

## 논문리뷰 : 과대산포가 존재하는 이항형 자료의 화귀분석방법

[Reference : 연세대학교 의과대학 의학통계학과교실]

서론 :

### 1\_ 문제점 거론 : 과대산포(Overdispersion)

Logistic Regression 을 통해 이분형 분석을 실시할 때에, 자료의 변동이 심한 문제로 Logistic Regression 으로 설명되지 못한 Variation 이 존재한다.

#### 1-1\_ 문제세부형태 :

- 개체간 변동(Inter – subject variation)
- 임의효과(Random effect)

: 반응변수들이 군집표본추출(Cluster sampling)에 의해 관찰되었을 때, 이항모수(Binomial parameter)가 군(Cluster)에서 군으로 변화함을 가정할 수 있다.

: 회귀계수에 대 한 추정치는 심각하게 편향되어(biased)있지는 않지만, 표준오차가 정확하지 않게 된다.

첨부 : 포아송 회귀에서의 일반적 문제와 유사한 양상이다. 모형을 통해 추출된 응답이 기대보다 과한 확률로 상대적 값을 반환한다.

$$D = -2[\log\text{-likelihood}(\text{model}) - \log\text{-likelihood}(\text{saturated})] = 2\sum[y_i \log(y_i / \hat{y}_i) - (y_i - \hat{y}_i)]$$

### 2\_ 문제해결방안 : overdispersed logistic regression model

#### 2\_1\_1 기존모형 정의 :

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

#### 2-1-2\_ Odds ratio 에 대한 95% 신뢰구간 정의

$$\text{Odds ratio} = \exp(\hat{\beta}_j)$$

$$95\% \text{ Confidence interval} = \exp(\hat{\beta}_j \pm 1.96 \times \text{sej}) \quad (j \text{ 번째 변수에 해당하는 회귀계수의 추정치 및 표준오차: } \hat{\beta}_j, \text{sej})$$

## 2-2-1\_ Overdispersed logistic regression model

Variance =  $\Phi p(1-p)$  ( $\Phi$  : 과대산포의 정도를 측정하는 모수) \*\*\*

**Table 1.** Classification of the overdispersed binomial model

Model	
$\Phi = 1$	Logistic regression model
$0 < \Phi < 1$	Underdispersed logistic regression model
$\Phi > 1$	Overdispersed logistic regression model

: Overdispersed logistic regression model 을 이용하여 회귀계수를 추정하면 통상적으로  $\hat{\beta}_j$  는 동일

반면, se가 다르게 구해진다. 그러므로 Odds ratio 는 동일하나 95%신뢰구간이 다르게 구해진다.

### 선행연구 참고 :

- Williams – 베타-이항 모형(beta – binomial model)에 기초
- Kim – transform -both - sides