MATH452/552 Generalised Linear Models

Group project

J. Park

20/10/2020

Readmission to hospital of newborn babies

The Langley study (Langley et al, 2002) investigated the impact of community neonatal services (CNS) on high risk neonatal survivors within the first year of life. CNS care was provided through:

1. Home visits offering nursing care and specialist advice for a minimum of five days a week for the first month after discharge from the neonatal unit (NNU),
2. Ad hoc advice to families until the infant was at least 12 months old and
3. A named nurse providing the link between primary and secondary health care services.

The aim of the Langley study was to see if units with a CNS could induce a more satisfactory service in the mother and baby than units without a CNS. Service was measured by a number of responses such as (a) the length of stay in the NNU from birth to the initial discharge and (b) the number of readmissions to the NNU during the first year of life, among others.

Data collection

Data were collected from eighteen NNUs, with surveys sent to every mother who had visited within the previous three years. In total there were complete replies from the parents of 1488 babies. Apart from information on CNS, data was collected on a number of other covariates.

The dataset, `neomod.dat`, contains the following variables:

Variable names	Description
cns	Whether CNS was provided; 0=no, 1=yes
size	NNU size; 0=small, 1=large
gest	Gestation period; 1=Less than 26 weeks, 2=26-29 weeks, 3=30-32 weeks, 4=33-36 weeks, 5=More than 36 weeks
bwt	Birth weight in kg.
emp.f	Whether father/partner employed; 0=no, 1=yes
emp.m	Whether mother employed; 0=no, 1=yes

Variable names	Description
edu	Age mother left full time education; 1=before 16, 2=16-17 yrs, 3=18-20 yrs, 4=after age 20
re.ad	Whether re-admitted to an NNU within a year; 0=no, 1=yes
los	Total length of stay in hospital in log(days) during the initial admission
sex	Sex of the baby, 0=female, 1=male
accom	Whether owner occupier, 0=no, 1=yes

Aim of the project

Your dataset will be a random sample of the original data.

The aim of the project is to build an appropriate statistical model in order to determine which, if any, of the covariates are associated with re-admission to hospital within a year.

Guidelines for analysis

- You should treat **re.ad** as the response variable in your model.
- You should ensure that ordinal or categorical covariates are coded as factors.
- You should decide upon a systematic strategy for model selection including consideration of interaction terms.
- You should note that **los** is itself a post-birth response variable, as such you should consider the effect of including or excluding it from the model.
- You should perform exploratory analysis and model diagnostics to ensure the model is appropriate. In particular, care should be taken to ensure the correct functional form is chosen for continuous covariates.

Guidelines for the report

- Your full report should be no more than **six** A4 pages in length, in 11pt font using standard margins and line spacing and including references and appropriate tabular and graphical material. You may use up to two pages of an appendix exclusive of page limit to include additional material, which may cover some detailed derivations or computational details. However, you may only use it to add values and to satisfy the curious readers, not to fill the pages!

- You should submit your **R** code with annotated comments in a separate text file (a simple text file, R file or R markdown file are all acceptable). Your analysis should be reproducible!

- The report should be submitted as a pdf file. You are welcome to use **R markdown** to write your report, but this is not compulsory.

- You should structure your report to be in the style of a scientific paper. It should have an abstract followed by sections corresponding to an introduction, methods section, results section (including exploratory analysis) and a conclusion/discussion. In doing so, you may use more informative section headings and structure the material in the way you feel appropriate.

- The methods section should contain the key aspects of the model(s) used and the strategy for model selection and model assessment. You do not need to copy large chunks of theory from the lecture notes.

- Marks will be awarded on the basis of the quality and correctness of the statistical analyses performed; the interpretation of results; presentation and choice of tables and figures; quality of writing; consistency of notation; use of literature.

Submission instructions

You should submit an electronic copy of your work to Moodle by **10am Monday in Week 7**.

On the first page (excluded from the page limit), you must provide the following information:

- your group number
- student ID numbers of each member and
- the module number

In addition, you need to include a statement of contribution of each member to share the mark.

An opportunity to moderate the mark for non-equal contributions will be given as an oral presentation by each member.

Below is an example of your declaration.

- statement of contribution:

  "NAME 1 . . . and NAME k have contributed equally to the whole project and agree to share the mark equally."

  "NAME 1 have contributed to . . . . and NAME k have contributed to . . . and agree that the mark should be distributed disproportionately as (1:A%,2:B%, . . .)."

  "NAME 1 have contributed to . . . . and NAME k have contributed to . . . and agree that the mark should be distributed disproportionately. Therefore, the members wish to attend an oral meeting."

References

Langley, D., Hollis, S., Friede, T., MacGregor, D. and Gatrell, A., 2002. Impact of community neonatal services: a multicentre survey. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, **87**: F204-F208.