<산업안전보건연구원 - Meeting 자료>

- 6월 12일 Version
- 작성자 : 이은경

<What TO DO>

1) Demographic Information "AGE", "DURATION" 변수 정의

<Share & Result>

0) 기본적으로 이상치가 정제된 Raw data 재생성

DB	table명	구성 변수	기본 정제 조건
고용보험 DB	"Demographic_raw" tbl	입사 시 연령(AGE) + 근속연수(Duration) + 취득일(ECNY_DT) + 상실일(OUT_DT) + INDI_ID + NO	- 취득일 ≠ 상실일 - 취득일 < 20190101 - INDI_ID ≠ '1000000000001'
사망 DB	"Death_raw" tbl	사망연령(DTH_AGE) + 사망연월(DTH_DATE1~DTH_DATE3) + NO+INDI_ID	- INDI_ID ≠ '1000000000001'
암 DB	"Cancer_raw" tbl	진단 일자(fdx1~fdx6) + 진단코드(icd10_1~icd10_6) + NO + INDI_ID	- INDI_ID ≠ '1000000000001'

1) "AGE", "DURATION" 변수 정의

1)-1. "Demographic_raw" data의 column이 모두 문자형으로 구성되어 있어 수치형으로 변환해줌.
--"Num_Demographic_raw" tbl

1)-2. 이전에 생성한 "mine.Personal_Information"(생년월일 정보) tbl + "Num_Demographic_raw" tbl joint를 통해 생년이 1940년과 1999년 사이 + DURATION이 0 이상인 객체들의 unique INDI_ID만 가져옴.
--- "Include_ID" tbl
(∵ 포함조건을 만족하는 객체들의 unique INDI_ID 가져오기 위함.)
↳ 생년이 1940년과 1999년 사이 + DURATION이 0 이상인 객체들의 N수 = 29,390,968

Problem)

Macro 이용해 기존에 생성한 연도별 / 성별 기준으로 나눈 tbl에 속한 객체들의 "AGE", "DURATION" 공변량 정의하려 하였으나, Memory error로 인해 Macro 결과 얻은 tbl을 저장할 수 없는 상황임.
--- url link 이용해 LIBNAME 지정하는 방안?