

〈당뇨병-망막변성〉

- Procedure에 자주 사용되는 sas table 정리

- 11월 8일 Version

작성자 : 이은경

〈Notice〉

: table이 생성된 과정 / 새로 추가한 변수 / PROC step에서 주로 쓰이는 변수명 정리하였음.

1) “dramd4” table

: 모든 분석 Data의 Base가 되는 table

[새로 생성한 주요변수]

① “cat” : ‘DR with AMD’(value가 1)와 ‘DR without AMD’(value가 0)를 구분하는 변수

② “bdr” : 기존의 “Baseline DR” 변수(chr)를 수치형으로 변환한 변수

-- 1 : mild / 2 : moderate / 3 : severe / 4 : otherwise

②-1 “bdr2” : “bdr” 변수 재정의

-- 1 : mild / 2 : moderate , severe / 4 : otherwise

③ “bad” : event(>=2-step prog or prog to PDR) 발생 여부 의미하는 변수 — Indicator variable

④ “A1cdiff” : ‘A1c_baseline - Avg_a1c’ 연산 통해 생성한 변수

⑤ “AMDG” : “AMD_grade” 변수 재정의 -- “AMD_grade” 변수값이 0이면 1로 변환한 step만 거침

⑥ “exccr” : exclusion 여부 나타내는 변수

⑦ “cnt” : 변수를 구별해주는 또 다른 변수 (변수 type이 수치형)

⑧ “prog” : “2-step progression” 여부 나타내는 변수

⑨ “pdr” : “Progressio to PDR” 여부 나타내는 변수

1)-1 “dramdexc” table

: “dramd” table에서 “exccr” 변수값이 1인 개체 제외한 tbl — 제외기준에 포함되는 객체 삭제

1)-2 “med” table

: “dramdexc” table에서 age, A1c_baseline, Avg_a1c, DM_duration 변수에 대해 순위를 매긴 변수가 추가된 table — subset analysis를 위한 과정

2) “all” table

: “dramdexc” tbl과 “med” tbl을 “cnt” 변수 기준으로 joint

: 이때, “med” tbl에 존재하는 순위 관련 변수 중 결측값이 존재하는데 “all” tbl을 생성하면서 array 구문을 통해 순위 중 결측값을 없앴(순위 관련 변수에서 기존 결측이었던 변수 값이 모두 9로 바뀜.)

변수 Age의 순위	변수 A1c_baseline의 순위	변수 avg_A1c의 순위	변수 DM_duration의 순위	ageb2	A1CB2	A1CA2	DMD2
0	.	.	0	1	9	9	1
0	.	.	0	1	9	9	1
0	1	1	0	1	2	2	1
0	1	1	1	1	2	2	2
0	1	1	1	1	2	2	2
0	.	0	0	1	9	1	1
0	.	0	0	1	9	1	1
1	.	.	1	2	9	9	2
1	.	.	1	2	9	9	2
0	1	1	0	1	2	2	1
0	1	1	0	1	2	2	1

2-1) “all” table update

① “pna” 변수 추가 : “ $\text{logit}(P(\text{event 발생})) = \beta_0$ ” model 적합한 후 $\hat{P}(\text{event 발생})$ 값을 저장
 -- “pna” 변수값은 모든 개체에 대해 동일함.

② “pda” 변수 추가 : “ $\text{logit}(P(\text{event 발생})) = \beta_0 + \beta^T X$ ” (이때, X에는 age, agesq, sex, bdr, DM_type, DM_duration, HTN, past_smoker, curr_smoker 변수들이 포함) model 적합 후 $\hat{P}(\text{event 발생} | X = x)$ 값을 저장

③ “wa” 변수 추가 : pna / pda 계산한 값 저장

3) “allprog” table

: “all” table에서 Baseline DR이 severe 미만, 즉, Baseline DR이 weak / moderate인 개체만 가져와 저장한 tbl

4) “all2” table

: “all” table에서 “avg_A1c”와 “A1c_baseline” 변수값이 결측인 개체들을 제외한 나머지 개체들만 가져와 저장한 tbl

4-1) “all2” table update

① “pna1” 변수 추가 : “ $\text{logit}(P(\text{event 발생})) = \beta_0$ ” model 적합한 후 $\hat{P}(\text{event 발생})$ 값을 저장
 -- “pna1” 변수값은 모든 개체에 대해 동일함.

② “pda1” 변수 추가 : “ $\text{logit}(P(\text{event 발생})) = \beta_0 + \beta^T X$ ” (이때, X에는 age, agesq, sex, bdr, DM_type, DM_duration, HTN, past_smoker, curr_smoker 변수들이 포함) model 적합 후 $\hat{P}(\text{event 발생} | X = x)$ 값을 저장

③ “rden” 변수 추가 : “ $g(\text{Avg_a1c}) = \beta_0 + \beta_1 X$ ” (이때, X에는 cat, age, agesq, sex, bdr, DM_type, DM_duration, A1c_baseline, HTN, past_smoker, curr_smoker 변수들이 포함됨 / $g(\)$ 는 변환 함수 의미) model 적합 후 얻은 표준화된 잔차 값들 저장

④ “rnum” 변수 추가 : “ $g(\text{Avg_a1c}) = \beta_0 + \beta_1 * (\text{cat})$ ” model 적합 후 얻은 표준화된 잔차 값들 저장

⑤ “wa1” 변수 추가 : standardized IPW를 의미하는 변수

-- “cat” = 1이면 $wa1 = pna1 / pda1$

-- “cat” = 0이면 $wa1 = (1-pna1) / (1-pda1)$

⑥ “wm1” 변수 추가 (generalized PS를 의미한다고 함)

: 각각 “rden”, “rnum” 변수를 이용해 “ $\exp(-5*((rnum \text{ or } rden)**2)/2.506)$ ” 계산 과정을 통해 얻은 “pden”, “pnum” 변수들의 ratio — 즉, “pnum” / “pden” 의미

⑦ “w” = “wa1” × “wm1” 계산 통해 얻은 변수

∴ 마지막에 “w” 변수값이 양수인 개체들만 가져오는 조건문 시행

— PS의 부호가 같은(?) 개체만 가져오려는 의도? -- **My guess**

5) allprog2 table

: 위에서 “all” table -> “all2” table 생성 -> “all2” table update 하는 과정을 “allprog” table에 적용해 얻은 tbl

-- Baseline DR이 severe 미만인 개체들 즉, Baseline DR이 weak / moderate인 개체들을 대상으로 2가지 종류(?)의 PS(standardized IPW & generalized PS)를 구하고 두 종류의 PS의 부호가 모두 같은(?) 개체들만 가져와 저장한 tbl이라 추측.