Assignment Statistics V (2017-2018)

May 3, 2018

Introduction

Self-reported life satisfaction is considered indicator of a person's general well-being and it may be affected by many influences. Because of its importance to individuals and the society, life satisfaction is heavily studied by academics and policy makers.

The data for this assignment come from the German socio-economic panel (SOEP). The goal of this assignment is to find out which factors are associated with life satisfaction at a later age. An overview of all variables in the data set (called `soep_statV.csv`) can be found in Table 1.

Life satisfaction of a sample of 1236 Germans was tracked during ages 55-60 (not all persons have been measured six times). Life satisfaction was measured on a scale from 0 (totally unhappy) to 10 (totally happy). The candidate factors that may explain the variation in life satisfaction can be found in Table 1.

Table 1. Variables in the `soep_statV.csv` dataset.

Variable	***Name***	***Values***
Person identifier	id	1236 unique identifying integer values
Life satisfaction	lifeSat	0 (totally unhappy) – 10 (totally happy)
Age (in years)	age	55 – 60
Employment status at 55	emplStat55	1=full time employed,2=part time empl.,3=unemployed
Satisfaction with health	healthSat	0 (low) – 10 (high)
Living together	livtog	1=married/living with partner,0=single,divorced,widow

Research questions

One obvious question is to identify the factors that are associated with life satisfaction across time for people in late middle age. However, this question is very broad, so here a few more specific questions you may want to look at:

- How much individual differences are there in life satisfaction trajectories from 55 to 60?
- To which extent can these individual differences be explained by contextual factors such employment status, having a partner, and satisfaction with health?

It may require a bit of creativity to come up with a suitable model. Generally it helps to start with a simple model and then gradually make it more (but not overly) complex. Also try to make insightful graphs that go together with the model.

There is a lot of room for personal input and creativity in this assignment. This freedom implies that although all groups use the same data, different results among groups are perfectly possible, or even expected.

You are by no means restricted to techniques covered in the lectures but using the techniques covered in an insightful way will earn you good grades. Of course, given the topic of the course, we require you to use Bayesian methods to answer the research questions.

Practical issues

- Form a group of (preferably) 3 students (because the total number of students is not divisible by three, we allow for one 2 person group). We have created a Google spreadsheet (click here) where you can write your name to form the groups (do this as soon as possible). (Here is the full url: https://docs.google.com/spreadsheets/d/1jZUZ5SubRpD3AuyPgN3rcQgDlzrlcT_vq6ou2h_79qU/edit?usp=sharing).
- This assignment and the data set will be posted on Toledo.
- The submission deadline is Monday 11 June, 2018, 4pm (first day of the exams), but submitting earlier is possible.
- To submit, send your paper and R code (see below) to Aniek Sies (aniek.sies@kuleuven.be), Evalyne Thauvoye (evalyne.thauvoye@kuleuven.be) and Richard Artner (richard.artner@kuleuven.be).
- We will have two assignment Q&A moments during the practical sessions of May 16 and May 23.

Content of the paper

- Title page (max 1 page): Include all names of the team members, and a fancy title, if you have one.
- Make sure that the title page can be removed easily without affecting the content so that we can grade your paper anonymously.
- A paper (max 1234 words, not including figure captions, table headings and reference list): The structure of the paper is free, really, but at least the following aspects should be covered:
    - the research question(s)
    - a brief description and motivation of the Bayesian method(s) (likelihood and priors) used
    - exploratory analyses
    - an assessment of the convergence
    - a sensitivity analysis
    - figures and/or table(s) with parameter estimates of the model(s)
    - insightful and clear graphical depiction(s) of the results
    - conclusion(s) about the research question(s)
    - limitation(s) of the study and of the conclusions
    - a reference list (if you have used additional literature)
- The paper should be written in English.
- The paper can be written in Rmarkdown, but that is not required (so you can use Word or any other text editor).
- If your analysis is exploratory (i.e., several research questions are formed while investigating and analyzing the data), then state this clearly.
- Send along with your paper also a file with the R and JAGS code that you used to analyze your data; here is some additional information on this file:
    - we will run and check your R code!
    - there is only a single R file with the of all your analyses
    - if there is, for some reason, more than one R file, clearly tell us which one we need to run to reproduce your results comment your code!
    - only send us those files we need

- if possible, make sure of a relative path when you need to read or save files; e.g., save(file="./output/output.RData", output)
- if relative paths do not work, then make sure we only need to change the working directory to get the code running
- if we do need to do anything else (install packages, download functions, etc.) do tell us explicitly (you can of course assume we have JAGS, rjags and runjags)
- when additional packages need to be installed, you can use the following code chunk (with the packages Cairo and coda as example packages; replace their names with your packages):

```
list.of.packages <- c("decision","coda")
new.packages <- list.of.packages[!(list.of.packages %in% installed.packages()[,"Package"])]
if(length(new.packages)){install.packages(new.packages,repos = "http://cran.us.r-project.org")}
lapply(list.of.packages, require, character.only = TRUE)
```

Good luck!