18 Additional Models

In this chapter we briefly discuss some models that are not linear in the parameters or that have an error structure different from that assumed in previous chapters.

18.1 NONLINEAR REGRESSION

A *nonlinear regression model* can be expressed as

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{18.1}$$

where $f(\mathbf{x}_i, \boldsymbol{\beta})$ is a nonlinear function of the parameter vector $\boldsymbol{\beta}$. The error term ε_i is sometimes assumed to be distributed as $N(0, \sigma^2)$. An example of a nonlinear model is the exponential model

$$y_i = \beta_0 + \beta_1 e^{\beta_2 x_i} + \varepsilon_i.$$

Estimators of the parameters in (18.1) can be obtained using the method of least squares. We seek the value of $\hat{\boldsymbol{\beta}}$ that minimizes

$$Q(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} [y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})]^2. \tag{18.2}$$

A simple analytical solution for $\hat{\boldsymbol{\beta}}$ that minimizes (18.2) is not available for nonlinear $f(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$. An iterative approach is therefore used to obtain a solution. In general, the resulting estimators in $\hat{\boldsymbol{\beta}}$ are not unbiased, do not have minimum variance, and are not normally distributed. However, according to large-sample theory, the estimators are almost unbiased, have near-minimum variance, and are approximately normally distributed.

Linear Models in Statistics, *Second Edition*, by Alvin C. Rencher and G. Bruce Schaalje

Inferential procedures, including confidence intervals and hypothesis tests, are available for the least-squares estimator $\hat{\boldsymbol{\beta}}$ obtained by minimizing (18.2). Diagnostic procedures are available for checking on the model and on the suitability of the large-sample inferential procedures.

For details of the above procedures, see Gallant (1975), Bates and Watts (1988), Seber and Wild (1989), Ratkowsky (1983, 1990), Kutner et al. (2005, Chapter 13), Hocking (1996, Section 11.2), Fox (1997, Section 14.2), and Ryan (1997, Chapter 13).

18.2 LOGISTIC REGRESSION

In some regression situations, the response variable y has only two possible outcomes, for example, high blood pressure or low blood pressure, developing cancer of the esophagus or not developing it, whether a crime will be solved or not solved, and whether a bee specimen is a "killer" (africanized) bee or a domestic honey bee. In such cases, the outcome y can be coded as 0 or 1 and we wish to predict the outcome (or the probability of the outcome) on the basis of one or more x's.

To illustrate a linear model in which y is binary, consider the model with one x:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i; \quad y_i = 0, 1; \quad i = 1, 2, \ldots, n. \tag{18.3}$$

Since y_i is 0 or 1, the mean $E(y_i)$ for each x_i becomes the proportion of observations at x_i for which $y_i = 1$. This can be expressed as

$$\begin{aligned} E(y_i) &= P(y_i = 1) = p_i, \\ 1 - E(y_i) &= P(y_i = 0) = 1 - p_i. \end{aligned} \tag{18.4}$$

The distribution $P(y_i = 0) = 1 - p_i$ and $\mathrm{P}(y_i = 1) = p_i$ in (18.4) is known as the *Bernoulli distribution*. By (18.3) and (18.4), we have

$$E(y_i) = p_i = \beta_0 + \beta_1 x_i. \tag{18.5}$$

For the variance of y_i, we obtain

$$\begin{aligned} \mathrm{var}(y_i) &= E[y_i - E(y_i)]^2 \\ &= p_i(1 - p_i). \end{aligned} \tag{18.6}$$

By (18.5) and (18.6), we obtain

$$\mathrm{var}(y_i) = (\beta_0 + \beta_1 x_i)(1 - \beta_0 - \beta_1 x_i),$$

and the variance of each y_i depends on the value of x_i. Thus the fundamental assumption of constant variance is violated, and the usual least-squares estimators $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ computed as in (6.5) and (6.6) will not be optimal (see Theorem 7.3d).

To obtain optimal estimators of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, we could use generalized least-squares estimators

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

as in Theorem 7.8a, but there is an additional challenge in fitting the linear model (18.5). Since $E(y_i) = p_i$ is a probability, it is limited by $0 \leq p_i \leq 1$. If we fit (18.5) by generalized least squares to obtain

$$\hat{p}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,$$

then $\hat{p}_i$ may be less than 0 or greater than 1 for some values of x_i. A model for $E(y_i)$ that is bounded between 0 and 1 and reaches 0 and 1 asymptotically (instead of linearly) would be more suitable. A popular choice is the *logistic regression model.*

$$p_i = E(y_i) = \frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}} = \frac{1}{1 + e^{-\beta_0-\beta_1 x_i}}. \tag{18.7}$$

This model is illustrated in Figure 18.1. The model in (18.7) can be linearized by the simple transformation

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i, \tag{18.8}$$

sometimes called the *logit transformation*.

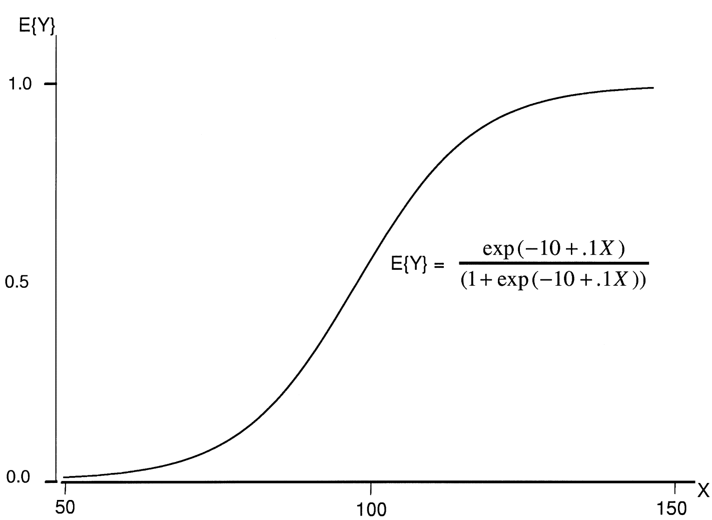

Figure 18.1 Logistic regression function.

The parameters β_0 and β_1 in (18.7) and (18.8) are typically estimated by the method of maximum likelihood (see Section 7.2). For a random sample y_1, y_2, $\ldots, y_n$ from the Bernoulli distribution with $P(y_i = 0) = 1 - p_i$ and $P(y_i = 1) = p_i$, the likelihood function becomes

$$
\begin{aligned}
L(\beta_0, \beta_1) = f(y_1, y_2, \ldots, y_n; \beta_0, \beta_1) &= \prod_{i=1}^{n} f_i(y_i; \beta_0, \beta_1) \\
&= \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}. \qquad (18.9)
\end{aligned}
$$

Taking the logarithm of both sides of (18.9) and using (18.8), we obtain

$$
\ln L(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \ln(1 + e^{\beta_0 + \beta_1 x_i}). \qquad (18.10)
$$

Differentiating (18.10) with respect to β_0 and β_1 and setting the results equal to zero gives

$$
\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{1}{1 + e^{-\hat{\beta}_0 - \hat{\beta}_1 x_i}} \qquad (18.11)
$$

$$
\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \frac{x_i}{1 + e^{-\hat{\beta}_0 - \hat{\beta}_1 x_i}}. \qquad (18.12)
$$

These equations can be solved iteratively for $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$.

The logistic regression model in (18.7) can be readily extended to include more than one x. Using the notation $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ and $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})'$, the model in (18.7) becomes

$$
p_i = E(y_i) = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i'\boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{x}_i'\boldsymbol{\beta}}},
$$

and (18.8) takes the form

$$
\ln\left(\frac{p_i}{1 - p_i}\right) = \mathbf{x}_i'\boldsymbol{\beta}, \qquad (18.13)
$$

where $\mathbf{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$. For binary y_i ($y_i = 0, 1; i = 1, 2, \ldots, n$), the mean and variance are given by (18.4) and (18.6). The likelihood function and the value of $\boldsymbol{\beta}$ that maximize it are found in a manner analogous to the approach used to find β_0 and β_1. Confidence intervals, tests of significance, measures of fit, subset selection procedures, diagnostic techniques, and other procedures are available.

Logistic regression has been extended from binary to a *polytomous* logistic regression model in which y has several possible outcomes. These may be ordinal such as large, medium, and small, or categorical such as Republicans, Democrats, and Independents. The analysis differs for the ordinal and categorical cases.

For details of these procedures, see Hosmer and Lemeshow (1989), Hosmer et al. (1989), McCullagh and Nelder (1989), Myers (1990, Section 7.4), Kleinbaum (1994), Stapleton (1995, Section 8.8), Stokes et al. (1995, Chapters 8 and 9), Kutner et al. (2005), Chapter 14, Hocking (1996, Section 11.4), Ryan (1997, Chapter 9), Fox (1997, Chapter 15), Christensen (1997), and McCulloch and Searle (2001, Chapter 5).

18.3 LOGLINEAR MODELS

In the analysis of categorical data, we often use loglinear models. To illustrate a loglinear model for categorical data, consider a two-way contingency table with frequencies (counts) designated as y_{ij} as in Table 18.1, with $y_{i.} = \sum_{j=1}^{s} y_{ij}$ and $y_{.j} = \sum_{i=1}^{r} y_{ij}$. The corresponding cell probabilities p_{ij} are given in Table 18.2, with $p_{i\cdot} = \sum_{j=1}^{s} p_{ij}$ and $p_j = \sum_{i=1}^{r} p_{ij}$.

The hypothesis that A and B are independent can be expressed as $H_0 : p_{ij} = p_{i.}p_{.j}$ for all i, j. Under H_0, the expected frequencies are

$$E(y_{ij}) = np_{i.}p_{.j}.$$

This becomes linear if we take the logarithm of both sides:

$$\ln E(y_{ij}) = \ln n + \ln p_{i.} + \ln p_{.j}.$$

TABLE 18.1 Contingency Table Showing Frequencies y_{ij} (Cell Counts) for an $r \times s$ Classification of Two Categorical Variables A and B

Variable	B_1	B_2	$\dots$	B_s	Total
A_1	y_{11}	y_{12}	$\dots$	y_{1s}	$y_{1\cdot}$
A_2	y_{21}	y_{22}	$\dots$	y_{2s}	$y_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	y_{r1}	y_{r2}	$\dots$	y_{rs}	$y_{r\cdot}$
Total	$y_{.1}$	$y_{.2}$	$\dots$	$y_{.s}$	$y_{..} = n$

TABLE 18.2 Cell Probabilities for an $r \times s$ Contingency Table

Variable	B_1	B_2	$\cdots$	B_s	Total
A_1	p_{11}	p_{12}	$\cdots$	p_{1s}	$p_{1.}$
A_2	p_{21}	p_{22}	$\cdots$	p_{2s}	$p_{2.}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
A_r	p_{r1}	p_{r2}	$\cdots$	p_{rs}	$p_{r.}$
Total	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot s}$	$p_{..} = 1$

To test H_0: $p_{ij} = p_{i.}p_{.j}$, we can use the likelihood ratio test. The likelihood function is given by the multinomial density

$$L(p_{11}, p_{12}, \ldots, p_{rs}) = \frac{n!}{y_{11}!y_{12}! \cdots y_{rs}!} p_{11}^{y_{11}} p_{12}^{y_{12}} \cdots p_{rs}^{y_{rs}}.$$

The unrestricted maximum likelihood estimators of p_{ij} (subject to $\sum_{ij} p_{ij} = 1$) are $\hat{p}_{ij} = y_{ij}/n$, and the estimators under H_0 are $\hat{p}_{ij} = y_{i.}y_{.j}/n^2$ (Christensen 1997, pp. 42–46). The likelihood ratio is then given by

$$\mathrm{LR} = \prod_{i=1}^{r} \prod_{j=1}^{s} \left(\frac{y_{i.}y_{.j}}{ny_{ij}} \right)^{y_{ij}}.$$

The test statistic is

$$-2 \ln \mathrm{LR} = 2 \sum_{ij} y_{ij} \ln \left(\frac{ny_{ij}}{y_{i.}y_{.j}} \right),$$

which is approximately distributed as $\chi^2[(r-1)(s-1)]$.

For further details of loglinear models, see Ku and Kullback (1974), Bishop et al. (1975), Plackett (1981), Read and Cressie (1988), Santner and Duffy (1989), Agresti (1984, 1990) Dobson (1990, Chapter 9), Anderson (1991), and Christensen (1997).

18.4 POISSON REGRESSION

If the response y_i in a regression model is a count, the Poisson regression model may be useful. The Poisson probability distribution is given by

$$f(y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, 2, \ldots.$$

The *Poisson regression model* is

$$y_i = E(y_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where the y_i's are independently distributed as Poisson random variables and $\mu_i = E(y_i)$ is a function of $\mathbf{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$. Some commonly used functions of $\mathbf{x}_i'\boldsymbol{\beta}$ are

$$\mu_i = \mathbf{x}_i'\boldsymbol{\beta}, \quad \mu_i = e^{\mathbf{x}_i'\boldsymbol{\beta}}, \quad \mu_i = \ln(\mathbf{x}_i'\boldsymbol{\beta}). \tag{18.14}$$

In each of the three cases in (18.14), the values of μ_i must be positive.

To estimate $\boldsymbol{\beta}$, we can use the method of maximum likelihood. Since y_i has a Poisson distribution, the likelihood function is given by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n} \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!},$$

where μ_i is typically one of the three forms in (18.14). Iterative methods can be used to find the value of $\hat{\boldsymbol{\beta}}$ that maximizes $L(\boldsymbol{\beta})$. Confidence intervals, tests of hypotheses, measures of fit, and other procedures are available. For details, see Myers (1990, Section 7.5) Stokes et al. (1995, pp. 471–475), Lindsey (1997), and Kutner et al. (2005, Chapter 14).

18.5 GENERALIZED LINEAR MODELS

Generalized linear models include the classical linear regression and ANOVA models covered in earlier chapters as well as logistic regression in Section 18.2 and some forms of nonlinear regression in Section 18.1. Also included in this broad family of models are loglinear models for categorical data in Section 18.3 and Poisson regression models for count data in Section 18.4. This expansion of traditional linear models was introduced by Wedderburn (1972).

A *generalized linear model* can be briefly characterized by the following three components.

1. Independent random variables $y_1, y_2, \ldots, y_n$ with expected value $E(y_i) = \mu_i$ and density function from the exponential family [described below in (18.15)].
2. A *linear predictor*

$$\mathbf{x}_i'\boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$

3. A *link function* that describes how $E(y_i) = \mu_i$ relates to $\mathbf{x}_i'\boldsymbol{\beta}$:

$$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}.$$

4. The link function $g(\mu_i)$ is often nonlinear.

A density $f(y_i, \theta_i)$ belongs to the *exponential family* of density functions if $f(y_i, \theta_i)$ can be expressed in the form

$$f(y_i, \theta_i) = \exp[y_i\theta_i + b(\theta_i) + c(y_i)]. \tag{18.15}$$

A scale parameter such as σ^2 in the normal distribution can be incorporated into (18.15) by considering it to be known and treating it as part of θ_i. Alternatively, an additional parameter can be inserted into (18.15). The exponential family of density functions provides a unified approach to estimation of the parameters in generalized linear models.

Some common statistical distributions that are members of the exponential family are the binomial, Poisson, normal, and gamma [see (11.7)]. We illustrate three of these in Example 18.5.

Example 18.5. The binomial probability distribution can be written in the form of (18.15) as follows:

$$\begin{aligned} f(y_i, p_i) &= \binom{n_i}{y_i} p_i^{y_i}(1 - p_i)^{n_i - y_i} \\ &= \exp\left[y_i \ln p_i - y_i \ln(1 - p_i) + n_i \ln(1 - p_i) + \ln\binom{n_i}{y_i}\right] \\ &= \exp\left[y_i \ln\left(\frac{p_i}{1 - p_i}\right) + n_i \ln(1 - p_i) + \ln\binom{n_i}{y_i}\right] \\ &= \exp[y_i\theta_i + b(\theta_i) + c(y_i)], \end{aligned} \tag{18.16}$$

where $\theta_i = \ln[p_i/(1 - p_i)]$, $b(\theta_i) = n_i \ln(1 - p_i) = -n_i \ln(1 + e^{\theta_i})$, and $c(y_i) = \ln\binom{n_i}{y_i}$.

The Poisson distribution can be expressed in exponential form as follows:

$$\begin{aligned} f(y_i, \mu_i) &= \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} = \exp[y_i \ln \mu_i - \mu_i - \ln(y_i!)] \\ &= \exp[y_i\theta_i + b(\theta_i) + c(y_i)], \end{aligned}$$

where $\theta_i = \ln \mu_i$, $b(\theta_i) = -\mu_i = -e^{\theta_i}$, and $c(y_i) = -\ln(y_i!)$.

The normal distribution $N(\mu_i, \sigma^2)$ can be written in the form of (18.15) as follows:

$$
\begin{aligned}
f(y_i, \mu_i) &= \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(y_i-\mu_i)^2/2\sigma^2} \\
&= \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(y_i^2-2y_i\mu_i+\mu_i^2)/2\sigma^2} \\
&= \exp\left[-\frac{y_i^2}{2\sigma^2} + \frac{y_i\mu_i}{\sigma^2} - \frac{\mu_i^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right] \\
&= \exp[y_i\theta_i + b(\theta_i) + c(y_i)],
\end{aligned}
$$

where $\theta_i = \mu_i/\sigma^2$, $b(\theta_i) = \sigma^2\theta_i^2/2$, and $c(y_i) = -y_i^2/2\sigma^2 - \frac{1}{2}\ln(2\pi\sigma^2)$. □

To obtain an estimator of β in a generalized linear model, we use the method of maximum likelihood. From (18.15), the likelihood function is given by

$$
L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \exp[y_i\theta_i + b(\theta_i) + c(y_i)].
$$

The logarithm of the likelihood is

$$
\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i\theta_i + \sum_{i=1}^{n} b(\theta_i) + \sum_{i=1}^{n} c(y_i). \tag{18.17}
$$

For the exponential family in (18.15), it can be shown that

$$
E(y_i) = \mu_i = -b'(\theta_i),
$$

where $b'(\theta_i)$ is the derivative with respect to θ_i. This relates θ_i to the link function

$$
g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}.
$$

Differentiating (18.17) with respect to each β_i, setting the results equal to zero, and solving the resulting (nonlinear) equations iteratively (iteratively reweighted least squares) gives the estimators $\hat{\beta}_i$. Confidence intervals, tests of hypotheses, measures of fit, subset selection techniques, and other procedures are available. For details, see McCullagh and Nelder (1989), Dobson (1990), Myers (1990, Section 7.6), Hilbe (1994), Lindsey (1997), Christensen (1997, Chapter 9), and McCulloch and Searle (2001, Chapter 5).

PROBLEMS

18.1 For the Bernoulli distribution, $P(y_i = 0) = 1 - p_i$ and $P(y_i = 1) = p_i$ in (18.4), show that $E(y_i) = p_i$ and $\text{var}(y_i) = p_i(1 - p_i)$ as in (18.5) and (18.6).

18.2 Show that $\ln [p_i/(1 - p_i)] = \beta_0 + \beta_1 x_i$ in (18.8) can be obtained from (18.7).

18.3 Verify that $\ln L(\beta_0, \beta_1)$ has the form shown in (18.10), where $L(\beta_0, \beta_1)$ is as given by (18.9).

18.4 Differentiate $\ln L(\beta_0, \beta_1)$ in (18.10) to obtain (18.11) and (18.12).

18.5 Show that $b(\theta_i) = -n \ln (1 + e^{\theta_i})$, as noted following (18.16).