Project 2: Comparison of hospital length of stay between insurers for patients with pediatric asthma

Project description

In today's healthcare environment, health insurance companies are increasingly pressuring hospitals to provide high quality health services at the lowest possible cost. The vast majority of all healthcare costs are for hospitalization. During the past decade, impatient costs of patients in hospitals have been reduced in two primary ways. First, the less severe cases are now being treated in the doctor's office or in hospital emergency rooms rather than being admitted to the hospital. Second, for cases admitted to the hospital, the lengths of hospital stays have been considerably shortened.

It is believed that some insurers have been more successful than others at minimizing hospital lengths of stay. To test this, a sample of hospital medical records was drawn for each of several illnesses from metropolitan hospitals operating in one state. The data for this project consists of information taken from the medical records of patients with asthma between the ages of 2 and 18 years old. You will use these data to answer the following three questions:

1. Are there differences in lengths of stay between insurers A and B?

2. Are there any differences in hospital characteristics between the two insurers? Specifically, are there differences between the two insurers for the teaching category of the hospitals, the proportion of public vs. private hospitals, and the size of the hospital as measured by the bedsize variable?

3. Do any observed differences in length of stay between insurers A and B persist after accounting for the differences in hospital characteristics that you considered in the question above?

Background information

The sample of medical records was drawn in two stages. At the first stage, 29 metropolitan hospitals were sampled with probabilities proportional to an estimate of the number of asthma admissions during a single year. At the second stage, 393 asthma cases insured by Insurer A were randomly selected and 396 asthma cases insured by Insurer B were randomly selected from the 29 hospitals.

Information was abstracted from each patient's medical record. Aside from the main variables of interest, insurer and length of stay, the additional information falls into four categories:

- *Patient severity variables* describe the severity of the patient's condition at admission to the hospital. Included in this category are comorbidities, additional conditions that may exacerbate the condition that is the major reason for admission. The length of stay is expected to be longer for more severely ill patients and for patients with comorbidities.
- *Demographic variables* describe the patient's age, sex, and race. The effect of demographic factors on length of stay for patients with pediatric asthma is of interest.
- *Hospital variables* describe the characteristics of the hospital in which the patient was treated. It is expected that length of stay will be lower for hospitals with more services and with a higher staff-to-bed ratio. The effects of size of the hospital, measured by the number of beds in the hospital (bedsize), and the categorization of that hospital as a teaching hospital or not (teachcat) are of interest.
- *Treatment variables* describe aspects of the patient's treatment during hospitalization. Treatment variables are not fully known until the hospital stay is complete. It is expected that length of stay will be longer for patients who have treatment complications and process of care failures (the hospital fails to adhere to standards of care given the patient's condition). Finally, patients with more diagnostic tests are probably more severely ill. Consequently, length of stay is expected to increase with the number of tests performed.

Getting started in R

You can use the following R command to read the dataset into R:

```
project2.dat = read.table(file="project2.txt", col.names=c("los","hospital","patient","insurer","sumcomrb", "cmpl11","cmpl12", "sevmod", "sevsev","histf01", "histf02","histf03","age","female","race","numserv","bedsize","owner","teachcat","anycomp", "highpoc","medpoc","lowpoc","diagtsts","ftetobed","pctins1"))
```

The data set contains the following variables:

CODE	DESCRIPTION
los	Length of stay (days)
hospital	Numerical code for the hospital
patient	Patient identification number
insurer	0=Insurer A, 1=Insurer B
sumcomrb	Number of comorbidities
cmpl11	Bronchitis present? (1=yes, 0=no)
cmpl12	Pneumonia present? (1=yes, 0=no)
sevmod	Asthma severity is moderate or higher (1=yes, 0=no)
sevsev	Asthma severity is severe (1=yes, 0=no)
histf01	History of respiratory failure (1=yes, 0=no)
histf02	Oral steroid dependent (1=yes, 0=no)
histf03	Two or more previous hospitalizations for asthma (1=yes, 0=no)
age	Patient age in years
female	Patient is female? (1=yes, 0=no)
race	Patient race (1=white, 2=Hispanic, 3=Black, 4=Asian/Pacific Islander, 5=unknown)
numserv	Number of hospital services
bedsize	Hospital bedsize category (1=1-99, 2=100-249, 3=250-400, 4=401-650)
owner	Ownership category (1=public, 2=private)
teachcat	Degree of teaching category (0=none, 1=minor, 2=major)

anycomp	Any treatment complications (1=yes, 0=no)
highpoc	Number of high level process of care failures
medpoc	Number of medium level process of care failures
lowpoc	Number of low level process of care failures
diagtsts	Number of diagnostic tests ordered or performed
ftetobed	Full time staff per bed
pctins1	Percent of annual patients insured by insurer B

You are acting as an agent asked to prepare a written and oral report to a group of hospital CEOs in the state that are interested in the role that insurers play in the length of hospital stays.

Your written report must include:

- A title page
- A table of contents
- An introduction to the problem being considered, using appropriate scientific language where reasonable
- Clear statement of the specific questions of interest
- Presentation of relevant graphical summaries
- Presentation of relevant numerical summaries
- Descriptions of methods used in the analysis, along with addressing diagnostics to check for validity of conditions (e.g. QQ-plots, residual plots, etc.)
- Use of at least one statistical test, with proper statement of the test and a conclusion based on a p-value
- Use of at least one estimator, together with an assessment of uncertainty in the estimator (e.g. a confidence interval)
- Conclusions and recommendations written from the point of view of the agent and appropriate for the target audience

Your recorded presentation should be 8–12 minutes long (don’t take this literally), and should summarize your written report. Essentially, your recorded presentation should make clear what your conclusions and recommendations are, and should support them with a summary of the analyses and discussion of your methods.