



기초 주어진 Data => 전립선암 (prostate cancer)

이론 y (결과) => Capsule penetration (전립선 투과 여부)

x (설명변수) => PSA, gleason

전립선 투과 여부
 => 양과 관련된 전립선 부피가
 크면, 전립선 투과 확률이 높아짐
 => 전립선 암의 정도를 판별하는 표준법

공통 용어 (전립선 전암화 + PSA 마크로 노드)
 권선1 권선2
 => 전립선암이 얼마나 빨리 퍼져나가는 지 정도

y는 $\begin{cases} 1 : \text{Capsule involvement (영양분)} \\ 0 : \text{not Capsule involvement (no 영양분)} \end{cases}$

generalized linear regression

가정 model $\leftarrow \text{glm}(y \sim x_1 + x_2, \text{family} = "binomial", \text{data} = df)$

summary(model)

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -7.691 + 0.21 \cdot \text{PSA} + 1.059 \cdot \text{gleason}$$

Capsule gleason이 각각 1점과 2점이면, y의 값이 1인 확률은 1.06%
 이 때 z값을 $e^{1.06\%}$ 가 되고 odds는 2.89이다.
 (한번)

odds ratio의 신뢰구간을 구해보자 -- 95%

$$\hat{p} \pm 1.96 \cdot \text{se}(\hat{p})$$

회귀분석의 결과값을 z test

(+) 회귀분석의 결과값을 z test

$$L(\beta_0, \beta_1) = p_1^{y_1} (1-p_1)^{1-y_1} \cdot p_2^{y_2} (1-p_2)^{1-y_2} \cdots p_n^{y_n} (1-p_n)^{1-y_n} =$$

• 결과값 => (x, y), ..., (x_n, y_n)

$$\bullet X = x_i \text{에 대해 } p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

$$\left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{1-y_i}$$

$$= \frac{[\exp(\beta_0 + \beta_1 x_i)]^{y_i}}{1 + \exp(\beta_0 + \beta_1 x_i)} = \frac{\exp(\beta_0 y_i + \beta_1 x_i y_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

$$\ln \frac{p}{1-p} = x$$

$$\frac{p}{1-p} = e^x$$

$$p = (1-p)e^x$$

$$p + p e^x = e^x$$

$$p = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}$$