Logistic regression,
explained using all the algebra you'll ever need

Statistical concepts for clinical investigators
David M. Thompson
Oklahoma Shared Translational and Clinical Resource
October 1, 2014

Logistic regression is widely used to explore a binary (yes/no) outcome.

> Logistic regression estimates odds ratios, for example the odds of the outcome in some high risk group compared to the odds of the outcome in some low risk group.

An example from:

> Katz BS, Fugate JE, Ameriso SF, Pujol-Lereis VA, Mandrekar J, Flemming KD, Kallmes DF, Rabinstein AA.
> Clinical worsening in reversible cerebral vasoconstriction syndrome.
> JAMA Neurol. 2014 Jan;71(1):68-73. doi: 10.1001/jamaneurol.2013.4639.
>
> "Univariate logistic regression analysis was performed to assess associations between patient characteristics and clinical worsening."

The most basic regression model takes a form that you encountered in secondary school, the equation of a straight line. It relates an outcome variable (y) to a predictor variable (x):

$y=mx+b$, where m is the slope and b is the y-intercept

In linear regression, the outcome (y) is continuous, so:

$y=mx+b$
$y=\beta_1 x+\beta_0$
$y= \beta_0+\beta_1 x$ (β_0 is the y-intercept and β_1 is the slope)

In logistic regression,

the outcome (y) is binary with just two possible values:

y=1 for patients with clinical worsening;

y=0 for patients without clinical worsening.

So, while the right side of the regression equation is the same, the left side of the equation is modified or transformed into the log of the odds:

$$\ln\left[\frac{p(y=1)}{p(y=0)}\right] = \beta 0 + \beta 1 x$$

The left side of the equation uses a "logit transformation," so this is "logistic regression."

The right side of the equation lists a single variable, so this is "univariate logistic regression."

Example: Association between clinical worsening and a history of migraine headaches.

This example skips no steps and shows the logic of logistic regression.

Table 1 shows that, among those with no clinical worsening, 9 of 39 (23.1%) had a history of migraine HA. Among those with clinical worsening, 7 of 20 (35%) had a history of migraine headaches.

Software that performs logistic regression produces the following logistic regression equation for **the log odds of clinical worsening**

$$= \ln[p(y=1)/p(y=0)] = \beta 0+\beta 1x$$
$$= -0.84 + 0.59X$$

T**he odds of clinical worsening** are obtained by exponentiating both sides of the equation:

$$p(y=1)/p(y=0) = \exp(-0.8362 + 0.5850*X)$$

For those with a history of migraine (X=1), these odds are $e^{(-0.8362 + 0.5850*1)} = e^{-0.2513)} = 0.7778$

For those with no history of migraine (X=0), these odds are $e^{(-0.8362 + 0.5850*0)} = e^{-0.8362)} = 0.4333$

The ratio of these odds is 0.7778/0.4333 = 1.7951

The odds of clinical worsening for those with a history of migraine are 1.795 times the odds of clinical worsening for those with no history of migraines. The 95% CI on this odds ratio, calculated from SAS PROC LOGISTIC, is (0.550, 5.858). A p-value on the hypothesis that the true odds are equal to 1 is 0.3324

Compare this estimate, its confidence interval, and the associated p value to those reported in Table 2 for "migraine headaches."