Written examination (2 pages) - 2 hours

Instructions
You can use your computer and the course material (lectures and exercise sheets). Upload your solutions directly into ILIAS (under the exam section of the course). The upload format should be similar to the exercises: a report in PDF, Open Office, Word (or any other format of your choice) that includes the *commented* R code you wrote, snapshots of any plots you made and a clear narrative about your results and conclusions. All these elements will be taken into account in the grading. Datasets can be found on ILIAS, under the exam tab. All types of offline/online communications/discussions (including online forums, instant messaging, social networks etc.) are prohibited. Using internet beyond ILIAS is **only** authorised to troubleshoot R code errors.

Exercise 1 (20/30 pts) – Brightening star

For this exercise, you will need to load the dataset `brightening_star.csv` available in ILIAS:

```
> data <- read.csv(file="brightening_star.csv", header=TRUE, sep=";")
```

Description

The flux of a star has been found to increase abruptly recently. No less than 26 observatories have performed measurements of this star at the same, 9 epochs (dates that have been standardised in the file). All these measurements have been compiled in the file `brightening_star.csv` mentioned above.
One issue is that each observatory, run by different teams and using different data analysis methods are affected by systematics effects and the measured values differ by a lot. Using a multi-level model will thus be useful to make accurate inference.
The science question is: what is the dependence of star flux on time?

Questions

1. Which would be the obvious cluster to use?
2. Design two multi-level models to address the question above, one with varying intercepts alone and one with both varying intercepts and varying slopes.
3. Run both models, present the results in the form that you find the most relevant and interpret the estimates. What is the largest source (intercept/slope?) of variation in the outcome?
4. Build a third model that includes the correlation between the varying intercepts and slopes. Interpret the value of the correlation and discuss the implication on the predictions that you could make, if you were to receive a sample of new observations for that star.
5. Using the posterior mean values of the parameters, simulate flux data from 5 new observatories. Simulating both the varying slopes and intercepts will be necessary.

Exercise 2 (10/30 pts) – Dysfunctional cosmic ray detectors

Start by loading the dataset `detectors.csv` available in ILIAS :

```
> data <- read.csv(file="detectors.csv", header=TRUE, sep=";")
```

Description

The dataset is about a campaign of 250 distinct experiments to detect cosmic rays (CR) around the globe. Each detector has different properties (e.g. old/new generation, number of arrays...). One issue is that we do not know whether each experiment was actually functioning when it was powered up. For each experiment, the dataset includes the number of CR that have been detected, whether the detector used was equipped with an amplifier to increase sensitivity, whether it used a next-generation detector, how many arrays each detector had and how many of these arrays were operating in parallel. Finally the timespan during which the experiment was powered-up is indicated in the final column.
The scientific question is how many cosmic rays are detected per hour per detector array, when the experiment is actually functioning.

Questions

1. Think about the data-generating process. Write down one possible likelihood to answer the science question.
2. Design a model to predict the number of detected CRs per detector array per second as a function of any variables that are found to be relevant, you can try several and use model comparison to identify those that are worth including.
3. Run the model and discuss your findings.