BIOS707 | Problem Set 01

Kuei-Yueh (Clint) Ko

Set environment

This report aim to recreate the table one and table two in the study *"Operation Timing and 30-Dayrtality After Elective General Surgery"*. The packages required in this report include **dplyr**, **tableone**, **labelled**.

```
### load required packages
library(dplyr)
library(tableone)
library(labelled)
```

Import data

The study data are available in a number of formats on the website. Load the RData object directly from the website.

```
### load R dataset
load(url("https://www.causeweb.org/tshs/datasets/surgery_timing.Rdata"))
head(stata_data)
```

```
##   ahrq_ccs  age gender             race asa_status   bmi baseline_cancer
## 1  <Other> 67.8      M        Caucasian       I-II 28.04              No
## 2  <Other> 39.5      F        Caucasian       I-II 37.85              No
## 3  <Other> 56.5      F        Caucasian       I-II 19.56              No
## 4  <Other> 71.0      M        Caucasian        III 32.22              No
## 5  <Other> 56.3      M African American       I-II 24.32             Yes
## 6  <Other> 57.7      F        Caucasian       I-II 40.30              No
##   baseline_cvd baseline_dementia baseline_diabetes baseline_digestive
## 1          Yes                No                No               Yes
## 2          Yes                No                No                No
## 3           No                No                No                No
## 4          Yes                No                No                No
## 5           No                No                No                No
## 6          Yes                No                No                No
##   baseline_osteoart baseline_psych baseline_pulmonary baseline_charlson
## 1                No             No                 No                 0
## 2                No             No                 No                 0
## 3                No             No                 No                 0
## 4                No             No                 No                 0
## 5                No             No                 No                 0
## 6                No            Yes                 No                 0
##   mortality_rsi complication_rsi ccsmort30rate ccscomplicationrate  hour
## 1         -0.63            -0.26     0.0042508          0.07226355  9.03
## 2         -0.63            -0.26     0.0042508          0.07226355 18.48
## 3         -0.49             0.00     0.0042508          0.07226355  7.88
## 4         -1.38            -1.15     0.0042508          0.07226355  8.80
## 5          0.00             0.00     0.0042508          0.07226355 12.20
## 6         -0.77            -0.84     0.0042508          0.07226355  7.67
```

```
##   dow month     moonphase mort30 complication
## 1 Mon   Nov     Full Moon     No           No
## 2 Wed   Sep      New Moon     No           No
## 3 Fri   Aug     Full Moon     No           No
## 4 Wed   Jun  Last Quarter     No           No
## 5 Thu   Aug  Last Quarter     No           No
## 6 Thu   Dec First Quarter     No           No
```

Table 01

Recreate table 1 of the study. The table 1 is **the summary of baseline risk factors for 32,001 general surgical patients**.

```
## intialization
df <- stata_data

## Make categorical variables factors
varsToFactor <- c("gender", "race", "asa_status",
          "baseline_cancer",     # Cancer
          "baseline_cvd",        # Cardiovascular/cerebrovascular disease
          "baseline_dementia",   # Dementia
          "baseline_diabetes",   # Diabetes
          "baseline_digestive", # Digestive disease
          "baseline_osteoart",   # Osteoarthritis
          "baseline_psych",      # Psychiatric disorder
          "baseline_pulmonary") # Pulmonary disease


df[varsToFactor] <- lapply(df[varsToFactor], factor)

## set levels
df$asa_status <- factor(df$asa_status, levels = c("I-II", "III", "IV-VI", ""))

## set label of variables
var_label(df$age)                <- "Age"
var_label(df$gender)             <- "Gender"
var_label(df$race)               <- "Race"
var_label(df$asa_status)         <- "ASA physical status"
var_label(df$bmi)                <- "Body mass index (kg/m^2)"
var_label(df$baseline_cancer)    <-  "Cancer"
var_label(df$baseline_cvd)       <- "Cardiovascular/cerebrovascular disease"
var_label(df$baseline_dementia)  <- "Dementia"
var_label(df$baseline_diabetes)  <- "Diabetes"
var_label(df$baseline_digestive) <- "Digestive disease"
var_label(df$baseline_osteoart)  <- "Osteoarthritis"
var_label(df$baseline_psych)     <- "Psychiatric disorder"
var_label(df$baseline_pulmonary) <- "Pulmonary disease"
var_label(df$baseline_charlson)  <- "Charison Comorbidity Index"
var_label(df$mortality_rsi)      <- "Risk Stratification Index (30-day mortality)"
var_label(df$complication_rsi)   <- "Risk Stratification Index (in-hospital complications)"

## Create a variable list
vars <- c("age",                # Age
```

```
          "gender",              # Gender
          "race",                # Race
          "asa_status",          # ASA physical status
          "bmi",                 # Body mass index (kg/m^2)
          "baseline_cancer",     # Cancer
          "baseline_cvd",        # Cardiovascular/cerebrovascular disease
          "baseline_dementia",   # Dementia
          "baseline_diabetes",   # Diabetes
          "baseline_digestive", # Digestive disease
          "baseline_osteoart",   # Osteoarthritis
          "baseline_psych",      # Psychiatric disorder
          "baseline_pulmonary", # Pulmonary disease
          "baseline_charlson",   # Charison Comorbidity Index
          "mortality_rsi",       # Risk Stratification Index (30-day mortality)
          "complication_rsi")    # Risk Stratification Index (in-hospital complications)

## nonnormal continuous variables
vars_nonnorm <- c("bmi", "baseline_charlson", "mortality_rsi", "complication_rsi")

## Create Table 1 stratified by trt
tableOne <- CreateTableOne(data = df, vars = vars)

## Format
print(tableOne,
      contDigits = 1,
      nonnormal  = vars_nonnorm,
      varLabels  = TRUE,
      explain    = FALSE,
      dropEqual  = TRUE)
```

```
##
##                                                    Overall
##   n                                                32001
##   Age                                               57.7 (15.0)
##   Gender
##                                                        3 ( 0.0)
##      F                                            17230 (53.8)
##      M                                            14768 (46.1)
##   Race
##                                                      480 ( 1.5)
##      African American                                3790 (11.8)
##      Caucasian                                      26488 (82.8)
##      Other                                           1243 ( 3.9)
##   ASA physical status
##      I-II                                           17261 (53.9)
##      III                                            13677 (42.7)
##      IV-VI                                           1055 ( 3.3)
##                                                         8 ( 0.0)
##   Body mass index (kg/m^2)                          28.2 [24.6, 32.8]
##   Cancer                                         10958 (34.2)
##   Cardiovascular/cerebrovascular disease         16176 (50.5)
##   Dementia                                         242 ( 0.8)
##   Diabetes                                        4166 (13.0)
##   Digestive disease                               7037 (22.0)
```

```
##   Osteoarthritis                                          5719 (17.9)
##   Psychiatric disorder                                    2910 ( 9.1)
##   Pulmonary disease                                       3493 (10.9)
##   Charison Comorbidity Index                               0.0 [0.0, 2.0]
##   Risk Stratification Index (30-day mortality)            -0.3 [-1.2, 0.0]
##   Risk Stratification Index (in-hospital complications)   -0.3 [-0.8, 0.0]
```

Table 02

Recreate table two of the study. The table 2 is **frequencies of case start times by hour of day, day of week, month of Year, and moon phase**.

```
### initialization
df <- stata_data

### preprocess and set factor levels
df$hour <- floor(df$hour) %>%
    as.character %>%
    sapply(., function(x){return(ifelse(x == "19", "18", x))}) %>%
    paste0(., ":00") %>%
    factor(., levels = paste0(6:18, ":00"))

# set week abbreviation and full name
week_abb = c("Mon", "Tue", "Wed", "Thu", "Fri")
week_name = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")

# Day of week
df$dow <- df$dow %>%
    match(., week_abb) %>%          # get the index of each element by matching abbr.
    `[`(week_name, .) %>%           # change the element into full name
    factor(., levels = week_name)   # specify factor level

# Month of year
df$month <- df$month %>%
    match(., month.abb) %>%         # get the index of each month by matching abbr.
    `[`(month.name, .) %>%          # change the lement into full name
    factor(., levels = month.name)  # specify factor level

# moon phase
df$moonphase <- factor(
    df$moonphase,                   # set as a factor
    levels = c(                     # specify factor level
        "New Moon",
        "First Quarter",
        "Full Moon",
        "Last Quarter"))

### set label of variables
var_label(df$hour)      <- "Factor operation hour"
var_label(df$dow)       <- "Day of week"
var_label(df$month)     <- "Factor month"
var_label(df$moonphase) <- "Phase of moon"
```

```r
### Create a variable list
vars <- c("hour", "dow", "month", "moonphase") #, "mort30", "complication")

### Create table two
tableTwo <- CreateTableOne(data = df, vars = vars)

## Format
print(tableTwo,
      contDigits = 1,
      varLabels  = TRUE,
      explain    = FALSE,
      dropEqual  = TRUE)
```

```
##
##                          Overall
##   n                      32001
##   Factor operation hour
##      6:00                  562 ( 1.8)
##      7:00                10631 (33.2)
##      8:00                 3807 (11.9)
##      9:00                 1664 ( 5.2)
##      10:00                2501 ( 7.8)
##      11:00                2855 ( 8.9)
##      12:00                2763 ( 8.6)
##      13:00                2623 ( 8.2)
##      14:00                2063 ( 6.4)
##      15:00                1267 ( 4.0)
##      16:00                 745 ( 2.3)
##      17:00                 356 ( 1.1)
##      18:00                 164 ( 0.5)
##   Day of week
##      Monday               7005 (21.9)
##      Tuesday              7008 (21.9)
##      Wednesday            6266 (19.6)
##      Thursday             5635 (17.6)
##      Friday               6087 (19.0)
##   Factor month
##      January              2670 ( 8.3)
##      February             2506 ( 7.8)
##      March                2697 ( 8.4)
##      April                2698 ( 8.4)
##      May                  2654 ( 8.3)
##      June                 2994 ( 9.4)
##      July                 2325 ( 7.3)
##      August               3177 ( 9.9)
##      September            3208 (10.0)
##      October              2689 ( 8.4)
##      November             2544 ( 7.9)
##      December             1839 ( 5.7)
##   Phase of moon
##      New Moon             7708 (24.1)
##      First Quarter        8100 (25.3)
##      Full Moon            8051 (25.2)
##      Last Quarter         8142 (25.4)
```

What does adjusted probability mean in figure 3 and figure 4

Question
Figures 3 and 4 report adjusted probability estimates. What were these probabilities adjusted for? Is this study reproducible based on the published article

My Answer
Based on the statistical analysis in the method part, the thirty-day mortality was adjusted based on Risk Stratification Index (RSI). However, the authors did not mention clearly how and why the probability was adjusted. Instead, the authors only cited the reference of the adjusted method. Since the description is not enough for reader to replicate easily, I believe the calculation of adjusted probability estimates in the study is not reproducible.