BIOS 522: Project 1

Falcons Group

2020-08-29

Contents

1 Standard Regression techniques questions

- Dr. Blum is interested in survival times of patients and would like to know the impact of treatment, age, and serum bilirubin as a categorical variable (<1.1, 1.1-3.3, and >3.3) on survival.
  - Use a linear model after excluding all censored observations
  - Use a linear model after treating censored times as death times
  - Use a logistic regression by defining a new outcomes as dead=1 and otherwise (survived or censored) as 0.
- For each of above models perform univariate and multivariate analyses(for the three covariates above).
- Interpret the estimates of coefficients of treatment, age, and serum bilirubin regardless of their significance.
- Now comment on the appropriateness of the data analyses Dr. Blum suggested. A critique.

2 Survival Regression techniques questions

- Now perform a parametric survival analysis (Weibull) and conduct the same univariate and multivariate analyses. Report same results for interpreting coefficients.
- Dr. Blum wants to know how to derive the i) estimates ii) standard errors of the coefficients that R outputs from the regression. What is the procedure? How can Dr. Blum recreate them herself? Attach this techinical section as an appendix to the report.

3 Analysis of the linear regression

3.1 Excluding censored observations

3.2 Treating censored times as death times

4 Analysis of the logistic regression

5 Critique of Dr. Blum's analyses

6 Parametric survival analysis

7 Technical Appendix: How to derive coefficient estimates and standard errors from parametric survival analyses

8   Code

```r
knitr::opts_chunk$set(
  echo = FALSE,            # don't show code
  warning = FALSE,         # don't show warnings
  message = FALSE,         # don't show messages (less serious warnings)
  cache = FALSE,           # set to TRUE to save results from last compilation
  fig.align = "center"     # center figures
)
library(data.table)
library(ggplot2)

set.seed(1)               # make random results reproducible
data <- fread("data.csv")

cols <- c("id", "time", "status", "drug", "age", "sex",
          "is_ascites", "is_hept", "is_spiders", "is_edema",
          "serum_bilirubin", "serum_chol", "albumin", "urine_copper",
          "alk_phosphatase", "sgot", "triglicerides", "platelets",
          "prothrombin", "histologic_stage")

names(data) <- cols

data <- data[, .(id, time, status, drug, age, serum_bilirubin)]
data[, serum_cat := cut(serum_bilirubin, c(0, 1.1, 3.3, max(serum_bilirubin)))]
# this R markdown chunk generates a code appendix
```