Untitled

SeungGyu Bak

2022-10-15

## 5.7 다음은 AIDS 증상을 완화하는 AZT의 효과를 설명하는 연구에서 얻어진 자료이다. 이 연구에서는 AIDS 바이러스가 침투한 후 면역기능이 저하된 338명의 재향 군인에 대하여 AZT를 즉시 복용하는 그룹과 T세포의 면역성이 심하게 약화될 대까지 기다리는 그룹으로 랜덤하게 구분하였다. 연습문제 4.10에 소개된 AIDS 증상, AZT 처리, 인종자료에 대해 R 출력 결과는 다음과 같다.

data(aids)  
fit <- glm(yes/(yes+no) ~ azt + race, weights = yes+no, family = binomial, data = aids)  
summary(fit)

##   
## Call:  
## glm(formula = yes/(yes + no) ~ azt + race, family = binomial,   
## data = aids, weights = yes + no)  
##   
## Deviance Residuals:   
## 1 2 3 4   
## -0.5547 0.4253 0.7035 -0.6326   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.07357 0.26294 -4.083 4.45e-05 \*\*\*  
## aztyes -0.71946 0.27898 -2.579 0.00991 \*\*   
## racew 0.05548 0.28861 0.192 0.84755   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 8.3499 on 3 degrees of freedom  
## Residual deviance: 1.3835 on 1 degrees of freedom  
## AIC: 24.86  
##   
## Number of Fisher Scoring iterations: 4

1 - pchisq(1.3851, 1)

## [1] 0.2392343

cbind(aids$azt, aids$race, fitted(fit), rstandard(fit, type = 'pearson'), residuals(fit, type = 'pearson'),  
 residuals(fit, type = 'deviance'))

## [,1] [,2] [,3] [,4] [,5] [,6]  
## 1 2 2 0.1496245 -1.179418 -0.5447025 -0.5546645  
## 2 1 2 0.2653998 1.179418 0.4281964 0.4252503  
## 3 2 1 0.1427012 1.179418 0.7239488 0.7034699  
## 4 1 1 0.2547241 -1.179418 -0.6219896 -0.6325896

### a. 모형에 대해 적합도 검정을 하고 결과에 대해 해석하라.

1 - pchisq(fit$deviance, fit$df.residual)

## [1] 0.2395008

주어진 사례에서 적합도 검정을 위한 통계량(잔차이탈도)는 1.3835이고, 자유도는 이다. 해당 통계량을 기준으로 적합도 검정을 한 결과 p값은 0.2395로 계산된다. 따라서 귀무가설을 기각할 수 없고, 주어진 회귀모형은 관측 데이터를 잘 설명하고 있다고 할 수 있다.

### b. 적합값의 상대적 크기가 AZT효과와 인종 효과에 대한 개별 검정 결과를 어덯게 반영하는지 설명하라.

로지스틱 회귀모형 적합 결과 변수로서 AZT는 유의수준 0.05에서 귀무가설을 기각할 수 있고, 인종은 귀무가설을 기각할 수 없는 것으로 나타났다. AZT 변수의 회귀계수는 -0.7195로, AZT를 복용한 환자는 그렇지 않은 환자에 비해 에이즈 증상 발현의 오즈가 배일 것으로 예상할 수 있다. 인종변수의 회귀계수는 0.0555로, 백인 환자의 경우 에이즈 증방상 발현이ㅡ 오즈가 배일 것으로 예상할 수 있다. 따라서 모형을 기초로 에이즈 발병률을 계산한 적합값은 동일한 인종 내에서 AZT 복용군의 에이즈 발병률이 미복용군의 에이즈 발병률에 비해 약 절반 정도의 수치를 가지고, AZT 복용 여부가 동일한 군 내에서 백인은 흑인에 비해 에이즈 발병률이 약간 높을 것이다.

사례에서 주어진 에이즈 발병률 적합값은 이러한 모형을 반영하고 있다. 즉, AZT 복용군의 에이즈 발병률은 미복용군에 비해 백인의 경우 , 흑인의 경우 으로 나타난다. 한편 백인의 에이즈 발병률을 흑인과 비교하면, AZT 복용군의 경우 , 미복용군의 경우 인 것으로 나타난다.

### c. 제시된 R의 출력값은 표준화잔차, Pearson 잔차 및 이탈도 잔차를 보여준다. 다른 잔차에 비해 표준화잔차를 사용하는 경우의 장점에 대해 설명하라.

위 표에서 피어슨 잔차와 이탈도잔차는 각 수준에 대해 다른 값을 보이고 있지만 표준화잔차는 동일한 값을 보이고 있다. 주어진 모형에서 잔차 자유도는 1이기 때문에 1개의 잔차만을 가져야 한다. 하지만 피어슨 잔차와 이탈도 잔차는 각각 다른 값을 보여주고 있어서 모형의 특성을 잘 보여주지 못하는 단점이 있다.

이처럼 로지스틱 회귀모형을 진단하는 지표로서 표준화잔차는 모수중복(개별 변수가 서로 분리되지 않고 교호작용 변수의 형태로만 식별되는 현상)을 보여줄 수 있기 때문에 피어슨 잔차 또는 이탈도 잔차보다 더 많은 정보를 제공해준다고 할 수 있다.