Practical 7:
Longitudinal data III

Data

1. The `cornea transplant data`.
   The data arise from a study of 80 cornea transplant patients. The transplant consisted of replacing a central disk of the patient's cornea with a disk taken from a donor's cornea. For most patients sutures (surgical stitches) were removed after about 1.5 years. To assess the visual quality of the patient after the transplant, a photo of the affected eye was taken at regular follow-up visits, with a visual score derived from each photo.

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

   ```
   id      patient ID
   visit   visit number
   date    Time since operation (years)
   dispcat Disparity between donor and patient's eye (categorical)
   medium  Medium where cornea was preserved
   score   Visual score
   ```

Questions

1. Load and familiarize yourself with the cornea transplant data. In particular examine the patient level (cluster) characteristics after creating an indicator variable that picks up only one record per patient. For example try:

   ```
   . egen pickone=tag(id)
   . tab dispcat medium if pickone
   ```

2. Examine the distribution of number of visits per patient using:

   ```
   . sort id visit
   . qui by id:gen totv=_N
   . egen totvcat=cut(totv), at(0,5,10,15) label
   . tab totvcat if pickone
   ```

3. Compare whether patients with different characteristics have more visits than others:

```
. tab totvcat dispcat if pickone,col
. tab totvcat medium if pickone,col
```

4. Plot the individual trajectories of the patients using:

```
. sort id date
. twoway (line score date,c(ascending)), ylabel(,angle(h)) ///
       ytitle(Visual score) legend(off)
```

   Do you think they evolve linearly with time?

5. Before starting to model the visual scores examine whether they need to be transformed. Because patients have different numbers of visits we need to do this separately by visit:

```
. hist score, by(visit)
```

6. Start building a mixed effects model for these data by specifying a (close to) 'saturated' model for the mean structure (i.e. a model that includes all potential explanatory variables, polynomial terms if some are continuous, and all their interactions). For the covariance structure allow random intercepts and random slopes for the time variable (`date`):

```
. gen date2=(date)^2
. mixed score date date2 i.medium i.dispcat c.date#i.medium c.date#i.dispcat   ///
    c.date2#i.medium c.date2#i.dispcat i.medium#i.dispcat ///
    i.medium#i.dispcat#c.date i.medium#i.dispcat#c.date2 ///
      || id: date, reml cov(unstr)
```

7. Consider now a different covariance structure, where the level 1 errors vary with donor-recipient disparity:

```
. mixed score date date2 i.medium i.dispcat c.date#i.medium c.date#i.dispcat   ///
    c.date2#i.medium c.date2#i.dispcat i.medium#i.dispcat ///
    i.medium#i.dispcat#c.date i.medium#i.dispcat#c.date2 ///
    || id: date, reml cov(unstr)  residuals(independent, by(dispcat))
```

8. Compare these two models using a (restricted) LRT.

9. Without further checks, let's assume that neither a level 2 heterogeneity model nor models with only random intercepts fit the data better than this last model. Now simplify the specification of the mean structure (remember to respect the hierarchy of the terms).

10. What is your final model? Interpret the results. Check whether the residuals are well behaved as follows:

```
predict u_slope u_inter, reffects
predict rst,rstandard
hist rst, name(plot1)
qnorm rst, name(plot2)
```

```
graph combine plot1 plot2
graph combine plot1a plot2a
hist u_inter if pickone
qnorm u_inter if pickone, name(plot1a)
hist u_slope if pickone
qnorm u_slope if pickone, name(plot2a)
graph combine plot1a plot2a
```