

# 〈산업보건연구원 - Meeting 자료〉

- 6월 5일 ~ 6월 7일 Version

- 작성자 : 이은경

## 〈What TO DO〉

: “Personal Information” table에 저장된 모든 객체들이 “신규 취득 여부 = Y”인 행을 가지고 있는지 확인  
: 연도(5년 단위)별, 성별로 split한 Data 별로 N수 파악하기

## 〈Share & Result〉

1) 모든 객체들이 “신규 취득 여부 = Y” 행을 가지고 있는지 확인

: 1개의 INDI\_ID당 “신규 취득 여부 = Y”인 행이 여러 개 존재하는 객체가 있음.

example)

8	3408642	20190304	Y	10000000000009
9	3442507	20190805	Y	10000000000009

→ “신규 취득 여부 = Y”인 행을 가지는 unique INDI\_ID는 총 “30,618,227”명이다.

/ 기존에 생성한 “Personal\_Information” table에 존재하는 unique INDI\_ID는 총 “31,360,814”명이다.

∴ 총 742,587명이 “신규 취득 여부 =Y” 관측값을 가지고 있지 않은 것으로 파악됨.

Question 1) “신규 취득 여부 = Y”가 의미하는 것이 처음 고용 보험에 가입되었음이 아닌지?  
Question 2) “신규 취득 여부 = Y” 관측치가 없는 객체의 경우 “최초 고용 보험 가입 일자”를 어떻게 정의해야 하는지?

2) 연도(5년 단위), 성별로 Data split 한 뒤, 각 table 별로 N수 파악

(이전 Meeting comment)

- 분석 연도를 “1940년~1999년”으로 limit

- “Personal\_Information.sas7bdat” file에 포함된 unique 객체 수는 30,603,851명이다.

⇒ 연도와 성별로 분리한 data table 명은 “dt\_연도\_성별” 형태이다.

YEAR 범위	성별	N수
1940~1944	Male	503,011
	Female	282,970
1945~1949	Male	672,897
	Female	409,409
1950~1954	Male	899,526
	Female	614,036
1955~1959	Male	1,383,918
	Female	1,024,800
1960~1964	Male	1,679,756
	Female	1,289,628
1965~1969	Male	1,842,086
	Female	1,449,134

YEAR 범위	성별	N수
1970~1974	Male	2,047,952
	Female	1,724,545
1975~1979	Male	1,844,648
	Female	1,672,550
1980~1984	Male	1,879,930
	Female	1,672,550
1985~1989	Male	1,483,705
	Female	1,387,813
1990~1994	Male	1,500,376
	Female	1,423,793
1995~1999	Male	899,285
	Female	984,922

: 연도별 / 성별 데이터를 분리한 뒤, N수를 파악해본 결과 대부분 남성 근로자의 수가 더 많고  
1970년~1974년에 태어난 근로자의 수가 가장 많은 것으로 파악된다.