BZAN 540 Project 2020

Due on April 23, 2:00 pm

This project is intended to be completed in groups of **four-five** students (of your choosing). A single member of each group should send me an e-mail with the names of the respective group members by Tuesday, April 21^{st}. Each group will present on the 23^{rd} during class time (8 minute presentations). Each group can delegate a group member (or members) to present.

1 Problem Overview

Hospital-acquired conditions (HACs) represent a significant safety concern for patients. It is estimated that on any given day, about 3% of hospitalized patients in the US experience one or more HAC and 11% of patients with HACs (representing about 72,000 per year) die during their hospitalization (Magill et al., 2018). To help reduce the prevalence of HACs, the Hospital-Acquired Condition Reduction Program (HACRP) was established as part of the Affordable Care Act (ACA). According to the HACRP, the Centers for Medicare and Medicaid Services (CMS) do not provide payment for treatment of certain reasonably preventable conditions when those conditions are acquired after the patient has been admitted to the hospital. Thus, the program aims to shift the financial burden of treating HAC to the provider, as an incentive for the providers to improve patient safety and prevent HACs.

While CMS has instituted financial penalties that incentivize hospitals to reduce HACs, the relationship between HACs and some patient outcome measures such as length of stay (`los`), 30 day post-discharge readmission risk (`read`), costs (`totchg`), etc., is still not well understood. In this project, you are tasked to take up two of these problems to study.

1.1 Pick two of these problems to study

(a) What is the association between HACs and 30-day post-discharge readmission risk (or effect of HACs on 30-day post-discharge readmission risk).

(b) What is the association between HACs and length of stay during the index hospitalization (or effect of HACs on length of stay).

(c) What is the association between HACs and total charges during the index hospitalization (or effect of HACs on total charges on discharge from index hospitalization).

1.2 Overview of the data

The `FL_data.csv` data file contains information on 346,685 in-patient hosptializations (i.e. hospital admissions) for patients admitted for one of three conditions: Heart Failure (`HF`), Heart Attach (`AMI`), and Pneumonia (`PN`), over the time period 2010 - 2014, for patients in the state of Florida. That is, each row corresponds to a hospitalization (hospital admission). The data file has the following variables:

```
FL_data = readRDS("FL_data.rds")
colnames(FL_data)
```

```
##  [1] "age"                   "nchronic"
##  [3] "ndx"                   "npr"
##  [5] "totchg"                "los"
##  [7] "count_dxnopoa_ns"      "visit"
##  [9] "ahour"                 "totalhospital_noofbed"
## [11] "teachstatus"           "tcontrol"
## [13] "rural"                 "pay1"
## [15] "condition"             "drg"
## [17] "atype"                 "aweekend"
## [19] "dqtr"                  "dshospid"
## [21] "female"                "medincstq"
## [23] "year"                  "race"
## [25] "tran_out"              "zipinc_qrtl"
## [27] "cm_anemdef"            "cm_chf"
## [29] "cm_chrnlung"           "cm_coag"
## [31] "cm_depress"            "cm_dm"
## [33] "cm_dmcx"               "cm_htn_c"
## [35] "cm_hypothy"            "cm_lytes"
## [37] "cm_neuro"              "cm_obese"
## [39] "cm_perivasc"           "cm_renlfail"
## [41] "read"                  "HACs"
```

The variable `HACs` is an indicator variable of presence/absence of hosptical-acquires condition(s) during an index hospitalization. The `count_dxnopoa_ns` variable indicates the number of hospital-acquired conditions (diagnostics) from which `HACs` is derived by: `ifelse(count_dxnopoa_ns==0,0,1)`. The variable `count_dxnopoa_ns` is computed by looking at the difference in the number of diagnostics at admission and discharge.

The variable `read` indicates wheather the associated index hospitalization (current hospitalization) was followed by a readmission within 30 days post-discharge (all cause readmissions excluding scheduled readmissions).

The variables `los` and `totchg` represent the length of stay and total charges for the corresponding index hospitalization, respectively.

The variables represented as `" cm_"` are indicators for presence/absence of commorbidities. For instance, `cm_chf` is an indicator variable for presence of `chronic heart failure`.

The variable `condition` is constructed from the `drg` (Diagnosis Related Group) variable . More speficically, `drg%in%c(291:293)` correspond to `condition='HF'`, `drg%in%c(193:195)` correspond to `condition='PN'`, `drg%in%c(280:282)` correspond to `condition='AMI'`.

More information about the variables can be found at https://www.hcup-us.ahrq.gov/db/state/siddist/siddistvarnote2014.jsp.

2 Project Report Structure

Your written report must contain an **executive report** (brief summary of findings), not to exceed three paragraphs, on a separate page (page 1). The body of the write up should, in no more than 12 pages, cover:

(1) The two problems you want to study using your analysis, i.e. problem statements.
(2) Brief and relevant descriptive analytics results - use visualizations when possible, and numerical summaries.
(3) Details of any data pre-processing steps that you undertake: transformations, handling of missing data, if any, etc.

(4) Statistical model(s) that you use, there is no need to describe the methods in detail in here.
(5) Results form your statistical model(s). This should be restricted to only relevant results. Brief interpretation of any interesting results should be provided. Focus on the highlights of your analysis results in relation to the problem(s) stated in (1). Comment on any effects of other relevant predictors.

You can include an appendix at the end of your write-up containing the following (this is beyond the body of the write up):

(7) Any analysis done to assess assumptions, address sensitivity of results, and any references utilized.

Presentation during class time on the 23rd: Each team will have 8 minutes to present their interesting findings. In your presentation use as many graphics as necessary that are clear. Your presentation should focus on:

(a) Problem statement(s)
(b) Any interesting descriptive analytics results, and any data pre-processing steps
(c) Brief mention of statistical methods utilized
(d) Summary of (interesting) results

Preparation of report - report should be prepared in R Markdown. Set `echo=FALSE` in your global `knitr` options. You should `knit` to a pdf file for cleaner and more condensed report. You can install the `tinytex` package in order to be able to `knit` to a pdf file (`install.packages('tinytex')`).

What to upload to Canvas: The R markdown file, The knitted pdf file, The slides you use for your presentation.

Magill et al. (2018). Changes in Prevalence of Health Care–Associated Infections in U.S. Hospitals. *The new england journal of medicine*, 379:1732-44. DOI:10.1056/NEJMoa1801550.