Conditional Logistic Regression

PubH 7450

©Wei Pan

Email: weip@biostat.umn.edu

Http: www.biostat.umn.edu/∼weip

§1 Matched Case-Control Study

- 1-M match: 1 case is matched with M controls;
  matching variable: potential confounder $\Longrightarrow$ ...;
  $M \geq 1$: to improve efficiency;
  there may be ≥ 1 risk factor/covariate to be investigated.
- More generally, $n_i - M_i$ match in set i.
- Example: Low birth weight data.
  Ref: Le, CT (1998). Applied Categorical Data Analysis.
  Example 5.9.
  1-3 matching; $n = 15$ matched sets;
  matching variable: mother's age;
  4 covariates: mother's body weights (in pounds); hypertension status; smoking status; uterine irritability.

TABLE 5.2. Low Birthweight Data

Matched Set	Case	MotherWeight	Hypertension	Smoking	U-Irritability
1	1	130	0	0	0
	0	112	0	0	0
	0	135	1	0	0
	0	270	0	0	0
2	1	110	0	0	0
	0	103	0	0	0
	0	113	0	0	0
	0	142	0	1	0
3	1	110	1	0	0
	0	100	1	0	0
	0	120	1	0	0
	0	229	0	0	0
4	1	102	0	0	0
	0	182	0	0	1
	0	150	0	0	0
	0	189	0	0	0
5	1	125	0	0	1
	0	120	0	0	1
	0	169	0	0	1
	0	158	0	0	0
6	1	200	0	0	1
	0	108	1	0	1
	0	185	1	0	0
	0	110	1	0	1
7	1	130	1	0	0
	0	95	0	1	0
	0	120	0	1	0
	0	169	0	0	0
8	1	97	0	0	1
	0	128	0	0	0
	0	115	1	0	0
	0	190	0	0	0
9	1	132	0	1	0
	0	90	1	0	0

2-1

§2 CLR

- For simplicity, first consider 1-m_i matching for set i.
- Notation: for matched set $i = 1, ...n$,
  case: x_i, $y_i = 1$;
  controls: x_{ij}, $y_{ij} = 0$, $j = 1, ..., m_i$;
- LR model:

$$\text{Logit}Pr(Y = 1|X, \text{ set } i) = \alpha_i + X'\beta.$$

  note **matched set-dependent** intercepts; why?
  interpretation of β:
- How to infer β?
- Use the standard likelihood $L(\alpha_1, ..., \alpha_n, \beta)$?
  How? Why or why not?

- Use conditional likelihood:

$$L_c = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} \frac{\exp(x_i'\beta)}{\exp(x_i'\beta) + \sum_{j=1}^{m_i} \exp(x_{ij}'\beta)}.$$

  L_i looks like ...

- Derivation:

$$p_i = Pr(Y_i = 1 | x_i, \text{ set } i) = \frac{\exp(\alpha_i + x_i'\beta)}{1 + \exp(\alpha_i + x_i'\beta)},$$

  $1 - p_i = ...$
  similarly, $p_{ij} = Pr(Y_{ij} = 1 | x_{ij}, \text{ set } i) = ...$
  $1 - p_{ij} = ...$

- Two key probabilities:
  A =Pr(the case has disease and controls do not in set i)
  $= p_i \prod_{j=1}^{m_i}(1 - p_{ij}) = \frac{\exp(\alpha_i + x_i'\beta)}{1+\exp(\alpha_i + x_i'\beta)} \prod_{j=1}^{m_i} \frac{1}{1+\exp(\alpha_i + x_{ij}'\beta)}$

  B=Pr(only one subject has disease in set i)

$= p_i \prod_{j=1}^{m_i}(1-p_{ij}) + (1-p_i)p_{i1}(1-p_{i2})...(1-p_{im_i}) + ... = ...$

- $$\begin{aligned} L_i &= \Pr(\text{the case has disease}|\text{only one has disease in set } i) \\ &= A/B = ... \end{aligned}$$

- More generally,

$$L_i = \frac{\prod_{j\in \text{Cases}} \exp(x'_{ij}\beta)}{\sum_{S\in R(n_i,m_i)} \prod_{k\in S} \exp(x'_{ik}\beta)},$$

where $R(n_i, m_i)$ is the set of all partitions of the n_i-m_i matched set into two parts with the first containing n_i subjects and the second m_i subjects.
L_i looks like ...

- How to operate on $L \Longrightarrow$ ...?
use PL, e.g. in SAS Proc Phreg.

- How? Convert the CLR into a PH regression problem:

  1) Create a dummy time variables such that its value for any case is always smaller than that of any control;

  2) For any case, its dummy time is for an event; for any control it is for an censoring;

  3) Stratified by set i; see §9.3 to be discussed;

  4) Fit a PHM.

- Example: Low birth weights data.

§3 Application to clustered binary data

- Multiple observations from the same cluster may be correlated.
  Multiple members from the same family from a familial study.
  Study on twins; study on two eyes, kidneys,...;
  Multiple observations on the same subject in a longitudinal study.
- A matched set is a cluster.
- Notation: for subject j in cluster i, we have a binary response Y_{ij}, and covariates X_{ij}.
- Logistic regression model:

$$\text{Logit}Pr(Y_{ij} = 1|X_{ij}) = \alpha_i + X'_{ij}\beta.$$

  Why α_i?
  Similar to a random-effects model?
  Different from a random-effects model?

- How to infer β?
  use the CL in CLR!
- An example: multi-center clinical trial data.