

# <건강검진코호트DB CoxPH model 적합 결과>

- 2월 22일 ~23일 Version

## <What TO DO>

- 변수 변환하지 않고(spline 적용하지 않고) 모든 변수 이용해 Coxph model 적합한 뒤 Exposure 결과 보기
- AGE, AGE^2, SEX 변수만을 공변량으로 하여 Multivariate Coxph model 적합

## <Share & Result>

0) “SODA”, “SNCK”, “과일, 채소 섭취 빈도” 변수의 경우 결측 비율이 높아(10% 이상) model 적합 때 설명변수로 사용하지 않음.

0)-①. 이변량인 “Outcome” 변수 새로 생성 - “Outcome\_category” 값이 “Last\_follow”이면 Outcome = 0 / 나머지는 Outcome = 1

0)-②. “BMI\_Category”에 따라 “Exposure” categorical 변수 새로 생성  
- “unweighted” = 1 ~ “Obesity” = 4

-----  
 $h(t; X) = h_0(t) \exp(X_1^T \beta_1 + \dots + X_P^T \beta_P)$  - Multivariate Coxph model 적합

1) 생성한 “모든 공변량 보정해” Coxph model 적합

1)-①. 변수 값 중 결측이 존재하는 행은 삭제됨.

- model 적합에 사용된 관측치 개수는 총 168,339명 중 133,182명
- 133,182명 중 Event가 발생한 객체는 13,387명

1)-②. CoxPH model 적합 결과 -- “수치형” BMI를 Exposure로 설정 (Categorical로 변환하지 않음.)

Parameter Estimate	P-value	HR
-0.05399	<.0001	0.947

1)-③. CoxPH model 적합 결과 — BMI 변수를 Categorical로 변환

BMI Category	Parameter Estimate	P-value	HR
Moderate	-0.55078	< .0001	0.577
Overweight	-0.54152	< .0001	0.582
Obesity	-0.707	< .0001	0.493

→ HR이 예상과 정반대의 결과가 나옴. (HR이 1이 넘을 것이라 예상하였으나, 현재 model 적합 결과를 해석해보면 BMI Category가 “Underweighted”일 때 Outcome이 발생할 확률이 다른 BMI Category일 때보다 유의미하게 높다.)

/ 반면, “현재 흡연상태”, “Dementia 병력 여부”, “OLIG\_PROTE\_CD” 변수의 HR Ratio가 다른 변수들에 비해 눈에 띄게 큰 값을 가짐을 확인함. (특히 “흡연상태” -- 그러나, 유의한 변수는 아님.)

2) “AGE”, “AGE^2”, “SEX” 공변량만 이용해 Coxph model 적합

2)-①. 변수 값 중 결측이 존재하는 행은 삭제됨.

-- model 적합에 사용된 관측치 개수는 총 168,339명 중 133,182명

-- 133,182명 중 Event가 발생한 객체는 13,387명

2)-②. CoxPH model 적합 결과 -- “수치형” BMI를 Exposure로 설정 (Categorical로 변환하지 않음.)

Parameter Estimate	P-value	HR
-0.05158	<.0001	0.950

2)-③. CoxPH model 적합 결과 -- BMI 변수를 Categorical로 변환

BMI Category	Parameter Estimate	P-value	HR
Moderate	-0.66171	< .0001	0.516
Overweight	-0.65635	< .0001	0.519
Obesity	-0.86779	< .0001	0.420

→ “생성한 모든 공변량”을 보정해 적합한 Coxph model 결과와 크게 다르지 않다.

[TRY TO DO]

: “PERSON\_ID” 변수를 Random effect로 지정하여(객체 자체가 가지는 효과가 존재할 것이라 예상) frailty model을 적합 하려 하였으나, 메모리 문제로 인해 모델 적합 자체가 안됨.

: “insulin” 처방 이력 나타내는 변수를 빼보고 model을 적합해 보았는데 결과가 크게 달라지진 않음.