STAT 796: Homework 1

Joshua Keller

Due Sunday, January 27 at 11:59pm on Canvas.

The file `ICU_raw.txt` on Canvas contains the data discussed in lecture on patients in the intensive care unit, but in a different format. This assignment asks you to perform various coding tasks related to modifying ("wrangling") the dataset and creating various summaries of the data.

Note: Include your R code for the entire assignment as an appendix at the end of your responses.

1. Read the file `ICU_raw.txt` into R using the `read_table2()` function from the `readr` package. (Nothing needs to be turned in for this question)
2. Documentation of the original coding for this dataset is provided in the file `ICU_Data_Code_Sheet` on Canvas. Rename and recode the variables in the dataset according to the following table.

Original Name	New Name	Re-coding Comments
STA	status	
AGE	age	
GENDER	female	
RACE	race	Factor with 1=White, 2=Black, 3=Other
SER	service	Factor with 0=medical, 1=surgical
CAN	cancer	
CRN	renal	
INF	infect	
CPR	cpr	
SYS	sbp	
HRA	hr	
PRE	prv_admit	
TYP	type	Factor with 0=elective, 1=emergency
FRA	fracture	
PO2	po2_g60	Define new variable as 1-PO2
PCO	pco2_g45	
BIC	bicarb_g18	Define new variable as 1 - BIC
CRE	creat_g2	
LOC	conscious	Factor with 0=no_coma_or_stupor, 1=stupor, 2=coma

Example code starter code is below. Note that since we are renaming and replacing *all* variables, we use `transmute()` instead of `mutate()`.

```
icu <- icu_raw %>%
  transmute(id=ID,
         died=STA,
         age=AGE,
         female=GENDER,
         race=factor(RACE,
                     levels=1:3,
                     labels=c("white", "black", "other")))
```

3. Compute the following summary statistics for the entire cohort:
   a. Average, Minimum, Maximum, and Standard Deviation of age (in years)
   b. Count of subjects with each service type (Medical/Surgical)

   c. Count of subjects with each procedure types (Elective/Emergency)
   d. Count of subjects of each race and of each sex
   e. Counts of subjects by consciousness at admission (Coma, Stupor, No Coma or Stupor)
4. Compute the following summary statistics for the cohort, grouped by vital status:
   a. Average, Minimum, Maximum, and Standard Deviation of age (in years)
   b. Count of service types (Medical/Surgical)
   c. Count of procedure types (Elective/Emergency)
   d. Race and sex counts
   e. Counts of consciousness at admission (Coma, Stupor, No Coma or Stupor)
5. Create the following graphical summaries of the data. Figures should include appropriate titles and labels.
   a. Scatterplot of systolic blood pressure (`sbp`) against age (`age`), using color and shape to identify observations by vital status at discharge.
   b. Boxplots of heart rate, grouped by consciousness at admission. [Use `geom_boxplot(aes(x=GROUPVAR, Y=PRIMARYVAR))`.]
   c. Histograms of age [Use `geom_histogram(aes(x=XVAR))`.]
   d. Density estimates of heart rate, grouped by consciousness at admission.