

Code pipeline 설명서

담당자	은경 이
상태	완료
마감일	@2023년 7월 30일
태그	My work

Code File 목록 & 설명 & 주요 table 명)

“0.Data_cleaning.sas”

: 데이터 클리닝 작업 수행하는 code file

[참고 file]

[전처리] 코딩 오류 제거 및 코호트 정의

[Data_EDA 확인사항 정리.pdf](#)

[주요 변수 명]

- “[Preproc1_new](#)” tbl

: Raw data에서 cleaning 작업이 완료된 table로, EDA 작업할 때 해당 table 이용하면 됨.

“1.Data_EDA.sas”

: “Entry” 변수(최초 고용 보험 등록 연도 나타내는 Dummy variable), 출생 연도, 성별, 폐암/백혈병 발병 유무 + 진단 일자, Outcome 별 time-invariant 개인 암 이력 변수 정의하는 내용 담긴 code file

[참고 file]

[20230627_Demographic_Outcome_reDefine.pdf](#)

[주요 변수 명]

- “[Demographic_Info](#)” tbl

: 생년월일 / 성별 / 최초 고용보험 등록일 / “Entry” 변수(최초 고용보험 등록 연도 Dummy variable) / 사망 여부 + 사망 시점 변수 다시 정의한 tbl

SEX	INDI_ID	BYEAR	DTH_DATE	Entry_1 999	Entry_2 004	Entry_2 009	Entry_2 014	Entry_2 019	Death

↳ 하나의 INDI_ID 당 하나의 row 가지는 tbl

- “Cancer” tbl

: Outcome(폐암, 백혈병 발병 여부) + Outcome 발생 시점 / Outcome 별 개인 암 이력 변수 정의

INDI_ID	lung_ cancer	lung_cancer_ date	leukemia	leukemia_ date	Lung_Cancer _History	Leukemia_Cancer _History

↳ 하나의 INDI_ID 당 하나의 row 가지는 tbl

- “N1_Raw” tbl

: “N1” model(공변량이 entry 변수 + time-invariant 개인 암 이력)을 적합하기 위해 필요한 data

↳ “Cancer” tbl + “Demographic_Info” tbl MERGE

↳ 하나의 INDI_ID 당 하나의 row 가지는 tbl

“2. lung_cancer_counting_format.sas”

: Outcome이 폐암 발생인 데이터 생성한 후, counting-process format 형성하는 과정 담은 code file

[참고 file]

2023.07.03 폐암 대상 Counting_process format 변환

[주요 변수 명]

- “Company_Info” tbl

: joint key인 “NO”와 사업장 정보 관련 변수만을 가져온 tbl

NO	OUT_CUZ	BIZ_INDU TY10	JSSFC_ CD	JSSFC_ NO	INDDIS _NO	ENROL _NO	BIZ_N M	BIZ_AD DRESS	BIZ_ZIP
----	---------	------------------	--------------	--------------	---------------	--------------	------------	-----------------	---------

- “lung_raw” tbl

: Outcome이 폐암인 반복 측정 자료로 “N1_raw”와 “company_Info” tbl을 INDI_ID 기준으로 MERGE 해 얻음.

- “*final_lung*” tbl

(Start, Stop] 형태를 추가한 N1 model 적합 용 데이터 (“개인 암 이력 변수”는 시간 불변 공변량 형태) / 해당 tbl을 구성하는 변수는 “*lung_raw*” tbl 변수 + Start + Stop이다.

- “*final_lung_plus_NO2*” tbl

: “*final_lung*” tbl에 각 obs를 구분해주는 “NO2” 변수 추가한 tbl

“3.lung_cancer_time_varying_cancer_history_define.sas”

: 폐암 Outcome 대상, 개인 암 이력을 “시간 가변 공변량”으로 변환 + counting-process format 수정하는 code file

[참고 file]

2023.07.07 개인 암 이력 time-varying covariate 변환

2023.07.10 개인 암 이력 수정 & Leukemia Outcome 대상 counting.process format 형성

[주요 변수 명]

- “*final_lung_time_varying_NO3*” tbl

: 개인 암 이력을 시간 가변 공변량으로 변환 + counting-process format 재정리 + (Start, Stop] 변수 추가 + 각 행 구분하는 unique key인 “NO3”(obs number) 추가 생성한 tbl

→ 폐암 Outcome 대상, Cox-ph model 적합할 때 사용한 분석용 데이터

“4.lung_cancer_outcome_N1_model_fitting_with_time_varying_cancer_history.sas”

: 폐암 발생이 Outcome일 때, 5년 단위로 출생 연도와 성별 기준으로 데이터 분리한 뒤, 공변량이 time-invariant 개인 암 이력 변수, “Entry” 변수인 Cox-ph model 적합하는 code (SAS MACRO) file

(참고 사항)

: Cox-ph model 적합 때 “개인 암 이력”이 시간 가변 공변량으로 들어간 경우, “COVSANDWICH” option은 sas에서 적용되지 않음.

[참고 file]

2023.07.10 개인 암 이력 수정 & Leukemia Outcome 대상 counting.process format 형성

“5. leukemia_couting_process_format.sas”

: 백혈병 발생이 Outcome인 데이터 생성 후, ,counting-process format 형성하는 과정 담은 code file — 3번 code file 과 코드 구성 다르지 않음.

[참고 file]

2023.07.10 개인 암 이력 수정 & Leukemia Outcome 대상 counting.process format 형성

[주요 변수 명]

- “*final_leukemia_NO2*” tbl

: 백혈병 Outcome 대상, “개인 암 이력” 이 시간 불변 공변량 형태인 N1 model 적합 용 최종 데이터(counting-process format으로 변환된 데이터) + (Start, Stop] 형태 추가 + 각 obs 구분해주는 “NO2”(obs number) 변수 추가되어 있음.

“6.leukemia_cancer_time_varying_cancer_history_define.sas”

: 백혈병 Outcome 대상, 개인 암 이력을 “시간 가변 공변량”으로 변환 + counting-process format 수정하는 code file

[참고 file]

2023.07.10 개인 암 이력 수정 & Leukemia Outcome 대상 counting.process format 형성

[주요 변수 명]

- “*final_leukemia_time_varying_NO3*” tbl

: 개인 암 이력을 시간 가변 공변량으로 변환 + counting-process format 재정리 + (Start, Stop] 변수 추가 + 각 행 구분하는 unique key인 “NO3”(obs number) 추가 생성한 tbl

→ 백혈병 Outcome 대상, Cox-ph model 적합할 때 사용한 분석용 데이터

“7.leukemia_outcome_N1_model_fitting_with_time_varying_cancer_history.sas”

: 백혈병 발생이 Outcome일 때, 5년 단위로 출생 연도와 성별 기준으로 데이터 분리한 뒤, 공변량이 time-invariant 개인 암 이력 변수, “Entry” 변수인 Cox-ph model 적합하는 code (SAS MACRO) file

(참고 사항)

: Cox-ph model 적합 때 “개인 암 이력”이 시간 가변 공변량으로 들어간 경우, “COVSANDWICH” option은 sas에서 적용되지 않음.

[참고 file]

2023.07.10 개인 암 이력 수정 & Leukemia Outcome 대상 counting.process format 형성