

<건강검진코호트DB Coding 결과 정리>

- 12월 20일 Version

- 작성자 : 이은경

<What TO DO>

: “SNCK” 관련 covariate 정의 (변수명은 “SNCK_FRQ”)

<Share>

1) 02~08년도 구강검진DB / 07~08년도 생애구강검진DB / 14~19년도 일반 + 생애구강검진DB에서 “SNACK” 관련 변수에 결측이 있는 것을 확인

--- SNCK FRQ가 0인 사람들과의 구분을 위해 값을 “9”로 대체

2) 14~19년도 일반 + 생애구강검진DB에 “모르겠다”를 의미하는 객관식 항목이 존재하여 이는 결측과 동일시 취급

3) 07~08년도 생애구강검진DB / 14~19년도 일반 + 생애구강검진DB 기록 이용해 SNACK 문진 항목 표준화한 후, 변수 값이 9인 관측치 삭제(for 한 객체에 대해 결측이 아닌 값으로 공변량 정의하기 위해)

4) 07~08년도 생애구강검진DB & 14~19년도 일반 + 생애구강검진DB로 “SNCK” 변수가 정의되는 객체 수는 6367명(결측 제외)

3)-①. 3)과정에서 “SNCK” 변수 정의한 후, “SNCK” 변수의 요약 통계량 확인

분위수(정의 5)	
레벨	분위수
100% 최댓값	4.0
99%	2.5
95%	2.5
90%	1.0
75% Q3	1.0
50% 중위수	0.0
25% Q1	0.0
10%	0.0
5%	0.0
1%	0.0
0% 최솟값	0.0

4) 3)-① 과정에서 얻은 요약 통계량치를 이용해 02~08년도 구강검진DB 변수 값 대체 방안 설립

-- 1(그렇다) -> 1.403 (3) 과정에서 얻은 SNCK 변수 값 중 3분위수 이상 값들의 평균치)

-- 2(아니다) -> 0 (3) 과정에서 얻은 SNCK 변수 값 중 1분위수 이하 값들의 평균치)

-- 3(모른다) -> 0.255 (3) 과정에서 얻은 SNCK 변수 값 중 1분위수 이상 3분위수 이하 값들의 평균치)

↳ “모르겠다”라는 항목은 인지하지 못했던 부분이나 후에 데이터 확인 통해 찾은 부분

-- 02~08년도 구강검진DB를 통해 “SNCK” 변수를 정의할 수 있는 객체는

3)에서 정의된 객체 제외한 134,845명 중 총 67,733명 (결측 제외)

∴ 즉, “SNCK” Covariate 가 정의되는 객체는 총 141,212명 중 74,100명이며, 해당 공변량의 Missing rate는 약 47.52%이다.