Practical 4: The random coefficient model

Data

1. The `GCSE data`.
   These data hold information on the General Certificate of Secondary Education (GCSE) score at age 16 and reading level before entering the school at age 11 (measured by the London Reading Test (LRT) score) of each pupil attending 65 schools, as well as other pupil and school level characteristics.

   Data are held in `gcse_selected.dta`. The variables are:

```
school          School identifier
student         Student identifier
gcse            Standardized GCSE score (multiplied by 10)
lrt             Standardized LRT score (multiplied by 10)
girl            Student female sex (1=yes, 0=no)
schgend         type of school (1=mixed gender, 2=boys only, 3=girls only)
```

Questions

1. Load and familiarize yourself with the GCSE data. Create the indicator variable that picks up only one record per school and tabulate type of school:

```
. egen pickone=tag(school)
. label define schgen 1 "mixed" 2 "boys only" 3 "girls only"
. label values schgen schgen
. tab schgen pickone,col
```

2. Generate the mean of the GCSE and LRT score for each school. Generate also a variable holding the average number of girls in each school. Then summarize them.

```
. egen ave_gcse = mean(gcse), by(school)
. egen ave_lrt = mean(lrt), by(school)
. egen ave_girl = mean(girl), by(school)
. summ ave_* if pickone
```

3. Repeat some of the analyses presented in the lecture. Drop school 48 and then fit the fixed effect model for GCSE on LRT:

```
. drop if school==48
. reg gcse lrt ibn.school, nocon
```

4. Replace `ibn.school` with `i.schgen` and allow for the constant term to be included. Does this school-level variable explain as much of the school variability as the individual indicators? What happens to the estimated slope for `lrt`?

5. Fit the random intercept model of `gcse` on `lrt` using REML:

   ```
   . mixed gcse lrt || school:, reml stddev
   ```

   What is the null hypothesis for the LRT test at the bottom of the output? Interpret its finding. Make a note of the value of the log-restricted likelihood at the maximum.

6. Add the the variable `schgen` to the model as an explanatory variable for the intercept. Test its significance.

7. Fit the random intercept and slope model of `gcse` on `lrt` using REML:

   ```
   . mixed gcse lrt i.schgen || school: lrt, reml cov(unstructured) stddev
   ```

   Make a note of the value of the log-restricted likelihood at the maximum.

8. Test the joint significance of the random slope variance and the random slope and intercept covariance (when the models include `schgen`) by hand (compared to the model that only includes random intercepts) and using the `lrtest` command. What do you conclude?

9. Retaining `schgen` in the model with both random intercepts and random slopes, compare the model with an unstructured matrix with a model assuming independence. what can you conclude?

10. Load the data again and examine school 48. Fit a linear regression model just on school 48. Compare its regression line with that of all the other schools as seen in the lecture:

    ```
    . reg gcse lrt if school==48
    . twoway (scatter gcse lrt)(lfit gcse lrt) if school==48
    ```

11. Refit the random intercept and random slope model with the variable `schgen` included as explanatory for the intercept:

    ```
    . mixed gcse lrt i.schgen  || school: lrt, reml cov(unstructured) stddev
    ```

    Have the results changed because of the inclusion of school 48?

12. Calculate the predicted level 2 residuals with:

    ```
    . predict ebslope ebinterc if pickone==1,reffects
    ```

    Note that `ebslope` and `ebinterc` are names chosen by the user: you could replace them with any other name of your choice!

    Check their distribution (even if they are not standardized). Examine the residuals for school 48. Are they similar to the results you obtained when fitting the regression model only on school 48?

13. Calculate the predicted standardised level 1 residuals with:

    ```
    . predict rst,rstandard
    ```

    Check their distribution. Examine the values for school 48.