Multiple Logistic Regression

Chapter 9, Lab 2: Template

OpenIntro Biostatistics

Topics

- Multiple logistic regression
- AIC (Akaike Information Criterion)

This lab introduces multiple logistic regression, a model for the association of a binary response variable with several predictor variable. The use of the Akaike Information Criterion (AIC) for comparing logistic regression models is also discussed.

The material in this lab corresponds to Section 9.xx in *OpenIntro Biostatistics*.

Introduction

Multiple logistic regression

The principles of simple logistic regression can be extended to a multiple logistic regression model with several predictors. Suppose that Y is a binary response variable, where $Y = 1$ represents the particular outcome of interest, and $X_1, X_2, \ldots, X_p$ are predictor variables.

The model for multiple logistic regression, where $p(x) = P(Y = 1|x_1, x_2, ..., x_p)$, is

$$\log\left[\frac{p(x)}{1 - p(x)}\right] = \beta_0 + \beta_1 x_2 + \beta_2 x_2 + \cdots + \beta_p x_p.$$

The estimated model equation has the form

$$\log\left[\frac{\hat{p}(x)}{1 - \hat{p}(x)}\right] = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p,$$

where $b_0, b_1, b_2, \ldots, b_p$ are estimates of the model parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$.

The coefficient b_j of a predictor x_j is the predicted change in the log of the estimated odds corresponding to a one unit change in x_j, when the values of all other predictors remain constant.

AIC (Akaike Information Criterion)

The **AIC (Akaike Information Criterion)** can be used to compare models. It is analogous to the adjusted R^2 for linear regression in that it penalizes a model for having a larger number of predictors.

A *lower* AIC is indicative of a more parsimonious model.

Background Information

Patients admitted to an intensive care unit (ICU) are either extremely ill or considered to be at great risk of serious complications. There are no widely accepted criteria for distinguishing between patients who should be admitted to an ICU and those for whom admission to other hospital units would be more appropriate. Thus, among different ICUs, there are wide ranges in a patient's chance of survival. When studies are done to compare effectiveness of ICU care, it is critical to have a reliable means of assessing the comparability of the different patient populations.

One such strategy for doing so involves the use of statistical modeling to relate empirical data for many patient variables to outcomes of interest. The following dataset consists of a sample of 200 subjects who were part of a much larger study on survival of patients following admission to an adult ICU.[1] The major goal of the study was to develop a logistic regression model to predict the probability of survival to hospital discharge.[2]

The following table provides a list of the variables in the dataset and their description. The data are accessible as the `icu` dataset in the `aplore3` package.

Variable	Description
`id`	patient ID number
`sta`	patient status at discharge, either `Lived` or `Died`
`age`	age in years (when admitted)
`gender`	gender, either `Male` or `Female`
`can`	cancer part of current issue, either `No` or `Yes`
`crn`	history of chronic renal failure, either `No` or `Yes`
`inf`	infection probable at admission, either `No` or `Yes`
`cpr`	CPR prior to admission, either `No` or `Yes`
`sys`	systolic blood pressure at admission, in mm Hg
`hra`	heart rate at admission, in beats per minute
`pre`	previous admission to an ICU within 6 months, either `No` or `Yes`
`type`	type of admission, either `Elective` or `Emergency`
`fra`	long bone, multiple, neck, single area, or hip fracture, either `No` or `Yes`
`po2`	PO_2 from initial blood gases, either `60` or `<=60`, in mm Hg
`ph`	pH from initial blood gases, either `>= 7.25` or `< 7.25`
`pco`	PCO_2 from initial blood gases, either `<= 45` or `>45`, in mm Hg
`bic`	HCO_3 (bicarbonate) from initial blood gases, either `>= 18` or `<18`, in mm Hg
`crea`	creatinine from initial blood gases, either `<= 2.0` or `> 2.0`, in mg/dL
`loc`	level of consciousness at admission, either `Nothing`, `Stupor`, or `Coma`

[1] From Hosmer D.W., Lemeshow, S., and Sturdivant, R.X. *Applied Logistic Regression.* 3^{rd} ed., 2013.

[2] Lemeshow S., et al. Predicting the outcome of intensive care unit patients. *Journal of the American Statistical Association* 83.402 (1988): 348-356.

1. Previously, age and CPR prior to ICU admission were each found to be statistically significantly associated with survival to discharge; an indicator for creatinine elevated beyond 2.0 mg/dL was also significantly associated with survival to discharge.

   Fit a single model to predict survival to discharge (`sta`) from age (`age`), CPR prior to admission (`cpr`), and an indicator of elevated creatinine level (`cre`).

   a) Write the model equation.

   b) Interpret the slope coefficients.

   c) Make predictions.

      i. Compare the odds of survival for those who did receive CPR prior to admission to those who did not receive CPR prior to admission.

      ii. Compare the odds of survival for those who had elevated creatinine level beyond 2.0 mg/dL to those who did not.

      iii. Calculate the odds of survival for a 65-year-old individual who did not receive CPR prior to admission and had creatinine level of 1.1 mg/dL. Is this individual more likely to survive than not survive?

      iv. Compare the odds of survival for a 30-year-old individual who received CPR prior to admission and had creatinine level of 1.5 mg/dL to the odds of survival for the individual described in part iii.

   d) Identify the significant predictors of survival to discharge from this model.

2. Fit a model for predicting survival to discharge from age, CPR prior to admission, and the interaction between age and CPR prior to admission.

   a) Interpret the slope of the interaction term.

   b) Assess whether this model is a better parsimonious model than the model fit in Question 1.

3. Consider two additional variables: the level of consciousness at admission (`loc`) and whether infection was probable (`inf`).

   a) Explore the relationship of each variable with survival to discharge.

      i. Create a two-way table; explore the relationship of level of consciousness at admission and survival to discharge. Summarize your findings.

      ii. Create a two-way table; explore the relationship of probable infection at admission and survival to discharge. Summarize your findings.

   b) Fit a logistic regression model to predict survival to discharge from age, prior CPR, probable infection, and level of consciousness.

      i. Write the model equation.

      ii. Interpret the slope coefficients.

      iii. Examine the output of `summary(glm())`. Do you notice anything surprising or unexpected?

c) Run the code in the template to create a binary version of `loc` that records whether or not an individual was conscious upon being admitted to the ICU. Fit the model from part b) using `loc.binary`.

   i. Interpret the slope coefficient of `loc.binary`.

   ii. Compare the odds of survival for a 50-year-old versus a 30-year-old, if infection was probable for both, and both received CPR prior to admission and entered the ICU conscious.

   iii. Suppose that a 70-year-old individual enters the ICU: CPR was not administered prior to admission, they entered the ICU conscious, and infection was not probable. Suppose that a 40-year-old enters the ICU: CPR was not administered prior to admission, they entered the ICU unconscious, and infection was not probable. Compare the odds of survival for these two individuals.

d) Assess whether the model from part c) is a better parsimonious model than the model fit in Question 1.