02424 Assignment 3

This is the third of three mandatory assignments for the course 02424. It must handed in using the Campusnet (time and date is given at Campusnet). The submissions most contain one collected attached file in Portable Document Format (PDF), other document formats will not be accepted.

The report of each assignment must be prepared in groups of 3, and the final grading wil be based on the reports and the (individual) oral exam.

When writing the report please explain carefully what you did in each step, back up your statements with quantitative measures when possible, explicitly write down all models used in mathematical notation, and last keep it short and concise.

Part 1: Strength of ready mixed concrete

The strength of ready mixed concrete depends on the cement which is used in the production of concrete at the production plants. The strength is increasing from the time of the production until it reaches a steady state value after a rather long period of time, maybe six months. The strength of the concrete is measured after 28 days for control and validation.

The cement is delivered in batches, and it is observed that the strength varies from batch to batch. Given the same batch of cement the strength is assumed to be Gaussian distributed.

In order to obtain a more efficient control for the strength of concrete an early value of the strength is measured after one week. The purpose of this exercise is to establish a model for the joint variation of the 7-days and 28-days strength such that an initial validation can be performed already one week after the production.

In the file `concrete.csv` both early (7-days) and 28-days observations of concrete strength are listed. Also the air temperature at the individual production day is shown. The batch number for the cement used in the production is shown and the date of the production is also shown.

Problem A

1. Plot the observations of the 7-days and 28-days strength as a function of time.

2. Estimate the mean concrete strength within batches.

3. Formulate a mixed effect model for the 28-days strength.

4. Test for dependency on the air temperature.

Problem B

Let us now introduce $Y = (Y_7, Y_{28})^T$, ie. a vector containing both the early and the 28-day strength of concrete.

5. Formulate a multivariate mixed effect model for the simultaneous variation of strength.
6. Estimate the model parameters.
7. Estimate the correlation between the levels of early and 28-day concrete strength, including a confidence interval.

Part 2: Clothing insulation count data

In this part you should analyse the data-set `dat_count3.csv` the data set is the same as you analyzed in Assignment 2. Table 1 give an overview of variables to be used for modelling.

Table 1: List of included variables in the data-set for part 2, all variable are within day and subject average.

Variable	Type	Description
`clo`	Continuous	Number of changes
`tOut`	Continuous	Outdoor temperature
`tInOp`	Continuous	Indoor operating temperature
`sex`	Factor	Sex of the subject
`subjId`	Factor	Identifier for subject
`time`	Continuous	Total measurement time
`day`	Factor	Day (within the subject)
`nobs`	Integer	Measurement number (within the day)

Problem A

1. Present the data with focus on difference between subject.
2. Develop a generalized linear mixed effect models, based on the Binomial and Poisson distribution, using `glmmTMB`. Compare the results.

Problem B

With the structure above, fit a Poisson-Gamma model (i.e. random effects are Gamma distributed), i.e. we consider the following model

$$Y_{ij}|U_i \sim Pois(\mu_{ij}U_i) \tag{1}$$
$$U_i \sim G(\alpha, \beta) \tag{2}$$
$$\mu_{ij} = e^{x_{ij}^T\theta} \tag{3}$$

using the Laplace approximation, note also the the model should be implemented such that $E[U_i] = 1$. Consider the following

1. Implement the model directly in R using the Laplace approximation or in TMB.
2. Check the accuracy by importance sampling.
3. Compare the result with the previous question.

Note that: If you use another package than TMB you should make a precise reference that document that it does as requested.

Problem C

The model in Problem B allows for an explicit calculation of the integral in the hierarchical likelihood. You should:

1. Consider the model with only sex as explanatory variable (you may also initially consider a model with only an intercept)
2. Show that the marginal distribution of $Y_i = [Y_{i,1}, ..., Y_{i,n_i}]$ (observations for one subject) can be written in terms of the negative binomial density (Hint: consult the proof of Theorem 6.1 and the formulation in Theorem 6.3 and Example 6.6, note that the there is a misprint in the example ($U = \log(V)$ should be $U = \exp(V)$)).
3. Explicitly write down the parameters for the negative binomial above.
4. Using the above find the explicit formulation of the marginal likelihood
5. Implement the above formulation in R and estimate the parameters.
6. Find the conditional mean and variance of the random effects (Hint: see (the proof of) Theorem 6.2), and compare with the random effects in the previous questions
7. If you use a R-package to solve this you should make precise references documenting that it does as requested

Problem D

1. Compare parameter estimates and radom effect for the implemented models.
2. Write a small conclusion of your findings in Problem 2.