Lab 8

Conditional Logistic Regression

Henian Chen, M.D., Ph.D.

Parameters Estimation for Linear Regression

Ordinary Least Squares (OLS)

Statistical methods based on the minimization of the squared differences between the observed and predicted values of Y.

Parameters Estimation for Logistic Regression

Likelihood

The probability that a score or set of scores could occur, given the values of a set of parameters in a model.

Maximum Likelihood (ML)

A method for the estimation of parameters based on the principle of maximizing the likelihood of the sample.

Parameters Estimation for Ordinary (unconditional) Logistic Regression

Unconditional Likelihood (L_u)

$L_u = \pi P(X_i)\ \pi[1-P(X_i)]$

$= \pi \exp(\alpha+\beta_i X_i)\ /\ \pi[1+\exp(\alpha+\beta_i X_i)]$

$\pi P(X_i)$ = the probability for the event

$\pi[1-P(X_i)]$ = the probability for not having the event

```
proc logistic data=case_control978 descending;
model status=alcgrp;
run;
```

Parameter	DF	Estimate	Standard Error	Chi-Square	Wald Pr > ChiSq
Intercept	1	-2.5911	0.1925	181.1314	<.0001
alcgrp	1	1.7641	0.2132	68.4372	<.0001

`OR=exp(1.7641)=5.836`

1:1 Conditional Logistic Regression

Description of Data (CASECONTROL11.DBF)

1:1 matched case-control study of low birth weight (<2500 grams) babies. 56 matched pairs. Controls were age-matched mothers who gave birth to a baby above 2500 grams.

ID	identification variable for each matched pair
Status	0= GE 2500 grams (controls), 1=<2500 grams (cases)
AGE	age of mother in years
LWT	pre-pregnant weight, weight in pounds at last menstrual period
RACE	race of mother (1=white, 2=black, 3= other)
SMOKE	smoking status during pregnancy (1=yes, 0=no)
PTD	history of premature labor (1=yes, 0=no)
HT	history of hypertension (1=yes, 0=no)
UI	presence of uterine irritability (1=yes, 0=no)

The goal of the analysis is to determine whether the low birth weight is associated with any of above explanatory variables.

Parameters Estimation for Conditional Logistic Regression (1)

We cannot use unconditional likelihood (L_u) for matched case-control study. Using L_u for matched case-control data, the α_i is the effect of the pair effect, the β_i is the effect of X_i. Since there are only two observations in each pair, you can't estimate the α_i without bias.

Parameters Estimation for Conditional Logistic Regression (2)

We have to include 56-1=55 dummy variables for the strata if we want to use unconditional logistic regression for matched case-control study. This would leave us with possibly 64 parameters being estimated for a data set with only 112 observations.

Furthermore, increasing the sample size will not help because an additional stratum parameter would have to be estimated for each additional matched set in the study sample.

Parameters Estimation for Conditional Logistic Regression (3)

Conditional Likelihood (L_c)

$L_c = \pi P(X_i)\ \pi[1-P(X_i)] \ / \ \Sigma\{\Sigma P(X_i)\ \pi[1-P(X_i)]\}$

$= \pi \exp(\Sigma \beta_i X_i) \ / \ \Sigma[\pi \exp(\Sigma \beta_i X_i)]$

$\pi P(X_i)$ =probability of pair's case having the event

$\pi[1-P(X_i)]$ = probability of pair's control not having the event

$\Sigma\{\Sigma P(X_i)\ \pi[1-P(X_i)]\}$ = the joint probability that either the case or control has the event.

Parameters Estimation for Conditional Logistic Regression (4)

Conditional Likelihood (L_c)

1. L_c can use the information contained in the matches
2. L_c can't estimate the intercept (α). With no α in conditional logistic regression model, we can't estimate the P(x).
3. L_c can estimate the odds ratio by using the β_i, so we can estimate the other effects (except intercept) in which we are interested.

L_c vs. L_u

We can use conditional likelihood (L_c) for unmatched case-control study:

1) L_c treats the unmatched case-control data as one stratum.
2) Conditional likelihood (L_c) always gives you the unbiased estimation.

We can ***not*** use unconditional likelihood (L_u) for matched case-control study. Because L_u omits the information inherent in the matching process. It is incorrect to treat 56 strata as one stratum.

1:1 Conditional Logistic Regression (1)

SAS Program

```
proc import datafile= 'a:casecotrol11.dbf' out=casecontrol11
            dbms=dbf replace;
run;
data casecontrol11;
set casecontrol11;
status1=1-status;
race1=0;
if race=2 then race1=1;
race2=0;
if race=3 then race2=1;
run;
proc phreg;
model status1 = age lwt race1 race2 smoke ptd ht ui
                /selection=forward ties=discrete rl;
strata=id;
run;
```


1:1 Conditional Logistic Regression (2)

Status1 (case=0,control=1): Probability of being a case is modeled

proc phreg: Procedure PHREG performs both Cox regression for survival data, and conditional logistic regression for matched case-control studies

selection=forward: start with a model containing none of the independent variables and then considers variables one by one for inclusion

ties= option specifies how to handle ties in the failure time
- = BRESOW: uses the approximate likelihood of Breslow. this is the default value
- = DISCRETE: replaces proportional hazards model by the discrete logistic model
- = EFRON: uses the approximate likelihood of Efron
- = EXACT: computes the exact conditional probability under the proportional hazards assumption

The DISCRETE method is required :1) if the dependent variable is discrete 2) there is more than one case in a matched set of case-control study

rl: estimate the 95% confidence limits

1:3 Conditional Logistic Regression

Description of Data (CASECONTROL13.TXT)

1:3 matched hospital based case-control study. Cases are women diagnosed with benign breast disease from two hospitals. Controls were selected from other patients at the same two hospitals.

ID	stratum number
SUBJECT	observation within a matched set (1=case, 2-4= controls)
AGEINTER	age of the subject at the interview
STATUS	diagnosis (1=case, 0=control)
MCHECK	regular medical checkup history (1=yes, 0=no)
AGEP	age at first pregnancy
AGEM	age at menarche
NONLIVEN	number of stillbirths, miscarriages, and other non live births
LIVEN	number of live births
WEIGHT	weight of the subject
AGELM	age at last menstrual period

1:3 Conditional Logistic Regression

SAS Program

```
proc import datafile= 'a:casecotrol13.txt' out=casecontrol13 dbms=tab
replace;
getnames=yes;
run;
data casecontrol13;
set casecontrol13;
status1=1-status;
run;
proc phreg;
model status1 = ageinter mcheck agep agem nonliven liven weight
                agelm /selection=forward ties=discrete rl;
strata=id;
run;
```


We can *not* use unconditional logistic regression for matched case-control study, but we can use conditional logistic regression for unmatched case-control study.

Unconditional model
proc logistic data=case_control978 descending;
model status=alcgrp;

Parameter	β	SE	OR	95% Confidence Limits	
alcgrp	1.7641	0.2132	*5.836*	3.843	8.864

Conditional model (*ties=discrete*)
proc phreg; model status1*status(0)= alcgrp / ties=discrete;

Parameter	β	SE	OR	95% Confidence Limits	
alcgrp	1.76231	0.21315	*5.826*	3.836	8.847

Conditional model (*default value for ties*)
proc phreg; model status1*status(0) = alcgrp;

Parameter	β	SE	OR	95% Confidence Limits	
alcgrp	1.47319	0.2008	*4.363*	2.944	6.467