

Practical 2: The random intercept model

Data

1. The GHQ data.

These data hold the response given by 12 students on two occasions to the 12-item General Health Questionnaire (GHQ), an instrument to measure psychological distress. The data are held in the file called `ghq` and its variables are:

| | |
|-------------------|--|
| <code>id</code> | Student identifier |
| <code>GHQ1</code> | General Health Questionnaire score- 1st occasion |
| <code>GHQ2</code> | General Health Questionnaire score- 2nd occasion |

Questions

1. Load and summarise the GHQ data. Are the data balanced? Calculate and summarize the mean score for each student with

```
. gen mean_GHQ=(GHQ1+GHQ2)/2  
. summ mean_GHQ
```

Check its values with `list` or `browse` and take a note of its mean and SD.

2. Reshape the data in long format with:

```
. reshape long GHQ, i(id) j(occasion)
```

and use the command `xtsum` to summarise the data again. Compare the results with those found in answering the previous question (it helps to create a new variable that holds the within person difference in GHQ score with `gen dif_GHQ= (GHQ-mean_GHQ)`).

3. Carry out an ANOVA of the repeated GHQ measures using the command `lone way`.

```
. lone way GHQ id
```

Make sure you understand the output. Compare the estimated SDs of the between and within effects with the sample statistics found in answering the previous question.

4. Fit a random effects model to these data using REML with the command `mixed`:

```
. mixed GHQ || id: , reml stddev
```

Compare the results with those found with `loneaway`.

5. Now re-fit the model using ML:

```
. mixed GHQ || id: ,ml stddev
```

Have any of the results changed?

6. Finally fit a fixed effect model to these data:

```
summ GHQ
. qui gen d_GHQ=GHQ-r(mean)
. reg d_GHQ ibn.id,nocons
```

To check the SD of the fitted intercepts you could save them and summarize them as follows (note that their mean is 0 because of the parametrization we have used):

```
. matrix A=e(b)
. svmat A, names(alpha)
. egen mean_id=rmean(alpha1-alpha12)
. egen SD_id=rsd(alpha1-alpha12)
. list mean_id SD_id in 1,noobs
```

Compare the results with those obtained earlier.

7. Fit a linear regression model on these data ignoring the clustering:

```
. reg GHQ
```

What happens?

8. Refit the simple regression model of question 7 but this time use the `robust cluster(id)` option. What happens?