

Rex workshop

파워 쿼리 및 KoGES

이보라 (서울대학교 보건환경연구소)

Contents

- 파워 쿼리 소개
- KoGES 소개
- KoGES 교육용 데이터



Make Analysis Easy and Fast

Contents

01 파워 쿼리 소개

02 KoGES 소개

03 KoGES 교육용 데이터



Excel Power Query

■ Excel BI (Business Intelligence)

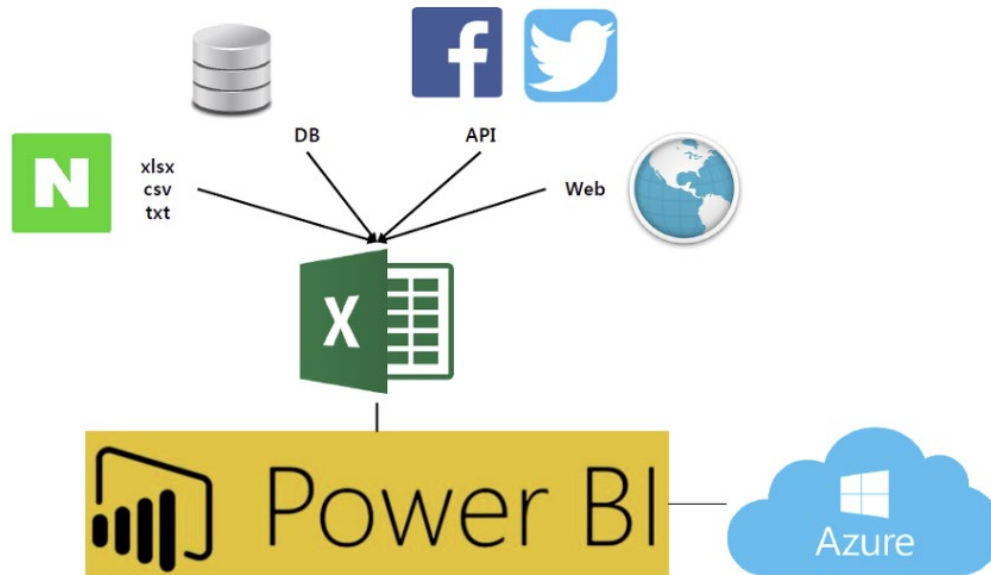
- 레코드의 제한 없이 빠르게 많은 양의 데이터를 분석할 수 있는 추가 기능

Excel BI 종류	기능
Power Query (파워 쿼리)	다양한 유형의 데이터를 검색, 연결, 결합 → 분석에 적합한 형태의 데이터로 변환
Power Pivot (파워 피벗)	1억 건 이상의 데이터를 빠르게 분석 데이터 모델 작성
Power View (파워 뷰)	데이터 시각화
Power Map (파워 맵)	지리 데이터를 맵으로 시각화

출처: <https://www.youtube.com/c/OfficeTutorKorea>

■ Power Query

- 여러 종류의 외부 데이터 소스에서 데이터를 가져와 정제, 변환, 재구성하는 엑셀 내 도구
- 엑셀 2010버전부터 제공된 새로운 COM 추가 기능 중 하나
- 2016버전부터는 추가기능이 아닌, [데이터] 탭에 등록되어 더 간편하게 활용 가능



■ 오피스 365 또는 엑셀 2016 버전 이상 사용자

- 리본 메뉴의 [데이터] 탭 > [데이터 가져오기 및 변환] / [쿼리 및 연결] 그룹의 명령 사용



■ 엑셀 2010, 2013 버전 사용자

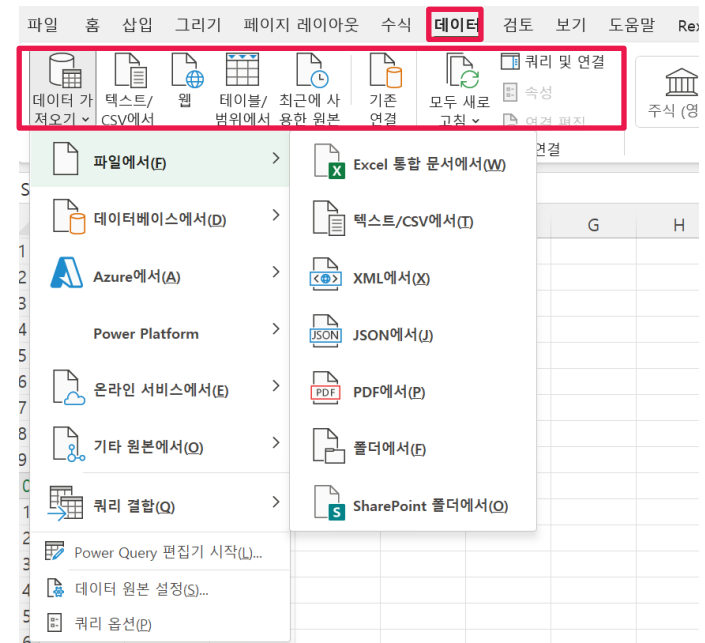
- 파워쿼리 추가기능 다운로드하여 설치 필요
- 다만 다음 조건 만족 시에만 설치 가능
 1. 윈도우 7 이상의 운영체제 및 인터넷 익스플로러 9 이상 사용
 2. 오피스 2010의 경우 Professional Plus 이상
- 제품 종류 및 버전 확인할 것
 - ✓ 2010 버전 : [파일] > [도움말] > Professional Plus 이상인지 / 32비트 또는 64비트인지 확인
 - ✓ 2013 버전 : [파일] > [계정] > [Excel 정보] 버튼 클릭 > 엑셀 버전 32비트 또는 64비트인지 확인
- 다음 사이트에서 사용중인 제품 종류 및 버전에 맞게 선택하여 다운로드
 - ✓ <https://www.microsoft.com/ko-kr/download/details.aspx?id=39379>
(구글에서 power query download 검색해도 링크로 연결됨)

■ 2010, 2013버전의 경우

- 외부데이터 가져오기
 - ✓ 외부파일로부터 편집할 데이터를 가져올 때 사용하는 명령
- 테이블에서
 - ✓ 현재 파일의 표를 쿼리로 생성할 때 사용
- 결합
 - ✓ 여러 표를 하나의 표로 합칠 때 사용
- 창 표시
 - ✓ 파일 내 쿼리 목록을 표시하는 작업 창
- 편집기 시작
 - ✓ 쿼리 편집기를 열고, 파일 내 쿼리를 관리
- 설정
 - ✓ 쿼리의 원본 데이터를 변경하거나 몇 가지 옵션 변경 가능

■ 2016 이상 버전의 경우

- [데이터] - [가져오기 및 변환] 그룹 내 하위메뉴로 많은 메뉴 제공



■ 파워쿼리 적용 전 체크 사항

- 데이터는 **표로 변경** (반드시는 아니지만, 가능하면 표로 인식시키기)
- 머리글은 1줄로 입력
- 병합된 셀이 없는지 확인
- 합계, 소계 등 **집계된 데이터는 원본데이터에 포함하지 말 것**

■ 현재 엑셀 시트 내에서 필요한 부분을 표로 지정 후, 해당 표 불러오기

파일 홈 삽입 그리기 페이지 레이아웃 수식 데이터 검토 보기 도움말 Acrobat Rex

피벗 테이블 추천 표 그림 아이콘 SmartArt 스크린샷 추가 기능 가져오기 내 추가 기능 추천 차트 지도

표 일러스트레이션 추가 기능 차트

A2 : X ✓ fx 부서

	A	B	C	D	E	F	G	H	I	J	K
1	직원 정보			생산 내역(단위 : mm)							
2	부서	사번	성명	250	255	260	265	270	275	280	285
3	생산1팀	S-0328	강영광	3	2	5	5	3	4	2	2
4	생산1팀	S-0324	문분홍	3	2	4	7	4	4	4	3
5	생산1팀	S-0341	박영재	2	2	6	7	4	4	2	3
6	생산1팀	S-0331	강다래	5	4	4	3	7	4	1	1
7	생산1팀	S-0384	안민주	3	4	4	8	8	2	4	3
8	생산1팀	S-0361	박단비	2	1	5				2	3
9	생산2팀	S-0390	이보석	4	4	4				2	2
10	생산2팀	S-0376	전술	2	3	5				3	2
11	생산2팀	S-0323	김민지	5	1	8				5	1
12	생산2팀	S-0322	성보람	5	5	7				2	3
13	생산2팀	S-0317	장소리	1	2	7	6	6	5	5	2
14	생산2팀	S-0325	정공주	2	2	3	8	5	4	5	1

표 만들기 ? X

표에 사용할 데이터를 지정하십시오(W).

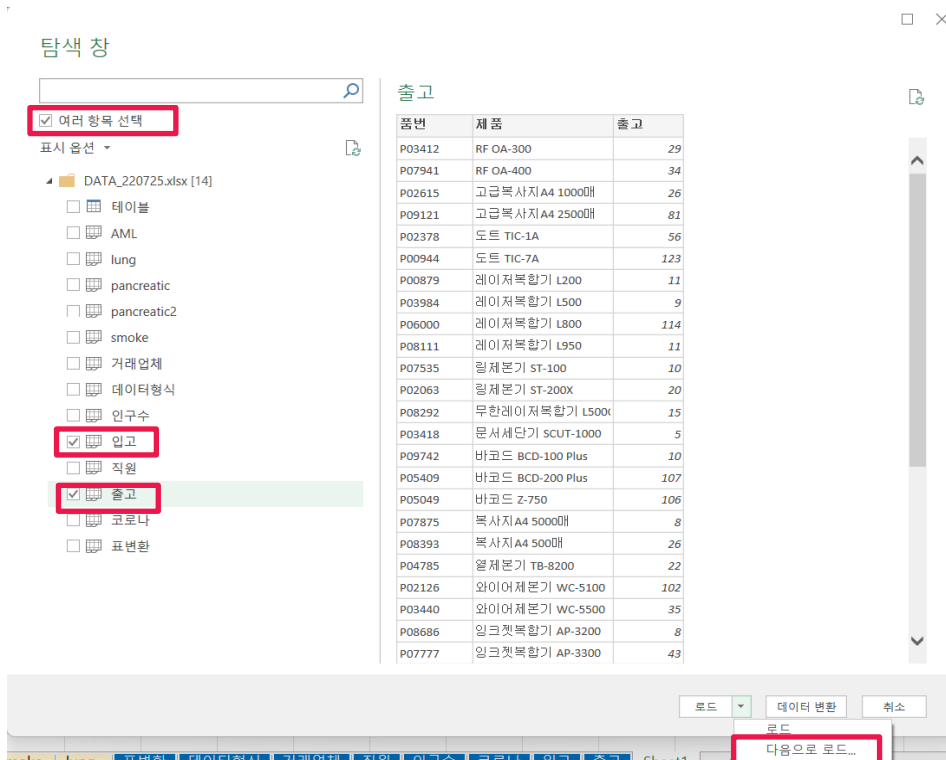
\$A\$2:\$K\$14

☒ 머리글 포함(M)

확인 취소

■ 외부 엑셀 파일에서 여러 시트의 데이터를 가져오기

- 새 시트 생성 후, [데이터]탭 - [데이터 가져오기] - [파일에서] - [Excel 통합문서에서] 클릭
- 데이터가 들어있는 엑셀 파일 선택 - [여러 항목 선택] 체크 - 필요한 시트 선택 - [다음으로 로드]
- [연결만 만들기] 선택 - [확인]



탐색 창

☒ 여러 항목 선택

표시 옵션 ▾

DATA_220725.xlsx [14]

- ☐ 데이터
- ☐ AML
- ☐ lung
- ☐ pancreatic
- ☐ pancreatic2
- ☐ smoke
- ☐ 거래업체
- ☐ 데이터형식
- ☐ 인구수
- ☒ 입고
- ☐ 직원
- ☒ 출고
- ☐ 코로나
- ☐ 표변환

출고

품번	제품	출고
P03412	RF OA-300	29
P07941	RF OA-400	34
P02615	고급복사지 A4 1000매	26
P09121	고급복사지 A4 2500매	81
P02378	도트 TIC-1A	56
P00944	도트 TIC-7A	123
P00879	레이저복합기 L200	11
P03984	레이저복합기 L500	9
P06000	레이저복합기 L800	114
P08111	레이저복합기 L950	11
P07535	잉크젯복합기 ST-100	10
P02063	잉크젯복합기 ST-200X	20
P08292	무한레이저복합기 L500i	15
P03418	문서채단기 SCUT-1000	5
P09742	바코드 BCD-100 Plus	10
P05409	바코드 BCD-200 Plus	107
P05049	바코드 Z-750	106
P07875	복사지 A4 5000매	8
P08393	복사지 A4 500매	26
P04785	열재분기 TB-8200	22
P02126	와이어채분기 WC-5100	102
P03440	와이어채분기 WC-5500	35
P08686	잉크젯복합기 AP-3200	8
P07777	잉크젯복합기 AP-3300	43

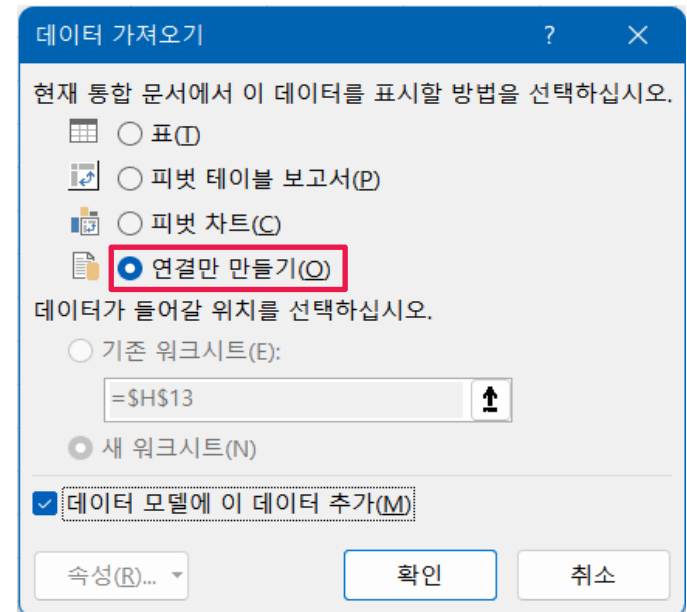
로딩

데이터 변환

취소

로딩

다음으로 로드...



데이터 가져오기

현재 통합 문서에서 이 데이터를 표시할 방법을 선택하십시오.

- ☐ 표(T)
- ☐ 피벗 테이블 보고서(P)
- ☐ 피벗 차트(C)
- ☒ 연결만 만들기(O)

데이터가 들어갈 위치를 선택하십시오.

☐ 기존 워크시트(E):

=H\$13

☒ 새 워크시트(N)

☒ 데이터 모델에 이 데이터 추가(M)

속성(R)...

확인

취소

■ 텍스트 파일 불러오기

파일 홈 삽입 그리기 페이지 레이아웃 수식 **데이터** 검토 보기 도움말 Acrobat Rex

데이터 가져오기 > 텍스트/CSV에서 > 최근에 사용한 원본 > 기존 연결 > 모두 새로 고침 > 쿼리 및 연결 > 속성 > 연결 편집 > 데이터 가져오기 및 변환 > 쿼리 및 연결 > 데이터 형식 > 정렬

주식 (영어) 지리 (영어)

A1

텍스트 파일.txt

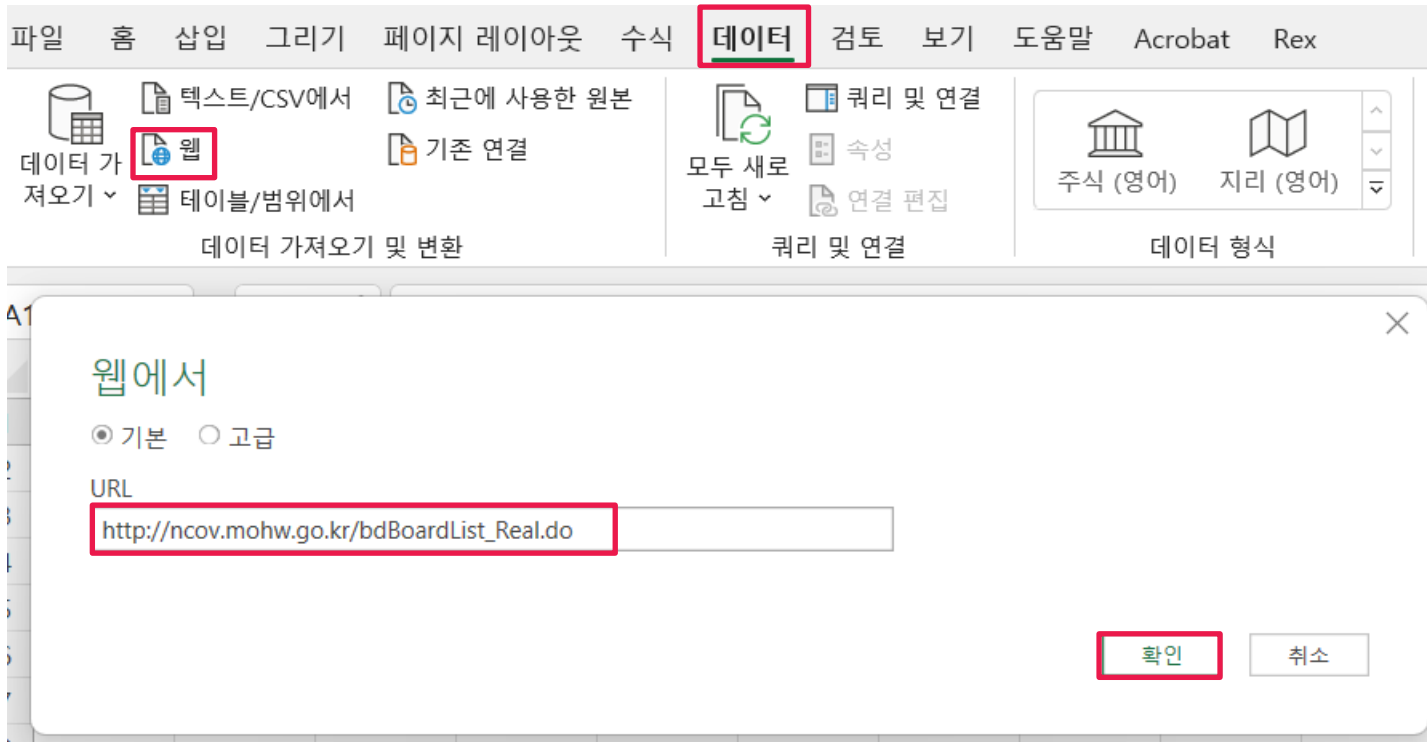
파일 원본: 949: 한국어 구분 기호: 탭 데이터 형식 검색: 첫 200개 행 기준

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11
항목	1949	1950	1960	1961	1962	1963	1964	1965	1966	
총인구(1,000명)	20189	21526	25012	25766	26513	27262	27984	28705	29436	

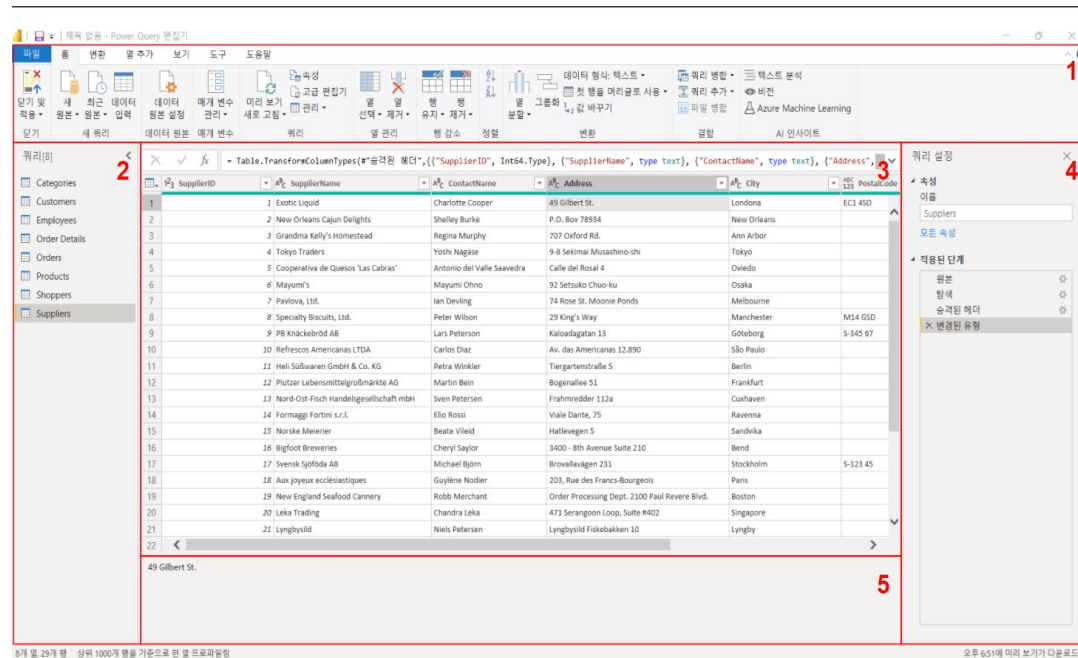
로딩 데이터 변환 취소

■ 웹 데이터 불러오기

- 구글에서 [코로나 현황] 검색 - [국내발생현황 < 발생동향 < 코로나바이러스감염증] 클릭 - [링크] 복사
- [데이터] 탭 - [데이터 가져오기 및 변환] 그룹 - [웹] 클릭 - [링크] URL에 붙여넣기 - [확인]



1. 쿼리 에디터 리본 (Query Editor Ribbon) - 데이터 수정 작업을 실행하는 명령 버튼 (복잡한 프로그래밍 언어를 간단히 버튼 클릭으로 수행할 수 있도록 도와줍니다)
2. 쿼리 창 - 현재 쿼리 에디터에 있는 쿼리 리스트
3. 작업 창 - 진행된 작업의 결과물을 보여주는 창
4. 쿼리 설정 - 현재 작업 중인 쿼리 이름과 적용된 단계를 보여주는 창
5. 상태 바 - 셀에 있는 내용을 보여주는 창





KoGES 소개

■ KoGES (한국인유전체역학조사사업)

- Korean Genome and Epidemiology Study (한국인유전체역학조사사업)
- 40세 이상의 일반 인구 집단 대상 대규모 코호트 구축을 위한 국가사업
- 2001년부터 설문조사 및 건강검진, 인체자원 수집 등의 업무 수행
- 한국인에게 가장 많이 발생하는 만성질환 (당뇨, 고혈압, 비만, 골다공증, 고지혈증, 대사증후군 등) 연구 목적
- 이중 영양조사는 대상자의 식습관, 식품 및 영양소 섭취량 등을 파악하기 위하여 실시
 - ✓ 태어나면서부터 지속적으로 우리 몸에 축적되어 질병 발생에 중요한 환경요인으로 인식됨
- SQFFQ를 통한 영양조사
 - ✓ KoGES 자체 개발한 반정량적 식품섭취빈도조사법(Semi-quantitative food frequency questionnaire)
 - ✓ 장기간의 영양소 섭취상태를 평가할 수 있어 코호트 연구에 많이 활용됨

■ 식품섭취빈도조사

- 식품의 목록과 기준분량을 제시한 후 일정기간동안 각 식품의 섭취빈도와 섭취량을 조사하는 방법
- 장기간의 식습관 파악 목적
- 한국인이 자주 섭취하는 식품의 목록을 제시
- 최근 1년간의 평균 섭취빈도 및 섭취량을 조사
- 성인의 식습관은 많이 변화하지 않는다는 가정 하에 조사
- 2004년부터 조사 실시

■ 식품섭취빈도조사지 개발 방법

- 1998년 국민건강영양조사 (Korea Health and Nutrition Examination Survey: KHNANES) 자료 이용
- 대상연령 (40-69세)의 다빈도 식품 추출
- 17개 주요 영양소에 대한 식품별 기여도와 변이정도 고려
 - ✓ 각 영양소별 기여 식품의 누적기여비율 (cumulative percent contribution: CPC)이 90%가 되는 수준까지 식품 선택
 - ✓ 누적 다중회귀계수(cumulative multiple regression coefficients: CMRC)가 높을수록 해당 식품의 존재 여부가 영양소 섭취를 설명하는데 영향을 크게 주는 것으로 해석
 - ✓ CMRC의 누적이 0.9가 되는 수준까지 식품 선택

Ⅰ 17개 영양소 목록 Ⅰ

영양소 (한글명)	영양소 (영문명)	단위
에너지	Energy	kcal
단백질	Protein	g
지방	Fat	g
탄수화물	Carbohydrate	g
섬유소	Fiber	g
칼슘	Ca	mg
인	P	mg
철	Fe	mg
나트륨	Na	mg
칼륨	K	mg
비타민 A	Vitamin A	R.E.
레티놀	Retinol	ug
베타카로틴	Carotene	ug
비타민 B ₁	Vitamin B ₁	mg
비타민 B ₂	Vitamin B ₂	mg
나이아신	Niacin	mg
비타민 C	Vitamin C	mg

■ 초판 (2001년)

- 2개의 누적다중회귀계수로부터 추출된 식품을 통합하여 식품목록을 선정한 후, 유사 식품은 동일 식품목록으로 구성
- 누락된 계절 식품 추가
- 최종 103개 식품 및 음식 목록을 포함한 식품섭취빈도조사지 초판 개발

| 설문지 초판 예시 |

※ 작년 한 해 동안 드셨던 음식

※ 다음 식품들을 지난 일년간 제시된 분량으로 얼마나 자주 섭취하였는지 √표시를 해 주세요.

식품명	섭취빈도									1회섭취분량			
	거의 안먹 음	월		주			일			기준량	더적 음	기준 량	더많 음
		1 회	2-3 회	1-2 회	3-4 회	5-6 회	1회	2회	3회				
쌀밥										1공기			
보리밥										1공기			
잡곡밥										1공기			
미숫가루/선식										3큰술(30g)			
라면										1인분			

| 설문지 초판 103개 식품/음식 항목 |

식품군	항목수	103개 식품/음식 항목
1 곡류	7	쌀밥, 보리밥, 잡곡밥, 미숫가루/선식, 흰떡/떡국, 백설기/인절미, 플레인크
2 국수와 빵류	10	라면, 국수(장국국수, 우동), 짜장면, 냉면/메밀국수, 만두, 식빵, 팔빵/호빵, 기타빵(크림빵, 카스텔라 등), 피자/햄버거, 초코파이/케익
3 과자류	3	과자류, 사탕/초콜렛, 차에 넣는 설탕
4 유제품	1	버터/마아카린
5 서류	3	감자, 고구마, 당면(잡채)
6 콩류 및 콩제품	4	콩/콩자반, 두부, 된장국/청국장, 묵
7 종실류	1	땅콩/아몬드/잣
8 채소류	23	배추김치, 깍두기/무김치, 나박김치/동치미, 기타김치(파김치, 갓김치, 고들빼김치 등), 풋고추, 고춧잎, 사금치, 상추, 들깻잎, 부추/미나리, 기타녹색채소(냉이, 쪽갓, 아욱, 취, 무청 등), 무(국, 조림)/단무지, 도라지/덕덕, 양파, 배추/배추국, 오이, 콩나물/숙주나물, 당근/당근주스, 늙은호박(호박죽)/호박즙, 애호박, 야채주스/녹즙, 고사리/고구마줄기, 장아찌
9 버섯류	2	느타리버섯, 기타버섯(양송이, 팽이, 표고)
10 과일류	12	감/꽃감, 귤, 참외/멜론, 바나나, 배, 사과/사과주스, 오렌지/오렌지주스, 수박, 복숭아/자두, 딸기, 포도/포도주스, 토마토/토마토주스
11 육류	9	계고기, 닭고기/닭다리/닭날개, 돼지고기구이(갈비, 등심 등), 삼겹살, 돼지고기찜(장조림, 족발, 순대), 햄/소시지, 쇠고기구이, 쇠고기국/갈비찜, 내장육
12 난류	1	계란/메추리알
13 생선류	7	생선회(가자미, 광어, 참치, 우럭 등), 갈치, 장어, 조기, 명태/동태/복어, 동푸른생선(고등어, 꽁치, 참치), 멸치/멸치볶음
14 기타어류 및 조개류	8	오징어/낙지, 참치(캔), 어묵/계맞살, 게, 조개(꼬막, 바지락, 조갯살), 굴, 새우, 갯갈
15 해조류	2	김, 다시마/미역
16 우유 및 유제품	5	우유, 요쿠르트/요플레, 아이스크림, 치즈, 차에 넣는 크림
17 음료류	5	청량음료(콜라, 사이다), 커피, 두유, 녹차, 기타음료(식혜, 유자차 등)

■ 식품섭취빈도조사지 개정판 (2004년)

- 초판 설문지를 활용하여 조사 이후 수정사항을 검토하여 반영된 설문지 개정판
- 계절에만 섭취하는 음식의 경우 계절변수 추가
- 대상자의 섭취량 추정을 돕기 위해 음식 사진 추가, 항목 변경 및 레시피 수정 등의 과정을 거침
- 총 103개의 항목에서 106개의 항목으로 설문지 수정

Ⅰ 설문지 개정판 예시 Ⅰ

다음은 지난 1년 동안 평균적으로 드신 음식과 식단에 관한 질문입니다. 각 항목마다 함께 제시된 그림을 참고로 하셔서 얼마나 자주 드시는지, 얼마만큼 드시는지 해당 칸에 표시해 주시기 바랍니다. 모든 항목에 응답을 하셔야 합니다. 거의 먹지 않거나 한달에 한 번보다 적게 먹는 경우에도 '거의 안먹음'에 표시를 하셔야 합니다. 두 가지 이상이 들어가 있는 경우 그 항목 모두를 고려하여 빈도를 표시해 주세요.





1
2
3

사진 1. 밥

	섭취빈도										1회 섭취분량
	거의 안먹음		월		주		일				
	1회	2-8회	1-2회	3-4회	5-6회	1회	2회	3회			
밥	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 사전 1-1 (4공기) <input type="checkbox"/> 사전 1-2 (1공기) <input type="checkbox"/> 사전 1-3 (1공기 밖)
	주로 드시는 밥의 종류는? <input type="checkbox"/> 알밥 <input type="checkbox"/> 잡곡밥 <input type="checkbox"/> 순밥과 잡곡밥을 비슷하게 먹는다 잡곡밥이 종류는? <input type="checkbox"/> 팥밥 <input type="checkbox"/> 기타잡곡밥 (기타서 잡수시는 것 뿐 아니라 회사나 식당에서 드시는 것도 포함하여 생각하십시오.)										
라면	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> 4공기 <input type="checkbox"/> 1공기 <input type="checkbox"/> 1공기 반

Ⅰ 설문지 개정판 106개 식품/음식 항목 Ⅰ

식품군	항목수	106개 식품/음식 항목
1	곡류	9
2	국수와 빵류	10
3	과자류	3
4	유제품	1
5	서육	3
6	콩류 및 콩제품	4
7	종실류	1
8	채소류	23
9	버섯류	2
10	과일류	12
11	육류	10
12	난류	1
13	생선류	7
14	기타어류 및 조개류	8
15	해조류	2
16	유유 및 유제품	5
17	음료류	5

■ 응답형태 (빈도, 분량, 섭취계절)

- 식품별 섭취빈도 (총 9단계)

- 대부분의 식품은 일 1회, 2회, 3회의 빈도로 조사함

	지난1년간 평균 섭취 빈도								
	거의 안 먹음	월		주			일		
		1회	2-3회	1-2회	3-4회	5-6회	1회	2회	3회
입력값	1	2	3	4	5	6	7	8	9

- 단 커피, 차에 넣는 설탕, 차에 넣는 프림, 녹차, 청량음료, 기타음료의 경우는
기호식품이므로 빈도를 일 1-2회, 3-4회, 5회 이상으로 조사함(개정판부터 적용)

	지난1년간 평균 섭취 빈도								
	거의 안 먹음	월		주			일		
		1회	2-3회	1-2회	3-4회	5-6회	1-2회	3-4회	5-6회
입력값	1	2	3	4	5	6	7	8	9

- 식품별 섭취분량 (총 3단계)

- 섭취 분량은 한국인이 평균적으로 섭취하는 분량을 기준으로 대, 중, 소의 분량을
각각 제시하여 사진 또는 제시한 분량을 보고 응답하게 함

	1회 평균 섭취 분량		
	소	중	대
입력값	1	2	3

- 식품별 섭취계절 (총 4단계)

✓ 과일은 섭취계절을 질문하여 빈도와 분량도 그
계절동안의 섭취빈도 및 분량으로 계산

✓ 3, 6, 9, 12개월로 조사

- 개정판부터 계절변수 추가됨

	해당 과일을 섭취하는 총 개월 수(계절을 조사하기도 함)			
	3개월	6개월	9개월	12개월
입력값	3	6	9	12

■ 결과 출력 데이터

- 106개 항목 및 23개 영양소의 1일 평균 섭취량 산출 가능

- 개인별 106개 항목의 일일 평균 섭취량

ID	F001	F002	F003	F004	F005	...	F106
개인 id	항목1 섭취량	항목2 섭취량	항목3 섭취량	항목4 섭취량	항목5 섭취량	...	항목106 섭취량

- 개인별 23개 영양소의 일일 평균 섭취량

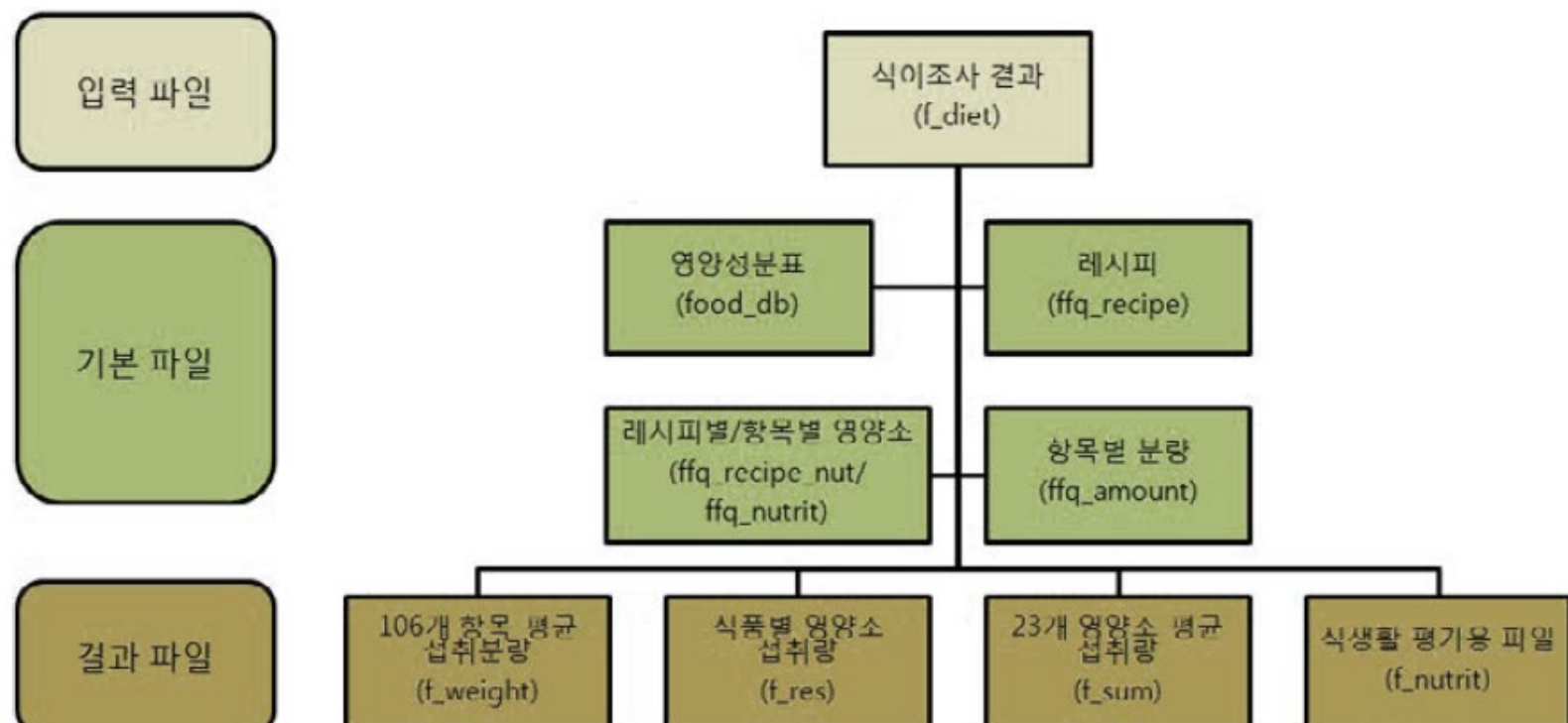
ID	S01	S02	S03	S04	S05	...	S24
개인 id	에너지	단백질	지방	탄수화물	칼슘	...	콜레스테롤
	Energy (Kcal)	Protein (g)	Fat (g)	Sugar (g)	Ca (mg)	...	Cholesterol (mg)

■ 개인별 섭취 영양소 계산 과정

- 영양소 계산 시 필요한 기본 파일, 입력 파일, 결과 파일의 종류 및 설명

파일분류	파일종류	파일명	파일설명
기본파일	식품성분표	Food_db	각 식품별 100g당 영양소 DB
	레시피	ffq_recipe	106개 항목의 구성 식품 및 음식의 레시피 (가중치 포함)
	레시피별 영양소	ffq_recipe_nut	106개 항목별 구성 식품 및 음식의 레시피 구성 무게별 영양소 함량 (100g 기준)
	항목별 영양소	ffq_nutrit	106개 항목별 영양소 함량 (100g 기준)
	항목별 분량	ffq_amount	106개 항목별 대, 중, 소 분량
입력파일	식이조사 결과	f_diet	개인별 106개 항목에 대한 섭취빈도 및 섭취분량
결과파일	항목별 섭취량	f_weight	개인별 106개 항목별 섭취 무게 (g)
	항목별 1일 영양소 섭취량	f_res	개인별 항목별 섭취한 영양소량
	1일 영양소 섭취량	f_sum	개인별 1일 섭취한 영양소량
	영양평가결과지용	f_nutrit	영양평가결과지용 데이터 포함 (권장량 대비 섭취비율 등)

■ 영양소 및 식품량 계산과정



■ 식품성분표 (Food_db)

- 각 식품별 100g 당 영양소의 DB
- 출처 : 한국영양학회 (2000)

CODE	FKOR	FENG	N01	N02	N03	N04	N05	...	N24
식품 코드	한글식품명	영문식품명	에너지	단백질	지질	탄수화물	칼슘		콜레스테롤
1001	고량미	Great millet	363	10.1	3.7	67.8	36	...	0
1002	귀리, 알곡, 도정곡	Oats, whole grain, milled grain	317	13	6.2	54.7	55	...	0
1003	귀리, 오트밀가루	Oatmeal powder	372	13.5	5.6	66.5	30	...	0
1004	기장, 알곡	Proso millet, whole grain	299	12.7	3.8	57.1	20	...	0
1005	기장, 도정곡	Proso millet, milled grain	352	11.5	2.1	70.4	11	...	0

■ 레시피 (ffq_recipe)

- 106개 식품리스트의 음식 및 식품의 레시피 (가중치 포함)

Item	Foodgroup	R_freq	median_amount	f_m	group_sum	Fcode	Foodname	R_weight	F_amount
식품 목록 번호	식품목록	외부자료* 섭취횟수	외부자료* 중간섭취량	R_freq* median_amount	항목별 F_M의 합	식품 코드	한글식품명	항목별group_sum/ 식품목록별 group_sum 합 (가중치)	항목별 식품무게 (항목별합은 100g)
P001	쌀밥	11221	220	2468620	2468620	1216	쌀, 멥쌀, 논벼, 쌀밥, 백미	1.00	100.0
P002	콩밥	1704	5.65	9627.6		4012	대두, 검정콩, 흑태	0.90	3.2
P002			170	289680	289307.6	1216	쌀, 멥쌀, 논벼, 쌀밥, 백미	0.90	96.8
P002		189	4.56	861.84		4013	대두, 검정콩, 서라태	0.10	2.6
P002			170	32130	32991.84	1216	쌀, 멥쌀, 논벼, 쌀밥, 백미	0.10	97.4
...
P006	라면	174	120	20880	20880	1031	라면, 인스턴트	1.00	100.0
P007	칼국수/장국국수/우동	85	252.29	18921.75		1030	국수, 삶은것	0.19	100.0
P007		277	270	74790		1037	소면, 삶은것	0.74	100.0
P007		23	300	6900	100611.75	1042	우동, 삶은것	0.07	100.0

■ 레시피별 영양소 (ffq_recipe_nut)

- 106개 항목별 구성 식품 및 음식의 레시피 구성 무게별 영양소 함량

Item	Fcode	Foodname	RECIPE_WEIGHT	B01	B02	B03	B04	B05	...	B24
식품 목록 번호	식품 코드	한글식품명	Recipe의 'R_weight×F_ amount'	계산된 에너지	단백질	지방	탄수 화물	칼슘	...	콜레 스테롤
F001	1216	쌀, 멧쌀, 논벼, 쌀밥, 백미	100.00	146.00	2.70	0.20	31.70	2.00	...	0
F002	1216	쌀, 멧쌀, 논벼, 쌀밥, 백미	87.11	127.17	2.35	0.17	27.61	1.74	...	0
F002	1216	쌀, 멧쌀, 논벼, 쌀밥, 백미	9.74	14.22	0.26	0.02	3.09	0.19	...	0
F002	4012	대두, 검정콩, 흑태	2.89	11.72	1.00	0.46	0.94	4.14	...	0
F002	4013	대두, 검정콩, 서리태	0.26	0.99	0.09	0.05	0.07	0.59	...	0

■ 항목별 무게 (ffq_amount)

- 106개 항목의 3단계 분량 무게 (대, 중, 소)

ITEM	FOODGROUP	AMOUNT1	AMOUNT2	AMOUNT3
식품목록번호	설문지 106개 항목	항목별 소 무게	항목별 중 무게	항목별 대 무게
F001	쌀밥	110	220	330
F002	콩밥	110	220	330
F003	기타잡곡밥	110	220	330
F004	반반콩밥	110	220	330
F005	반반잡곡밥	110	220	330
F006	라면	60	120	180
F007	칼국수/장국수/우동	150	300	450
F008	짜장면/짬뽕	215	430	645
F009	냉면/메밀국수	55	110	165
F010	만두/만두국	100	200	300

■ 항목별 영양소 (ffq_nutrit)

- 106개 항목의 100g 당 영양소 DB

ITEM	FOODGROUP	S01	S02	S03	S04	S05	...	S24
식품목록번호	식품목록	에너지	단백질	지방	탄수화물	칼슘	...	콜레스테롤
F001	쌀밥	146.00	2.70	0.20	31.70	2.00	...	0
F002	콩밥	154.10	3.71	0.70	31.70	6.66	...	0
F003	기타잡곡밥	175.48	3.46	0.31	37.87	3.26	...	0
F004	반반콩밥	150.05	3.20	0.45	31.70	4.33	...	0
F005	반반잡곡밥	160.74	3.08	0.26	34.79	2.63	...	0

■ 식이조사결과 (f_diet)

- 개인별 106개 항목의 섭취빈도 및 섭취분량

ID	Rice	Mixed rice	F001_FQ	F001_AM	F002_FQ	F002_AM
개인 id	주로 드시는 밥의 종류는? 1) 쌀밥 2) 잡곡밥 3) 쌀밥과 잡곡밥 비슷하게	잡곡밥의 종류는? 1) 콩밥 2) 기타잡곡밥	항목1의 섭취빈도	항목1의 섭취량	항목2의 섭취빈도	항목2의 섭취량

| 밥류 설문지 형태 |

	지난 1년간 평균 섭취 빈도									평균 1회 섭취분량
	거의 안먹음	월 1회	월 2-3회	주 1-2회	주 3-4회	주 5-6회	일 1회	일 2회	일 3회	
밥	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	<input type="checkbox"/> 7	<input type="checkbox"/> 8	<input type="checkbox"/> 9	<input type="checkbox"/> 1: 쌀밥 1-2 (1공기) <input type="checkbox"/> 2: 쌀밥 1-2 (2공기) <input type="checkbox"/> 3: 쌀밥 1-3 (3공기)
주로 드시는 밥의 종류는? ① 쌀밥 ② 잡곡밥 ③ 쌀밥과 잡곡밥을 비슷하게 먹는다										
잡곡밥의 종류는? ① 콩밥 ② 기타잡곡밥										
(집에서 잘수시는 것뿐 아니라 회사나 식당에서 드시는 것도 포함하여 생각하십시오.)										

| 응답에 따른 섭취한 밥의 종류 |

Rice	Mixed Rice	밥의 종류	
1	77777	F001	쌀밥
2	1	F002	콩밥
	2	F003	잡곡밥
3	1	F004	쌀밥+콩밥
	2	F005	쌀밥+잡곡밥

■ 처리 규칙

- Rice 항목에서 '①쌀밥'에 표시하면 F001
- Rice 항목에서 '②잡곡밥'에 표시하고, Mixed rice는 ①콩밥에 표시하면 F002
- Rice 항목에서 '②잡곡밥'에 표시하고, Mixed rice는 ②기타잡곡밥에 표시하면 F003
- Rice 항목에서 '③쌀밥과 잡곡밥을 비슷하게 먹는다'에 표시하면서 Mixed rice는 ①콩밥에 표시하면 F004
- Rice 항목에서 '③쌀밥과 잡곡밥을 비슷하게 먹는다'에 표시하고, Mixed rice는 ②기타잡곡밥에 표시하면 F005에 분량과 빈도를 입력함(코호트역학정보시스템과 CAPI 시스템에서는 자동으로 그 항목만 활성화되어 입력이 됨)
- F***_fq의 변수는 각 식품 항목별로 섭취빈도를 의미함
- '거의 안 먹음'은 1, '월 1회'는 2, '월 2-3회'는 3, '주 1-2회'는 4, '주 3-4회'는 5, '주 5-6회'는 6, '일 1회'는 7, '일 2회'는 8, '일 3회'는 9로 입력한다.
- F***_am의 변수는 각 식품 항목별로 섭취분량을 의미함
- '소'는 1, '중'은 2, '대'는 3으로 입력한다.
- F***_ss의 변수는(과일항목인 95번부터 106번에만 존재) 각 항목별 섭취 개월수를 의미함.
- '3개월' 동안 섭취하면 3, '6개월' 동안 섭취하면 6, '9개월' 동안 섭취하면 9, '12개월' 동안 섭취하면 12로 입력한다.

■ 항목별 섭취량 (f_weight)

- 개인별 106개 항목의 일일 평균 섭취량

ID	F001	F002	F003	F004	F005	...	F106
개인 id	항목1 섭취량	항목2 섭취량	항목3 섭취량	항목4 섭취량	항목5 섭취량	...	항목106 섭취량

■ 항목별 1일 영양소 섭취량 (f_res)

- 개인별 106개 항목별 섭취 영양소량

ID	ITEM	B01	B02	B03	B04	B05	...	S24
개인 id	106개식품 항목별	에너지	단백질	지방	탄수화물	칼슘	...	콜레스테롤
	...	Energy (Kcal)	Protein (g)	Fat (g)	Sugar (g)	Ca (mg)	...	Cholesterol (mg)

■ 일일 영양소 섭취량 (f_sum)

- 23개 영양소별 개인당 일일 평균 섭취량

ID	S01	S02	S03	S04	S05	...	S24
개인 id	에너지	단백질	지방	탄수화물	칼슘	...	콜레스테롤
	Energy (Kcal)	Protein (g)	Fat (g)	Sugar (g)	Ca (mg)	...	Cholesterol (mg)

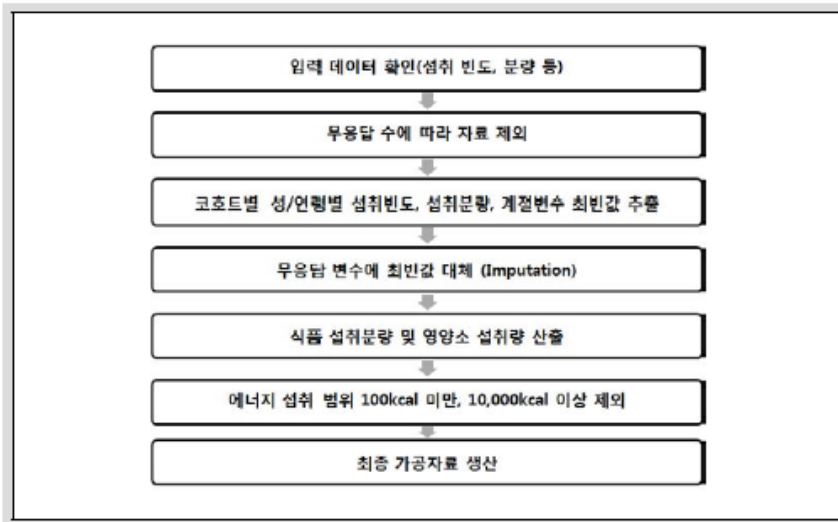
■ 영양평가 결과지용 파일 (f_nutrit)

- 개인별 식생활 평가 결과 내용만 모은 파일
 - S** : 영양소별 일일 평균 섭취량
 - N** : 영양소별 섭취 권장량
 - P** : 영양소별 권장량 대비 섭취비율 (%)
 - C** : 섭취비율에 따른 평가를 의미
 - PP : 단백질로부터 섭취한 에너지 비율(%)
 - FF : 지방으로부터 섭취한 에너지 비율(%)
 - CC : 탄수화물로부터 섭취한 에너지 비율(%)
- ✓ PP, FF, CC는 열량영양소인 protein, fat, carbohydrate로부터 섭취한 에너지가 총 에너지 섭취량에서 차지하는 비율(%)

| 23개 영양소 목록 |

변수명	영양소명	영문명(단위)
S01	에너지	Energy(Kcal)
S02	단백질	Protein(g)
S03	지방	Fat(g)
S04	탄수화물	Carbohydrate(g)
S05	칼슘	Ca(mg)
S06	인	P(mg)
S07	철	Fe(mg)
S08	칼륨	K(mg)
S09	비타민 A	Vitamin A (R.E.)
S10	나트륨	Na(mg)
S11	비타민 B1	Vitamin B1(mg)
S12	비타민 B2	Vitamin B2(mg)
S13	니아신	Niacin(mg)
S14	비타민 C	Vitamin C(mg)
S15	아연	Zinc(mg)
S16	비타민 B6	Vitamin B6(mg)
S17	엽산	Folate(ug)
S18	레티놀	Retinol(ug)
S19	카로틴	Carotene(ug)
S20	회분	Ash(mg)
S21	섬유	Fiber(g)
S23	비타민 E	Vitamin E(mg)
S24	콜레스테롤	Cholesterol(mg)

| 영양데이터 정제 단계 |



■ 식품별 섭취빈도 자료

- 일정기간으로 환산된 식품별 식품섭취빈도 (1일, 1개월 등)
- 분량을 표준화하여 환산된 식품섭취빈도임

■ 개인별 식품섭취량

- 식품별 섭취량의 상대적 비교
- 식품별 혹은 식품군별 권장량 대비 섭취 비율

■ 개인별 1일 영양소 섭취량

- 성별, 연령별 영양섭취기준 대비 영양소 섭취 비율
- 영양소별 기여 식품 도출

코호트	기반/추적	공개자료				
		테이블명		테이블 설명	변수 수 ^a	대상자 수 ^b
		한글	영문			
지역사회기반 코호트	기반, 1기 ('01-'02년)	식품섭취빈도조사	AS1_17_FFQ	섭취빈도 및 분량	206	10,030
		식품섭취빈도조사(결측치대체적용)	AS1_17_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	206	9,704
		식품섭취빈도조사(식품섭취무게)	AS1_17_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	103	
		식품섭취빈도조사(영양소)	AS1_17_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
	2차추적, 3기 ('05-'06년)	식품섭취빈도조사	AS3_19_FFQ	섭취빈도 및 분량	224	7,515
		식품섭취빈도조사(결측치대체적용)	AS3_19_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	7,482
		식품섭취빈도조사(식품섭취무게)	AS3_19_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	AS3_19_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
도시기반 코호트	기반 ('04-'13년)	식품섭취빈도조사	CT1_13_FFQ	섭취빈도 및 분량	224	173,209
		식품섭취빈도조사(결측치대체적용)	CT1_13_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	170,144
		식품섭취빈도조사(식품섭취무게)	CT1_13_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	CT1_13_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
	1차추적, CAPI ('12-'16년)	식품섭취빈도조사	CT2_12_FFQ	섭취빈도 및 분량	224	65,617
		식품섭취빈도조사(결측치대체적용)	CT2_12_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	65,318
		식품섭취빈도조사(식품섭취무게)	CT2_12_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	CT2_12_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	

코호트	기반/추적	공개자료				
		테이블명		테이블 설명	변수 수 ^a	대상자 수 ^b
		한글	영문			
농촌기반 코호트	기반 ('05-'11년)	식품섭취빈도조사	NC1_19_FFQ	섭취빈도 및 분량	224	28,337
		식품섭취빈도조사(결측치대체적용)	NC1_19_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	28,160
		식품섭취빈도조사(식품섭취무게)	NC1_19_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	NC1_19_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
	1차추적 ('07-'14년)	식품섭취빈도조사	NC2_19_FFQ	섭취빈도 및 분량	224	12,463
		식품섭취빈도조사(결측치대체적용)	NC2_19_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	12,375
		식품섭취빈도조사(식품섭취무게)	NC2_19_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	NC2_19_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
	2차추적 ('08-'16년)	식품섭취빈도조사	NC3_19_FFQ	섭취빈도 및 분량	224	11,399
		식품섭취빈도조사(결측치대체적용)	NC3_19_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	9,182
		식품섭취빈도조사(식품섭취무게)	NC3_19_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	NC3_19_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
쌍둥이 및 가족 코호트	기반 ('05-'13년)	식품섭취빈도조사	TW1_15_FFQ	섭취빈도 및 분량	224	3,202
		식품섭취빈도조사(결측치대체적용)	TW1_15_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	3,059
		식품섭취빈도조사(식품섭취무게)	TW1_15_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	TW1_15_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	
	1-3차추적 ('08-'14년)	식품섭취빈도조사	TW2_12_FFQ	섭취빈도 및 분량	224	2,030
		식품섭취빈도조사(결측치대체적용)	TW2_12_FFQIMPUTATION	섭취빈도 및 분량 결측값 대체 적용	224	2,024
		식품섭취빈도조사(식품섭취무게)	TW2_12_FFQWEIGHT	결측값 적용한 1일 평균 식품 섭취량	106	
		식품섭취빈도조사(영양소)	TW2_12_FFQNUTRI	결측값 적용한 1일 평균 영양소 섭취량	23	

^a 지역사회기반 코호트 기반조사(1기)에서는 FFQ 초판(103개 항목)을 이용하여 조사되었으며 이외 자료에서는 FFQ 개정판(106개 항목)을 이용하여 조사되어, 공개자료의 변수 수에 차이가 있음

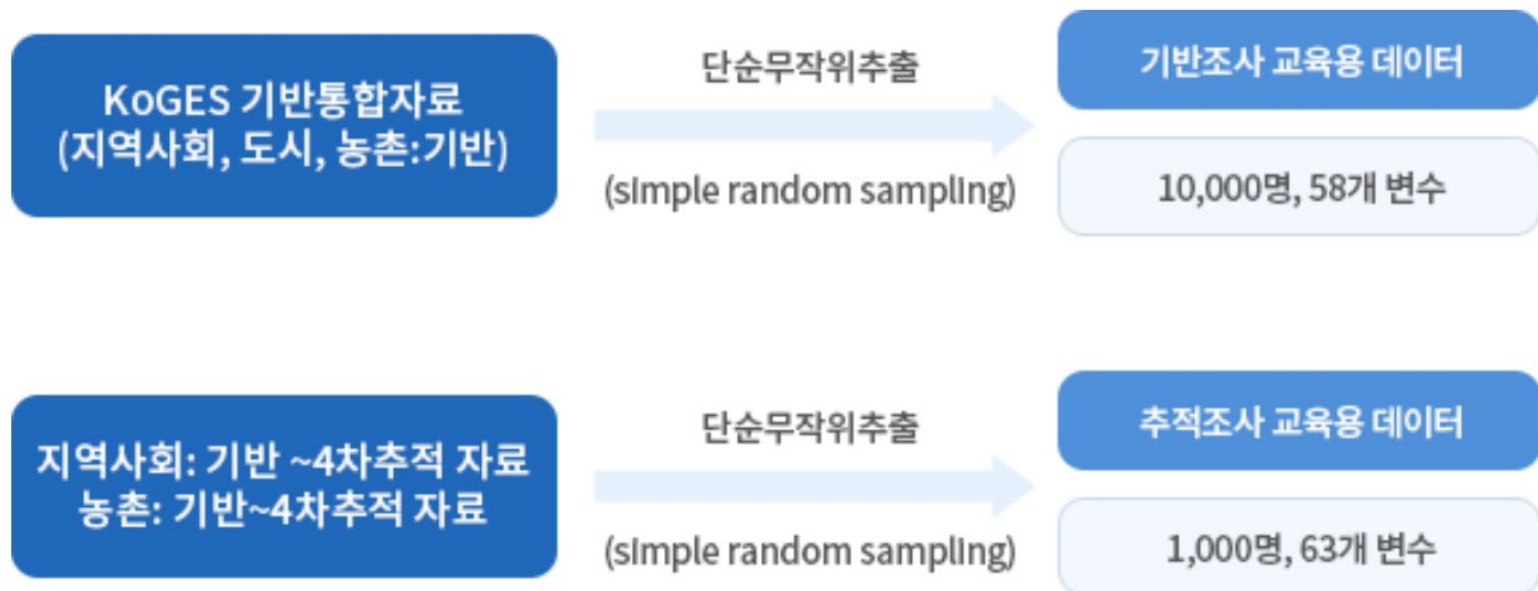
^b 식품섭취빈도조사 테이블의 대상자 수는 각 코호트 참여자 수이며, 식품섭취빈도조사(결측치대체적용, 식품섭취무게, 영양소) 테이블의 대상자 수는 무응답 빈도에 따른 제외가 반영된 참여자 수임



KoGES 교육용 데이터

■ 제공처

- <https://www.kdca.go.kr/contents.es?mid=a40504070100>
- 기반조사 / 추적조사 교육용 데이터
- 각각 코호트 자료에서 10,000명, 1,000명을 단순 무작위 추출하여 일부 자료를 변형하여 생성
- 60개 내외 변수로 구성됨



■ 교육용 데이터 구성

교육용 데이터 다운로드

구분		데이터 다운로드	코드북 다운로드
기반조사 교육용 데이터		 base_data1	
		 base_data2	
		 base_data3	
추적조사 교육용 데이터	기반	 follow 01 data	
	1차추적	 follow 02 data	
	2차추적	 follow 03 data	
	3차추적	 follow 04 data	
	4차추적	 follow 05 data	

※ 본 데이터는 교육프로그램의 교육을 위한 샘플 데이터로서 해당 목적으로만 활용하시기 바랍니다.

KoGES 데이터 분석 예시 - 기반조사

3장

KoGES 기반조사 자료 분석하기

1. 기반조사 교육용 데이터 이해하기	26
1-1. 자료 구성	26
1-2. 코드북	27
2. 자료 불러오기	28
2-1. 자료 불러오기(Excel 파일 불러오기)	28
2-2. 불러온 자료 확인하기	29
3. 자료 결합하기	30
4. 자료 분석 준비하기	33
4-1. 기본코드 결측치 처리하기	33
5. 자료 분석하기	36
5-1. 분석 대상자 선정	36
5-2. 변수 생성	38
5-3. 빈도분석	42
5-4. 평균분석	48
5-5. t-검정	53
5-6. 분산분석	55
5-7. 선형 회귀분석	58
5-8. 로지스틱 회귀분석	65

4장

KoGES 추적조사 자료 분석하기

1. 추적조사 교육용 데이터 이해하기	72
1-1. 자료 구성	72
1-2. 코드북	73
2. 자료 불러오기	74
2-1. 자료 불러오기(Excel 파일 불러오기)	74
2-2. 불러온 자료 확인하기	75
3. 자료 결합하기	76
4. 자료 분석 준비하기	79
4-1. 기본코드 결측치 처리하기	79
4-2. 날짜변수 유형 변환하기	82
5. 자료 분석하기	84
5-1. 분석 대상자 선정	84
5-2. 변수 생성	86
5-3. 총 관찰인년(Person-years) 산출	102
5-4. 생존함수 추정(생존분석)	104
5-5. Cox 비례위험모형	107

부 록

1. SPSS 기본 화면 구성	118
2. ACCESS 자료 불러오기	120
3. 분석 목적에 따른 통계분석 방법 요약	122

	지역사회 기반	도시기반	농촌기반	쌍둥이 및 가족	국내 이주자	국제협력	
기반	10,030 (‘01~‘02)	173,209 (‘04~‘13)	28,337 (‘05~‘11)	3,202 (‘05~‘13)	7,191 (‘06~‘14)	일본 1,063 (‘05~‘07, ‘11)	중국 2,493 (‘05~‘06, ‘08)
1차 추적	8,603 (‘03~‘04)	65,617 (‘12~‘16)	12,463 (‘07~‘14)	2,030 (‘08~‘14)	1,824 (‘12~‘14)	773 (‘08~‘09, ‘13)	964 (‘08~‘11)
2차 추적	7,515 (‘05~‘06)		11,399 (‘08~‘16)	940 (‘09~‘14)		549 (‘10~‘11)	
3차 추적	6,688 (‘07~‘08)		6,423 (‘11~‘16)	165 (‘12~‘14)		520 (‘12~‘13)	
4차 추적	6,665 (‘09~‘10)		1,449 (‘14~‘16)				
5차 추적	6,238 (‘11~‘12)						
6차 추적	5,906 (‘13~‘14)						
7차 추적	6,318 (‘15~‘16)						
8차 추적	6,157 (‘17~‘18)						



〈그림11. KoGES 분량자료 형태〉

〈표2. KoGES 조사 항목- 설문조사〉

항목	코호트	일반인 기반 코호트			유전-환경 모델 코호트		
		지역사회기반	도시기반	농촌기반	쌍둥이 및 가족	국내이주자	국제협력 I
생활습관	인구사회학	●	●	●	●	●	●
	흡연	●	●	●	●	●	●
	음주	●	●	●	●	●	●
	신체활동	●	●	●	●	●	●
	체중변화	●	-	-	-	●	△
질환경	과거력	●	●	●	●	●	●
	악물력	●	-	-	●	●	●
	보충제	●	●	●	●	●	△
	수술력	-	-	-	-	●	●
	가족력	●	●	●	●	●	●
	호흡순환기질환	●	-	△	-	●	-
식이조사	식품섭취빈도조사법	●	●	●	●	●	●
	24시간 회상법	-	●	-	-	-	-
	식습관	●	●	●	●	●	●
여성력		●	●	●	●	●	●
사회심리		●	●	●	-	●	●

● 전체 참여자, △ 일부참여자

〈표3. KoGES 조사 항목 - 신체계측〉

항목	코호트	일반인 기반 코호트			유전-환경 모델 코호트		
		지역사회기반	도시기반	농촌기반	쌍둥이 및 가족	국내이주자	국제협력 I
신장		●	●	●	●	●	●
체중		●	●	●	●	●	●
허리둘레		●	●	●	●	●	●
엉덩이둘레		●	●	●	●	●	●
혈압		●	●	●	●	●	●
맥박		●	●	●	-	●	●

● 전체 참여자, △ 일부참여자

〈표4. KoGES 조사 항목 - 혈액검사〉

항목	코호트	일반인 기반 코호트			유전-환경 모델 코호트		
		지역사회기반	도시기반	농촌기반	생동이 및 가족	국내이주자	국제협력 I
Glucose (공복)		●	●	●	●	●	●
Glucose (당부하 60분)		●	-	-	-	-	-
Glucose (당부하 120분)		●	-	-	-	△	-
Insulin (공복)		●	-	△	●	-	-
Insulin (60분)		●	-	-	-	-	-
Insulin (120분)		●	-	-	-	-	-
HbA1c		●	△	△	△	-	△
Total protein		●	●	△	●	●	△
Uric acid		-	●	△	●	-	△
Creatinine		●	●	●	●	●	●
BUN		●	●	●	△	●	●
Albumin		●	●	●	●	●	●
Total Bilirubin		●	△	△	△	-	-
AST		●	●	●	●	●	●
ALT		●	●	●	●	●	●
γ-GTP		●	●	●	●	●	●
Total Cholesterol		●	●	●	●	●	●
HDL-Cholesterol		●	●	●	●	●	●
Triglyceride		●	●	●	●	●	●
hs-CRP		●	●	●	●	△	△
W.B.C blood		●	△	●	●	●	●
R.B.C blood		●	△	●	●	●	●
Hemoglobin		●	●	●	●	●	●
Hematocrit		●	△	●	●	●	●
Platelet		●	△	△	●	●	△
평균적혈구용적		-	△	△	●	●	△
평균적혈구혈색소		-	△	△	●	●	△
평균적혈구혈색농도		-	△	△	●	●	△
혈액형		-	-	-	-	●	△
Calcium		●	△	-	●	-	△
Sodium		●	-	-	-	-	△
Potassium		●	-	-	-	-	△
Chloride		●	-	-	-	-	△
Vitamin B12		●	-	-	-	-	-
Folate		●	-	-	-	-	-

● 전체 참여자, △ 일부참여자

〈표5. KoGES 조사 항목 - 소변검사〉

항목	코호트	일반인 기반 코호트			유전-환경 모델 코호트	
		지역사회기반	도시기반	농촌기반	국내이주자	국제협력 I
PH		●	●	●	●	△
Nitrite		●	-	-	●	△
S.G		●	-	-	●	-
Protein		●	●	●	●	●
Glucose		●	●	●	●	●
Ketone		●	-	●	●	△
Bilirubin		●	-	-	●	●
Blood		●	●	●	●	●
Urobilinogen		●	-	-	●	△
Color		●	-	-	●	-
R.B.C.		●	-	-	●	-
W.B.C.		●	-	-	●	△
E.P cells		●	-	-	●	-
Casts		●	-	-	●	-
Bacteria		●	-	-	●	-
Crystals		●	-	-	●	-
Others		●	-	-	●	-

● 전체 참여자, △ 일부참여자

■ 기반조사 교육용 데이터

- 총 10,000명의 58개 변수가 3개의 테이블로 나뉘어 구성되어 있음



〈그림12. KoGES 기반조사 교육용 데이터 구성〉

■ 자료 결합하기



〈그림14. KoGES 역학자료 결합 예시(1)〉

테이블1 - 기본정보			테이블2 - 생활습관			기본정보 + 생활습관				
ID	SEX	AGE	ID	DRINK	SMOKE	ID	SEX	AGE	DRINK	SMOKE
NO_1	1	55	NO_1	1	2	NO_1	1	55	1	2
NO_2	2	60	NO_2	2	2	NO_2	2	60	2	2
NO_3	2	58	NO_3	2	1	NO_3	2	58	2	1
NO_4	1	63	NO_4	1	1	NO_4	1	63	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

〈그림15. KoGES 역학자료 결합 예시(2)〉

- 기반조사 교육용 데이터를 구성하는 3개 데이터셋을 하나의 데이터셋으로 결합한 후, DATA_MERGE_BASE로 저장해보자.
 - base_data1.csv ~ base_data3.csv까지 쿼리로 로드
 - base_data1 + base_data2 병합하는 쿼리 작성
 - base_data1 + base_data2 + base_data3 병합하는 쿼리 작성

■ 기본코드 결측치 처리하기

- 설문문항의 '미상/무응답', '설문문항간 상하위 관계(해당없음)', '해당변수의 조사유무(조사안함)', '추적조사 참여유무(추적조사 미참여)' 등을 구분하기 위하여 아래와 같은 기본코드가 적용되어 있음

〈표7. KoGES 기본코드의 종류와 정의〉

구분	코드명	코드	코드 정의		
결측	미상/무응답	99999	Null값(missing value) 또는 조사항목 상의 미상/무응답 값		
	해당없음	77777	조사항목에 대해 응답의 대상이 아닌 경우 예)		
			변수명	변수설명	변수값(코드)
			HTN	고혈압 과거력 - 진단여부 (1=아니오, 2=예)	1
	HTNAG	고혈압 과거력 - 처음 진단 나이	77777		
	조사안함	66666	특정 조사단위에 조사되지 않은 항목의 경우		
추적조사 미참여		55555	반복추적조사 통합자료에서 해당 차수의 조사에 참여하지 않은 경우		

■ 기본코드 결측치 처리하기

- 음주 여부(T_DRINK), 신장(T_HEIGHT) 변수를 분석하면 결측 포함되어 아래와 같이 나옴
- 따라서 기본코드 처리 필요

음주 여부 변수 빈도분석

T_DRINK					
		빈도	퍼센트	유효 퍼센트	누적 퍼센트
유호	1	5015	50.1	50.1	50.1
	2	478	4.8	4.8	54.9
	3	4446	44.5	44.5	99.4
	99999	61	.6	.6	100.0
전체		10000	100.0	100.0	

음주 여부(T_DRINK, 범주형) :
1=비음주, 2=과거 음주, 3=현재 음주

신장 변수 평균분석

기술통계량					
	N	최소값	최대값	평균	표준편차
T_HEIGHT	10000	117	99999	619.30	6756.158
유호 N(무결측)	10000				

신장(T_HEIGHT, 연속형) :
()cm

- 기반조사 교육용 데이터에 포함된 변수의 기본코드 결측치 처리하기
 - [값 바꾸기] 기능을 통해 99999, 77777, 66666, 55555를 null로 변경

■ 당뇨병 또는 비만 관련 변수가 결측인 대상자 제외하기

변수명	변수 설명	변수값 설명	변수 유형
T_DM	당뇨병 의사진단	1=아니오, 2=예	범주형
T_GLU0	혈당	() mg/dL	연속형
T_BMI	체질량지수(BMI)	() kg/m ²	연속형

- 해당 변수에 대해 필터링 적용

■ 비만도 및 당뇨병 여부 변수 생성하기

변수명	변수 설명	변수값 설명	변수 유형
T_BMI	체질량지수(BMI)	() kg/m ²	연속형
T_GLU0	혈당	() mg/dL	연속형
T_DM	당뇨병 의사진단	1=아니오, 2=예	범주형

아래 기준에 따라 체질량지수(T_BMI) 변수를 이용하여 비만도 변수를, 당뇨병 의사진단(T_DM)과 공복 혈당(T_GLU0) 변수를 이용하여 당뇨병 여부 변수를 생성해보자.

비만도* 변수 생성 기준	당뇨병 여부 변수 생성 기준
<ul style="list-style-type: none"> · 저 체 중 : 체질량지수(BMI) < 18.5 kg/m² · 정상 체 중 : 체질량지수(BMI) 18.5 ~ 23 kg/m² · 과 체 중 : 체질량지수(BMI) 23 ~ 25 kg/m² · 비 만 : 체질량지수(BMI) ≥ 25 kg/m² <p>* 세계보건기구(WHO) 아시아-태평양 지역 기준</p>	<ul style="list-style-type: none"> · 당뇨병 : 공복혈당 ≥ 126 mg/dL** 이상 또는 과거 의사로부터 당뇨병을 진단 받은 적이 있는 경우 (당뇨병 의사진단 '예') · 정상 : 당뇨병에 해당하지 않은 모든 경우 <p>** Report of a WHO/IDF consultation (2006) 기준</p>

- 사용자 지정 열 추가를 통해 변수 생성

- 성별, 나이, 비만도, 당뇨병 변수 선택 후, 성별에 따른 비만도와 당뇨병에 유의한 차이가 나는지 검정하시오.
- 성별에 따른 연령 및 혈당에 차이가 있는지 검정하시오.

KoGES 데이터 분석 예시 - 추적조사

■ 추적조사 교육용 데이터

- 기반조사 ~ 4차 추적조사자료로 구성
- 총 1,000명의 63개 변수가 5개의 테이블로 나뉘어 구성되어 있음



〈그림16. KoGES 추적조사 교육용 데이터 구성〉

■ 자료 결합하기

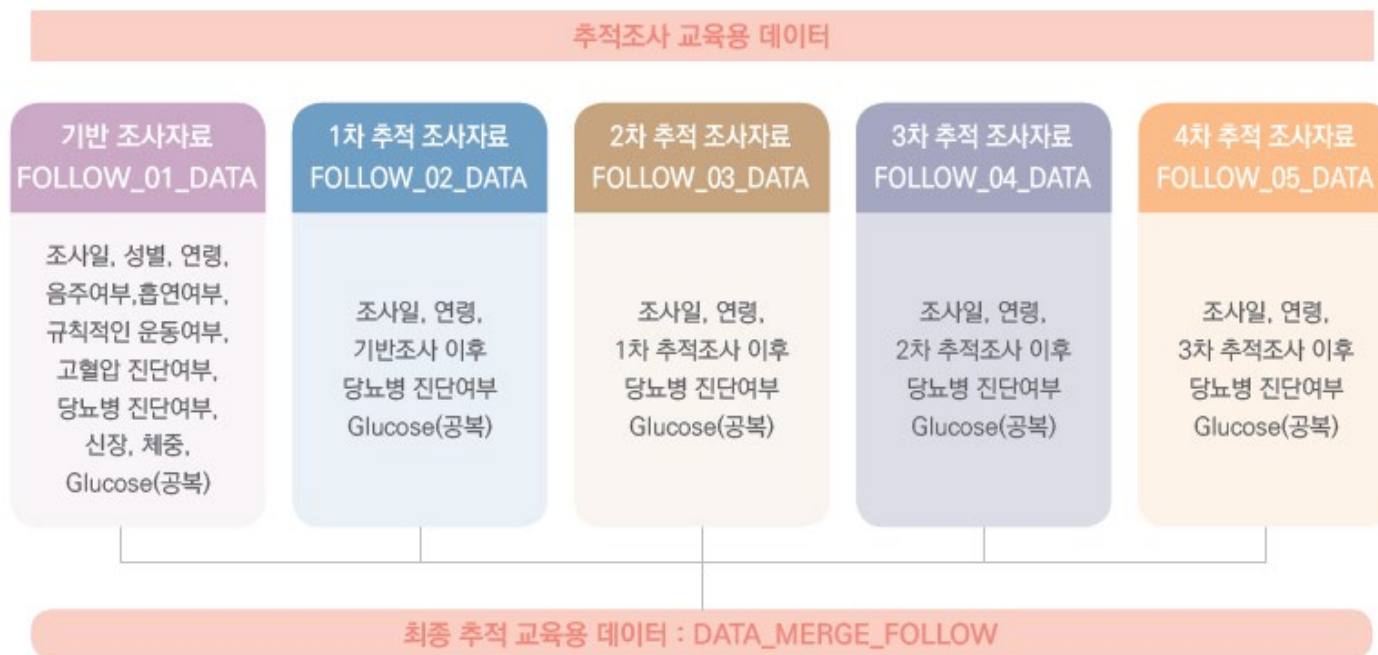
기반			1차 추적		
ID	A1_ HTN	A1_ HTN_AG	ID	A2_ HTN	A2_ HTN_AG
NO_1	1	77777	NO_1	2	55
NO_2	2	60	-	-	-
NO_3	2	58	NO_3	1	77777
NO_4	1	77777	NO_4	2	63
NO_5	1	77777	-	-	-

+
→

기반			1차 추적	
ID	A1_ HTN	A1_ HTN_AG	A2_ HTN	A2_ HTN_AG
NO_1	1	77777	2	55
NO_2	2	60	-	-
NO_3	2	58	1	77777
NO_4	1	77777	2	63
NO_5	1	77777	-	-

〈그림18. KoGES 역학자료 결합 예시〉

- 추적조사 교육용 데이터 5개에서 아래와 같이 일부 변수만 선택하여 하나의 데이터 셋으로 결합하라.



〈그림19. 분석에 사용할 최종 추적조사 교육용 데이터〉

■ 기본코드 결측치 처리하기

- 설문문항의 '미상/무응답', '설문문항간 상하위 관계(해당없음)', '해당변수의 조사유무(조사안함)', '추적조사 참여유무(추적조사 미참여)' 등을 구분하기 위하여 아래와 같은 기본코드가 적용되어 있음

〈표8. KoGES 기본코드의 종류와 정의〉

구분	코드명	코드	코드 정의									
결측	미상/무응답	99999	Null값(missing value) 또는 조사항목 상의 미상/무응답 값									
	해당없음	77777	조사항목에 대해 응답의 대상이 아닌 경우 예)									
			<table><tr><th>변수명</th><th>변수설명</th><th>변수값(코드)</th></tr><tr><td>HTN</td><td>고혈압 과거력 - 진단여부 (1=아니오, 2=예)</td><td>1</td></tr><tr><td>HTNAG</td><td>고혈압 과거력 - 처음 진단 나이</td><td>77777</td></tr></table>	변수명	변수설명	변수값(코드)	HTN	고혈압 과거력 - 진단여부 (1=아니오, 2=예)	1	HTNAG	고혈압 과거력 - 처음 진단 나이	77777
			변수명	변수설명	변수값(코드)							
HTN	고혈압 과거력 - 진단여부 (1=아니오, 2=예)	1										
HTNAG	고혈압 과거력 - 처음 진단 나이	77777										
조사안함	66666	특정 조사단위에 조사되지 않은 항목의 경우										
추적조사 미참여		55555	반복추적조사 통합자료에서 해당 차수의 조사에 참여하지 않은 경우									

■ 날짜변수 유형 변환하기

- 조사일 변수(T01_EDATE ~ T05_EDATE)는 YYYYMM으로 입력되어 있음
- 이를 날짜 유형으로 변경해야 함
- 조사일 변수는 연과 월까지 정보만 있으므로, '일'에 대한 값은 연구자가 임의로 1일, 15일, 30일 등으로 지정

■ 각 차수별 당뇨병 여부 변수를 생성한 후, 추적조사기간 동안 당뇨병 발생 여부 변수 생성

변수명	변수 설명	변수값 설명	변수 유형
T02_DM ~ T05_DM	1차 추적, 당뇨병 의사진단 ~ 4차 추적, 당뇨병 의사진단	1=아니오, 2=예	범주형
T02_GLU0 ~ T05_GLU0	1차 추적, 공복혈당 ~ 4차 추적, 공복혈당	() mg/dL	연속형

변수 생성 절차

단계	사용 변수	설명	생성변수	설명
I. 추적조사 각 차수 별 당뇨병 여부 변수 생성	① 당뇨병 의사진단 변수 다른 변수로 코딩 변경: 1=아니오, 2=예 → 0=아니오, 1=예			
	T02_DM~T05_DM	1 ~ 4차 추적, 당뇨병 의사진단 (1=아니오, 2=예)	T02_DM_D~T05_DM_D	1 ~ 4차 추적, 당뇨병 의사진단 (0=아니오, 1=예)
	② 당뇨병 여부 변수 생성: 당뇨병 의사진단 '예' 또는 공복혈당 $\geq 126\text{mg/dL}$			
	T02_DM_D~T05_DM_D	1 ~ 4차 추적, 당뇨병 의사진단 (0=아니오, 1=예)	T02_DM_D~T05_DM_D	1 ~ 4차 추적, 당뇨병 여부 (0=아니오, 1=예)
T02_GLU0~T05_GLU0	1 ~ 4차 추적, 공복혈당 (mg/dL)			
II. 추적조사 기간 동안 당뇨병 발생 여부 변수 생성	③ 1 ~ 4차 추적 당뇨병 여부 합 변수 생성: 1 ~ 4차 추적 당뇨병 여부 합			
	T02_DM_D~T05_DM_D	1 ~ 4차 추적, 당뇨병 여부 (0=아니오, 1=예)	DM_SUM	당뇨병 여부 합 (0, 1, 2, 3, 4)
	④ 당뇨병 발생 여부 변수 생성: 당뇨병 여부 합 $\geq 1 \rightarrow$ 당뇨병 발생 '예'			
	DM_SUM	당뇨병 여부 합 (0, 1, 2, 3, 4)	DM_D	당뇨병 발생 여부 (0=아니오, 1=예)

■ 연구 종료 시점을 구하여 당뇨병 발생까지 시간 변수 생성

변수명	변수 설명	변수값 설명	변수 유형
EDATE_1 ~ EDATE_5	기반, 조사일 ~ 4차 추적, 조사일	YYYY/MM/15	날짜형
T02_DM_D ~ T05_DM_D	1차 추적, 당뇨병 여부 ~ 4차 추적, 당뇨병 여부	0=아니오, 1=예	범주형
DM_D	당뇨병 발생 여부	0=아니오, 1=예(당뇨병)	범주형

■ 연구 종료 시점(END_DATE) 산출

- T02_DM=1 이면 END_DATE = T02_EDATE
- T02_DM=0 이고 T03_DM=1 이면 END_DATE = T03_EDATE
- ...
- 그 외에는 max(T02_EDATE, T03_EDATE, T04_EDATE, T05_EDATE)

Rexsoft

감사합니다

| 문의



<https://rexsoft.org>



rexsoft.help@gmail.com

홈페이지 '질문과 답변' 게시판을 통해 Rex 설치, 다운로드, 기능 등 사용문의를 남겨주세요.
신속하고 친절한 상담을 통해 사용자들의 궁금증에 답변해 드립니다.



Make Analysis Easy and Fast