Stat 4620: Day 23

(4/16)

Sec. 9.6: Application of Likelihood Estimation to Linear Regression

Example: There is often interest in the relationship between two variables. For example, let

$$Y = \text{Calculus grade and } X = \text{ Math aptitude test score}$$

How well does knowing X predict Y? Given data for n students $(x_1, y_1)(x_2, y_2), \ldots, (x_n, y_n)$, the **regression model** may be written as follows:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n$$

where $\{\epsilon_i\}$ are independent $N(0, \sigma^2)$. An equivalent way of writing the model is

$$\{Y_i\} \text{ are independent, and } Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

This allows likelihood estimation

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y_i - (\beta_0 + \beta_1 x_i)]^2}{2\sigma^2}}$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{\sum [y_i - (\beta_0 + \beta_1 x_i)]^2}{2\sigma^2}}$$

and

$$\ln L(\beta_0, \beta_1, \sigma^2) = \frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum [y_i - (\beta_0 + \beta_1 x_i)]^2}{2\sigma^2}$$

If we take partial derivatives, this gives a set of three equations in three unknowns. Solution of the linear equations results in the **least squares estimates**

$$\hat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\sigma}^2 = \frac{\sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n}$$

Sec: Likelihood Estimation in Logistic Regression (GLM)

Example: A sample of $n = 25$ programmers who were asked to complete a certain coding task within a specified length of time. The table below presents data on whether they were successful ($Y = 1$) or not ($Y = 0$), and a measure of their programming experience in months.

Person	Months	Y	Success
1	14	0	No
2	29	0	No
3	6	0	No
4	25	1	Yes
5	18	1	Yes
6	4	0	No
7	18	0	No
8	12	0	No
9	22	1	Yes
:			
:			
22	4	0	No
23	28	1	Yes
24	22	1	Yes
25	8	1	Yes

We want to answer the following questions:

1. Is experience a predictor of success?
2. How strongly? Or similarly, what is the "effect size"?

0.1 The generalized linear model

Let

$$\begin{aligned}\pi_i = P[Y_i = 1|x_i] &= g(\beta_0 + \beta_1 x_i) \qquad (g \text{ is called the } \textit{link function}) \\ &= \frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}} \qquad (\text{ logistic link})\end{aligned}$$

(Plot here)

The Bernoulli probability function is $\begin{array}{c|cc} y_i & 1 & 0 \\ \hline P(Y_i = y_i \mid x_i) & \pi_i & 1-\pi_i \end{array}$. We may write this more compactly as

$$P(Y_i = y_i \mid x_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i}, \quad y_i = 0, 1$$

Likelihood estimation:

$$\begin{aligned}L(\beta_0, \beta_1) &= \prod_{i=1}^{n} P[Y_i = y_i|x_i] = \prod \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\ &= P[Y_1 = 0 \mid x_1 = 14] \cdot P[Y_2 = 0 \mid x_2 = 29] \cdots P[Y_{25} = 1 \mid x_{25} = 8] \\ &= \left(1 - \frac{e^{\beta_0+\beta_1 14}}{1+e^{\beta_0+\beta_1 14}}\right)\left(1 - \frac{e^{\beta_0+\beta_1 29}}{1+e^{\beta_0+\beta_1 29}}\right)\cdots\left(\frac{e^{\beta_0+\beta_1 8}}{1+e^{\beta_0+\beta_1 8}}\right)\end{aligned}$$

Find $(\hat{\beta}_0, \hat{\beta}_1)$ that maximizes $L(\beta_0, \beta_1)$. The mle estimates are $(\hat{\beta}_0 = -3.0597, \hat{\beta}_1 = 0.1615)$, and the mle of the logistic prediction function is

$$\hat{\pi}_i = \frac{e^{-3.0597 + .1615\, x_i}}{1 + e^{-3.0597 + .1615\, x_i}}$$

Table: SAS Output: Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.0597	1.2594	5.9029	0.0151
months	1	0.1615	0.0650	6.1760	0.0129

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
months	1.175	1.035	1.335

To calculate the standard errors, recall that

$$V(\hat{\boldsymbol{\theta}}) \doteq \mathbf{I}_n^{-1}(\boldsymbol{\theta}))$$

so

$$\text{SE}(\hat{\theta}_i) = \sqrt{\left[\mathbf{I}_n^{-1}(\boldsymbol{\theta})\right]_{ii}}$$

where

$$\begin{aligned}
\mathbf{I}_n(\boldsymbol{\theta}) &= -E\left[\frac{\partial^2}{\boldsymbol{\theta}^2} l(\boldsymbol{\theta})\right] \\
&= -E\left[\begin{array}{cc} \frac{\partial^2}{\partial \beta_0^2} \ln L(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ln L(\beta_0, \beta_1) \\ \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ln L(\beta_0, \beta_1) & \frac{\partial^2}{\partial \beta_1^2} \ln L(\beta_0, \beta_1) \end{array}\right]
\end{aligned}$$

For example,

$$\begin{aligned}
\ln L(\beta_0, \beta_1) &= \sum \ln \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right)^{y_i} \left(1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right)^{1 - y_1} \\
&= \sum y_i \ln \left(\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right) + \sum (1 - y_i) \ln \left(1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}\right)
\end{aligned}$$

0.2 Odds ratio

Recall that

$$P[Y_i = 1 \mid x_i] = \pi_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

and

$$P[Y_i = 0 \mid x_i] = 1 - \pi_i = 1 - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} = \frac{1+e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}} - \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$
$$= \frac{1}{1+e^{\beta_0+\beta_1 x_i}}$$

Then

$$\text{Odds}_i = \frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1 x_i}$$

The odds ratio provides a convenient measure of effect size. For example, by how much does an additional month of experience increase the odds of success?

$$OR = \frac{\text{Odds}(Y=1 \mid x=10)}{\text{Odds}(Y=1 \mid x=9)} = \frac{e^{\beta_0+\beta_1 10}}{e^{\beta_0+\beta_1 9}} = e^{\beta_1}$$

For our programming example,

$$\hat{OR} = e^{\hat{\beta}_1} = e^{.1615} = 1.175$$

Furthermore,

$$OR = \frac{\text{Odds}(Y=1 \mid x=20)}{\text{Odds}(Y=1 \mid x=19)} = \frac{e^{\beta_0+\beta_1 20}}{e^{\beta_0+\beta_1 19}} = e^{\beta_1}$$

so we can say that *the odds of success increases by 17.5% for every additional month of programming experience.* In general

$$\frac{\text{Odds}(Y=1 \mid x=b+\delta)}{\text{Odds}(Y=1 \mid x=b)} = e^{\beta_1 \delta}$$

Example (con't.) What is the effect of an additional 6 months of programming experience?

Solution.

$$e^{\beta_1 \delta} = e^{.1615(6)} = 2.635$$

Comment: Note that $e^{.1615(6)} = \left[e^{.1615}\right]^6 = [1.175]^6$ ■

0.3 Confidence interval for odds ratio

First we construct a 95% confidence interval for β_1. Let (L, U) be the lower and upper endpoints of the confidence interval for β_1

$$\hat{\beta}_1 \pm t_{.975,n-2}\text{SE}_{\hat{\beta}_1}$$

Then

$$.95 = P\left(L \le \beta_1 \le U\right) = P\left(e^L \le e^{\beta_1} \le e^U\right)$$

so that

$$\left(e^L, e^U\right) = \left(e^{\hat{\beta}_1 - t\,\text{SE}_{\hat{\beta}_1}}, e^{\hat{\beta}_1 + t\,\text{SE}_{\hat{\beta}_1}}\right)$$

is a 95% confidence interval for the odds ratio.

Example (Association between homocysteine and risk of stroke)
A study involving $n = 2258$ patients used logistic regression models, odds ratios (OR), and their

associated 95% confidence intervals (CI) to estimate an independent effect of homocysteine (Hcy) on the risk of stroke. According to Table 2 of the paper, the odds ratio was

$$OR = 1.024\ (CI : 1.015 - 1.034)$$

for

$$\text{Model 1: } \pi_i = \frac{e^{\beta_0+\beta_1 \ln HCy}}{1 + e^{\beta_0+\beta_1 \ln Hcy}}$$

The Hcy effects are similar when adjusted for age and sex:

$$\text{Model 2: } \pi_i = \frac{e^{\beta_0+\beta_1 \ln Hcy+\beta_2 Age+\beta_3 Sex}}{1 + e^{\beta_0+\beta_1 \ln Hcy+\beta_2 Age+\beta_3 Sex}}$$

where Sex=1 if male, Sex=0 if female. This now requires likelihood maximization over $p = 4$ parameters. The estimated coefficient $\hat{\beta}_1$ will change, but odds ratio is interpreted similarly. Since

$$\text{Odds}_i = \frac{\pi_i}{1 - \pi_i} = e^{\beta_0+\beta_1 \ln Hcy+\beta_2 Age+\beta_3 Sex}$$

then the odds ratio of stroke *for every additional unit increase* in ln Hcy is

$$\frac{\text{Odds}(Y = 1 \mid \ \ln Hcy + 1, Age, Sex)}{\text{Odds}(Y = 1 \mid \ \ln Hcy, Age, Sex)} = \frac{e^{\beta_0+\beta_1 \ln Hcy+1+\beta_2 Age+\beta_3 Sex}}{e^{\beta_0+\beta_1 \ln Hcy+\beta_2 Age+\beta_3 Sex}} = e^{\beta_1}$$

A categorical analysis was done with the ln Hcy variable divided into Low (< 1.09), Middle ($1.09 - 1.23$) and High (> 1.23). Two indicator variables W_1 and W_2 were created and coded as follows:

$$(W_1, W_2) = \begin{cases} (0,0) & \text{if} \quad \ln Hcy < 1.09 \\ (1,0) & \text{if} \quad 1.09 \leq \ln Hcy \leq 1.23 \\ (0,1) & \text{if} \quad 1.23 < \ln Hcy \end{cases}$$

Note that W_1 is an indicator for middle values and W_2 is an indicator for high values. The following model was fit to data:

$$\pi_i = \frac{e^{\beta_0+\beta_1 W_1+\beta_2 W_2}}{1 + e^{\beta_0+\beta_1 W_1+\beta_2 W_2}}$$

or equivalently

$$\text{Odds}_i = \frac{\pi_i}{1 - \pi_i} = e^{\beta_0+\beta_1 W_1+\beta_2 W_2}$$

Then

$$\frac{\text{Odds}(Y = 1 \mid \text{Middle})}{\text{Odds}(Y = 1 \mid \text{Low })} = \frac{e^{\beta_0+\beta_1(1)+\beta_2(0)}}{e^{\beta_0+\beta_1(0)+\beta_2(0)}} = e^{\beta_1}$$

and

$$\frac{\text{Odds}(Y = 1 \mid \text{High})}{\text{Odds}(Y = 1 \mid \text{Low })} = \frac{e^{\beta_0+\beta_1(0)+\beta_2(1)}}{e^{\beta_0+\beta_1(0)+\beta_2(0)}} = e^{\beta_2}$$

Here, Low is the *reference category* against which we compare Middle and High. The choice of reference category is arbitrary, and chosen for simplicity of interpretation. Table 3 gives these odds ratios, stratified by gender, age, and other variables of interest.

Caution: OR=4.0 means a quadrupling of odds, not a quadrupling of probabilities.

Prob	Odds
1/10	$(1/10)/(9/10) = 1:9$
1/4	$(1/4)/(3/4) = 1:3$
1/2	$(1/2)/(1/2) = 1:1$
3/4	$(3/4)/(1/4) = 3:1$
9/10	$(9/10)/(1/10) = 9:1$