

<산업보건연구원 - Meeting 자료>

- 6월 15일 Version

- 작성자 : 이은경

<What TO DO>

- 1) Memory error 방지하기 위해 생성한 데이터 저장 경로 변경
+ 데이터 생성 과정 때 work library에서 이전 반복 작업 때 사용한 데이터 삭제하는 code 추가
- 2) Demographic Information 공변량 정의한 table final version 생성
+ “AGE”, “DURATION” 변수에 “ECNY_DT”, “OUT_DT” 변수 추가

<Share & Result>

0) 기존 code의 문제점 발견

: 연령이 음수 혹은 100 초과, 근속연수가 음수인 관측치를 가지는 객체 전부의 자료를 제외해야 하는데, 그렇지 않음.

↳ 제외조건 만족하는 unique INDI_ID는 56명

1) 하드 디스크 용량 부족 + Ram 용량 부족 문제 해결하고자 데이터 저장 경로 변경

: PC에 용량이 충분한 D 드라이브 발견 → 정제된 데이터 저장 경로를 D 드라이브로 변경

1)-1. Ram 용량 부족 문제 해결하기 위해 work library 경로 바꾸고자 함.

↳ 권한 문제로 바꿀 수 없었음.

∴ cohort를 분리해 각 부분 모집단마다 사용하는 raw data가 다르므로 “proc datasets library=work kill;” 구문 사용해 work library memory 비우고 다음 코드 실행시키려고 함.

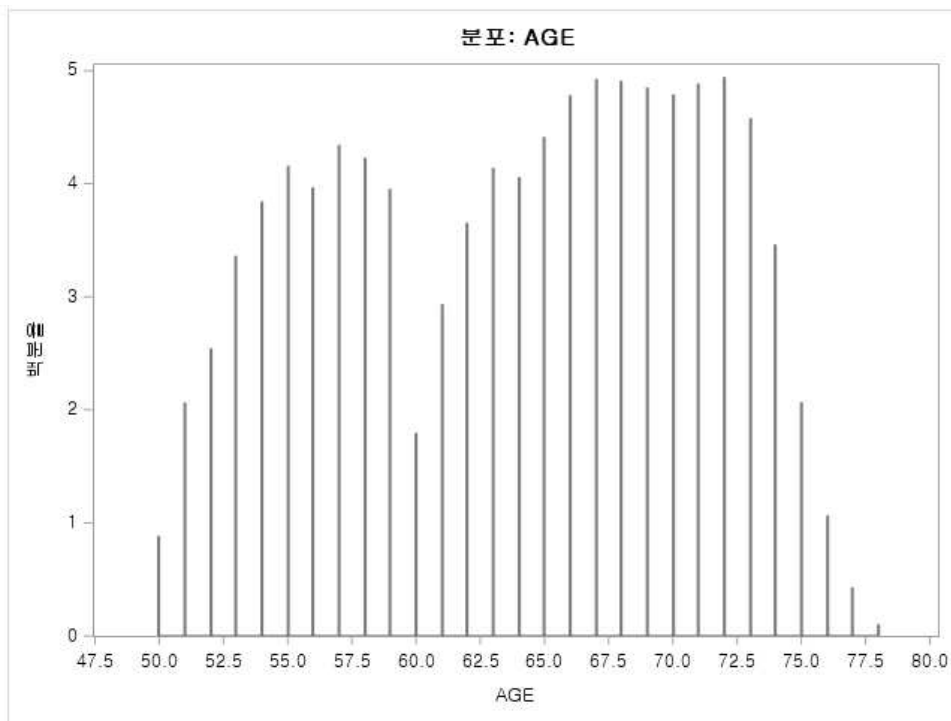
2) Demographic Information 공변량 정의한 table final version 생성

VIEWTABLE: Dir1.Temp_1975_m						
	연번	취득일	상실일	INDI_ID	AGE	DURATION
1	3104506	20040818	20050319	1000016572032	29	1
2	1060698	20040817	20040910	1000016572033	29	0
3	193218	20040817	20050315	1000016572034	29	1
4	1734329	20040817	20050110	1000016572035	29	1
5	704735	20040817	20050811	1000016572036	29	1
6	441226	20040817	20041116	1000016572037	29	0
7	1392297	19990730	19990910	1000016572040	24	0
8	2570583	20000609	20080701	1000016572040	25	8
9	2217074	20090325	20090901	1000016572040	34	0
10	1240008	20140701	20150301	1000016572040	39	1
11	510717	19980801	19981114	1000016572041	23	0
12	210624	20000317	20050401	1000016572041	25	5
13	2258252	20070223	20081231	1000016572041	32	1
14	4129561	20090101	20100201	1000016572041	34	1
15	2887953	20100916	20101226	1000016572041	35	0
16	127129	20110401	20110527	1000016572041	36	0
17	5445756	20110601		1000016572041	36	9
18	1784975	20001030	20010531	1000016572042	25	1
19	1771618	20020111	20020312	1000016572042	27	0
20	3103157	20020501	20020531	1000016572042	27	0
21	4233889	20090401	20100401	1000016572043	34	1
22	2914794	20180201		1000016572043	43	2

Question1) Demographic Information 정의한 table을 살펴보니, 고용보험 상실일(OUT_DT)가 빈 칸인 관측치가 존재, 이를 '2018/12/31'로 변경?

Question2) 태어난 연도가 1940 ~ 1959년 사이인 객체들 대상, 처음 고용보험 취득일(ECNY_DT) 때의 나이가 이미 정년에 가까운 나이이다. 이로 인해 반복 측정치가 1개인 객체들이 다수 존재 — 모형 적합 때 문제점이 되진 않을까?

example) 1940 ~ 1945년 출생 & 남성의 AGE Histogram



Question3) 생년이 1995 ~ 1999년인 객체들의 경우, 근속연수가 0(즉, 1년을 채우지 못함) 관측치가 다수 존재 --- 근속연수 이상치 정제 조건에 0도 포함?

Question4) 연번(NO), 취득일, 상실일, DURATION, AGE 값이 모두 동일한 관측치가 여러 행이 존재하는 객체 발견 — 발생할 수 있는 일인지 / How to deal with?

53597415	373412	19980618	20000401	1000001273396	55	2
53597416	373412	19980618	20000401	1000001273396	55	2
53597417	373412	19980618	20000401	1000001273396	55	2
53597418	373412	19980618	20000401	1000001273396	55	2
53597419	373412	19980618	20000401	1000001273396	55	2
53597420	373412	19980618	20000401	1000001273396	55	2
53597421	373412	19980618	20000401	1000001273396	55	2
53597422	373412	19980618	20000401	1000001273396	55	2
53597423	373412	19980618	20000401	1000001273396	55	2
53597424	373412	19980618	20000401	1000001273396	55	2
53597425	373412	19980618	20000401	1000001273396	55	2
53597426	373412	19980618	20000401	1000001273396	55	2
53597427	3814311	19950701	19961231	1000001273397	52	1
53597428	3814311	19950701	19961231	1000001273397	52	1
53597429	3814311	19950701	19961231	1000001273397	52	1
53597430	3814311	19950701	19961231	1000001273397	52	1
53597431	3814311	19950701	19961231	1000001273397	52	1
53597432	3814311	19950701	19961231	1000001273397	52	1
53597433	3814311	19950701	19961231	1000001273397	52	1
53597434	3814311	19950701	19961231	1000001273397	52	1
53597435	3814311	19950701	19961231	1000001273397	52	1
53597436	3814311	19950701	19961231	1000001273397	52	1
53597437	3814311	19950701	19961231	1000001273397	52	1
53597438	3814311	19950701	19961231	1000001273397	52	1
53597439	3814311	19950701	19961231	1000001273397	52	1