

GRAPHIC REPRESENTATION OF LOGISTIC REGRESSION MODELS

Edward L. Spitznagel, John K. Gohagan, and Donald G. Sessions
Washington University

ABSTRACT

Logistic regression and discriminant analysis are valuable tools in medical diagnosis and other disciplines involving prediction of a binary variable. Once the model or models have been fit, there remains the important task of conveying their meaning to the client. This paper demonstrates several alternatives in graphical presentation of these results.

INTRODUCTION

In analyzing data from a tumor registry of head and neck cancer patients, one of our goals has been to predict recurrence of cancer. Application of logistic regression has been fairly successful: For example, in studying tumors from three primary sites (supraglottic, glottic, and inferior hypopharynx), the logistic regression model incorporating site, TNM classification, and age had the following measures of association:

LR chi-square = 269.0 (with 14 df and $n=815$)
Harrell's R = 0.471
Fraction of concordant pairs = 0.815

In ordinary regression, measures of association have been around so long and been discussed so widely that their interpretation is straightforward, at least for statisticians. And for the non-statistician, a simple scatterplot of the data usually conveys an adequate feeling for how well the model fits and how accurate predictions from it will be.

With logistic regression, things are more complicated. First, the measures of association are not as familiar as their ordinary regression counterparts are. Second, an important goal of logistic regression is the classification of individual cases by comparison with a cutpoint, and any measure that speaks to this goal will necessarily depend on the choice of cutpoint.

With the latter consideration in mind, we began looking for graphical methods by which association could be assessed, taking into account the effect of the cutpoint. This article shows some of the possibilities. In all cases, relatively few lines of SAS code are needed to produce them.

THE GRAPHICS

In all, we present six figures, three related to density functions and three having to do with sensitivity versus specificity, done in the traditional ROC format.

The data underlying all six is the same, consisting of 815 head and neck cancer cases, of which 316 had recurrence. For the 316 cases of recurrence, the mean logit was 0.39, with a standard deviation of 1.36. For the 499 cases of non-recurrence, the mean logit was -1.22, with a standard deviation of 1.15.

Figure 1 shows normal density functions matched to these means and standard deviations. The amount of overlap between these density curves shows clearly that no cutpoint will achieve perfect discrimination between the two groups. In the case of our data, the logit values were approximately normal within each group. It is a good idea to check for approximate normality before using this kind of portrayal, but with multiple predictors contributing to the logit, normality is likely to be the rule rather than the exception.

Figure 1 demonstrates that separation will be imperfect wherever the cutpoint is placed, but it does not show the relative frequency of errors unless the two groups are of equal size (i.e., prevalence = 50%). Figure 2 consists of a modification of Figure 1 in which the areas under the curves reflect the relative sizes of the two groups. From this plot, one can read immediately the point at which maximum correct classification occurs (at the intersection of the two curves, or in economic language, the point at which the two marginal misclassification rates are equal).

Figure 3 goes one step further and graphs the percent correctly classified as a function of logit cutpoint. As auxiliary curves, we have included the percent correctly classified belonging to each group. The curve decreasing as one moves from left to right is the percent correctly classified belonging to the recurrence group. The sum of the heights of the two auxiliary curves is the upper curve, the total percent correctly classified.

TOOLS

It is clear from the graph but still worth mentioning that the cutpoint of maximum correct classification is not the point at which the two group curves cross. That is, maximum correct classification does not occur when the number of misclassifications from each group is the same. It is also worth mentioning that Figure 3 does not show the cutpoint for minimum cost due to misclassification unless the penalty for each type of misclassification (false negative and false positive in diagnostic terms) is the same. The graph can, however, be easily modified to incorporate costs, and drawn with a dip in the center to reflect minimum cost.

The next three figures all are concerned with portraying the classification information using Relative Operating Characteristic (ROC) curves. An ROC curve relates the rate of true positive classifications to the rate of false positive classifications. Due to inconsistencies in the literature of the meanings of these terms, we rephrase in diagnostic language: An ROC curve is a graph of sensitivity versus the complement of specificity. That is, the ROC curve lets us view how sensitivity increases as we relax specificity. Note that ROC curves do not incorporate differences in group size, or, equivalently, prevalence.

Figure 4 is the ROC curve for the fitted normal densities as graphed in Figure 1. The dashed diagonal line is a reference corresponding to a model producing no separation between the two groups. The further the ROC curve lies above this line and up into the corner, the greater the separation. The area below and to the right of the ROC curve (up to a maximum value of 1) corresponds to the "fraction of concordant pairs" statistic 0.815 quoted above and also to the U statistic of the Mann-Whitney test.

Figure 5 is the ROC curve produced from the actual data, without fitting any distribution. Except for the bumpiness, which understandably increases at the ends of the curve, the agreement between Figures 4 and 5 is very good, as one would expect if the logit scores in each of the two groups were close to normally distributed.

Figure 6 is a variant on Figure 5, consisting of the same curve plotted on twin probit scales, as suggested by Swets and Pickett [1]. Of all six plots, this is the one that yields the most technical information. Departures from linearity in this graph correspond to departures from normality in one or both of the classification groups. There is, in fact, just a hint of such departure, to the effect that the overlapping tails of the two densities might be slightly smaller than expected, with a corresponding increase in correct classification rates in that region.

Following is a partial list of ingredients used in producing these graphs:

The LOGIST procedure
The MEANS procedure
The GPLOT procedure

The PROBNORM function
The PROBIT function

The LOGIST procedure creates an output data set containing predicted probabilities in a variable named \underline{P} . These can be converted back to logits by the transformation

$$\text{LOGIT} = \text{LOG}(\underline{P}/(1-\underline{P}));$$

The MEANS procedure was used to create an output data set containing means and standard deviations of the logit variable, from which the density functions were plotted. The GPLOT procedure was used to produce all six figures. The PROBNORM function was used to produce the percent correct curves in Figure 3, and the PROBIT function was used to produce the ROC curve plotted on twin probit scales in Figure 5.

CONCLUDING REMARKS

Although our examples have been in terms of logistic regression, the methods apply equally well to discriminant analysis. A number of other variations on the themes presented above are possible. To mention one, the various spline interpolation features of the GPLOT procedure can be used to produce curves that reflect non-normal data, but with the statistical noise smoothed away.

REFERENCES

- [1] Swets, J.A. and Pickett, R.M. (1982) Evaluation of Diagnostic Systems, Academic Press, New York
- [2] Harrell, F.E. (1983) The LOGIST Procedure. SUGI Supplemental Library User's Guide, 181-202.

Addresses:

Edward L. Spitznagel
Department of Mathematics
Washington University
St. Louis, Missouri 63130

John K. Gohagan
Department of Preventive Medicine
Washington University School of Medicine
St. Louis, Missouri 63110

Donald G. Sessions
Department of Otolaryngology
Washington University School of Medicine
St. Louis, Missouri 63110

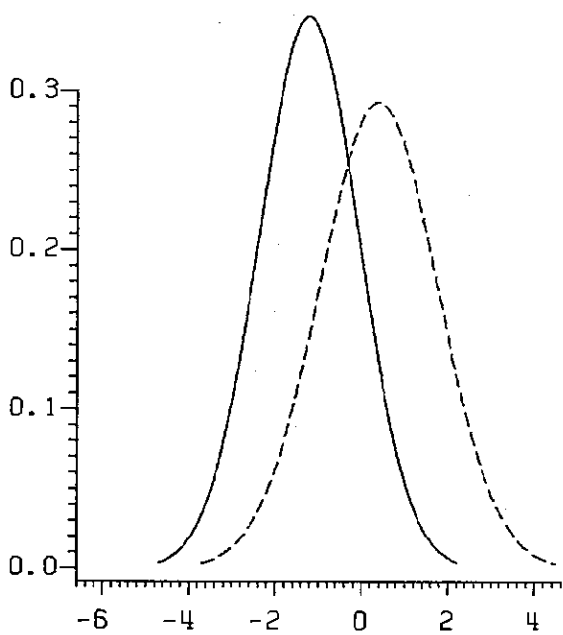


Figure 1. Comparison of normal density functions matched on means and standard deviations. The function for recurrence is to the right.

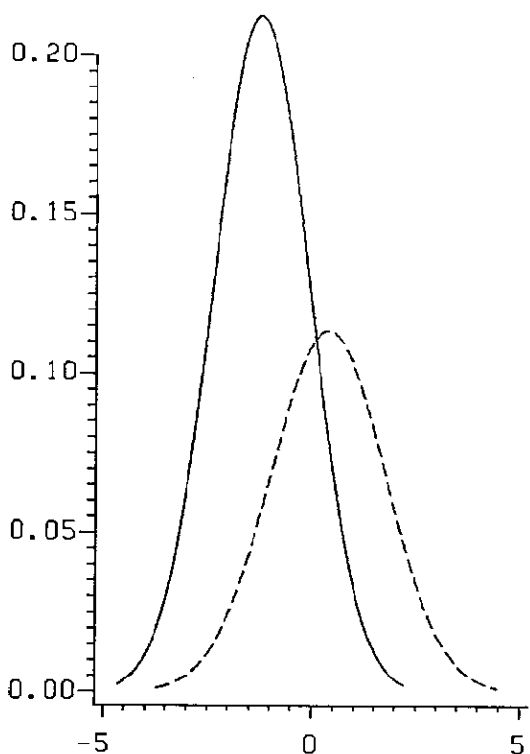


Figure 2. Comparison of density functions scaled to match sizes within the groups. Total area under both curves is equal to 1.

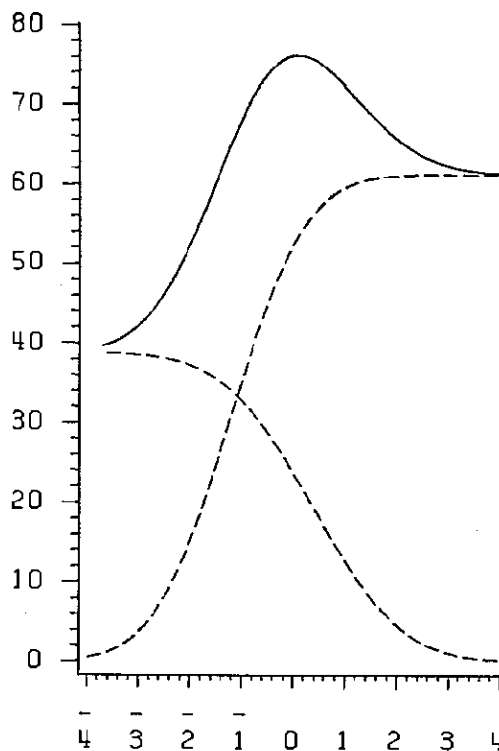


Figure 3. Percent correctly classified as a function of the logit outpoint. Dashed curves indicate the contribution of true positives and true negatives.

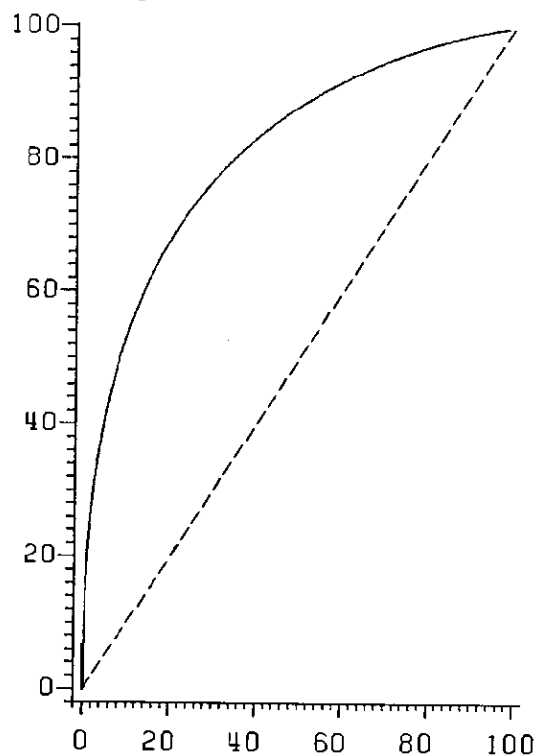


Figure 4. ROC curve produced from the fitted normal density functions.

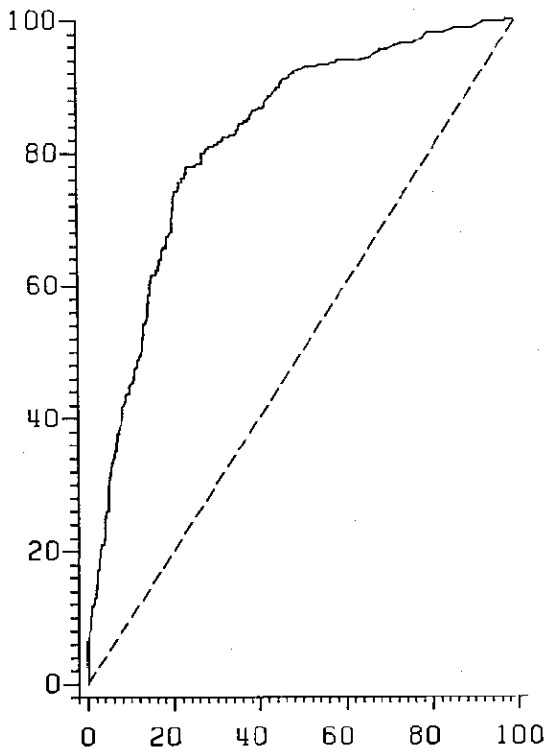


Figure 5. ROC curve produced from the actual classifications.

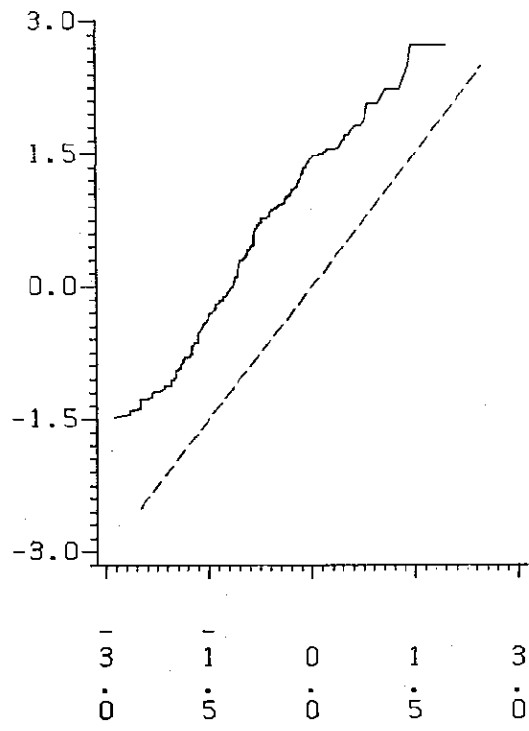


Figure 6. ROC curve produced from the actual classifications, plotted on probit scales. Linearity indicates normality of logit values. Inequality of variances is reflected in departure from parallelism with the diagonal.