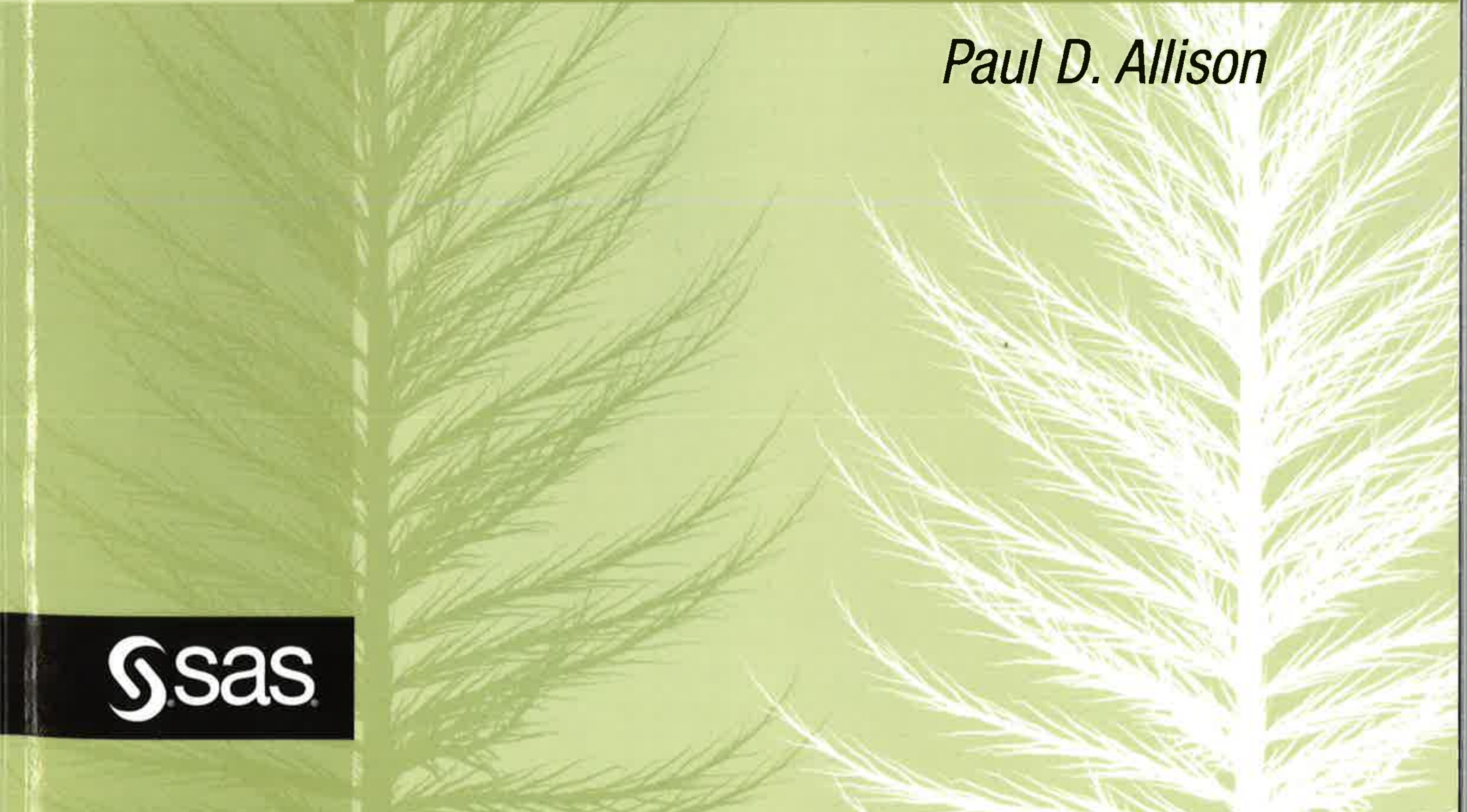
SAS® Press

Logistic Regression Using SAS®

Theory and Application

Second Edition

Paul D. Allison



Chapter 2

Binary Logistic Regression with PROC LOGISTIC: Basics

2.1 Introduction

A great many variables in the social sciences are dichotomous—employed vs. unemployed, married vs. unmarried, guilty vs. not guilty, voted vs. didn't vote. It's hardly surprising, then, that social scientists frequently want to estimate regression models in which the dependent variable is a dichotomy. Nowadays, most researchers are aware that there's something wrong with using ordinary linear regression for a dichotomous dependent variable, and that it's better to use logistic or probit regression. But many of them don't know what it is about linear regression that makes dichotomous variables problematic, and they may have only a vague notion of why other methods are superior.

In this chapter, we focus on logistic regression (a.k.a. logit analysis) as an optimal method for the regression analysis of dichotomous (binary) dependent variables. Along the way, we'll see that logistic regression has many similarities to ordinary linear regression analysis. To understand and appreciate the logistic model, we first need to see why ordinary linear regression runs into problems when the dependent variable is dichotomous.

2.2 Dichotomous Dependent Variables: Example

To make things tangible, let's start with an example. Throughout this chapter, we'll be examining a data set consisting of 147 death penalty cases in the state of New Jersey. In all of these cases, the defendant was convicted of first-degree murder with a recommendation by the prosecutor that a death sentence be imposed. Then a penalty trial was conducted to determine whether the defendant would receive a sentence of death or life imprisonment. Our dependent variable DEATH is coded 1 for a death sentence, and 0 for a life sentence. The aim is to determine how this outcome was influenced by various characteristics of the defendant and the crime.

Many potential independent variables are available in the data set, but let's consider three of special interest:

BLACKD	Coded 1 if the defendant was black, otherwise 0.
WHITVIC	Coded 1 if the victim was white, otherwise 0.
SERIOUS	A rating of the seriousness of the crime, as evaluated by a panel of judges.

The variable SERIOUS was developed in an auxiliary study in which panels of trial judges were given written descriptions of each of the crimes in the original data set. These descriptions did not mention the race of the defendant or the victim. Each judge evaluated 14 or 15 cases and ranked them from least serious to most serious. Each case was ranked by four to six judges. As used in this chapter, the SERIOUS score is the average ranking given to each case, ranging from 1 (least serious) to 15 (most serious).

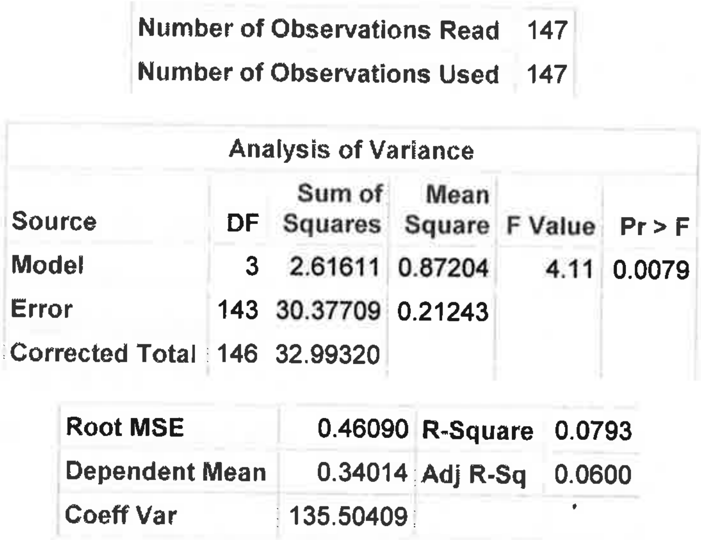
Using the REG procedure in SAS, I estimated a linear regression model that uses DEATH as the dependent variable and the other three as independent variables. The SAS code is

```
PROC REG DATA=penalty;
  MODEL death=blackd whitvic serious;
RUN;
```

Results are shown in Output 2.1. Neither of the two racial variables have coefficients that are significantly different from 0. Not surprisingly, the coefficient for SERIOUS is highly significant—more serious crimes are more likely to get the death penalty.

Should we trust these results, or should we ignore them because the statistical technique is incorrect? To answer that question we need to see *why* linear regression is regarded as inappropriate when the dependent variable is a dichotomy. That's the task of the next section.

Output 2.1 Linear Regression of Death Penalty on Selected Independent Variables

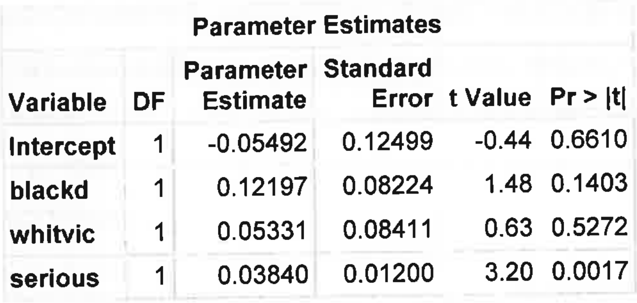


Number of Observations Read	147
Number of Observations Used	147

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.61611	0.87204	4.11	0.0079
Error	143	30.37709	0.21243		
Corrected Total	146	32.99320			

Root MSE	0.46090	R-Square	0.0793
Dependent Mean	0.34014	Adj R-Sq	0.0600
Coeff Var	135.50409		



Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.05492	0.12499	-0.44	0.6610
blackd	1	0.12197	0.08224	1.48	0.1403
whitvic	1	0.05331	0.08411	0.63	0.5272
serious	1	0.03840	0.01200	3.20	0.0017

2.3 Problems with Ordinary Linear Regression

Not so long ago, it was common to see published research that used ordinary least squares (OLS) linear regression to analyze dichotomous dependent variables. Some people didn't know any better. Others knew better, but didn't have access to good software for alternative methods. Now, every major statistical package includes a procedure for logistic regression, so there's no excuse for applying inferior methods. No reputable social science journal would publish an article that used OLS regression with a dichotomous dependent variable.

Should all the earlier literature that violated this prohibition be dismissed? Actually, most applications of OLS regression to dichotomous variables give results that are qualitatively quite similar to results obtained using logistic regression. There are exceptions, of course, so I certainly wouldn't claim that there's no need for logistic regression. But as an approximate method, OLS linear regression does a surprisingly good job with dichotomous variables, despite clear-cut violations of assumptions.

What are the assumptions that underlie OLS regression? While there's no single set of assumptions that justifies linear regression, the list in the box below is fairly standard. To keep things simple, I've included only a single independent variable *x*, and I've presumed that *x* is "fixed" across repeated samples (which means that every sample has the same set of *x* values). The *i* subscript distinguishes different members of the sample.

Assumptions of the Linear Regression Model

1. $y_i = \alpha + \beta x_i + \varepsilon_i$
2. $\mathrm{E}(\varepsilon_i) = 0$
3. $\mathrm{var}(\varepsilon_i) = \sigma^2$
4. $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$
5. $\varepsilon_i \sim \mathit{Normal}$

Assumption 1 says that y is a linear function of x plus a random disturbance term ε, for all members of the sample. The remaining assumptions all say something about the distribution of ε. What's important about assumption 2 is that $\mathrm{E}(\varepsilon)$ (the expected value of ε) does *not* vary with x, implying that x and ε are uncorrelated. Assumption 3, often called the *homoscedasticity* assumption, says that the variance of ε is the same for all observations. Assumption 4 says that the random disturbance for one observation is uncorrelated with the random disturbance for any other observation. Finally, assumption 5 says that the random disturbance is normally distributed. If all five assumptions are satisfied, ordinary least squares estimates of α and β are unbiased and have minimum sampling variance (minimum variability across repeated samples).

Now suppose that y is a dichotomy with possible values of 1 or 0. It's still reasonable to claim that assumptions 1, 2, and 4 are true. But if 1 and 2 are true for a dichotomy, then 3 and 5 are *necessarily* false. First, let's consider assumption 5. Suppose that y_i=1. Then assumption 1 implies that $\varepsilon_i=1-\alpha-\beta x_i$. On the other hand, if y_i=0, we have $\varepsilon_i=-\alpha-\beta x_i$. Because ε_i can only take on two values, it's impossible for it to have a normal distribution (which has a continuum of values and no upper or lower bound). So assumption 5 must be rejected.

To evaluate assumption 3, it's helpful to do a little preliminary algebra. The expected value of y_i is, by definition,

$$E(y_i) = 1 \times \Pr(y_i = 1) + 0 \times \Pr(y_i = 0).$$

If we define $p_i = \Pr(y_i=1)$, this reduces to

$$E(y_i) = p_i$$

In general, for any dummy variable, its expected value is just the probability that it is equal to 1. But assumptions 1 and 2 also imply another expression for this expectation. Taking the expected values of both sides of the equation in assumption 1, we get

$$\begin{aligned} E(y_i) &= E(\alpha + \beta x_i + \varepsilon_i) \\ &= E(\alpha) + E(\beta x_i) + E(\varepsilon_i) \\ &= \alpha + \beta x_i . \end{aligned}$$

Putting these two results together, we get

$$p_i = \alpha + \beta x_i , \tag{2.1}$$

which is sometimes called the *linear probability model.* As the name suggests, this model says that the probability that y=1 is a linear function of x. Regression coefficients have a straightforward interpretation under this model. A 1-unit change in x produces a change of β in the probability that y=1. In Output 2.1, the coefficient for SERIOUS was .038. So we can say that each 1-point increase in the SERIOUS scale (which ranges from 1 to 15) is associated with an increase of .038 in the probability of a death sentence, controlling for the other variables in the model. The BLACKD coefficient of .12 tells us that the estimated probability of a death sentence for black defendants is .12 higher than for nonblack defendants, controlling for other variables.

Now let's consider the variance of ε_i. Because x is treated as fixed, the variance of ε_i is the same as the variance of y_i. In general, the variance of a dummy variable is $p_i(1-p_i)$. Therefore, we have

$$\text{var}(\varepsilon_i) = p_i(1-p_i) = (\alpha + \beta x_i)(1 - \alpha - \beta x_i).$$

We see, then, that the variance of ε_i must be different for different observations and, in particular, it varies as a function of x. The disturbance variance is at a maximum when p_i=.5 and gets small when p_i is near 1 or 0.

We've just shown that a dichotomous dependent variable in a linear regression model necessarily violates assumptions of homoscedasticity (assumption 3) and normality (assumption 5) of the error term. What are the consequences? Not as serious as you might think. First of all, we don't need these assumptions to get *unbiased* estimates. If just assumptions 1 and 2 hold, ordinary least squares will produce unbiased estimates of α and β. Second, the normality assumption is not needed if the sample is reasonably large. The central limit theorem assures us that coefficient estimates will have a distribution that is approximately normal even when ε is *not* normally distributed. That means that we can still use a normal table to calculate p-values and confidence intervals. If the sample is small, however, these approximations could be poor.

Violation of the homoscedasticity assumption has two undesirable consequences. First, the coefficient estimates are no longer *efficient*. In statistical terminology, this means that there are alternative methods of estimation with smaller standard errors. Second, and more serious, the standard error estimates are no longer *consistent* estimates of the true standard errors. That means that the estimated standard errors could be biased (either upward or downward) to unknown degrees. And because the standard errors are used in calculating test statistics, the test statistics could also be problematic.

Fortunately, the potential problems with standard errors and test statistics are easily fixed. Beginning with SAS 9.2, PROC REG offers a *heteroscedasticity consistent covariance estimator* that uses the method of Huber (1967) and White (1980), sometimes known as the "sandwich" method because of the structure of the matrix formula. This method produces consistent estimates of the standard errors even when the homoscedasticity assumption is violated. To implement the method in PROC REG, simply put the option HCC on the MODEL statement:

```
PROC REG DATA=penalty;
  MODEL death=blackd whitvic serious / HCC;
RUN;
```

Now, in addition to the output in Output 2.1, we get the corrected standard errors, t-statistics, and p-values shown in Output 2.2. In this case, the correction for heteroscedasticity makes almost no difference in the results.

Output 2.2 Linear Regression of Death Penalty with Correction for Heteroscedasticity

Parameter Estimates

						Heteroscedasticity Consistent		
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	-0.05492	0.12499	-0.44	0.6610	0.11959	-0.46	0.6468
blackd	1	0.12197	0.08224	1.48	0.1403	0.08197	1.49	0.1390
whitvic	1	0.05331	0.08411	0.63	0.5272	0.08315	0.64	0.5224
serious	1	0.03840	0.01200	3.20	0.0017	0.01140	3.37	0.0010

Although the HCC standard errors are an easy fix, be aware that they have inherently more sampling variability than conventional standard errors (Kauermann and Carroll 2001), and may be especially unreliable in small samples. For large samples, however, they should be quite satisfactory.

In addition to these technical difficulties, there is a more fundamental problem with the assumptions of the linear model. We've seen that when the dependent variable is a dichotomy, assumptions 1 and 2 imply the linear probability model

$$p_i = \alpha + \beta x_i$$

While there's nothing intrinsically wrong with this model, it's a bit implausible, especially if x is measured on a continuum. If x has no upper or lower bound, then for any value of β there are values of x for which p_i is either greater than 1 or less than 0. In fact, when estimating a linear probability model by OLS, it's quite common for predicted values generated by the model to be outside the (0, 1) interval. (That wasn't a problem with the regression in Output 2.1, which implied predicted probabilities ranging from .03 to .65.) Of course, it's impossible for the true values (which are probabilities) to be greater than 1 or less than 0. So the only way the model could be true is if a ceiling and floor are somehow imposed on p_i, leading to considerable awkwardness both theoretically and computationally.

These problems with the linear model led statisticians to develop alternative approaches that make more sense conceptually and also have better statistical properties. The most popular of these approaches is the logistic model, which is estimated by maximum likelihood. Before considering the full model, let's examine one of its key components—the *odds* of an event.

2.4 Odds and Odds Ratios

To appreciate the logistic model, it's helpful to have an understanding of *odds* and *odds ratios.* Most people regard probability as the "natural" way to quantify the chances that an event will occur. We automatically think in terms of numbers ranging from 0 to 1, with a 0 meaning that the event will certainly not occur and a 1 meaning that the event certainly will occur. But there are other ways of representing the chances of event, one of which—the odds—has a nearly equal claim to being "natural."

Widely used by professional gamblers, the odds of an event is the ratio of the expected number of times that an event will occur to the expected number of times it will not occur. An odds of 4 means we expect 4 times as many occurrences as non-occurrences. An odds of 1/5 means that we expect only one-fifth as many occurrences as non-occurrences. In gambling circles, odds are sometimes expressed as, say, "5 to 2," but that corresponds to the single number 5/2.

There is a simple relationship between probabilities and odds. If p is the probability of an event and O is the odds of the event, then

$$O = \frac{p}{1-p} = \frac{probability\ of\ event}{probability\ of\ no\ event}$$

$$p = \frac{O}{1+O}. \tag{2.2}$$

This functional relationship is illustrated in Table 2.1.

Table 2.1 Relationship between Odds and Probability

Probability	Odds
.1	.11
.2	.25
.3	.43
.4	.67
.5	1.00
.6	1.50
.7	2.33
.8	4.00
.9	9.00

Note that odds less than 1 correspond to probabilities below .5, while odds greater than 1 correspond to probabilities greater than .5. Like probabilities, odds have a lower bound of 0. But unlike probabilities, there is no upper bound on the odds.

Why do we need the odds? Because it's a more sensible scale for multiplicative comparisons. If I have a probability of .30 of voting in the next election, and your probability of voting is .60, it's meaningful to claim that your probability is twice as great as mine. But if my probability is .60, it's impossible for your probability to be twice as great. There's no problem on the odds scale, however. A probability of .60 corresponds to odds of .60/.40=1.5. Doubling that yields odds of 3. Converting back to probabilities gives us 3/(1+3)=.75.

This leads us to the odds ratio, which is a widely used measure of the relationship between two dichotomous variables. Consider Table 2.2, which shows the cross-tabulation of race of defendant by death sentence for the 147 penalty-trial cases. The numbers in the table are the actual numbers of cases that have the stated characteristics.

Table 2.2 Death Sentence by Race of Defendant for 147 Penalty Trials

	Blacks	Non-blacks	Total
Death	28	22	50
Life	45	52	97
Total	73	74	147

Overall, the estimated odds of a death sentence are 50/97= .52. For blacks, the odds are 28/45 = .62. For nonblacks, the odds are 22/52 = .42. The ratio of the black odds to the nonblack odds is 1.47. We may say, then, that the odds of a death sentence for blacks are 47% greater than for nonblacks. Note that the odds ratio in a 2 × 2 table is also equal to the *cross-product ratio*, which is the product of the two main-diagonal frequencies divided by the product of the two off-diagonal frequencies. In this case, we have (52 × 28)/(22 × 45) = 1.47.

Of course, we can also say that the odds of a death sentence for nonblacks are 1/1.47 = .63 times the odds of a death sentence for blacks. Similarly, the odds of a *life sentence* for blacks are .63 times the odds for nonblacks. So, depending on which categories we're comparing, we either get an odds ratio greater than 1 or its reciprocal, which is less than 1.

Implicit in much of the contemporary literature on categorical data analysis is the notion that odds ratios (and various functions of them) are less sensitive to changes in the marginal frequencies (for example, the total number of death and life sentences) than other measures of association. In this sense, they are frequently regarded as fundamental descriptions of the relationship between the variables of interest. As we shall see, odds ratios are directly related to the parameters in the logistic regression model.

2.5 The Logistic Regression Model

Now we're ready to introduce the *logistic regression model*, otherwise known as the *logit model*. As we discussed earlier, a major problem with the linear probability model is that probabilities are bounded by 0 and 1, but linear functions are inherently unbounded. The solution is to transform the probability so that it's no longer bounded.

Transforming the probability to an odds removes the upper bound. If we then take the logarithm of the odds, we also remove the lower bound. Setting the result equal to a linear function of the explanatory variables, we get the logistic model. For *k* explanatory variables and *i* = 1,…, *n* individuals, the model is

$$\log\left[\frac{p_i}{1-p_i}\right] = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} \qquad (2.3)$$

where p_i is, as before, the probability that y_i=1. The expression on the left-hand side is usually referred to as the *logit* or *log-odds*. (Natural logarithms are used throughout this book. However, the only consequence of switching to base-10 logarithms would be to change the intercept α.) As in ordinary linear regression, the *x*'s may be either quantitative variables or dummy (indicator) variables.

Unlike the usual linear regression model, there is no random disturbance term in the equation for the logistic model. That doesn't mean that the model is deterministic because there is still room for random variation in the probabilistic relationship between p_i and y_i. Nevertheless, as we shall see later, problems may arise if there is unobserved heterogeneity in the sample.

We can solve the logit equation for p_i to obtain

$$p_i = \frac{\exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})}{1 + \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik})} \qquad (2.4)$$

Exp(x) is the exponential function, equivalent to e^x. In turn, e is the exponential constant, approximately equal to 2.71828. Its defining property is that log(e^x)=x. We can simplify further by dividing both the numerator and denominator by the numerator itself:

$$p_i = \frac{1}{1+\exp(-\alpha - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_k x_{ik})} \tag{2.5}$$

This equation has the desired property that no matter what values we substitute for the β's and the x's, p_i will always be a number between 0 and 1.

If we have a single x variable with $\alpha = 0$ and $\beta = 1$, the equation can be graphed to produce the S-shaped curve in Figure 2.1. As x gets large or small, p gets close to 1 or 0 but is never equal to these limits. From the graph, we see that the effect of a unit change in x depends on where you start. When p is near .50, the effect is large; but when p is near 0 or 1, the effect is small. More specifically, the slope of the curve is given by the derivative of p_i with respect to the covariate x_i.

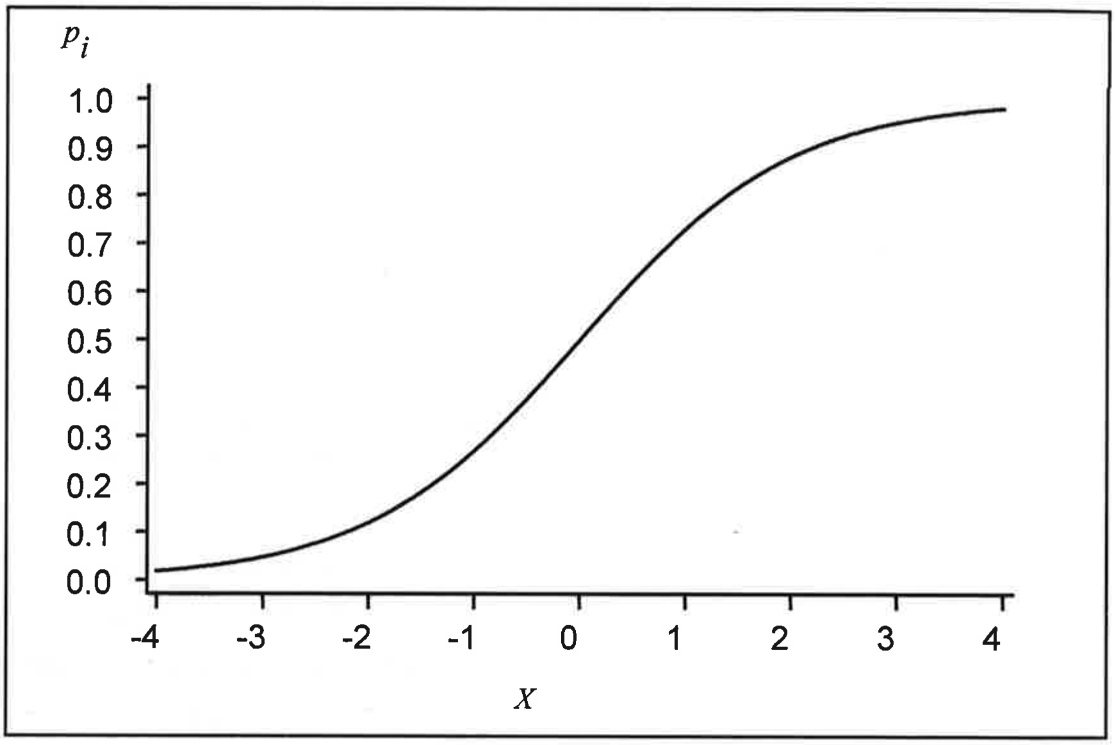
$$\frac{\partial p_i}{\partial x_i} = \beta p_i (1 - p_i). \tag{2.6}$$

This is known as the "marginal effect" of x on the event probability, which will be discussed in Section 2.8. When β=1 and p=.5, a 1-unit increase in x produces an increase of .25 in the probability. When β is larger, the slope of the S-shaped curve at p=.5 is steeper. When β is negative, the curve is flipped horizontally so that p is near 1 when x is small and near 0 when x is large. The derivative in equation (2.6) also applies when there is more than one x variable, although then it's a partial derivative.

There are alternative models that have similar S-shaped curves, most notably the probit and complementary log-log models. I'll discuss them briefly in the next chapter. But, for several reasons, the logistic model is more popular:

- The coefficients have a simple interpretation in terms of odds ratios.
- The logistic model is intimately related to the loglinear model, which is discussed in Chapter 10.
- The logistic model has desirable sampling properties, which are discussed in Section 3.13.
- The model can be easily generalized to allow for multiple, unordered categories for the dependent variable.

Figure 2.1 Graph of Logistic Model for a Single Explanatory Variable



2.6 Estimation of the Logistic Model: General Principles

Now that we have a model for dichotomous dependent variables, the next step is to use sample data to estimate the coefficients. How that's done depends on the type of data you're working with. If you have *grouped data*, there are three readily available methods: ordinary least squares, weighted least squares, and maximum likelihood.

Grouped data occurs when the explanatory variables are all discrete and the data is arrayed in the form of a contingency table. We'll see several examples of grouped data in Chapter 4. Grouped data can also occur when data are collected from naturally occurring groups. For example, suppose that the units of analysis are business firms and the dependent variable is the probability that an employee is a full-time worker. Let P_i be the observed proportion of employees who work full-time in firm *i*. To estimate a logistic model by OLS, we could simply take the logit transformation of *P,* which is log[*P*/(1-*P*)], and regress the result on characteristics of the firm and on the average characteristics of the employees. A weighted least squares (WLS) analysis would be similar, except that the data would be weighted to adjust for heteroscedasticity. The SAS procedure CATMOD does WLS estimation for grouped data (in addition to maximum likelihood).

Maximum likelihood (ML) is the third method for estimating the logistic model for grouped data and the *only* method in general use for *individual-level data.* With individual-level data, we simply observe a dichotomous dependent variable for each individual along with measured characteristics of the individual. OLS and WLS can't be used with this kind of data unless the data can be grouped in some way. If y_i can only have values of 1 and 0, it's impossible to apply the logit transformation—you get either minus infinity or plus infinity. To put it another way, any transformation of a dichotomy is still a dichotomy.

Maximum likelihood is a very general approach to estimation that is widely used for all sorts of statistical models. You may have encountered it before with loglinear models, latent variable models, or event history models. There are two reasons for this popularity. First, ML estimators are known to have good properties in large samples. Under fairly general conditions, ML estimators are consistent, asymptotically efficient, and asymptotically normal. Consistency means that as the sample size gets larger the probability that the estimate is within some small distance of the true value also gets larger. No matter how small the distance or how high the specified probability, there is always a sample size that yields an even higher probability that the estimator is within that distance of the true value. One implication of consistency is that the ML estimator is approximately unbiased in large samples. Asymptotic efficiency means that, in large samples, the estimates will have standard errors that are, approximately, at least as small as those for any other estimation method. And, finally, the sampling distribution of the estimates will be approximately normal in large samples, which means that you can use the normal and chi-square distributions to compute confidence intervals and *p*-values.

All these approximations get better as the sample size gets larger. The fact that these desirable properties have only been proven for large samples does *not* mean that ML has bad properties for small samples. It simply means that we usually don't *know* exactly what the small-sample properties are. And in the absence of attractive alternatives, researchers routinely use ML estimation for both large and small samples. Although I won't argue against that practice, I do urge caution in interpreting *p*-values and confidence intervals when samples are small. Despite the temptation to accept *larger p*-values as evidence against the null hypothesis in small samples, it is actually more reasonable to demand *smaller* values to compensate for the fact that the approximation to the normal or chi-square distributions may be poor.

The other reason for ML's popularity is that it is often straightforward to derive ML estimators when there are no other obvious candidates. One case that ML handles very nicely is data with categorical dependent variables.

The basic principle of ML is to choose as estimates those parameter values that, if true, would maximize the probability of observing what we have, in fact, observed. There are two steps to this: (1) write down an expression for the probability of the data as a function of the unknown parameters, and (2) find the values of the unknown parameters that make the value of this expression as large as possible.

The first step is known as constructing the *likelihood function*. To accomplish this you must specify a model, which amounts to choosing a probability distribution for the dependent variable and choosing a functional form that relates the parameters of this distribution to the values of the explanatory variables. In the case of the logistic model, the dichotomous dependent variable is presumed to have a binomial distribution with a single "trial" and a probability of "success" given by p_i. Then p_i is assumed to depend on the explanatory variables according to equation (2.3), which is the logistic model. Finally, we assume that the observations are independent across individuals.

The second step—maximization—typically requires an iterative numerical method, which means that it involves successive approximations. Such methods are usually more computationally demanding than a non-iterative method like ordinary least squares. For those who are interested, I will work through the basic mathematics of constructing and maximizing the likelihood function in Chapter 3. Here I focus on some of the practical aspects of ML estimation with SAS.

2.7 Maximum Likelihood Estimation with PROC LOGISTIC

The most popular SAS procedure for doing ML estimation of the logistic regression model is PROC LOGISTIC. SAS has several other procedures that will also do this, and we will meet some of them in later chapters. But LOGISTIC has been around the longest, and it has the largest set of features that are likely to be used by most data analysts.

Let's estimate a logistic model analogous to the linear probability model that we examined in Section 2.2. Minimal SAS code for this model is

```
PROC LOGISTIC DATA=penalty;
  MODEL death(EVENT='1')=blackd whitvic serious;
RUN;
```

Of course, there are also numerous options and special features that we'll consider later.

One option that I've specified in the MODEL statement is EVENT='1', after the dependent variable. The default in LOGISTIC is to estimate a model predicting the *lowest* value of the dependent variable. Consequently, if I had omitted EVENT='1', the result would be a logistic model predicting the probability that the dependent variable DEATH is equal to 0. The EVENT='1' option reverses this so that the model predicts the probability that the dependent variable is equal to 1. (The single quotes around 1 are necessary because PROC LOGISTIC treats the dependent variable as a character variable rather than as a numeric variable.)

An equivalent (and popular) way to accomplish this is to use the option DEATH(DESCENDING), which tells LOGISTIC to model the "higher" value of DEATH rather than the lower. But what is considered higher rather than lower can depend on other options that are chosen, so it's safer to be explicit about which value of the dependent variable is to be modeled. If you forget the EVENT='1' option, the only consequence is to change the signs of the coefficients. As long as you realize what you've done, you shouldn't need to rerun the model.

Results are shown in Output 2.3. The "Model Information" table is pretty straightforward, except for the "Optimization Technique," which is reported as Fisher's scoring. This is the numerical method used to maximize the likelihood function, and we will see how it works in Chapter 3. After the "Response Profile" table, we see the message "Probability modeled is death=1." If we had not used the EVENT='1' option, this would have said "death=0", so it's important to check this so that you correctly interpret the signs of the coefficients. Next we are told that the convergence criterion was satisfied, which is a good thing. Iterative methods don't always converge, and we'll talk about why that happens and how to deal with it in Chapter 3.

The next table reports three different "Model Fit Statistics:" AIC, SC, and -2 Log L. Values of these fit statistics are displayed for two different models, a model with an intercept but no covariates (predictors), and a model that includes all the specified predictors (covariates). Usually, we can ignore the INTERCEPT ONLY column. The most fundamental of the fit statistics, -2 Log L, is simply the maximized value of the logarithm of the likelihood function multiplied by –2. We'll see how this number is obtained in Chapter 3. Higher values of -2 Log L mean a worse fit to the data. But keep in mind that the overall magnitude of this statistic is heavily dependent on the number of observations. Furthermore, there is no absolute standard for what's a good fit, so one can only use this statistic to compare different models fit to the same data set.

The problem with -2 Log L is that models with more covariates tend to fit better by chance alone. The other two fit statistics avoid this problem by penalizing models that have more covariates. Akaike's Information Criterion (AIC) is calculated as

$$AIC = -2\log L + 2k$$

where k is the number of parameters (including the intercept). So, in this example, there are four parameters, which adds 8 to the -2 Log L.

The Schwarz Criterion (SC), also known as the Bayesian Information Criterion (BIC), gives a more severe penalization for additional parameters:

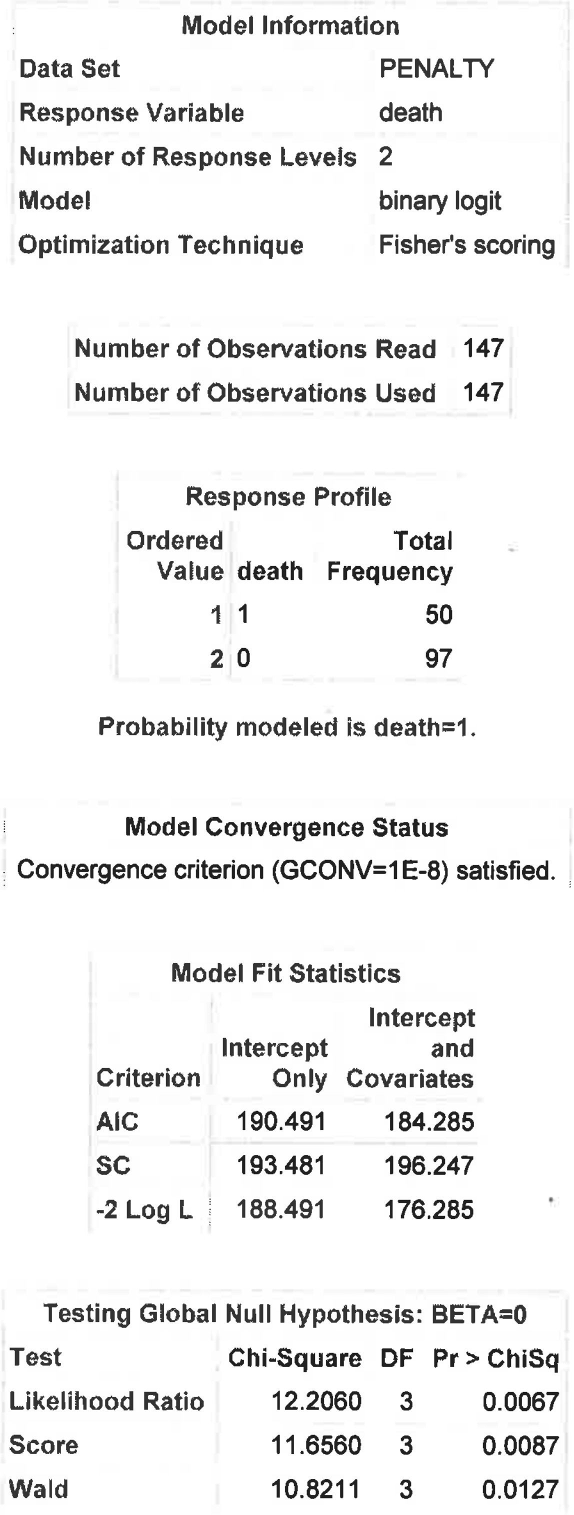
$$SC = -2\log L + k\log n$$

where n is the sample size. In this example, n = 147 and log n = 4.99. So, to get the SC value, we add 4 times 4.99 = 19.96 to the -2 Log L.

Both of the penalized statistics can be used to compare models with different sets of covariates. The models being compared do not have to be nested in the sense of one model being a special case of another. However, these statistics cannot be used to construct a formal hypothesis test, so the comparison is only informal.

The next table is "Testing Global Null Hypothesis: BETA=0." Within this table there are three chi-square statistics with values of 12.206, 11.656, and 10.8211. All three statistics are testing the same null hypothesis—that *all* the explanatory variables have coefficients of 0. (In a linear regression, this hypothesis is usually tested by means of an overall *F*-test.) The three degrees of freedom for each statistic correspond to the three coefficients for the independent variables. In this case, the associated *p*-values are around .01, so we can reject the null hypothesis and conclude that at least one of the coefficients is not 0.

Output 2.3 PROC LOGISTIC Output for Death Penalty Data



```
              Model Information
Data Set                      PENALTY
Response Variable             death
Number of Response Levels     2
Model                         binary logit
Optimization Technique        Fisher's scoring

    Number of Observations Read   147
    Number of Observations Used   147

          Response Profile
      Ordered               Total
        Value  death    Frequency
            1  1               50
            2  0               97

      Probability modeled is death=1.

           Model Convergence Status
 Convergence criterion (GCONV=1E-8) satisfied.

          Model Fit Statistics
                             Intercept
                 Intercept         and
      Criterion       Only  Covariates
      AIC          190.491     184.285
      SC           193.481     196.247
      -2 Log L     188.491     176.285

     Testing Global Null Hypothesis: BETA=0
Test              Chi-Square   DF   Pr > ChiSq
Likelihood Ratio     12.2060    3       0.0067
Score                11.6560    3       0.0087
Wald                 10.8211    3       0.0127
```

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-2.6516	0.6748	15.4424	<.0001
blackd	1	0.5952	0.3939	2.2827	0.1308
whitvic	1	0.2565	0.4002	0.4107	0.5216
serious	1	0.1871	0.0612	9.3342	0.0022

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
blackd	1.813	0.838	3.925
whitvic	1.292	0.590	2.832
serious	1.206	1.069	1.359

Association of Predicted Probabilities and Observed Responses

Percent Concordant	67.2	Somers' D	0.349
Percent Discordant	32.3	Gamma	0.351
Percent Tied	0.5	Tau-a	0.158
Pairs	4850	c	0.675

Why do we need three chi-square statistics? The first one is the *likelihood ratio chi-square* obtained by comparing the log-likelihood for the fitted model with the log-likelihood for a model with *no* explanatory variables (intercept only). It is calculated by taking twice the positive difference in the two log-likelihoods. In fact, LOGISTIC reports –2 Log L for each of those models, and the chi-square is just the difference between those two numbers. The *score* statistic is a function of the first and second derivatives of the log-likelihood function under the null hypothesis. The Wald statistic is a function of the coefficients and their covariance matrix. In large samples, there's no reason to prefer any one of these statistics, and they will generally be quite close in value. In small samples or samples with extreme data patterns, there is some evidence that the likelihood ratio chi-square is superior (Jennings 1986), especially when compared with the Wald test.

Next we get to the heart of the matter—the "Analysis of Maximum Likelihood Estimates." As with linear regression, we get coefficient estimates, their estimated standard errors, and test-statistics for the null hypotheses that each coefficient is equal to 0. The test statistics are labeled “Wald Chi-Square.” They are calculated by dividing each coefficient by its standard error and squaring the result. If we omitted the squaring operation (as many software packages do), we could call them *z* statistics, and they would have a standard normal distribution under the null hypothesis. In that case, the *p*-values calculated from a normal table would be exactly the same as the chi-square *p*-values reported here.

We see that the variable SERIOUS has a highly significant coefficient, while the coefficient for the variable WHITVIC is clearly not significant. With a *p*-value of .13, BLACKD is approaching conventional significance levels but doesn’t quite make it. Now compare these *p*-values with those in Output 2.1 for an ordinary linear regression analysis. While not identical, they are remarkably similar. In this case at least, logistic regression and OLS regression lead us to exactly the same qualitative conclusions. It’s more difficult to compare coefficient estimates across the two methods, however, and I won’t attempt that in this chapter.

The odds ratios in the next table are obtained by simply exponentiating the coefficients in the first column, that is, calculating exp(β). They are very important in the interpretation of logistic regression coefficients, and we will discuss that interpretation in the next section. The 95 percent confidence intervals are obtained as follows. First, we get 95 percent confidence intervals around the original β coefficients in the usual way. That is, we add and subtract 1.96 standard errors. To get confidence intervals around the odds ratios, we exponentiate those upper and lower confidence limits.

The last section of Output 2.3, labeled “Association of Predicted Probabilities and Observed Responses,” is an attempt to measure the explanatory power of the model. I’ll discuss these statistics in Section 3.7.

2.8 Interpreting Coefficients

When logistic regression first became popular, a major complaint by those who resisted its advance was that the coefficients had no intuitive meaning. Admittedly, they’re not as easy to interpret as coefficients in the linear probability model. For the linear probability model, a coefficient of .25 tells you that the predicted probability of the event increases by .25 for every 1-unit increase in the explanatory variable. By contrast, a logit coefficient of .25 tells you that the log-odds increases by .25 for every 1-unit increase in the explanatory variable. But who knows what a .25 increase in the log-odds means?

The basic problem is that the logistic model assumes a nonlinear relationship between the probability and the explanatory variables, as shown in Figure 2.1. The change in the probability for a 1-unit increase in an independent variable varies according to where you start. Things become much simpler, however, if we think in terms of odds rather than probabilities.

In Output 2.3, we saw the estimated β coefficients and their associated statistics in the "Analysis of Maximum Likelihood Estimates" table. Except for their sign, they *are* hard to interpret. Let's look instead at the numbers in the "Odds Ratio Estimates" table which, in this example, are obtained from the parameter estimates by computing e^{β}. (When there are CLASS variables in the model, the odds ratios may be computed differently, depending on the parameterization). These might be better described as *adjusted* odds ratios because they control for other variables in the model. Recall that BLACKD has a value of 1 for black defendants and 0 for everyone else. The odds ratio of 1.813 tells us that the predicted odds of a death sentence for black defendants are 1.813 times the odds for nonblack defendants. In other words, the odds of a death sentence for black defendants are 81% *higher* than the odds for other defendants. This compares with an *unadjusted* odds ratio of 1.47 found in Table 2.2. Although the adjusted odds ratio for BLACKD is not statistically significant (the 95% confidence interval includes 1, corresponding to no effect), it is still our best estimate of the effect of this variable.

For the dummy variable WHITVIC, which indicates white victim, the odds ratio is 1.292. This implies that the predicted odds of death are about 29% higher when the victim is white compared to the odds when the victim is not white. Of course, the coefficient is far from statistically significant so we wouldn't want to put much confidence in this value. What about the coefficient for the variable SERIOUS, which *is* statistically significant at the .01 level? Recall that this variable is measured on a 15-point scale. For quantitative variables, it's helpful to subtract 1 from the odds ratio and multiply by 100, that is, calculate $100(e^{\beta}-1)$. This tells us the *percent change* in the odds for each 1-unit increase in the independent variable. In this case, we find that a 1-unit increase in the SERIOUS scale is associated with a 21% increase in the predicted odds of a death sentence. Note that if a β coefficient is significantly different from 0, then the corresponding odds ratio is significantly different from 1. There is no need for a separate test for the odds ratio.

Interpretation of coefficients in terms of odds ratios is certainly the easiest way to approach the logistic model. On the other hand, odds ratios can sometimes be misleading if the probabilities are near 1 or 0. Suppose that in a wealthy suburban high school, the probability of graduation is .99, which corresponds to odds of 99. When financial aid is increased for needy students, the probability of graduation goes up to .995, which implies odds of 199. Apparently, the odds of graduation have more than *doubled* under the new program even though only half a percent more students are graduating. Is this a meaningful increase? That depends on many nonstatistical issues.

For those who insist on interpreting logistic models in terms of probabilities, there are several graphical and tabular methods available (Long 1997). Perhaps the simplest approach is to make use of equation (2.6):

$$\frac{\partial p_i}{\partial x_i} = \beta p_i(1 - p_i).$$

This equation says that the change in the probability for a 1-unit increase in *x* depends on the logistic regression coefficient for *x,* as well as on the value of the probability itself. For this to be practically useful, we need to know what probability we are starting from. If we have to choose one value, the most natural is the overall proportion of cases that have the event. In our example, 50 out of 147 defendants got the death penalty, so the overall proportion is .34. Taking .34 times 1–.34, we get .224. We can multiply each of the coefficients in Output 2.3 by .224, and we get:

BLACKD	.133
WHITVIC	.057
SERIOUS	.046

We can then say that, on average, the probability of a death sentence is .133 higher if the defendant is black compared with nonblacks, .057 higher if the victim is white compared with non-white, and .046 higher for a 1-unit increase on the SERIOUS scale. These are sometimes called "marginal effects," and they are of considerable interest in some fields. Of course, these numbers only give a rough indication of what actually happens for a given change in the *x* variable. Note, however, that they are very similar to the coefficients obtained with the OLS regression in Output 2.1.

Instead of choosing a single value for p_i, we can calculate a predicted probability for each individual, using equation (2.5). Then we can use the derivative formula to generate marginal effects for each individual. You can easily get them with PROC QLIM (part of the

SAS/ETS product) using the OUTPUT statement with the MARGINAL option. Here's the code:

```
PROC QLIM DATA=penalty;
  ENDOGENOUS death~DISCRETE(DIST=LOGISTIC);
  MODEL death = blackd whitvic serious;
  OUTPUT OUT=a MARGINAL;
PROC PRINT DATA=a(OBS=10);
  VAR meff_p2_blackd meff_p2_whitvic meff_p2_serious;
RUN;
```

QLIM reports coefficients, standard errors, and test statistics (not shown) that are identical to those produced by PROC LOGISTIC. PROC PRINT produces the table shown in Output 2.4, for the first 10 observations in the output data set. For each variable, we get the predicted change in the probability of the death penalty for a 1-unit increase in that variable, for a particular individual based on that individual's predicted probability.

Output 2.4 Marginal Effects Produced by PROC QLIM for the First 10 Cases

Obs	Meff_P2_blackd	Meff_P2_whitvic	Meff_P2_serious
1	0.13068	0.056312	0.041070
2	0.14660	0.063172	0.046073
3	0.08712	0.037541	0.027380
4	0.12615	0.054358	0.039645
5	0.14692	0.063309	0.046173
6	0.10523	0.045347	0.033072
7	0.09613	0.041422	0.030210
8	0.14880	0.064118	0.046763
9	0.12982	0.055942	0.040800
10	0.13682	0.058957	0.042999

2.9 CLASS Variables

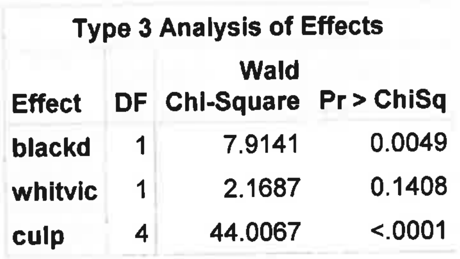
As with several other SAS regression procedures, PROC LOGISTIC has a CLASS statement that allows you to specify that a variable should be treated as categorical (nominal). When a CLASS variable is included as an explanatory variable in the MODEL statement, LOGISTIC automatically creates a set of "design variables" to represent the levels of the CLASS variable. Keep in mind that when a predictor variable is an indicator (dummy) variable, like BLACKD or WHITVIC which only have values of 0 or 1, there is no need to declare it to be a CLASS variable. In fact, putting an indicator variable on the CLASS statement can produce misleading results. That's because the CLASS statement might recode the variable in unexpected ways, as we will see. So the CLASS statement should be reserved for categorical variables with more than two categories, or for dichotomous variables that have character values like "yes" and "no."

Here's an example with the death-penalty data. The data set contains the variable CULP, which has the integer values 1 to 5 (5 denotes high culpability and 1 denotes low culpability, based on a large number of aggravating and mitigating circumstances defined by statute). Although we could treat this variable as an interval scale, we might prefer to treat it as a set of categories. To do this, we run the following program:

```
PROC LOGISTIC DATA=penalty;
  CLASS culp /PARAM=REF;
  MODEL death(EVENT='1') = blackd whitvic culp ;
RUN;
```

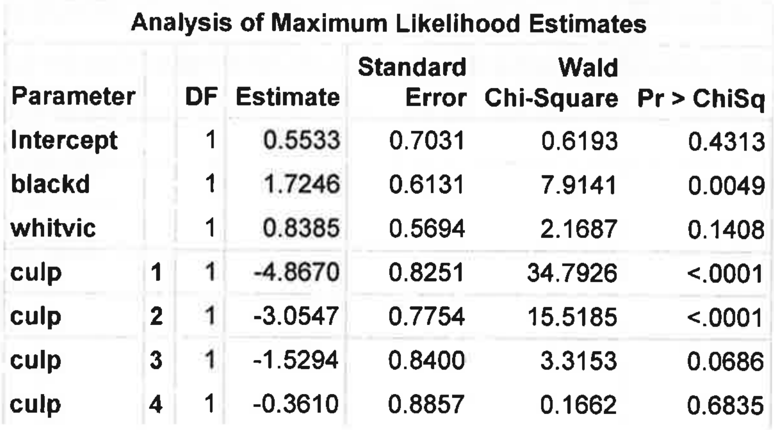
The CLASS statement declares CULP to be a classification variable. You can have more than one variable on the CLASS statement. The PARAM=REF option tells LOGISTIC to create a set of four dummy (indicator) variables, one for each value of CULP except the highest one (CULP=5). Results are shown in Output 2.5.

Output 2.5 Use of a CLASS Variable in PROC LOGISTIC



Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
blackd	1	7.9141	0.0049
whitvic	1	2.1687	0.1408
culp	4	44.0067	<.0001



Analysis of Maximum Likelihood Estimates

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	0.5533	0.7031	0.6193	0.4313
blackd		1	1.7246	0.6131	7.9141	0.0049
whitvic		1	0.8385	0.5694	2.1687	0.1408
culp	1	1	-4.8670	0.8251	34.7926	<.0001
culp	2	1	-3.0547	0.7754	15.5185	<.0001
culp	3	1	-1.5294	0.8400	3.3153	0.0686
culp	4	1	-0.3610	0.8857	0.1662	0.6835

Because the variable CULP has five possible values, LOGISTIC has created four dummy variables, one for each of the values 1 through 4. As in other procedures that have CLASS variables, the default is to take the highest value as the omitted category. We'll see how to change that in a moment. Thus, each of the four coefficients for CULP is a comparison between that particular value and the highest value. More specifically, each coefficient can be interpreted as the log-odds of the death penalty for that particular value of CULP minus the log-odds for CULP=5, controlling for other variables in the model.

When you have a CLASS variable in a model, LOGISTIC provides an additional table, labeled "Type 3 Analysis of Effects." For variables that are *not* CLASS variables, this table is completely redundant with the standard table below it—the chi-squares and *p*-values are exactly the same. For CLASS variables, on the other hand, it gives us something very useful: a test of the null hypothesis that *all* of the coefficients pertaining to this variable are 0. In other words, it gives us a test of whether CULP has any impact on the probability of the death penalty. In this case, we clearly have strong evidence that CULP makes a difference. What's particularly attractive about this test is that it is invariant to the choice of the omitted category, or even to the choice among very different methods for constructing design variables (which I'll discuss in a moment).

The pattern for the four coefficients is just what one might expect. Defendants with CULP=1 are much less likely to get the death sentence than those with CULP=5. Each increase of CULP is associated with an increase in the probability of a death sentence. Notice also that when CULP is included in the model, the coefficient for BLACKD (black defendant) is much larger than it was in Output 2.3 and is now statistically significant.

If you don't like the default reference category (5 in this example), how do you change it? If you want it to be the lowest value of CULP rather than the highest value, just use

```
CLASS culp / PARAM=REF DESCENDING;
```

If you want it to be some particular value, say 3, use

```
CLASS culp(REF='3') / PARAM=REF;
```

Suppose you want to compare two categories of a CLASS variable, such as CULP=2 and CULP=3. You could always accomplish this by rerunning the model, and making one of those values the reference category. But it's usually easier to use a TEST statement or a CONTRAST statement, and you can have as many of those as you want. The following two statements test the same null hypothesis, that there is no difference in the coefficients for CULP=2 and CULP=3:

```
CONTRAST 'Culp2 vs. Culp3' culp 0 1 -1 0;
TEST culp2=culp3;
```

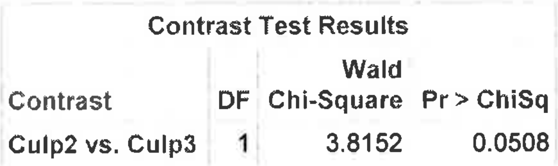
The CONTRAST statement requires a label, which can be any text enclosed by quotes. This is annoying if you only have a single CONTRAST, but very useful if you have more than one because the label helps you distinguish the different tests in the output. Here's how the CONTRAST statement works: we know that CULP has four coefficients. The code "CULP 0 1 -1 0" tells SAS to multiply the first coefficient by 0, the second coefficient by 1, the third coefficient by -1, and the fourth coefficient by 0, add up the results, and test whether the sum is equal to 0. Of course, this is equivalent to testing whether there is a difference between the second and third coefficients.

The TEST statement is a little more straightforward. To refer to a coefficient, you just append the value of the variable to the variable name itself. A TEST statement may optionally have a label, but the label comes before TEST and cannot have any spaces. Thus, we could have written

```
Culp2_vs_Culp3: TEST culp2=culp3;
```

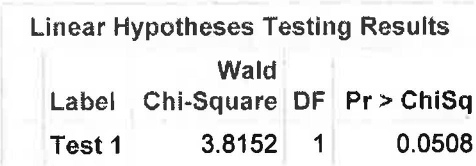
The output from the CONTRAST and TEST statements is nearly identical (see Output 2.6). In this case, the difference in the log-odds of a death penalty for those in category 2 vs. those in category 3 is not quite significant at the .05 level. Note that this difference (and its statistical significance) is invariant to the choice of the reference category.

Output 2.6 Results from TEST and CONTRAST Statements



Contrast Test Results

Contrast	DF	Wald Chi-Square	Pr > ChiSq
Culp2 vs. Culp3	1	3.8152	0.0508



Linear Hypotheses Testing Results

Label	Wald Chi-Square	DF	Pr > ChiSq
Test 1	3.8152	1	0.0508

What happens if you leave off the PARAM=REF option? Unfortunately, the default for the CLASS statement in PROC LOGISTIC is quite different from the default in many other regression procedures, such as GLM, GENMOD, PHREG, or LIFEREG. This could lead some analysts to misinterpret the results. In those procedures, the default is to produce a set of dummy variables, much like we did using the PARAM=REF option. The default in LOGISTIC, on the other hand, is to produce design variables that are sometimes described as analysis of variance coding or effect coding.

Output 2.7 shows what you get when you simply write CLASS CULP without using the PARAM option. The "Class Level Information" table tells us how the four design variables are constructed. Clearly these are not indicator variables, because each one can takes on values of 1, 0, or -1. The first design variable has a valuę of 1 when CULP=1, a value of -1 when CULP=5, and values of 0 for the other values of CULP. The other three are constructed in a similar fashion, except that the CULP value that is assigned a 1 changes for each design variable.

Output 2.7 Results from CLASS Statement Using Default Design Variables

```
   Class Level Information

                  Design
Class  Value     Variables
culp   1        1   0   0   0
       2        0   1   0   0
       3        0   0   1   0
       4        0   0   0   1
       5       -1  -1  -1  -1
```

```
    Type 3 Analysis of Effects

                    Wald
Effect   DF   Chi-Square   Pr > ChiSq
blackd    1       7.9141       0.0049
whitvic   1       2.1687       0.1408
culp      4      44.0067       <.0001
```

```
              Analysis of Maximum Likelihood Estimates

                                Standard        Wald
Parameter     DF   Estimate        Error  Chi-Square   Pr > ChiSq
Intercept      1    -1.4092       0.6066      5.3972       0.0202
blackd         1     1.7246       0.6131      7.9141       0.0049
whitvic        1     0.8385       0.5694      2.1687       0.1408
culp      1    1    -2.9046       0.5007     33.6535       <.0001
culp      2    1    -1.0923       0.4645      5.5304       0.0187
culp      3    1     0.4330       0.5292      0.6696       0.4132
culp      4    1     1.6014       0.6026      7.0625       0.0079
```

```
          Odds Ratio Estimates

                  Point        95% Wald
Effect         Estimate   Confidence Limits
blackd            5.610      1.687     18.657
whitvic           2.313      0.758      7.061
culp   1 vs 5     0.008      0.002      0.039
culp   2 vs 5     0.047      0.010      0.215
culp   3 vs 5     0.217      0.042      1.124
culp   4 vs 5     0.697      0.123      3.954
```

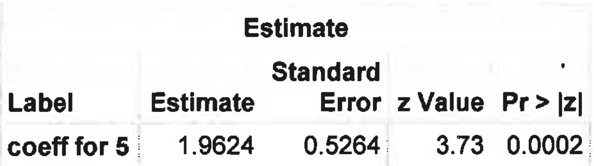
As can be seen in the “Type 3 Analysis of Effects” table in Output 2.7, this alternative coding has no effect on the overall test for CULP. And skipping down to the “Odds Ratio Effects” table, it also has no effect here either. Just as when we use PARAM=REF, each odds ratio compares a particular category with category 5. But there are certainly major changes in the CULP coefficients reported in the “Analysis of Maximum Likelihood Estimates” table, along with their standard errors and *p*-values. What’s important to understand is that these coefficients are *not* comparisons between each of the first four values of CULP and value 5. Rather they are comparisons between each category of CULP and the overall average (roughly speaking) of the log-odds of getting the death penalty, adjusting for other variables in the model. Thus, if you are in category 1 of CULP, your log-odds of getting the death penalty is 2.9046 *below* average. If you are in category 4, your log-odds is 1.6014 *above* average.

What about category 5? How does it differ from the average? To get that value, you must add together the coefficients for values 1 through 4 and then change the sign. That ensures that all five coefficients sum to zero. You could do that by hand calculation, but it’s a lot easier to let SAS do it with the ESTIMATE statement:

```
ESTIMATE 'coeff for 5' culp -1 -1 -1 -1;
```

This says to take the four coefficients for CULP, multiply each by -1, and then add them together. Results are shown in Output 2.8. We see that those with value 5 of CULP have a death penalty log-odds that is 1.9624 higher than average. We also get a standard error and a *z*-test of the null hypothesis that this effect is 0.

Output 2.8 Results from ESTIMATE Statement



Estimate

Label	Estimate	Standard Error	z Value	Pr > |z|
coeff for 5	1.9624	0.5264	3.73	0.0002

Some people like this approach to creating design variables, but I’m not a fan. I prefer comparisons with an explicit reference category. PROC LOGISTIC also offers several other methods for creating design variables, including ordinal, polynomial, and orthogonalized versions of all the methods.

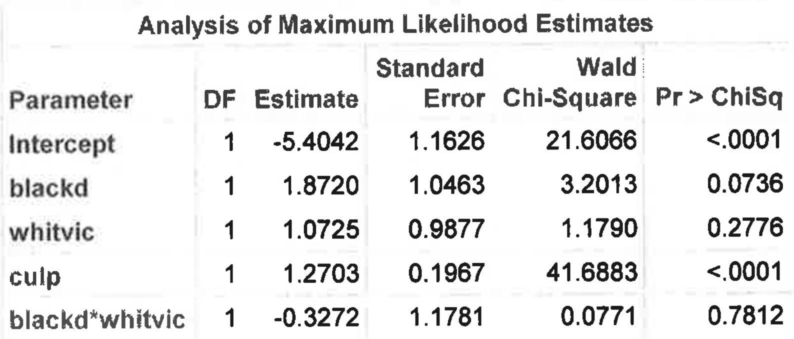
2.10 Multiplicative Terms in the MODEL Statement

Regression analysts often want to build models that have *interactions* in which the effect of one variable depends on the level of another variable. The most popular way of doing this is to include a new explanatory variable in the model, one that is the product of the two original variables. With PROC LOGISTIC, rather than creating a new variable in a DATA step, you can specify the product directly in the MODEL statement. For example, some criminologists have argued that black defendants who kill white victims may be especially likely to receive a death sentence. We can test that hypothesis for the New Jersey data with this program:

```
PROC LOGISTIC DATA=penalty;
  MODEL death(EVENT='1') = blackd whitvic culp blackd*whitvic;
RUN;
```

This produces the table in Output 2.9.

Output 2.9 Multiplicative Variables in PROC LOGISTIC



Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.4042	1.1626	21.6066	<.0001
blackd	1	1.8720	1.0463	3.2013	0.0736
whitvic	1	1.0725	0.9877	1.1790	0.2776
culp	1	1.2703	0.1967	41.6883	<.0001
blackd*whitvic	1	-0.3272	1.1781	0.0771	0.7812

With a *p*-value of .78, the product term is clearly not significant and can be excluded from the model. The MODEL statement also has a short-hand notation that allows you specify both the interaction and the two main effects:

```
MODEL death(EVENT='1') = culp blackd|whitvic;
```

The product syntax in LOGISTIC also makes it easy to construct polynomial functions. For example, to estimate a *cubic* equation you can specify a model of the form

```
MODEL y = x x*x x*x*x;
```

Or, equivalently

```
MODEL y = x|x|x;
```

This fits the model $\log(p/(1-p)) = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

If you are a researcher or student with experience in multiple linear regression and want to learn about logistic regression, Paul Allison's *Logistic Regression Using SAS®: Theory and Application, Second Edition*, is for you! Informal and nontechnical, this book both explains the theory behind logistic regression, and looks at all the practical details involved in its implementation using SAS. Several real-world examples are included in full detail. This book also explains the differences and similarities among the many generalizations of the logistic regression model. The following topics are covered: binary logistic regression, logit analysis of contingency tables, multinomial logit analysis, ordered logit analysis, discrete-choice analysis, and Poisson regression. Other highlights include discussions on how to use the GENMOD procedure to do loglinear analysis and GEE estimation for longitudinal binary data. Only basic knowledge of the SAS DATA step is assumed. The second edition describes many new features of PROC LOGISTIC, including conditional logistic regression, exact logistic regression, generalized logit models, ROC curves, the ODDSRATIO statement (for analyzing interactions), and the EFFECTPLOT statement (for graphing nonlinear effects). Also new is coverage of PROC SURVEYLOGISTIC (for complex samples), PROC GLIMMIX (for generalized linear mixed models), PROC QLIM (for selection models and heterogeneous logit models), and PROC MDC (for advanced discrete choice models).

Free Code on the Web!
support.sas.com/authors

Paul D. Allison is Professor of Sociology at the University of Pennsylvania, and President of Statistical Horizons LLC. He is the author of *Logistic Regression Using SAS®: Theory and Application*, *Survival Analysis Using SAS®: A Practical Guide*, and *Fixed Effects Regression Methods for Longitudinal Data Using SAS®*. Paul has also written numerous statistical papers and published extensively on the subject of scientists' careers. He frequently teaches public short courses on the methods described in his books. You can visit his Web site at www.statisticalhorizons.com.

SAS® Publishing

support.sas.com/publishing
support.sas.com/authors

ISBN 978-1-59994-641-2
9 781599 946412