

§6. Conditional Logistic Regression (CLR) for Matched or Stratified Data

§6.1 Overview

Have considered “unconditional” regression, e.g., logistic

Results in estimate for intercept (α) which corresponds to the baseline risk (risk in group with no risk factors present) and estimates of “slope” (β 's) which represent departure from baseline risk when one or more factors is present.

If there 5 centers in a study, unconditional regression will need 4 dummy variables to account for differences across those centers.

For example, suppose we have a simple two-arm randomized study.

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \alpha_0 + \alpha_1 \cdot h_1 + \alpha_2 \cdot h_2 + \alpha_3 \cdot h_3 + \alpha_4 \cdot h_4 + \beta \cdot tx$$

Then we assume that there are 5 possible “intercepts” or baseline risks in the control groups across the 5 hospitals (0,...,4) and then a common treatment effect.

But this method requires that we estimate 5 parameters for the hospital. For 30 hospitals, this means 30 parameters.

Copyright © 2005 Trustees of the University of Pennsylvania

“Conditional” regression – what do we mean?

- (a) Have levels of strata (age category) or centers (hospitals) or pairs (matched)
- (b) Wish to estimate association of a within-strata exposure (drug exposure) and outcome
- (c) Have different intercept for each stratum, but these levels are “nuisance” parameters – we do not care about them, and we cannot estimate them using this method. They are “conditioned out” of the analysis
- (d) We want to condition these out in order to address the problem of heterogeneity in the baseline risk (the risk of outcome in the reference group of patients)

Example:

In a group of 30 ICUs we are analyzing the association between a new antibiotic and outcome, and we have randomly selected equal numbers of treated and controls within each clinic.

Each ICU serves a somewhat different population, and the different populations each has a somewhat different risk of outcome among the controls. But we do not care about estimating or modeling the baseline risk. We are concerned only about estimating the effect of treatment.

These ICUs form “strata”

ICU=1		Antibiotic	
Infection		+	-
	+	25	50
	-	75	50

For this ICU the baseline (reference) risk =0.5. OR = 0.33

*
*
*

ICU=30		Antibiotic	
Infection		+	-
	+		20
	-		80

For this ICU the baseline (reference) risk =0.2. But we are not going to estimate the baseline risks. We are interested only in the overall OR of the association of antibiotic and infection.

If we were to use unconditional logistic regression (logit or logistic), then we would have to include 29 indicator (dummy) variables for the 30 ICUs to allow for differences in baseline (reference group) risks across the ICUs.

Conditional logistic regression works in nearly the same way as regular logistic regression, except we need to specify which individuals belong to which matched set (e.g., which pair) or stratum.

The theory is similar: we can derive a likelihood and maximize it, etc. From a practical perspective, the only difference is the need to specify the matched set or the stratum to which each person belongs.

Comparisons within the matched set

Recall from stratified analysis what subtables (matched set) are informative:

	E+	E-		E+	E
D+	1	0		0	1
D-	0	1		0	1
	Informative			Uninformative	

For the conditional analyses to be able to estimate effect of exposure (E), must have variation of E within the matched set.

So, for CLR we are making comparisons of E+ vs E- *within* the matched sets

In unconditional regression, the comparisons are across (*among*) sets.

§6.2 Example: (Kelsey et al—Table 7.11) Layout for a pair-matched case-control study of prolapsed lumbar disc, with place of residence and whether or not a person drives considered as risk factors.

				Control			
				Suburban residence		City residence	
				Does drive	Does not drive	Does drive	Does not drive
Case	Suburban residence	Does drive	63	4	32	22	
		Does not drive	1	2	1	2	
	City residence	Does drive	29	1	20	14	
		Does not drive	7	0	10	9	

McNemar's Test (collapsing the table) $OR_{MLE} = 41/19 = 2.16$

		Case		
		Does not drive	Does drive	
Control	Does not drive	13	41	54
	Does drive	19	144	163
		32	185	217

STATA commands and output for this example:

Check the data

```
. tab driving
driving |          Freq.    Percent    Cum.
-----|-----
      0 |           86     19.82     19.82
      1 |          348     80.18    100.00
-----|-----
    Total |          434    100.00
```

```
. tab suburbs
suburbs |          Freq.    Percent    Cum.
-----|-----
      0 |          200     46.08     46.08
      1 |          234     53.92    100.00
-----|-----
    Total |          434    100.00
```

```
. tab casecon
casecon |          Freq.    Percent    Cum.
-----|-----
      0 |          217     50.00     50.00
      1 |          217     50.00    100.00
-----|-----
    Total |          434    100.00
```

Running the regression models

Generating the coefficients

```
. clogit casecon driving, strata(pairid)
```

```
Iteration 0: Log Likelihood =-150.41294
Iteration 1: Log Likelihood =-146.29234
Iteration 2: Log Likelihood =-146.28399
Iteration 3: Log Likelihood =-146.28399
```

```
Conditional logistic regression
```

```
Number of obs = 434
chi2(1) = 8.26
Prob > chi2 = 0.0041
Pseudo R2 = 0.0275
```

```
Log Likelihood = -146.28399
```

```
-----+-----
casecon |      Coef.   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
driving |   .7691331   .2775281    2.771   0.006   .2251881   1.313078
-----+-----
```

Estimating odds ratios

Model with driving alone

```
. clogit casecon driving, strata(pairid) or
```

```
Iteration 0: Log Likelihood =-150.41294
Iteration 1: Log Likelihood =-146.29234
Iteration 2: Log Likelihood =-146.28399
Iteration 3: Log Likelihood =-146.28399
```

```
Conditional logistic regression
```

```
Number of obs = 434
chi2(1) = 8.26
Prob > chi2 = 0.0041
Pseudo R2 = 0.0275
```

```
Log Likelihood = -146.28399
```

```
-----+-----
casecon | Odds Ratio   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
driving |   2.157895   .5988763    2.771   0.006   1.252558   3.717599
-----+-----
```

Model with suburbs alone

```
. clogit casecon suburbs, strata(pairid) or
```

```
Iteration 0: Log Likelihood =-150.41294
Iteration 1: Log Likelihood =-148.26942
Iteration 2: Log Likelihood =-148.26893
```

```
Conditional logistic regression
```

```
Number of obs = 434
chi2(1) = 4.29
Prob > chi2 = 0.0384
Pseudo R2 = 0.0143
```

```
Log Likelihood = -148.26893
```

casecon	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
suburbs	1.540541	.325236	2.047	0.041	1.018519	2.330114

Model with both main effects (driving and suburbs)

```
. clogit casecon driving suburbs, strata(pairid) or
```

```
Iteration 0: Log Likelihood =-150.41294
Iteration 1: Log Likelihood = -145.6509
Iteration 2: Log Likelihood =-145.64015
Iteration 3: Log Likelihood =-145.64015
```

```
Conditional logistic regression
```

```
Number of obs = 434
chi2(2) = 9.55
Prob > chi2 = 0.0085
Pseudo R2 = 0.0317
```

```
Log Likelihood = -145.64015
```

casecon	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
driving	1.930682	.5675763	2.238	0.025	1.085118	3.43514
suburbs	1.291056	.291563	1.131	0.258	.8293075	2.0099

Testing the driving * suburbs interaction using xi:

```
. xi: clogit casecon i.driving*i.suburbs, strata(pairid) or
i.driving      Idrivi_0-1 (naturally coded; Idrivi_0 omitted)
i.suburbs      Isubur_0-1 (naturally coded; Isubur_0 omitted)
i.driving*i.suburbs IdXs_#-# (coded as above)

Iteration 0:  Log Likelihood =-150.41294
Iteration 1:  Log Likelihood =-145.61448
Iteration 2:  Log Likelihood =-145.60245
Iteration 3:  Log Likelihood =-145.60245

Conditional logistic regression                Number of obs =   434
                                              chi2(3)         =    9.62
                                              Prob > chi2    = 0.0221
Log Likelihood = -145.60245                  Pseudo R2      = 0.0320
```

```
-----+-----
casecon | Odds Ratio   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
Idrivi_1 |   1.996329   .6366903     2.168  0.030     1.068459   3.729978
Isubur_1 |   1.558695   1.122786     0.616  0.538     .3798472   6.396072
IdXs_1_1 |   .8140014   .6085232    -0.275  0.783     .1880582   3.523369
-----+-----
```

Could do the likelihood ratio test or just look at the Z-test for the single interaction term. Likelihood ratio has broader use – can apply to sets of interaction terms as when one of the factors has more than one level (we have seen this before).

“xi” syntax: clogit work the same as regular logit:

```
. xi: clogit casecon i.driving i.suburbs, strata(pairid) or
i.driving      Idrivi_0-1 (naturally coded; Idrivi_0 omitted)
i.suburbs      Isubur_0-1 (naturally coded; Isubur_0 omitted)

Iteration 0:  Log Likelihood =-150.41294
Iteration 1:  Log Likelihood = -145.6509
Iteration 2:  Log Likelihood =-145.64015
Iteration 3:  Log Likelihood =-145.64015

Conditional logistic regression                Number of obs =   434
                                              chi2(2)         =    9.55
                                              Prob > chi2    = 0.0085
Log Likelihood = -145.64015                  Pseudo R2      = 0.0317
```

```
-----+-----
casecon | Odds Ratio   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
Idrivi_1 |   1.930682   .5675763     2.238  0.025     1.085118   3.43514
Isubur_1 |   1.291056   .291563      1.131  0.258     .8293075   2.0099
-----+-----
```

Some notes on implementation:

In all of these results, there is no intercept (`_cons`) estimate.

Obtain only estimates for “slopes” (betas), which represent the effect of the exposures of interest, driving and suburbs, conditioning on the matching factors, or the effect of exposures within the matched pairs.

By choosing the variable “paired” as the matching factor (conditioning factor) we are controlling for all of the factors that go into the match

If matching variable were sibship, for example, then would be conditioning on all of the factors that make two sibs alike (except for where they live and whether they drive). This might be weight, height, anatomy of the back and spine.

§6.3 Conditional vs unconditional logistic regression

Question: Instead of using the conditional analysis on thematching factors, we use unconditional logistic regression controlling for region and age group. What happens?

Breslow et al. 1978 Oesophageal cancer study matched on village, age, gender

Variables in equation	Fully matched			7 regions, 4 age groups		
	beta	se	OR	beta	se	OR
Social class	-1.125	0.254	0.325	-0.808	0.212	0.446
Ownership of garden	-0.815	0.250	0.443	-0.614	0.222	0.541
Consumption of raw green vegetables	-0.552	0.220	0.576	-0.459	0.203	0.632
Consumption of cucumbers	-0.640	0.217	0.527	-0.539	0.196	0.553
Log likelihood	-187.69			-388.27		

We have seen this before. If you do not account for the matching, can lead to bias of the results to null. [Some info is missing.]. We have seen this bias before when we considered the issue of “noncollapsibility”. The conditional analysis will be farther from the null, but the standard errors will be somewhat larger for the fully conditional analysis.

Notice, also, that the conditional analysis has a higher (less negative) log likelihood, which suggests a somewhat better “fit”.

So, if match in design, should match in analysis (or at least try it out and see if it makes a difference.)

Note: Do not present results this way. Convert coefficients to OR s and compute confidence intervals. Do not require the reader to do this work.

§6.4 Conditional logistic regression to estimate interaction terms

Conditional regression involves a stratified analysis on the matching or stratification factors. One cannot estimate the effect of matching factors alone. These are nuisance parameters that are not estimated.

But *one can estimate an interaction* between a matching factor and another risk factor. We might want to do this in a “pre-post” design.

Example: (From Hennessy S, et al. Unpublished. 2002)
Courtesy of the Centers for Education and Research on Therapeutics (CERTs), Agency for Healthcare Research and Quality.

A study of the effectiveness of three programs to reduce prescribing of antibiotics for respiratory infections.

Question: Can we reduce unnecessary prescribing of antibiotics.

Setting: 28 primary care physicians.

Design: 3 interventions: high, moderate, none. “pre-post” design

Endpoint: The number of persons who were examined and the fraction who received a prescription for antibiotics.

Statistical problem:

We want to ensure that we compare the pre-intervention with the post-intervention experience within each practice, and then we wish to combine those individual comparisons into an overall effect size.

How do we do that?

If we used ordinary logistic regression, we would lose the ability to adjust the post-test rates of prescribing by the pre-test rates.

So:

```
. tab group prepost
```

group		prepost		Total
		0	1	
High	1	130	109	239
Low	2	97	99	196
Control	3	208	238	446
Total		435	446	881

First, we need to make the reference group the baseline using

```
char group[omit] 3
```

```
. xi: clogit abx_prescribed i.group*prepost, group(md) or
i.group      _Igroup_1-3      (naturally coded; _Igroup_3 omitted)
i.group*prepost  _IgroXprepo_#      (coded as above)
```

note: multiple positive outcomes within groups encountered.

note: 1 group (32 obs) dropped due to all positive or all negative outcomes.

note: _Igroup_2 omitted due to no within-group variance.

note: _Igroup_1 omitted due to no within-group variance.

Iteration 0: log likelihood = -461.84413

Iteration 1: log likelihood = -458.3484

Iteration 2: log likelihood = -458.34304

```
Conditional (fixed-effects) logistic regression      Number of obs      =      846
LR chi2(3)                                          =      9.76
Prob > chi2                                         =      0.0207
Pseudo R2                                          =      0.0105
Log likelihood = -458.34304
```

abx_prescr~d	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
prepost	1.145387	.2355383	0.66	0.509	.7654392	1.713934
_IgroXprep~1	.3567393	.1304362	-2.82	0.005	.1742297	.7304316
_IgroXprep~2	.7763375	.3098916	-0.63	0.526	.3550384	1.697562

Note first, that the main effects of group fall out of the model. These are not estimated because MD is a stratification factor, and the MD determines the group effect. So, we cannot estimate the main effect of group (the type of intervention) and STATA knows this.

But is that an important question? Do we want to know whether overall (pre and post combined) the groups differed in their prescribing? No, we want to know whether the groups differed in their effect over time.

Second, this output is not completely helpful. Why? Because there are two interaction terms. What do we do?

We can obtain the effect of the high group=1 against the baseline directly from the output.

```
_IgroXprep~1 | .3567393 .1304362 -2.82 0.005 .1742297 .7304316
```

The odds of prescribing an antibiotic dropped 65% from the pre to the post period when this group is compared to the experience of the controls (Group=3). (Of course, one must not interpret the OR as a relative risk because the percentages are very high). We obtained a statistically significant effect in Group=1. This happened to be the high intervention group.

We can next determine the relative effect of the low intensity group.

```
_IgroXprep~2 | .7763375 .3098916 -0.63 0.526 .3550384 1.697562
```

There were only slight and non-significant differences between the low intensity group and the controls.

Finally, we have to use `lincom` to get the contrast between the high and low groups.

```
. lincom _IgroXprepo_1-_IgroXprepo_2, or
( 1) _IgroXprepo_1 - _IgroXprepo_2 = 0.0
```

```
-----
abx_prescr~d | Odds Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
(1) | .4595157   .209798    -1.70   0.089    .1877902    1.124418
```

So, the low intensity group did not fare much better than the controls, the high intensity group did much better than the controls, and the two intervention groups did not differ significantly.

We could reparameterize the model to use any group as the baseline or reference, and then use `lincom` to obtain any contrast we like.

All we need to do is invoke the syntax “char varname [omit] x” where x is the baseline level of the variable varname (in our case group). But using the control as baseline makes sense.

Summary: we can use conditional logistic regression to compare treatments in prepost designs. When the individual patients are not followed over time, but only the groups are followed and the patients change with time, they are sometimes called “repeated cross sectional studies”.

§6.5 Conditional Logistic Regression (summary)

1. Appropriate analysis of matched studies requires a stratified (by the matching factor) model in which each matching stratum has its own intercept α_k , where there are k matching strata. The intercept is just another way of allowing each matching stratum to have its own risk of event/outcome.
2. The k intercepts are NOT computed – i.e., we DO NOT use dummy variables to adjust for each level of matching.

Suppose there are k matched sets of 2×2 tables. Then:

$$\text{logit}(y) = \alpha_k + \beta x, \text{ or}$$

$$pr(y = 1|x) = \frac{\exp\{\alpha_k + \beta x\}}{1 + \exp\{\alpha_k + \beta x\}}$$

3. Because matching terms are subsumed in the intercept, which is not computed, one cannot model (usually) the matching factors.

One can, however, estimate finer stratification factors within one of the matching factors. So, if age is a matching category in 10 year intervals, one can still model (using special parameterization) age within the strata.

4. Interactions and product terms – one cannot estimate the matching factors alone, but *one can estimate an interaction* between a matching factor and another risk factor.
5. With the many strata, stratification produces sparse data (but the CLR algorithm is designed to handle sparse data).

6. CLR model cannot handle large strata well – computation becomes difficult. In these cases use unconditional logistic regression and use a “fixed effect” (dummy or indicator variable) for each stratum. This is a type of conditional regression, but there is a separate intercept for each stratum:

For example, you have 10 strata and a binary treatment:

Then the model is:

$$\text{logit}(E(y)) = \alpha + \alpha_2 S_2 + \dots + \alpha_{10} S_{10} + \beta x$$

If one assumes that each strata also has a separate slope (treatment effect), then need a separate slope for each stratum, so need a stratum*slope factor for the model. End up with many covariates, so have to have much data.

But this fixed effects regression generated biased estimates of effect size when there are many strata. Therefore, use with caution.

7. When strata are very fine, one cannot use ordinary LR because there are too many fixed effects (one for each matched-pair or matched group) to estimate. MLE leads to *seriously biased* estimates. Exercise extreme caution when there are many strata compared to the size of the sample. (Point, as above, is NOT TO USE many dummy variables to code the matching strata).

8. For extremely sparse data, there are exact CLR algorithms (LogXact) analogous to exact methods for stratified 2 by 2 tables.
9. CLR is an extension of McNemar’s test, just as logistic regression is an extension of the analysis of 2 by 2 tables or stratified 2 by 2 tables.
10. Can have (a) 1:M matched data, with more than one control per case (or more than one unexposed to exposed), (b) M need not be the same for each matched set of observations, such as when a control is lost. If the case is lost, then the entire matched set is lost, however. If all subjects within a matched set are concordant on exposure, that set is lost.
11. Loss of a matched set can occur if, for example, have a missing X for one observation in a matched set. So, must exercise great care to have complete data for as many observations as possible in matched analysis.
12. Case control and prospective studies.

The example we use (from Kelsey) involves a matched case-control study, but these methods can be used for prospective paired studies (pre-post design, study of twins or family members).

References: Rothman and Greenland pp 420-22; Breslow and Day, Vol I, chap VII (1980)

End of Vol II, Part 4