Quantitative Geographical Analysis
**Logistic Regression**
*Background paper for April 9, 2001 seminar discussion.*
Elvin Wyly

## 1. Overview.

Multiple regression techniques are the mainstay of a broad swath of the multivariate quantitative literature. Yet they suffer from many limitations; recall, for example, the seven key assumptions of the general linear model outlined last week. In practice, one or more of these assumptions will not be met in most real-world applications; nevertheless, researchers often press on, reporting regression coefficient estimates that are biased or unreliable. At a minimum, it is essential to test for these assumptions and to report the likely effects on the results and interpretations for your particular study.

Other limitations of the regression model cannot be ignored. Ordinary least squares regression is appropriate only when the dependent variable is measured on a continuous, interval-ratio scale. For problems in which the dependent variable is an outcome, other methods are required. Regression models in which the dependent variable is measured on a nominal scale are referred to as categorical models; among the most common are various types of *logistic regression* techniques. Logistic regression models can be conceptualized as probabilistic rather than deterministic: the goal is to determine how one or more independent variables affect the probability or likelihood of a particular outcome.

In this background paper, we first consider a bit of the history of some of the mathematical concepts that eventually found their way into probability modeling. We then examine some of the techniques involved in calibrating a logistic regression equation. The final section presents a sample code file for logistic regression in SAS.

## 2. A Short History of *e*.

The logistic regression framework is normally represented by something like this:

$$\ln\left[\frac{P_{robability}}{1-P_{robability}}\right]=\beta_0+\beta_1X_1+\beta_2X_2+\ldots\beta_nX_n+r_i$$

We are trying to predict the odds of a certain outcome (i.e., the ratio of the probability of a certain outcome to the difference between one and this probability) as a function of a set of independent variables, $X_1$, $X_2 ... X_n$, along with a constant term ($B_o$) and a unique residual $r$ for each observation $i$.[1]

[1] In most of the notation we've used up to this point, the error term has been symbolized by *e*. We use *r* here to denote the residual in order to minimize confusion between *e* (error) and *e*(the universal constant, the base of the natural logs).

This looks similar to the multiple linear regression framework, except for the fact that the dependent variable is the natural log of the odds ratio.

But where do 'natural logs' come from, and what makes them 'natural'?

The answer is more interesting than you might think. Among the many concepts that figure prominently in the history of mathematics, *e*, the base of the natural logs, occupies a curious and privileged position, at the intersection of several important debates and simultaneous discoveries. Much of what follows is taken from Eli Maor's lively and accessible history of *e*; which I highly recommend as an addition to the mathematics section of your [voluminous] personal collection of recreational reading.[2]

The number now known as *e* originated from several independent sources in the early years of the seventeenth century. The first origin was an eminently practical consideration with enormous historical and geographical repercussions. At the time, the European transition from localized, feudal economic relations to a more integrated mercantile regime involved intense competition among rival powers. The Portugese had dominated Atlantic trade and exploration since the Treaty of Tordesillas (1494) reconciled competing claims of Portugal and Spain in the Americas, but by the end of the 1500s the Dutch had become serious players. The Dutch East India Company was founded in 1602 by the government of the United Provinces of the Netherlands, in part as an explicit response to the English East India Company. Dutch hegemony lasted until a series of wars between the 1650s and the 1670s that favored England. These broad outlines in 'world leadership cycles' comprise only the simplest features of an extraordinarily complex terrain of struggle among governments and varied interests among shippers, investors, and merchants.

This intense competition in the race to dominate trade might remind us that 'globalization' is simply the latest word for geographical processes that have deep historical roots.[3] The struggle was also, not surprisingly, bound up with all sorts of innovations in cartography, navigation, transportation, storage, and finance. Sometime in the early seventeenth century, someone stumbled upon a few fascinating properties of interest calculations.

Consider the case where we save (or lend) an amount, say $P$, at an agreed annual interest rate $r$ (say, 0.08 for a rate of 8 percent). If the interest is compounded annually, after the first year the account has grown to $P(1+r)$; after the second year, we have $P(1+r)(1+r)$, or $P(1+\mathrm{r})^2$. In more general terms, the relation can be expressed as

$$A=P(1+r)^t$$

---

[2] Maor, Eli. (1994). *e: The Story of a Number*. Princeton: Princeton University Press.

[3] Conversely, one might read the comparatively insular postwar history of United States economic relations as a deviation from a more historically persistent pattern: that is, the "golden age" of American capitalism was not the norm to which subsequent bouts of restructuring are to be compared.

Where $A$ is the amount after compounding, $P$ is the principal invested (or loaned), $r$ is the annual rate of interest, and $t$ is the number of years. If the lender or bank compounds the interest more than once a year, however, then we need to make two changes in the equation:

$$A=P(1+r/n)^{nt}$$

Where $n$ is the number of times per year the interest is compounded. If $n$=1, of course, then the two equations above yield identical results. But consider the effect of more frequent compounding:

<u>Investing $100 at an annual interest rate of 8 percent:</u>

| *Compounding* | *n* | *r/n* | *A* |
|---|---|---|---|
| Annually | 1 | .08 | $108.00 |
| Semiannually | 2 | .04 | $108.16 |
| Quarterly | 4 | .02 | $108.24 |
| Monthly | 12 | .00666667 | $108.29 |
| Weekly | 52 | .00153846 | $108.32 |
| Daily | 365 | .00021918 | $108.33 |

It is worth noting that daily compounding yields just 33 cents more than annual compounding.[4]

Sooner or later, it was inevitable that some mischievious prankster would try to find out what would happen if $r$=1, that is, if the annual interest rate could be boosted to a loan-shark-league 100 percent. In that case, consider what happens if we just put in $1 for one year. Then the equation becomes

$$A=(1+1/n)^{n}$$

If we simply vary $n$, the frequency of compounding during our one-year loan period, we get:

| $n$ | $(1+1/n)^n$ |
|---|---|
| 1 | 2 |
| 2 | 2.25 |
| 3 | 2.37037 |
| 4 | 2.44141 |
| 5 | 2.48832 |
| 10 | 2.59374 |
| 50 | 2.69159 |

[4] There are two ways in which more frequent compounding yields significant returns to the lender or investor. The first is through simple multiplication -- i.e., as P grows very large, even tiny differences in the 1+r/n term can result in sizeable sums. The second is through mechanisms that add to P midstream through the compounding period. This is the magic that allows credit card companies to boost interest charges by crafting ever more innovative charges for late payments, cash advances, and other transactions. Some credit card companies have taken to charging fees to those customers who pay off their balances every month.

| | |
|---|---|
| 100 | 2.70481 |
| 1,000 | 2.71692 |
| 10,000 | 2.71815 |
| 100,000 | 2.71827 |
| 1,000,000 | 2.71828 |
| 10,000,000 | 2.71828 |

No matter how many times the interest is compounded, the amount does not seem to increase by very much. But it never stops increasing, because the second term of the equation always adds another fraction, no matter how small. In fact, this number is *irrational*, meaning that the decimals are non-repeating, and they never stop. In turn, if the decimals never stop, it means that it is not possible to 'solve' the equation for A.

Now, we can recognize the problem as a simple case of a "limit." The basic problem had been around for centuries, of course. Perhaps the earliest illustration came from the philosopher Zeno of Elea, who in the fourth century B.C. proposed four paradoxes to show the inability of mathematics to deal with the concept of infinity. One of these was the "runner's paradox," which presumably showed that motion is impossible. In order for a runner to proceed from point A to point B, she must first pass the midpoint of a line segment connecting the origin and the destination; she must then halve the remaining distance; and so on. Since the original line segment between A and B can be divided into an infinite number of line segments of non-zero length, the runner never reaches her destination. And yet she does.

Around the same time that an anonymous observer came up with the puzzling regularity of the one-hundred percent interest rate, Jonathan Napier published his Description of the Wonderful Canon of Logarithms (1614). Napier, a Scottish landowner, inventor, and son of Sir Archibald Napier, had written the contemporary equivalent of a tabloid exposé of the Catholic Church, but had few other academic credentials. But he was a tireless inventor, and, like many scholars and would-be scholars of the time, concerned to find ways of reducing the tedium of calculations with large numbers. In 1544, the German mathematician Michael Stifel formalized the simple relation between the terms of a geometric progression and the corresponding exponents. If we take any two numbers of the progression 1, $q$, $q^2$, $q^3$..., then the product of these two terms is the same as if we had added the respective exponents. So $q^3 \times q^2 = q^5$. Similarly, dividing the numbers is the same as subtracting the exponents, so that $q^3 / q^2 = q$. If the exponent of the second term is larger than that of the first, we have $q^2 / q^3 = q^{2-3} = q^{-1}$. Negative exponents had been proposed as early as the fourteenth century, but only became widely used in Newton's time. We handle them by defining $q^{-m}=1/q^m$. It also turns out that when we try to calculate $q^m/q^n = q^{m-n}$ when $m=n$, we obtain $q^0=1$. For example, $2^2/2^2=4/4=1$, or $2^0$.

This means we can now extend a geometric progression in both ways:

$$...,q^{-3}, q^{-2}, q^{-1}, q^0=1,q, q^2, q^3,...$$

Consider the powers of 2:

| *n* | $2^n$ |
|---|---|
| -3 | 1/8 |
| -2 | 1/4 |
| -1 | 1/2 |
| 0 | 1 |
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 16 |
| 5 | 32 |
| 6 | 64 |
| 7 | 128 |
| 8 | 256 |
| 9 | 512 |
| 10 | 1,024 |

If we want to calculate 8 x 128, that is, $2^3 \times 2^7$ without the tedium, we can simply look up the value for $2^{3+7}$, or $2^{10}$, which is 1,024. What Napier did was to recognize that any number could be used as a "base", and then the relations between the exponents could be used to perform enormously complicated calculations with out all the tedium. Napier also filled in the gaps in the entries in table of this sort. A few years later, Henry Briggs, a professor of Geometry at Gresham College in London, met with Napier and proposed several modifications. Briggs devised a new set of tables using the base 10.

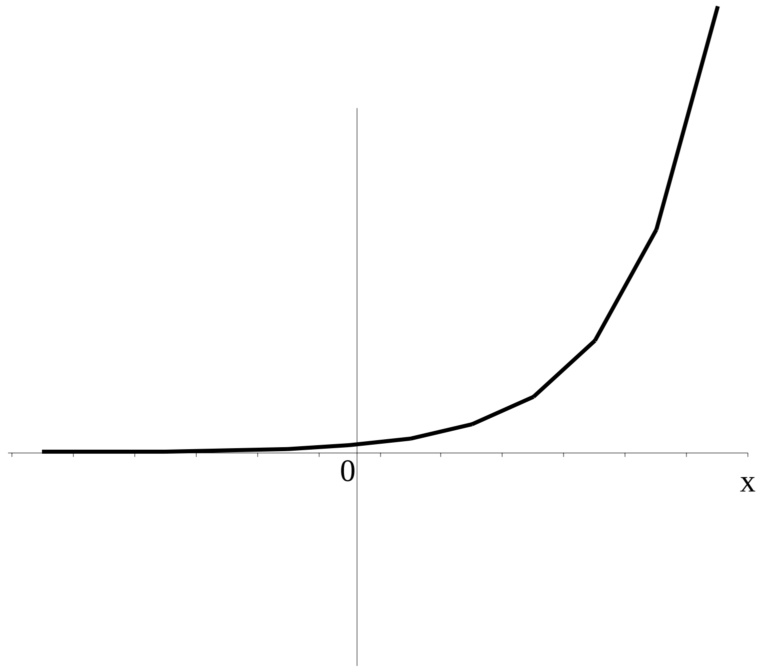
At around the same time, Isaac Newton (1642-1727) and Gottfried Wilhelm Leibniz (1646-1716) were working on a set of problems that led to what we now know as calculus. The initial foundations for some of their work came from much earlier, due to the insights of Archimedes of Syracuse (287-212 B.C.). But one of the key developments of the calculus -- the derivative -- revealed yet another intriguing finding.

Suppose we consider an exponential function of the form $y=b^x$. If we choose a base (b) of 2 and limit ourselves to integer values of x, we obtain this:

| *x* | $2^x$ |
|---|---|
| -3 | 1/8 |
| -2 | 1/4 |
| -1 | 1/2 |
| 0 | 1 |
| 1 | 2 |
| 2 | 4 |
| 3 | 8 |
| 4 | 16 |
| 5 | 32 |
| 6 | 64 |
| 7 | 128 |

| | |
|---|---|
| 8 | 256 |
| 9 | 512 |
| 10 | 1,024 |

And if we graph these values, we obtain something like this:




When the derivative of this function was worked out, it looked like this.

(If your elementary calculus is as rusty as mine was, you'll have to consult a readable refresher. Maori takes you through the details on pages 100-102.)

$$\frac{dy}{dx} = derivative = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

$$\frac{\Delta y}{\Delta x} = b^{x+\Delta x} - b^{x}$$

$$\frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{b^{x}(b^{\Delta x} - 1)}{\Delta x}$$

$$\frac{dy}{dx} = b^{x} \lim_{\Delta x \to 0} \frac{b^{\Delta x} - 1}{\Delta x}$$

Take a close look at the last line. What the equation says is that *the derivative of an exponential function is proportional to the function itself*. The natural question, then, is there any value of b that would make the proportionality on the right-hand side of the equation equal to 1? Maori provides the details in an appendix, but the result is:

$$1 = \lim_{\Delta x \to 0} \frac{b^{\Delta x} - 1}{\Delta x}$$
$$b = \lim_{\Delta x \to 0} (1 + \Delta x)^{1/\Delta x}$$

If we replace the term $1/\Delta x$ with m, we get:

$$b = \lim_{m \to \infty} (1 + 1/m)^m$$

Note that this brings us back to the original puzzle of the dollar invested at a generous return of 100 percent per annum, compounded continuously (and infinite number of times). The equation above is the same as $A=(1+1/n)^n$ with $n$ approaching infinity. The limit is 2.71828...

This is *e*, the universal constant. It is not known precisely why it is designated with the letter e, but it was probably due to the work of Leonhard Euler (1707-1783), who likely was referring to the *e*xponential functions. In any event, the finding is remarkable: if the number e is chosen as a base for the exponential function, then the function is equal to its own derivative.

This finding has wide-ranging implications for phenomena in which the rate of change (the derivative) depends upon the initial state of a system. Examples include radioactive decay (ranging from billions of years for certain isotopes of uranium to rare forms of radium that last only a few milliseconds before decaying to other, more stable elements); the lessening intensity of sound waves as distance increases; and the growth of certain types of populations.

It is also worth noting that if money is compounded continuously, the balance after t years is $A=Pe^{rt}$.

## 3. A Simple Illustration of Logistic Regression.[5]

Suppose you own a few acres on a nice little stream on the Delmarva peninsula.[6] Suppose you have more money than I do. It's a nice little place, right on the Chesapeake Bay, but every now and then a tropical storm makes its way up the Eastern Seaboard, and a large swath of your property is inundated for a day or two. You're interested in putting in a few improvements on your property, and you'd like to have a sense of where you're most likely to have a problem with recurrent flooding. So you turn to the flood insurance maps distributed by FEMA, the Federal Emergency Management Agency. They delineate the boundaries of "100-

[5] The general approach adopted in this illustration is borrowed and adapted from Subhash Sharma (1996). *Applied Multivariate Techniques*. New York: John Wiley and Sons. See Chapter 10, particularly pages 317-321.
[6] Delaware, Maryland, and Virginia.

year" flood plains, which means the area which, on average, they expect to be completely inundated only once every 100 years. The concept is a slippery one, and landowners are often chagrined to experience two or three "100-year" floods in a twenty-year period; so it's a probability, not a certainty.

In any event, many of these maps are quite old, and they're drawn at a very small scale, so that it's hard to see the precise boundaries. So you'd like to assess the accuracy of this map, based on the floods you've seen in the backyard in the past few years. You set up a sampling lattice, and for each point you code two variables: a) whether the point flooded at least once in the last five years (0=no, 1=yes), and b) whether the point lies in the floodplain defined on the old, questionable map (0=no, 1=yes). Suppose you have two dozen sampling points:

| Sample point | FLOOD | MAP |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 1 | 1 |
| 3 | 1 | 1 |
| 4 | 1 | 1 |
| 5 | 1 | 1 |
| 6 | 1 | 1 |
| 7 | 1 | 1 |
| 8 | 1 | 1 |
| 9 | 1 | 1 |
| 10 | 1 | 1 |
| 11 | 1 | 0 |
| 12 | 1 | 0 |
| 13 | 0 | 1 |
| 14 | 0 | 0 |
| 15 | 0 | 0 |
| 16 | 0 | 0 |
| 17 | 0 | 0 |
| 18 | 0 | 0 |
| 19 | 0 | 0 |
| 20 | 0 | 0 |
| 21 | 0 | 0 |
| 22 | 0 | 0 |
| 23 | 0 | 0 |
| 24 | 0 | 0 |

We can summarize the information in this list in another way, using a contingency table:

| | *Located in map floodplain?* | | |
|---|---|---|---|
| *Flooded at least once in last five years?* | Yes | No | Total |
| Yes | 10 | 2 | 12 |
| No | 1 | 11 | 12 |
| Total | 11 | 13 | 24 |

This table, by itself, boosts your confidence in the old map sitting on your dining-room table, as you glance out at the Bay. For only three sample points (the ones off the diagonal of the two-by-two matrix) does the flood map provide erroneous predictions of flood events, at least as measured over the last five years. But is there any way to quantify your confidence in these findings?

*Probabilities and Odds*

1. The probability that any sample point was flooded in the last five years is:

$$P(FLOOD) = 12/24 = 0.50$$

2. The probability that any sample point was flooded given that it is located in the mapped floodplain (Y, for "mapped"):

$$P(FLOOD|Y) = 10/11 = 0.909$$

3. The probability that any sample point was flooded given that it is *no*t located in the mapped floodplain (N, for "Not mapped"):

$$P(FLOOD|N)=2/13=0.154$$

Another way of expressing these relationships is with *odds*. Odds present the same information as probabilities, but in a slightly different way.

1. The odds of any sample point being flooded are

$$odds(FLOOD) = 12/12 =1$$

   which is another way of saying that the odds are even, that is, 1 to 1.

2. Odds of a point getting flooded given that it is mapped in the floodplain are

$$odds(FLOOD|Y)=10/1 = 10$$

   meaning that the odds are ten to one. The odds of a mapped sample point being flooded are ten times larger than its chance of not being flooded.

3. Odds of a point getting flooded given that it lies outside the mapped area:

$$odds\ (FLOOD|N)=2/11=0.182$$

It is also a fairly simple matter to convert odds to probabilities, and back again. Probabilities and odds are just two complementary ways of expressing the same thing:

$$P(FLOOD|Y) = \frac{odds(FLOOD|Y)}{1 + odds(FLOOD|Y)}$$
$$P(FLOOD|Y) = \frac{10}{1+10} = 0.909$$

$$odds(FLOOD|Y) = \frac{P(FLOOD|Y)}{1 - P(FLOOD|Y)}$$
$$odds(FLOOD|Y) = \frac{0.909}{1-0.909} = 10$$

At this point it would be convenient to develop an equation relating the odds of flooding to a predictor variable, such as whether the point is located in the mapped floodplain. The simplest approach would be to use the multiple regression framework we discussed last week. Unfortunately, this approach suffers from three fatal flaws when the dependent variable is discrete. First, predicted probability values from an ordinary least squares solution are not bounded, and commonly fall outside the range (0 to 1) that has any meaning. Second, OLS solutions are heteroscedastic (meaning that error variance is not constant for all values of the independent variables), creating problems for the standard statistical tests for the slope coefficients. Third, fitting an OLS model presumes linearity. This problem is not serious when the independent variables are binary (0/1 dichotomies), but comes into play when continuous independent variables are measured. As Wrigley points out in his chapter, most phenomena conform to a nonlinear, S-shaped curve, such that an increase in the probability from, say, 97 percent to 98 percent is much more difficult to achieve than an increase from 49 to 50 percent (see Wrigley, pp. 26-27).

Natural logarithms provide one way of achieving this nonlinear, S-shaped curve. Let's take two of the odds equations that appear above,

odds(FLOOD|Y)=10/1 = 10
odds (FLOOD|N)=2/11=0.182

and then take the natural log of both sides of these equations:

ln[odds(FLOOD|Y)] = ln (10) = 2.303
ln[odds(FLOOD|N)] = ln(0.182) = -1.704

We can combine these equations. When the area is not mapped (N), we have

ln[odds(FLOOD)] = -1.704

When the area is in the mapped floodplain, the log of the odds is equal to 2.303, or (-1.704 + 4.007). So our general equation is:

ln[odds(FLOOD|MAP)] = -1.704 + 4.007 * MAP

And since the terms in that equation can also be expressed in terms of a ratio of probabilities, we have

$$\ln\left[\frac{P}{1-P}\right] = -1.704 + 4.007\,MAP + r$$

where the *r* term is just a unique error term (residual) for each observation in the sample. Notice the similarity between this specification and the general form of a multiple regression equation. Although the relationship between predictor variables and the probability of a certain outcome is not linear, the *relationship between the log of the odds and these predictors is linear*. The log of the odds is often called a "logit," and so this equation is referred to as logistic regression. The logistic reqression equation is the same as a multiple linear regression, just with the log of the odds serving as the dependent variable. The general form of the logistic specification is:

$$\ln\left[\frac{P_{robability}}{1-P_{robability}}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n + r_i$$

Estimating the parameters for this equation is considerably more complicated than for the case of ordinary least squares multiple linear regression. But once the estimates are obtained, we can calculate a logit (the term on the left-hand side of the equation), and use another equation to calculate the probability value:

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots \beta n X n)}}$$

In some textbooks, the probability value is shown as

$$P = \frac{e^{(\beta_0 + \beta_1 X_1 + \ldots \beta n X n)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \ldots \beta n X n)}}$$

Both of these equations give the same result.

But how do we find the parameter estimates for the logistic regression equation? Wrigley's chapter outlines two solutions. Weighted least squares, which is a modified version of the least squares approach we've seen before, is suited for situations in which both the independent and dependent variables are categorical. But when one or more of the independent variables is continuous (measured on an interval or ratio scale), *maximum likelihood estimation* must be used.

It is not necessary for you to understand the details of maximum likelihood estimation as laid out in Wrigley's text. Very few social scientists understand the procedure on the same level as Wrigley. The general idea, however, is important. As we saw in the background paper for King's ecological solution to the ecological inference problem, maximum likelihood refers to a family of methods where equations are solved by finding the combination of values that were most likely to have resulted in the observed patterns in a sample. The simple example we used last time was a coin toss, which yielded a very simple function; finding the maximum of this curve, either by calculus or trial-and-error, gives us the "maximum likelihood" solution. Every hypothesized relation -- an initial equation in which you specify a certain outcome as a function of a certain set of independent variables -- has a unique likelihood function, which expresses the likelihood of obtaining the observed pattern of outcomes with different combinations of parameters. The task, therefore, is to find the parameters of this function at its maximum. Calculus can be used for simple functions, but for more complicated specifications an iterative search procedure is used.

To illustrate how we assess model fit and the reliability of the parameter estimates, let's venture out of your luxurous Delmarva retreat and consider a more serious set of problems.

## 4. A More Serious Example.[7]

Let's abandon the disciplinary boundaries for a moment, and borrow a dataset from the medical field. From a study on the survival of patients following admission to an adult intensive care unit (ICU), we have information on a sample of 200 patients. The study used logistic regression to predict the likelihood of survival until discharge from the hospital. A total of nineteen predictor variables were observed, but not all of them were used on the survival model. The dependent variable is STA, which is coded 0 for those who survived until discharge, and 1 for those who died in the hospital.[8] Forty of the 200 patients died. Our goal is to find the observed characteristics of patients upon admissions to the ICU that serve as the best predictors of whether they will survive their hospital stay. We know that asking this question raises all sorts of thorny issues in terms of the ultimate use of the information once it is obtained (patient insurance, doctor and hospital malpractice insurance, pre-screening, etc.); but it is clear that the results are not likely to be ignored as uninteresting, trivial statistical blather.

The SAS code file that reads in the data and requests a logistic regression is on the next page, followed by a bit of interpretation.

---

[7] The dataset included in this example is taken from the Data and Story Library, developed by the Department of Statistics at Carnegie Mellon University. See [http://lib.stat.cmu.edu/DASL/Datafiles/ICU.html].

[8] The default setting in SAS yields regression models that predict the likelihood of observing a 0 compared to a 1; this option can be changed either by specifying a different option in the software, or coding a "reverse dummy variable" in the data step.

**********

## *LOGIT.SAS*

## *Logistic Regression of Intensive Care Unit Data*

```
libname qga "c:\sasproj\qga";

data qga.icu;
input ID STA AGE SEX RAC SER CAN CRN INF CPR SYS HRA PRE TYP FRA PO2 PH PCO BIC CRE LOC;
label ID="patient id";
label STA="vital status (0=lived, 1=died)";
label AGE="Patient age in years";
label SEX="Patient sex (0=Male,1=Female)";
label RAC="Patient race (1=White,2=Black,3=Other)";
label SER="Serv at ICU adm (0 = Medical,1=Surgical)";
label CAN="Cancer part of prsnt prob?(0=No,1=Yes)";
label CRN="Hist chronic renal failure(0=No,1=Yes)";
label INF="Infection prob at ICU adm (0=No,1=Yes)";
label CPR="CPR prior to ICU adm (0 = No, 1=Yes)";
label SYS="Systol blood press at ICU adm (in mm Hg)";
label HRA="Heart rate at ICU adm (beats/min)";
label PRE="Prev adm to ICU w/i 6 mo (0=No,1=Yes)";
label TYP="Type of adm (0=Elective,1=Emergency)";
label FRA="Serious fracture (0 = No,1=Yes)";
label PO2="PO2/initial blood (0 = >60,1=60)";
label PH="PH/initial blood (0 = 7.25,1<7.25)";
label PCO="PCO2/initial blood (0 = 45,1=>45)";
label BIC="Bicarbonate/initial blood (0=18,1=<18)";
label CRE="Creatinine/initial blood (0=2.0,1=>2.0)";
label LOC="consc at adm (0=consc,1=stupor,2=coma)";

cards;
8 0 27 2 1 1 1 1 2 1 142 88 1 2 1 1 1 1 1 1 1
12 0 59 1 1 1 1 1 1 1 112 80 2 2 1 1 1 1 1 1 1
14 0 77 1 1 2 1 1 1 1 100 70 1 1 1 1 1 1 1 1 1
28 0 54 1 1 1 1 1 2 1 142 103 1 2 2 1 1 1 1 1 1
32 0 87 2 1 2 1 1 2 1 110 154 2 2 1 1 1 1 1 1 1
38 0 69 1 1 1 1 1 2 1 110 132 1 2 1 2 1 1 2 1 1
40 0 63 1 1 2 1 1 1 1 104 66 1 1 1 1 1 1 1 1 1
41 0 30 2 1 1 1 1 1 1 144 110 1 2 1 1 1 1 1 1 1
42 0 35 1 2 1 1 1 1 1 108 60 1 2 1 1 1 1 1 1 1
50 0 70 2 1 2 2 1 1 1 138 103 1 1 1 1 1 1 1 1 1
51 0 55 2 1 2 1 1 2 1 188 86 2 1 1 1 1 1 1 1 1
53 0 48 1 2 2 2 1 1 1 162 100 1 1 1 1 1 1 1 1 1
58 0 66 2 1 2 1 1 1 1 160 80 2 1 1 1 1 1 1 1 1
61 0 61 2 1 1 1 2 1 1 174 99 1 2 1 1 2 1 2 2 1
73 0 66 1 1 1 1 1 1 1 206 90 1 2 1 1 1 1 1 2 1
75 0 52 1 1 2 1 1 2 1 150 71 2 1 1 1 1 1 1 1 1
82 0 55 1 1 2 1 1 2 1 140 116 1 1 1 1 1 1 1 1 1
84 0 59 1 1 1 1 1 2 1 48 39 1 2 1 2 1 2 2 1 3
92 0 63 1 1 1 1 1 1 1 132 128 2 2 1 1 1 1 1 1 1
96 0 72 1 1 2 1 1 1 1 120 80 2 1 1 1 1 1 1 1 1
98 0 60 1 1 1 1 1 2 2 114 110 1 2 1 1 1 1 1 1 1
100 0 78 1 1 2 1 1 1 1 180 75 1 1 1 1 1 1 1 1 1
102 0 16 2 1 1 1 1 1 1 104 111 1 2 1 1 1 1 1 1 1
111 0 62 1 1 2 1 2 1 1 200 120 1 1 1 1 1 1 1 1 1
112 0 61 1 1 1 1 1 2 1 110 120 1 2 1 1 1 1 1 1 1
136 0 35 1 1 1 1 1 1 1 150 98 1 2 1 1 1 1 1 1 1
137 0 74 2 1 2 1 1 1 1 170 92 1 1 1 1 1 2 1 1 1
143 0 68 1 1 2 1 1 1 1 158 96 1 1 1 1 1 1 1 1 1
153 0 69 2 1 2 1 1 1 1 132 60 1 2 1 1 1 1 1 1 1
170 0 51 1 1 1 1 1 1 1 110 99 1 2 1 1 1 1 1 1 1
173 0 55 1 1 2 1 1 1 1 128 92 1 1 1 1 1 1 1 1 1
180 0 64 2 3 2 1 1 2 1 158 90 2 2 1 1 1 1 1 1 1
184 0 88 2 1 2 1 1 2 1 140 88 2 2 1 1 1 1 1 1 1
186 0 23 2 1 2 1 1 1 1 112 64 1 2 2 1 1 1 1 1 1
187 0 73 2 1 2 2 1 1 1 134 60 1 1 1 1 1 2 1 1 1
190 0 53 1 3 2 1 1 1 1 110 70 2 1 1 1 1 1 1 1 1
191 0 74 1 1 2 1 1 1 1 174 86 1 1 1 1 1 1 1 1 1
207 0 68 1 1 2 1 1 1 1 142 89 1 1 1 1 1 1 1 1 1
211 0 66 2 1 1 1 1 2 1 170 95 2 2 1 1 1 1 1 1 1
214 0 60 1 1 2 2 1 2 1 110 92 1 1 1 1 1 1 1 1 1
219 0 64 1 1 2 1 1 2 1 160 120 1 1 1 1 1 1 1 1 1
225 0 66 1 2 2 2 1 2 1 150 120 1 1 1 1 1 2 1 1 1
237 0 19 2 1 2 1 1 2 1 142 106 1 2 2 1 1 1 1 1 1
247 0 18 2 1 1 1 1 1 1 146 112 1 2 1 1 1 1 1 1 1
249 0 63 1 1 2 1 1 2 1 162 84 2 2 1 1 1 1 1 1 1
260 0 45 1 1 1 1 1 1 1 126 110 1 2 1 1 1 1 1 1 1
```

```
266 0 64 1 1 1 1 1 1 1 162 114 1 2 1 1 1 1 1 1 1
271 0 68 2 1 1 1 1 2 1 200 170 2 2 1 1 1 1 1 1 1
276 0 64 2 1 1 1 1 2 1 126 122 1 2 1 2 1 2 1 1 1
277 0 82 1 1 2 1 1 1 1 135 70 1 1 1 1 1 1 1 1 1
278 0 73 1 1 2 1 1 1 1 170 88 1 1 1 1 1 1 1 1 1
282 0 70 1 1 1 1 1 1 1 86 153 2 2 1 1 1 2 1 1 1
292 0 61 1 1 2 1 1 2 1 68 124 1 2 1 1 1 1 1 1 1
295 0 64 1 1 2 2 1 2 1 116 88 1 1 1 1 1 1 1 1 1
297 0 47 1 1 2 2 1 2 1 120 83 1 1 1 1 1 1 1 1 1
298 0 69 1 1 2 1 1 1 1 170 100 1 1 1 1 1 1 1 1 1
308 0 67 2 1 1 1 1 2 1 190 125 1 2 1 1 1 1 1 1 1
310 0 18 1 1 2 2 1 1 1 156 99 1 1 1 1 1 1 1 1 1
319 0 77 1 1 2 1 1 2 1 158 107 1 1 1 1 1 1 1 1 1
327 0 32 1 2 2 1 1 1 1 120 84 1 2 1 1 1 1 1 1 1
333 0 19 2 1 2 1 1 2 1 104 121 2 1 1 1 1 1 1 1 1
335 0 72 2 1 2 1 1 1 1 130 86 1 2 1 1 1 1 1 1 1
343 0 49 1 1 1 1 1 2 1 112 112 1 2 1 1 1 1 1 1 1
357 0 68 2 1 2 1 1 1 1 154 74 1 1 1 1 1 1 1 1 1
362 0 82 1 1 2 1 2 2 1 130 131 1 2 1 1 1 1 1 1 1
365 0 32 2 3 1 1 1 2 2 110 118 1 2 1 1 1 1 1 1 1
369 0 78 2 1 2 1 1 2 1 126 96 1 2 1 1 1 1 1 1 1
370 0 57 1 1 1 1 1 2 1 128 104 1 2 1 1 1 2 1 1 1
371 0 46 2 1 2 2 1 1 1 132 90 1 2 1 1 1 1 1 1 1
376 0 23 1 1 1 1 1 2 1 144 88 1 2 1 1 1 1 1 1 1
378 0 55 1 1 1 1 1 1 1 132 112 1 2 1 1 1 1 1 1 1
379 0 18 1 1 2 1 1 1 1 112 76 1 2 2 1 1 1 1 1 1
381 0 20 1 1 2 1 1 1 1 164 108 1 2 1 1 1 1 1 1 1
382 0 75 2 1 2 1 1 1 1 100 48 1 1 1 1 1 1 1 1 1
398 0 79 1 1 2 1 1 2 1 112 67 1 1 1 1 1 1 1 1 1
401 0 40 1 1 2 1 1 1 1 140 65 1 2 2 1 1 1 1 1 1
409 0 76 1 1 2 1 1 2 1 110 70 1 2 1 1 1 1 1 1 1
413 0 66 2 1 2 1 1 2 1 139 92 1 1 1 1 1 1 1 1 1
416 0 76 1 1 1 1 1 2 1 190 100 1 2 1 1 1 1 1 1 1
438 0 80 2 1 2 1 1 1 1 162 44 1 2 1 1 1 1 1 1 1
439 0 23 2 1 1 1 1 2 1 120 88 1 2 1 1 1 1 1 1 1
440 0 48 1 2 2 1 1 2 1 92 162 2 2 1 1 1 1 1 1 1
455 0 67 1 2 2 1 1 1 1 90 92 2 1 1 1 1 1 1 1 1
462 0 69 2 1 2 1 1 1 1 150 85 1 2 1 1 1 1 1 1 1
495 0 65 1 3 2 1 1 1 1 208 124 1 1 1 1 1 1 1 1 1
498 0 72 1 1 2 1 1 1 1 126 88 1 1 1 1 1 1 1 1 1
502 0 55 1 1 1 1 1 1 1 190 136 1 2 1 2 2 2 1 1 1
505 0 40 1 1 1 1 1 1 1 130 65 1 2 1 1 1 1 1 1 1
508 0 55 2 1 1 1 1 2 1 110 86 1 2 1 1 1 1 1 1 1
517 0 34 1 1 2 1 1 1 1 110 80 1 2 2 1 1 1 1 1 1
522 0 47 2 1 2 1 1 1 1 132 68 1 2 1 1 1 1 1 1 1
525 0 41 2 1 1 1 1 2 1 118 145 1 2 1 1 2 1 2 1 1
526 0 84 2 1 1 1 2 2 1 100 103 1 2 1 1 1 1 2 2 1
546 0 88 2 1 2 1 1 1 1 110 46 2 1 1 1 1 1 1 1 1
548 0 77 2 1 2 2 1 1 1 212 87 1 1 1 1 1 2 1 1 1
550 0 80 1 1 1 1 1 1 1 122 126 1 2 1 2 1 1 2 1 1
552 0 16 1 1 2 1 1 1 1 100 140 1 2 2 1 1 1 1 1 1
560 0 70 1 1 2 1 1 1 1 160 60 1 1 1 1 1 1 1 1 1
563 0 83 2 1 2 1 1 2 1 138 91 1 2 1 1 1 1 1 1 1
573 0 23 1 2 1 1 1 1 1 130 52 1 2 1 1 1 1 1 1 1
575 0 67 2 1 1 1 1 1 2 120 120 1 2 1 1 2 2 1 1 1
584 0 18 1 1 2 2 1 1 1 130 140 1 1 1 1 1 1 1 1 1
597 0 77 2 1 1 1 1 2 1 136 138 1 1 1 2 2 2 1 1 1
598 0 48 2 1 1 1 1 1 2 128 96 1 2 1 1 1 1 1 1 1
601 0 24 2 2 1 1 1 1 1 140 86 1 2 1 1 1 1 1 1 1
605 0 71 2 1 1 1 1 2 1 124 106 1 2 1 1 1 1 1 1 1
607 0 72 1 1 2 1 1 1 1 134 60 1 2 1 1 1 1 1 1 1
619 0 77 2 1 2 1 2 1 1 170 115 2 1 1 1 1 1 1 1 1
620 0 60 1 1 2 1 1 2 1 124 135 1 2 1 1 1 1 1 1 1
639 0 46 1 1 2 2 1 1 1 110 128 1 1 1 1 1 1 1 1 1
644 0 65 2 1 1 1 1 1 1 100 105 1 2 1 1 1 1 1 1 1
645 0 36 1 1 1 1 1 1 1 224 125 1 2 1 1 1 1 1 1 1
648 0 68 1 1 2 1 1 1 1 112 64 1 1 1 1 1 1 1 1 1
655 0 58 1 1 1 1 1 1 1 154 98 1 2 1 1 1 1 1 1 1
659 0 76 2 1 1 1 1 2 1 92 112 1 2 1 1 1 1 1 1 1
669 0 41 2 2 1 1 1 1 1 110 144 1 2 1 1 1 1 2 2 1
670 0 20 1 3 1 1 1 1 1 120 68 1 2 1 1 1 1 1 1 1
674 0 91 1 1 1 1 2 2 1 152 125 1 2 1 1 1 1 1 1 1
675 0 75 1 1 2 1 1 1 1 140 90 1 2 1 1 1 1 1 1 1
676 0 25 2 1 1 1 1 1 1 131 135 1 2 1 1 1 1 2 1 1
709 0 70 1 1 1 1 1 2 1 78 143 1 2 1 2 1 1 1 1 1
713 0 47 1 1 2 1 1 1 1 156 112 1 2 1 1 1 1 1 1 1
727 0 75 1 3 2 1 1 1 1 144 120 1 2 1 1 1 1 1 2 1
728 0 40 1 2 1 1 1 2 1 160 150 2 2 2 1 1 1 1 1 1
732 0 71 1 1 1 1 1 2 1 148 192 1 2 1 2 2 2 1 1 1
746 0 70 2 1 1 1 1 2 1 90 140 1 2 1 2 1 1 2 1 1
749 0 58 1 1 2 1 1 1 1 148 95 2 2 1 1 1 1 1 1 1
```

```
754 0 54 1 1 2 1 1 1 1 136 80 1 1 1 1 1 1 1 1 1
761 0 77 1 1 2 1 1 1 1 128 59 1 1 1 1 1 1 1 1 1
763 0 55 1 1 2 2 1 2 1 138 140 1 1 1 1 1 1 1 1 1
764 0 21 1 1 2 1 1 1 1 120 62 1 2 1 1 1 1 1 1 1
765 0 53 1 2 1 1 2 1 2 170 115 1 2 1 1 1 1 1 1 1
766 0 31 2 1 1 2 2 2 2 146 100 1 2 1 1 2 2 1 1 1
772 0 71 1 1 2 2 1 1 1 204 52 1 1 1 1 1 1 1 1 1
776 0 49 1 2 1 1 1 1 1 150 100 1 2 1 1 1 1 1 1 1
784 0 60 2 2 1 1 1 2 1 116 92 2 2 1 1 1 1 1 1 1
794 0 50 1 1 1 1 1 2 1 156 99 1 2 1 2 1 2 1 1 1
796 0 45 2 1 2 1 1 1 1 132 109 1 2 2 1 1 1 1 1 1
809 0 21 1 1 2 1 1 1 1 110 90 1 2 1 1 1 1 1 1 1
814 0 73 2 1 2 1 1 1 1 130 83 1 2 1 1 1 1 1 1 1
816 0 28 1 1 2 1 1 2 1 122 80 2 1 2 1 1 1 1 1 1
829 0 17 1 1 2 1 1 1 1 140 78 1 2 2 1 1 1 1 1 1
837 0 17 2 3 1 1 1 1 1 130 140 1 2 1 1 1 1 1 1 1
846 0 21 2 1 2 1 1 1 1 142 79 1 2 1 1 1 1 1 1 1
847 0 68 2 1 2 2 1 1 1 91 79 1 1 1 1 1 1 1 1 1
863 0 17 1 3 2 1 1 1 1 136 78 1 2 1 1 1 1 1 1 1
867 0 60 1 1 1 1 1 2 1 108 120 1 2 1 1 1 1 1 1 1
875 0 69 1 1 2 1 1 1 1 169 73 1 2 1 1 1 1 1 1 1
877 0 88 2 1 1 1 2 1 1 190 88 1 2 1 1 1 1 1 1 1
880 0 20 1 1 2 1 1 1 1 120 80 1 2 1 1 1 1 1 1 1
881 0 89 2 1 2 1 1 1 1 190 114 1 2 1 1 1 2 1 1 3
889 0 62 2 1 1 1 1 1 1 110 78 1 2 1 1 1 1 1 1 1
893 0 46 1 1 1 1 2 2 1 142 89 1 2 1 1 2 1 2 1 1
906 0 19 1 1 2 1 1 2 1 100 137 1 2 1 1 1 1 1 1 1
912 0 71 1 1 1 1 1 2 1 124 124 1 2 1 2 2 2 1 1 1
915 0 67 1 1 2 1 1 1 1 152 78 1 1 1 1 1 1 1 1 1
923 0 20 1 1 2 1 1 1 1 104 83 1 2 1 1 1 1 1 1 1
924 0 73 2 2 1 1 2 1 1 162 100 1 2 1 1 1 1 1 1 1
925 0 59 1 1 1 1 1 1 1 100 88 1 2 1 1 1 1 1 1 1
929 0 42 1 1 2 1 1 1 1 122 84 1 2 2 1 1 1 1 1 1
4 1 87 2 1 2 1 1 2 1 80 96 1 2 2 2 2 2 1 1 1
27 1 76 2 1 2 1 1 2 1 128 90 2 2 1 1 1 1 1 1 1
47 1 78 1 1 1 1 1 2 1 130 132 1 2 1 1 1 1 2 1 1
52 1 63 1 1 1 1 2 2 1 112 106 2 2 1 2 1 1 1 1 1
127 1 19 1 1 2 1 1 1 1 140 76 1 2 1 1 1 1 1 1 1
145 1 67 2 1 1 1 1 2 1 62 145 1 2 1 1 1 1 1 2 1
154 1 53 2 1 1 1 1 2 1 148 128 1 2 1 1 2 2 1 1 1
165 1 92 1 1 1 1 1 2 1 124 80 1 2 1 1 1 1 2 1 1
195 1 57 1 1 1 1 1 2 2 110 124 1 2 1 1 1 1 1 1 3
202 1 75 2 1 2 2 1 1 1 130 136 1 1 1 1 1 1 1 1 1
204 1 91 1 1 1 1 1 2 1 64 125 1 2 1 1 1 2 1 1 1
208 1 70 1 1 2 1 1 1 1 168 122 1 1 1 2 1 1 1 1 2
222 1 88 1 1 1 1 1 2 2 141 140 1 2 1 1 1 1 1 1 1
238 1 41 1 1 2 1 1 2 1 140 58 1 2 1 1 1 1 1 1 3
241 1 61 1 1 1 1 1 1 1 140 81 1 2 1 1 1 1 1 1 1
273 1 80 1 1 2 1 1 1 1 100 85 1 2 1 1 1 1 1 1 1
285 1 40 1 1 1 1 1 2 1 86 80 2 2 1 1 1 1 1 1 1
299 1 75 1 1 1 1 1 2 1 90 100 1 2 1 1 1 1 1 1 2
331 1 63 2 1 2 1 2 2 2 36 86 1 2 2 1 1 1 1 2 3
346 1 75 2 1 1 2 1 1 1 190 94 1 2 1 1 1 1 1 1 1
380 1 20 1 1 2 1 1 1 1 148 72 1 2 2 1 1 1 1 1 1
384 1 71 1 1 1 1 1 1 1 142 95 1 2 1 1 1 1 1 1 1
412 1 51 2 1 2 1 1 2 1 134 100 2 2 1 1 1 1 1 1 2
427 1 65 1 1 1 1 1 1 1 66 94 1 2 1 1 1 1 1 1 3
442 1 69 2 3 1 1 2 1 1 170 60 2 2 1 2 1 1 1 1 1
461 1 55 1 1 2 1 2 2 1 122 100 2 2 1 1 1 1 1 1 1
468 1 50 2 1 2 2 1 1 1 120 96 1 2 1 1 1 1 1 1 1
490 1 78 1 1 1 1 1 2 1 110 81 1 2 1 1 1 1 1 1 1
518 1 71 2 1 1 1 1 1 2 70 112 1 2 1 1 1 1 1 1 3
611 1 85 2 1 2 1 1 1 1 136 96 1 2 1 1 1 1 1 1 1
613 1 75 1 1 1 1 2 2 1 130 119 1 2 1 1 2 1 2 2 1
666 1 65 2 1 1 1 1 2 2 104 150 1 2 1 1 1 2 1 1 3
671 1 49 1 1 1 1 1 2 2 140 108 1 2 1 1 1 1 2 1 1
706 1 75 2 1 1 1 2 2 2 150 66 1 2 1 1 1 1 1 2 3
740 1 72 2 1 1 1 1 1 1 90 160 1 2 1 1 1 1 1 1 1
751 1 69 1 1 1 1 2 1 1 80 81 1 2 1 1 1 1 1 1 3
752 1 64 1 1 1 2 1 2 1 80 118 1 2 1 2 1 1 1 2 1
789 1 60 1 1 1 1 1 2 1 56 114 2 2 1 1 2 1 2 1 1
871 1 60 1 3 2 1 2 2 1 130 55 1 2 1 1 1 1 1 1 2
921 1 50 2 2 1 1 1 1 1 256 64 1 2 1 1 1 1 1 1 2
;
run;

proc logistic data=qga.icu;
        model sta=age sex;
        title "model 1 baseline";
        run;
proc logistic data=qga.icu;
```

```
        model sta=age sex can hra typ fra loc cpr inf ser;
        title "model 2: add medical variables";
        run;
proc logistic data=qga.icu;
        model sta=age sex can hra typ fra loc cpr inf ser pre;
        title "model 3: add prev icu";
        run;
```

## *OUTPUT*
## *Logistic Regression of Intensive Care Unit Data*

```
                                        model 1 baseline         10:42 Monday, April 9, 2001   5

                                     The LOGISTIC Procedure

             Data Set: QGA.ICU
             Response Variable: STA       vital status (0=lived, 1=died)
             Response Levels: 2
             Number of Observations: 200
             Link Function: Logit


                                        Response Profile

                                   Ordered
                                     Value     STA     Count

                                         1       0       160
                                         2       1        40


              Model Fitting Information and Testing Global Null Hypothesis BETA=0

                                              Intercept
                             Intercept          and
            Criterion          Only         Covariates    Chi-Square for Covariates

            AIC               202.161        198.305           .
            SC                205.459        208.200           .
            -2 LOG L          200.161        192.305          7.856 with 2 DF (p=0.0197)
            Score                .              .             7.180 with 2 DF (p=0.0276)


                           Analysis of Maximum Likelihood Estimates

                      Parameter   Standard      Wald         Pr >      Standardized       Odds
   Variable     DF     Estimate     Error     Chi-Square   Chi-Square     Estimate         Ratio

   INTERCPT    1       3.0454     0.8190      13.8254       0.0002            .            .
   AGE         1      -0.0276     0.0107       6.7005       0.0096      -0.304989       0.973
   SEX         1       0.0113     0.3718       0.0009       0.9757       0.003034       1.011


                  Association of Predicted Probabilities and Observed Responses

                            Concordant = 62.4%          Somers' D = 0.259
                            Discordant = 36.5%          Gamma     = 0.262
                            Tied       =  1.0%          Tau-a     = 0.083
                            (6400 pairs)                c         = 0.630
```

```
                         model 2: add medical variables                              6
                                                              10:42 Monday, April 9, 2001

                              The LOGISTIC Procedure

          Data Set: QGA.ICU
          Response Variable: STA       vital status (0=lived, 1=died)
          Response Levels: 2
          Number of Observations: 200
          Link Function: Logit


                                  Response Profile

                            Ordered
                              Value     STA     Count

                                  1       0       160
                                  2       1        40


           Model Fitting Information and Testing Global Null Hypothesis BETA=0

                                      Intercept
                         Intercept       and
         Criterion         Only       Covariates    Chi-Square for Covariates

         AIC              202.161       164.197          .
         SC               205.459       200.479          .
         -2 LOG L         200.161       142.197       57.964 with 10 DF (p=0.0001)
         Score               .             .          58.505 with 10 DF (p=0.0001)


                         Analysis of Maximum Likelihood Estimates

                    Parameter   Standard      Wald        Pr >      Standardized       Odds
Variable    DF      Estimate      Error    Chi-Square  Chi-Square     Estimate        Ratio

INTERCPT    1       13.7570     3.4112      16.2636      0.0001           .             .
AGE         1       -0.0392     0.0138       8.0913      0.0044      -0.432895       0.962
SEX         1        0.4884     0.4603       1.1261      0.2886       0.131036       1.630
CAN         1       -2.1649     0.8711       6.1769      0.0129      -0.358970       0.115
HRA         1       0.00379    0.00907       0.1748      0.6758       0.056073       1.004
TYP         1       -2.7604     0.9852       7.8506      0.0051      -0.673342       0.063
FRA         1       -0.7634     0.9093       0.7049      0.4012      -0.111140       0.466
LOC         1       -1.8571     0.5778      10.3300      0.0013      -0.469682       0.156
CPR         1       -0.4316     0.8516       0.2569      0.6123      -0.058812       0.649
INF         1       -0.3837     0.4656       0.6790      0.4099      -0.104671       0.681
SER         1        0.3022     0.5200       0.3376      0.5612       0.083303       1.353
                         model 2: add medical variables                              7
                                                              10:42 Monday, April 9, 2001

                              The LOGISTIC Procedure

               Association of Predicted Probabilities and Observed Responses

                    Concordant = 84.7%          Somers' D = 0.698
                    Discordant = 14.9%          Gamma     = 0.701
                    Tied       =  0.4%          Tau-a     = 0.225
                    (6400 pairs)                c         = 0.849
```

```
                          model 3: add prev icu       10:42 Monday, April 9, 2001   8

                         The LOGISTIC Procedure

          Data Set: QGA.ICU
          Response Variable: STA       vital status (0=lived, 1=died)
          Response Levels: 2
          Number of Observations: 200
          Link Function: Logit


                            Response Profile

                         Ordered
                           Value     STA     Count

                               1       0       160
                               2       1        40


           Model Fitting Information and Testing Global Null Hypothesis BETA=0

                                     Intercept
                        Intercept       and
         Criterion        Only       Covariates    Chi-Square for Covariates

         AIC            202.161       163.805         .
         SC             205.459       203.385         .
         -2 LOG L       200.161       139.805       60.356 with 11 DF (p=0.0001)
         Score             .             .          60.356 with 11 DF (p=0.0001)


                         Analysis of Maximum Likelihood Estimates

                Parameter   Standard      Wald        Pr >      Standardized       Odds
Variable   DF    Estimate     Error    Chi-Square  Chi-Square     Estimate        Ratio

INTERCPT   1      15.1049     3.5621      17.9819      0.0001            .          .
AGE        1      -0.0418     0.0145       8.3256      0.0039    -0.462539      0.959
SEX        1       0.5910     0.4732       1.5596      0.2117     0.158546      1.806
CAN        1      -2.4174     0.8982       7.2440      0.0071    -0.400845      0.089
HRA        1      0.00497    0.00900       0.3047      0.5809     0.073529      1.005
TYP        1      -2.7804     0.9827       8.0055      0.0047    -0.678216      0.062
FRA        1      -0.9107     0.9191       0.9819      0.3217    -0.132583      0.402
LOC        1      -1.9129     0.5773      10.9799      0.0009    -0.483776      0.148
CPR        1      -0.6132     0.8685       0.4985      0.4802    -0.083552      0.542
INF        1      -0.2495     0.4741       0.2769      0.5987    -0.068069      0.779
SER        1       0.4067     0.5244       0.6015      0.4380     0.112110      1.502
PRE        1      -0.8991     0.5679       2.5061      0.1134    -0.177440      0.407
                          model 3: add prev icu       10:42 Monday, April 9, 2001   9

                         The LOGISTIC Procedure

             Association of Predicted Probabilities and Observed Responses

                         Concordant = 85.9%          Somers' D = 0.720
                         Discordant = 13.9%          Gamma     = 0.722
                         Tied       =  0.2%          Tau-a     = 0.232
                         (6400 pairs)                c         = 0.860
```