

THE ANALYSIS OF HIERARCHICAL AND OTHER DEPENDENT DATA

Computer Practicals

LSHTM MSc in Medical Statistics

Practical 0: Revision of linear regression modelling

Data and programs for the course

As this is the first computer practical of the course you will need to **start by copying all the data and programs needed for the course**. These are in `U:\Download\Teach\MedSt_Hier`. Please copy them all onto a dedicated folder in your own workspace. Remember to make sure that you set the working directory of Stata to be that folder so that you can easily access data and ado programs.

Data

We are going to use the dataset already introduced in the lectures.

The `births` dataset is in the Stata file called `births.dta` and includes records of 500 independent births in a London hospital. The variables are:

<code>id</code>	identity number of mother and baby
<code>bweight</code>	birth weight in gms
<code>lowbw</code>	coded 1 for birth weight < 2500, 0 otherwise
<code>gestwks</code>	gestational age in weeks
<code>preterm</code>	coded 1 for gestwks < 37
<code>matage</code>	maternal age in years
<code>hyp</code>	hypertension in mother 1=yes, 0=no
<code>sex</code>	sex of baby 1=male, 2=female
<code>sexalph</code>	sex of baby "male", "female"

Questions

1. Start Stata, load the `births` dataset and examine its content with `describe`.
2. Do some housekeeping: label the contents of the variables `sex` and `hyp` with

```
. label define hyp 0 "normal" 1 "hypertensive"  
. label value hyp hyp  
. label define sex 1 "boy" 2 "girl"  
. label value sex sex
```

3. We are going to focus on the relationship between birth weight and maternal hypertension. Start with some graphical exploration:

```
. graph hbox bweight, over(hyp) by(sex,total)
```

What do you see? Compute some summary statistics equivalent to this plot using the `tabstat` command.

4. Use ANOVA and the t-test to assess whether babies born to hypertensive mothers have the same mean birthweight as those born to mothers with normal blood pressure. What are your conclusions?
5. Visually check whether the linear relationship between birthweight and gestational age is modified by maternal hypertension, for example with:

```
. scatter bweight gest, by(hyp) ytitle("Birthweight (g)") ylabel(,angle(h))
```

What do you think?

6. Fit univariable (i.e. separate) linear regression models of birthweight on maternal hypertension and on gestational period (possibly centred, as discussed in the lecture). Then fit a model that includes them both and then another that also includes their interaction. What do you conclude?
7. Assess the appropriateness of its assumptions via graphical examination of the residuals using `predict` to extract the residuals.