Exercise 2.3 Conditional logistic regression

1. Using the same data as for Exercise 2.1 (**2.1_framingham_exercises.dta),** generate a new age category variable with age in decades **agecat2** (< 40, 40-49, 50-59, 60-70). Select a matched case-control sample of diabetes cases and one control per case, matched on **agecat2.** Make sure to set a "seed" before the sampling so that you can reproduce the exact same selection if necessary.

    Save as a temporary data set. We will use these data to examine the association of prevalent diabetes with prevalent CHD.

    (i)   Create the _paired_ 2-by-2 table of the association between CHD and Diabetes for individuals aged 40 years and above.

    (ii)   Conduct the McNemar test of association using your software or your own calculations based on the counts in the table, or both.

    (iii)   Calculate the paired OR with the 95% confidence interval using your software or your own calculations based on the counts in the table, or both.

2. The dataset "**RSV_data_exercise.dta**" contains data from the 1:2 matched case-control study of severe RSV disease in infants in Kenya (Nyiro JU, Sande CJ, et al. Absence of Association between Cord Specific Antibody Levels and Severe Respiratory Syncytial Virus (RSV) Disease in Early Infants: A Case Control Study from Coastal Kenya. PLoS One. 2016 Nov 16;11(11):e0166706). The variables of interest for this exercise are:
matchset:	a unique number identifying a case and their controls
controlno:	control number (1,2)
case:	indicator for case (1) or control (0)
ntitre0:	antibody titre (on log scale)
bwt:	birth weight (Kg)
underwt:	indicator for low birth weight
gest_agew:	gestational age (weeks)
premature:	indicator for prematurity

    (i)   Open this data set and create an indicator for a high titre ($>=$ the median rounded to 2 decimal place). Save a temporary copy of the data.

    (ii)   Drop one control from each set (Note that it may not be necessary to use a "seed" this time if we make some reasonable assumption!). Save a temporary copy of the data.

    Examine the 2-by-2 table of the case-control pairs and from this table, compute the paired OR and 95% confidence interval for the association between high titre and RSV.

    (iii)   Repeat the computation of the paired OR in (ii) using conditional logistic regression.

    (iv)   For the continuous measure of antibody levels, use the temporary data set from (i) with all the infants from the 1:2 matched study to reproduce the results described in the following text from the paper by Nyiro et al:

The overall reduction in risk of RSV disease per $\log_2$PRNTunit increase in cord blood antibody titres in all infants was 21% (OR = 0.79, 95% CI 0.47±1.31; P = 0.4).

The $\log_2$PRNT antibody titre in cord blood showed an increase with increasing gestational age. However, adjusting for prematurity (i.e. gestational age < 37 weeks) in the conditional logistic regression analysis did not have an appreciable effect on the outcome (OR = 0.67, 95% CI 0.19±2.4; P = 0.5).

HINTS

Ex 1: Stata immediate commands csi (cross-sectional data) and cci (case-control data)
Ex 3:
Ex 4: Test for trend
OpenEpi: Dose-Response command
Stata: tabodds command

Ex 5: Stata immediate commands csi (cross-sectional data) and cci (case-control data)

Ex 13: Stata tabodds command