Practical 5:
Longitudinal data I

Data

1. The `Beta-blocker trial`
   The data come from a randomised, double blind, placebo-controlled trial to establish the efficacy of beta-blockers for reducing blood pressure in patients with abnormally high blood pressure, or *hypertension*.
   The data are held in the file called `beta_blocker.dta`. The variables are:

```
-------------------------------------------------------------------------------
id              patient identifier
treat           treatment: 1=active; 0=placebo
pre             pre randomisation diastolic BP in mmHg
dbp1            diastolic blood pressure post randomisation visit 1
dbp2            diastolic blood pressure post randomisation visit 2
dbp3            diastolic blood pressure post randomisation visit 3
dbp4            diastolic blood pressure post randomisation visit 4
dbp5            diastolic blood pressure post randomisation visit 5
dbp6            diastolic blood pressure post randomisation visit 6
-------------------------------------------------------------------------------
```

Questions

1. Load and familiarise yourself with the data. What type of data are these?

   Summarise the variables `pre` and `dbp1-dbp6` separately by treatment group using `tabstat`:

```
. label define treat 0 "P" 1 "A"
. label value treat treat
. tabstat pre dbp*,by(treat) s(count mean sd)
```

   What do you notice about the number of observations as the time between baseline and measurement increases? Can you guess whether there is any treatment difference and/or time effect on the blood pressure measurements?

2. Produce a plot of mean DBP by treatment as follows:

```
. preserve
. collapse pre dbp1-dbp6,by(treat)
. rename pre dbp0
. reshape long dbp, i(treat) j(time)
. gen dbp_act=dbp if treat==1
. gen dbp_pl=dbp if treat==0
. twoway (line dbp_act time,sort) (line dbp_pl time, lpat(dash) sort)
. restore
```

Note that `preserve` and `restore` allows you to collapse the data to produce the plot but then return to their original format.

3. Calculate the pairwise correlation among all the post-treatment DBP measures, by treatment:

```
. bysort treat:pwcorr dbp*, obs
. pwcorr dbp*, obs
```

Do you detect any patterns?

4. Now reshape the data in long format in order to model the repeated observations per individual and flag one observation only per patient with:

```
. reshape long dbp, i(id) j(time)
. egen pickone=tag(id)
```

Check that the reshaping has worked without corrupting the data.

5. Check the distribution of DBP at each time point:

```
. hist dbp,by(time)
```

Is it normal?

6. Fit the following random intercepts model with `mixed`:

$$\text{DBP}_{ij} \quad = \quad (\beta_0 + u_{0j}) + \beta_1 \text{time}_i + \beta_2 \text{treat}_j + \beta_3 (\text{pre}_j - \overline{\text{pre}}) + e_{ij}$$

where $\overline{\text{pre}}$ is the overall mean of baseline DBP, $u_{0j} \sim N(0, \sigma_u^2)$, $e_{ij} \sim \text{IIN}(0, \sigma_e^2)$ and $u_{0j} \perp e_{ij}$. (Be careful when you calculate mean DBP at baseline!)

Interpret the meaning of all estimated coefficients.

7. Calculate the level 1 (repeated observations) and level 2 (patient) residuals, using the `predict` command:

```
. predict r_inter, reffects
. predict stres, rstandard
```

Plot them and check the model's assumptions.