Logit and Probit Models for Categorical Response Variables

The general linear model that we have studied thus far in the course has worked well for many continous response variables. However, they are not appropriate for a categorical response variable such as a binary response variable. Suppose the response Y_i has only two outcomes, which we will label 0 and 1, and we consider the model

$$Y_i = \alpha + \beta X_i + \varepsilon_i,$$

where the errors ε_i are independent and satisfy $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. If we define the success probability π as $\pi_i = \Pr(Y = 1|X = x_i)$ then we have $E(Y|x_i) = \pi_i(1) + (1-\pi_i)(0) = \pi_i$. Using this result, if we take the expected value of our model above we have

$$\pi_i = \alpha + \beta X_i.$$

This model is called the linear-probability model. As noted in the text, this model is untenable for a variety of reasons: i) The errors ε_i cannot have a Normal distribution, ii) the variance of the errors is not constant, and more importantly, iii) the predicted probabilities from this model π_i are not confined to the $[0, 1]$ interval.

Transformations of π

Instead of modeling π directly, it makes more sense to model a transformation of the linear function $\alpha+\beta X_i$ so that the transformed values are confined to the $[0, 1]$ interval. Cumulative distribution functions (CDFs) are perfect for this situation (especially if they are strictly increasing), so we should use a model of the form $\pi_i = P(\alpha + \beta X_i)$, where P is a CDF. If P is strictly increasing, then we can invert it to rewrite the model as: $P^{-1}(\pi_i) = \alpha + \beta X_i$. The transformation P is often chosen to be the CDF of the standard-normal distribution, giving the linear probit model: $\pi_i = \Phi(\alpha + \beta X_i)$, or the CDF of the logistic distribution, giving the linear logit model $\pi_i = \Lambda(\alpha + \beta X_i)$. When properly standardized these two distributions are nearly indistinguishable for all but large data sets, but the logit model is more commonly used due to less computational complexity and the interpretation that

$$\frac{\pi_i}{1-\pi_i} = \exp(\alpha + \beta X_i) \text{ or equivalently}$$
$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta X_i.$$

The quantity $\pi_i/(1-\pi_i)$ is the odds that $Y_i = 1$, and $\log[\pi_i/(1-\pi_i)]$ is called the logit of π_i.

Logit and Probit Models for Multiple Regression

The models above generalize to multiple regression, using η_i as the linear predictor we have for the logit model

$$\begin{aligned}\pi_i &= \Lambda(\eta_i) = \Lambda(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) \\ &= \frac{1}{1 + \exp[-(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})]},\end{aligned}$$

or equivalently

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}.$$

One way to interpret that model coefficients is via the odds

$$\begin{aligned}\frac{\pi_i}{1-\pi_i} &= \exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}) \\ &= e^{\alpha}(e^{\beta_1})^{X_{i1}} \cdots (e^{\beta_k})^{X_{ik}},\end{aligned}$$

where the regression coefficient e^{β_j} is the multiplicative effect on the odds of increasing X_j by 1, with the other Xs constant. Assuming a sample of n observations $y_1, y_2, \ldots, y_n$ is independent, then the likelihood function for the model is

$$
\begin{aligned}
p(y_1, y_2, \ldots, y_n) &= \prod_{i=1}^{n} p(y_i) \\
&= \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{(1-y_i)} \\
&\prod_{i=1}^{n} \left(\frac{\pi_i}{1-\pi_i} \right)^{y_i} (1-\pi_i) \\
&\prod_{i=1}^{n} (\exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}))^{y_i} \\
&\times \left(\frac{1}{1 + \exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})} \right).
\end{aligned}
$$

The usual (large-sample) tests and confidence intervals for maximum likelihood estimators can be used, such as a Wald statistic for the null hypothesis $H_0 : \beta_j = \beta_j^{(0)}$,

$$
Z_0 = \frac{B_j - \beta_j^{(0)}}{SE(B_j)},
$$

and its associated confidence interval. Tests of sets of coefficients can be conducted by defining a full model L_1 and null model L_0 and using the generalized likelihood-ratio test statistic

$$
G_0^2 = 2(\log L_1 - \log L_0),
$$

which has an asymptotic chi-squared distribution under H_0.