Odds Ratio

Sayantee Jana

Odds Ratio

$$OR = Odds1 \ / \ Odds2 = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Alternative measure

- Relative Risk(RR) : RR = p1 / p2

Why use OR?

- Easy to interpret
- Popular since in most practical situations we have binary data
- **invariant under study design**
- Good **mathematical properties**

Example :

Cross-Classification of Smoking by Lung Cancer

	Lung Cancer	
Smoking status	Cases	Controls
Y	688	650
N	21	59
Total	709	709

- Odds of LC for smokers = 688/21= Ω 1 ,(say)
- Odds of No LC for smokers=650/59 = Ω2,(say)
- **OR= Ω 1/ Ω 2**=2.97≈3
- s.e. (log(OR)) = $\sqrt{\frac{1}{688}+\frac{1}{650}+\frac{1}{21}+\frac{1}{59}}$ =3.847
- P value=Pr(log(OR)/s.e.(OR)>obs value|H_0)= 0

Interpretations

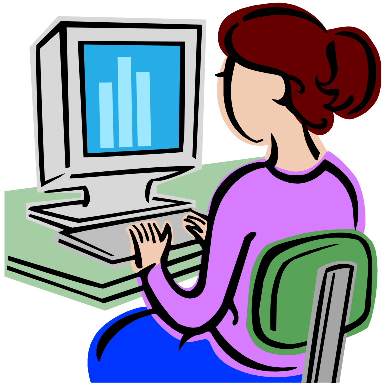
- OR=1 => disease and exposure status are **independent**

- OR>1 => smokers are more likely to have LC than non-smokers i.e. **Positive association** between exposure and disease

- OR<1 => non-smokers are more likely to have LC than smokers i.e. **negative association** between exposure and disease

Software

R :Command: summary(glm(Y ~ X , data = , family=binomial(link="logit")))



Different ways of calculating OR

From contingency tables

	Y	
X	Y=0	Y=1
X=0	a	b
X=1	c	d

- H_0 : OR=1
- OR= (a*d)/(b*c)

$$s.e.(\log(OR)) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$$

- Z_teststat=log(OR)/s.e.(OR)

Logistic Regression Model : Gender vs Status

Let $\pi_0 = P(J|F)$, $\pi_1 = P(J|M)$

$\log\{\pi/(1-\pi)\} = \text{intercept} + \text{gender} * \beta$

joining status= 1, if joined
joining status=0, if not joined

gender = 1 , F
= 0 , M

log of odds(π_0)= intercept + β
log of odds(π_1)= intercept

log odds ratio comparing F to M
= log odds for F-log odds for M
=(intercept +β.1)-(intercept +β.0)

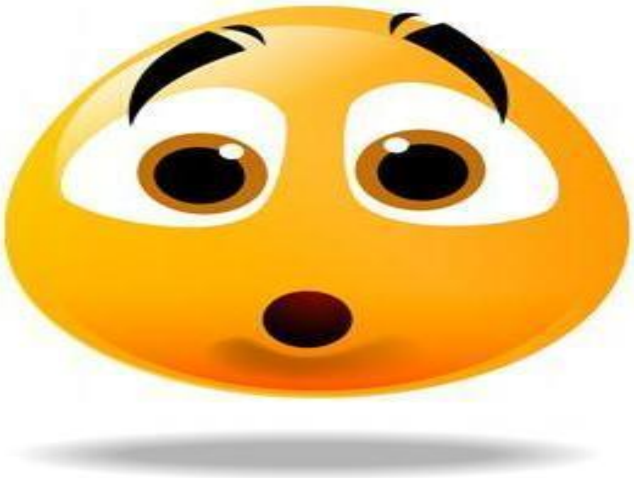
Ho: **log(OR) = 0 ≈ ß = 0**

Maximum likelihood equation

- m= no. of F
- n=no of M

$$l = \prod_{1}^{m} \pi_0 \prod_{1}^{n} \pi_1$$

Why Logistic Regression instead of 2x2 tables ?



- multiple categories.
- multiple covariates
- Dependent variable is always a binary outcome.
- Independent variables may be categorical or quantitative.