06강. 로지스틱회귀모형 [2]

■ 주요용어

용어	해설
민감도(sensitivity)와 특이도(specificity)	- P = True Positives+ False Negatives N = False Positives+ True Negatives로 정의할 때 • 민감도(sensitivity) = True Positives/P • 특이도(specificity) = 1-False Negatives/N
ROC 곡선	로지스틱회귀분석에서 예측력을 살펴볼 목적으로 작성하며, 0과 1사이의 모든 cutoff 값에 대해서 이에 해당하는 (민감도, (1-특이도)) 좌표를 나타낸 그래프를 말함
일치성 지수 (concordance index)	ROC 곡선 아래의 면적을 말하며, 로지스틱회귀모형의 예측검정력 측도로 사용됨

정리하기

1. 범주형 예측변수를 표현하는 지시변수(가변수)
   - 두 개의 예측변수 X, Z 와 반응변수Y 가 각각 (0,1)의 값을 갖는 이항변수인 경우(x, z : 지시변수 또는 가변수(dummy variable)라고 함)
   - X와 Z의 주효과를 갖는 로지스틱회귀모형

     $\log it[P(Y=1)]=\alpha+\beta_1 x+\beta_2 z$
   - 모형 $\log it[P(Y=1)]=\alpha+\beta_1 x+\beta_2 z$ 에서 가변수 값에 따른 로짓값

x	z	로짓
0	0	α
1	0	$\alpha+\beta_1$
0	1	$\alpha+\beta_2$
1	1	$\alpha+\beta_1+\beta_2$

   - z가 주어졌을 때 $x=1$에서 "성공"일 오즈는 $x=0$에서 "성공"일 오즈의 $\exp(\beta_1)$배임. $\dfrac{\exp(\alpha+\beta_1)}{\exp(\alpha)}=\exp(\beta_1)$, $\dfrac{\exp(\alpha+\beta_1+\beta_2)}{\exp(\alpha+\beta_2)}=\exp(\beta_1)$
   - 모형에서 교호작용이 없다는 것은 Z의 두 수준에서 구한부분 분할표에 대한 오즈비 값들이 동일하다는 것을 의미함 ⇔ 동질연관성 만족

2. 2×2×K 분할표에 대한 Cochran-Mantel-Haenszel 검정법
   - multi-center clinical trials 사례: 처리(X), 반응(Y), center(Z)
   - $\log it[P(Y=1)]=\alpha+\beta x+\beta_1 c_1+\beta_2 c_2+\cdots+\beta_{k-1}c_{k-1}$

     여기서 x는 X의 두 수준에 대한 가변수임
   - 모형에 대한 다른 형태 표현: $\log it[P(Y=1)]=\alpha+\beta x+\beta_k^z$

     β_k^z : 센터 k의 효과(첫 번째 또는 마지막 센터에 대한 상대적 크기로 표현)

     x : X의 두 수준에 대한 가변수
   - $\exp(\beta)$: K개 분할표에서 Z를 통제했을 때의 X, Y의 공통 오즈비

     [Z를 통제했을 때 X, Y간 조건부독립성 성립 ⇔ $\beta=0$ (X, Y의 오즈비=1)]
   - “$H_0:\beta=0$”에 대한 검정을 통해서 조건부 독립성 검정 가능

     ⇒ 가능도비 검정 또는 Wald 검정

     ⇒ Cockran-Mantel-Haenszel 검정(CMH검정)

3. 다중 로지스틱회귀모형

- Y : 이항반응변수, $\pi = P(Y=1)$
- $x_1, x_2, \cdots, x_k$: k개의 설명변수
- $\text{logit}[P(Y=1)] = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$

$$\Leftrightarrow \ \pi = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k)}$$

  - β_i: 다른 변수들을 통제할 경우, x_i가 미치는 효과
  - e^{β_i}: 다른 설명변수가 고정될 때 X_i가 한 단위 증가할 때 오즈 증가 비(ratio)

4. ROC 곡선 작성

- 분류표

실제/예측	예측		
	$\hat{Y}=1$	$\hat{Y}=0$	계
$Y=1$	True Positives	False Positives	P
$Y=0$	False Positives	True Positives	N

- 민감도(sensitivity) = True Positives/P
- 특이도(specificity) = 1-False Negatives/N
- ROC 곡선: 로지스틱회귀분석에서 예측력을 살펴볼 목적으로 작성하며, 0과 1 사이의 모든 cutoff 값에 대해서 이에 해당하는 (민감도, (1-특이도)) 좌표를 나타낸 그래프를 말함
- 일치성 지수(concordance index): ROC 곡선 아래의 면적을 말하며, 로지스틱회귀모형의 예측검정력 측도로 사용됨

과제하기

구분	내용
과제 주제	- 박태성 & 이승연 (2020) 152쪽 문제 4.9 - 박태성 & 이승연 (2020) 155쪽 문제 4.14
목적	6주차 강의 내용을 복습하고, 로지스틱회귀모형을 실제 데이터에 적용함으로써 자료 분석에 대한 심층적인 이해를 목적으로 함.
제출 기간	7주차 강의 후 1주 후 일요일 밤 12시까지
참고 자료	
기타 유의사항	