Homework 3

BSTA 513/613

Your name here - update this!!!!

2024-05-09

Directions

- [Download the .qmd file here.](#)

- You will need to download the datasets. Use [this link to download](#) the homework datasets needed in this assignment. If you do not want to make changes to the paths set in this document, then make sure the files are stored in a folder named "data" that is housed in the same location as this homework `.qmd` file.

- Please upload your homework to [Sakai](#). **Upload both your `.qmd` code file and the rendered `.html` file**

  - **Please rename you homework as `Lastname_Firstinitial_HW0.qmd`. This will help organize the homeworks when the TAs grade them.**

  - **Please also add the following line under subtitle: "BSTA 512/612": author: First-name Last-name with your first and last name so it is attached to the viewable document.**

- For each question, make sure to include all code and resulting output in the html file to support your answers.

- Show the work of your calculations using R code within a code chunk. Make sure that both your code and output are visible in the rendered html file. This is the default setting.

- If you are computing something by hand, you may take a picture of your work and insert the image in this file. You may also use LaTeX to write it inline.

- Write all answers in complete sentences as if communicating the results to a collaborator. This means including a sentence summarizing results in the context of the research study.

1

Questions

Question 1

This question is taken from the Hosmer and Lemeshow textbook. The ICU study data set consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult intensive care unit (ICU). The dataset should be available within Course Materials. The major goal of this study was to develop a logistic regression model to predict the probability of survival to hospital discharge of these patients. In this question, the primary outcome variable is vital status at hospital discharge, STA. Clinicians associated with the study felt that a key determinant of survival was the patient's age at admission, AGE. We will be building to a multivariable logistic regression model while adjusting for cancer part of the present problem (CAN), CPR prior to ICU admission (CPR), infection probable at ICU admission (INF), and level of consciousness at ICU admission (LOC).

A code sheet for the variables to be considered is displayed in Table 1.5 below (from the Hosmer and Lemeshow textbook, pg. 23). We refer to this data set as the ICU data.

Table 1.5 Code Sheet for the Variables in the ICU Data

Variable	Description	Codes/Values	Name
1	Identification code	ID number	ID
2	Vital status at hospital discharge	0 = Lived 1 = Died	STA
3	Age	Years	AGE
4	Gender	0 = Male 1 = Female	GENDER
5	Race	1 = White 2 = Black 3 = Other	RACE
6	Service at ICU admission	0 = Medical 1 = Surgical	SER
7	Cancer part of present problem	0 = No 1 = Yes	CAN
8	History of chronic renal failure	0 = No 1 = Yes	CRN
9	Infection probable at ICU admission	0 = No 1 = Yes	INF
10	CPR prior to ICU admission	0 = No 1 = Yes	CPR
11	Systolic blood pressure at ICU admission	mm Hg	SYS
12	Heart rate at ICU admission	Beats/min	HRA
13	Previous admission to an ICU within 6 months	0 = No 1 = Yes	PRE
14	Type of admission	0 = Elective 1 = Emergency	TYPE
15	Long bone, multiple, neck, single area, or hip fracture	0 = No 1 = Yes	FRA
16	PO_2 from initial blood gases	0 = >60 1 = $\leq$60	PO2
17	PH from initial blood gases	0 = $\geq$7.25 1 = <7.25	PH
18	PCO_2 from initial blood gases	0 = $\leq$45 1 = >45	PCO
19	Bicarbonate from initial blood gases	0 = $\geq$18 1 = <18	BIC
20	Creatinine from initial blood gases	0 = $\leq$2.0 1 = >2.0	CRE
21	Level of consciousness at ICU admission	0 = No coma or deep stupor 1 = Deep stupor 2 = Coma	LOC

Part a

From the above list (AGE, CAN, CPR, INF, and LOC) of independent variables, identify if each is a continuous, binary, or multi-level (>2) categorical variable.

Part b

For the binary and multi-level categorical variables, please identify a reference group for each. Include justification for the reference group.

Part c

Refer back to Part c from Homework 2's Question 4. Interpret the odds ratio for age in the simple logistic regression model. Please include the 95% confidence interval.

Part d

Compute the predicted probability of hospital discharge for a subject who is 63 years old. Compute the 95% confidence interval for the predicted probability and interpret the predicted probability.

Part e

For the categorical variables (binary and multi-group), please mutate the variables within the ICU dataset to set your chosen reference groups.

Part f

Write down the equation for the logistic regression model of STA on CPR.

Part g

Using the `glm()` function, obtain the maximum likelihood estimates of the coefficient parameters of the logistic regression model in Part f. Using these estimates, write down fitted logistic regression model.

Part h

Write a sentence interpreting the odds ratio for the coefficients in Part g's model. Please include the 95% confidence interval.

Part i

Write down the equation for the logistic regression model of STA on LOC.

Part j

Using the `glm()` function, obtain the maximum likelihood estimates of the coefficient parameters of the logistic regression model in Part i. Present the coefficient estimates. No need to write out the fitted regression equation.

Please take note of the warnings that you receive from fitting the `glm()` model and any large coefficient estimate with large confidence intervals. In this case, we have a category within LOC that has very few observations. (We will discuss this more in Lesson 13: Numerical Problems)

Check the number of observations that have a deep stupor and death at discharge and the number of observations that have a deep stupor and live at discharge. You can do this using the `table()` function to create a contingency table.

Part k

Write a sentence interpreting the odds ratio of death for the indicator of coma. Please include the 95% confidence interval.