November 10 2016
Morten Frydenberg

POSTGRADUATE COURSE IN
LINEAR REGRESSION MODELS FOR CONTINUOUS AND BINARY DATA
Homework

Part A
(Identical to Exercise 7.1. Day 7 in the Postgraduate course in Basic Biostatistics)

The file *postterm.dta* contains the post term delivery data used at the lectures day 7 (day7.pdf). Below we will go through the analysis of these data in Stata.
We will use parity=0 as reference.

1. Make a two by two table showing the association between parity and post term delivery. Calculate, by hand, the odds for post term delivery in each of the two parity groups (no CI's).

2. Run the Stata command `cs ptd parity, or woolf`.
   Find all the numbers on page 7.

3. Run the Stata command `logit ptd b0.parity`.
   Find the log odds for the reference group.
   Find the log odds ratio with standard error and confidence intervals on page 14.
   The output also contains two z-tests, what do we test here?

4. Run the Stata command `logit ptd b0.parity, or`.
   Find the odds ratio with confidence interval and the z-test.
   ln(0.9) =- 0.10536 so run `lincom 1.parity +0.10536`
   to get the test for OR=0.9 on page 15.

Generate a new variable `age30= age-30`.
We will now look at the association between age and the risk of post term delivery among women with parity ==0.

5. Run the Stata commands
   ```
   logit ptd age30  if parity==0
   logit, or
   lincom _cons,or.
   ```
   Find all estimates, confidence intervals, and the tests shown on page 23.

6. Calculate, based on the output, the odds ratio (with CI) for post term delivery comparing two women (both having given birth to their first child), who differ 10 years in age.
   Run the Stata commands `lincom 10*age30,or` and find the result .

Back to comparing the parity groups.

7. Run the Stata command `logit ptd b0.parity##c.age30.`
   Find the estimates with confidence interval and the z-test from page 32.

8. Run the Stata command `logit ptd b0.parity age30.`
   Find the estimates with confidence interval from page 34.
   Run the Stata command `logit, or.`
   Find the age-adjusted OR with confidence interval from page 34.
   Comment on the z-test.

Part B

Here you will be introduced to the `binreg` command, relative risk models and risk difference models. The exercise contains some plots, so I have made a do- file with the codes `HomeworkPartB.do`

9. There is another command in Stata (`binreg,or`) that will estimate a logistic regression. Run

```
logit  ptd b0.parity age30,or
binreg ptd b0.parity age30,or
```

   and compare the output.

10. After a binary regression model like this you can find the estimated probability for the event by:

```
predict p_or if e(sample), mu
```

    Do this and make a plot illustrating the relationship between parity, age and the outcome postterm delivery (use the do file).

The `binreg` commando can also be used to work with regression models that use relative risks or risk differences as measures of association.

11. Run the commands

```
cs ptd parity , or woolf
binreg ptd b0.parity ,rd
binreg ptd b0.parity ,rr
```

    find the relative risk in the output from the second command

and the risk difference in the in the output from the last command.

12. Similarly, you can estimate a relative risk model:

$$\ln\left[\Pr(postterm)\right] = \alpha_0 + \alpha_1 \cdot [parity > 0] + \alpha_2 \cdot [age - 30]$$

Or equivalently:

$$\Pr(postterm) = \exp\{\alpha_0 + \alpha_1 \cdot [parity > 0] + \alpha_2 \cdot [age - 30]\}$$
$$= \exp\{\alpha_0\} \cdot RR_{parity} \cdot RR_{age}^{[age-30]}$$

By `binreg ptd b0.parity age30,` **`rr`**
Do this and try to understand the estimates:
You can also find and plot the estimated probabilities as you did when you used a logistic regression model , consult the do file for details.

13. And we can fit a risk difference model

$$\Pr(postterm) = \gamma_0 + \gamma_1 \cdot [parity > 0] + \gamma_2 \cdot [age - 30]$$

By `binreg ptd b0.parity age30,` **`rd`**
Do this and try to understand the estimates:
You can also find and plot the estimated probabilities as you did, when you used a logistic regression model, consult the do file for details.

14. In all of the above we have modelled the risk of postterm delivery.
We can also consider at the "risk" of not having a postterm delivery.
We can do this by using `notpostterm` as the dependent variable:

```
generate notpostterm=1-ptd
binreg notpostterm  b0.parity c.age30, or
binreg notpostterm  b0.parity c.age30, rr
binreg notpostterm  b0.parity c.age30, rd
```

Compared the output of this with what you found in 9), 12) and 13).