

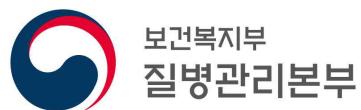


국가승인통계  
[승인번호 제117002호]

# 국민건강영양조사

## 원시자료 분석지침서(SAS)

최초 공개 : 2013년 5월  
최근 업데이트 및 수정 : 2019년 5월





## 이용자를 위하여

- 국민건강영양조사는 「국민건강증진법」 제16조에 근거하여 시행하는 국민의 건강행태, 만성질환 유병현황, 식품 및 영양섭취실태에 관한 법정조사이며, 「통계법」 제17조에 근거한 지정통계(승인번호 제117002호)입니다.
- 국민건강영양조사는 제1기(1998)부터 제3기(2005)까지 3년 주기로 실시하였고, 이후 연중 조사체계로 개편되어 제4기(2007~2009)부터 현재까지 매년 실시하고 있습니다. 본 조사는 질병관리본부 연구윤리심의위원회 승인을 받아 수행하였고, 조사익년도 12월까지 보도자료 배포, 통계집 발간, 원시자료 공개를 통해 조사결과를 공표하고 있습니다.

※ 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr/>) 자료실을 방문하시면 보도자료, 통계집, 조사표, 조사지침서, 질관리 보고서, 표본설계 보고서, 기타 연구용역 결과보고서 등의 자료를 다운로드 받으실 수 있습니다.

<질병관리본부 연구윤리심의위원회 승인번호>

	1차년도	2차년도	3차년도
제4기 (2007~2009)	2007-02CON-04-P	2008-04EXP-01-C	2009-01CON-03-2C
제5기 (2010~2012)	2010-02CON-21-C	2011-02CON-06-C	2012-01EXP-01-2C
제6기 (2013~2015)	2013-07CON-03-4C	2013-12EXP-03-5C	- (*)
제7기 (2016~2018)	- (*)	- (*)	

(\*) 국민건강영양조사는 생명윤리법 제2조제1호 및 동법 시행규칙 제2조제2항제1호에 따라 국가가 직접 공공복리를 위해 수행하는 연구에 해당하여 연구윤리심의위원회 심의를 받지 않고 수행(문의: 건강영양조사과 ☎ 043-719-7494)

- 질병관리본부는 개인정보보호법 및 통계법을 준수하여 조사자료에서 개인을 추정할 수 없도록 비식별 조치된 자료만을 제공하고 있으며, 해당 자료는 국민건강영양조사 홈페이지에서 다운로드 받으실 수 있습니다. 학술연구 등의 목적에 한해서는 제한공개자료(조사일, 조사지역)의 이용이 가능하며, 이를 위해 상세 연구계획서의 별도 제출, 정보보호서약서 제출, 단계별 심의·인정·인정·인정 등이 완료된 연구 건별로 질병관리본부의 엄격한 자료 관리 하에 해당 자료의 제한적 활용이 허용됩니다.

- 국민건강영양조사 제4기(2009–2009) 원시자료부터 다음과 같은 비식별조치가 새롭게 적용되었으며, 이는 2019년 3월을 기준으로 시행되어 해당 원시자료 분석결과와 기발간 보고서의 해당 수치 간 차이가 발생할 수 있으므로 유의하여 이용하시기 바랍니다.

비식별조치 신규 적용사항	해당 변수(변수명)
탑코딩(top-coding)	만나이(age), 가구총소득(ainc_1), 월평균 가구총소득(ainc), 이환 질환별 의사진단연령(□□□_ag), 음주시작연령(BD2)
바텀코딩(bottom-coding)	가구총소득(ainc_1), 월평균 가구총소득(ainc)
재범주화	가구원수(cfam), 가구세대구성코드(genertn), 기초생활수급 여부(allownc), 주택형태(live_t), 건강보험 종류(tins)

※상세 내용은 변수설명서 부분 참조

- 국민건강영양조사 원시자료 이용을 위해서는 반드시 『통계자료 이용자 준수사항 이행 서약서』 및 『보안서약서』의 서명, 제출이 필요합니다(규정 전문은 부록 참조). 해당 절차는 국민건강영양조사 홈페이지에서 원시자료 다운로드 시 진행됩니다.

※또한 국민건강영양조사 원시자료를 이용하여 작성된 논문, 포스터, 발표자료 등이 있는 경우에는 해당 파일을 국민건강영양조사 홈페이지에 등록하여 주시기 바랍니다.

〈원시자료 출처 표기 예시〉

구분		출처표기(Citation)
제3기 (2005)	국문	국민건강영양조사 제3기(2005), 질병관리본부
	영문	The Third Korea National Health and Nutrition Examination Survey (KNHANES III), 2005, Korea Centers for Disease Control and Prevention
제7기 2차년도 (2017)	국문	국민건강영양조사 제7기 2차년도(2017), 질병관리본부
	영문	The Seventh Korea National Health and Nutrition Examination Survey (KNHANES VII-2), 2017, Korea Centers for Disease Control and Prevention

- 국민건강영양조사 자료는 전수조사가 아닌 표본조사의 자료이므로 목표모집단인 대한민국 국민에 대한 결과로 확대하여 해석이 필요한 경우에는 복합표본설계 요소를 반영하여 분석하도록 권고하고 있습니다. 복합표본분석 관련 상세 내용은 『국민건강영양조사 및 청소년 건강행태조사 자료활용 워크숍 자료집』, 『국민건강영양조사 분석지침서(SPSS)』 등을 참고할 수 있고, 해당 자료는 국민건강영양조사 홈페이지에서 다운로드 가능합니다. 기타 원시자료 또는 자료분석 문의는 건강영양조사과 통계팀(☎ 043-719-7467~70)으로 주시기 바랍니다.

※자료활용 워크숍 자료집 및 분석지침서는 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr/>)에서 다운로드 가능합니다.

# 목 차

01	국민건강영양조사 자료 개요	1
02	복합표본 자료분석 개요	11
03	분석DB 준비단계	19
	• 자료결합	20
	- 자료의 가로결합	20
	- 자료의 세로결합	24
	• 식품섭취조사 개인단위 분석DB 생성	27
	• 통합기증치 산출	32
	• 기타	38
	- 라이브러리 생성	38
	- 데이터셋 활용하기(SET문, KEEP문)	40
04	복합표본 자료분석	43
	• 개요	44
	• 자료분석 시 유의사항	46
	• 복합표본 빈도분석	47
	• 복합표본 평균분석	52
	• 복합표본 회귀분석	56
	• 복합표본 로지스틱 회귀분석	67
05	연령표준화	83
	부록	89
	• 예제 변수 생성 및 산출	90
	• SAS 이용시 참고사항	95
	참고문헌	97



01

---

## 국민건강영양조사 자료 개요

## □ 국민건강영양조사 원시자료 DB 구성

### ○ 원시자료 다운로드

- 제1기(1998)~제7기 1,2차년도(2016–2017) 국민건강영양조사 원시자료는 국민건강영양조사 홈페이지(<http://knhanes.cdc.go.kr/>)에서 SPSS, SAS DB형태로 다운로드 가능하다.

### < 제1기(1998) ~ 제7기 2차년도(2017) 원시자료 구성 >

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

지침서	원시자료 DB (SAS, SPSS)
· 원시자료 이용지침서 및 분석지침서 · 식품섭취조사 식품코드 및 음식코드	<ul style="list-style-type: none"> <li>- HN<sub>YR</sub>_ALL : 건강설문, 검진, 영양</li> <li>- HN<sub>YR</sub>_IJMT : 손상 및 의료이용(건강설문)</li> <li>- HN<sub>YR</sub>_OE : 구강검사(검진)</li> <li>- HN<sub>YR</sub>_DXA : 골밀도 및 체지방검사(검진)</li> <li>- HN<sub>YR</sub>_ENT : 이비인후검사(검진)</li> <li>- HN<sub>YR</sub>_EYE : 안검사(검진)</li> <li>- HN<sub>YR</sub>_24RC : 식품섭취조사(영양)</li> <li>- HN<sub>YR</sub>_FFQ : 식품섭취빈도조사(영양)</li> </ul>
<b>제6기(2013–2015)</b>	
· 원시자료 이용지침서 및 분석지침서 · 식품섭취조사 식품코드 및 음식코드	<ul style="list-style-type: none"> <li>- HN<sub>YR</sub>_ALL : 건강설문, 검진, 영양</li> <li>- HN<sub>YR</sub>_IJMT : 손상 및 의료이용(건강설문)</li> <li>- HN<sub>YR</sub>_OE : 구강검사(검진)</li> <li>- HN<sub>YR</sub>_24RC : 식품섭취조사(영양)</li> <li>- HN<sub>YR</sub>_FFQ : 식품섭취빈도조사(영양)</li> </ul>
<b>제7기(2016–2017)</b>	
· 원시자료 이용지침서 및 분석지침서 · 식품섭취조사 식품코드 및 음식코드	<ul style="list-style-type: none"> <li>· <b>1차년도(2016)</b> <ul style="list-style-type: none"> <li>- HN<sub>YR</sub>_ALL : 건강설문, 검진, 영양</li> <li>- HN<sub>YR</sub>_24RC : 식품섭취조사(영양)</li> <li>- HN<sub>YR</sub>_FFQ : 식품섭취빈도조사(영양)</li> </ul> </li> <li>· <b>2차년도(2017)</b> <ul style="list-style-type: none"> <li>- HN<sub>YR</sub>_ALL : 건강설문, 검진, 영양</li> <li>- HN<sub>YR</sub>_24RC : 식품섭취조사(영양)</li> </ul> </li> </ul>

\*<sub>YR</sub>은 해당연도 두자리수 숫자

\*2008년 7월 안검사(HN<sub>YR</sub>\_EYE), 이비인후검사(HN<sub>YR</sub>\_ENT), 골밀도 및 체지방검사(HN<sub>YR</sub>\_DXA) 신규도입

\*2012년 식품섭취빈도조사 도입

## ○ 분석변수 및 범위 결정

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

## ○ 분석변수의 상세내용 확인

- 「원시자료 이용지침서」에서 변수명, 해당변수 포함된 DB명, 코딩내용 확인한다.
- 대부분의 주요변수는 기본DB(HN<sub>YR</sub>\_ALL, <sub>YR</sub>은 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, 총, 집락)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))

자료의 개수 (변수의 개수)	제1기 (1998)	제2기 (2001)	제3기 (2005)	제4기		제5기			제6기			제7기		
				(2007)	(2008)	(2009)	(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	
<b>▪ 건강설문조사</b>														
가구조사 (HNYR_ALL)	39,060 (47)	37,769 (47)	34,145 (57)	4,594 (49)	9,744 (60)	10,533 (62)	8,958 (40)	8,518 (48)	8,058 (40)	8,018 (56)	7,550 (41)	7,380 (43)	8,150 (41)	8,126 (49)
건강면접조사 (HNYR_ALL)	39,060 (78)	37,769 (816)	34,145 (355)	4,148 (771)	9,281 (771)	10,051 (774)	8,388 (449)	7,895 (441)	7,337 (429)	7,277 (330)	6,888 (369)	6,904 (411)	7,797 (461)	7,700 (506)
건강행태조사 (HNYR_ALL)	10,808 (111)	9,170 (146)	23,000 (162)	-	-	-	-	-	-	-	-	-	-	-
순상및의료이용 (HNYR_IJMT)	39,060 (216)	37,769 (548)	34,145 (1,510)	4,148 (830)	9,281 (1,233)	10,051 (874)	8,388 (609)	7,895 (697)	7,337 (425)	7,277 (256)	6,888 (324)	6,904 (248)	-	-
이환카드 (HNYR_DSPRAW)	39,060 (350)	37,769 (414)	-	-	-	-	-	-	-	-	-	-	-	-
<b>▪ 검진조사</b>														
검진조사 (HNYR_ALL)	9,771 (70)	9,702 (68)	7,597 (130)	4,246 (74)	9,307 (109)	10,078 (108)	8,473 (143)	8,055 (134)	7,713 (130)	7,571 (206)	7,167 (174)	6,976 (172)	7,803 (194)	7,707 (191)
구강검사 (HNYR_OE)	-	-	-	4,246 (335)	9,307 (346)	10,078 (348)	8,473 (341)	7,836 (308)	7,353 (314)	7,507 (314)	7,167 (314)	6,977 (314)	-	-
안검사 (HNYR_EYE)	-	-	-	-	4,846 (212)	9,760 (217)	8,141 (209)	7,791 (216)	7,444 (163)	-	-	-	-	-
이비인후검사 (HNYR_ENT)	-	-	-	-	4,592 (132)	10,065 (152)	8,313 (142)	7,887 (123)	7,421 (104)	-	-	-	-	-
골밀도검사 (HNYR_DXA)	-	-	-	-	3,583 (156)	7,920 (159)	7,043 (140)	2,757 (114)	-	-	-	-	-	-
<b>▪ 영양조사</b>														
영양조사 (HNYR_ALL)	11,267 (99)	10,000 (122)	9,004 (176)	4,099 (205)	8,641 (208)	9,397 (211)	8,027 (132)	7,715 (132)	7,024 (92)	7,242 (96)	6,803 (96)	6,630 (96)	7,049 (75)	7,169 (75)
식품섭취조사 (HNYR_24RC)	454,841 (36)	430,890 (42)	420,794 (45)	197 (48)	418,628 (51)	462,127 (51)	476,964 (72)	451,014 (72)	415,443 (55)	472,354 (85)	446,753 (85)	447,718 (100)	518,061 (102)	526,382 (100)
식품섭취빈도조사 (HNYR_FFQ)	-	-	-	-	-	-	-	-	-	3,984 (401)	4,152 (402)	3,810 (402)	3,302 (402)	3,371 (402)

※ DB명에서 YR은 해당연도 2자리 숫자 (예를 들어 제7기 2차년도(2017) 식품섭취조사 DB는 HN17\_24RC)

※ HNYR\_ALL DB에서 조사부문별, 조사영역별 변수의 개수는 공통변수(시도, 동읍면, 주택유형, 조사구, 소득분위수, 성별, 만나이, 문산추정 총, 교육수준, 직업분류, 가중치 등)를 가구조사 변수의 개수로만 포함하여 계산

※ 제4기(2007~2009) 이후 손상및의료이용 자료(DB명 HNYR\_IJMT)는 연도별 최대이용건수에 따라 변수의 개수 변동, 제7기 1차년도(2016)부터 손상및의료이용 자료는 손상이용건수별 자료만 존재하여 HNYR\_ALL DB에 포함

(최대이용건수)	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17
- 손 상:	4회	10회	5회	3회	3회	4회	5회	6회	4회	4회	10회
- 입원이용:	12회	6회	8회	14회	13회	11회	6회	12회	20회	-	-
- 외래이용:	13회	22회	15회	14회	20회	26회	14회	14회	14회	-	-
- 약국이용:	7회	12회	8회	10회	12회	10회	10회	7회	-	-	-

## 〈 제7기(2016-2017) 조사항목 〉

조사부문	조사영역	대상연령	조사항목
건강설문	가구조사	만19세이상	성, 연령, 결혼상태, 가구원수, 세대유형, 가구소득, 건강보험 가입, 민간보험 가입, 출생시/현재 국적('16년), 치매진단('16년)
	교육	만1세이상	학력, 출업 여부
		만19세이상	부모 학력
	경제활동	만15세이상	경제활동여부, 미취업사유, 취업형태, 종사상지위, 근로시간형태, 직업, 직장내 지위
		만19세이상	최장 직업
	이환	만1세이상	최근 2주간 이환, 만성질환별(성인 27개, 소아청소년 10개) 이환
	의료이용	만1세이상	(병의원) 미충족의료, 외래이용, 입원이용
	건강검진	만19세이상	건강검진 수진 여부, 건강검진 수진 종류('16년), 암검진 수진 여부, 암검진 수진 종류('16년)
	예방접종	만1세이상	인플루엔자 예방접종 여부
	활동제한	만1세이상	활동제한여부, 활동제한이유, (월간) 외병경험/일수, (월간) 결근결석경험/일수
	삶의 질	만19세이상	주관적 건강인지, 건강관련 삶의 질 측정도구(EuroQol-5 Dimension, EQ-5D: 운동능력, 자기관리, 일상활동, 통증불편, 불안우울)
	손상	만1세이상	손상 경험발생기전내외처, 손상으로 인한 외병결석
	흡연	만6~11세	가정내공장소 실내 간접흡연
		만12~18세	평생흡연, 현재흡연, 흡연량, 가정내공장소 실내 간접흡연
		만19세이상	평생흡연, 현재흡연, 과거흡연, 처음흡연시작 연령, 흡연량, 금연시도, 금연계획, 금연기간, 금연방법, 가정내공장소 실내 간접흡연, 니코틴의존('16년), 금연캠페인경험('17년), 전자담배배종류별 사용, 전자담배 사용이유('16년), 니코틴대체용품 사용
	음주	만12세이상	평생음주, 음주시작 연령, 음주빈도, 음주량, 폭음빈도, 간접폐해(만19세이상)
	신체활동	만12~18세	하루 60분 신체활동 실천
		만12세이상	국제신체활동설문(Global Physical Activity Questionnaire, GPAQ: 일/여가 고강도 운동강도 신체활동, 이동시 활동, 앓아서 보내는 시간)('17년, 만19세이상), 걷기('17년, 만19세이상), 앓아서 보내는 시간, 근력운동
		만65세이상	가속도계 측정('16년, 만19~64세)
	정신건강	만12세이상	주중주말 잠든 시간, 일어난 시각, 스트레스인지, 우울감경험('16년, 만12~18세), 자살생각('16년, 만12~18세) 폐획면도, 정신문제 상담경험, 우울증선별도구(Patient Health Questionnaire-9, PHQ-9) ('16년, 만19세이상)
영양	안전의식	만1~5세	자동차 보호장구 착용
		만1~11세	자동차 앞좌석 이용('17년), 자전거 헬멧 착용
		만12세이상	동승차량 자동차 안전벨트 착용, 자전거요토바이 헬멧 착용('16년)
		만19세이상	운전시 안전벨트 착용('17년), 자동차요토바이자전거 음주운전 경험, 음주운전 차량 동승
	비만 및 체중조절	만6세이상	주관적 체형 인지, 체중 변화조절
	여성건강	만10세이상	현재 월경 여부, 조경 연령
		만15세이상	임신 경험, 출산 경험
		만19세이상	모유수유 경험여부수면간, 폐경 연령, 경구피임약 복용 경험
	구강건강	만1세이상	칫솔질 여부, 치아 손상, 구강검진, 치과 이용, (치과) 미충족의료
		만12세이상	구강용품 사용
		만19세이상	저작 불편, 발음 불편
식생활조사	식생활조사	만1세이상	끼니별 식사 빈도, 끼니별 돌반식사 여부대상, 외식 빈도, 식이보충제 복용 경험
		초등학생이상	영양표시 인지여용행여부 심상형목, 영양교육 및 상담 경험
		만1~3세	출생체중, 수유 방법여간, 이유식, 일반우유, 영아기 식이보충제 섭취 정보
	식품안정성조사	식생활관리자	가구의 식품안정성 확보
	식품섭취빈도조사	만19~64세	112개 음식항목의 섭취 빈도와 1회 섭취량('16년)
	식품섭취조사	만1세이상	조사1일전 섭취 음식(식이보충제 포함)의 종류 및 섭취량
		조리자	조사1일전 섭취 음식 중 직접 조리한 음식의 식품재료 및 재료량, 음식 총량

## 6 제1장 분석 준비단계

조사부문	조사영역	대상연령	조사항목
검진	신체계측	만1세이상	신장, 체중, 허리둘레
	혈압 및 맥박	만10세이상	수축기혈압, 이완기혈압, 맥박수
	혈액검사	만10세이상	(혈당) 공복혈당, 당화혈색소 (지질) 총콜레스테롤, 중성지방, HDL콜레스테롤, LDL콜레스테롤 (신장) 혈중요소질소, 크레아티닌 (간염) B형간염표면항원, ALT, AST, C형간염항체 (빈혈) 해모글로빈, 헤마토크리트, 적혈구 수, 백혈구 수 (기타) 고감도C반응단백, 요산
		만10세이상	(중금속) 납, 수은, 카드뮴, 니켈('17년)(일부 표본) <sup>1)</sup> (비타민) 비타민A, E, 엽산(일부 표본)
	소변검사	만6세이상	크레아티닌, 코티닌(전수), NNAL*(1/2표본) * 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol
		만10세이상	단백, 당, 잠혈, 비중, 산도, 유로빌리노겐, 케톤, 빌리루빈, 아질산염, 나트륨, 요칼륨
	구강검사	만1세이상	지아 상태, 치료 필요, 보철물 상태별요, 치주조직 상태, 지아반점도, 주관적 구강건강 상태, 치아통증 경험, 교정치료 경험
	폐기능검사	만40세이상	노력성 폐활량, 1초간 노력성 호기량
	안검사	('16년)만5~18세	시력 및 굴절검사
		('17년)만40세이상	시력 및 굴절검사, 안저촬영검사, 빛간섭단층촬영검사, 생체계측검사, 안압검사, 시야검사
	악력검사	만10세이상	악력
	이비인후검사	만40세이상	소음노출설문
	가족력	만10세이상	만성질환 가족력(부모형제)

1) 제7기 1차년도(2016) 1/2 표본, 제7기 2차년도(2017) 1/3 표본만 조사

## □ 국민건강영양조사 보고서

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

### < 국민건강영양조사 보고서 목록 >

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

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

구분	진행보고서 <sup>1)</sup>	결과보고서 <sup>2)</sup>	표본설계보고서 <sup>3)</sup>
<b>제5기 3차년도(2012)</b>			
	<ul style="list-style-type: none"> <li>- 국민건강영양조사 제5기 3차년도(2012) 골관절염검사 전문조사원(방사선사) 교육 및 질관리</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 구강검사 전문조사원(공중보건치과의) 교육 및 질관리</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 안(눈)검사 전문조사원(안과전공의) 교육 및 질관리</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 영양조사 질 관리 및 식품섭취조사 자료처리시스템 개발을 위한 기초 자료 생산</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 이비인후(귀, 코, 목)검사 전문조사원(이비인후과전공의) 교육 및 질관리</li> <li>- 국민건강영양조사 및 한국인유전체역학조사사업 진단의학검사 수행기관 질 관리</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 진단의학검사(임상검사)</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 혈압측정 질관리</li> <li>- 국민건강영양조사 제5기 3차년도(2012) 호흡기계검사 전문조사원(간호사, 방사선사) 교육 및 질관리</li> </ul>		
<b>제6기 (2013–2015)</b>			
	<ul style="list-style-type: none"> <li>· 2013 국민건강통계: 제6기 1차년도 (2013)</li> <li>· 2014 국민건강통계: 제6기 2차년도 (2014)</li> <li>· 2015 국민건강통계: 제6기 3차년도 (2015)</li> </ul>	<ul style="list-style-type: none"> <li>· 제4기(2007–2009) 가중치 산출 및 제5기(2010–2012) 표본설계</li> <li>· 국민건강영양조사 제6기 표본설계 및 제5기 가중치 산출</li> </ul>	
<b>질관리보고서</b>			
<b>제6기 1차년도(2013)</b>			
	<ul style="list-style-type: none"> <li>- 국민건강영양조사 제6기(2013–2015) 진단의학검사(임상검사)</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 구강검사 전문조사원(공중보건치과의) 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 골관절염 전문조사원(방사선사) 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 영양조사 질관리 및 전자조사표 도입을 위한 사전조사</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 진단의학검사 수행기관 질관리</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 호흡기계검사 전문조사원(간호사, 방사선사) 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 1차년도(2013) 혈압측정 질관리</li> </ul>		
<b>제6기 2차년도(2014)</b>			
	<ul style="list-style-type: none"> <li>- 국민건강영양조사 제6기(2013–2015) 진단의학검사(임상검사)</li> <li>- 국민건강영양조사 제6기 2차년도(2014) 구강검사 전문조사원(공중보건치과의) 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 2차년도(2014) 진단의학검사 수행기관 질관리</li> <li>- 국민건강영양조사 제6기 2차년도(2014) 폐기능검사 전문조사원 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 2차년도(2014) 혈압측정 질관리</li> </ul>		
<b>제6기 3차년도(2015)</b>			
	<ul style="list-style-type: none"> <li>- 국민건강영양조사 제6기(2013–2015) 진단의학검사(임상검사)</li> <li>- 국민건강영양조사 제6기 3차년도(2015) 구강검사 전문조사원 교육 및 질관리</li> <li>- 국민건강영양조사 제6기 3차년도(2015) 진단의학검사 질관리</li> <li>- 국민건강영양조사 제6기 3차년도(2015) 혈압측정 질관리</li> </ul>		
<b>제7기 (2016–2017)</b>			
	<ul style="list-style-type: none"> <li>· 2016 국민건강통계: 제7기 1차년도 (2016)</li> <li>· 2017 국민건강통계: 제7기 2차년도 (2017)</li> </ul>		<ul style="list-style-type: none"> <li>· 국민건강영양조사 제7기(2016–2018) 표본설계 및 소지역 추정방안 개선</li> </ul>
<b>질관리보고서</b>			
<b>제7기 1차년도(2016)</b>			
	<ul style="list-style-type: none"> <li>- 국민건강영양조사 제7기(2016–2018) 진단의학검사(임상검사)</li> <li>- 국가 흡연폐해 비아오모니터링(2016–2018)</li> <li>- 국민건강영양조사 제7기 1차년도(2016) 구강검사 전문조사원 교육 및 질관리</li> <li>- 국민건강영양조사 제7기 1차년도(2016) 진단의학검사 질관리</li> <li>- 국민건강영양조사 제7기 1차년도(2016) 폐기능검사 질관리</li> <li>- 국민건강영양조사 제7기 1차년도(2016) 혈압측정 질관리</li> <li>- 국민건강영양조사 제7기 1차년도(2016) 결핵감염률조사 지원 및 질관리</li> <li>- 국민건강영양조사 영양심층 분석 연구</li> <li>- 한국형 건강관련 삶의질 측정도구(HINT-8)의 가치평가연구</li> </ul>		

## 10 제1장 분석 준비단계

구분	진행보고서 <sup>1)</sup>	결과보고서 <sup>2)</sup>	표본설계보고서 <sup>3)</sup>
	<b>제7기 2차년도(2017)</b> <ul style="list-style-type: none"><li>- 국민건강영양조사 제7기(2016~2018) 진단의학검사(임상검사)</li><li>- 국가 흡연폐해 바이오모니터링(2016~2018)</li><li>- 국민건강영양조사 제7기 2차년도(2017) 구강검사 전문조사원 교육 및 질관리</li><li>- 국민건강영양조사 제7기 2차년도(2017) 진단의학검사 질관리</li><li>- 국민건강영양조사 제7기 2차년도(2017) 폐기능검사 질관리</li><li>- 국민건강영양조사 제7기 2차년도(2017) 혈압측정 질관리</li><li>- 국민건강영양조사 제7기 2차년도(2017) 안검사 지원 및 질관리</li></ul>		

1)조사 수행체계, 자료수집방법, 조사항목, 지침서 등

2)결과표(당해년도 현황 및 연도별 추이, 결과 해석, 지표정의, 대상자 현황, 분석방법 등)

3)모집단, 표본추출률, 표본추출방법, 모수추정, 사후관리 등

02

---

복합표본

자료분석 개요

## □ 국민건강영양조사 복합표본설계 개요

### ○ 표본설계

- 국민건강영양조사는 표본의 대표성 및 추정의 정확성 향상을 위해 복합표본설계방법인 다단계층화집락획분법으로 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	11~12월 (2개월)	11~12월 (2개월)
표본크기	<ul style="list-style-type: none"> <li>▪ 건강면접조사: 200조사구, 12,000가구</li> <li>▪ 건강행태/검진/영양조사: 200조사구, 4,000가구</li> </ul>	<ul style="list-style-type: none"> <li>▪ 건강면접조사: 600조사구, 12,000가구 (통합조사구 200, 비통합조사구 400)</li> <li>▪ 건강행태/검진/영양조사: 200통합조사구, 4,000가구</li> </ul>
추출틀	인구주택총조사(1995), 신축아파트목록(1997)	인구주택총조사(2000)
추출단위	조사구 → 가구	조사구 → 가구
총화변수	통합시(군), 동읍면, 주택유형	통합시도, 동읍면, 주택유형
구분	제3기(2005)	제4기(2007~2009)
조사기간	4~6월 (3개월)	2007년 7~12월(6개월), 2008~2009년 1~12월(12개월)
표본크기	<ul style="list-style-type: none"> <li>▪ 건강면접조사: 600조사구, 12,000가구 (통합조사구 200, 비통합조사구 400)</li> <li>▪ 건강행태/검진/영양조사: 200통합조사구, 4,000가구</li> </ul>	600조사구, 13,800가구 (연간 200조사구, 4,600가구) ※2007년은 반년조사로 100조사구, 2,300가구 완료
추출틀	인구주택총조사(2000), 신축아파트목록(2001)	인구주택총조사(2005)
추출단위	조사구 → 가구	동읍면 → 조사구 → 가구
총화변수	시도, 동읍면, 주택유형	시도, 동읍면, 주택유형(내재적층: 권역, 연령)
구분	제5기(2010~2012)	제6기(2013~2015)
조사기간	1~12월 (12개월)	1~12월 (12개월)
표본크기	총 576조사구, 11,520가구(연간 192조사구, 3,840가구)	총 576조사구, 11,520가구(연간 192조사구, 3,840가구)
추출틀	주민등록인구(2009), 아파트목록(2008)	인구주택총조사(2010)
추출단위	조사구 → 가구	조사구 → 가구
총화변수	시도, 동읍면, 주택유형 (내재적층: 일반- 성별, 연령; 아파트- 평당기객, 평균평수 등)	시도, 동읍면, 주택유형 (내재적층: 성별, 연령, 주거면적, 가구주 학력)
구분	제7기(2016~2018)	
조사기간	1~12월 (12개월)	
표본크기	총 576조사구, 13,248가구(연간 192조사구, 4,416가구)	
추출틀	인구주택총조사(2010), 공동주택공시가격(2011~14)	
추출단위	조사구 → 가구	
총화변수	시도(세종시 추가), 동읍면, 주택유형(일반/아파트/신축아파트) (내재적층: 성별, 연령, 주거면적 비율)	

### ○ 복합표본분석방법

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

### 〈 복합표본설계 추정식 〉

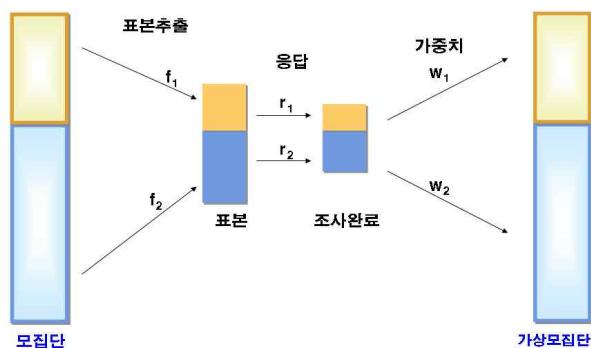
	단순임의설계	복합표본설계
평균 (비율) 추정	$\hat{Y} = \frac{\sum_{k=1}^n y_k}{n}$	$\hat{Y} = \frac{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \times y_{hij}}{\sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}}$
분산추정	$\hat{V}(\hat{Y}) = \frac{\sum_{k=1}^n (y_k - \hat{Y})^2}{n-1}$	$\hat{V}(\hat{Y})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

## □ 가중치

### ○ 가중치 산출방법

- 국민건강영양조사 가중치는 추정치가 우리나라 전체 모집단을 대표하도록 부여하는 확대승수로, 추출률, 응답률, 모집단 분포 등을 반영하여 산출한다.
- 국민건강영양조사 제3기(2005) 이후부터 가중치는 설계가중치 계산, 무응답률 조정, 사후보정, 극단가중치 처리를 거쳐 최종 산출한다.

### < 가중치의 구성 >



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

(예) 제5기 2차년도(2011): 모집단 조사구중 총 192개 표본조사구 추출, 표본조사구내 적절가구중 20개 표본가구 추출, 표본가구에서 부적절가구원(만1세미만, 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

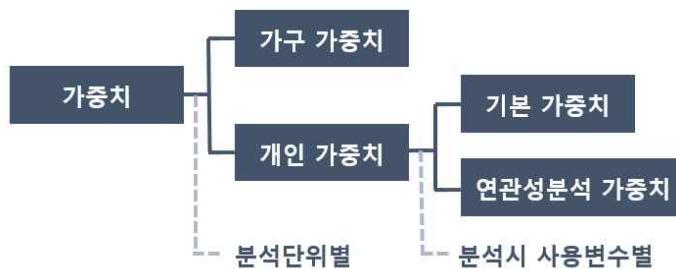
※ 사후보정 결과 가구가중치는 조사 참여가구의 가중치 합이 우리나라 전체 가구수와 같아지도록, 부문별 개인가중치 (건강설문·면진조사, 영양조사 등)는 조사 참여자수의 가중치 합이 우리나라 전체 인구수와 같아짐

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

- 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( $Q3 - Q1$ )

## ○ 가중치 종류

### < 국민건강영양조사 가중치 분류 >



- 참여가구별로 부여하는 가구가중치와 참여자에게 부여하는 개인가중치로 구분되며, 개인가중치는 기본 가중치와 연관성 가중치로 분류된다.
- 각 조사부문(또는 조사항목) 참여자에게 부여한 기본 가중치는 조사부문별(건강설문조사, 검진조사, 영양조사) 가중치와 조사항목별 대상자수 또는 참여자수 고려가 필요한 경우 조사항목별(중금속검사, 폐기능검사 등) 가중치로 구분된다.
  - 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 건강설문 조사 가중치(wt\_itv) 부여
  - 제5기(2010- 2012)~제7기 2차년도(2017) : 건강설문조사 및 검진조사 완료자가 거의 일치하여 건강 설문조사 또는 검진조사 참여자에 대해 건강설문-검진조사 가중치(wt\_itvex) 부여  
※동일한 조사부문 가중치라 하더라도 변수명 변경에 주의하여 분석
  - 중금속검사 : 검진조사 참여자 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 참여자수와 폐기능검사 참여자수 차이가 발생하여 별도의 폐기능검사 가중치(wt\_pft) 부여

- 안검사, 이비인후검사, 흉부X-선검사, 골밀도 및 체지방검사는 2008년 7월 신규도입, 골관절염검사는 2009년 8월 신규도입, 골밀도 및 체지방검사는 2011년 5월 완료 등, 검진항목이 연중 신규도입 또는 완료되어 전체 검진조사 참여자수와 차이가 발생하는 경우 별도 가중치 부여
- NNAL검사 : 만6세이상 검진조사 참여자 중 매년 약 3,000명을 선정하여 검사하므로 NNAL검사 참여자에 대해 별도의 NNAL검사 가중치(wt\_nn) 부여

### 〈 조사부문 및 항목별 기본 가중치 〉

분석변수가 속한 조사부문	가중치 변수명 : 앞3자리 "wt_"생략						
	제1기 (1998)	제2기 (2001)	제3기 (2005)	제4기 (2007)	제5기 (2009)	제6기 (2010)	제7기 (2017)
■ 가구 가중치	-	-	hs	hs	hs	hs	hs
■ 건강설문조사 가중치 <sup>1)</sup>							
건강면접조사	itv		itv	itvex	itvex	itvex	itvex
건강행태조사	bhv						
■ 검진조사 가중치							
검진조사 기본가중치 <sup>1)</sup>	ex	exhs	-	-	-	-	-
검진보통조사	ex	exhs	-	-	-	-	-
안검사, 이비인후검사 <sup>2)</sup>	-	-	-	ex1	ex	itvex	itvex
흉부X-선검사 <sup>2)</sup>	-	-	-	ex1	ex	itvex	ex1
골밀도검사 <sup>2)</sup>	-	-	-	ex1	ex	itvex	ex1
체지방검사 <sup>2)</sup>	-	-	-	ex1	dw	itvex	ex1
골관절염검사 <sup>3)</sup>	-	-	-	-	ex1	itvex	itvex
폐기능검사	-	-	-	pft	pft	pft	pft
중금속검사, 감상선검사 <sup>4)</sup>	-	-	hm	-	hm	hm	hm
NNAL검사 <sup>5)</sup>	-	-	-	-	-	-	nn
■ 영양조사 가중치							
식생활조사	life						
영유아식생활조사	ylife		ntr		ntr	ntr	ntr
24시간회식조사	24rc						
식품섭취빈도조사 <sup>6)</sup>	ffq						-

1) 제5기 1차년도(2010)부터 건강설문조사 가중치와 검진조사 가중치를 건강설문&검진조사 가중치로 통합

2) 안검사, 이비인후검사, 흉부X-선검사, 골밀도검사, 체지방검사: (2008년) 7월 신규도입 연도에서 별도 가중치 사용  
(2009년~) 검진조사 기본가중치 사용

- 흉부X-선검사: 2013년 6월 조사완료 연도에서 별도 가중치 사용

- 골밀도검사: 2011년 5월 조사완료 연도에서 별도 가중치 사용

- 체지방검사: (2009년) 총 200조사구 중 199조사구만 조사완료 연도에서 별도 가중치 사용  
(2011년) 5월 조사완료 연도에서 별도 가중치 사용

- 안검사: (2017년) 4월 신규도입 연도에서 별도 가중치 사용

3) 골관절염검사: (2009년) 8월 신규도입 연도에서 별도 가중치 사용, (2010~2013년) 검진조사 기본가중치 사용

4) 중금속검사, 감상선검사: 다음과 같이 검진조사 대상자 중 일부를 선정하여 검사하였으므로 별도 가중치 사용

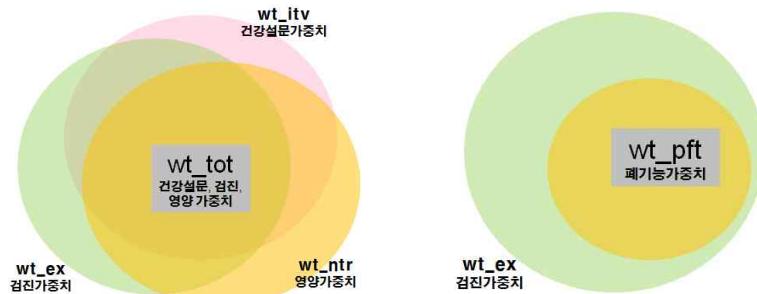
- **중금속검사:** **아연, 촉1gE, 질먼지진드기/바퀴/개털, 알레르겐:**  
(2008~09년) 만20세이상 2,000명, (2010~13년) 만10세이상 2,400명 (2010년) 중금속검사 대상자 대상 실시→ 중금속검사 가중치 사용  
(2016년) 만10세이상 대상자의 1/3표본

- **감상선검사:** **【감상선자극호르몬/유리티록신/갑상선과산화효소활성제】** 만40세이상 2,400명 (2008~09년 6월, 2014년) 09년 포함해 검진조사 기본가중치 사용  
【요오드】 (2013년~2014년 6월) 만10세이상 2,400명 (2009년 7월~2011년) 중금속검사 대상자만 실시  
(2014년 7~12월) 만10세이상 전체, → '10~'11년 중금속가중치 사용  
(2015년) 만6세이상 2,650명

5) **NNAL검사:** 검진조사 대상자 중 일부를 선정하여 검사하였으므로 별도 가중치 사용, (2016~2017년) 만6세이상, 약3,000명

6) **식품섭취빈도조사:** 제7기 2차년도(2017) 조사 미실시

### < 부문별 포함관계에 따른 연관성분석 가중치의 산출 예시 >



\* 가중치변수명은 제4기(2007~2009) 기준

- 연관성 가중치는 2개 이상의 조사부문(또는 조사항목)에 모두 참여한 대상자에게 가중치를 부여하였다.

- 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, 「부문별 포함관계에 따른 연관성분석 가중치의 산출」 그림과 같이 건강설문조사, 검진조사, 영양조사 3개 부문 참여자는 건강설문-검진-영양조사 연관성분석 가중치(wt\_tot)를 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)~제7기 1,2차년도(2016~2017) : 검진조사와 건강설문조사의 참여자수가 거의 일치 하여 두 조사의 가중치를 통합하였으므로, 두 조사부문 중 1개 조사부문의 변수가 분석에 포함되는 경우 와 2개 조사부문의 변수가 분석에 포함되는 경우 모두 건강설문-검진조사 가중치(wt\_itvex)를 사용

### < 조사부문별 연관성 가중치 >

분석변수들이 속한 조사부문	가중치 변수명 : 앞3자리 "wt_" 생략					
	제3기 (2005)	제4기 (2007)	제5기 (2008)	제6기 (2011)	제7기 (2012)	제7기 (2013)
• 건강설문(건강면접/건강행태) AND 검진	bhvex	itvex	itvex	itvex	itvex	itvex
• 건강설문( ), AND 영양	bhvnt	itvnt				
• 검진 AND 영양	exnt	exnt	tot	tot	tot	tot
• 건강설문( ) AND 검진 AND 영양	tot	tot				
• 폐기능 AND						
(건강설문( ) OR 검진)7)	-	pft	pft	pft	pft	pft
(건강설문( ) OR 검진) AND 영양	-	pfnt	pfnt	pfnt	pfnt	pfnt
중금속(감상선)	-	-	pfhm	pfhm	pfhm	pfhm
중금속(감상선) AND 영양	-	-	pfhmn	pfhmn	pfhmn	pfhmn
• 중금속(감상선) AND						
(건강설문( ) OR 검진)	hm	-	hm	hm	hm	hm
(건강설문( ) OR 검진) AND 영양	-	-	hmnt	hmnt	hmnt	hmnt
• NNAL AND						
(건강설문( ) OR 검진)	-	-	-	-	-	-
(건강설문( ) OR 검진) AND 영양	-	-	-	-	-	-

7) 예시: (폐기능검사 AND 건강설문조사 항목) 또는 (폐기능검사 AND 검진조사 항목) 또는 (폐기능검사 AND 건강설문조사 AND 검진조사 항목)을 함께 분석하는 경우 사용하는 가중치

- 검진조사 가중치(제4기 wt\_ex, 제5기~제7기 2차년도(2017) 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  
(예) 제5기 2차년도(2011) 자료에서 흡연(건강설문조사), 혈압(검진조사), 영양소(영양조사) 변수를 함께 분석하는 경우에는 건강설문-검진-영양 가중치(wt\_tdt)를 사용하고, 흡연(건강설문조사), 혈압(검진조사), 폐기능(검진조사) 자료를 함께 분석하는 경우에는 폐기능검사 가중치(wt\_pft)를 사용

## □ 증변수 : 제1기(1998) ~ 제7기 2차년도(2017)

- 일부 표본설계층을 통합하여 생성한 분산추정층(변수명: kstrata)을 기본 증변수로 사용한다.
- 단, 2008년 7월 신규도입된 안검사, 이비인후검사, 골밀도 및 체지방검사, 흉부X-선 검사 자료분석 시에는 2008년 하반기 분산추정층(변수명: kstrata1)를, 2009년 8월 신규도입된 골관절염검사 자료분석 시에는 2009년 하반기 분산추정층(변수명: kstrata2)을 사용한다.  
※kstrata1, kstrata2는 1개 조사구만 있는 일부 표본설계층을 통합한 분산추정용 층이므로, 신규도입 항목에 대해 도입연도에 kstrata를 사용하여 분석할 경우 분산 추정 편의 발생 가능
- 기수 간, 기수 내 자료를 통합하는 연도 통합 분석시에는 2008년 및 2009년 신규도입 항목의 경우 신규도입한 연도의 자료와 이후의 자료를 통합하게 되므로, 2008년 하반기 분산추정층 변수(변수명: kstrata1)이 아닌 기존 분산추정층(변수명: kstrata)를 사용한다.

## □ 집락변수 : 제1기(1998) ~ 제7기 2차년도(2017)

- 집락변수는 1차 표본추출단위를 의미하며 제1기(1998)~제7기 1,2차년도(2016-2017) 조사구 변수(변수명: PSU)를 사용한다.

※제4기(2007-2009) 1차 추출단위 동읍면, 2차 추출단위 조사구이나, 1개 동읍면당 1개 조사구만을 추출하기 때문에 집락 변수에 조사구 지정

# 03

## 분석DB 준비 단계

- 자료결합
  - 자료의 가로결합
  - 자료의 세로결합
- 식품섭취조사 개인단위 분석DB 생성
- 통합가중치 산출
- 기타
  - 라이브러리 생성
  - 데이터셋 활용하기(SET문, KEEP문)

## □ 자료결합

- 국민건강영양조사 원시자료는 1998년부터 2017년까지의 자료가 공개되어 있으며 각 연도에는 기본 DB 및 분야별 DB로 구성되어 있다. 자료 분석 시 여러 연도의 자료를 합쳐서 분석 하거나 필요한 변수가 기본 DB와 분야별 DB에 각각 존재할 경우, 분석하기 전 해당 DB들을 자료결합을 통해 1개의 새로운 DB로 생성해야 한다.

## □ 자료의 가로결합

- 국민건강영양조사 원시자료는 기본 DB 및 분야별 DB로 구성되어 있어 자료분석시 필요한 변수가 기본 DB와 분야별 DB에 각각 존재할 경우, 해당 DB들을 가로결합하여 1개의 새로운 DB로 생성한 후 분석해야 한다.
- 국민건강영양조사 원시자료에서 개인별 조사자료는 개인아이디(ID)로 구분되어 있고, DB별 개인아이디가 동일한 경우는 동일인의 자료이다. 두 개의 DB를 동일인에 대하여 가로결합을 하기 위해 각각의 DB를 ID(기준변수)로 정렬한 후 **MERGE문**을 사용하여 새로운 하나의 DB를 생성해야 한다.
- 기본형태

```
PROC SORT DATA=데이터셋;
  BY </옵션> 기준변수;
RUN;

DATA 새로운데이터셋이름;
  MERGE 기준데이터셋1 기준데이터셋2 … 기준데이터셋n;
  BY 기준변수;
RUN;
```

- PROC SORT DATA= 데이터셋  
가로결합을 하기 전에 결합하고자 하는 데이터셋들을 기준이 되는 기준변수로 정렬해야한다. 정렬하고자 하는 데이터셋의 이름을 **데이터셋**에 지정한다.

- BY </옵션> 기준변수

데이터셋을 정렬시키기 위한 기준변수명을 **기준변수**에 지정한다.

옵션에는 descending(내림차순)을 사용할 수 있으며 생략시에는 ascending(오름차순) 옵션이 사용된다.

- DATA 새로운데이터셋이름

두 개 이상의 데이터셋을 하나의 새로운 데이터셋으로 만들 데이터셋의 이름을 **새로운데이터셋이름**에 지정한다.

- MERGE 기존데이터셋들

가로결합 할 기존 데이터셋들의 이름을 **기존데이터셋들**에 나열한다.

- BY 기준변수

가로결합 할 두 개 이상의 데이터셋에서 기준이 되는 기준변수의 이름을 **기준변수**에 지정한다.

## 자료의 가로결합 예제

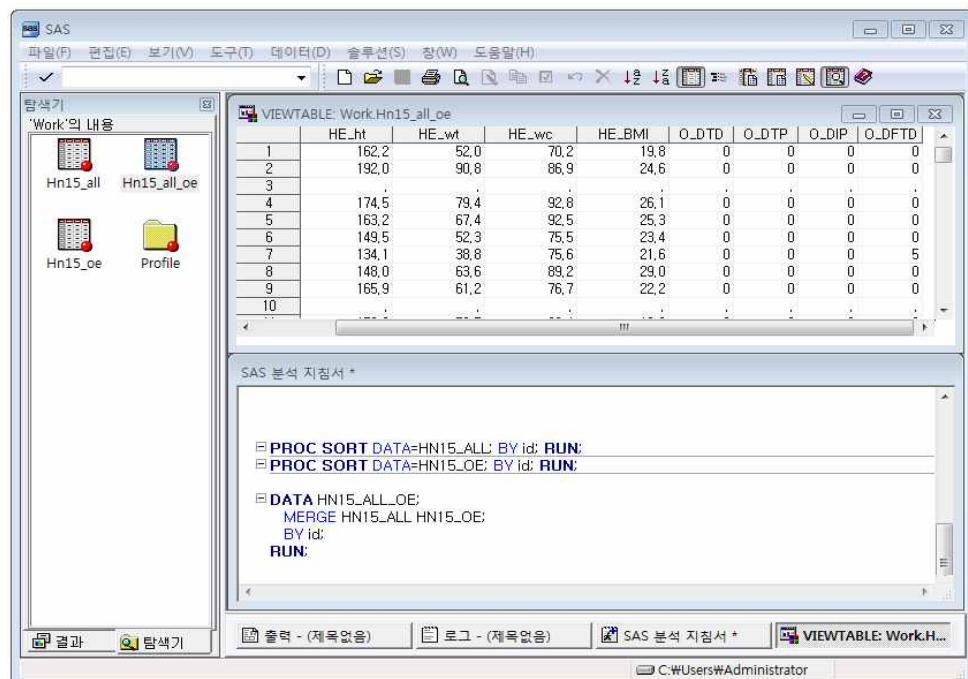
□ 2015년 국민건강영양조사 기본 DB(HN15\_ALL)와 구강DB(HN15\_OE)를 하나의 DB(HN15\_ALL\_OE)로 가로결합하고자 한다.

- 1) 가로결합을 하기 전 PROC SORT를 이용하여 결합할 두개의 DB(HN15\_ALL, HN15\_OE)를 기준변수인 ID로 오름차순 정렬한다.

```
PROC SORT DATA=HN15_ALL; BY id; RUN;
PROC SORT DATA=HN15_OE; BY id; RUN;
```

- 2) 오름차순으로 정렬된 데이터셋(HN15\_ALL, HN15\_OE)들을 MERGE문을 이용하여 새로운 데이터셋(HN15\_ALL\_OE)으로 가로결합한다. 이때 동일 대상자의 자료를 연계하기 위해 기준변수인 ID를 이용하여 두 DB의 데이터셋을 가로결합한다.

```
DATA HN15_ALL_OE;
  MERGE HN15_ALL HN15_OE;
  BY id;
RUN;
```



**< HN15\_ALL과 HN15\_OE의 가로결합 >**

work.HN15_ALL				work.HN15_OE		work.HN15_ALL_OE						
ID	HE_ht	HE_wt	HE_wc	ID	Q_DTD	Q_DTP	ID	HE_ht	HE_wt	HE_wc	Q_DTD	Q_DTP
1	156.4	69.8	87.5	1	0	0	1	156.4	69.8	87.5	0	0
2	162.5	73.2	84.6	2	0	1	2	162.5	73.2	84.6	0	1
3	113.0	19.9	48.0				3	113.0	19.9	48.0	.	.
4	166.0	56.3	67.5	4	0	4	4	166.0	56.3	67.5	0	4
5	180.8	75.7	86.9				5	180.8	75.7	86.9	.	.

+

=

## □ 자료의 세로결합

- 국민건강영양조사 원시자료는 1998년부터 2017년까지의 자료가 공개되어 있다. 세부 집단별 분석 또는 낮은 유병률 자료 분석 시, 1개년도의 자료만으로는 자료수가 충분하지 않은 경우 기수간 혹은 기수내의 다른 조사년도와의 자료결합을 고려해 볼 수 있다.
- 기수간 혹은 기수내의 다른 조사년도와의 자료를 결합하여 하나의 DB로 만들고자 할 때 **SET문**을 사용한다.
- 기본형태

**DATA** 새로운데이터셋이름;

**SET** 기존데이터셋1 기존데이터셋2 … 기존데이터셋n;

**RUN;**

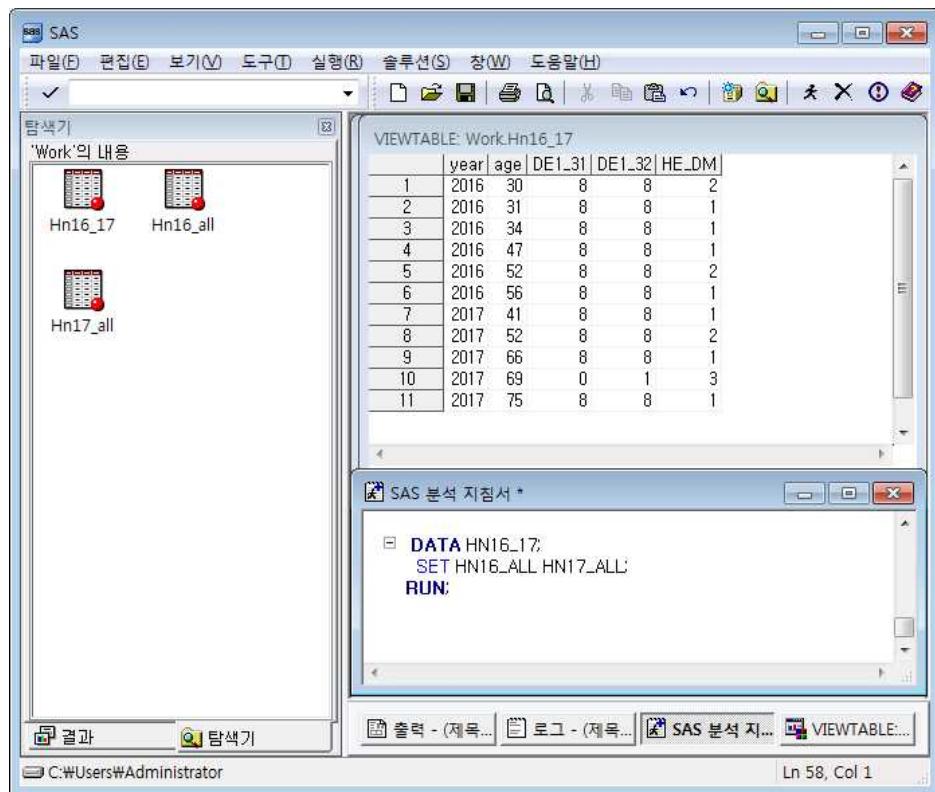
- DATA 새로운데이터셋이름  
세로결합하여 생성할 새로운 데이터셋의 이름을 **새로운데이터셋이름**에 지정 한다.
- SET 기존데이터셋들  
세로결합시 필요한 데이터셋들의 이름을 **기존데이터셋들**에 나열한다.

## 자료의 세로결합 예제

- 국민건강영양조사 제7기 1,2차년도(2016~2017년) 기본 DB(HN16\_ALL, HN17\_ALL)를 하나의 통합DB(HN16\_17)로 세로결합하고자 한다.

- 1) 통합할 데이터셋(HN16\_ALL, HN17\_ALL)들을 SET 문을 이용하여 새로운 데이터셋 (HN16\_17)으로 세로결합한다.

```
DATA HN16_17;
  SET HN16_ALL HN17_ALL;
RUN;
```



< HN16\_ALL, HN17\_ALL의 세로결합 >

work.HN16\_ALL

ID	year	age	DE1_31	DE1_32	HE_DM
1	2016	47	8	8	2
2	2016	31	8	8	1
3	2016	30	8	8	3
4	2016	52	0	1	3
5	2016	56	8	8	1
6	2016	34	8	8	1

+

work.HN17\_ALL

ID	year	age	DE1_31	DE1_32	HE_DM
1	2017	66	0	1	3
2	2017	52	8	8	2
3	2017	75	0	1	3
4	2017	69	8	8	1

work.HN16\_17

ID	year	age	DE1_31	DE1_32	HE_DM
1	2016	47	8	8	2
2	2016	31	8	8	1
3	2016	30	8	8	3
4	2016	52	0	1	3
5	2016	56	8	8	1
6	2016	34	8	8	1
1	2017	66	0	1	3
2	2017	52	8	8	2
3	2017	75	0	1	3
4	2017	69	8	8	1

## □ 식품섭취조사 개인단위 분석DB 생성

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

< 식품섭취조사DB(HNyr\_24RC) 자료 예시 >

ID	끼니구분	음식명	총부피	섭취부피	식품재료	섭취량	에너지	단백질	지방	탄수화물
A101020101	1	혼합잡곡밥	943	300	혼합잡곡	52.15	193	3.9	0.6	41.3
A101020101	1	혼합잡곡밥	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, 끽고추	300	50	깻소금	0.5	3	0.1	0.3	0
A101020101	1	멸치볶음, 끽고추	300	50	식용유	2	18	0	0	0
A101020101	1	멸치볶음, 끽고추	300	50	풋고추	21.71	4	0.3	0.1	0.8

- 식품섭취조사DB(HNyr\_24RC)를 이용하여 개인별 섭취량, 에너지 및 영양소를 합산한 개인이 1일간 섭취한 섭취량은 기본 DB(HNyr\_ALL)에 포함되어 있다. 따라서 개인별 섭취량, 에너지 및 영양소자료를 활용시에는 기본 DB를 사용하며, 이외에 식품군별 분석, 끼니별 분석, 특정 식품 분석 등을 위해서는 식품섭취조사DB를 사용하여 해당 변수를 생성해야 한다. 이때 개인당 여러 개의 자료를 하나의 자료로 변환하는 과정이 필요하다.

## 식품섭취조사 개인단위 분석DB생성 예제

- 2017년 식품섭취조사DB(HN17\_24RC)에서 개인별 1일 채소류섭취량을 산출하고자 한다.

### □ 분석변수 요약

구분	변수명	변수설명	내용
기존변수	ID	개인아이디	
	NF_INTK	식품섭취량	1일 식품섭취량(g)
	N_KINDG1	식품군분류1	01.곡류 12.유지류(식물) 02.감자·전분류 13.기타(식물) 03.당류 15.육류 04.두류 16.난류 05.종실류 17.어패류 06.채소류 18.우유류 07.버섯류 19.유지류(동물) 08.과일류 20.기타(동물) 09.해조류 21.음료류 11.양념류 22.주류
생성변수	vegetable	채소류 식품섭취량	1일 채소류 식품섭취량(g)
	vegetable_sum	개인별 총 채소류 식품섭취량	개인별 1일 채소류 식품섭취량의 합

- 개인별 1일 채소류 섭취량을 산출과정은 다음과 같다.

- 식품섭취조사DB에서 식품군분류1(N\_KINDG1)이 채소류(코드:06)인 식품섭취량을 합하여 개인별 채소류 식품섭취량 산출
- 식품섭취량(NF\_INTK)에서 채소류 섭취량 변수 생성
- 채소류 식품섭취량을 합하여 개인별 1일 채소류 섭취량 변수 생성

### 1) 식품섭취량(NF\_INTK)에서 채소류 섭취량 변수(vegetable) 생성

식품섭취조사DB에서 식품군분류1(N\_KINDG1)이 채소류(코드:06)인 식품섭취량 값들만 갖도록 채소류 식품섭취량 변수(vegetable)를 생성한다.

```

DATA hn17_24rc_1;
SET hn17_24rc;

if N_KINDG1='06' then vegetable=NF_INTK;

keep id N_KINDG1 vegetable NF_INTK;

RUN;

```

VIEWTABLE: Work.Hn17\_24rc\_1

	ID	N_KINDG1	NF_INTK	vegetable
1	A601169401	06	0,02	0,0221152
2	A601169401	06	0,10	0,100747
3	A601169401	06	0,22	0,2236092
4	A601169401	17	1,86	.
5	A601169401	11	0,06	.
6	A601169401	11	0,05	.
7	A601169401	06	1,54	1,53898
8	A601169401	04	68,57	.
9	A601169401	06	8,47	8,47000674
10	A601169401	06	3,03	3,03302629
11	A601169401	06	0,38	0,38193664
12	A601169401	11	0,09	.
13	A601169401	11	7,02	.
14	A601169401	17	4,04	.

## 2) 채소류 식품섭취량을 합하여 개인별 1일 채소류 섭취량 변수 생성

SUMMARY 프로시저를 이용하여 개인별 1일 채소류 식품섭취량 변수를 생성한다. 이때 총합을 산출하기 위해 SUMMARY 프로시저에서 OUTPUT 문의 통계량 옵션인 SUM을 사용하며, 해당 변수명을 ‘vegetable\_sum’이라고 지정하고 식품섭취조사 DB가 아닌 새로운 DB(hn17\_24rc\_2)를 생성한다.

\*만약 SUM 뒤에 변수명을 지정하지 않는 경우, 산출하고자하는 변수의 변수명으로 총합이 저장된다.

```

PROC SUMMARY DATA=hn17_24rc_1 nway;
VAR vegetable;
BY id;
OUTPUT OUT=hn17_24rc_2(drop=_type_ _freq_)
SUM(vegetable)=vegetable_sum;
RUN;

```

	ID	vegetable_sum
1	A601169401	246,05429838
2	A601169402	165,2329797
3	A601177901	432,25420718
4	A601177902	126,66455255
5	A601177903	0,1792786421
6	A601208501	181,26207729
7	A601208502	388,08775758
8	A601240801	166,00165793

### 3) 기본 DB와 개인별 1일 채소류 섭취량DB 가로결합

식품섭취조사DB를 통해 개인별 1일 채소류 섭취량 변수(vegetable\_sum)를 생성하면, 가로결합을 통해 기본 DB에 생성한 개인별 1일 채소류 섭취량 변수를 추가한다.

- ① 가로결합을 하기 전에 기본 DB와 개인별 1일 채소류 식품섭취량 변수가 존재하는 개인별 1일 채소류 섭취량 DB를 기준변수인 ID로 정렬한다.

```
PROC SORT DATA=hn17_ALL; BY id; RUN;
PROC SORT DATA=hn17_24rc_2; BY id; RUN;
```

- ② 기준변수인 ID로 정렬한 두 개의 DB를 MERGE문을 이용하여 가로결합한다.

※자료의 가로결합 과정은 「자료의 가로결합(p20~23)」 참조

```
DATA hn17_24rc_3;
MERGE hn17_ALL hn17_24rc_2;
BY id;
RUN;
```

	ID	psu	wt_ntr	kstrata	vegetable_sum
120	A604378501	A604	3466,3095461	713	350,67597625
121	A604378502	A604	3382,8225253	713	395,84816692
122	A604378503	A604	5385,6437532	713	169,85245332
123	A604378504	A604	4879,3650368	713	38,995401254
124	A604387001	A604	4546,2907648	713	422,62763158
125	A604387002	A604	6239,1646857	713	220,43682477
126	A604387004	A604	4634,2711537	713	
127	A604415901	A604	4091,3549141	713	180,38782874
128	A604415902	A604	6966,6433782	713	78,719764305
129	A604415903	A604	4879,3650368	713	74,193971166
130	A604421001	A604	4546,2907648	713	425,95531299

#### 4) 채소류 식품섭취량이 결측인 경우 0으로 변환

식품섭취조사 완료자 중에 개인별 1일 채소류 식품섭취량이 결측값(.)인 경우, 조사에는 참여했지만 채소류 식품군을 섭취하지 않은 대상자이므로 해당 섭취량의 값을 0으로 변환해야 한다.

```
DATA hn17_24rc_4;
SET hn17_24rc_3;
ARRAY aaa[*] vegetable_sum;
DO i=1 TO dim(aaa);
  IF aaa{i}=. THEN aaa(i)=0;
END;
RUN;
```

	ID	psu	wt_ntr	kstrata	vegetable_sum
120	A604378501	A604	3466,3095461	713	350,67597625
121	A604378502	A604	3382,8225253	713	395,84816692
122	A604378503	A604	5385,6437532	713	169,85245332
123	A604378504	A604	4879,3650368	713	38,995401254
124	A604387001	A604	4546,2907648	713	422,62763158
125	A604387002	A604	6239,1646857	713	220,43682477
126	A604387004	A604	4634,2711537	713	0
127	A604415901	A604	4091,3549141	713	180,38782874
128	A604415902	A604	6966,6433782	713	78,719764305
129	A604415903	A604	4879,3650368	713	74,193971166
130	A604421001	A604	4546,2907648	713	425,95531299

## □ 통합가중치 산출

□ 국민건강영양조사 원시자료에는 각 조사부문별 기본 가중치와 연관성 가중치가 있다.

당해년도 자료분석 시에는 원시자료에 제공되는 가중치를 사용하지만, 연도별 자료통합(세로통합) 분석 시에는 다음과 같이 **통합 가중치를 산출**한 후 분석을 실시해야 한다.

※ 자료의 통합DB 생성과정은 「자료의 세로통합(p24~26)」, 조사부문별 기본 가중치는 「가중치(p16)」 참조

## □ 기수내 자료통합

제4기(2007–2009), 제5기(2010–2012), 제6기(2013–2015), 제7기 1,2차년도(2016–2017)의 기수내 자료를 통합하는 경우, 기존 가중치에 연도별 조사구수 비율을 곱하여 통합 가중치를 산출하여 사용한다. 예를 들어, 제4기(2007–2009) 3개년도의 자료를 통합하는 경우, 조사구수가 2007년 100개, 2008년 200개, 2009년 200개이므로 통합 가중치는 다음 표와 같이 계산한다.

〈 기수내 자료통합 시 통합가중치 계산식 〉

구 분		소계	1차년도	2차년도	3차년도
제 4 기	(조사연도) 조사구수	500	(2007) 100개	(2008) 200개	(2009) 200개
	2007–2008(2년) 통합	300	$w_{07-08} = w_{07} \times (100/300)$	$w_{07-08} = w_{08} \times (200/300)$	-
	2008–2009(2년) 통합	400	-	$w_{08-09} = w_{08} \times (200/400)$	$w_{08-09} = w_{09} \times (200/400)$
	2007–2009(3년) 통합	500	$w_{07-09} = w_{07} \times (100/500)$	$w_{07-09} = w_{08} \times (200/500)$	$w_{07-09} = w_{09} \times (200/500)$
제 5 기	(조사연도) 조사구수	576	(2010) 192개	(2011) 192개	(2012) 192개
	2010–2011(2년) 통합	384	$w_{10-11} = w_{10} \times (192/384)$	$w_{10-11} = w_{11} \times (192/384)$	-
	2011–2012(2년) 통합	384	-	$w_{11-12} = w_{11} \times (192/384)$	$w_{11-12} = w_{12} \times (192/384)$
	2010–2012(3년) 통합	576	$w_{10-12} = w_{10} \times (192/576)$	$w_{10-12} = w_{11} \times (192/576)$	$w_{10-12} = w_{12} \times (192/576)$
제 6 기	(조사연도) 조사구수	576	(2013) 192개	(2014) 192개	(2015) 192개
	2013–2014(2년) 통합	384	$w_{13-14} = w_{13} \times (192/384)$	$w_{13-14} = w_{14} \times (192/384)$	-
	2015–2016(2년) 통합	384	-	$w_{14-15} = w_{14} \times (192/384)$	$w_{14-15} = w_{15} \times (192/384)$
	2013–2015(3년) 통합	576	$w_{13-15} = w_{13} \times (192/576)$	$w_{13-15} = w_{14} \times (192/576)$	$w_{13-15} = w_{15} \times (192/576)$
제 7 기	(조사연도) 조사구수	576	(2016) 192개	(2017) 192개	(2018) 192개
	2016–2017(2년) 통합	384	$w_{16-17} = w_{16} \times (192/384)$	$w_{16-17} = w_{17} \times (192/384)$	-

※  $w_{yr}$ : 해당 조사연도( $yr$ ) 가중치,  $w_{yr_1 - yr_2}$ : 조사연도  $yr_1 \sim yr_2$  간 자료통합 시 가중치

## □ 기수간 자료통합

제4기(2007–2009), 제5기(2010–2012), 제6기(2013–2015), 제7기 1,2차년도(2016–2017) 기수 내 뿐만 아니라 기수 간 자료 통합도 가능하다. 기수간 자료를 통합하는 경우, 연도별 조사구 자료가 그 연도를 대표한다고 간주하여 연간 조사자료에 값 1을 부여하고, 이를 기준으로 각 연도의 조사기간에 비례하는 값을 부여한다. 단, 2007년은 연중 반기(7~12월) 동안만 조사된 점을 고려하여 다른 연도가 1의 통합 비율을 갖는다면, 2007년이 1/2의 통합 비율을 갖도록 부여한다.

〈 기수간 자료통합 시 통합가중치 계산식 〉

구 분	소 계	제4기			제5기			제6기			제7기	
		(2007)	(2008)	(2009)	(2010)	(2011)	(2012)	(2013)	(2014)	(2015)	(2016)	(2017)
조사기간 값	10.5	0.5	1	1	1	1	1	1	1	1	1	1
2007–2012년 (6년) 통합	5.5	$w_{07-12} = w_{07} \times \frac{0.5}{5.5} = w_{08} \times \frac{1}{5.5}$	$w_{07-12} = w_{08} \times \frac{1}{5.5}$	$w_{07-12} = w_{09} \times \frac{1}{5.5}$	$w_{07-12} = w_{10} \times \frac{1}{5.5}$	$w_{07-12} = w_{11} \times \frac{1}{5.5}$	$w_{07-12} = w_{12} \times \frac{1}{5.5}$	–	–	–	–	–
2011–2015년 (5년) 통합	5	–	–	–	–	$w_{11-15} = w_{11} \times \frac{1}{5}$	$w_{11-15} = w_{12} \times \frac{1}{5}$	$w_{11-15} = w_{13} \times \frac{1}{5}$	$w_{11-15} = w_{14} \times \frac{1}{5}$	$w_{11-15} = w_{15} \times \frac{1}{5}$	–	–
2014–2017년 (4년) 통합	4	–	–	–	–	–	–	–	$w_{14-17} = w_{14} \times \frac{1}{4}$	$w_{14-17} = w_{15} \times \frac{1}{4}$	$w_{14-17} = w_{16} \times \frac{1}{4}$	$w_{14-17} = w_{17} \times \frac{1}{4}$

※ 2007년은 연중 하반기만 조사, 2008년부터 1년간 조사

※  $w_{yr}$ : 해당 조사연도( $yr$ ) 가중치,  $w_{yr_1 - yr_2}$ : 조사연도  $yr_1 \sim yr_2$  간 자료통합 시 가중치

## □ 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 적용한다.

〈 안검사 및 이비인후검사 자료통합시 통합가중치 계산식 〉

구 分	소계	제4기		제5기			제6기	
		(2008)	(2009)	(2010)	(2011)	(2012)	(2013)	(2014)
기수내	조사구수	1,268	108	200	192	192	192	192
기수내	제4기 (2년) 통합	308	$w_{08-09} = w_{08} \times \frac{108}{308}$	$w_{08-09} = w_{09} \times \frac{200}{308}$	–	–	–	–
기수내	제5기 (3년) 통합	576	–	–	$w_{10-12} = w_{10} \times \frac{192}{576}$	$w_{10-12} = w_{11} \times \frac{192}{576}$	$w_{10-12} = w_{12} \times \frac{192}{576}$	–
기수내	제6기 (2년) 통합	384	–	–	–	–	$w_{13-14} = w_{13} \times \frac{192}{384}$	$w_{13-14} = w_{14} \times \frac{192}{384}$

구 분	소계	제4기		제5기			제6기	
		(2008)	(2009)	(2010)	(2011)	(2012)	(2013)	(2014)
기 수 간	조사기간 값	6.5	0.5	1	1	1	1	1
	2008-2012년 (5년) 통합	4.5	$w_{08-12} = w_{08} \times \frac{0.5}{4.5}$	$w_{08-12} = w_{09} \times \frac{1}{4.5}$	$w_{08-12} = w_{10} \times \frac{1}{4.5}$	$w_{08-12} = w_{11} \times \frac{1}{4.5}$	$w_{08-12} = w_{12} \times \frac{1}{4.5}$	-
	2010-2014 (5년) 통합	5	-	-	$w_{10-14} = w_{10} \times \frac{1}{5}$	$w_{10-14} = w_{11} \times \frac{1}{5}$	$w_{10-14} = w_{12} \times \frac{1}{5}$	$w_{10-14} = w_{13} \times \frac{1}{5}$

\* 2008년 7월 안·이비인후 검사항목 도입 이후, 2009년부터 1년간 조사

\*  $w_{yr}$ : 해당 조사연도( $yr$ ) 가중치,  $w_{yr1-yr2}$ : 조사연도  $yr_1 \sim yr_2$  간 자료통합 시 가중치

## □ 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.

### 〈 골밀도검사 및 체지방검사 자료통합시 통합가중치 계산식 〉

통합예시	소계	제4기		제5기	
		(2008)	(2009)	(2010)	(2011)
기 수 내	조사구수	580	108	(골밀도) 200	192
		579		(체지방) 199	
	제4기 (2년) 통합	308 (골밀도)	$w_{08-09} = w_{08} \times \frac{108}{308}$	$w_{08-09} = w_{09} \times \frac{200}{308}$	-
기 수 간	제5기 (2년) 통합	307 (체지방)	$w_{08-09} = w_{08} \times \frac{108}{307}$	$w_{08-09} = w_{09} \times \frac{199}{307}$	-
	조사기간 값	2.5	0.5	1	1
	2008-2011년 (4년) 통합	580 (골밀도)	$w_{08-11} = w_{08} \times \frac{108}{580}$	$w_{08-11} = w_{09} \times \frac{200}{580}$	$w_{08-11} = w_{10} \times \frac{192}{580}$
	2008-2010년 (3년) 통합	579 (체지방)	$w_{08-11} = w_{08} \times \frac{108}{579}$	$w_{08-11} = w_{09} \times \frac{199}{579}$	$w_{08-11} = w_{10} \times \frac{192}{579}$
		2.5	$w_{08-10} = w_{08} \times \frac{0.5}{2.5}$	$w_{08-10} = w_{09} \times \frac{1}{2.5}$	$w_{08-10} = w_{10} \times \frac{1}{2.5}$

\*  $w_{yr}$ : 해당 조사연도( $yr$ ) 가중치,  $w_{yr1-yr2}$ : 조사연도  $yr_1 \sim yr_2$  간 자료통합 시 가중치

## 통합가중치 산출 예제1 : 기수내

- 국민건강영양조사 제7기 1,2차년도(2016–2017) 기본 DB(HN16\_ALL, HN17\_ALL)를 이용하여 통합DB를 생성한 후, 건강설문캡진 통합가중치 wt\_ex\_pool를 산출하고자 한다.

- 1) 통합할 데이터셋을 SET 문을 이용하여 하나의 데이터셋으로 세로결합한다.

※ 자료의 통합DB 생성과정은 「자료의 세로통합(p24~26)」 참조

```
DATA HN16_17_all;
  SET HN16_all HN17_all;
RUN;
```

- 2) 건강설문캡진 통합가중치 계산식을 확인한다.

- 건강설문캡진 통합가중치 계산식

년도		2016	2017
건강설문캡진 가중치		wt_itvex	wt_itvex
통합 가중치	통합비율	1	1
	가중치 × $\frac{\text{통합비율}}{\text{통합비율합}}$	wt_itvex × 1/2	wt_itvex × 1/2

- 3) 위와 같은 통합비율로 각 연도에 해당하는 건강설문캡진 통합가중치를 산출한다.

```
DATA HN16_17_all1;
  SET HN16_17_all;
  if year=2016 then wt_ex_pool=wt_itvex*1/2;
  if year=2017 then wt_ex_pool=wt_itvex*1/2;
RUN;
```

※if–then 문 : if 조건문 then 실행문

※제4기, 제5기, 제6기도 동일한 방법으로 통합가중치를 산출하며, 이때 각 기수에 해당하는 통합가중치 계산식을 확인한 후 적용한다.

## 통합가중치 산출 예제2 : 기수간(체지방검사)

- 국민건강영양조사 제4기 2,3차년도(2008–2009) 와 제5기 1,2차년도(2010–2011) 골밀도 및 체지방검사 DB(HN08\_dxa~HN11\_dxa) 이용하여 통합DB를 생성한 후, 체지방 통합가중치 wt\_dw\_pool를 산출하고자 한다.

- 1) 통합할 데이터셋을 SET 문을 이용하여 하나의 데이터셋으로 세로결합한다.

※ 자료의 통합DB 생성과정은 「자료의 세로통합(p24~26)」 참조

```
DATA HN08_11_dw;
  SET HN08_dxa HN09_dxa HN10_dxa HN11_dxa;
RUN;
```

- 2) 체지방 통합가중치 계산식을 확인한다.

- 체지방 통합가중치 계산식

년도		2008	2009	2010	2011
검진조사(체지방) 가중치		wt_ex1	wt_dw	wt_itvex	wt_ex1
통합 가중치	조사구수	108	199	192	80
	통합비율	108/579	199/579	192/579	80/579
	가중치 × $\frac{\text{통합비율}}{\text{통합비율합}}$	wt_ex1 × 108/579	wt_dw × 199/579	wt_itvex × 192/579	wt_ex1 × 80/579

- 3) 위와 같은 통합비율로 각 연도에 해당하는 체지방 통합가중치를 산출한다.

```
DATA HN08_11_dw1;
  SET HN08_11_dw;
  if year=2008 then wt_dw_pool=wt_ex1*108/579;
  if year=2009 then wt_dw_pool=wt_dw*199/579;
  if year=2010 then wt_dw_pool=wt_itvex*192/579;
  if year=2011 then wt_dw_pool=wt_ex1*80/579;
RUN;
```

※if–then 문 : if 조건문 then 실행문

### 통합가중치 산출 예제3 : 기수간(골밀도검사)

- 국민건강영양조사 제4기 2,3차년도(2008–2009) 와 제5기 1,2차년도(2010–2011) 골밀도 및 체지방검사 DB(HN08\_dxa~HN11\_dxa) 이용하여 통합DB를 생성한 후, 골밀도 통합가중치 wt\_dxa\_pool를 산출하고자 한다.

- 1) 통합할 데이터셋을 SET 문을 이용하여 하나의 데이터셋으로 세로결합한다.

※ 자료의 통합DB 생성과정은 「자료의 세로통합(p24~26)」 참조

```
DATA HN08_11_dxa;
SET HN08_dxa HN09_dxa HN10_dxa HN11_dxa;
RUN;
```

- 2) 골밀도 통합가중치 계산식을 확인한다.

- 골밀도 통합가중치 계산식

년도		2008	2009	2010	2011
통합 가중치	검진조사(골밀도) 가중치	wt_ex1	wt_ex	wt_itvex	wt_ex1
	조사구수	108	200	192	80
	통합비율	108/580	200/580	192/580	80/580
	가중치 × $\frac{\text{통합비율}}{\text{통합비율합}}$	wt_ex1 × 108/580	wt_ex × 200/580	wt_itvex × 192/580	wt_ex1 × 80/580

- 3) 위와 같은 통합비율로 각 연도에 해당하는 골밀도 통합가중치를 산출한다.

```
DATA HN08_11_dx a1;
SET HN08_11_dx a;

if year=2008 then wt_dx a_pool=wt_ex1*108/580;
if year=2009 then wt_dx a_pool=wt_ex*200/580;
if year=2010 then wt_dx a_pool=wt_itvex*192/580;
if year=2011 then wt_dx a_pool=wt_ex1*80/580;

RUN;
```

※if–then 문 : if 조건문 then 실행문

## □ 라이브러리 생성

- SAS에서는 SAS 데이터셋을 자동으로 WORK 라이브러리에 저장되며, SAS 종료시 WORK 라이브러리에 저장된 데이터셋은 자동 삭제된다. 따라서 분석시 생성한 데이터셋을 다시 사용하려면 **LIBNAME문**을 통해 영구 라이브러리를 생성하여 SAS 데이터셋을 저장해야 한다.
- 다운로드 받은 원시자료를 분석하기 위해 LIBNAME문을 이용하여 저장된 데이터셋을 불러올 수도 있으며, 분석시 생성한 데이터셋을 다시 사용하기 위해 저장할 수도 있다.
- 기본형태

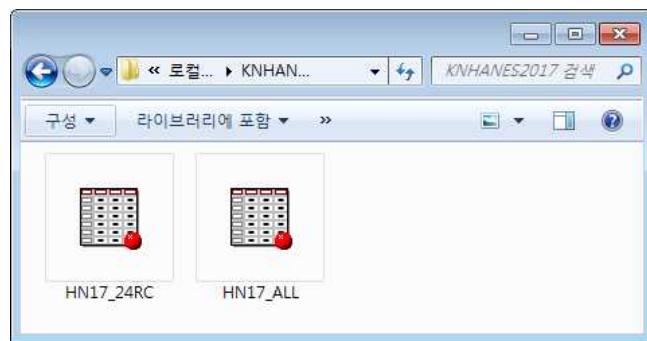
**LIBNAME** 라이브러리명 'SAS 파일 위치(파일 저장 경로)';

- 라이브러리명은 반드시 영문자나 밑줄(\_)로 시작되어야하며, 알파벳 대소문자 구별 없이 최대 8자까지 가능하다.

## 라이브러리 생성 예제

□ 로컬디스크(D드라이브)에 저장된 2017년 국민건강영양조사 DB를 SAS프로그램에 불러오는 과정은 다음과 같다.

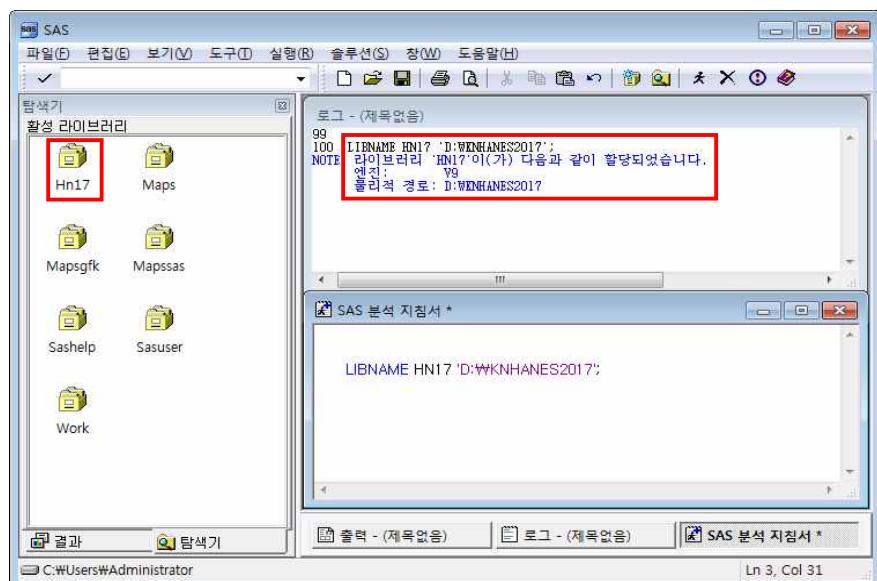
- 1) 국민건강영양조사 홈페이지에서 다운받은 2017년 DB의 저장된 경로(D:\KNHANES2017)를 확인한다.



- 2) 저장된 DB의 경로를 확인하고 LIBNAME문을 통해 SAS 데이터셋을 'HN17' 라이브러리명으로 지정하여 DB를 불러온다. 그러면 'HN17'로 할당된 라이브러리를 통해 'D:\KNHANES2017' 폴더에 저장된 데이터셋을 불러올 수 있다.

```
LIBNAME HN17 "D:\KNHANES2017";
```

※라이브러리를 지정하지 않으면 임시 라이브러리인 Work에 불러들여진다.



## □ 데이터셋 활용하기(SET문, KEEP문)

- SET 문은 자료의 세로결합 시 사용될 뿐만 아니라 기존에 영구라이브러리에 존재하는 SAS 데이터셋을 새로운 데이터셋으로 불러오고자 할 때도 **SET문**을 사용한다.  
※ 자료의 세로결합 과정은 「자료의 세로결합(p24~26)」 참조

- 기존 SAS 데이터셋의 변수들 중 필요한 변수로 구성된 SAS 데이터셋을 생성하는 이유는 전체DB를 사용하는 것보다 분석 실행속도를 단축시킬 수 있기 때문이다. 이렇게 필요한 변수로 구성 SAS 데이터셋을 생성하고자 할 때 **KEEP문**이 사용하여 필요한 변수들을 지정~~파열~~한다.

- 기본형태

```
DATA 새로운데이터셋이름;  
SET 라이브러리명.기존데이터셋이름;  
KEEP 선택변수명;  
RUN;
```

- DATA 새로운데이터셋이름  
선택한 변수들로 구성된 새로운 데이터 이름을 **새로운데이터셋이름**에 지정한다.
- SET 기존데이터셋이름  
데이터셋의 이름을 **기존데이터셋이름**에 지정한다. 단, 영구 라이브러리에 저장된 데이터셋을 불러들일 때는 기존 데이터셋 이름 앞에 라이브러리명을 지정해줘야 한다.
- KEEP 선택변수명  
기존 데이터셋에 들어있는 변수들 중 새로운 데이터셋에 포함하고자 하는 변수명을 **선택변수명**에 지정한다.

## 데이터셋 활용[SET문, KEEP문] 예제

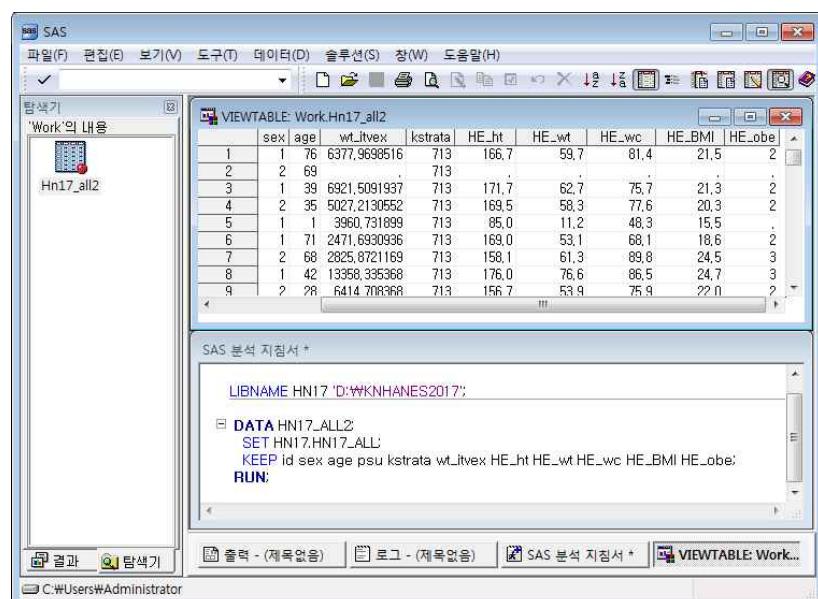
- 2017년 국민건강영양조사 기본 DB(HN17\_ALL)를 이용하여 신체계측 관련 변수로 구성된 DB(HN17\_ALL2)를 생성하고자 한다.

- 1) 영구 라이브러리인 HN17에 저장된 기본 DB(HN17\_ALL)를 SET 문을 이용하여 임시 라이브러리로 불러온다. 이때 DB에 신체계측 관련 변수만 존재하도록 신체 계측 관련 변수를 KEEP문에 지정한다.

```
DATA HN17_ALL2;
  SET HN17.HN17_ALL;
  KEEP id sex age psu kstrata wt_itvex HE_ht HE_wt HE_wc HE_BMI HE_obe;
RUN;
```

- 변수 요약

구분	내용
기존 데이터셋 (기본 DB)	HN17.HN17_ALL
선택변수 (신체계측 분석DB 관련 변수)	id, sex, age, psu, kstrata, wt_itvex, HE_ht, HE_wt, HE_wc, HE_BMI, HE_obe
새로운 데이터셋 (신체계측 분석DB)	HN17_ALL2





# 04

## 복합표본 자료분석

- 개요
- 자료분석 시 유의사항
- 복합표본 빈도분석
- 복합표본 평균분석
- 복합표본 회귀분석
- 복합표본 로지스틱 회귀분석

## □ 개요

- 국민건강영양조사는 복합표본설계되어 단순임의추출자료의 분석과는 다르게 복합 표본설계 요소(총, 군집, 가중치)를 반영하여 자료분석을 하도록 권고하고 있다.
- SAS에서 복합표본 자료분석을 위해 사용되는 프로시저는 SURVEYFREQ, SURVEYMEANS, SURVEYREG, SURVEYLOGISTIC이며, 각 프로시저에서는 복합표본설계 요소인 총, 집락, 가중치를 STRATA, CLUSTER, WEIGHT 문에 반드시 지정할 수 있도록 되어있다.

복합표본설계 SAS 프로시저	예제
PROC SURVEYFREQ	<pre>PROC SURVEYFREQ &lt;옵션&gt;;   STRATA 총변수;   CLUSTER 집락변수;   WEIGHT 가중치변수;   : RUN;</pre>
PROC SURVEYMEANS	<pre>PROC SURVEYMEANS &lt;옵션&gt;;   STRATA 총변수;   CLUSTER 집락변수;   WEIGHT 가중치변수;   : RUN;</pre>
PROC SURVEYREG	<pre>PROC SURVEYREG &lt;옵션&gt;;   STRATA 총변수;   CLUSTER 집락변수;   WEIGHT 가중치변수;   : RUN;</pre>
PROC SURVEYLOGISTIC	<pre>PROC SURVEYLOGISTIC &lt;옵션&gt;;   STRATA 총변수;   CLUSTER 집락변수;   WEIGHT 가중치변수;   : RUN;</pre>

- 다음 복합표본 분석에서 사용할 예제는 다음과 같다.

복합표본설계 SAS 프로시저	구분	예제
PROC SURVEYFREQ	빈도분석	만 30세 이상에서 성별에 따른 고혈압 유병률
	Chi-square 검정	만 30세 이상에서 성별과 고혈압 유병여부에 관한 연관성 검정
PROC SURVEYMEANS	기술통계	만 30세 이상 체질량지수(BMI)의 평균
PROC SURVEYREG	T검정	만 30세 이상에서 성별에 따른 수축기혈압의 평균차이 검정
	분산분석	만 30세 이상에서 연령그룹에 따른 수축기혈압의 평균차이 검정
	공분산분석	연령, BMI 보정시 만 30세 이상에서 성별에 따른 수축기혈압의 평균차이 검정
PROC SURVEYLOGISTIC	로지스틱 (연속형 독립변수)	만 30세 이상에서 체질량지수(BMI)가 고혈압 유병여부에 미치는 영향
	로지스틱 (범주형 독립변수)	만 30세 이상에서 비만 유병여부가 고혈압 유병여부에 미치는 영향
	다중로지스틱	만 30세 이상에서 연령, 현재 흡연여부, 비만 유병여부가 고혈압 유병여부에 미치는 영향

## □ 자료분석 시 유의사항

### 분석 유의사항 1 : SURVEY 프로시저에서 by, where문 사용시 추정치의 표준오차 편향 발생 => 집단변수 생성 후 관심집단 분석

국민건강영양조사의 표본은 복합표본설계방법인 다단계총화집락획를추출방법에 의해 선정되었고, 복합표본설계 요소인 층, 집락, 가중치 정보가 원시자료DB 전체에 포함되어 있다. 따라서 **자료분석 시 전체 자료가 아닌 일부자료만으로 분석DB를 생성할 경우**(즉, 비해당 자료를 분석DB에서 삭제하거나 by, where문 등으로 관심집단을 구분하는 경우) 삭제된 자료에 포함된 복합표본설계 정보가 누락됨에 따라 **추정치의 분산(표준오차)에 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)를 생성한 후, 생성된 관심집단 변수를 집단변수로 지정하여 분석한다. 복합표본 교차분석에서는 집단변수를 table문에, 복합표본 평균분석, 회귀분석, 로지스틱 회귀분석에서는 집단변수를 domain문에 지정하여 분석한다.

### 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 SAS 9.30이상 버전에서 SURVEY 프로시저를 이용한 자료분석 시에는 반드시 **NOMCAR(Not missing completely at random)<sup>8)</sup>** 옵션을 사용해야한다. 해당 옵션을 지정하지 않을 경우 분석변수에 결측인 자료의 표본설계정보가 누락되어 추정치의 분산추정량에 편향이 발생 할 수 있으므로, 옵션지정 후 분석한다.

8) Paul Gorrell. Survey Analysis: Options for Missing Data. Proceeding of Northeast SAS Users Group Conference, 2009.

## □ 복합표본 빈도분석

□ SURVEYFREQ 프로시저에서는 1차원 이상의 빈도표 및 백분율 등을 제공한다.

□ 기본형태

```
PROC SURVEYFREQ <옵션1>;
  STRATA 층변수;
  CLUSTER 집락변수;
  WEIGHT 가중치변수;
  TABLES 집단변수*분석변수 </옵션2>;
  RUN;
```

- <옵션1>

□ DATA= 데이터셋

분석할 SAS 데이터셋 이름을 **데이터셋**에 지정한다.

□ NOMCAR

분석변수에 결측인 자료의 표본설계정보가 누락되어 추정치의 분산추정량에 편향이 발생하지 않도록 **NOMCAR**옵션을 지정한다.

- STRATA 층변수

층변수명을 **층변수**에 지정한다.

- CLUSTER 집락변수

집락변수명을 **집락변수**에 지정한다.

- WEIGHT 가중치변수

가중치변수명을 **가중치변수**에 지정한다.

- TABLES 집단변수\*분석변수

관심있는 집단의 변수명과 분석하고자하는 변수명을 각각 집단변수와 분석변수에 지정한다.

## ▷ 옵션2&gt;

옵션	설명
CHISQ	Rao- Scott 카이제곱 독립성검정 결과 출력
LRCHISQ	우도비검정 결과 출력
WCHISQ	Wald검정 결과 출력
WLCHISQ	Wald로그선형검정 결과 출력
CELLCHI2	각 셀의 $(\text{실제빈도} - \text{기대빈도})^2 / (\text{기대빈도})$ 출력
CL	신뢰구간 출력
CLWT	가중빈도의 신뢰구간 출력
CV	변동계수 출력
CVWT	가중빈도의 변동계수 출력
DEFF	퍼센트의 설계효과 출력
EXPECTED	각 셀의 기대빈도 출력
DEVIATION	각 셀의 $(\text{실제빈도} - \text{기대빈도})$ 출력
PEARSONRES	Pearson 잔차 출력
COL	열 빈도 및 전체빈도 출력
ROW	행 빈도 및 전체빈도 출력
VAR	퍼센트의 분산 출력
VARWT	가중빈도의 분산 출력
NOCELLPERCENT	빈도를 출력하지 않음
NOFREQ	실제빈도를 출력하지 않음
NOPERCENT	셀 빈도 및 전체빈도를 출력하지 않음
NOPRINT	분할표를 출력하지 않음
NOSTD	모든 표준오차를 출력하지 않음
NOTOTAL	전체 total값을 출력하지 않음
NOWT	가중빈도 및 가중빈도 표준오차를 출력하지 않음

## 복합표본 빈도분석 예제

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 성별에 따른 고혈압 유병률을 산출하고, 만 30세 이상에서 성별과 고혈압 유병여부에 관한 연관성 검정을 하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_HP	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
	age	연령	연속형
	sex	성별	1. 남자 2. 여자
생성변수	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

※복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

※생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYFREQ DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
TABLES age30*sex*HE_HP / ROW CHISQ;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Data Summary	
<b>Number of Strata</b>	27
<b>Number of Clusters</b>	192
<b>Number of Observations</b>	8127
<b>Number of Observations Used</b>	7709
<b>Number of Obs with Nonpositive Weights</b>	418
<b>Sum of Weights</b>	51035252

▶ 2017년 기본 DB는 27개의 층(kstrata), 192개의 집락(psu)으로 구성되었고 국민건강영양조사에 참여한 8,127명 중 검진·건강설문조사에 참여한 대상자는 7,709명이다.

## ② 분할표

Table of sex by HE_HP								
Controlling for age30=1								
sex	HE_HP	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent
1	1	742	5617171	296218	16.1850	0.6599	33.0496	1.2586
2		719	5427976	255225	15.6399	0.5886	31.9364	1.1288
3		936	5951057	245000	17.1471	0.5930	35.0140	1.0867
	Total	2397	16996204	538292	48.9720	0.7011	100.000	
2	1	1366	8962175	411650	25.8232	0.7427	50.6060	1.3296
2	2	676	3864108	199365	11.1339	0.5014	21.8191	0.8888
2	3	1013	4883442	208553	14.0709	0.6338	27.5749	1.1980
	Total	3055	17709724	537037	51.0280	0.7011	100.000	
	Total	2108	14579346	627406	42.0082	1.0565		
	2	1395	9292083	350617	26.7738	0.7063		
	3	1949	10834498	351060	31.2180	0.9199		
	Total	5452	34705928	959158	100.000			

▶ 만 30세 이상에서의 고혈압 유병률은 전체 31.2%, 남자 35.0%, 여자 27.6%이다.

### ③ Rao-Scott 카이제곱 독립성 검정

$H_0$  : 만 30세 이상에서 성별에 따른 고혈압 유병여부는 차이가 없다.

$H_1$  : 만 30세 이상에서 성별에 따른 고혈압 유병여부는 차이가 있다.

Rao-Scott Chi-Square Test	
Pearson Chi-Square	176.2039
Design Correction	1.3647
Rao-Scott Chi-Square	129.1129
DF	2
Pr > ChiSq	<.0001
F Value	64.5565
Num DF	2
Den DF	330
Pr > F	<.0001
Sample Size = 7709	

- ▶ Rao-Scott Chi-Square는 129.1129, P-value는 <0.0001이므로 만 30세 이상에서 성별에 따른 고혈압 유병여부는 차이가 있는 것으로 나타났다.

## □ 복합표본 평균분석

- SURVEYMEANS 프로시저에서는 연속형 변수의 일변량 통계량(평균, 비율, 표준오차, 신뢰구간 추정 등)을 제공한다.

- 기본형태

```
PROC SURVEYMEANS <옵션1>;
  STRATA 층변수;
  CLUSTER 집락변수;
  WEIGHT 가중치변수;
  VAR 분석변수;
  DOMAIN 집단변수;
  RUN;
```

- <옵션1>

□ DATA= 데이터셋

분석할 SAS 데이터셋 이름을 **데이터셋**에 지정한다.

□ NOMCAR

분석변수에 결측인 자료의 표본설계정보가 누락되어 추정치의 분산추정량에 편향이 발생하지 않도록 **NOMCAR**옵션을 지정한다.

□ 통계량

통계량	설명
ALL	기하평균을 제외한 모든 통계량 출력
CLM	신뢰계수 출력
CV	변동계수 출력
MAX	최대값 출력
MEAN	평균 출력
MEDIAN	중앙값 출력
MIN	최소값 출력
NOBS	결측치 제외 관측값의 수 출력
Q1	1분위수 출력
Q3	3분위수 출력
STDEERR	표준오차 출력

- STRATA 층변수

층변수명을 **층변수**에 지정한다.

- CLUSTER 집락변수

집락변수명을 **집락변수**에 지정한다.

- WEIGHT 가중치변수

가중치변수명을 **가중치변수**에 지정한다.

- VAR 분석변수

구하고자 하는 평균의 변수명을 **분석변수**에 지정한다.

- DOMAIN 집단변수

관심있는 집단의 변수명을 **집단변수**에 지정한다.

※자료분석 시 유의사항 「분석 유의사항 1(p46)」 참조

## 복합표본 평균분석 예제

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상 체질량지수(BMI)의 평균을 산출하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_BMI	체질량지수(체중(kg) / 신장 <sup>2</sup> (m <sup>2</sup> ))	연속형
	age	연령	연속형
	sex	성별	1. 남자 2. 여자
생성변수	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

※복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문-검진조사 가중치(wt\_itvex)

※생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYMEANS DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
VAR HE_BMI;
DOMAIN age30 age30*sex;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Data Summary	
<b>Number of Strata</b>	27
<b>Number of Clusters</b>	192
<b>Number of Observations</b>	8127
<b>Number of Observations Used</b>	7709
<b>Number of Obs with Nonpositive Weights</b>	418
<b>Sum of Weights</b>	51035252

▶ 2017년 기본 DB는 27개의 층(kstrata), 192개의 집락(psu)으로 구성되었고 국민건강영양조사에 참여한 8,127명 중 검진·건강설문조사에 참여한 대상자는 7,709명이다.

## ② 기술통계량

Statistics								
Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean			
HE_BMI	체질량지수	7684	23.028175	0.076312	22.8775010		23.1788492	
Domain Statistics in age30								
age30	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean		
0	HE_BMI	체질량지수	2234	20.909722	0.159403	20.5949892	21.2244549	
1	HE_BMI	체질량지수	5450	24.014806	0.059472	23.8973823	24.1322291	
Domain Statistics in age30*sex								
age30	성별	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	
0	1	HE_BMI	체질량지수	1123	21.568429	0.210395	21.1530152	21.9838418
	2	HE_BMI	체질량지수	1111	20.184671	0.181636	19.8260406	20.5433022
1	1	HE_BMI	체질량지수	2399	24.468985	0.076605	24.3177329	24.6202370
	2	HE_BMI	체질량지수	3051	23.577839	0.091842	23.3965020	23.7591751

▶ 검진·건강설문조사 참여자 7,709명 중 HE\_BMI가 결측이 아닌 대상자 7,684명의 체질량지수(BMI) 평균은 23.00이다.

▶ 만 30세 이상 대상자 5,450명의 체질량지수(BMI) 평균은 24.00이고, 이 중 남자 2,399명의 체질량지수 평균은 24.5, 여자 3,051명의 체질량지수 평균은 23.60이다.

## 복합표본 회귀분석

- SURVEYREG 프로시저는 복합표본설계 자료에서 연속형 종속변수와 독립변수 사이의 연관성을 알아보기 위해 사용한다.
- 단순임의표본설계 자료분석 시 그룹간 평균차이 검정(T-검정, 분산분석, 공분산분석)을 할 경우 각 분석방법에 따른 프로시저를 이용하여 분석을 수행하지만, 복합표본설계 자료에서는 분석방법의 구분 없이 SURVEYREG 프로시저를 이용하여 분석한다.

종속변수	독립변수	분석방법	복합표본 자료분석
연속형	2그룹 범주형	T - 검정	PROC SURVEYREG
	3그룹 이상 범주형	분산분석	
	연속형, 범주형	회귀분석	
	연속형, 범주형	공분산분석	

- 또한, 종속변수에 어떤 영향을 줄 수 있는 연속형 변수를 통제하여 보정평균을 산출하거나, 연령표준화를 산출할 경우에도 SURVEYREG 프로시저를 이용하여 분석한다.

※연령표준화 산출방법은 「5장 연령표준화(p84~88)」 참조

- 기본형태

```
PROC SURVEYREG <옵션1>;
  STRATA 층변수;
  CLUSTER 집락변수;
  WEIGHT 가중치변수;
  CLASS 범주형독립변수;
  MODEL 종속변수=독립변수1 ... 독립변수K </옵션2>;
  DOMAIN 집단변수;

  *보정평균 산출시 LSMEANS문 또는 ESTIMATE문을 이용;
  LSMEANS 범주형독립변수 </옵션3>;
  ESTIMATE 독립변수1(divisor=n) ... 독립변수K(divisor=n) </옵션4>;
RUN;
```

- <옵션1>

□ DATA= 데이터셋

분석할 SAS 데이터셋 이름을 **데이터셋**에 지정한다.

□ NOMCAR

분석변수에 결측인 자료의 표본설계정보가 누락되어 추정치의 분산추정량에 편향이 발생하지 않도록 **NOMCAR**옵션을 지정한다.

- STRATA 층변수

층변수명을 **층변수**에 지정한다.

- CLUSTER 집락변수

집락변수명을 **집락변수**에 지정한다.

- WEIGHT 가중치변수

가중치변수명을 **가중치변수**에 지정한다.

- CLASS 범주형독립변수

모형(model)의 독립변수 중 범주형변수가 있는 경우 CLASS문에 **범주형독립변수**를 지정한다.

※독립변수 중 범주형변수에 속하는 변수를 CLASS 명령문에 지정하지 않고 분석할 경우 연속형변수로 가정하여 분석된다.

- MODEL 종속변수=독립변수1 독립변수2 ... 독립변수K

분석모형을 “**Model 종속변수=독립변수1 ... 독립변수K**”에 지정한다.

Model문에는 하나의 모형만 지정할 수 있으며 독립변수들의 순서는 분석결과에 영향을 미치지 않는다.

□<옵션2>

옵션	설명
ADJRSQ	수정 결정계수 값 출력
ANOVA	ANOVA테이블 출력
CLPARM	추정 회귀계수의 95% 신뢰구간 출력
COVB	추정 회귀계수의 공분산 출력
SOLUTION	추정 회귀계수 출력
STB	표준화된 추정 회귀계수 출력
VADJUST=NONE	분산추정시 자유도 미보정

### - DOMAIN 집단변수

관심있는 집단의 변수명을 **집단변수**에 지정한다.

※자료분석 시 유의사항 「분석 유의사항 1(p46)」 참조

### - LSMEANS 범주형독립변수

LSMEANS문은 보정평균 추정 및 검정과 그룹별 추정치간의 다중비교 분석시 사용된다. 모형(model)에 포함된 독립변수 중 결과를 보고자하는 집단변수(범주형) 또는 직접 비교를 원하는 독립변수(범주형)를 **범주형독립변수**에 지정한다.

이때 범주형 보정변수의 가중비율은 **OM옵션**을 지정하면 자동으로 계산되어 적용되지만, 연속형 보정변수의 가중평균은 직접 산출하여 **AT옵션**에 지정해야 한다.

#### ▷ 옵션3>

옵션	설명
ADJUST=검정법	그룹별 추정치 간의 다중비교(검정법:BON, SCHEFFE 등)
AT	연속형 보정변수 가중평균값 지정
CL	신뢰구간 출력
E	보정행렬L 출력
OM	범주형 변수의 가중비율 자동 적용

### - ESTIMATE 독립변수1(divisor=n) ... 독립변수K(divisor=n)

ESTIMATE문은 보정평균 추정 및 검정과 그룹별 추정치간의 다중비교 분석시 사용되며, 모형(model)에 포함된 독립변수들을 “**독립변수1...독립변수K**”에 지정한다. LSMEANS문과 달리 가중평균뿐만 아니라 가중비율도 직접 지정할 수 있으며, 보정변수가 연속형 변수일 경우 가중평균을, 범주형 변수일 경우 가중비율을 산출하여 Estimate문에 지정해야 한다.

#### ▷ 옵션4>

옵션	설명
ADJUST=검정법	그룹별 추정치 간의 다중비교(검정법:BON, SCHEFFE 등)
CL	신뢰구간 출력
E	보정행렬L 출력

## 복합표본 회귀분석 예제1 : T검정

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 성별에 따른 수축기혈압의 평균차이 검정을 하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_sbp	수축기혈압(mmHg)	연속형
	age	연령	연속형
	sex	성별	1. 남자 2. 여자
생성변수	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

※복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

※생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYREG DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS sex;
MODEL HE_sbp=sex / SOLUTION CLPARM VADJUST=none;
DOMAIN age30;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Domain Summary	
Number of Observations	8127
Number of Obs with Nonpositive Weights	418
Number of Observations in Domain	5465
Number of Observations Not in Domain	2244
Sum of Weights in Domain	34814977
Weighted Mean of HE_sbp	118.81680
Weighted Sum of HE_sbp	4126225194

▶ 2017년 국민건강영양조사 검진·건강설문조사에 참여한 대상자 중 만 30세 이상인 대상자는 5,465명이다.

## ② 회귀모형과 회귀계수 추정에 대한 유의성 검정

$H_0$  : 만 30세 이상에서 성별에 따른 수축기혈압 평균은 차이가 없다.

$H_1$  : 만 30세 이상에서 성별에 따른 수축기혈압 평균은 차이가 있다.

Fit Statistics		Tests of Model Effects			
Effect	Num DF	F Value	Pr > F		
R-Square	0.01209				
Adjusted R-Square	0.01196				
Root MSE	16.2528				
Denominator DF	165				

Estimated Regression Coefficients						
Parameter	Estimate	Standard Error	t Value	Pr >  t	95% Confidence Interval	
Intercept	117.054741	0.48581311	240.95	<.0001	116.095529	118.013952
sex 1	3.595746	0.50718751	7.09	<.0001	2.594332	4.597160
sex 2	0.000000	0.00000000	.	.	0.000000	0.000000

▶ 추정모형 :  $E(HE_{sbp}) = 117.05 + 3.60 \times 1 (sex=1)$

▶ 모형의 설명력은 P-value < 0.0001로 매우 유의하고,  $R^2$ 은 0.012로 성별이 수축기혈압 변동 중 약 1.2% 설명이 가능하다.

▶ 회귀계수의 추정값을 검정해 보면 유의확률이 P-value < 0.0001로 매우 유의하게 나타났다. 만 30세 이상에서 성별에 따른 수축기혈압의 평균은 남자는 120.65( $117.05 + 3.60$ ) mmHg, 여자는 117.05( $117.05 + 0.00$ ) mmHg이며, 여자 대비 남자가 3.60mmHg 높은 것으로 나타났다.

## 복합표본 회귀분석 예제2 : 분산분석

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 연령그룹에 따른 수축기혈압의 평균차이 검정을 하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_sbp	수축기혈압(mmHg)	연속형
	age	연령	연속형
생성변수	age_g	10세단위 연령구분	3. 30~39세 4. 40~49세 5. 50~59세 6. 60~69세 7. 70세이상
	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

\*복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문-검진조사 가중치(wt\_itvex)

\*생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYREG DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS age_g;
MODEL HE_sbp=age_g / SOLUTION CLPARM VADJUST=none;
DOMAIN age30;
RUN;
```

## 분석 결과

## ① 회귀모형과 회귀계수 추정에 대한 유의성 검정

$H_0$  : 만 30세 이상에서 연령그룹에 따른 수축기혈압 평균은 차이가 없다.  
 $H_1$  : 만 30세 이상에서 연령그룹에 따른 수축기혈압 평균은 차이가 있다.

Fit Statistics		Tests of Model Effects			
	R-Square	Effect	Num DF	F Value	Pr > F
	0.1295	Model	4	110.08	<.0001
Root MSE	15.2593	Intercept	1	115285	<.0001
Denominator DF	165	age_g	4	110.08	<.0001

Estimated Regression Coefficients						
Parameter	Estimate	Standard Error	t Value	Pr >  t	95% Confidence Interval	
Intercept	129.010027	0.73782552	174.85	<.0001	127.553231	130.466824
age_g 3	-17.269030	0.87747896	-19.68	<.0001	-19.001564	-15.536495
age_g 4	-14.227536	0.90604698	-15.70	<.0001	-16.016476	-12.438595
age_g 5	-9.259634	0.80751625	-11.47	<.0001	-10.854031	-7.665237
age_g 6	-4.427132	0.82015544	-5.40	<.0001	-6.046485	-2.807780
age_g 7	0.000000	0.00000000	.	.	0.000000	0.000000

- ▶ 추정모형 :  $E(HE_{sbp}) = 129.01 - 17.27 \times I(age_g=3) - 14.23 \times I(age_g=4) - 9.26 \times I(age_g=5) - 4.43 \times I(age_g=6)$
- ▶ 모형의 설명력은 P-value < 0.0001로 매우 유의하고,  $R^2$ 은 0.1295로 연령그룹이 수축기혈압의 변동 중 약 13.0% 설명이 가능하다.
- ▶ 회귀계수의 추정값을 검정해 보면 유의확률이 P-value < 0.0001로 매우 유의하게 나타났다. 만 30세 이상에서 연령그룹에 따른 수축기혈압의 평균은 70세 이상( $age_g=7$ )에 비해 연령이 낮을수록 수축기혈압의 평균이 감소하는 것으로 나타났다.  
 : 70대이상 129.01mmHg, 30대는 111.74(129.01 - 17.27) mmHg

### 복합표본 회귀분석 예제3 : 공분산분석, 보정평균

□ 2017년 국민건강영양조사 자료를 이용하여 연령그룹, **BMI**로 보정 시, 만 30세 이상에서 성별에 따른 수축기혈압의 평균차이 검정을 하고, 연령그룹, **BMI**로 보정 시, 만 30세 이상에서 성별에 따른 수축기혈압 보정평균을 산출하고자 한다.

분석 변수

구분	변수명	변수설명	내용
기준변수	HE_sbp	수축기혈압(mmHg)	연속형
	HE_BMI	체질량지수(체중(kg) / 신장 <sup>2</sup> (m <sup>2</sup> ))	연속형
	age	연령	연속형
	sex	성별	1. 남자 2. 여자
생성변수	age_g	10세단위 연령구분	3. 30~39세 4. 40~49세 5. 50~59세 6. 60~69세 7. 70세이상
	age30	30세 이상	0. 30세 미만 1. 30세 이상

※복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

※생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

## 분석 프로그램

## ① 공분산분석

```
PROC SURVEYREG DATA=ALL17 NOMCAR;
  STRATA kstrata;
  CLUSTER PSU;
  WEIGHT wt_itvex;
  CLASS sex age_g;
  MODEL HE_sbp=sex age_g HE_BMI / SOLUTION CLPARM VADJUST=none;
  DOMAIN age30;
RUN;
```

## ② 보정평균1(LSMEANS문 이용)

```
PROC SURVEYREG DATA=ALL17 NOMCAR;
  STRATA kstrata;
  CLUSTER PSU;
  WEIGHT wt_itvex;
  CLASS sex age_g;
  MODEL HE_sbp=sex age_g HE_BMI / SOLUTION CLPARM VADJUST=none;
  DOMAIN age30;
  LSMEANS sex / AT HE_BMI=24.010027 OM E CL DIFF ADJUST=BON;
RUN;
```

## ③ 보정평균2(ESTIMATE문 이용)

```
PROC SURVEYREG DATA=ALL17 NOMCAR;
  STRATA kstrata;
  CLUSTER PSU;
  WEIGHT wt_itvex;
  CLASS sex age_g;
  MODEL HE_sbp=sex age_g HE_BMI / SOLUTION CLPARM VADJUST=none;
  DOMAIN age30;
  ESTIMATE '남자' intercept 1 sex 1 0 age_g 0.216999 0.247302 0.241752
            0.156981 0.136966 HE_BMI 24.010027,
            '여자' intercept 1 sex 0 1 age_g 0.216999 0.247302 0.241752
            0.156981 0.136966 HE_BMI 24.010027 / E CL ADJUST=BON;
RUN;
```

※LSMEANS문과 ESTIMATE문에 사용된 가중평균, 가중비율은 부록의 「예제변수 생성 및 산출 : 가중평균, 가중비율 산출(p93~94)」 참고





## □ 복합표본 로지스틱 회귀분석

- SURVEYLOGISTIC 프로시저는 복합표본설계 자료에서 범주형 종속변수와 독립변수 사이의 연관성을 알아보기 위해 사용한다.
- 복합표본 로지스틱 회귀분석을 통해 종속변수가 2개범주(이분형), 3개이상 범주 또는 순서형 범주에 대한 분석이 가능하며, 종속변수에 영향을 줄 수 있는 보정변수를 반영하여 분석할 경우에도 SURVEYLOGISTIC 프로시저를 이용하여 분석한다.

독립변수	종속변수	분석방법	복합표본설계 SAS 분석
연속형 범주형	명목형(이분형)	이분형 로지스틱 회귀분석	PROC SURVEYLOGISTIC
	명목형(3범주이상)	다중명목 로지스틱 회귀분석	

- 기본형태

```

PROC SURVEYLOGISTIC <옵션1>;
  STRATA 층변수;
  CLUSTER 집락변수;
  WEIGHT 가중치변수;
  CLASS 범주형독립변수 </옵션2>;
  MODEL 종속변수<옵션3>=독립변수1 ... 독립변수K </옵션4>;
  DOMAIN 집단변수;

*보정분을 산출시 LSMEANS문 또는 ESTIMATE문을 이용;
  LSMEANS 범주형독립변수 </옵션5>;
  ESTIMATE 독립변수1(divisor=n) ... 독립변수K(divisor=n) </옵션6>;
RUN;

```

- <옵션1>

□ DATA= 데이터셋

분석할 SAS 데이터셋 이름을 **데이터셋**에 지정한다.

## □ NOMCAR

분석변수에 결측인 자료의 표본설계정보가 누락되어 추정치의 분산추정량에 편향이 발생하지 않도록 **NOMCAR**옵션을 지정한다.

- STRATA 총변수

총변수명을 **총변수**에 지정한다.

- CLUSTER 집락변수

집락변수명을 **집락변수**에 지정한다.

- WEIGHT 가중치변수

가중치변수명을 **가중치변수**에 지정한다.

- CLASS 범주형독립변수

모형(model)의 독립변수 중 범주형변수가 있는 경우 CLASS문에 **범주형독립변수**를 지정한다.

※독립변수 중 범주형변수에 속하는 변수를 CLASS 명령문에 지정하지 않고 분석할 경우 연속형변수로 가정하여 분석된다.

## ▣ 옵션2>

옵션2를 생략할 경우 기준변수는 가장 높은 범주를 기준으로 자동분석되며 기준을 변경하고자 할 경우 “Param=Ref Ref=기준값” 옵션을 사용한다.

- MODEL 종속변수</옵션3> = 독립변수1 독립변수2 … 독립변수K </옵션4>

분석모형을 “**Model 종속변수= 독립변수1 … 독립변수K**”에 지정한다.

Model문에는 하나의 모형만 지정할 수 있으며 독립변수들의 순서는 분석결과에 영향을 미치지 않는다.

## ▣ 옵션3>

종속변수의 옵션으로 descending, event, order 등을 사용한다. 로지스틱 회귀모형의 기준(관심확률) 변경은 “model 종속변수명(descending)” 또는 “model 종속변수명(event='1')”로 지정한다.

### ▷ 옵션4>

옵션	설명
VADJUST=NONE	분산추정시 자유도 미보정
CLPARM	추정 계수 신뢰구간 출력
CLODDS	오즈비 신뢰구간 출력
CORRB	추정 계수의 상관계수 행렬 출력
COVB	추정 계수의 공분산 행렬 출력
DF=INFINITY	카이제곱검정 결과 출력
RSQUARE	결정계수 값 출력
ST B	표준화된 추정 계수 출력

- DOMAIN 집단변수

관심있는 집단의 변수명을 **집단변수**에 지정한다.

※자료분석 시 유의사항 「분석 유의사항 1(p46)」 참조

- LSMEANS 범주형독립변수

LSMEANS문은 보정평균 추정 및 검정과 그룹별 추정치간의 다중비교 분석시 사용된다. 모형(model)에 포함된 독립변수 중 결과를 보고자하는 집단변수(범주형) 또는 직접 비교를 원하는 독립변수(범주형)를 **범주형독립변수**에 지정한다.

이때 범주형 보정변수의 가중비율은 **OM옵션**을 지정하면 자동으로 계산되어 적용되지만, 연속형 보정변수의 가중평균은 직접 산출하여 **AT옵션**에 지정해야 한다.

### ▷ 옵션5>

옵션	설명
ADJUST=검정법	그룹별 추정치 간의 다중비교(검정법:BON, SCHEFFE 등)
AT	연속형 보정변수 가중평균값 지정
CL	신뢰구간 출력
E	보정행렬L 출력
OM	범주형 보정변수의 가중분율 자동 적용
ILINK	보정분율, 표준오차 재계산

- ESTIMATE 독립변수1(divisor=n) ... 독립변수K(divisor=n) </옵션6>

ESTIMATE문은 보정평균 추정 및 검정과 그룹별 추정치간의 다중비교 분석시 사용되며, 모형(model)에 포함된 독립변수들을 “**독립변수1...독립변수K**”에 지정한다. LSMEANS문과 달리 가중평균뿐만 아니라 가중비율도 직접 지정할 수 있으며, 보정변수가 연속형 변수일 경우 가중평균을, 범주형 변수일 경우 가중비율을 산출하여 Estimate문에 지정해야 한다.

**▷ 옵션6>**

옵션	설명
ADJUST = 검정법	그룹별 추정치 간의 다중비교(검정법:BON, SCHEFFE 등)
CL	신뢰구간 출력
E	보정행렬 L 출력
LINK	보정분율, 표준오차 재계산

## 복합표본 로지스틱 회귀분석 예제1 : 연속형 독립변수

□ 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 체질량지수가 고혈압 유병여부에 미치는 영향을 알아보고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_hp	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
	HE_BMI	체질량지수(체중(kg) / 신장 <sup>2</sup> (m <sup>2</sup> ))	연속형
생성변수	hp1	고혈압 유병여부 재분류	1. 고혈압 0. 정상, 고혈압전단계
	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

\*복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

\*생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYLOGISTIC DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
DOMAIN age30;
MODEL hp1(EVENT='1')=HE_BMI / VADJUST=none DF=INFINITY;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Domain Summary	
Number of Observations	8127
Number of Obs with Nonpositive Weights	418
Number of Observations in Domain	5465
Number of Observations not in Domain	2244
Sum of Weights in Domain	34814977

▶ 2017년 국민건강영양조사 검진·건강설문조사에 참여한 대상자 중 만 30세 이상인 대상자는 5,465명이다.

## ② 회귀모형과 회귀계수 추정에 대한 유의성 검정, 오즈비

Testing Global Null Hypothesis: BETA=0				
Test	Chi-Square	DF	Pr > ChiSq	
Likelihood Ratio	1763599,18	1	<.0001	
Score	177,5037	1	<.0001	
Wald	175,7700	1	<.0001	

NOTE: First-order Rao-Scott design correction 1.1882 applied to the likelihood ratio test.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4,6217	0,3012	235,4878	<.0001
HE_BMI	1	0,1575	0,0119	175,7700	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
HE_BMI	1,171	1,144	1,198

- ▶ 모형의 설명력은 P-value<0.0001로 체질량지수가 고혈압 유병여부와 유의한 연관성이 있는 것으로 나타났다.
- ▶ 회귀계수의 추정값을 검정해 보면  $\beta>0$ 로 체질량지수가 증가할수록 고혈압 유병 가능성이 높다.
- ▶ 만 30세 이상에서 체질량지수가 1단위 증가할 때마다 고혈압 유병 오즈비가 1.17배 높다.

## 복합표본 로지스틱 회귀분석 예제2 : 범주형 독립변수

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 비만 유병여부가 고혈압 유병여부에 미치는 영향을 알아보고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_obe	비만 유병여부	1. 저체중 2. 정상 3. 비만
	HE_hp	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
	age	연령	연속형
생성변수	obe	비만 유병여부	1. 정상, 저체중 2. 비만
	hp1	고혈압 유병여부	1. 고혈압 0. 정상, 고혈압전단계
	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

\*복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문-검진조사 가중치(wt\_itvex)

\*생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYLOGISTIC DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS obe(PARAM=REF REF='1');
DOMAIN age30;
MODEL hp1(EVENT='1')=obe / VADJUST=none DF=INFINITY;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Domain Summary	
Number of Observations	8127
Number of Obs with Nonpositive Weights	418
Number of Observations in Domain	5465
Number of Observations not in Domain	2244
Sum of Weights in Domain	34814977

Number of Observations Read	8127
Number of Observations Used	3995
Sum of Weights Read	34814977
Sum of Weights Used	22097965

Response Profile			Class Level Information		
Ordered Value	hp1	Total Frequency	Total Weight	Class	Value
1	0	2480	16846518	obe	1
2	1	993	5251447		2

- ▶ 2017년 국민건강영양조사 검진·건강설문조사에 참여한 대상자 중 만 30세 이상인 대상자는 5,465명이다. 이 중 종속변수인 고혈압 유병여부(hp1)의 값이 존재하는 대상자는 3,473(2,480+993)명이고, 고혈압 유병자는 993명이다.
- ▶ 범주형 독립변수인 비만 유병여부(obe)에서 ‘정상, 저체중’을 기준으로 되어있다.

## ② 회귀모형과 회귀계수 추정에 대한 유의성 검정, 오즈비

Testing Global Null Hypothesis: BETA=0					
Test	Chi-Square		DF	Pr > ChiSq	
Likelihood Ratio	221545.068		1	<.0001	
Score	27.5335		1	<.0001	
Wald	28.8691		1	<.0001	

NOTE: First-order Rao-Scott design correction 1.2657 applied to the likelihood ratio test.

Type 3 Analysis of Effects					
Effect	DF	Wald Chi-Square	Pr > ChiSq	Wald Chi-Square	Pr > ChiSq
obe	1	28.8691	<.0001		

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.3808	0.0720	368.0005	<.0001
obe	2	1 0.5404	0.1006	28.8691	<.0001

Odds Ratio Estimates					
Effect	Point Estimate	95% Wald Confidence Limits			
obe 2 vs 1	1.717	1.410 2.091			

- ▶ 모형의 설명력은 P-value<0.0001로 비만유병여부가 고혈압 유병여부와 유의한 연관성이 있는 것으로 나타났다.
- ▶ 회귀계수의 추정값을 검정해 보면 ‘정상, 저체중’에 비해 ‘비만’인 경우 유의률이 P-value<0.0001로 매우 유의하게 나타났고,  $\beta>0$ 로 ‘정상, 저체중’에 비해 ‘비만’인 경우 고혈압 유병 가능성이 높다.
- ▶ 만 30세 이상에서 ‘정상, 저체중’에 비해 ‘비만’의 경우 고혈압 유병가능성은 1.72배 높다.

## 복합표본 다중 로지스틱 회귀분석 예제

- 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 연령, 현재 흡연 여부로 보정 시, 비만 유병여부가 고혈압 유병여부에 미치는 영향에 대해 알아보자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_obe	비만 유병여부	1. 저체중 2. 정상 3. 비만
	HE_hp	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
	age	연령	연속형
생성변수	sm1	현재 흡연여부	0. 비흡연 1. 흡연
	obe	비만 유병여부	1. 정상, 저체중 2. 비만
	hp1	고혈압 유병여부	1. 고혈압 유병 0. 정상, 고혈압전단계
	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

\*복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문-검진조사 가중치(wt\_itvex)

\*생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

```
PROC SURVEYLOGISTIC DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS obe(REF='1') sm1(REF='0')/ PARAM=ref;
DOMAIN age30;
MODEL hp1(EVENT='1')=obe age sm1 / VADJUST=none DF=INFINITY;
RUN;
```

## 분석 결과

## ① DATA SUMMARY

Domain Summary			
Number of Observations			8127
Number of Obs with Nonpositive Weights			418
Number of Observations in Domain			5465
Number of Observations not in Domain			2244
Sum of Weights in Domain			34814977
Number of Observations Read			8127
Number of Observations Used			3435
Sum of Weights Read			34814977
Sum of Weights Used			21851572
Response Profile			
Ordered Value	hp1	Total Frequency	Total Weight
1	0	2454	16654485
2	1	981	5197087
Class Level Information			
Class	Value	Design Variables	
obe	1	0	
	2	1	
sm1	0	0	
	1	1	

▶ 2017년 국민건강영양조사 검진·건강설문조사에 참여한 대상자 중 만 30세 이상인 대상자는 5,465명이다. 이 중 종속변수인 고혈압 유병여부 (hp1)의 값이 존재하는 대상자는 3,435(2,454+ 981)명이고, 고혈압 유병자는 981명이다.

▶ 범주형 독립변수인 ‘비만 유병여부(obe)’과 ‘현재 흡연여부(sm1)’에서 ‘정상, 저체중’과 ‘비흡연’을 기준으로 설정되어있다.

## ② 회귀모형과 회귀계수 추정에 대한 유의성 검정, 오즈비

Testing Global Null Hypothesis: BETA=0				
Test	Chi-Square		DF	Pr > ChiSq
Likelihood Ratio	4249388.27		2.9183	<.0001
Score	450.9270		3	<.0001
Wald	480.6870		3	<.0001

NOTE: Second-order Rao-Scott design correction 0.0280 applied to the Likelihood Ratio test.

Type 3 Analysis of Effects				
Effect	DF	Wald Chi-Square	Pr > ChiSq	
obe	1	19.8220	<.0001	
age	1	440.8673	<.0001	
sm1	1	12.9561	0.0003	

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.3127	0.2619	581.0950	<.0001
obe	2	0.5044	0.1133	19.8220	<.0001
age	1	0.0869	0.00414	440.8673	<.0001
sm1	1	0.4702	0.1306	12.9561	0.0003

Odds Ratio Estimates				
Effect	Point Estimate	95% Wald Confidence Limits		
obe 2 vs 1	1.656	1.326	2.068	
age	1.091	1.082	1.100	
sm1 1 vs 0	1.600	1.239	2.067	

- ▶ 모형의 설명력은 P-value<0.0001로 연령, 현재 흡연여부 보정 시, 비만 유병여부가 고혈압 유병여부와 유의한 연관성이 있는 것으로 나타났다.
- ▶ 회귀계수의 추정값을 검정해 보면 연령, 현재 흡연여부로 보정 시, ‘정상, 저체중’에 비해 ‘비만’에서의 유의확률이 P-value<0.0001로 매우 유의하게 나타났다.
- ▶ 만 30세 이상에서 연령, 현재 흡연여부로 보정 시, ‘정상그룹’에 비해 ‘비만그룹’의 고혈압 유병오즈비가 1.656배 높다.

## 복합표본 로지스틱 회귀분석 예제 : 보정분율

□ 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 연령, 현재 흡연 여부로 보정 시, 비만 유병여부에 따른 고혈압 유병률을 산출하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기준변수	HE_obe	비만 유병여부	1. 저체중 2. 정상 3. 비만
	HE_hp	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
	age	연령	연속형
생성변수	sm1	현재 흡연여부	0. 비흡연 1. 흡연
	obe	비만 유병여부	1. 정상, 저체중 2. 비만
	hp1	고혈압 유병여부	1. 고혈압 0. 정상, 고혈압전단계
	age30	30세를 기준으로 연령 구분	0. 30세 미만 1. 30세 이상

\*복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

\*생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

### 분석 프로그램

#### ① 보정분율1(LSMEANS문 이용)

```
PROC SURVEYLOGISTIC DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS obe(REF='1') sm1(REF='0')/ PARAM=glm;
DOMAIN age30;
MODEL hp1(EVENT='1')=obe age sm1 / VADJUST=none DF=INFINITY;
LSMEANS obe / AT age=52.055618 OM E CL ILINK ADJUST=BON;
RUN;
```

#### ② 보정분율2(ESTIMATE문 이용)

```
PROC SURVEYLOGISTIC DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER PSU;
WEIGHT wt_itvex;
CLASS obe(REF='1') sm1(REF='0')/ PARAM=glm;
DOMAIN age30;
MODEL hp1(EVENT='1')=obe age sm1 / VADJUST=none DF=INFINITY;
ESTIMATE '비만' intercept 1 obe 1 0 sm1 0.193524 0.806476 age 52.055618,
          '정상' intercept 1 obe 0 1 sm1 0.193524 0.806476 age 52.055618
          / E CL ILINK ADJUST=BON;
RUN;
```

※ LSMEANS문, ESTIMATE문의 경우 'Param=glm' 을 지정할 수 있다.

※ LSMEANS문과 ESTIMATE문에 사용된 기중평균, 기중비율은 부록의 「예제변수 생성 및 산출 : 기중평균, 기중비율 산출(p93~94)」 참고

## 분석 결과

## ① 보정분율1(LSMEANS문 이용) 산출결과

Coefficients for obe Least Squares Means Using WORK.ALL17 Margins At age=52.05562													
Parameter				obe		sm1		Row1		Row2			
Intercept: $hpi=0$									1		1		
obe 2				2						1			
obe 1				1							1		
만나이									52.056		52.056		
sm1 1						1			0.1935		0.1935		
sm1 0						0			0.8065		0.8065		

obe Least Squares Means													
obe	Margins	age	Estimate	Standard Error	z Value	Pr >  z	Alpha	Lower	Upper	Mean	Standard Error of Mean	Lower Mean	Upper Mean
2	WORK.ALL17	52.06	-1.1944	0.1017	-11.75	<.0001	0.05	-1.3936	-0.9951	0.2325	0.01814	0.1988	0.2699
1	WORK.ALL17	52.06	-1.6988	0.07575	-22.43	<.0001	0.05	-1.8473	-1.5504	0.1546	0.009901	0.1362	0.1750

Differences of obe Least Squares Means Adjustment for Multiple Comparisons: Bonferroni													
obe	_obe	Margins	age	Estimate	Standard Error	z Value	Pr >  z	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper
2	1	WORK.ALL17	52.06	0.5044	0.1133	4.45	<.0001	<.0001	0.05	0.2824	0.7265	0.2824	0.7265

## ② 보정분율2(ESTIMATE문 이용) 산출결과

Estimate Coefficients													
Parameter				obe		sm1		Row1		Row2			
Intercept: $hpi=0$									1		1		
obe 2				2						1			
obe 1				1							1		
만나이								52.056	52.056				
sm1 1						1		0.1935	0.1935				
sm1 0						0		0.8065	0.8065				

Estimates Adjustment for Multiplicity: Bonferroni														
Label	Estimate	Standard Error	z Value	Pr >  z	Adj P	Alpha	Lower	Upper	Adj Lower	Adj Upper	Mean	Standard Error of Mean	Lower Mean	Upper Mean
비만	-1.1944	0.1017	-11.75	<.0001	<.0001	0.05	-1.3936	-0.9951	-1.4223	-0.9665	0.2325	0.01814	0.1988	0.2699
정상	-1.6988	0.07575	-22.43	<.0001	<.0001	0.05	-1.8473	-1.5504	-1.8686	-1.5290	0.1546	0.009901	0.1362	0.1750

## □ 보정분율1, 2에 대한 결과해석

연령, 현재 흡연여부 보정 시, 만 30세 이상에서 비만 유병여부에 따른 고혈압 유병률은 정상, 저체중인 경우 15.5%, 비만인 경우 23.3%로 나타났다.



05

---

## 연령표준화

## □ 연령표준화 배경<sup>9)</sup>

- 연령은 보건통계에서 가장 중요한 공변량 중의 하나로 동일시점에서 연령분포가 다른 집단을 비교하거나, 사망질환유병 여부처럼 연령과 분석변수가 밀접한 연관성이 있는 경우 연령효과와 원인변수의 효과를 구분하여 해석할 필요가 있다.
- 이러한 경우 결과의 연령표준화를 통해 연령효과를 보정한 후, 결과변수에 대한 원인 변수 효과를 검토해야 한다.

## □ 연령표준화 방법

- 표준인구
  - 조율(crude rate)을 표준인구 비율로 보정하여 표준화율(age-adjusted rate)을 산출하며, 표준인구 자료의 종류 및 연령구분에 따라 상이한 결과가 산출되므로 연 구목적에 적합한 표준인구 선택 및 연령구분이 필요하다.
  - 「2016년 국민건강통계」 산출 시에는 2005년 추계인구(통계청 「장래인구추계」, 2016년 12월 공표)를 표준인구로 사용하였다.

### 〈국민건강영양조사 표준인구〉

연령(특정연령이상)	인구수	연령(10세간격)	인구수	비율
1세 이상	47,735,012	01~09세	5,308,304	0.111
10세 이상	42,426,708	10~18세	5,915,170	0.124
19세 이상	36,511,538	19~29세	8,262,905	0.173
20세 이상	35,881,240	30~39세	8,627,773	0.181
30세 이상	28,248,633	40~49세	8,206,397	0.172
40세 이상	19,620,860	50~59세	5,147,501	0.108
50세 이상	11,414,463	60~69세	3,635,784	0.076
		70세이상	2,631,178	0.055
		계	47,735,012	1.000

\*2005년 추계인구, 통계청 장래인구추계(2016년 12월 공표)

9) [http://www.cdc.gov/nchs/tutorials/nhanes/NHANESAnalyses/AgeStandardization/age\\_standardization\\_intro.htm](http://www.cdc.gov/nchs/tutorials/nhanes/NHANESAnalyses/AgeStandardization/age_standardization_intro.htm)

□ 기준 연령대

- 국민건강영양조사 보고서에서는 10세 간격(1~9세, 10~18세, 19~29세, 30~39세, …; 60~69세, 70세 이상)으로 연령대를 구분하였다.
- 10세단위보다 더 세분화된 연령구분(5세단위 등)을 사용할 경우 연령보정 효과가 더 클 수는 있으나, 연령대별 자료수가 작은 경우 해당 연령대 결과의 신뢰성이 낮을 수 있으므로 분석변수의 자료수를 검토 후 구분하였다.

□ 표준화 유병률 및 표준오차 산출식

- 표준화 유병률 및 표준화 표준오차는 연령별 유병률과 표준인구비율을 반영하여 산출하였다.

$$\text{표준화 유병률} = \sum_{i=1}^n \omega_i \times \hat{p}_i = \omega^T \times \hat{p}$$

$$\text{표준화 유병률 표준오차} = \sqrt{\omega^T \text{cov}(\hat{p}) \omega}$$

$\omega_i$  : 연령그룹  $i$ 의 표준인구 비율,  $\omega$  : 표준인구 비율 벡터,

$\hat{p}_i$  : 연령그룹  $i$ 의 유병률 추정치,  $\hat{p}$  : 유병률 추정치 벡터,

$\text{cov}(\hat{p})$  : 유병률 추정치의 공분산 행렬

□ 연령표준화 산출 방법

- 산출지표의 분모가 조사대상자인 경우 또는 건강설문조사와 영양조사에서 특정집단으로 제한된 경우, 연령대별 표준인구(2005년 추계인구)에 근거하여 표준화율을 산출하였다.

\*건강설문조사 지표의 경우 제4기(2007~2009) 조사항목 변경으로 2005년 조율 및 특정집단으로 제한된 표준인구 비율을 산출할 수 없는 경우가 많으므로 분모가 조사대상자인 경우와 동일하게 산출

〈 분모가 조사대상자인 경우 표준화 : 2017년 고혈압 표준화 유병률(만30세이상) 〉

(단위: 명, %)

기준 연령	표준인구	표준인구 합	유병률	표준 오차	기대인구 (추정유병자)	표준인구 비율	표준인구비율 기준 유병률	표준화 유병률
	①	②	③	④	①×③	①/②	①/②×③	
30~39	8,627,773	28,248,633	11.3	1.2	974,938	0.305422673	3.451276205	①/②×③ 의 합
40~49	8,206,397	28,248,633	19.2	1.4	1,575,628	0.290505987	5.577714943	
50~59	5,147,501	28,248,633	32.1	1.5	1,652,348	0.182221242	5.849301879	
60~69	3,635,784	28,248,633	46.9	1.9	1,705,183	0.128706547	6.036337036	
70+	2,631,178	28,248,633	64.7	1.8	1,702,372	0.093143551	6.026387776	
<b>소계</b>	<b>28,248,633</b>				<b>7,610,469</b>		26.94101784	
<b>※분모가 조사대상자인 경우 표준인구 :</b> 2005년 추계인구, 통계청 장래인구추계(2016년 12월 공표)								
								<b>26.9</b>

- 표준화 유병률 및 표준오차는 SAS의 SURVEYREG 프로시저를 사용하여 산출가능하며, 다음 ‘연령표준화율 산출 프로그램 예제’를 통해 SAS를 이용한 표준화 유병률 및 표준오차를 구하고자 한다.

## 연령표준화 예제

□ 2017년 국민건강영양조사 자료를 이용하여 만 30세 이상에서 고혈압 표준화 유병률을 산출하고자 한다.

### 분석 변수

구분	변수명	변수설명	내용
기존변수	HE_hp	고혈압 유병여부	1. 정상 2. 고혈압전단계 3. 고혈압
생성변수	age_g	10세단위 연령구분	3. 30~39세 4. 40~49세 5. 50~59세 6. 60~69세 7. 70세이상
	age30	30세를 기준으로 연령 구분	1. 30세 미만 2. 30세 이상
	hp	고혈압 유병여부	0. 정상, 고혈압전단계 1. 고혈압
	hp1	표준화를 산출을 하고자 30세 이상 연령에만 해당하는 고혈압 유병여부(hp) 재코딩	0. 정상 100. 고혈압

※복합표본설계 요소 변수 : 분산추정 층(kstrata), 조사구(psu), 건강설문–검진조사 가중치(wt\_itvex)

※생성변수는 부록의 「예제변수 생성 및 산출 : 예제변수 생성(p90~92)」 참조

## 분석 프로그램

```

PROC SURVEYREG DATA=ALL17 NOMCAR;
STRATA kstrata;
CLUSTER psu;
WEIGHT wt_itvex;
CLASS age_g;
DOMAIN age30;
MODEL hp1=age_g / NOINT VADJUST=NONE;
ESTIMATE '고혈압 표준화 유병률'
    age_g 8627773 8206397 5147501 3635784 2631178/divisor=28248633;
    /* 30~39세 40~49세 50~59세 60~69세 70세이상           30세이상합*/
RUN;

```

## 분석 결과

Estimate					
Label	Estimate	Standard Error	DF	t Value	Pr >  t
고혈압표준화유병률	26.9465	0.7269	165	37.07	<.0001

표준화율 표준화율 표준오차

▶ 30세이상의 고혈압 표준화 유병률은 26.9%이고, 표준오차는 0.70이다.



---

## 부록

- 예제변수 생성 및 산출
- SAS 이용시 참고사항

## 예제변수 생성 및 산출

### □ 예제변수 생성

□ 「4장 복합표본 자료분석」에서 사용된 생성변수의 정의 및 산출 프로그램은 다음과 같다.

#### 1 변수 age30

생성 변수	설명	연령이 30세 이상인 경우
	내용	0. 30세미만 1. 30세이상
사용 변수	설명	age : 연령
	내용	연속형
SAS 프로그램		if 30 <= age then age30= 1; else age30= 0;

#### 2 변수 age\_g

생성 변수	설명	10세단위 연령대
	내용	3. 30대 4. 40대 5. 50대 6. 60대 7. 70대이상
사용 변수	설명	age : 연령
	내용	연속형
SAS 프로그램		if 30 <= age <= 39 then age_g= 3; else if 40 <= age <= 49 then age_g= 4; else if 50 <= age <= 59 then age_g= 5; else if 60 <= age <= 69 then age_g= 6; else if 70 <= age then age_g= 7;



## 6 변수 subg1

생성 변수	설명	관심집단
	내용	연령이 30세 이상이면서 수축기혈압, 체질량 지수의 값이 결측이 아닌 경우
사용 변수	설명	age : 연령
	내용	연속형
	설명	HE_SBP : 수축기혈압
	내용	연속형
	설명	HE_BMI : 체질량 지수
SAS 프로그램	내용	연속형
	<code>if 30&lt;=age &amp; HE_SBP^ = . &amp; HE_BMI^ = . then subg1=1; else subg1=0;</code>	

## 7 변수 subg2

생성 변수	설명	관심집단
	내용	연령이 30세 이상이면서 고혈압 유병여부, 현재 흡연여부 그리고 체질량 지수 값이 결측이 아닌 경우
사용 변수	설명	age : 연령
	내용	연속형
	설명	hp1 : 고혈압 유병여부
	내용	0. 정상, 고혈압전단계 1. 고혈압
	설명	sm1 : 현재 흡연여부
	내용	1. 현재흡연 2. 과거흡연, 비흡연
	설명	HE_BMI : 체질량 지수
SAS 프로그램	내용	연속형
	<code>if 30&lt;=age &amp; hp1^ = . &amp; sm1^ = . &amp; HE_BMI^ = . then subg2=1; else subg2=0;</code>	

## □ 가중평균, 가중비율 산출

- 「4장 복합표본 자료분석」에서 사용된 가중평균, 가중비율 산출 프로그램 및 결과는 다음과 같다.
- 복합표본 회귀분석 예제

### 1 연속형 보정변수의 가중평균

관심집단	subg1 : 연령이 30세 이상이면서 수축기혈압, 체질량 지수의 값이 결측이 아닌 경우																					
보정변수	HE_BMI : 체질량 지수																					
SAS 프로그램	<pre>PROC SURVEYMEANS DATA=ALL17 NOMCAR; STRATA kstrata; CLUSTER psu; WEIGHT wt_itvex; VAR HE_BMI; DOMAIN subg1; RUN;</pre>																					
분석결과	<p style="text-align: center;"><b>Domain Statistics in subg1</b></p> <table border="1"> <thead> <tr> <th>subg1</th> <th>Variable</th> <th>Label</th> <th>N</th> <th>Mean</th> <th>Std Error of Mean</th> <th>95% CL for Mean</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>HE_BMI</td> <td>체질량지수</td> <td>2241</td> <td>20.930673</td> <td>0.160073</td> <td>20.6146176 21.2467289</td> </tr> <tr> <td>1</td> <td>HE_BMI</td> <td>체질량지수</td> <td>5443</td> <td>24.010027</td> <td>0.058915</td> <td>23.8937019 24.1263511</td> </tr> </tbody> </table>	subg1	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean	0	HE_BMI	체질량지수	2241	20.930673	0.160073	20.6146176 21.2467289	1	HE_BMI	체질량지수	5443	24.010027	0.058915	23.8937019 24.1263511
subg1	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean																
0	HE_BMI	체질량지수	2241	20.930673	0.160073	20.6146176 21.2467289																
1	HE_BMI	체질량지수	5443	24.010027	0.058915	23.8937019 24.1263511																

### 2 범주형 보정변수의 가중비율

관심집단	subg1 : 연령이 30세 이상이면서 수축기혈압, 체질량 지수의 값이 결측이 아닌 경우																																																															
보정변수	age_g : 10세단위 연령대																																																															
SAS 프로그램	<pre>PROC SURVEYFREQ DATA=ALL17 NOMCAR; STRATA kstrata; CLUSTER psu; WEIGHT wt_itvex; TABLE subg1*age_g / row; RUN;</pre>																																																															
분석결과	<p style="text-align: center;"><b>Table of subg1 by age_g</b></p> <table border="1"> <thead> <tr> <th>subg1</th> <th>age_g</th> <th>Frequency</th> <th>Weighted Frequency</th> <th>Std Err of Wgt Freq</th> <th>Percent</th> <th>Std Err of Percent</th> <th>Row Percent</th> <th>Std Err of Row Percent</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>3</td> <td>908</td> <td>7506467</td> <td>493405</td> <td>21.5610</td> <td>1.1768</td> <td>21.6692</td> <td>1.1802</td> </tr> <tr> <td></td> <td>4</td> <td>1136</td> <td>8580714</td> <td>445700</td> <td>24.6466</td> <td>0.9365</td> <td>24.7702</td> <td>0.9384</td> </tr> <tr> <td></td> <td>5</td> <td>1211</td> <td>8363124</td> <td>345760</td> <td>24.0216</td> <td>0.7935</td> <td>24.1421</td> <td>0.7945</td> </tr> <tr> <td></td> <td>6</td> <td>1098</td> <td>5443549</td> <td>244611</td> <td>15.6357</td> <td>0.6701</td> <td>15.7141</td> <td>0.6774</td> </tr> <tr> <td></td> <td>7</td> <td>1090</td> <td>4747376</td> <td>262982</td> <td>13.6360</td> <td>0.7669</td> <td>13.7044</td> <td>0.7723</td> </tr> <tr> <td></td> <td>Total</td> <td>5443</td> <td>34641230</td> <td>960021</td> <td>99.5009</td> <td>0.1308</td> <td>100.000</td> <td></td> </tr> </tbody> </table>	subg1	age_g	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	1	3	908	7506467	493405	21.5610	1.1768	21.6692	1.1802		4	1136	8580714	445700	24.6466	0.9365	24.7702	0.9384		5	1211	8363124	345760	24.0216	0.7935	24.1421	0.7945		6	1098	5443549	244611	15.6357	0.6701	15.7141	0.6774		7	1090	4747376	262982	13.6360	0.7669	13.7044	0.7723		Total	5443	34641230	960021	99.5009	0.1308	100.000	
subg1	age_g	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent																																																								
1	3	908	7506467	493405	21.5610	1.1768	21.6692	1.1802																																																								
	4	1136	8580714	445700	24.6466	0.9365	24.7702	0.9384																																																								
	5	1211	8363124	345760	24.0216	0.7935	24.1421	0.7945																																																								
	6	1098	5443549	244611	15.6357	0.6701	15.7141	0.6774																																																								
	7	1090	4747376	262982	13.6360	0.7669	13.7044	0.7723																																																								
	Total	5443	34641230	960021	99.5009	0.1308	100.000																																																									

□ 복합표본 로지스틱 회귀분석 예제

**1** 연속형 보정변수의 가중평균

관심집단	subg2 : 연령이 30세 이상이면서 고혈압 유병여부, 현재 흡연여부 그리고 체질량 지수 값이 결측이 아닌 경우																																				
보정변수	age : 연령																																				
SAS 프로그램	<pre>PROC SURVEYMEANS DATA=ALL17 NOMCAR; STRATA kstrata; CLUSTER psu; WEIGHT wt_itvex; VAR age; DOMAIN subg2; RUN;</pre>																																				
분석결과	<table border="1"> <thead> <tr> <th colspan="9">Domain Statistics in subg2</th> </tr> <tr> <th>subg2</th> <th>Variable</th> <th>Label</th> <th>N</th> <th>Mean</th> <th>Std Error of Mean</th> <th colspan="3">95% CL for Mean</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>age</td> <td>만나미</td> <td>4274</td> <td>32.473725</td> <td>0.545087</td> <td>31.3974791</td> <td>33.5499702</td> <td></td> </tr> <tr> <td>1</td> <td>age</td> <td>만나미</td> <td>3435</td> <td>52.055618</td> <td>0.480446</td> <td>51.1070037</td> <td>53.0042332</td> <td></td> </tr> </tbody> </table>	Domain Statistics in subg2									subg2	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean			0	age	만나미	4274	32.473725	0.545087	31.3974791	33.5499702		1	age	만나미	3435	52.055618	0.480446	51.1070037	53.0042332	
Domain Statistics in subg2																																					
subg2	Variable	Label	N	Mean	Std Error of Mean	95% CL for Mean																															
0	age	만나미	4274	32.473725	0.545087	31.3974791	33.5499702																														
1	age	만나미	3435	52.055618	0.480446	51.1070037	53.0042332																														

**2** 범주형 보정변수의 가중비율

관심집단	subg2 : 연령이 30세 이상이면서 고혈압 유병여부, 현재 흡연여부 그리고 체질량 지수 값이 결측이 아닌 경우																																													
보정변수	sm1 : 현재흡연여부																																													
SAS 프로그램	<pre>PROC SURVEYFREQ DATA=ALL17 NOMCAR; STRATA kstrata; CLUSTER psu; WEIGHT wt_itvex; table subg2*sm1 / row; RUN;</pre>																																													
분석결과	<table border="1"> <thead> <tr> <th colspan="9">Table of subg2 by sm1</th> </tr> <tr> <th>subg2</th> <th>sm1</th> <th>Frequency</th> <th>Weighted Frequency</th> <th>Std Err of Wgt Freq</th> <th>Percent</th> <th>Std Err of Percent</th> <th>Row Percent</th> <th>Std Err of Row Percent</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0</td> <td>2865</td> <td>17622768</td> <td>624743</td> <td>51.3626</td> <td>0.8586</td> <td>80.6476</td> <td>0.8705</td> </tr> <tr> <td></td> <td>1</td> <td>570</td> <td>4228803</td> <td>220940</td> <td>12.3251</td> <td>0.5725</td> <td>19.3524</td> <td>0.8705</td> </tr> <tr> <td></td> <td>Total</td> <td>3435</td> <td>21851572</td> <td>710175</td> <td>63.6877</td> <td>0.7989</td> <td>100.000</td> <td></td> </tr> </tbody> </table>	Table of subg2 by sm1									subg2	sm1	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent	1	0	2865	17622768	624743	51.3626	0.8586	80.6476	0.8705		1	570	4228803	220940	12.3251	0.5725	19.3524	0.8705		Total	3435	21851572	710175	63.6877	0.7989	100.000	
Table of subg2 by sm1																																														
subg2	sm1	Frequency	Weighted Frequency	Std Err of Wgt Freq	Percent	Std Err of Percent	Row Percent	Std Err of Row Percent																																						
1	0	2865	17622768	624743	51.3626	0.8586	80.6476	0.8705																																						
	1	570	4228803	220940	12.3251	0.5725	19.3524	0.8705																																						
	Total	3435	21851572	710175	63.6877	0.7989	100.000																																							

## SAS 이용시 참고사항

### □ Extendobscounter 옵션

□ Extendobscounter 옵션은 데이터셋 옵션으로, SAS 9.3 이후 버전에서 생성된 SAS DB를 9.2 이하 버전에서 사용할 경우, DB저장 시에 다음 옵션을 실행하여 DB를 저장해야 한다.

□ 형태

```
options Extendobscounter=no;  
DATA 새로운 데이터셋(eoc=no); SET 데이터셋; run;
```





---

## 참고문헌

1. Paul Gorrell. Survey Analysis: Options for Missing Data. Proceeding of Northeast SAS Users Group Conference, 2009. Available at:  
<http://www.nesug.org/Proceedings/nesug09/sa/sa03.pdf>
2. Pushpal K Mukhopadhyay, Anthony B. An, Randall D. Tobias, and Donna L. Watts. Try, Try Again: Replication-Based Variance Estimation Methods for Survey Data Analysis in SAS® 9.2. Proceeding of Northeast SAS Users Group Conference, 2008. Available at:  
<http://www.nesug.org/proceedings/nesug08/sa/sa18.pdf>
3. Xiuhua Chen. Survey Logistic Regression: Some SAS® and SUDAAN® Comparisons. Proceeding of Northeast SAS Users Group Conference, 2006. Available at:  
<http://www.nesug.org/proceedings/nesug06/po/po12.pdf>
4. Arlene B. Siller, Linda Tompkins. The Big Four: Analyzing Complex Sample Survey Data Using SAS®, SPSS®, STATA®, and SUDAAN®. Proceeding of Northeast SAS Users Group Conference. Available at:  
<http://www.nesug.org/proceedings/nesug05/pos/pos3.pdf>
5. Xiuhua Chen, Paul Gorrell. Variance Estimation With Complex Surveys: Some SAS–SUDAAN Comparisons. Proceeding of Northeast SAS Users Group Conference. Available at:  
<http://www.nesug.org/proceedings/nesug04/an/an02.pdf>
6. Centers for Disease Control and Prevention. Continuous NHANES Web tutorial. Available at:  
[http://www.cdc.gov/nchs/tutorials/nhanes/index\\_current.htm](http://www.cdc.gov/nchs/tutorials/nhanes/index_current.htm)  
<http://www.cdc.gov/nchs/tutorials/nhanes/surveydesign/info3.htm>

7. Katherine L. Baisden, Paul Hu. The Enigma of Survey Data Analysis: Comparison of SAS□ Survey Procedures and SUDAAN Procedures. Proceeding of SAS Users Group International(SUGI) Global Forum, 2006. Available at:  
<http://www2.sas.com/proceedings/sugi31/194-31.pdf>
8. SAS Institute, Inc, SAS OnlineDoc□, Version 9.3, Cary, NC: SAS Institute Inc., 2002–2008
9. SAS Institute, Inc, SAS OlineDoc□, SAS / STAT(R) 13.2 User's Guide, 2014
10. SAS Institute, Inc, SAS OlineDoc□, Version 9.4, User's Guide, 2014
11. SAS Institute, Inc, SAS OlineDoc□, SAS / STAT(R) 14.2 User's Guide, 2016