9

Nonlinear Regression

9.1

Overview

In Chapter 2 we made a distinction between linear and nonlinear regression functions. Chapters 3 through 8 were concerned with inferences for, and applications of, simple and multiple *linear* regression models. Recall that in a linear regression model the regression function $\mu_Y(x_1, \ldots, x_k)$ is a *linear function of the unknown parameters,* whereas in a nonlinear regression model the regression function is *not a linear function of the unknown parameters.* In this chapter we present some commonly used nonlinear regression models, discuss point and interval estimation for unknown parameters, and provide some examples. In Section 9.2 we list several commonly used regression models. Section 9.3 gives the statistical assumptions underlying the inference procedures for nonlinear regression models and discusses parameter estimation, confidence intervals, and tests for them. In Section 9.4 we illustrate how nonlinear regression functions can sometimes be reformulated as linear regression functions by applying suitable transformations. Section 9.5 contains chapter exercises. Our discussion of nonlinear regression is limited, and we give only a brief introduction to the subject.

The computations required for nonlinear regression analyses are not feasible without the use of a computer, and most major statistical packages have routines for nonlinear regression. In the laboratory manuals we explain the use of the computer for nonlinear regression analysis.

9.2

Some Commonly Used Families of Nonlinear Regression Functions

While simple and multiple linear regression functions are adequate for modeling a wide variety of relationships between response variables and predictor variables, many situations require nonlinear functions. Certain types of nonlinear regression functions have served, and will continue to serve, as useful models for describing various physical and biological systems. We list a few of these situations for the case of a single predictor variable.

1 The following functions have been considered in modeling the relationship between crop yield Y and the spacing between rows of plants, concentration Y of a drug in the bloodstream and time X after the drug is injected when this concentration is measured, the rate Y of a chemical reaction and the amount X of catalyst used, and many other relationships.

$$\mu_Y(x) = \frac{1}{(\beta_1 + \beta_2 x)^{\beta_3}} \tag{9.2.1}$$

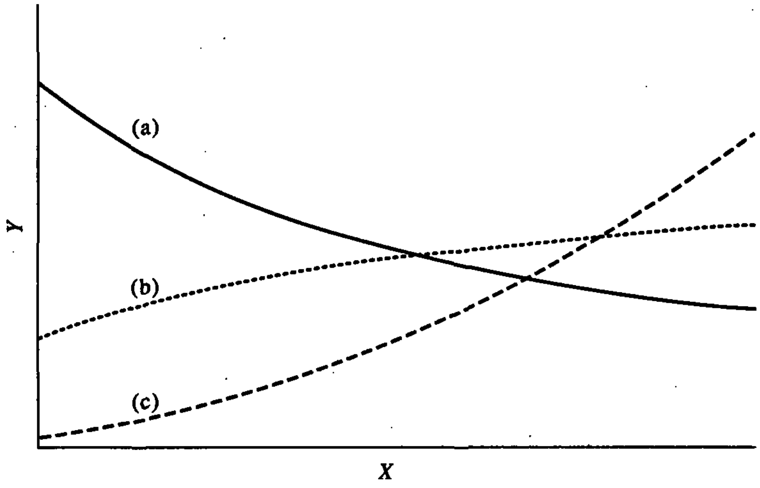
$$\mu_Y(x) = \frac{1}{\beta_1 + \beta_2 x + \beta_3 x^2} \tag{9.2.2}$$

$$\mu_Y(x) = \frac{1}{\beta_1 + \beta_2 x^{\beta_3}} \tag{9.2.3}$$

Typical members of the families of curves (9.2.1)–(9.2.3) are displayed in Figures 9.2.1–9.2.3, respectively.

FIGURE 9.2.1

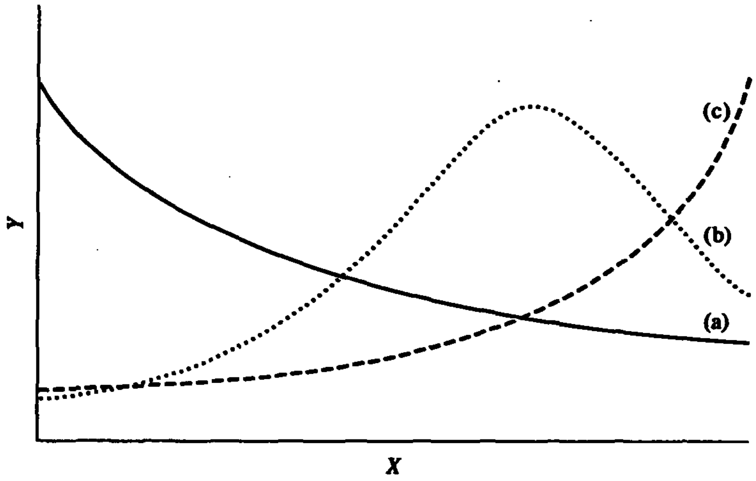
Three members of the family of curves $\mu_Y(x) = \frac{1}{(\beta_1+\beta_2 x)^{\beta_3}}$



(a) $\beta_1 = 0.6, \beta_2 = 1, \beta_3 = -1$; (b) $\beta_1 = 0.1, \beta_2 = 1, \beta_3 = 0.3$; (c) $\beta_1 = 0.2, \beta_2 = 1, \beta_3 = 2$.

FIGURE 9.2.2

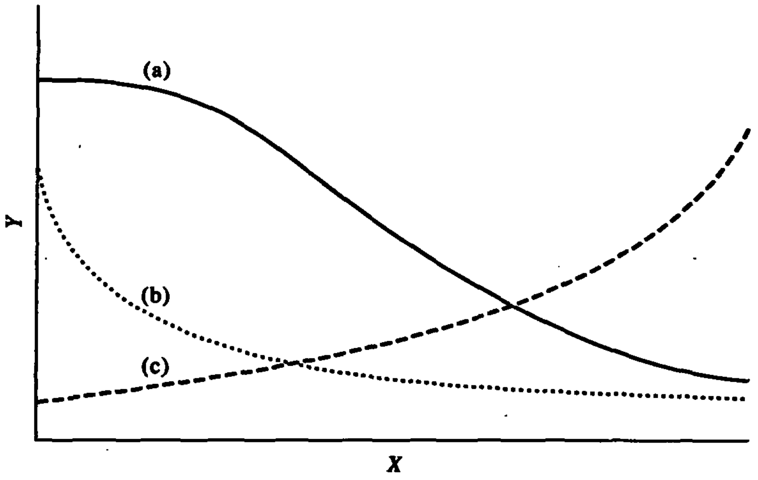
Three members of the family of curves $\mu_Y(x) = \frac{1}{\beta_1+\beta_2 x+\beta_3 x^2}$



(a) $\beta_1 = 1, \beta_2 = 3, \beta_3 = -0.2$; (b) $\beta_1 = 8.94, \beta_2 = -22.4, \beta_3 = 16$;
(c) $\beta_1 = 8, \beta_2 = -8, \beta_3 = 1$.

FIGURE 9.2.3

Three members of the family of curves $\mu_Y(x) = \frac{1}{\beta_1+\beta_2 x^{\beta_3}}$



(a) $\beta_1 = 1, \beta_2 = 6, \beta_3 = 3$; (b) $\beta_1 = 1.2, \beta_2 = 9, \beta_3 = 0.9$;
(c) $\beta_1 = 10, \beta_2 = -8.8, \beta_3 = 0.5$.

2 S-shaped curves, often referred to as sigmoidal curves, arise in various applications, including bioassay, signal detection theory, engineering, and economics. Various types of growth data often conform to sigmoidal curves. Some of the nonlinear regression functions that have been used in such situations include

$$\mu_Y(x) = \beta_1 e^{-e^{-(\beta_2+\beta_3 x)}} \tag{9.2.4}$$

$$\mu_Y(x) = \frac{\beta_1}{1 + e^{-(\beta_2+\beta_3 x)}} \tag{9.2.5}$$

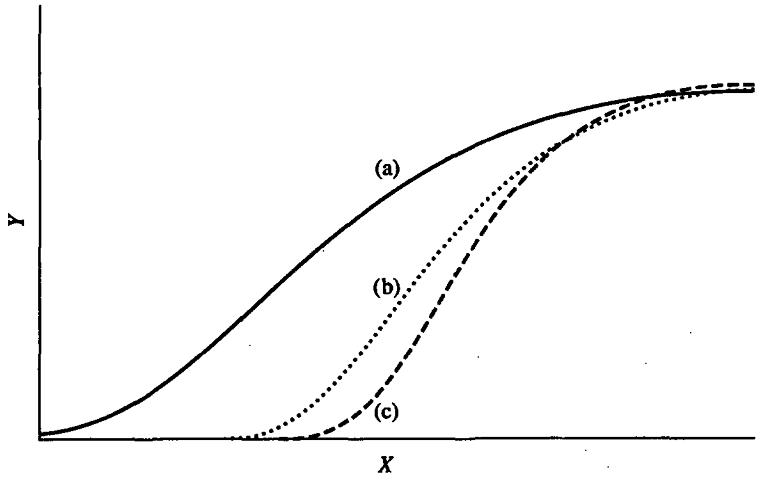
and

$$\mu_Y(x) = \frac{\beta_1}{[1 + e^{-(\beta_2+\beta_3 x)}]^{\beta_4}} \tag{9.2.6}$$

The model in (9.2.4) is often called the *Gompertz model*, the model in (9.2.5) is usually referred to as a *logistic regression model*, and the model in (9.2.6) is called *Richard's model*. Typical curves belonging to the families (9.2.4)–(9.2.6) are shown in Figures 9.2.4–9.2.6, respectively.

FIGURE 9.2.4

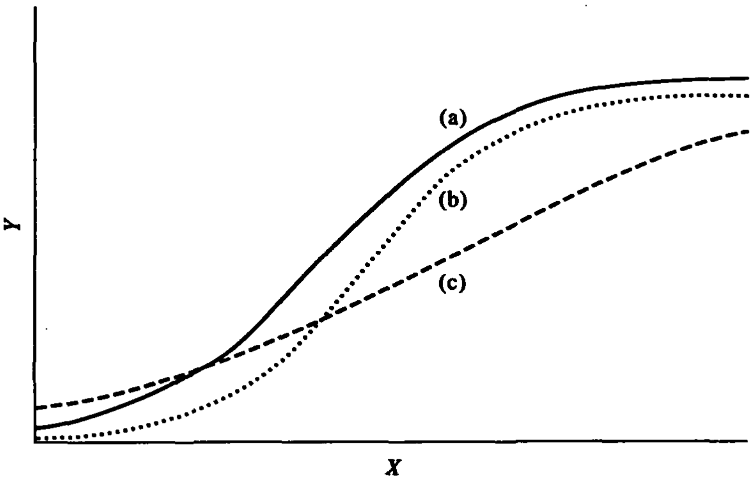
Three members of the family of curves $\mu_Y(x) = \beta_1 e^{-e^{-(\beta_2+\beta_3 x)}}$



(a) $\beta_1 = 1, \beta_2 = -1.5, \beta_3 = 5$; (b) $\beta_1 = 1, \beta_2 = -3.5, \beta_3 = 7$;
(c) $\beta_1 = 1, \beta_2 = -5, \beta_3 = 9$.

FIGURE 9.2.5

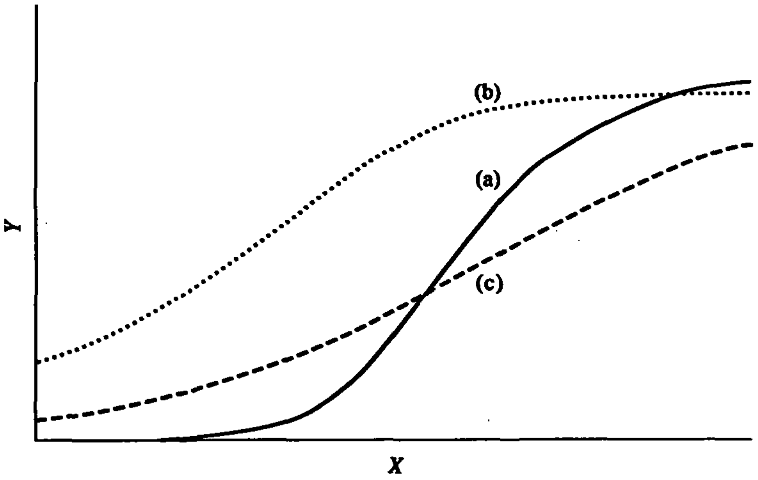
Three members of the family of curves $\mu_Y(x) = \frac{\beta_1}{1+e^{-(\beta_2+\beta_3 x)}}$



(a) $\beta_1 = 1, \beta_2 = -3.22, \beta_3 = 8$;
(b) $\beta_1 = 0.95, \beta_2 = -4.61, \beta_3 = 10$;
(c) $\beta_1 = 1, \beta_2 = -2.3, \beta_3 = 4$.

FIGURE 9.2.6

Three members of the family of curves $\mu_Y(x) = \frac{\beta_1}{[1+e^{-(\beta_2+\beta_3 x)}]^{\beta_4}}$



(a) $\beta_1 = 1, \beta_2 = -3.22, \beta_3 = 8, \beta_4 = 3.33$;
(b) $\beta_1 = 0.95, \beta_2 = -4.61, \beta_3 = 10, \beta_4 = 0.33$;
(c) $\beta_1 = 1, \beta_2 = -2.3, \beta_3 = 4, \beta_4 = 1.25$.

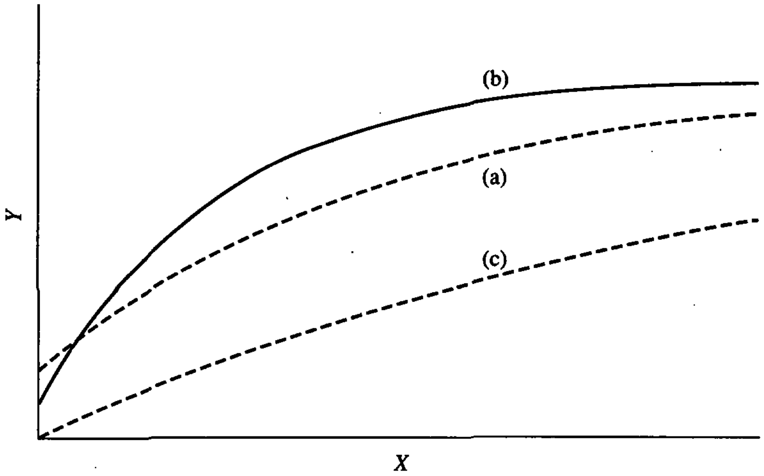
3 When the response variable Y steadily increases (decreases) with the independent variable X but the magnitude of the *rate of increase (decrease)* becomes smaller and smaller, with the response variable ultimately approaching a constant value called the *asymptote*, the family of curves defined by

$$\mu_Y(x) = \beta_1 + \beta_2 e^{-\beta_3 x} \tag{9.2.7}$$

has been found to provide useful nonlinear regression models. Three members of the family in (9.2.7) are displayed in Figure 9.2.7. Typical applications where such models are useful include the study of yield as a function of rate of application of fertilizer, mortality rate as a function of time, amount of chemical converted in a reaction as a function of time, etc.

FIGURE 9.2.7

Three members of the family of curves $\mu_Y(x) = \beta_1 + \beta_2 e^{-\beta_3 x}$



(a) $\beta_1 = 1, \beta_2 = -0.9, \beta_3 = 4$; (b) $\beta_1 = 1, \beta_2 = -0.8, \beta_3 = 2$;
(c) $\beta_1 = 1, \beta_2 = -1, \beta_3 = 0, 9$.

Numerous other useful families of nonlinear regression functions exist, but we have limited ourselves to presenting some of the simplest and the most commonly used functions. Note that although all the applications just discussed involve only a single predictor variable, the models can be extended to the case of multiple predictor variables in a variety of ways. For more information on useful nonlinear regression models, you should refer to the book by Ratkowsky [29].

In the next section we turn our attention to the statistical assumptions for nonlinear regression and procedures for point estimation, confidence intervals, and tests.

Problems 9.2

9.2.1 Find two examples from your field of specialization where the regression function of a response variable with one or more predictor variables is a *nonlinear* function of unknown parameters.

9.2.2 In Problem 9.2.1 investigate the possibility that one of the several nonlinear regression functions given in (9.2.1)–(9.2.7) might provide an appropriate model.

9.2.3 Which of the following regression functions are linear and which are nonlinear? Here β_0, β_1, β_2, and β_3 represent unknown parameters.

a $\mu_Y(x) = \beta_0 + \beta_1 e^{2x}$

b $\mu_Y(x) = \beta_0 + \beta_1 x + \beta_2 \log(x) \quad (x > 0)$

c $\mu_Y(x) = 6\beta_0 + x^{\beta_1} \quad (x > 0)$

d $\mu_Y(x) = \beta_0 + \beta_1 e^{x} + \beta_2 e^{-x}$

e $\mu_Y(x) = \beta_0 + \beta_1 e^{\beta_2 x} + \sin(\beta_3 x)$

9.3

Statistical Assumptions and Inferences for Nonlinear Regression

The most commonly used set of assumptions for nonlinear regression is the same as assumptions (A) for linear regression, the only exception being that the regression function $\mu_Y(x_1, \ldots, x_k)$ is a *nonlinear function of the unknown parameters* instead of a linear function of the parameters. For the record the complete set of assumptions is given in Box 9.3.1.

BOX 9.3.1 Assumptions (A) for Nonlinear Regression

The $(k + 1)$-variable population $\{(Y, X_1, \ldots, X_k)\}$ is the study population under investigation.

(Population) Assumption 1 The mean of the subpopulation of Y values determined by $X_1 = x_1, \ldots, X_k = x_k$ is denoted by $\mu_Y(x_1, \ldots, x_k)$, and is a *nonlinear function of unknown parameters.* At times we find it useful to write $\mu_Y(x_1, \ldots, x_k; \beta_1, \ldots, \beta_p)$ for the regression function to emphasize the fact that it depends on the parameters $\beta_1, \ldots, \beta_p$.

(Population) Assumption 2 The standard deviations of the Y values are the same for each subpopulation determined by specified values of the predictor variables $X_1, \ldots, X_k$. This common standard deviation of all the subpopulations is denoted by $\sigma_{Y|X_1,\ldots,X_k}$, but for simplicity of notation we simply write σ when there is no possibility of confusion.

(Population) Assumption 3 Each subpopulation of Y values, determined by specified values of the predictor variables $X_1, \ldots, X_k$ is Gaussian.

(Sample) Assumption 4 The sample (of size n) is selected either by simple random sampling or by sampling with preselected values of $X_1, \ldots, X_k$.

(Sample) Assumption 5 All sample values $y_i, x_{i,1}, \ldots, x_{i,k}$ for $i = 1, \ldots, n$ are observed without error.

Remarks Assumptions (A) for nonlinear regression are the same as assumptions (A) for linear regression except (population) assumption 1. Sample assumption 5 states that all sample values are observed without error. However, if only the response variable Y is measured with error, the procedures discussed in this chapter are still applicable for making inferences about the unknown parameters β_i. However, the standard error of the parameter estimates will tend to be larger and the confidence intervals will tend to be wider. *But if the predictor variables are measured with errors (that are not negligible), then the inference procedures discussed in this chapter are generally not applicable.*

Parameter Estimation

In practice, depending on the application, we assume that the form of the nonlinear regression function $\mu_Y(x_1, \ldots, x_k)$ is known but it contains unknown parameters $\beta_1, \ldots, \beta_p$. For instance, based on subject matter knowledge we may know that $\mu_Y(x)$ is of the form $\beta_1 + \beta_2 e^{\beta_3 x}$, but we may not know the values of β_1, β_2, and β_3. Observe that in the case of a linear regression function, i.e., when $\mu_Y(x_1, \ldots, x_k)$ has the form

$$\mu_Y(x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

the value of p is equal to $k + 1$ (equal to k if there is no intercept in the model). However, in nonlinear regression models, it is not always the case that p is equal to $k + 1$ or k; for the nonlinear function $\mu_Y(x) = \beta_1 + \beta_2 e^{\beta_3 x}$, we have $k = 1$ but $p = 3$.

We now discuss point estimation for the unknown parameters in a nonlinear regression function using the method of least squares.

Least Squares Estimates of $\beta_1, \ldots, \beta_p$

A popular method for estimating the unknown parameters in a nonlinear regression function is the *method of least squares*. This is described in Section 4.4 for multiple linear regression. According to this method, the estimates of $\beta_1, \ldots, \beta_p$ are obtained by minimizing the quantity $\sum_{i=1}^n e_i^2$, the sum of squares of errors of prediction, where e_i is given by

$$e_i = y_i - \mu_Y(x_{i,1}, \ldots, x_{i,k})$$

As usual, the least squares estimates of $\beta_1, \ldots, \beta_p$ are denoted by $\hat{\beta}_1, \ldots, \hat{\beta}_p$. The estimated value of the subpopulation mean $\mu_Y(x_1, \ldots, x_k)$ is denoted by

$\hat{\mu}_Y(x_1, \ldots, x_k)$. It is referred to as the **fitted value** corresponding to $x_1, \ldots, x_k$, and is obtained by substituting the least squares estimates of the parameters into the regression function. This is algebraically expressed by the equation

$$\hat{\mu}_Y(x_1, \ldots, x_k) = \mu_Y(x_1, \ldots, x_k; \hat{\beta}_1, \ldots, \hat{\beta}_p) \tag{9.3.1}$$

The quantity $\hat{e}_i$ defined by

$$\hat{e}_i = y_i - \hat{\mu}_Y(x_{i,1}, \ldots, x_{i,k}) \tag{9.3.2}$$

is called the **residual** corresponding to sample item i.

The minimum value for the sum of squares of errors of prediction corresponding *to the least squares estimates* $\hat{\beta}_1, \ldots, \hat{\beta}_p$ *is denoted by* SSE, *an abbreviation for the* more complete notation $SSE(X_1, \ldots, X_k)$. Thus

$$SSE = \sum_{i=1}^{n} \hat{e}_i^2 = \sum_{i=1}^{n} [y_i - \hat{\mu}_Y(x_{i,1}, \ldots, x_{i,k})]^2 \tag{9.3.3}$$

and, as in linear regression, we refer to SSE as the *sum of squared errors.* The quantity MSE, which is an abbreviation for the more complete notation $MSE(X_1 \ldots, X_k)$, is given by

$$MSE = \frac{SSE}{(n-p)} \tag{9.3.4}$$

and is called the *mean squared error,* and it is an unbiased estimate of σ^2. The corresponding estimate of σ is given by

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-p}} = \sqrt{MSE} \tag{9.3.5}$$

Computation of Least Squares Estimates

In the case of multiple linear regression, the least squares estimates of the parameters $\beta_1, \ldots, \beta_p$ can be computed quite easily using formula (4.4.8). However, the estimation of parameters in nonlinear regression models usually requires the use of *iterative methods* on digital computers, and explicit formulas for the estimates are generally not available. Most commonly available statistical software packages provide routines for calculating $\hat{\beta}_1, \ldots, \hat{\beta}_p$. The use of the computer for nonlinear regression analysis is discussed in the laboratory manual that accompanies the book.

To use any of these nonlinear regression programs, you must supply, in addition to the data, a set of starting values or initial guesses for $\beta_1, \ldots, \beta_p$. It is often helpful if the starting values are close to the actual least squares estimates $\hat{\beta}_i$. However, you may not have such initial guesses. Sometimes you can obtain good initial estimates of $\hat{\beta}_i$, or at least the signs of $\hat{\beta}_i$, based on theoretical considerations or by plotting the sample data.

In Example 9.3.1 we use the statistical package SAS for estimating the parameters for a specific nonlinear regression problem.

EXAMPLE 9.3.1

It is well known that when a beam of light is passed through a chemical solution, a certain fraction of the incident light will be absorbed or reflected and the remainder will be transmitted. The intensity of the transmitted light decreases as the concentration of the chemical solution increases. This fact is often used to determine concentrations of various chemicals in solutions.

In Table 9.3.1 the data are results from an experiment in which several solutions of known concentrations of pure chemical were used to measure the amount of transmitted light to determine the relationship between the optical readings Y and the concentrations X. The data are also stored in the file **light.dat** on the data disk.

TABLE 9.3.1
Light Data

Observation Number	Optical Reading Y (in arbitrary units)	Concentration X (in milligrams/liter)
1	2.86	0.0
2	2.64	0.0
3	1.57	1.0
4	1.24	1.0
5	0.45	2.0
6	1.02	2.0
7	0.65	3.0
8	0.18	3.0
9	0.15	4.0
10	0.01	4.0
11	0.04	5.0
12	0.36	5.0

Suppose assumptions (A) for nonlinear regression are satisfied with $\mu_Y(x)$ given by

$$\mu_Y(x) = \beta_1 + \beta_2 e^{-\beta_3 x} \tag{9.3.6}$$

where β_1, β_2, and β_3 are unknown parameters. On the basis of the data above, we want to estimate β_1, β_2, and β_3. Our initial guesses supplied to the computer program are $\beta_1 = 0.0$, $\beta_2 = 2.0$, and $\beta_3 = 0.5$. These initial guesses were obtained by plotting the data and interpreting geometrically the meanings of the parameters β_1, β_2, and β_3. Exhibit 9.3.1 shows the output from the statistical package SAS using a routine called NLIN (short for *Nonlin*ear regression). A plot of the data points along with the fitted curve is shown in Figure 9.3.1.

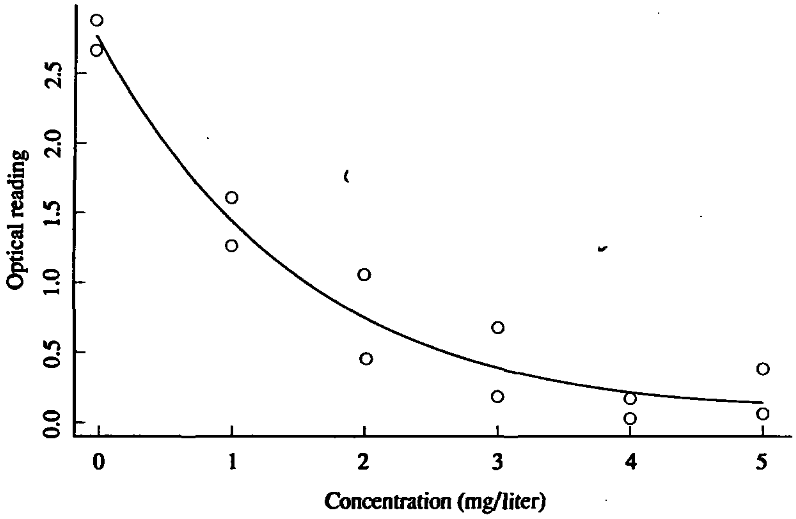
EXHIBIT 9.3.1

SAS Output for Example 9.3.1

```
                    The SAS System    0:00 Saturday,    Jan 1, 1994


Non-Linear Least Squares Summary Statistics  Dependent Variable LIGHT

Source              DF   Sum of Squares      Mean Square
Regression           3     20.542872863      6.847624288
Residual             9      0.460427137      0.051158571                  (9.3.7)

 Uncorrected Total  12     21.003300000
 (Corrected Total)  11     10.605891667

Parameter     Estimate      Asymptotic            Asymptotic 95%
                            Std. Error          Confidence Interval
                                                Lower          Upper

BETA1     0.028763192  0.17163881268     -0.3595140815  0.4170404648     (9.3.8)
BETA2     2.723273503  0.21054950823      2.2469733725  3.1995736340     (9.3.9)
BETA3     0.682773200  0.14160078051      0.3624472546  1.0030991454     (9.3.10)
```

FIGURE 9.3.1

Light Data



The final parameter estimates $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$, are given in (9.3.8)–(9.3.10). They are (after rounding to four decimals)

$$\hat{\beta}_1 = 0.0288 \qquad \hat{\beta}_2 = 2.7233 \qquad \hat{\beta}_3 = 0.6828 \qquad (9.3.11)$$

The estimate of σ^2 is given in (9.3.7) as part of an analysis of variance table along the row labeled `Residual` in the column labeled `Mean Square` and is equal to 0.0512. Thus $\hat{\sigma} = \sqrt{0.0512} = 0.2263$. ■

We now turn our attention to confidence intervals and test for nonlinear regression.

(Approximate) Confidence Intervals and Tests of Hypotheses

Exact confidence interval procedures or exact hypothesis tests are generally not available for parameters in nonlinear regression models. However, approximate inference procedures are available. In practice, the computations required for carrying out approximate hypothesis tests or obtaining approximate confidence intervals are best performed using a suitable computer program. Any computer program for calculating the estimates of parameters in a nonlinear regression function usually outputs an *approximate standard error* (ASE), sometimes also referred to as an *asymptotic standard error,* for each parameter estimate. The approximation is usually quite good if the number of observations in the sample is large. Confidence intervals for $\beta_1, \ldots, \beta_p$ may be computed using (4.6.1), and tests may be carried out according to Box 4.7.1, with approximate standard errors in place of exact standard errors. Confidence intervals and tests for $\sigma = \sigma_{Y|X_1,\ldots,X_k}$ can be computed using (4.6.13) and Box 4.7.2, respectively, without any modifications. The degrees of freedom to use for table-values are, as usual, equal to

$$n - p = n - (\text{number of } \beta \text{ parameters in the model})$$

The details are as follows. Suppose $\beta_1, \ldots, \beta_p$ are unknown parameters in a nonlinear regression function. An *approximate* $100(1-\alpha)\%$ confidence interval for β_j is given by the confidence statement

$$C[\hat{\beta}_j - t_{1-\alpha/2:n-p} ASE(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t_{1-\alpha/2:n-p} ASE(\hat{\beta}_j)] \approx 1 - \alpha \qquad \textbf{(9.3.12)}$$

where $ASE(\hat{\beta}_j)$ is the approximate standard error for $\hat{\beta}_j$. An approximate level α test of the hypothesis

$$\text{NH: } \beta_j = q \qquad \text{versus} \qquad \text{AH: } \beta_j \ne q \qquad \textbf{(9.3.13)}$$

(where q is a specified number) is conducted as follows:

$$\text{Compute } t_C = \frac{\hat{\beta}_j - q}{ASE(\hat{\beta}_j)}. \qquad \text{Reject NH if } |t_C| > t_{1-\alpha/2:n-p} \qquad \textbf{(9.3.14)}$$

One-sided tests of hypotheses are handled in the usual manner. We illustrate these computations in Example 9.3.2.

EXAMPLE 9.3.2

Here we continue with Example 9.3.1. From the computer output in Exhibit 9.3.1 (see (9.3.8)–(9.3.10) under the column labeled `Asymptotic Std. Error`) we get

$$ASE(\hat{\beta}_1) = 0.1716 \qquad ASE(\hat{\beta}_2) = 0.2105 \qquad ASE(\hat{\beta}_3) = 0.1416 \tag{9.3.15}$$

a Suppose we want to test NH: $\beta_3 = 0$ against AH: $\beta_3 \neq 0$ using $\alpha = 0.05$. Note that if β_3 is indeed zero, then $\mu_Y(x)$ does not depend on the predictor variable X. To carry out the test we calculate

$$t_C = \frac{0.6828 - 0}{0.1416} = 4.82 \tag{9.3.16}$$

Because $t_{0.975:9} = 2.262$, we reject the null hypothesis at the 5% level. The (approximate) P-value for the test is between 0 and 0.001 (using Table T-2 in Appendix T with 9 degrees of freedom).

b If we want to test NH: $\beta_2 \geq 3$ against AH: $\beta_2 < 3$ using $\alpha = 0.10$, we calculate

$$t_C = \frac{2.7233 - 3}{0.2105} = -1.314 \tag{9.3.17}$$

Because $t_{0.90:9} = 1.383$, we cannot reject the null hypothesis at $\alpha = 0.10$. The approximate P-value for this test is between 0.1 and 0.2.

c A two-sided 90% confidence interval for β_3 is $\hat{\beta}_3 \pm t_{0.95:9}$ ASE $(\hat{\beta}_3)$, leading to the confidence statement

$$C[0.4232 \leq \beta_3 \leq 0.9424] \approx 0.90 \tag{9.3.18}$$

d To compute an approximate 90% upper confidence bound for β_2 we first compute an approximate 80% two-sided confidence interval for β_2, and then use only the upper limit. We get

$$C[2.4322 \leq \beta_2 \leq 3.0144] \approx 0.80$$

from which it follows that

$$C[\beta_2 \leq 3.0144] \approx 0.90$$

e Suppose we wish to compute an approximate 95% two-sided confidence interval for σ. From (4.6.13) we have

$$C\left[\sqrt{\frac{SSE}{\chi^2_{0.975:9}}} \leq \sigma \leq \sqrt{\frac{SSE}{\chi^2_{0.025:9}}}\right] \approx 0.95$$

From (9.3.7) (under the column labeled `Sum of Squares`) we get $SSE = 0.46043$. From Table T-3 in Appendix T we obtain $\chi^2_{0.025:9} = 2.700$ and $\chi^2_{0.975:9} = 19.023$. Thus the required confidence interval is given by

$$C[0.156 \leq \sigma \leq 0.413] \approx 0.95 \quad \blacksquare$$

Note that the SAS output gives asymptotic 95% two-sided confidence intervals for each of the three parameters (see (9.3.8)–(9.3.10) under the column labeled `Asymptotic 95% Confidence Interval`).

In general, due to the approximate nature of the inference procedures for nonlinear regression problems, the actual confidence coefficients associated with the confidence intervals and P-values for tests discussed in this section may be quite different from the stated values. If critical decisions have to be made, the investigator should consult a professional statistician.

Problems 9.3

9.3.1 Consider the experiment discussed in Example 9.3.1. If we know that β_2 is nonnegative, show that no matter what the concentration of the chemical is in the solution, the average optical reading will never be less than β_1. Obtain an approximate 95% two-sided confidence interval for this parameter. Also obtain one-at-a-time 95% two-sided confidence intervals for β_2 and β_3.

9.3.2 In nondestructive testing of aluminum blocks, an electromagnetic probe is used to detect flaws below the surface. The sensitivity Y of the probe is known to be related to the thickness X of the wire used to construct the coil in the probe. An investigator interested in understanding this relationship has collected the data given in Table 9.3.2, which are also stored in the file **coil.dat** on the data disk.

TABLE 9.3.2
Coil Data

Observation Number	Sensitivity Y (a unitless quantity)	Wire Thickness X (in millimeters)
1	1.51	0.05
2	1.49	0.06
3	1.47	0.07
4	1.43	0.08
5	1.35	0.09
6	1.19	0.10
7	0.96	0.11
8	0.85	0.12
9	0.65	0.13
10	0.64	0.14
11	0.58	0.15
12	0.56	0.16
13	0.52	0.17
14	0.53	0.18
15	0.49	0.19
16	0.50	0.20

For the thickness values considered in the experiment, suppose assumptions (A) hold with the regression function of Y on X given by

$$\mu_Y(x) = \beta_1(1 - e^{-e^{-(\beta_2+\beta_3 x)}}) \tag{9.3.19}$$

Using the SAS output given in Exhibit 9.3.2, answer the following questions.

a What are the least squares estimates of the parameters β_1, β_2, β_3, and σ?

b Plot the estimated regression function $\hat{\mu}_Y(x)$ and on the same graph show the observed data points. You may do the plot manually or by using a computer program of your choice. Verify that, on the average, the sensitivity decreases as the thickness increases.

c If β_1 and β_3 are known to be positive, then show that no matter how thin the wire used, the average sensitivity can never exceed $\beta_1(1 - e^{-e^{-\beta_2}})$. Denote this upper bound for the sensitivity by θ. Estimate θ.

d Obtain one-at-a-time 95% two-sided approximate confidence intervals for β_1 and β_2. Using these estimates, obtain an approximate confidence interval for θ with confidence coefficient greater than or equal to 0.90 (use the Bonferroni method).

EXHIBIT 9.3.2

SAS Output for Problem 9.3.2

```
                  The SAS System                   0:00 Saturday,  Jan 1, 1994

Non-Linear Least Squares Summary Statistics    Dependent Variable Sensitvy

Source                     DF  Sum of Squares       Mean Square
Regression                  3    15.975543503       5.325181168
Residual                   13     0.136656497       0.010512038
Uncorrected Total          16    16.112200000
(Corrected Total)          15     2.569800000

Parameter      Estimate         Asymptotic             Asymptotic 95%
                                Std. Error          Confidence Interval
                                                    Lower          Upper
   BETA1     1.94834511       0.4724494031     0.9276808412    2.969009384
   BETA2    -1.26991572       0.6808804835    -2.740867911     0.201036258
   BETA3    14.36306754       2.7037907546     8.5218862331   20.204248838
```

9.3.3 An experiment was performed to evaluate the rate of absorption and the rate of removal of a certain drug administered intravenously on human subjects. The investigator collects blood samples at half hour intervals for six hours. The concentrations Y of the drug in the blood are obtained by a laboratory analysis of the blood samples that were taken at preselected times X after the drug was injected. The data for one particular subject are given in Table 9.3.3 and are also stored in the file **absorpt.dat** on the data disk.

TABLE 9.3.3
Drug Absorption Data

Observation Number	Concentration Y (in micrograms/deciliter)	Time X (in hours)
1	1.8	0.5
2	2.9	1.0
3	6.0	1.5
4	8.8	2.0
5	6.6	2.5
6	3.8	3.0
7	2.9	3.5
8	1.5	4.0
9	1.1	4.5
10	0.5	5.0
11	1.1	5.5
12	0.2	6.0

Suppose assumptions (A) in Box 9.3.1 for nonlinear regression are satisfied with $\mu_Y(x)$ given by

$$\mu_Y(x) = \frac{1}{\beta_1 + \beta_2 x + \beta_3 x^2} \qquad (9.3.20)$$

The coefficients β_1, β_2, and β_3 may be different for different subjects, but here we are considering the data for a single subject. Use the computer output given in Exhibit 9.3.3 to do the following.

a Obtain the least squares estimates of the parameters $\beta_1, \beta_2, \beta_3$, and σ.

b Plot Y against X and on the same graph show a plot of the estimated regression function.

c Using the sample regression function, estimate the number of hours it takes for the drug to reach peak concentration in the bloodstream. Use the graph in part (b) to verify your answer.

d Estimate the number of hours required for the body to eliminate enough of the drug so that the average drug concentration in the bloodstream falls below 1 microgram/deciliter. Use the graph in part (b) to verify your answer.

EXHIBIT 9.3.3

SAS Output for Problem 9.3.3

```
                        The SAS System          0:00 Saturday, Jan 1, 1994

Non-Linear Least Squares Summary Statistics      Dependent Variable concentr


   Source                 DF Sum of Squares    Mean Square
   Regression              3   195.23040871    65.07680290
   Residual                9     1.22959129     0.13662125
   Uncorrected Total      12   196.46000000
   (Corrected Total)      11    81.14000000

Parameter     Estimate      Asymptotic           Asymptotic 95%
                             Std Error         Confidence Interval
                                                Lower       Upper
   BETA1    0.8183215325  0.06534472402 0.67050023924 0.96614282570
   BETA2    -.6776058218  0.06162515696 -.81701279819 -.53819884543
   BETA3    0.1632703464  0.01450101285 0.13046649717 0.19607419557
```

9.4

Linearizable Models

In some situations it may be possible to transform a nonlinear regression function $\mu_Y(x)$ using appropriate transformations of the response variable, the predictor variables, the parameters, or any combination of these, such that the transformed function is linear in the unknown parameters. If the transformed variables satisfy assumptions (A) or (B) for multiple linear regression, then the results of Chapters 3 and 4 can be applied to the transformed problem. Using the results for the transformed problem, we can often obtain results for the original problem.

As an example, consider the model

$$\mu_Y(x) = \beta_1^* e^{-\beta_2^* x} \tag{9.4.1}$$

where we know that β_1^* is positive. This model is a special case of the model given in (9.2.7) with the β_1 term set to zero. By taking the logarithm to the base e of both sides, we get the transformed function $\ln[\mu_Y(x)] = \ln(\beta_1^*) - \beta_2^* x$. Now let $\ln(\beta_1^*) = \beta_0$ and $-\beta_2^* = \beta_1$. We thus have

$$\ln[\mu_Y(x)] = \beta_0 + \beta_1 x$$

which is linear in the unknown parameters. This suggests that if we set $Z = \ln(Y)$, the regression function of Z on X will be approximately linear and will be given by

$$\mu_Z(x) \approx \beta_0 + \beta_1 x \tag{9.4.2}$$

Thus we can use the theory of Chapter 3 to study this approximate regression function of Z on X and make approximate inferences about the parameters $\beta_0 = \ln(\beta_1^*)$ and $\beta_1 = -\beta_2^*$. This in turn will lead to inferences about β_1^* and β_2^*, the parameters of interest in the original problem.

More specifically, if the data are $(y_1, x_1), \ldots, (y_n, x_n)$, we let $z_i = \ln(y_i)$ and get at the transformed data $(z_1, x_1), \ldots, (z_n, x_n)$. If the investigator is confident that the transformed data satisfy assumptions (A) or (B) for straight line regression (at least approximately), then the theory of Chapter 3 can be used to draw inferences about $\mu_Z(x)$ in (9.4.2). Thus the estimates of β_0 and β_1 in (9.4.2) are those given in Chapter 3 (see (3.4.8) and (3.4.9)).

$$\hat{\beta}_1 = \frac{\sum(z_i - \bar{z})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \tag{9.4.3}$$

and

$$\hat{\beta}_0 = \bar{z} - \hat{\beta}_1\bar{x} \tag{9.4.4}$$

where $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_1$. So we get

$$\hat{\beta}_2^* = -\hat{\beta}_1 \tag{9.4.5}$$

and

$$\hat{\beta}_1^* = exp(\bar{z} - \hat{\beta}_1\bar{x}) \tag{9.4.6}$$

Some examples of linearizable models and their linear representations are given in Table 9.4.1. You are encouraged to think of other examples.

In some cases an investigator is not confident that assumptions (A) or (B) hold (even approximately) for the transformed variables. In these cases, the parameter

TABLE 9.4.1

Original Regression Function	**Linearizing Transformation**	**Suggested Transformation of Y**
$\mu_Y^*(x) = \beta_1^* e^{\beta_2^* x}$	$\ln[\mu_Y^*(x)] = \ln(\beta_1^*) + \beta_2^* x = \beta_0 + \beta_1 x$ where $\beta_0 = \ln(\beta_1^*)$ and $\beta_1 = \beta_2^*$	$Z = \ln(Y)$, *for* $Y > 0$
$\mu_Y^*(x) = (\beta_1^*)^x$	$\ln[\mu_Y^*(x)] = x\ln(\beta_1^*) = \beta_1 x$ where $\beta_1 = \ln(\beta_1^*)$	$Z = \ln(Y)$, *for* $Y > 0$
$\mu_Y^*(x) = \dfrac{1}{\beta_1^* - \beta_2^* x}$	$\dfrac{1}{\mu_Y^*(x)} = \beta_1^* - \beta_2^* x = \beta_0 + \beta_1 x$ where $\beta_0 = \beta_1^*$ and $\beta_1 = -\beta_2^*$	$Z = \dfrac{1}{Y}$, *for* $Y \neq 0$
$\mu_Y^*(x) = \dfrac{1}{1 + e^{-(\beta_1^* + \beta_2^* x)}}$	$\ln\left(\dfrac{\mu_Y^*(x)}{1 - \mu_Y^*(x)}\right) = \beta_1^* + \beta_2^* x = \beta_0 + \beta_1 x$ where $\beta_0 = \beta_1^*$ and $\beta_1 = \beta_2^*$	$Z = \ln\left(\dfrac{Y}{1 - Y}\right)$ *for* $0 < Y < 1$

estimates obtained by performing a linear regression analysis on the transformed data may be useful as starting values for nonlinear regression programs. This is a commonly used strategy. We illustrate this in Example 9.4.1.

EXAMPLE 9.4.1

A study was conducted using several subjects to determine the relationship between optical signal contrasts and visual responses in humans. Each subject was asked to view an image of a photographic slide projected on a screen and determine whether or not a specified object was present in the image. Ten different slides were used, each containing the specified object, but the optical contrast X between the object and the background was different on each slide. The ten slides were presented in random order, and each slide was seen by the subject 100 times. The response recorded was the proportion Y of times the subject reported having seen the object. Data for one particular subject are given in Table 9.4.2 and are also stored in the file **contrast.dat** on the data disk. Suppose that assumptions (A) in Box 9.3.1 are satisfied with the regression function $\mu_Y(x)$ given by

$$\mu_Y(x) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x)}} \tag{9.4.7}$$

This function is a special case of the function in (9.2.5). Although this is a nonlinear regression function, we observe that $\ln[\mu_Y(x)/(1 - \mu_Y(x))]$ is equal to $\beta_1 + \beta_2 x$, a linear function of the unknown parameters. Hence we apply the transformation $z_i = \ln[y_i/(1 - y_i)]$ to the original data, which leads to the transformed data set $(z_1, x_1), \ldots, (z_{10}, x_{10})$ given in Table 9.4.3. The transformed Y variable is denoted by Z. The regression function of Z on X should be, at least approximately, a straight line function; i.e.,

$$\mu_Z(x) \approx \beta_1 + \beta_2 x \tag{9.4.8}$$

TABLE 9.4.2

Contrast Data

Observation Numbers	Proportion of Time the Subject Saw the Object Y	Constant Between the Object and the Background X
1	0.02	0.000
2	0.06	0.005
3	0.10	0.010
4	0.18	0.015
5	0.35	0.020
6	0.56	0.025
7	0.78	0.030
8	0.86	0.035
9	0.94	0.040
10	0.99	0.045

TABLE 9.4.3

Transformed Contrast Data

Observation	Z	X
1	−3.89182	0.000
2	−2.75154	0.005
3	−2.19722	0.010
4	−1.51635	0.015
5	−0.61904	0.020
6	0.24116	0.025
7	1.26567	0.030
8	1.81529	0.035
9	2.75154	0.040
10	4.59512	0.045

We fit a straight line to the transformed data and obtain

$$\hat{\beta}_1 = -3.9627 \qquad \hat{\beta}_2 = 174.755 \tag{9.4.9}$$

The fitted straight line and the observations $\{(z_i, x_i)\}$ are displayed in Figure 9.4.1. The original data points $\{(y_i, x_i)\}$ and the estimated regression function

$$\hat{\mu}_Y(x) = \frac{1}{1+e^{-(\hat{\beta}_1+\hat{\beta}_2 x)}} = \frac{1}{1+e^{-(-3.9627+174.755x)}} \tag{9.4.10}$$

are shown in Figure 9.4.2.

FIGURE 9.4.1

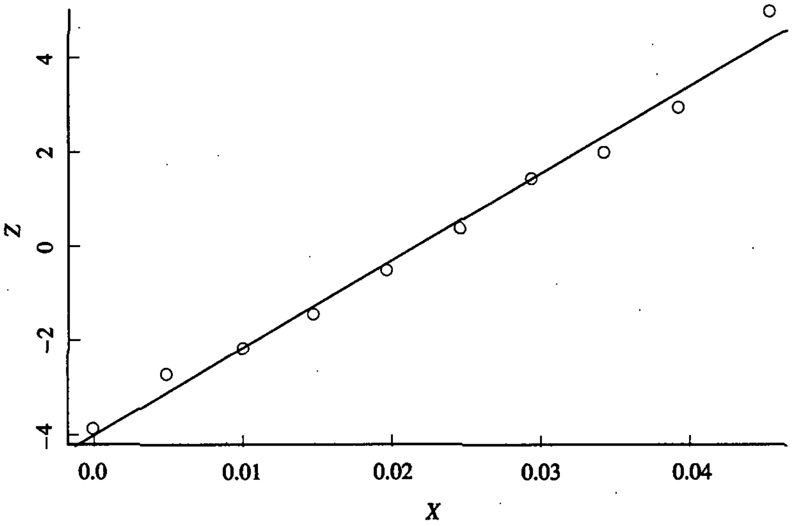
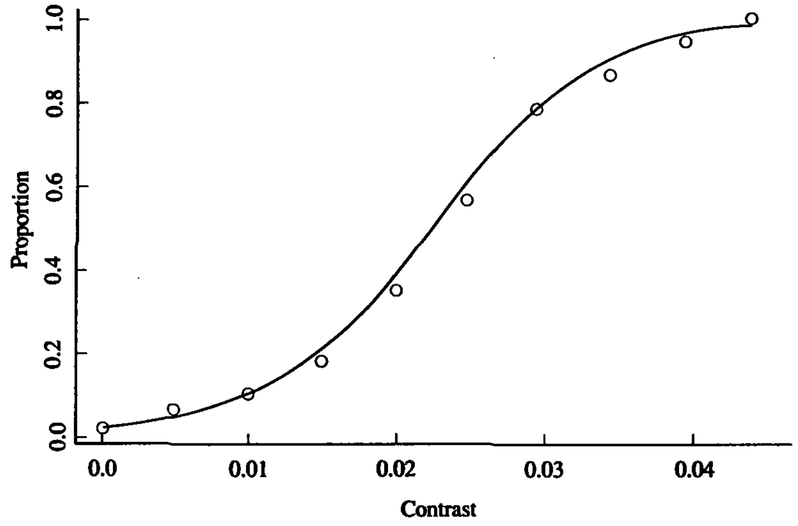


FIGURE 9.4.2



We can also estimate β_1 and β_2 directly, without using a linearizing transformation, by making use of a suitable nonlinear regression program as discussed in Section 9.3. In Exhibit 9.4.1 we present a SAS output for fitting the nonlinear regression function in (9.4.7).

EXHIBIT 9.4.1

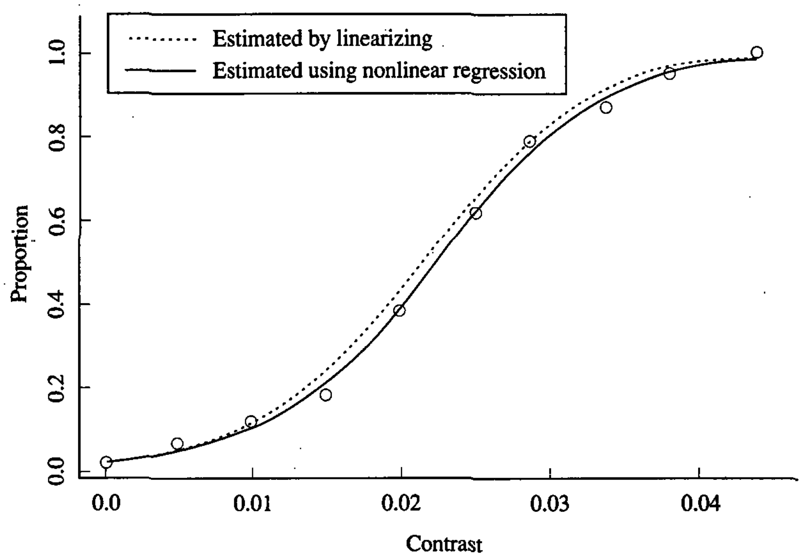
SAS Output for Example 9.4.1

```
                           The SAS System  0:00 Saturday,  Jan 1, 1994
Non-Linear Least Squares Summary Statistics  Dependent Variable Y

   Source                  DF Sum of Squares  Mean Square
   Regression               2   3.6923266027  1.8461633013
   Residual                 8   0.0018733973  0.0002341747                (9.4.11)

   Uncorrected Total       10   3.6942000000
   (Corrected Total)        9   1.3516400000

Parameter    Estimate      Asymptotic         Asymptotic 95%
                           Std. Error       Confidence Interval
                                              Lower      Upper

   B1      -4.0261321    0.1290043165       -4.3236    -3.7286           (9.4.12)
   B2     171.6643713    5.2990073124      159.4447   183.8840           (9.4.13)
```

From the *SAS* output we see that the estimates $\hat{\beta}_1 = -4.0261$ and $\hat{\beta}_2 = 171.6644$, given in (9.4.12) and (9.4.13), respectively, differ only slightly from the estimates obtained from the linearization approach. A plot of the original data points $\{(y_i, x_i)\}$ and the two estimated regression curves (one obtained by linearization and the other by nonlinear regression analysis) are displayed in Figure 9.4.3.

You may experiment with different starting values for β_1 and β_2, but you will see that the estimates obtained from the linearization approach lead to good starting values. ■

FIGURE 9.4.3



Problems 9.4

9.4.1 Consider the experiment discussed in Problem 9.3.3.

a Verify that $1/(\mu_Y(x))$ is a linear function of the unknown parameters. Use this fact to arrive at a suitable transformation of the response variable Y so that linear regression techniques may be used with the transformed data.

b Analyze the data for Problem 9.3.3 after applying the transformation in part (a) to obtain estimates of the unknown parameters β_1, β_2, and β_3. You can use the MINITAB output containing the results from a regression of $Z = 1/Y$ on X and X^2 given in Exhibit 9.4.2.

c Compare the estimates in part (b) with those obtained in Problem 9.3.3.

EXHIBIT 9.4.2

MINITAB Output for Problem 9.4.1

```
Note:  Z = 1/Y.          Y = concentr.   X = time

The regression equation is
Z = 1.40 - 1.29 X + 0.276 X^2

Predictor          Coef      Stdev   t-ratio        p
Constant         1.4026     0.7977      1.76    0.113
X               -1.2903     0.5642     -2.29    0.048
X^2             0.27640    0.08450      3.27    0.010

s = 0.7718         R-sq + 74.3%   R-sq(adj) = 68.6%

Analysis of Variance

Source      DF        SS         MS         F        p
Regression   2    15.5380     7.7690     13.04    0.002
Error        9     5.3611     0.5957
Total       11    20.8991
```

9.4.2 Consider the data of Problem 9.3.2. For illustrative purposes, suppose we know that the actual value of β_1 is 2.0, so the regression function of Y on X is

$$\mu_Y(x) = 2(1 - e^{-e^{-(\beta_2+\beta_3 x)}})$$

a Verify that $-\ln\left[-\ln\left(1 - \frac{\mu_Y(x)}{2}\right)\right]$ is linear in the unknown parameters. Use this fact to arrive at a suitable transformation of the response variable Y so that linear regression techniques may be applied to the transformed data.

b Analyze the data for Problem 9.3.2 after applying the transformation in part (a) to obtain estimates of the unknown parameters β_2 and β_3. You may use the MINITAB output containing the results from a regression of $Z = -\ln\left[-\ln\left(1 - \frac{Y}{2}\right)\right]$ on X given in Exhibit 9.4.3. Note that the regression function of Z on X is *approximately* equal to

$$\mu_Z(x) = \beta_2 + \beta_3 x$$

c Compare your answers in part (b) with those obtained in Problem 9.3.2.

EXHIBIT 9.4.3

MINITAB Output for Problem 9.4.2

```
The regression equation is
Z = -1.08 + 13.0 X

Predictor          Coef          Stdev          t-ratio          p
Constant        -1.0801         0.1274            -8.48      0.000
X               13.0413         0.9566            13.63      0.000

s = 0.1764        R-sq = 93.0%        R-sq(adj) = 92.5%

Analysis of Variance

SOURCE        DF          SS           MS            F           P
Regression     1      5.7826       5.7826       185.88       0.000
Error         14      0.4355       0.0311
Total         15      6.2181
```

9.5 Exercises

9.5.1 The concentration Y of a drug in human serum, after injection into the bloodstream, will reach a peak value in a certain amount of time X, after which it will begin to diminish with time. To understand how long after injecting a particular drug the peak concentration will be attained, and how much longer after that the concentration will return to normal levels, an experiment was conducted using human subjects. The concentration of the drug in the serum was measured immediately following injection and then after time lapses of 30 minutes, 1 hour, 2 hours, 4 hours, 8 hours, 16 hours, 24 hours, and 48 hours on each of the subjects. Table 9.5.1 gives the data for one of the subjects. These data are also stored in the file **serum.dat** on the data disk. Suppose that assumptions (A) for nonlinear regression are satisfied with the regression function of Y on X equal to

$$\mu_Y(x) = \frac{1}{\beta_1 + \beta_2 x + \beta_3 x^2} + \beta_4$$

This function is a slight generalization of the function in (9.2.2). Do the following (you can use the computer output given in Exhibit 9.5.1 whenever necessary).

a Plot the concentration against time.

b What are the estimated values of β_1, β_2, β_3, and β_4?

c Display the fitted curve in the graph of part (a).

d Using the fitted curve, estimate the time it takes for the drug concentration to reach a maximum value for this particular subject.

e Using the fitted curve, estimate the time it takes for the drug concentration to drop below 50 milligrams/liter for this particular subject.

TABLE 9.5.1

Serum Data

Observation Number	Drug Concentration in Serum (in mg/liter) Y	Time after Injection (in hours) X
1	67	0.0
2	70	0.5
3	82	1.0
4	85	2.0
5	104	4.0
6	129	8.0
7	171	16.0
8	162	24.0
9	18	48.0

EXHIBIT 9.5.1

SAS Output for Problem 9.5.1

```
                           The SAS System   0:00 Saturday, Jan 1, 1994
Non-Linear Least Squares Summary Statistics   Dependent Variable Y

   Source                 DF  Sum of Squares    Mean Square

   Regression              4    106518.91654    26629.72914
   Residual                5        85.08346       17.01669
   Uncorrected Total       9    106604.00000
   (Corrected Total)       8     18988.00000

Parameter     Estimate        Asymptotic          Asymptotic 95%
                               Std Error        Confidence Interval
                                                Lower          Upper
   BETA1    0.00656892      0.001475260     0.00277670    0.0103611407
   BETA2   -0.00028246      0.000112385    -0.00057135    0.0000064315
   BETA3    0.00000749      0.000003193    -0.00000071    0.0000157026
   BETA4  -80.65213231     34.151101655  -168.43906150    7.1347968761
```

9.5.2 An investigator conducted a laboratory study to evaluate the ability of human observers to detect the presence of layered haze in the atmosphere. She generated several photographic slides of a scenic vista (Hopi Point, Grand Canyon) with various levels of layered haze artificially superimposed using computer imagery, causing different levels of degradation in the visibility of the scene. Subjects were shown the projected images on a screen and were asked to respond whether or not they saw any layered haze. There were eight levels of haze quantified by optical contrast values of 0, 0.005, 0.008, 0.012, 0.015, 0.020, 0.030, and 0.040, respectively. The contrast value of 0 corresponds to a slide of the scene with no superimposed haze. The slides were shown in a randomized order. Each subject saw each slide 100 times during the experiment. The data pertaining to one of the subjects are listed in Table 9.5.2 and are also stored in the file **haze.dat** on the data disk.

TABLE 9.5.2
Haze Data

Observation Number	Proportion of Times Subject Reported Presence of Layered Haze Y	Contrast Value of Slide X
1	0.34	0.000
2	0.44	0.005
3	0.50	0.008
4	0.59	0.012
5	0.72	0.015
6	0.93	0.020
7	0.99	0.030
8	1.00	0.040

Suppose that the regression function of Y on X is given by

$$\mu_Y(x) = \frac{\beta_1}{[1 + e^{-(\beta_2+\beta_3 x)}]^{\beta_4}}$$

and that assumptions (A) for nonlinear regression are satisfied. Do the following (you can use the computer output given in Exhibit 9.5.2 wherever necessary).

a Plot Y versus X.

b What are the estimated values of β_1, β_2, β_3, and β_4? What is the estimated value of the subpopulation standard deviation σ?

c Display the fitted curve along with the sample data in a graph.

d Using the fitted curve, estimate the contrast value x_0 such that for values of X greater than or equal to x_0, this subject will report seeing the haze at least 90% of the time; i.e., $\mu_Y(x_0) \geq 0.90$. *Hint:* Locate the value 0.90 along the Y axis and find the corresponding value x_0 along the X axis using the fitted curve. You can also do this algebraically by solving the following equation for x_0:

$$0.90 = \frac{\hat{\beta}_1}{[1 + e^{-(\hat{\beta}_2+\hat{\beta}_3 x_0)}]^{\hat{\beta}_4}}$$

e Using the fitted curve, estimate the proportion of affirmative responses by the subject even when shown the slide with no superimposed haze. (This is referred to as false alarm rate.) Also construct an approximate 85% confidence interval for this subject's false alarm rate. (*Hint:* First compute an approximate 85% simultaneous confidence interval for β_1, β_2, and β_4 using the Bonferroni method.)

EXHIBIT 9.5.2

SAS Output for Problem 9.5.2

```
                                The SAS System   0:00 Saturday, Jan 1, 1994

Non-Linear Least Squares Summary Statistics   Dependent Variable Y

   Source                  DF Sum of Squares    Mean Square

   Regression               4   4.2698528454      1.0674632113
   Residual                 4   0.0008471546      0.0002117887
   Uncorrected Total        8   4.2707000000
   (Corrected Total)        7   0.4756875000

Parameter     Estimate      Asymptotic          Asymptotic 95%
                             Std Error       Confidence Interval
                                             Lower         Upper

   BETA1       0.995000      0.01029050     0.9664292     1.0235704
   BETA2    -115.929282     10.75382750  -145.7862897   -86.0722743
   BETA3    5402.347473    571.24393936     38 414938  6988.3534528
   BETA4       0.009437      0.00077196     0.0072935     0.0115800
```