



Logistic 회귀분석

Logistic 회귀분석 정의

- 확률 모델로서 사건의 발생 가능성을 예측하는데 사용
- **Binary 한 값**으로 예측하는 분석 방법
- 어떤 독립변수가 연구 결과에 영향을 미치는지 파악할 수 있음

Logistic 회귀분석 함수

- Logistic 회귀는 선형 회귀 기반으로 구성

➤ $t = \beta_0 + \beta_1 x$

- 0과 1 사이의 값을 갖기 위해 Sigmoid 함수 적용

➤ $\sigma(t) = \frac{1}{1 + e^{-t}}$

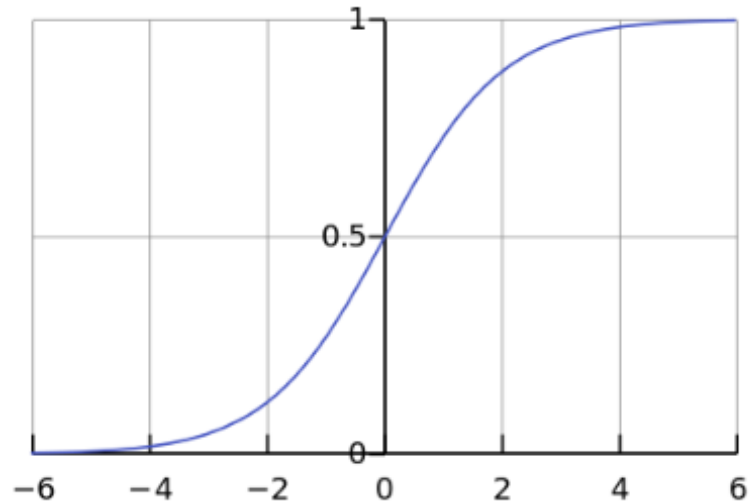
- Logistic 함수

➤ $F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$

β : 회귀계수
 x : 독립변수
 $F(x)$: 종속변수

Logistic Graph

- 일정한 기울기가 변함에 따라 그래프의 커브도 변함



Odds Ratio

- 교차비 or $\text{Exp}(p)$
- 한 집단이 다른 집단에 비해 **성공할 승산 비**에 대한 측정량
- Odds ratio는 Logistic 함수를 통해 얻어짐

$$\text{odds ratio} = \frac{\text{Probability of event}}{1 - \text{Probability of event}}$$

Odds Ratio 예시

- 흡연자가 비흡연자보다 폐암 발생 확률이 더 높은가?

Y/X	흡연자	비흡연자
폐암 발병	82 // p(1)	221 // p(0)
폐암 발병 안함	227 // 1-p(1)	1362 // 1-p(0)

- x=0인 가능성에 대한 x=1인 가능성의 비율

$$\frac{\frac{p(1)}{1-p(1)}}{\frac{p(0)}{1-p(0)}} = 1.824$$

- 흡연자가 비흡연자보다 폐암 발병 확률이 약 1.8배 정도 높다

Logistic 회귀분석 결론

분석하고자 하는 대상이 **2집단으로** 나누어져 있을 때,

ex) 성공과 실패

개별 data들이 **어느 집단으로 분류**되는지

이를 예측하는 **Model을 개발**하는데 사용되는 통계적 알고리즘

Logistic 회귀분석 예제

ex) 지역과 학력에 따른 스마트폰 구매의사 여부 확인

지역 : 1 = 강남

2 = 강북

학력 : 1 = 고졸

2 = 대졸

구매의사 : 1 = 구매

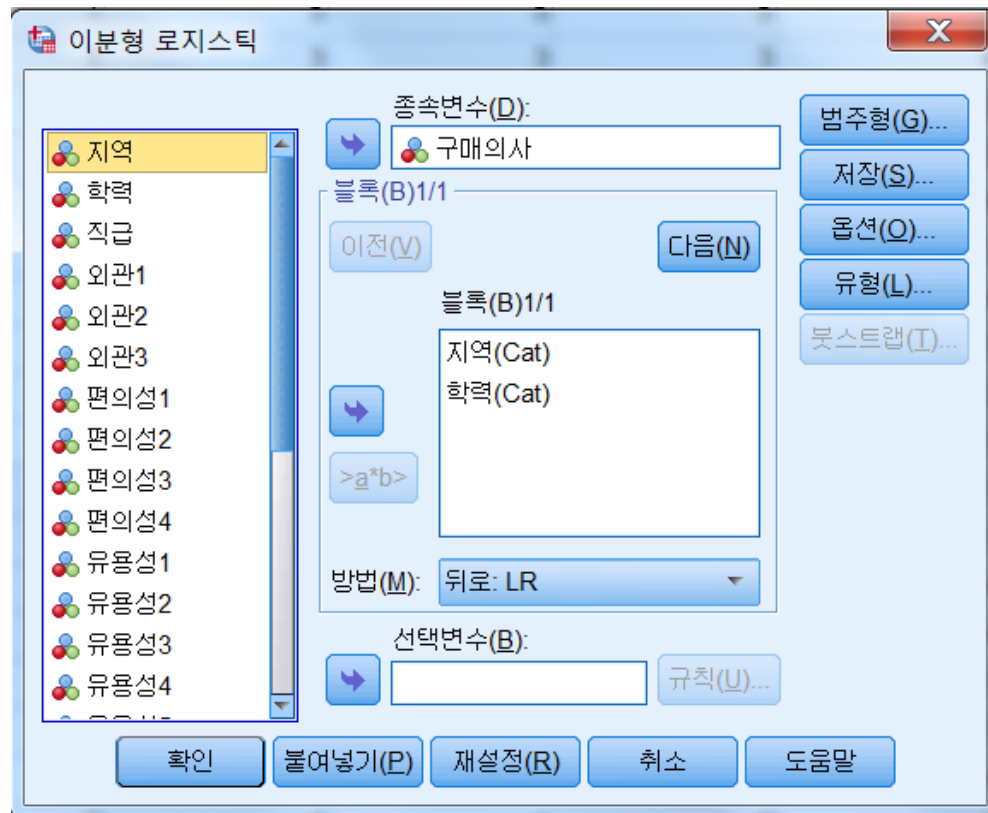
2 = 구매안함

	지역	학력	구매의사
7	1	1	1
8	1	1	1
9	1	1	1
10	1	1	1
11	1	1	2
12	1	1	2
13	1	1	2
14	1	1	2
15	1	1	2
16	1	1	2
17	1	1	2
18	1	1	2
19	1	1	2
20	1	2	2
21	1	2	2
22	1	2	2
23	1	2	2
24	1	2	2
25	2	2	2
26	2	2	2
27	2	2	2

Logistic 회귀분석 예제

ex) 지역과 학력에 따른 스마트폰 구매의사의 여부 확인

➤ 종속변수에 구매의사를 넣고 독립변수에 지역과 학력을 넣음



Logistic 회귀분석 예제

케이스 처리 요약

가중되지 않은 케이스 ^a		N	퍼센트
선택 케이스	분석에 포함	325	100.0
	결측 케이스	0	.0
	전체	325	100.0
비선택 케이스		0	.0
전체		325	100.0

종속변수 인코딩

원래 값	내부 값
구매	0
구매안함	1

Logistic 회귀분석 예제 결론

	변수	B	S.E.	유의수준	Exp(B)
지역	강남 -> 강북	-3.953	.428	.000	0.019
학력	고졸 -> 대졸	.313	.855	.714	1.367

- ‘지역’에 대해서는 유의한 것으로 확인
- 즉 ‘지역’에 대해 -3.953의 음의 상관관계가 있음 확인
- ‘학력’에 대해서는 유의하지 않으므로 구매의사에 학력이 영향을 끼친다고 결론지을 수 없음