18.5 입력.



- BMI그룹(BMI_group1=2) : 비만

대상변수에 BMI_group1, **숫자표현식**에 2, **조건**에 HE_BMI >= 25 입력.



- BMI그룹(BMI_group1=3) : 정상

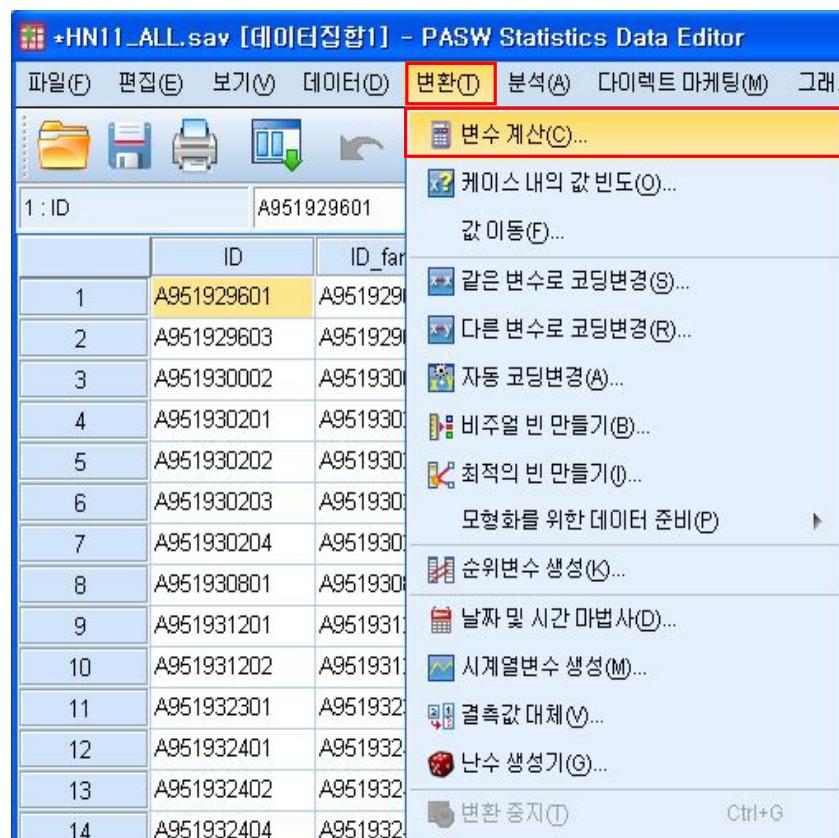
대상변수에 BMI_group1, **숫자표현식**에 3, **조건**에 $\text{HE_BMI} \geq 18.5 \text{ AND } \text{HE_BMI} < 25$ 입력



4 부모집단 변수 생성

생성변수	설명	
GRP	지표설명	연령은 30세 이상, 흡연여부, BMI그룹, 공복혈당이 결측이 아닌 집단
	변수유형	범주형 1.부모집단
	조건식	IF (age>=30 & (1<=BMI_group1<=3) & (1<=SM<=2) & fst_glu>=0) GRP=1. 사용변수 (age:연령, SM:흡연여부, BMI_group1:BMI그룹, fst_glu:공복혈당)

- 메뉴 창의 **변환** ▶ **변수계산** 을 이용하여 부모집단 변수 생성



- 아래와 같이 대상변수, 숫자표현식, 조건을 입력후 확인을 선택한다.

※ 입력과정은 공복혈당 변수 생성 참조(p90~91).

- 부모집단(GRP=1) : 관심집단

대상변수에 GRP, **숫자표현식**에 1, **조건**에 $age \geq 30 \ \& \ (1 \leq BMI_group1 \leq 3)$
 $\& \ (1 \leq SM \leq 2) \ \& \ fst_glu \geq 0$ 입력.



자료 분석

- **분석** ► **복합표본** ► **일반선형모형** 을 선택한 후, 분석 계획 파일을 선택하면 다음과 같은 복합표본 일반선형모형 대화상자가 나타난다.

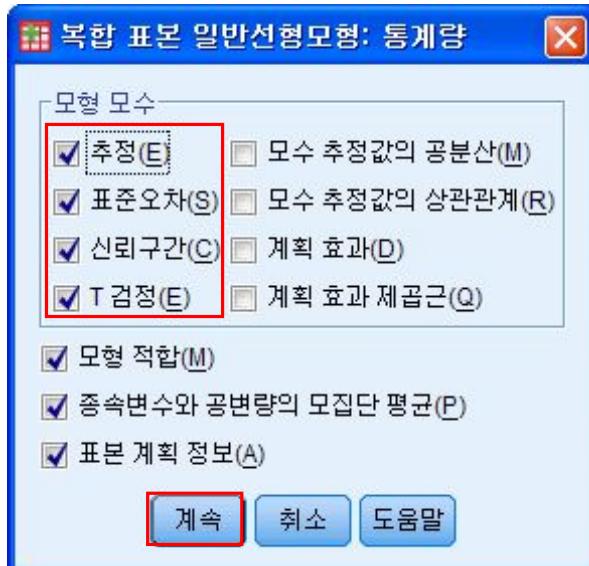
※ 분석계획파일의 복합표본설계 요소(가중치:wt_itvex, 층:Kstrata, 군집:PSU) 지정

- 분석에 사용될 변수를 해당 창에 지정한 후, 보고자 하는 옵션을 선택한다.

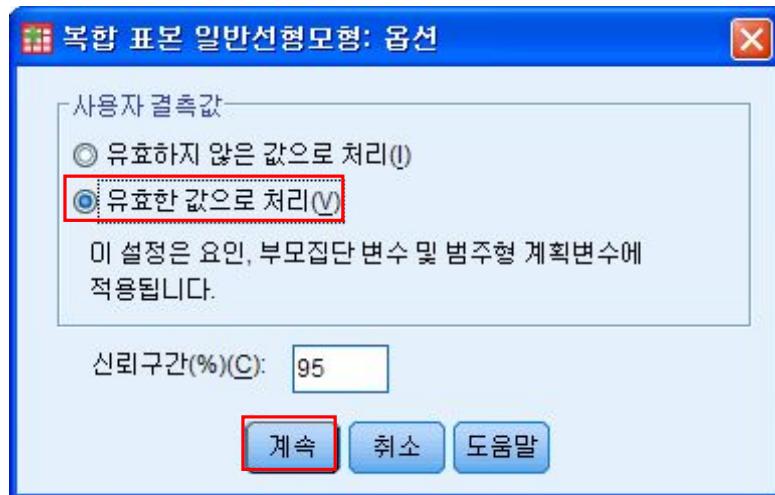
구분	지정	설명
종속변수	fst_glu	공복시간이 8시간이상인 경우의 공복혈당
요인분석	SEX	성별
	SM	흡연여부
	BMI_group1	BMI그룹
공변량	AGE	연령
부모집단	GRP	부모집단 변수
	1	부모집단의 분석범주



- 통계량은 **추정**, **표준오차**, **신뢰구간**, **T검정** 을 선택한 후 **계속** 을 클릭한다.



- 사용자 결측값은 **유효한 값으로 처리** 를 선택한 후 **계속** 을 클릭한다.



- 원하는 옵션을 선택 한 후, 최종적으로 복합표본 일반선형모형창에서 확인을 클릭하면 다음과 같은 결과가 산출된다.

분석 결과

모형 요약^a

R 제곱	.053
------	------

부모집단: GRP = 1.00

a. 모형: fst_glu
= (절편) + SM +
BMI_group1 +
sex + age

모형 효과 검정^a

소스	df1	df2	Wald F	유의확률
(정확 모델)	5.000	163.000	34.029	.000
(절편)	1.000	167.000	1135.682	.000
SM	1.000	167.000	1.835	.177
BMI_group1	2.000	166.000	32.408	.000
sex	1.000	167.000	20.391	.000
age	1.000	167.000	62.693	.000

부모집단: GRP = 1.00

a. 모형: fst_glu = (절편) + SM + BMI_group1 + sex + age

모수 추정값^b

모수	추정값	표준오차	95% 신뢰구간		가설검정		
			하한	상한	t	자유도	유의확률
(절편)	75.302	2.273	70.814	79.790	33.123	167.000	.000
현재흡연	[SM=1.00]	1.578	1.165	-.722	3.878	1.355	167.000
과거,비흡연	[SM=2.00]	.000 ^a177
저체중	[BMI_group1=1.00]	-13.191	1.996	-17.132	-9.251	-6.609	167.000
정상	[BMI_group1=2.00]	4.703	1.216	2.302	7.105	3.866	167.000
비만	[BMI_group1=3.00]	.000 ^a
남자	[sex=1]	5.946	1.317	3.346	8.545	4.516	167.000
여자	[sex=2]	.000 ^a
나이	age	.364	.046	.273	.455	7.918	167.000

부모집단: GRP = 1.00

a. 중복된 모수이므로 0으로 설정됩니다.

b. 모형: fst_glu = (절편) + SM + BMI_group1 + sex + age

결과해석

흡연여부(SM)를 제외한 나머지 변수는 p-value<0.05로 유의하게 나타났다, 비만그룹 대비 저체중 그룹의 공복혈당은 13.191mg/dL 낮았으며, 여자 대비 남자의 공복혈당은 5.946mg/dL 높았고, 연령이 1세 증가함에 따라 공복혈당은 0.364mg/dL 높았다.

▶ 복합표본 일반선형모형 예제 명령문

*변수생성 : 공복혈당.

IF (HE_fst>=8) fst_glu=HE_glu.
EXECUTE.

*변수생성 : 흡연여부.

*#1 현재흡연자.

IF ((BS1_1=2) & (BS3_1=1 or BS3_1=2)) SM=1.
EXECUTE.

*#2 비흡연, 과거흡연자.

IF (BS1_1=1 or (BS1_1=2 & BS3_1=3)) SM=2.
EXECUTE.

*변수생성 : BMI 그룹.

*#1 저체중.

IF (HE_BMI > 0 and HE_BMI <18.5) BMI_group1=1.
EXECUTE.

*#2 비만.

IF (HE_BMI >= 25) BMI_group1=2.
EXECUTE.

*#3 정상.

IF (HE_BMI >= 18.5 and HE_BMI < 25) BMI_group1=3.
EXECUTE.

*변수생성 : 일반선형모형 부모집단.

IF (age>=30 & (1<=BMI_group1<=3) & (1<=SM<=2) & fst_glu>=0)
GRP=1.
EXECUTE.

* 복합 표본 일반선형모형.

```
CSGLM fst_glu BY SM BMI_group1 sex WITH age
/PLAN FILE='D:\제5기 2차년도(2011)\SPSS\HN11_all_cs.csaplan'
/DOMAIN VARIABLE=GRP(1)
/MODEL SM BMI_group1 sex age
/INTERCEPT INCLUDE=YES SHOW=YES
/STATISTICS PARAMETER SE CINTERVAL TTEST
/PRINT SUMMARY VARIABLEINFO SAMPLEINFO
/TEST TYPE=F PADJUST=LSD
/MISSING CLASSMISSING=INCLUDE
/CRITERIA CLEVEL=95.
```

※명령문 활용에 대한 상세내용은 p39 참조.

복합표본 로지스틱 회귀모형

잠깐!! 분석 전에는 자료분석 유의사항(p53~55)을 반드시 확인해주시기 바랍니다.

- 복합표본 일반선형모형 프로시저에서는 복합표본설계자료에서 종속변수가 범주형일 때 종속변수와 독립변수간의 관련성을 분석한다.
- 복합표본설계 로지스틱 회귀분석에서는 종속변수가 2개범주(이분형), 3개이상 범주 또는 순서형 범주에 대한 분석이 가능하다.

표. 복합표본 로지스틱 회귀분석 분석방법

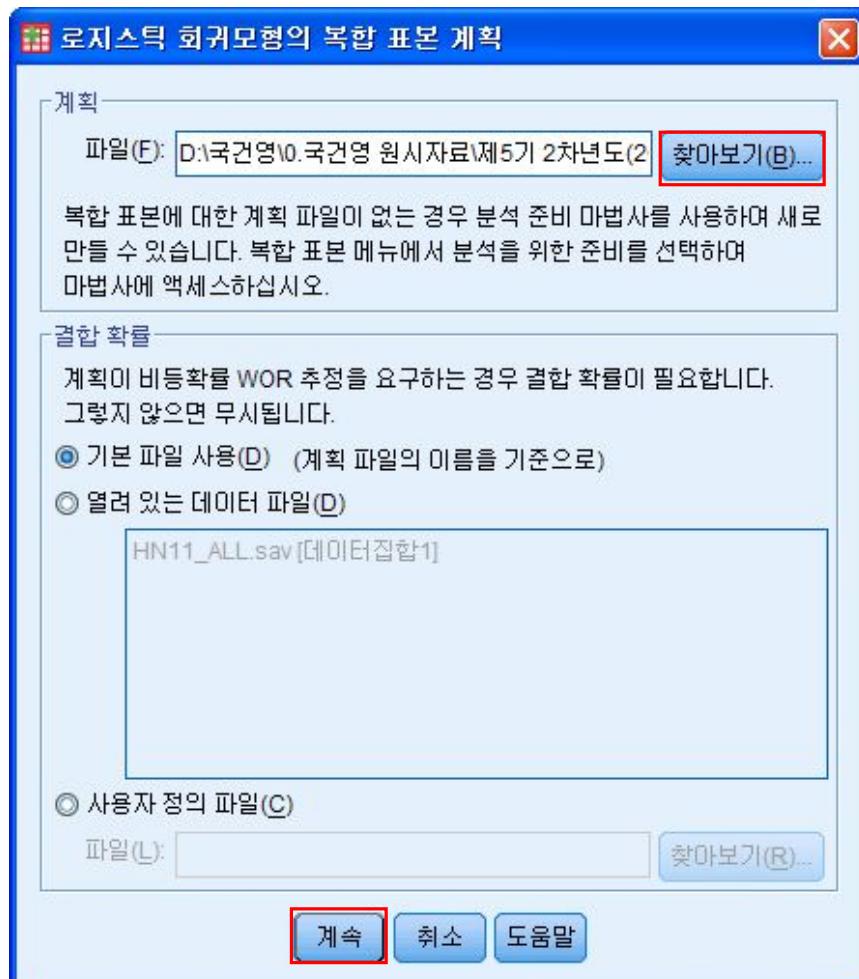
종속변수	독립변수	분석방법	SPSS 복합표본 프로시저
명목형 (이분형)	연속형 명목형 순서형	이분형 로지스틱회귀분석	복합표본 로지스틱 회귀모형
명목형 (3범주이상)		다중명목형 로지스틱회귀분석	

분석 준비

- 분석 ► 복합표본 ► 로지스틱 회귀모형 을 선택한다.



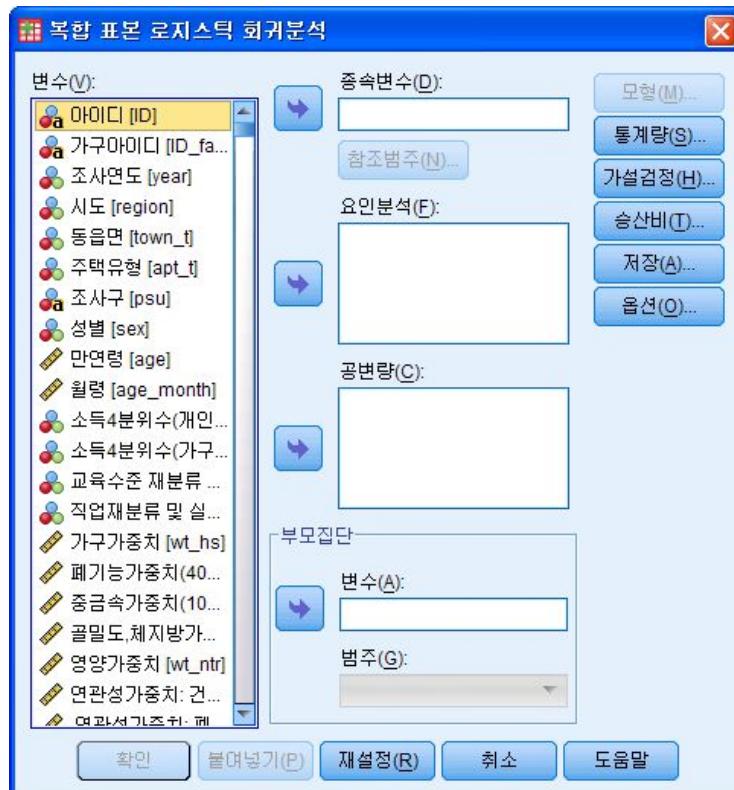
- 분석에 사용될 계획 파일을 **찾아보기**를 통해 선택한 후, **계속**을 클릭한다.



로지스틱 회귀분석 대화상자

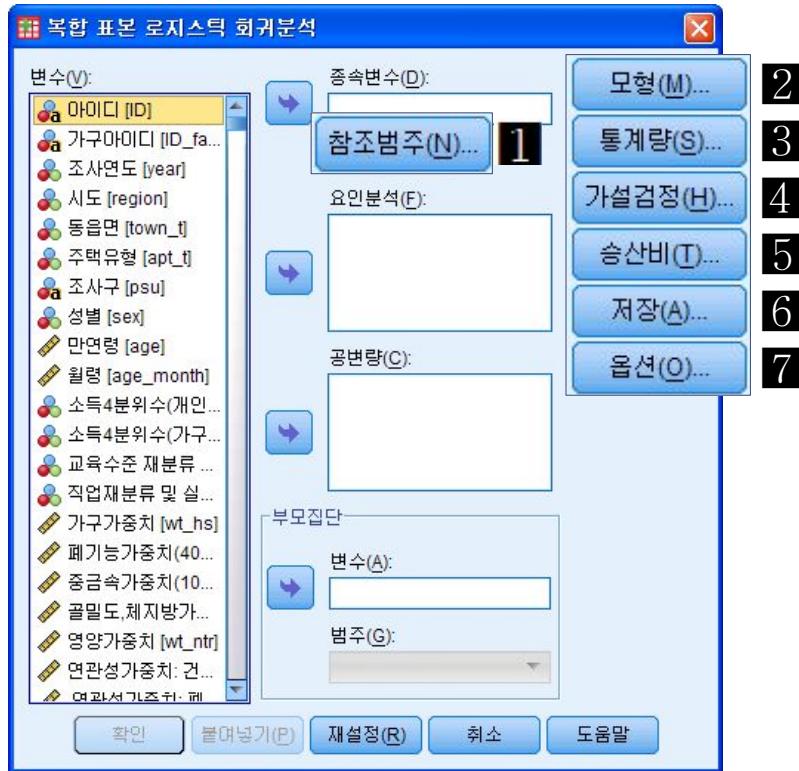
- 계획파일 선택을 완료하면 다음과 같은 복합표본 로지스틱 회귀분석 대화상자가 나타난다.

종속변수	독립변수의 영향을 받는 변수(범주형)	
요인분석	범주형 독립변수	
공변량	연속형 독립변수	
부모집단	변수	분석하고자 하는 집단을 구분하는 변수
	범주	분석하고자 하는 집단을 구분하는 변수 중 해당집단의 범주 값



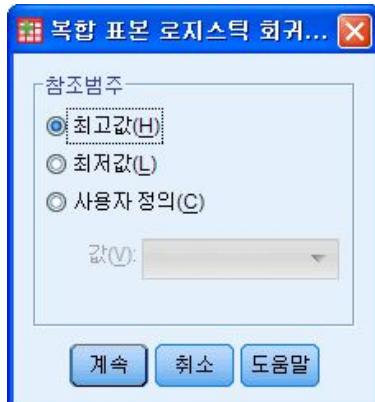
로지스틱 회귀분석 옵션 대화상자

- 복합표본 로지스틱 회귀분석 대화상자에서 분석변수를 선택한 다음, 그 외 필요한 옵션들을 옵션 대화상을 통해 선택한다.



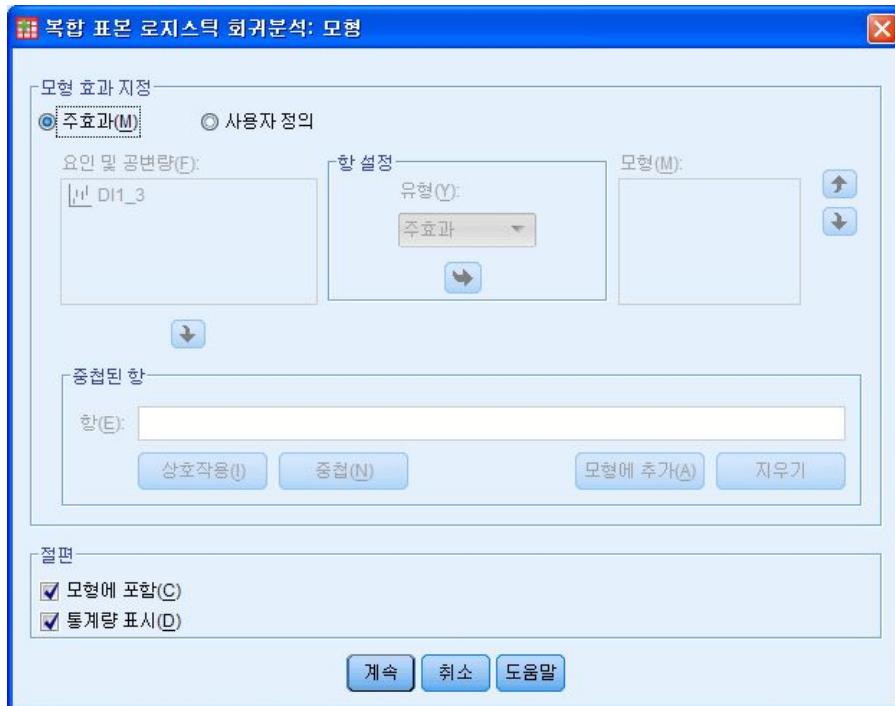
1 참조범주

종속변수로 선택한 변수에 대해 기준범주를 선택할 수 있다. 디폴트값은 최고값이다.



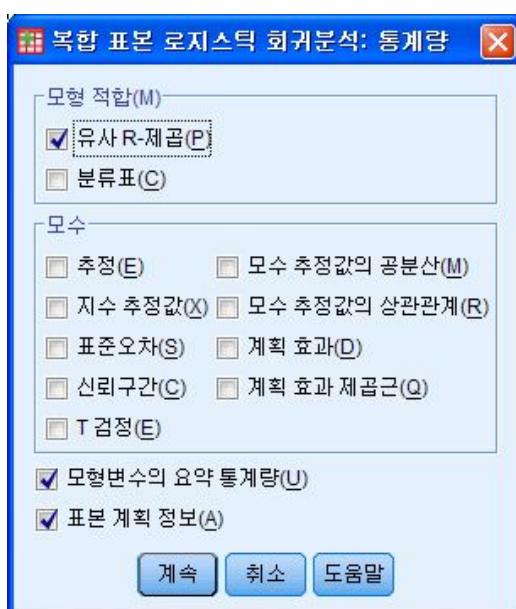
2 모형

선택모형을 만들고자 할 경우, 모형을 지정해 줄 수 있다. 디폴트값으로는 주대화 상자에서 지정한 요인변수와 공변량 변수를 이용하여 주효과 모형을 제시한다.



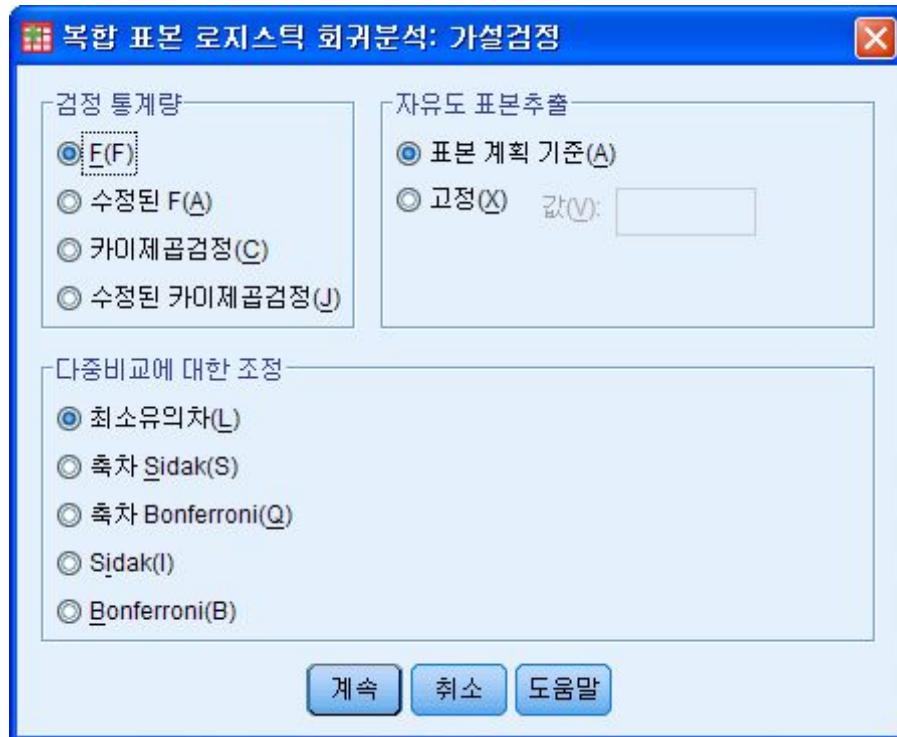
3 통계량

산출하고자 하는 하위 옵션을 선택 할 수 있다. 디폴트값은 유사 R-제곱, 모형변수의 요약통계량, 표본 계획정보이다.



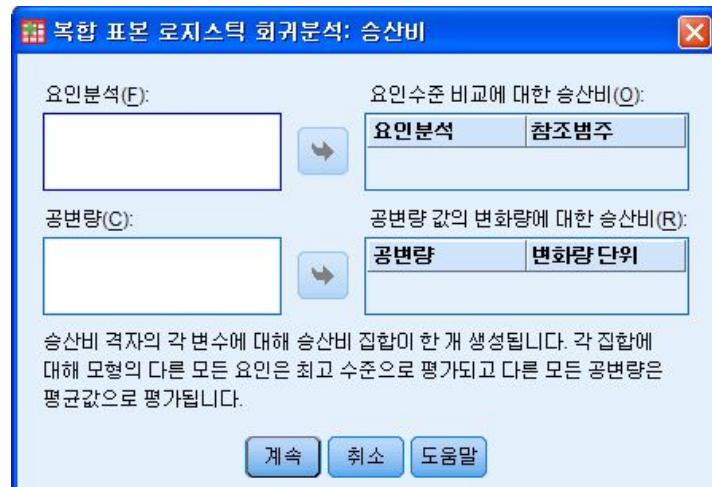
4 가설검정

가설을 검정하는데 사용되는 검정통계량, 다중비교에 대한 조정, 자유도 표본추출방법을 지정할 수 있다.



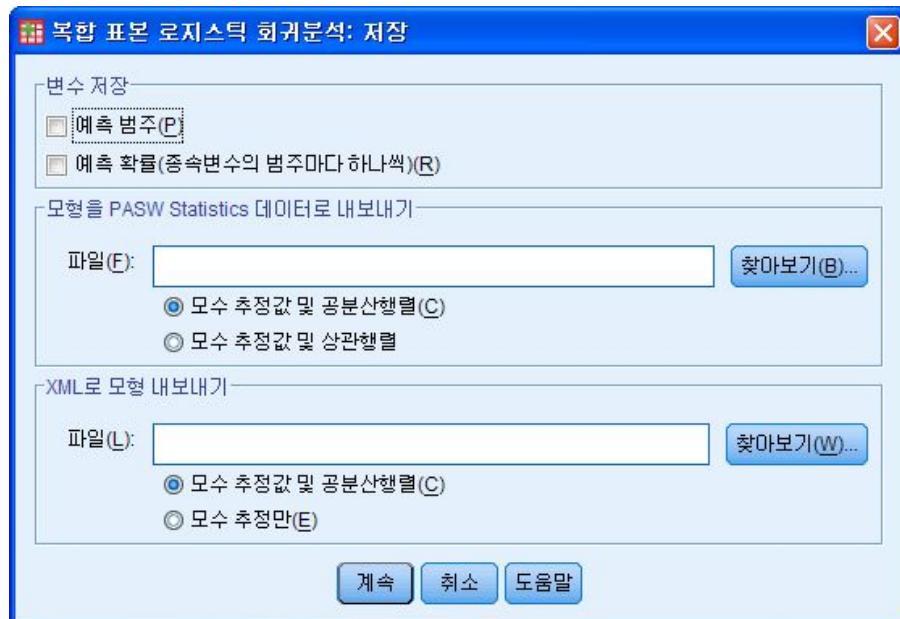
5 승산비

승산비에서는 요인과 공변량에 대한 승산비를 산출할 수 있다.



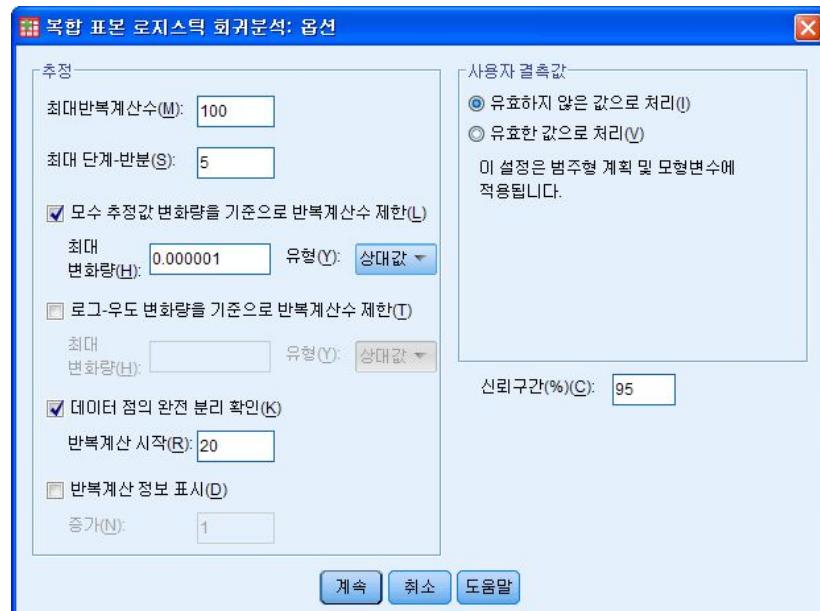
6 저장

분석 시 사용된 모형 예측값과 잔차를 새로운 변수들로 저장하여 내보내기를 할 수 있다. 내보내기를 원할 경우 출력형태를 선택 할 수 있다.



7 옵션

분석에 사용되는 변수의 결측값의 사용 여부를 선택할 수 있다. 국민건강영양조사의 경우 결측값은 유효하게 처리한다. 따라서 사용자 결측값을 유효한 값으로 처리를 선택한다.



복합표본 로지스틱 회귀모형 예제

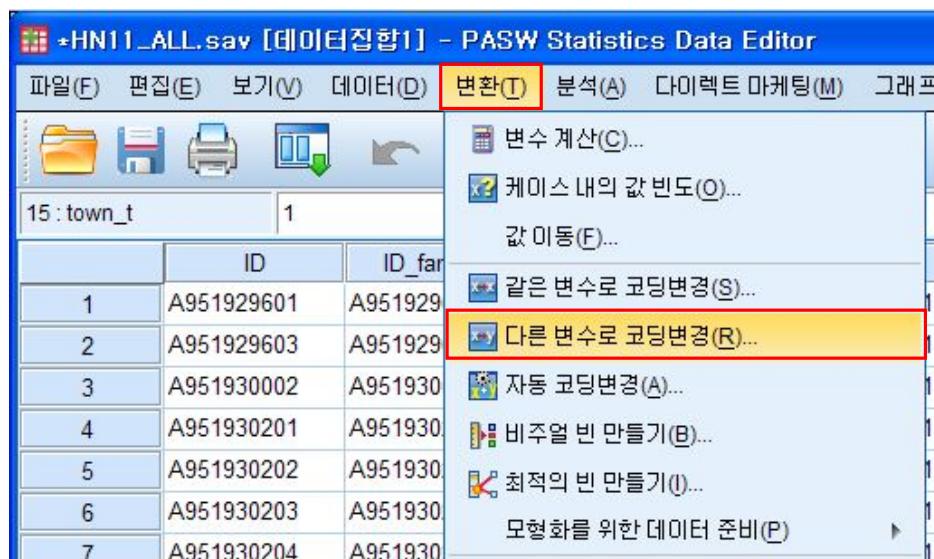
- 2011년 국민건강영양조사 자료를 이용하여 30세이상 남자의 AUDIT점수가 현재 흡연여부에 미치는 영향을 알아보고자 한다.
- 분석변수 요약
 - 기존 변수 : 현재흡연여부(sm_presnt), AUDIT 점수(AUDIT)
 - 생성 변수 : 부모집단(SEX_30)

구분	변수명	변수설명	내용
분석대상	SEX_30	부모집단(30세이상)	1. 남자 2. 여자
종속변수	sm_presnt	현재흡연여부	1. 현재흡연 0. 과거흡연,비흡연
독립변수	AUDIT	AUDIT 점수	연속형

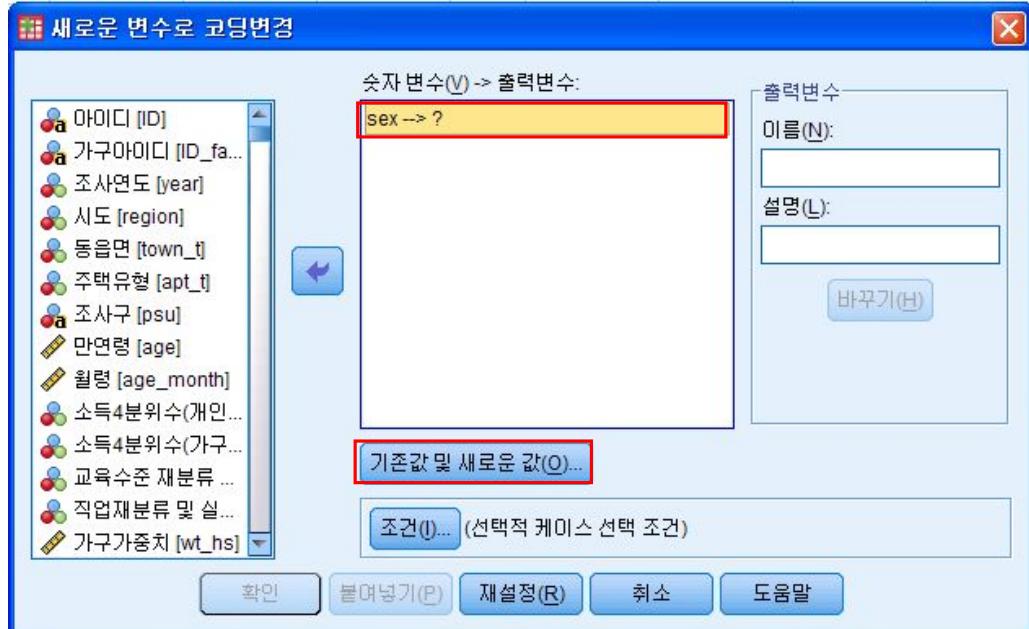
- 분석에 앞서, 생성변수를 만든 후 복합표본설계 로지스틱 회귀모형분석을 실시한다.

1 부모집단변수 생성

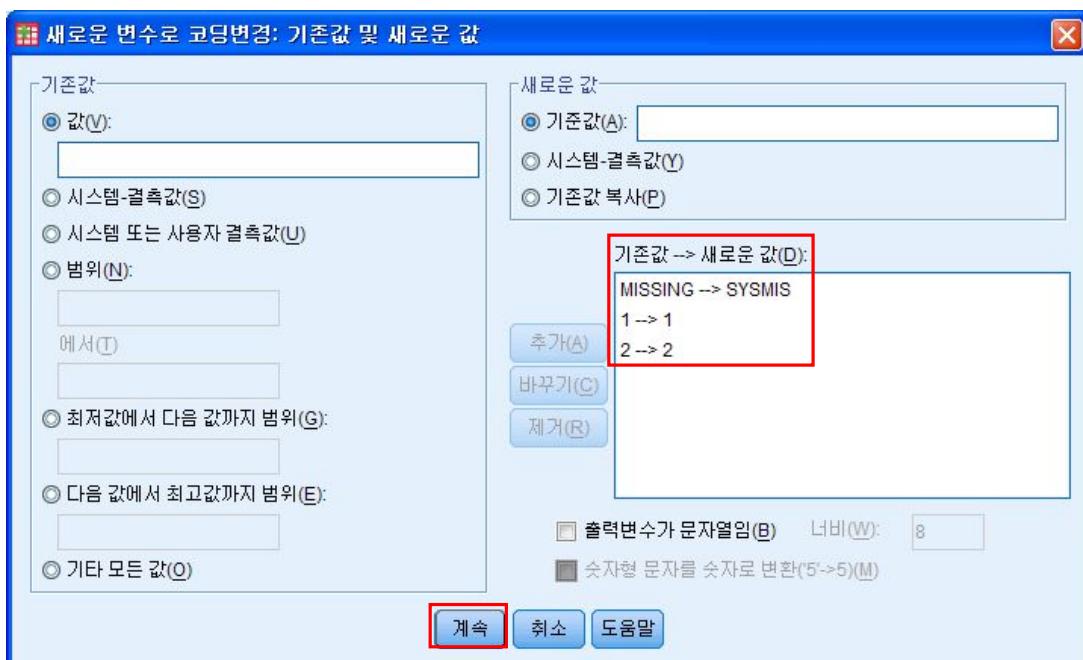
- 메뉴 창의 **변환** ▶ **다른 변수로 코딩변경** 을 선택한다.



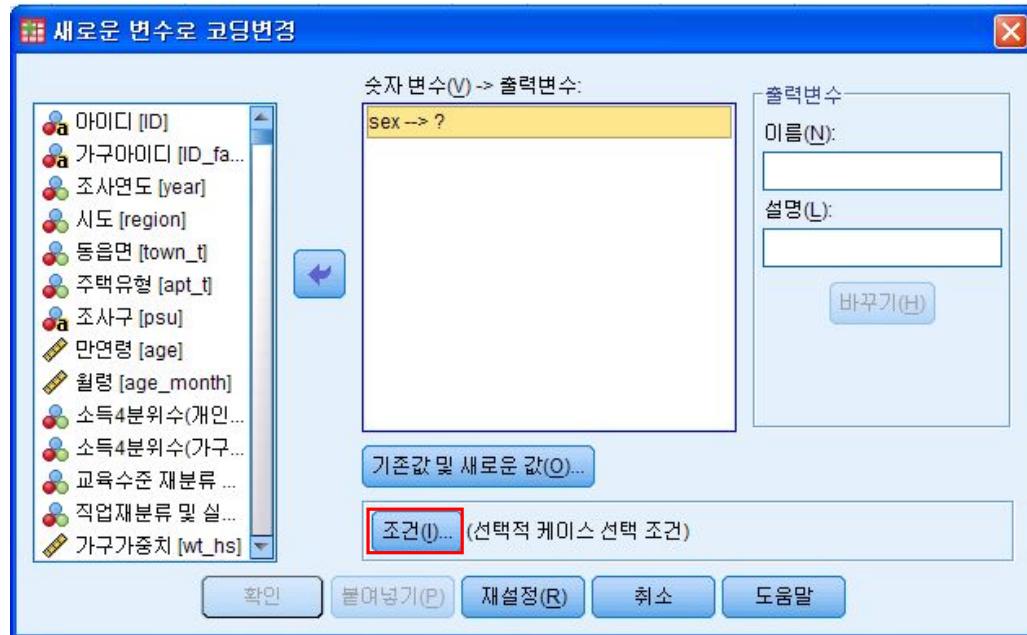
- 생성변수를 만들기 위해 기존변수인 sex를 선택하여 **숫자변수->출력변수** 창으로 옮겨놓은 후, **기존값 및 새로운 값** 을 클릭한다.



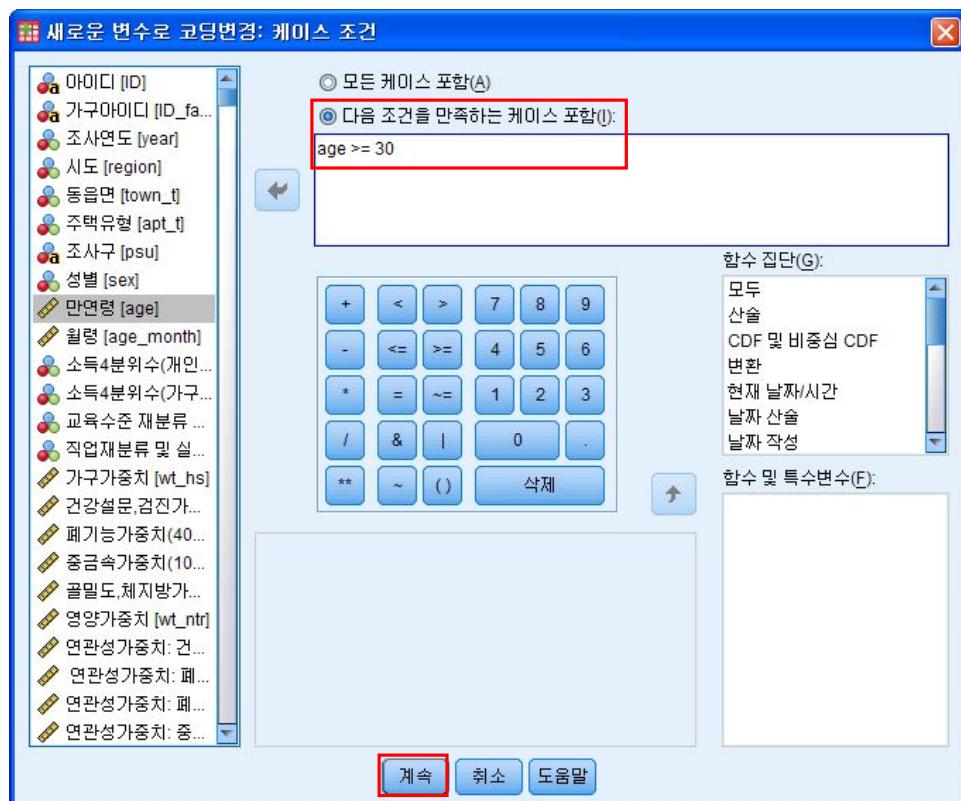
- 기존값(sex변수)를 이용하여 새로운 값을 지정 한 후 **계속** 을 클릭한다.
※기존값이 시스템 또는 사용자 결측값 인 경우 새로운 값 또한 시스템-결측값 으로 지정해 준다.



- 기존값 및 새로운 값을 설정한 후, 30세 이상에서만 부모집단 변수를 만들기 위해 **조건** 을 클릭한다.

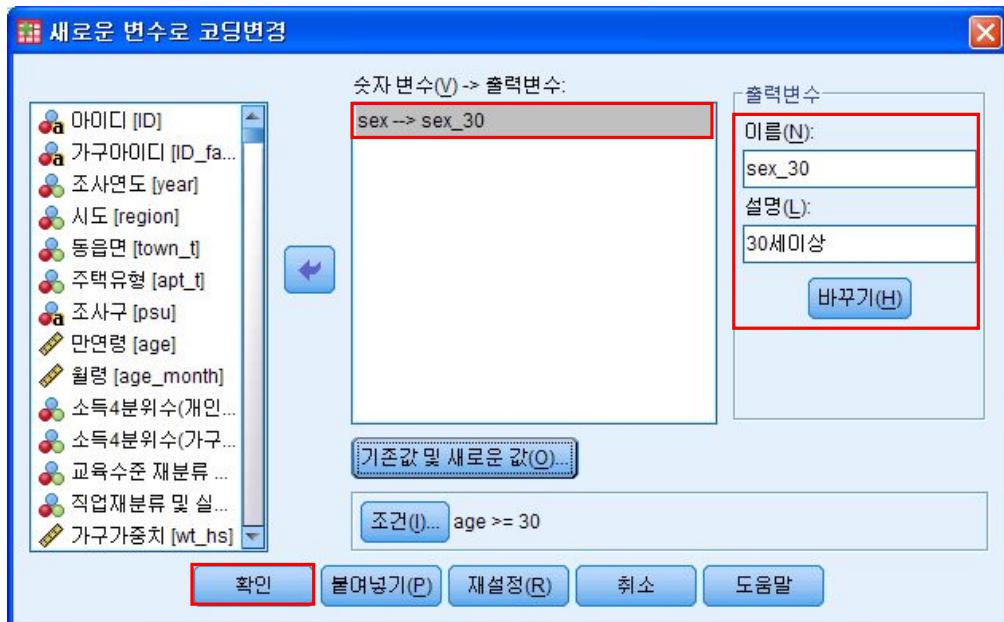


- **다음 조건을 만족하는 케이스 포함** 을 선택하면 아래창이 활성화 된다. 창이 활성화되면 30세 이상($age >= 30$)을 입력한 후 **계속** 을 클릭한다.



- 출력변수

: 이름에 생성할 변수명(sex_30)을 입력하고 **바꾸기**를 선택 한 후 **확인**을 클릭하면 새로운 변수가 생성된다. 설명은 연구자가 필요시에 작성한다.



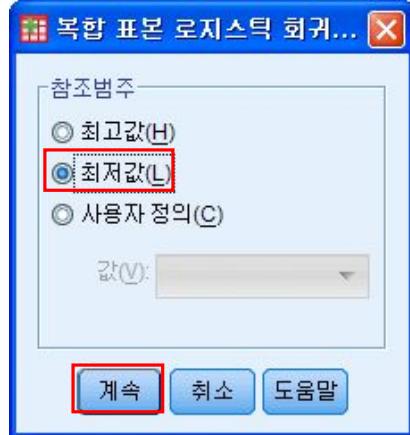
자료 분석

- **분석** ► **복합표본** ► **로지스틱 회귀모형** 을 선택한 후, 분석계획파일을 선택하면 다음과 같은 복합표본 로지스틱 회귀분석 대화상자가 나타난다.
※ 분석계획파일의 복합표본설계 요소(가중치:wt_itvex, 층:Kstrata, 군집:PSU) 지정
- 분석에 사용될 변수를 해당 창에 지정한 후, 보고자 하는 옵션을 선택한다.

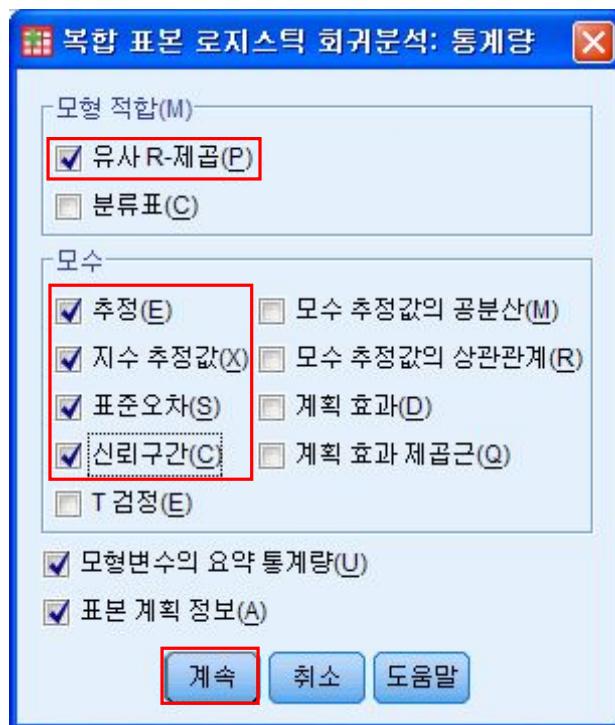
구분	지정	설명
종속변수	sm_presnt	현재흡연여부
공변량	AUDIT	AUDIT 점수
부모집단	변수	부모집단 변수
	범주	1 부모집단의 분석범주



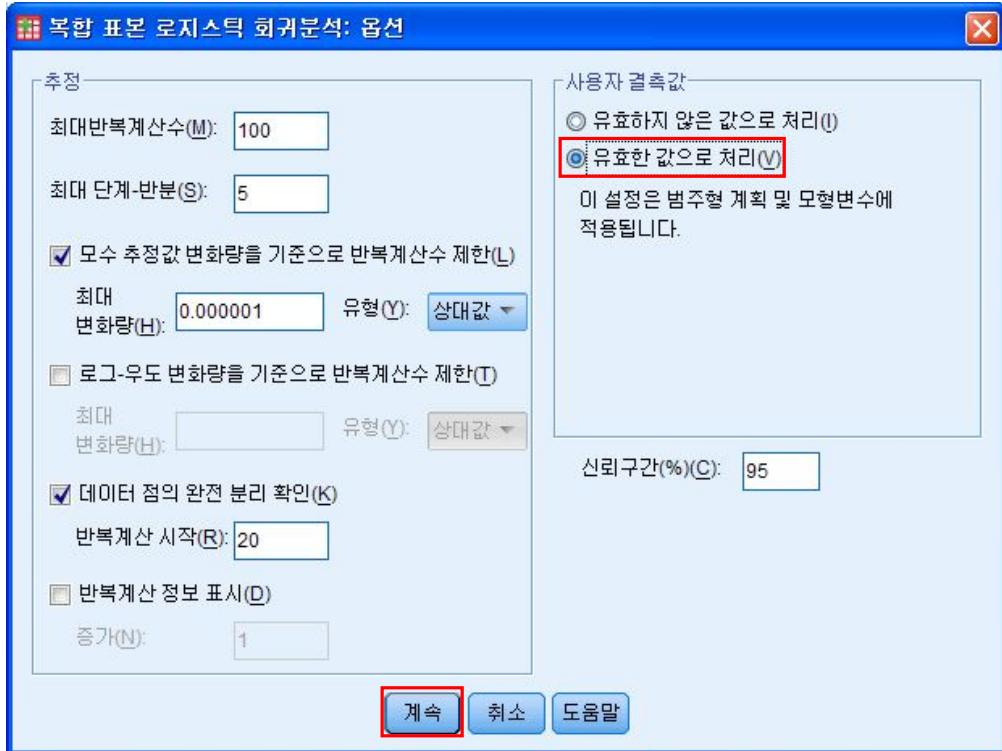
- 참조범주는 비흡연자(sm_presnt=0)를 기준으로 하기 위해 **최저값**을 선택한 후 **계속** 을 클릭한다.



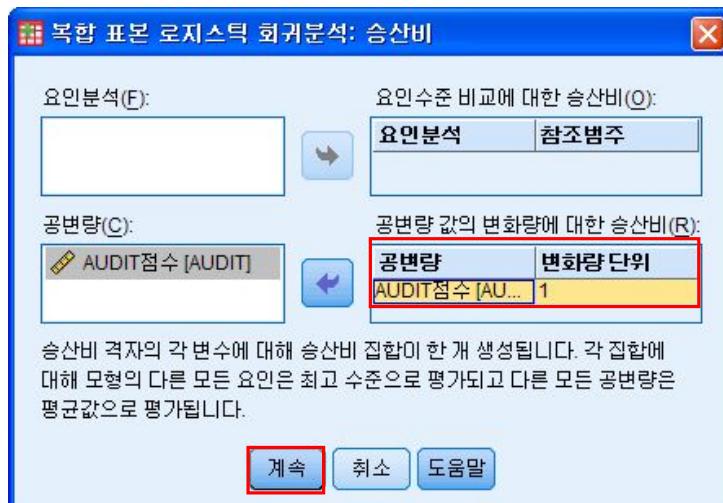
- 통계량은 **유사 R-제곱**, **추정**, **지수 추정값**, **표준오차**, **신뢰구간**, **모형변수의 요약 통계량**, **표본 계획 정보**를 선택하고, **계속** 을 클릭한다.



- 옵션은 사용자 결측값을 **유효한 값으로 처리**를 선택하고, **계속** 을 클릭한다.



- 승산비는 공변량 값의 변화량에 대한 승산비를 산출하기 위해 공변량인 **AUDIT점수**를 선택하고, **계속** 을 클릭한다.



- 원하는 옵션을 선택 한 후, 최종적으로 복합 표본 로지스틱 회귀분석창에서 확인을 클릭하면 다음과 같은 결과가 산출된다.

분석 결과

복합 표본: 로지스틱

모형 효과 검정

소스	df1	df2	Wald F	유의 확률
(정확 모델)	1.000	167.000	55.123	.000
(절편)	1.000	167.000	51.463	.000
AUDIT	1.000	167.000	55.123	.000

부모집단: 30세이상 = 1.30세이상 남자
증속변수: 현재흡연여부 (참조범주 = 0)
모형: (절편), AUDIT

모수 추정값

현재흡연여부	모수	B	표준오차	95% 신뢰구간		Exp(B)에 대한 95% 신뢰구간
				하한	상한	
1 (절편)	-.650	.091	.829	-.471	.522	.436
AUDIT	.057	.008	.042	.072	1.058	1.043

부모집단: 30세이상 = 1.30세이상 남자
증속변수: 현재흡연여부 (참조범주 = 0)
모형: (절편), AUDIT

승산비^a

변경 단위	현재흡연여부	승산비	95% 신뢰구간	
			하한	상한
AUDIT점수	1.000	1	1.058	1.043

부모집단: 30세이상 = 1.30세이상 남자
증속변수: 현재흡연여부 (참조범주 = 0)
모형: (절편), AUDIT

a. 계산에 사용된 요인 및 공변량은 다음 값에서 고정됨:
AUDIT점수=10.23

결과해석

30세 이상의 남자에서 $p\text{-value} < 0.05$ 로 AUDIT점수와 흡연은 서로 관련이 있었으며, AUDIT점수가 1점 증가함에 따라 오즈비(odds ratio)가 1.058배 증가하는 것으로 나타났다.

▶ 복합표본 로지스틱 회귀모형 예제 명령문

* 변수생성 : 로지스틱회귀모형 부모집단.
DO IF (age >= 30).
RECODE sex (MISSING=SYSMIS) (1=1) (2=2) INTO sex_30.
END IF.
VARIABLE LABELS sex_30 '30세이상'.
EXECUTE.

* 복합 표본 로지스틱 회귀분석.

CSLOGISTIC sm_presnt(LOW) WITH AUDIT
/PLAN FILE='D:\제5기 2차년도(2011)\SPSS\HN11_all_cs.csaplan'
/DOMAIN VARIABLE=sex_30(1)
/MODEL AUDIT
/INTERCEPT INCLUDE=YES SHOW=YES
/STATISTICS PARAMETER EXP SE CINTERVAL
/TEST TYPE=F PADJUST=LSD
/ODDSRATIOS COVARIATE=[AUDIT(1)]
/MISSING CLASSMISSING=INCLUDE
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE]
LCONVERGE=[0] CHKSEP=20 CILEVEL=95
/PRINT SUMMARY VARIABLEINFO SAMPLEINFO.

※ 명령문 활용에 대한 상세내용은 p39 참조.