SAS code를 이용해 임상정보, FFQ, Lifestyle 데이터 결측값, 빈도 확인하기

**1. data set 확인**

- 암호 제거 (암호가 걸려 있으면 프로그램이 불러오지 못함)

- 변수: 파일의 1행, 한글, 공백 X, 너무 길어도 X, 줄 바꿈 X

- 데이터: 줄바꿈 X, 날짜는 하나의 형식으로 통일 (일반 or 날짜)

- 행열 개수 파악

★ 대상자 파일의 **임상정보**: 1행에 SAS 분석용 변수, 2행에 원래 변수가 있음. 원래 변수를 삭제한 별도의 파일을 만들어서 진행하는 것을 권장 (암호 해제도 필수)

🡪 2023.12.05 기준/ SAS 분석용 변수가 추가된 대상자 파일: H13, H25, H251

★ **Lifestyle**: 1,2행이 버전이 다른 변수인 경우 있음. 둘 중 하나만 남긴 별도의 파일을 만들어서 진행하는 것을 권장

**2. Code 파일 열기**

현재 code 위치: \\172.20.216.34\연구서버\연구검체\H13\_구강\_Oral\data check using SAS

**3. Code 실행**

1) **libname** **🡪** 불러올 파일의 폴더 위치를 입력해 새로운 이름을 가진 library로 생성.

libname data "\\172.20.216.34\연구서버\연구검체\H13\_구강\_Oral";

2) **import** **🡪** 불러올 파일의 **경로**(위의 library와 동일)와 **파일이름**, **시트 이름**을 적고 새로운 이름의 데이터셋으로 생성

**proc** **import** datafile="\\172.20.216.34\연구서버\연구검체\H13\_구강\_Oral\H13\_구강암센터\_Oral\_대상자파일\_20231127\_no\_PW.xlsx" dbms=excel out=data.OC replace;Sheet="임상정보";**run**;

**☆ library 이름 일치하는지 확인 - 다르면 실행 안됨**

**☆** 라이브러리 - 활성 라이브러리 - libname으로 생성한 새로운 라이브러리 – ‘새로운 이름’ 으로 저장한 데이터셋 클릭하여 데이터 확인 (행/열 개수 일치하는지, 오류 없이 불러와졌는지)

3) **Freq** **🡪** 데이터셋의 변수명을 삽입하여 각 변수들의 빈도수 파악 (엑셀 파일에서 복사/붙여 넣기 하면 편함)

**proc** **freq** data= data.OC;

table 변수명 나열 (공백으로 구분)

; **run**;

**☆ 불러온 데이터셋 이름 확인 - 다르면 실행 안됨**

**4. 결과 확인**

- Freq 결과 확인 후 마우스 오른쪽 클릭으로 전체 선택-복사-붙여넣기(엑셀)로 별도 저장