로지스틱 회귀분석

# 로지스틱 회귀분석 결과를 확인하기 위한 사전 지식

로지스틱 회귀분석은 회귀분석의 일종이지만, 출력결과에서 확인해야 할 기준이 조금 다르다.

|  |  |  |
| --- | --- | --- |
| **구분** | **일반 회귀분석** | **로지스틱 회귀분석** |
| 회귀모형 검정 | F | Χ2(chi-square test) |
| 회귀계수 검정 | T | Wald 통계량 |
| 회귀식의 설명력 | R2 | Cox 와 Snell 의 R2  Negelkerke 의 R2 |

또한 기존의 회귀분석에서 사용되지 않는 용어가 나온다.

|  |  |
| --- | --- |
| **용어** | **내용** |
| 오즈(odds) | 어떤 사건이 일어날 확률과 일어나지 않을 확률 간의 비율  오즈 = |
| 오즈비(OR: Odds Ratio) | 관측치가 발생할 확률과 발생하지 않을 오즈 간의 비율 |
| 상대적 위험도(RR: Relative Risk) | 관측치와 예측치 간의 연관성 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **구분** | | **예측** | | **합계**  **(오즈)** | **발생률** |
| O | X |
| 관측 | O | a | b |  |  |
| X | c | d |  |  |
| 합계 | | a+c | b+d | a+b+c+d |  |

관측치와 예측치를 비교하여 관측치가 O일 때 예측치가O일 확률이 a, 관측치가 0이지만, 예측치가 X일 확률이 b이다. 여기서 사건이 일어날 확률이 a를 일어나지 않을 확률인 b로 나눈 값을 관측치 O의 ‘오즈(odds)’ 라 한다. 또한, 관측치가 X인데 예측치가 O일 확률이 c, 예측치가 x 이고 일어나지 않을 확률이 d이므로, 관측치 X의 오즈는 c를 d로 나눈 값으로 나타낼 수 있다.

여기서 구한 각 오즈 간의 비율을 ‘오즈비(odds ratio)’라 한다. 즉 관측치 O에 대한 오즈를 관측치 X로 나눈 값으로 구성비를 가늠할 수 있다. 즉 오즈비는 로 나타낼 수 있다. 이때 분자에는 관측치 O를 O로 예측하고 관측치 X를 X로 정확하게 예측한 값이 포함되어 있고, 분모에는 관측치 O를 X로 예측하고 관측치 X를 O로 예측한, 정확하지 못한 값들이 포함되어 있다.

오즈비(OR) = = =

상대적 위험도(relative risk)는 관측치 O에서 예측치가 O일 확률인 를 관측치가 X에서 예측치가 O일 확률인 로 나누어 구한다.

상대적 위험도(RR) = =

이상의 개념으로 보면 상대적 위험도가 전체의 수치를 모두 이용하므로 유용하게 사용할 수 있다고 판단할 수 있다. 그러나 오즈비를 사용하는 이유는 연구를 진행하면서 전체 모수를 이용하는 경우가 매우 드물기 때문이다. 즉 표본을 이용해 전체를 대변하는 수치를 나타내므로 연구에서는 오즈비를 이용하여 판단해야 한다.

# 로지스틱 회귀분석 Tip

* 이분형 로지스틱 회귀변수의 변수입력

범주형 버튼을 클릭하여 ‘로지스틱 회귀: 범주형 변수 정의’를 하는 이유는 ‘학력’ 변수의 범주가 3개 이상이므로 이 범주형 변수를 더미 변수화 해야 하기 때문이다.

* 이분형 로지스틱 회귀분석의 범주형 변수 설정

대비 변경 항목의 참조범주의 ‘마지막’은 변수를 더미화 했을 때 어떤 것을 기준변수로 할 것인지를 표시한 것이다. 이 부분은 출력결과를 보면 확실하게 이해될 것이다.

* 이분형 로지스틱 회귀분석의 변수 투입 방법

방법을 ‘뒤로’ 설정하는 이유는 단계적 회귀분석과 위계적 회귀분석에서 설명하였다. (단계적 / 위계적 회귀분석 방법의 ‘후진’ 과 같은 말이다.) ‘뒤로’에는 다시 변수의 수와 표본의 개수에 따라 ‘ 조건, LR , Wald’ 로 나뉘며, 이 중 가장 보편적인 LR(likelihood ratio)을 선택한다.

* 옵션 설정 배경

1. 분류도표 : ‘구매와 구매 안암’ 과 같이 두 집단이 완전히 구분될 때, 0.5를 기준으로 좌/우 나뉘어 분포하는 도표를 나타낸다.

2. Hosmer-Lemeshow 적합도 : 로지스틱 회귀모형의 전체적인 적합도를 판단하는 검정으로, 유의확률이 유의수준 보다 크면 좋은 모형이다. 이때의 귀무가설이 ‘모형이 적합하지않다’ 이기 때문이다.