Lab 009 - Logistic Regression

EPIB607 - Inferential Statistics[a]

[a] McGill University

This lab introduces simple logistic regression, a model for the association of a binary response variable with a single predictor variable. Logistic regression generalizes methods for two-way tables and allows for the use of a numerical predictor.

1. Introduction

Odds and probabilities
    If the probability of an event A is p, the odds of the event are

$$\frac{p}{1-p}.$$

Given the odds of an event A, the probability of the event is

$$\frac{\text{odds}}{1+\text{odds}}.$$

Simple logistic regression
    Suppose that Y is a binary response variable and X is a predictor variable, where $Y = 1$ represents the particular outcome of interest.
    The model for a single variable logistic regression, where $p(x) = P(Y = 1 | X = x)$, is

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

The estimated model equation has the form

$$\log\left[\frac{\hat{p}(x)}{1-\hat{p}(x)}\right] = b_0 + b_1 x,$$

where b_0 and b_1 are estimates of the model parameters β_0 and β_1.

2. 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
cre	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*. 3rd 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.

3. Odds and probabilities

1. Create a two-way table of survival to discharge by whether CPR was administered prior to admission. The template provides code for re-leveling the `sta` variable such that `0` corresponds to `Died` and `1` corresponds to `Lived`.

   a) Calculate the odds of survival to discharge for those who did not receive CPR prior to ICU admission. Is someone who did not receive CPR prior to admission more likely to survive to discharge than to not survive to discharge?

   b) Calculate the odds of survival to discharge for those who received CPR prior to ICU admission. Is someone who received CPR prior to admission more likely to survive to discharge than not?

   c) Calculate the odds ratio of survival to discharge, comparing patients who receive CPR prior to admission to those who do not receive CPR prior to admission.

2. Creatinine level in the data are recorded as being either less than or equal to 2.0 mg/dL or greater than 2.0 mg/dL. A typical creatinine level is between 0.5 - 1.0 mg/dL, and elevated creatinine may be a sign of renal failure.

   a) Calculate the odds of survival to discharge for patients who have a creatinine level less than or equal to 2.0 mg/dL. From the odds, calculate the probability of survival to discharge for these patients.

   b) Calculate the probability of survival to discharge for patients who have a creatinine level greater than 2.0 mg/dL. From the probability, calculate the odds of survival to discharge for these patients.

   c) Compute and interpret the odds ratio of survival to discharge, comparing patients with creatinine > 2.0 mg/dL to those with creatinine ≤ 2.0 mg/dL.

4. Simple logistic regression

3. Fit a logistic regression model to predict survival to discharge from prior CPR.

   a) Write the model equation estimated from the data.

   b) Interpret the intercept. Confirm that the interpretation coheres with the answer to Question 1, part a).

   c) Interpret the slope coefficient. Compute the exponential of the slope coefficient and confirm that this matches the answer to Question 1, part c).

   d) Compute and interpret an odds ratio that summarizes the association between survival to discharge and prior CPR.

   e) Is there evidence of a statistically significant association between survival to discharge and prior CPR at $\alpha = 0.05$?

4. Fit a logistic regression model to predict survival to discharge from an indicator of elevated creatinine.

    a) Write the model equation estimated from the data.

    b) Interpret the intercept and slope coefficient.

    c) Compute and interpret an odds ratio that summarizes the association between survival to discharge and an indicator of elevated creatinine.

    d) Is there evidence of a statistically significant association between survival to discharge and an indicator of elevated creatinine at $\alpha = 0.05$?

5. Fit a logistic regression model to predict survival to discharge from age.

    a) Write the model equation estimated from the data.

    b) Does the intercept have a meaningful interpretation in the context of the data?

    c) Interpret the slope coefficient.

    d) Calculate the odds of survival to discharge for a 70-year-old individual. Is a 70-year-old individual more likely to survive than not?

    e) Calculate the odds ratio of survival to discharge comparing a 45-year-old individual to a 70-year-old individual.