Statistical Methods in Epidemiology
Lab 8.
Matched Case – Control Studies

I. Matched studies

Matched sets should be coded as individual records which contain a variable d, for case/control status, and a variable, `set`, to identify which matched case-control set each subject belongs to. For a 1:m matched study there will be 1+m records for each matched set. Matched case control studies can be analyzed with Mantel-Haenszel by using case-control sets as strata.

i) Endometrial cancer, 1:1 matched sets

In the study of endometrial cancer and the medical use of estrogens, described in Breslow and Day (1980) each case was matched to 4 control women who were alive and living in the retirement community at the time the case was diagnosed, who were born within one year of the case, had the same marital status and who had entered the community at approximately the same time.

The variable `age3` codes subjects in three board groups (55-64, 65-74 and 75+) using the age of the case in each set, so subjects in the same matched set always have the same value for `age3`. The variable `agegrp` does the same thing for 5 year age groups. Breslow and Day start by analyzing the 1:1 study formed by using only the first control in each set, which is in the file `bdendo11`. The full data may be found in the file `bdendo` which includes the following variables:

`set:`	Case – control set: a numeric vector
`d:`	Case or control: a numeric vector (1=case, 0=control)
`gall:`	Gall bladder disease: a factor with levels `No Yes`.
`hyp:`	Hypertension: a factor with levels `No Yes`.
`ob:`	Obesity: a factor with levels `No Yes`.
`est:`	Use of estrogens: a factor with levels `No Yes`.
`dur:`	Duration of conjugated estrogen therapy: an ordered factor with levels `0<1<2<3<4`.
`non:`	Use of non estrogen drugs: a factor with levels `No Yes`.
`duration:`	Months of estrogen therapy: a numeric vector.
`age:`	A numeric vector.
`cest:`	Conjugated estrogen dose: an ordered factor with levels `0<1<2<3`.
`agegrp:`	A factor with levels `55-59 60-64 65-69 70-74 75-79 80-84`.
`age3:`	A factor with levels `<64 65-74 75+`

1. Examine the effect of estrogen with Mantel-Haenszel using the data in bdendo11. Find the number of matched pairs in which the case is exposed but not the control, and the number of matched pairs in which the control is exposed but not the case. Verify that the Mantel-Haenszel estimate is the ratio of these two numbers.

2. Because cases and controls were matched on age, the effect of age cannot be estimated from the study. You can verify it by trying to find the two effects of `age3`.

3. Controlling the effect of estrogen for `age3` makes no difference because age has already been controlled for in the `set` variable. However, it does give information about whether age modifies the effect of estrogen. Try this and check whether age modifies the effect of estrogen.

4. Any attempt to control for further variables, other than those used in the matching, will subdivide the matched sets into new strata and, because strata without cases and strata without controls are uninformative, there may be little information left. To illustrate this, try to control the effect of estrogen from hypertension. The number of strata goes up from 63 to 94, but only 18 contribute to the effect of estrogen.

ii) Endometrial cancer, 1:4 matched sets

The full study including all 4 controls per case can be analyzed in the same way. Remember that all matched sets contribute to the effect of an exposure except those in which all controls are exposed or all controls are unexposed.

1. Find the effect of exposure to estrogen. How many matched sets contribute to the effect of estrogen?
2. Try to find the effects of `age3`.
3. Check whether `age3` modifies the effect of estrogen.
4. Find the effect of estrogen controlled for `hyp`. How many extra strata are created? How many contribute to the effect of estrogen?

II. Conditional logistic regression

1. Read in the data from bdendo11 and examine the effect of estrogen with conditional logistic regression.

2. Try to find the effects of `age3`.

3. Examine whether the effect of estrogen is modified by `age3`. To do this you will need to fit a model with interactions between `est` and `age3`, and test the interaction parameters. Try this first using a likelihood ratio test and then using a Wald test.

4. Do the same thing using the variable `age` which is metric.

5. Examine the effect of estrogen controlled for `hyp`. There is no loss of information due to creating new strata as there was with Mantel – Haenszel.

6. Examine whether `hyp` modifies the effect of estrogen using by fitting the model with interaction, and testing the interaction parameters.

7. We can analyze the 1:4 study in the same way but, the Mantel-Haenszel methods are no longer precisely the same as maximum conditional likelihood, although they usually agree quite closely. Read in the data in `bdendo` and analyze them using conditional logistic regression.

III. Conditional vs. unconditional logistic

1. Read in the leprosy data and obtain the effect of BCG controlled for age (categorical) using ordinary logistic regression.

2. Create a variable `grp` which takes the value 1 for all subjects and use this to obtain the effect of BCG controlled for age (categorical) using conditional logistic regression. How close are the answers?