PM511B Project

Date Assigned: Mar 22 2019
Date Due: Apr 19 2019

The STATA dataset project_spring2019 provides data on persons who experienced a traumatic injury and were seen initially in the hospital emergency room, then admitted to the hospital. The dataset contains the following variables:

Variable	Description	Codes
sid	Patient ID	
died	In-hospital mortality	0 = survived 1 = died
Use these variables (age through gcs) to develop your prediction model		
age	Patient age (years)	
male	Gender	0 = female 1 = male
race	Race/ethnicity	1 = Asian 2 = African-American 3 = Non-Hispanic white 4 = Hispanic white
All variables below are initial measures obtained in the emergency room		
sbp	Systolic blood pressure (mm Hg) Healthy resting systolic BP is about 120	
rr	Respiratory rate (breaths per minute) Normal is about 12-20	
gcs	Glasgow Coma Scale: a neurologic measure of level of consciousness. Lower scores indicate lower level of consciousness. Sometimes used to estimate extent of brain injury. Severe injury: GCS<9 Moderate injury: GCS 9-12 Minor or no injury: GCS>12	Range 1-15
Use variable below to determine whether subject data is used in estimation (training) vs. validation		
val	Use in estimation vs. validation sample	0 = estimation sample 1 = validation sample
Use the measure below ONLY for Part 10		
asaps	A measure of physical fitness for surgery: 1: Healthy 2: Mild systemic disease 3: Severe systemic disease 4: Severe systemic disease that is a constant threat to life 5: Not expected to survive without operation	Range 1-5

General instructions:

1. This project is to be completed alone. Any evidence of the following will result in an F in the course and will be referred to student judicial affairs:
   Sharing of programs or text with other students
   Direct use of text from textbook, lectures, lab materials or other sources

2. The project should be turned in on Blackboard by the due date.

3. **Turn in two project files**:

   a. Responses to the questions below, which may include text, tables and figures.

      Title this file **"pm511bproject_text_lastname_firstname"**
      This document should be no longer than 20 pages, using Arial Font 11. Any documents longer than 20 pages will not be accepted.

   b. Your stata smcl file(s), organized by the question (e.g. 1a, 1b, etc.)

      Title this file **"pm511bproject_smcl_lastname_firstname"**
      Suggestion: When you have completed your responses and have an error-free working program, run the program, saving to the smcl file referenced above.

      Do NOT wait to the last minute to figure out how to save your program statements and results to this smcl file!

Assignment: Investigators want to use these data to develop a predictive model for in-hospital mortality among patients with trauma. **Use the subjects with val=0 (the estimation sample)** in these data to develop a predictive model for in-hospital mortality.

1. Provide a **publication-quality** descriptive table of the sample (with val=0) on the variables above. Report your summary statistics (means, frequencies, etc.) separately by mortality (i.e., your table will have 2 columns – died and survived).

   a. Write a **short** paragraph describing the sample and possible differences by mortality.

2. Using the listed variables above (age through gcs), develop a predictive model (among subjects with val=0), considering main effects and 2-way interactions as possible model terms. Pay attention to how you model continuous variables. Show all of your steps in developing your predictive model. At each step, provide a rationale for the modeling choices you have made.

3. Provide an appropriate test of goodness of fit. Interpret the result.

4. Complete model diagnostics and re-fit your model if needed. Explain any issues you might find in this process and how you will deal with it in your model.

5. Provide a **publication-quality** table reporting your resulting model (variables, beta (SE), p-value).

6. Provide the model formula for predicting the probability of dying in the hospital.

7. Provide an ROC curve, an estimate of the area under the ROC curve (AROC, with 95% CI) and a classification table for your model.

8. Evaluate the usefulness of your predictive model **in the validation sample (val=1).** Write a short conclusion.

9. Write a paragraph on your statistical methods, including model development, for the investigator to use in the manuscript. **This should be publication quality text.**

10. Investigators hypothesized they could better predict in-hospital mortality in this population by adding a measure of fitness for surgery assessed in the ER prior to surgery.   This measure is "asaps" in the dataset. Again **using just the val=0 sample:**

   a. Decide if the new asaps measure would be appropriate to add to your predictive model that you developed above.   Do NOT modify your predictive model above (other than possibly adding the new asaps measure. Attend to proper modeling of the variable.

   b. Assuming you have found that the new measure adds to your original prediction model, test whether addition of asaps adds significant prediction to the model in terms of the AROC (i.e., compare your full prediction model (with asaps) to your original prediction model).

   c. Write a short conclusion, comparing your original prediction model to the enhanced prediction model.