UNDERSTANDING STATISTICS, 1(1), 31–70

TEACHING ARTICLES

Logistic Regression Analysis and Reporting: A Primer

Chao-Ying Joanne Peng and Tak-Shing Harry So
Department of Counseling and Educational Psychology
Indiana University–Bloomington

Logistic regression, being well suited for analyzing dichotomous outcomes, has been increasingly applied in social science research. That potential expanded usage demands that researchers, editors, and readers be coached in terms of what to expect in an article that used the logistic regression technique: What tables should be included? What assumptions tested? What figures or charts should be expected? In this article we seek to answer these questions with an illustration of logistic regression applied to a real world data set. Results were evaluated and diagnosed in terms of the overall test of the model, interpretability and statistical significance of each predictor, goodness-of-fit statistics, predictive power, accuracy of prediction, and identification of potential outliers. Guidelines are offered for modeling strategies and reporting standards in logistic regression. Furthermore, 6 statistical packages were employed to perform logistic regression. Their strengths and weaknesses were noted in terms of flexibility, accuracy, completeness, and usefulness.

Keywords: logistic regression, categorical variables, binary outcome, statistical packages, statistical computing

Logistic regression was first proposed in the 1970s as an alternative technique to overcome limitations of ordinary least squares (OLS) regression in handling dichotomous outcomes. It became available in statistical packages in the early 1980s. Logistic regression has been widely employed in epidemiological research, where

Requests for reprints should be sent to Chao-Ying Joanne Peng, Department of Counseling and Educational Psychology, School of Education, Room 4050, 201 North Rose Avenue, Indiana University, Bloomington, IN 47405–1006. E-mail: peng@indiana.edu

often the outcome variable is presence or absence of some disease state (e.g., Yarandi & Simpson, 1991). Meanwhile, the use of logistic regression continues to grow in social sciences (e.g., Chuang, 1997; Janik & Kravitz, 1994; Tolman & Weisz, 1995) and educational research (e.g., Okun, Benin, & Brandt-Williams, 1996; St. John, Paulsen, & Starkey, 1996), especially in higher education (Austin, Yaffee, & Hinkle, 1992; Cabrera, 1994; Peng & So, in press; Peng, So, Stage, & St. John, in press).

Increasing volumes of literature written about logistic regression also contribute to the growing use of logistic regression in social sciences research. Logistic regression textbooks by Hosmer and Lemeshow (2000), Kleinbaum (1994), McCullagh and Nelder (1989), and Menard (1995) were published within the last 12 years. Other textbooks of multivariate statistics (e.g., Afifi & Clark, 1990; Ryan, 1997; Tabachnick & Fidell, 1996, 2001) have begun to include chapters on logistic regression in their recent editions. Because logistic regression does not require that data are drawn from a multivariate normal distribution with equal variances and covariances for all variables (Cleary & Angel, 1984; Efron, 1975; Lei & Koehly, 2000; Press & Wilson, 1978), it is less restrictive than linear discriminant function analysis for modeling categorical outcomes. Thus, social sciences researchers have recognized logistic regression as a viable alternative method to linear discriminant function (Tabachnick & Fidell, 2001, p. 521).

Despite the popularity of logistic regression modeling and the ease with which researchers are able to apply this technique using statistical software, confusion continues to exist over terms, concepts, modeling approaches, and interpretations. A recent review of 52 articles, published between 1988 and 1999 in three higher education journals, revealed lack of standards in the practice and reporting of logistic regression (Peng et al., in press). Specifically, inconsistency was found in the ratio of observations to predictors, modeling approaches, assessment of regression models, examinations of interactions among predictors, and presentations of results. The level of completeness and accuracy of supplementary analyses was uneven across studies. Logistic regression results have been reported in terms of logit, odds, odds ratio, relative risk, predicted probability, marginal probability (also called marginal effect, partial effect, or partial change), and change in predicted probability (also called delta-*p*). These terms are not equivalent; thus, their meanings are not interchangeable.

With the wide availability of sophisticated statistical software installed on high-speed computers, the anticipated use of this technique is increasing. That potential expanded usage demands that researchers, editors, and readers be coached in terms of what to expect in an article that used the logistic regression technique: What tables should be included? What assumptions tested? What figures or charts should be expected? This article seeks to answer these questions with an illustration of logistic regression applied to a real world data set. The remainder of this article is divided into seven sections: (a) Logistic Regression Models, (b) Illustration of Logistic Regression Analysis, (c) Evaluations of A Logistic Regression Model,

(d) Outliers and Diagnostic Statistics, (e) Modeling Strategy and Reporting, (f) Evaluations of Six Logistic Regression Procedures, and (g) Summary. A glossary of terms pertaining to logistic regression is found in Appendix A.

LOGISTIC REGRESSION MODELS

Logistic regression is well suited for studying the relation between a categorical or qualitative outcome variable and one or more predictor variables. In the simplest case of one predictor X (say, IQ score) and one dichotomous outcome variable Y (say, diagnosed to be learning disabled), the logistic model predicts the logit of Y from X. The logit is the natural logarithm (ln) of odds of Y. The simple logistic model has the form:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \log(odds) = logit = \alpha + \beta x. \quad (1)$$

$$\text{Hence, } \pi = \text{Probability}(Y = \text{outcome of interest} \mid X = x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}, \quad (2)$$

where π is the probability of the outcome of interest, or the "event", under variable Y, α is the Y intercept, and β is the slope parameter. X can be categorical or continuous, whereas Y is always categorical. Although a categorical variable may yield two or more possible categories, we focus on dichotomous outcomes only. Illustrations presented in this article can be extended to polytomous variables with ordered or unordered (i.e., nominal) outcomes.

Figure 1 presents two logistic functions for $\alpha = 0$, $\beta = 0.2$ and $\alpha = 0$, $\beta = 0.4$. It illustrates four unique characteristics:

1. Unless $\beta = 0$, the binary logistic regression maps the regression line onto the interval (0,1), which is compatible with the logical range of probabilities.
2. The regression line is monotonically increasing if $\beta > 0$ (and monotonically decreasing if $\beta < 0$).
3. The function takes on the value of 0.5 at $x = -\alpha/\beta$ (the point of inflection) and is symmetric to the point of $(-\alpha/\beta, 0.5)$.
4. While holding α as a constant, the logistic curve's steepness is determined by the absolute value of β. If β is held constant, the magnitude of α determines the median location of the curve (see number 3).

Within the inferential framework, the null hypothesis states that β equals zero in the population. Rejecting such a null hypothesis implies that a relation exists between X and Y. If a predictor is binary, such as gender, the exponentiated β (=

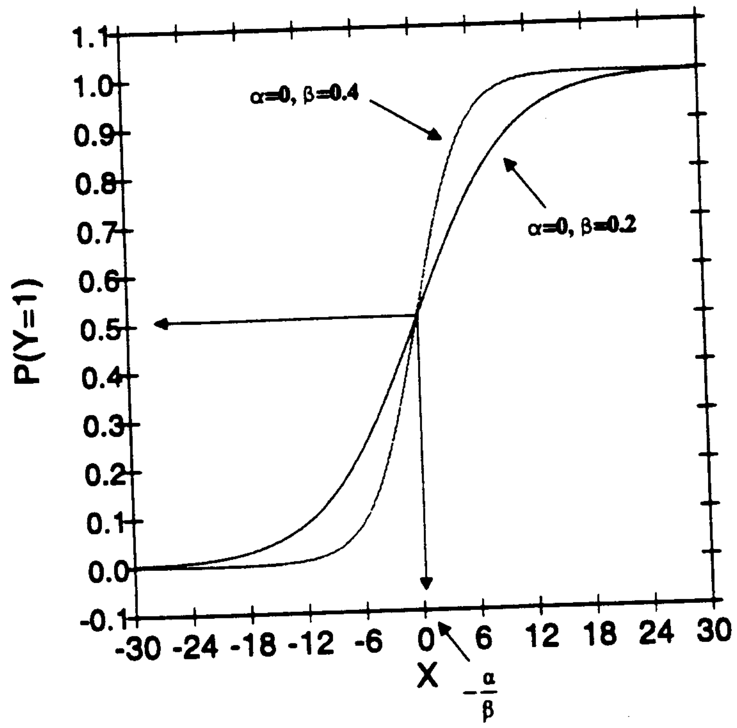

FIGURE 1 Univariate logistic regression model based on $\alpha = 0$ and $\beta = 0.2$ overlayed with $\alpha = 0$, $\beta = 0.4$.

TABLE 1
Sample Data for Weight of Mothers and Weight of New Born Infants

	Weight of Mothers	
Weight of Infants	*Low*	*Normal*
Low	73	15
Normal	23	11

e^{β}) is the odds ratio, namely, the ratio of two odds (see Appendix A). Consider an example in which the distribution of a dichotomous outcome variable (an infant is born with normal or low weight) is paired with a dichotomous predictor variable (mother's weight is normal or below normal). Example data are included in Table 1. A test of independence using chi-square could be applied. The results yield $\chi^2(1, N = 122) = 3.4268$. Alternatively, one might prefer to assess a low-weight mother's odds of giving birth to a low-weight baby versus a normal-weight baby, relative to a normal-weight mother's odds; the result is an odds ratio of 2.328. The odds ratio suggests that mothers who are below normal weight are 2.328 times more likely to deliver a low-weight baby than normal, compared to mothers who weigh normally. The odds ratio is derived from two odds (73/23 for normal-weight mothers and 15/11 for under-weight mothers); its natural logarithm, that is, ln (2.328), is a logit that equals 0.845. The value of 0.845 would be the regression coefficient (i.e., β) of the predictor (mother's weight), if logistic regression were used to model the outcome of an infant's weight.

If a predictor is continuous, such as mother's weight in pounds, Peterson (1984) suggested that delta-*p* (or change in the probability) be used in interpreting the logistic regression result. Using the previous example, delta-*p* would mean the increase in probability of giving birth to low-weight babies if the mother's weight decreases from 120 pounds to 100 pounds, or from 140 pounds to 120 pounds. These interpretations are easy to understand. Consequently, logistic regression has become increasingly appealing to social sciences researchers.

Extending the logic of the simple logistic regression to multiple predictors, one may construct a complex logistic regression as follows:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k. \tag{3}$$

Therefore,

$$\begin{aligned}\pi &= \text{Probability}(Y = \text{outcome of interest} | X_1 = x_1, X_2 = x_2, \ldots X_k = x_k) \\ &= \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k}},\end{aligned} \tag{4}$$

where π is once again the probability of the event, α is the Y intercept, βs are slope parameters, and Xs are a set of predictors. α and βs are estimated by the maximum likelihood method. This method is designed to maximize the likelihood of obtaining the data given its parameter estimates. The interpretation of βs is rendered using either the odds ratio (for categorical predictors) or the delta-*p* (for continuous predictors). The null hypothesis states that all βs equal zero. A rejection of this null hypothesis implies that at least one β does not equal zero in the population. Unlike discriminant function analysis, logistic regression does not assume that Xs are distributed as a multivariate normal distribution with equal covariance matrix across all levels of Y. Instead, it assumes that the binomial distribution describes the distribution of the errors $= Y - \hat{Y}$. The binomial distribution is also the assumed distribution for the conditional mean of the dichotomous outcome with the probability given by Equations 2 and 4. This assumption is satisfied as long as the same probability is maintained across the range of predictor values.

ILLUSTRATION OF LOGISTIC REGRESSION ANALYSIS

In this section, we describe a data set for which logistic regression analysis is suitable to predict dichotomous outcomes. Six logistic regression algorithms implemented in statistical packages were employed to perform logistic regression. Two analysis issues, data formats and specification of interactions in the model, are discussed and treated.

Data

The "married women labor force participation" data (hereafter abbreviated as MWLFP) were provided by Mroz (1987). The data set describes profiles of 752 married White women who were recruited in 1975 for the panel study of income dynamics conducted at the University of Michigan. Logistic regression was applied to explain women's decision to enter the paid labor force in 1975 with their demographic information. The outcome variable (*lfp*) was coded 1 for women who worked for pay in 1975 and 0 otherwise. The predictors were: women's age (*age*), number of children under the age of 5 (*k5*), number of children between ages 6 and 18 (*k618*), the household's total income minus the wife's income (*inc*), the wife's estimated wage rate (*wg*), and two dichotomous variables indicating, respectively, whether the wife (*wc*) and the husband (*hc*) spent at least 1 year in college. Table 2 presents descriptive information of these eight variables.

Logistic Regression Modeling

A logistic regression model was fit to the MWLFP data to explain the predicted odds of women entering the paid labor force (i.e., *lfp* = 1) in 1975. The model included four main effects—*k5*, *k618*, *hc*, *wc*—plus one categorical variable (*newage*) and its interaction with *wc*.

$$\begin{gathered}\text{predicted logit } (lfp = 1) = \alpha + \beta_1 \times k5 + \beta_2 \times k618 + \beta_3 \times newage1 \\ + \beta_4 \times newage2 + \beta_5 \times newage3 + \beta_6 \times newage4 + \beta_7 \times newage5 + \\ \beta_8 \times hc + \beta_9 \times wc + \beta_{10} \times (wc{*}newage1) + \\ \beta_{11} \times (wc{*}newage2) + \beta_{12} \times (wc{*}newage3) + \beta_{13} \times (wc{*}newage4) \\ + \beta_{14} \times (wc{*}newage5).\end{gathered}$$

The variable *newage* was transformed from the continuous variable age according to a 5-year increment (i.e., 30 to 34, 35 to 39, …, and 55 to 60). It was represented by five dummy variables (*newage1* through *newage5*) with the last category (women above 54 years old) designated as the reference group.

TABLE 2
Descriptive Statistics for the Married Women Labor Force Participation Data

Variable Name	*M*	*SD*	*Minimum*	*Maximum*
age[a]				
Full sample	42.547	8.073	30	60
Working women	41.988	7.722	30	60
Non-working women	43.283	8.468	30	60
hc[b]				
Full sample	0.392	0.489	0	1
Working women	0.415	0.493	0	1
Non-working women	0.363	0.482	0	1
inc[c]				
Full sample	20.156	11.619	1.120	96
Working women	18.981	10.564	1.120	91
Non-working women	21.698	12.728	1.500	96
k5[d]				
Full sample	0.238	0.524	0	3
Working women	0.141	0.392	0	2
Non-working women	0.366	1.327	0	3
k618[e]				
Full sample	1.352	1.321	0	8
Working women	1.349	1.317	0	8
Non-working women	1.357	1.327	0	7
lfp[f]				
Full sample	0.567	0.496	0	1
wc[g]				
Full sample	0.281	0.450	0	1
Working women	0.335	0.473	0	1
Non-working women	0.209	0.407	0	1
wg[h]				
Full sample	3.56	2.64	0.13	25.00
Working women	4.17	3.31	0.13	25.00
Non-working women	2.76	0.81	0.99	5.80

Note. $N = 752$. Working women sample: $n = 427$. Non-working women sample: $n = 325$.
[a]Wife's age in years. [b]If husband attended college 1, otherwise 0. [c]Family income excluding wife's wages (by $1,000). [d]Number of children ages 5 or younger. [e]Number of children ages 6 to 18. [f]If wife is in the paid labor force, 1 otherwise 0. [g]If wife attended college 1, otherwise 0. [h]Wife's estimated wage rate.

The logistic model was applied to the MWLFP data using algorithms implemented in SAS®, SPSS®, SYSTAT®, BMDP®, MINITAB®, and STATA®. These algorithms were:

1. The LOGISTIC procedure in SAS Release 8 (SAS Institute, 1999).
2. The LOGISTIC REGRESSION command in SPSS Release 10 (SPSS, 1999a).

3. The LOGIT command in SYSTAT Release 9 (SPSS, 199b).
4. The LR command in BMDP Release 7.1 (BMDP Statistical Software, 1992).
5. The BLOGISTIC command in MINITAB Release 13 (Minitab, 2000).
6. The LOGISTIC command in STATA Release 6 (StataCorp, 1999).

For the purpose of discussion, "statistical package" refers to SAS, SPSS, SYSTAT, BMDP, MINITAB, and STATA software. The term *procedure* refers to a procedure or main command in a statistical package that performs logistic regression, such as the LOGISTIC procedure in SAS, the LOGISTIC REGRESSION command in SPSS, and so forth.

Six statistical packages yielded very similar estimates for parameters and standard errors. All predictors reached the significance level of 0.05, except for *k618*, *hc*, and the interactions of *wc* with *newage1* through *newage5* (Table 3). The result implied that the odds for married women to enter the paid labor force in 1975 were related to the number of young children (5 years old or under), their age groups, and whether the women had some college education.

Data Formats

Three data formats are acceptable for logistic regression. They are (a) the raw data format, (b) the frequency data format, and (c) the covariate pattern (or event/trial) format. These formats are illustrated in Tables 4, 5, and 6 using 10 cases and the first six variables from Table 2. The raw data format in Table 4 records each case (i.e., woman) as a row and her scores on the outcome and predictors as columns. Data stored in this format contain the richest information. McCullagh and Nelder (1989) suggested using this format for logistic regression if the serial order of observations is relevant.

The frequency data format uses a single row to represent multiple observations that share identical outcomes and predictors, whereas the frequency information is stored in a separate variable, such as *count* in Table 5. Thus, the first row in Table 5 replaces Cases 1 and 4 in Table 4. With one or two additional commands, most statistical packages accept data stored in this format (see Appendix B Data requirements).

The covariate pattern format records patterns of predictors by rows. It replaces the outcome variable (*lfp*) with two new variables: *trial* and *event*. *Trial* keeps track of observations that have identical predictor values, whereas *event* records the number of observations having the event outcome; see Table 6. Five cases (2, 5, 7, 8, and 10) from Table 4 shared an identical predictor (covariate) pattern; they form the second row in Table 6. Out of these five women, three were working in 1975. Hence, *event* = 3 and *trial* = 5. Unlike the raw or the frequency data format, the covariate pattern format is not acceptable to either SPSS LOGISTIC REGRESSION or SYSTAT LOGIT, but is acceptable to SPSS PROBIT (see Appendix B Data requirements).

TABLE 3
Summary of the Logistic Regression Results by Six Packages

	SAS LOGISTIC		SPSS LOGISTIC REGRESSION		SYSTAT LOGIT		BMDP LR		MINITAB BLOGISTIC		STATA LOGISTIC	
	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
CONSTANT	−0.5935[a]	0.3325	−0.5935[a]	0.3325	−0.593[b]	0.332	−0.5935[c]	0.332	−0.5935[d]	0.33250	−0.5934696[d]	0.3324873
k5	−1.3710[a]**	0.1955	−1.3709[a]**	0.1955	−1.371[b]**	0.196	−1.371[c]**	0.1960	−1.3710[d]**	0.19550	1.370959[d]**	0.1955443
k618	−0.1097[a]	0.0703	0.1097[a]	0.0703	−0.110[b]	0.070	−0.1097[c]	0.0703	−0.10971[d]	0.07026	−0.1097122[d]	0.0702560
Age												
Newage1	1.8425[a]**	0.4271	1.8425[a]**	0.4271	1.843[b]**	0.427	1.843[c]**	0.4270	1.8425[d]**	0.42700	1.84252[d]**	0.4270950
Newage2	1.6956[a]**	0.4331	1.6956[a]**	0.4331	1.696[b]**	0.433	1.696[c]**	0.4330	1.6956[d]**	0.43310	1.695584[d]**	0.4331121
Newage3	1.1662[a]**	0.4175	1.1662[a]**	0.4175	1.166[b]**	0.417	1.166[c]**	0.4170	1.1662[d]**	0.41750	1.166183[d]**	0.4174900
Newage4	0.9405[a]*	0.3854	0.9405[a]*	0.3854	0.940[b]*	0.385	0.9405[c]*	0.3850	0.9405[d]*	0.38540	0.9404907[d]*	0.3854029
Newage5	0.4613[a]	0.3999	0.4613[a]	0.3999	0.461[b]	0.400	0.4613[c]	0.4000	0.4613[d]	0.39990	0.4613459[d]	0.3999121
Hc	−0.1205[a]	0.1938	0.1205[a]	0.1938	0.120[b]	0.194	0.1205[c]	0.1940	0.1205[d]	0.19380	−0.1204579[d]	0.1938383
Wc	1.3834[a]*	0.6541	1.3834[a]*	0.6541	1.383[b]*	0.654	1.383[c]*	0.6540	1.3834[d]*	0.65410	1.383379[d]*	0.6541175
wc × age												
wc × newage1	−0.5837[a]	0.7544	−0.5838[a]	0.7544	−0.584[b]	0.754	−0.584[c]	0.7540	−0.5837[d]	0.75440	−0.5837369[d]	0.7544270
wc × newage2	−0.7515[a]	0.7788	−0.7515[a]	0.7788	−0.752[b]	0.779	−0.7515[c]	0.7790	−0.7515[d]	0.77880	−0.7515159[d]	0.7787781
wc × newage3	−0.9163[a]	0.7637	−0.9163[a]	0.7637	−0.916[b]	0.764	−0.9163[c]	0.7640	−0.9163[d]	0.76370	−0.9162765[d]	0.7636721
wc × newage4	−0.0448[a]	0.7769	−0.0448[a]	0.7769	−0.045[b]	0.777	0.04478[c]	0.7770	−0.0448[d]	0.77690	−0.447769[d]	0.7769299
wc × newage5	−0.4655[a]	0.7934	−0.4665[a]	0.7934	−0.465[b]	0.793	−0.4655[c]	0.7930	−0.4655[d]	0.79340	−0.4654552[d]	0.7933746
Likelihood ratio test[g]	92.538**		92.538**		92.538**		—		92.538**		92.54**	
Score test[g]	87.547**		87.547**		—		—		—	—		
Wald test[g]	76.8380**		—		—		—		—		—	
AIC	966.078											
SC	1035.238											
Somers's D_{xy}	0.397		—		—		—		0.40			
Goodman–Kruskal Gamma	0.405		—		—		—		0.41			
Kendall's Tau-*a*	0.195		—		—			—		0.19		
c statistic	0.698		—		—		0.6987		—		0.6982	

(continued)

TABLE 3 (Continued)

	SAS LOGISTIC		SPSS LOGISTIC REGRESSION		SYSTAT LOGIT		BMDP LR		MINITAB BLOGISTIC		STATA LOGISTIC	
	Parameter Estimate	*SE*	*Parameter Estimate*	*SE*	*Parameter Estimate*	*SE*	*Parameter Estimate*	*SE*	*Parameter Estimate*	*SE*	*Parameter Estimate*	*SE*
Pearson goodness-of-fit statistic	173.9[h]		751.250[f,j]		751.255[i]		—		173.876[h]		—	
Deviance goodness-of-fit statistic	206.9**[h]		936.078[g,j]**		936.078[i]**		206.885[h]**		206.885[h]**		—	
Hosmer & Lemeshow goodness-of-fit statistic[j]	6.63		7.5899		6.123		4.106		4.788		7.13	
McFadden's Rho^2	—		—		0.090		—		—		—	
Cox & Snell R^2	0.1158		0.116		—		—		—		—	
Nagelkerke R^2 (max-rescaled R^2)	0.1553		0.155		—		—		—		—	
Pseudo R^2	—		—		—		—		—	0.0900		
Brown statistics												
General	—		—		—		0.797		0.797		—	
Symmetric[e]	—		—		—		—		0.000		—	

Note. *SE* = standard error of the parameter estimated. *newage1* = *age* ≤ 34; *newage2* = 34 < *age* ≤ 39; *newage3* < 39 *age* ≤ 44; *newage4* < *age* ≤ 49; *newage5* = 49 < *age* ≤ 54; AIC = Akaike Information Criterion; SC = Schwarz Criterion;

[a]The test of the hypotheses of $\beta = 0$ are based on Wald chi-square. [b]The test of the hypotheses of $\beta = 0$ are based on *t* ratio. [c]The test of the hypotheses of $\beta = 0$ are based on "APPROX. CHI-SQ. REMOVE." [d]The test of the hypotheses of $\beta = 0$ are based on *z* ratio. [e]The *df* and *p* level are not reported by SPSS. We derived the *df* from $N - k - 1$, where k = the number of predictors; the *p* level can therefore be obtained from a chi-square distribution with 737 degrees of freedom. [f]The value is listed as "–2 log likelihood;" the *df* and the *p* level are not reported by SPSS. We derived the *df* from $N - k - 1$, where k = the number of predictors; the *p* level was obtained from a chi-square distribution with 737 degrees of freedom. [g]$df = 14$. [h]$df = 145$. [i]$df = 737$. [j]$df = 8$.

*$p < .05$. **$p < .01$.

TABLE 4
Data Listed in the Raw Data Format

Case	*lfp*	*k5*	*k618*	*age*	*hc*	*wc*
1	1	0	0	30	0	0
2	1	1	2	36	0	0
3	0	1	0	43	1	1
4	1	0	0	30	0	0
5	1	1	2	36	0	0
6	1	1	0	43	1	1
7	0	1	2	36	0	0
8	1	1	2	36	0	0
9	0	2	6	39	0	0
10	0	1	2	36	0	0

Note. This format is based on profiles of 10 women (or cases) extracted from the Married Women Labor Force Participation data. *lfp* = outcome variable; *k5* = number of children under the age of 5; *k618* = number of children between ages 6 and 18; *age* = women's age; *hc* = husband has at least 1 year of college; *wc* = wife has at least 1 year of college.

TABLE 5
Data Listed in the Frequency Data Format

lfp	*k5*	*K618*	*age*	*hc*	*wc*	*count*
1	0	0	30	0	0	2
0	1	2	36	0	0	2
1	1	2	36	0	0	3
0	1	0	43	1	1	1
1	1	0	43	1	1	1
0	2	6	39	0	0	1

Note. This format is based on profiles of 10 women (or cases) extracted from the Married Women Labor Force Participation data. *lfp* = outcome variable; *k5* = number of children under the age of 5; *k618* = number of children between ages 6 and 18; *age* = women's age; *hc* = husband has at least 1 year of college; *wc* = wife has at least 1 year of college.

TABLE 6
Data Listed in the Covariate Pattern (or Events/Trials) Format

k5	*k618*	*age*	*hc*	*wc*	*event*	*trial*
0	0	30	0	0	2	2
1	2	36	0	0	3	5
1	0	43	1	1	1	2
2	6	39	0	0	0	1

Note. This format is based on profiles of 10 women (or cases) extracted from the Married Women Labor Force Participation data. *k5* = number of children under the age of 5; *k618* = number of children between ages 6 and 18; *age* = women's age; *hc* = husband has at least 1 year of college; *wc* = wife has at least 1 year of college.

Different data formats do not affect the estimation of parameters. However, diagnostic statistics are computed differently depending on the data format. Hosmer and Lemeshow (2000) suggested that statistical tests performed on goodness-of-fit chi-squares or diagnostic statistics be based on covariate patterns.

Interactions Between Predictors

Interactions between predictors, such as *wc*newage1* in the logistic model, speak to the multiplicative effect between two or more predictors. Determining if interactions are present in the model is particularly important when one predictor is a risk factor. In this case, the impact of the risk factor on the outcome needs to be estimated accurately. Say a risk factor, for example, truancy, interacts with another predictor, say age, then the slope coefficient of truancy is estimated at each level of age. In this instance, age is referred to as an effect modifier because its presence in the model modifies the impact of truancy on outcomes. Interactions are specified by cross-products in logistic models. BMDP, MINITAB, and STATA impose a model restriction that requires main effects (say, *wc* and *newage1*) be in the model whenever their cross-product is also in the model. SAS imposes this restriction only when a selection method is specified. This restriction may be removed in BMDP and SAS, but not in MINITAB or STATA (see Appendix B specification of model).

EVALUATIONS OF A LOGISTIC REGRESSION MODEL

Evaluations of a logistic regression model include the overall model evaluations, statistical tests of individual predictors, goodness-of-fit statistics, and validations of predicted probabilities. Each is illustrated next for the logistic model.

Overall Model Evaluations

A logistic model is said to provide a better fit to the data if it demonstrates an improvement over the intercept-only model (also called the null model, which has no predictors). Such an improvement is examined by inferential and descriptive statistics. The inferential statistics include three tests: the likelihood ratio, Score, and Wald tests. The likelihood ratio test is a test based on the difference in deviancies: the deviance without any predictor in the model (or the intercept-only model) minus the deviance with all predictors in the model. The Score test is based on the distribution of the *k*-derivatives of the fitted model's likelihood function with regard to all parameters. The Wald test is obtained from a vector–matrix calculation that involves the parameter vector, its transpose, and the inverse of its variance matrix (Hosmer & Lemeshow, 2000). All three test statistics are distributed as chi-squares with degrees of freedom equal to the number of predictors. For these data, these test results are similar as far as significance levels are concerned (Table 3). Among the six packages examined, BMDP computes none of these statistics, whereas SAS computes all three (see Appendix B). Two descriptive statistics—the Akaike Information Criterion and the Schwarz Criterion—may be used to compare different models derived from the same sample (SAS, 1999), different models from different samples, and nested or nonnested models (Long, 1997). A smaller value, including negative values, implies a better model fit. Only SAS computes these two indexes.

Statistical Tests of Individual Predictors

Individual parameter estimates are tested by the likelihood ratio test, the Wald statistic, or the Score test. The likelihood ratio test is a test based on the difference in deviancies: the deviance without the predictor in the model minus the deviance with the predictor in the model. The Wald statistic is formed from the ratio of the estimated slope parameter over its standard error. According to Jennings (1986), Long (1997), and Tabachnick and Fidell (2001), the likelihood ratio test is more powerful than the Wald test, whereas the Score test is a normal approximation to the likelihood ratio test. BMDP is the only package that performs the likelihood ratio test. The other five compute the Wald test, although SPSS LOGISTIC REGRESSION and STATA LOGISTIC carry out the likelihood ratio test during stepwise logistic modeling (see Appendix B Results). For categorical predictors, MINITAB, SAS, and SPSS automatically perform an overall test of design variables transformed from the same categorical predictor. In SYSTAT, the same test is requested by the CONSTRAINT subcommand. This subcommand may also be applied to test two or more slope parameters simultaneously against zero. Such a test is carried out in SAS LOGISTIC by the TEST statement. Both CONSTRAINT and TEST work also in stepwise modeling if the multiple predictors, to be tested simultaneously, are already selected into the model.

If all observations could be perfectly or nearly perfectly separated by one or several of the predictors via a linear combination, there is no need for the logistic model. Hence, the maximum likelihood estimates become nonunique and infinite in this rare, but special, condition. Complete or quasicomplete separation is most likely to occur with small data sets. SAS is the only package that prints a warning on the output when a complete or quasicomplete separation is detected. Other packages render clues on this problem by producing unusually large parameter estimates or standard errors for predictors on which data are completely or quasicompletely separated.

Goodness-of-Fit Statistics

Goodness-of-fit statistics assess the fit of a logistic model against the data. Four inferential tests and four descriptive measures are provided by six packages we reviewed.

Inferential tests. The four tests are the Brown chi-square test, the Pearson chi-square test, the deviance-based test, and the Hosmer–Lemeshow (H–L) test. The Brown test treats a model's fit as a special case of Prentice's family of generalized response models (Brown, 1982; Prentice 1976). The generalized response model has two parameters and can be used to model many of the possible relations between predictors, X, and the probability of a positive binary outcome, $P(X)$. When both parameters equal one, the generalized response model is a logistic model. The null hypothesis for the Brown test states that both parameters equal one; this implies that the logistic model is adequate. Thus, for this data, the nonsignificant Brown statistic (= 0.797, $df = 2, p = 0.671$) implied that the logistic model fit the data as well as the extended model. This result was obtained from BMDP and MINITAB. Furthermore, MINITAB provides an additional test (with $df = 1$) for symmetric alternative models. This test was also nonsignificant ($df = 1, p = 0.984$). Thus, it was concluded that the logistic model sufficiently explained the data.

The Pearson chi-square and deviance-based goodness-of-fit statistics are computed by most packages (see Appendix B results). For the logistic model, the two statistics are reported in Table 3. Insignificant statistics imply a good fit of the model. Only when these statistics are calculated from covariate patterns and the number of observations in each covariate pattern is mostly greater than one, can they be regarded as indexes of goodness-of-fit (Hosmer & Lemeshow, 2000; McCullagh & Nelder, 1989). Unfortunately, SPSS and SYSTAT calculate both statistics from the raw data. Because the chi-square distribution is an m-asymptotic approximation to the true sampling distribution only when m (= number of observations per covariate pattern) is sufficiently large (Hosmer & Lemeshow, 2000, pp. 145–147), researchers should not assess these two statistics computed from either SPSS or SYSTAT against a chi-square distribution.

The H–L statistic is a Pearson chi-square statistic, calculated from a $2 \times g$ table of observed and estimated expected frequencies, where g is the number of groups formed from the estimated probabilities. Ideally, each group should have an equal number of observations. The H–L statistics for the logistic model, calculated by different packages, range from 4.106 to 7.5899 (Table 3). None of these values reached significance at $\alpha = 0.05$ on a χ^2 distribution with 8 degrees of freedom. This indicates that the model fits the MWLFP data well. Differences in the H–L statistic were attributable to the way ties on estimated probabilities were handled. The H–L statistic is routinely reported by BMDP. It must be requested by researchers in other packages. MINITAB, SYSTAT, and STATA permit researchers to specify the number of groups used in the calculation. Hosmer and Lemeshow (2000) suggested that no fewer than 6 groups be employed; 10 are commonly used. SYSTAT further allows researchers to establish cutoff points on estimated probabilities by which groups are formed.

There are limitations with the H–L test. First, the test is conservative, lacking statistical power in certain cases to detect a model's poor fit. Second, even when the test is significant, indicating that a model does not fit the data well, it does not shed light on where and why data are not well fitted by the model. According to Hosmer and Lemeshow (2000, pp. 150–151) and Ryan (1997, p. 279), the H–L statistic is too conservative to reject the null hypothesis when groups are fewer than six or expected cell frequencies are less than five.

Descriptive measures. Four descriptive measures of goodness-of-fit are provided by all packages we examined. They are variations of the R^2 concept defined for the OLS regression model (Table 3). R^2 has a clear definition in linear regression, that is, the proportion of the variation in the dependent variable that can be explained by predictors in the model. Numerous formulas have been devised to yield an equivalent of this concept for the logistic model. None, however, renders the meaning of variance explained (Long, 1997, pp. 104–109; Menard, 2000). Furthermore, none corresponds to predictive efficiency and none can be tested in an inferential framework (Menard, 2000).

Among the various R^2 analogs proposed for logistic regression, Menard's (2000) empirical study seemed to suggest that McFadden's (1973) Rho^2 is preferred over others. It is implemented in SYSTAT only and defined as the difference between the initial and the model −2 log-likelihood statistics, divided by the initial −2 log-likelihood. The McFadden Rho^2 is conceptually similar to the OLS R^2, relatively independent from the base rate, and comparable across models that comprise different predictors, yet the same outcome variable. It is not necessarily linearly related to the percentage of correct classifications in empirical studies (Menard, 2000). SAS and SPSS provide two R^2 indexes defined by Cox and Snell (1989) and Nagelkerke (1991), respectively. STATA computes a pseudo R^2.

Validations of Predicted Probabilities

As was explained earlier, binary logistic regression predicts the logit of an event outcome by a set of predictors. Because the logit is the natural log of the odds, or probability/(1 – probability), it can be transformed back to the probability scale and become the predicted result of logistic regression. The predicted probabilities can be revalidated with the actual outcome to determine if high probabilities are indeed associated with events and low probabilities with nonevents. The degree to which predicted probabilities match with actual outcomes is expressed either as a measure of association or a classification table. There are altogether four measures of association and three classification tables that are provided by the six packages.

Measures of association. The four measures are Tau-a, Gamma, Somers's D statistic, and the c statistic. The Tau-a statistic is Kendall's rank-order correlation coefficient without adjustments for ties. The Gamma statistic is based on Kendall's coefficient but adjusts for ties. Gamma is more useful and appropriate than Tau-*a* when there are ties on both outcomes and predicted probabilities, as was the case with the MWLFP data. The Gamma statistic for the model is 0.405. It is interpreted as 40.5% fewer errors made in predicting which of two women participated in the labor force by utilizing the estimated probabilities than by chance alone (Demaris,1992). There are two problems associated with the Gamma statistic: (a) It has a tendency to overstate the strength of association between estimated probabilities and outcomes (Demaris, 1992), and (b) a value of zero does not necessarily imply independence when the data structure exceeds a 2 × 2 format (Siegel & Castellan, 1988).

Somers's *D* is an extension of Gamma—a better index too—whereby one variable is designated as the dependent variable and the other the independent variable (Siegel & Castellan, 1988). There are two asymmetric forms of Somers's *D* statistic: D_{xy} and D_{yx}. Only D_{yx} correctly represents the degree of association between the outcome (y), designated as the dependent variable, and the estimated probability (x), designated as the independent variable (Demaris, 1992). SAS and MINITAB compute only D_{xy}, (Table 3), although this incorrect index can be corrected to D_{yx} in SAS (Peng & So, 1998). For BMDP, SPSS, and SYSTAT, the D statistics may be computed by the 4F, CROSSTAB, and XTAB procedures, respectively. Both D_{xy} and D_{yx} may be used to compare the fit of different models—the greater the Somers's *D*, the better the fit.

The *c* statistic is the proportion of observation pairs with different observed outcomes for which the model correctly predicts a higher probability for observations with the event outcome than the probability for nonevent observations. For this model, the *c* statistic is 0.689. This means that for 68.9% of all pairs of women, one worked in 1975 and the other did not, the model correctly assigned a higher probability to working women. The *c* statistic ranges from 0.5 to 1. A 0.5 value means that the model is no better than assigning observations randomly into outcome categories. A value of 1 means the model assigns higher probabilities to all observations with the event outcome, compared to nonevent observations. SAS routinely reports the *c* statistic; BMDP and STATA compute this index by request. It is obtained in MINITAB, SPSS, and SYSTAT by dividing Somers's D_{xy} with 2 and adding 0.5 (SAS,1999; SPSS, 1999a).

Classification tables. There are three types of classification tables to show the validity of predicted probabilities: (a) the prediction success table, (b) the histogram of predicted probabilities, and (c) the two-way classification table. The prediction success table, available only in SYSTAT, assigns each observation into an outcome category according to its probability of belonging in that category. Hence, entries in this table are probabilities of belonging, rather than frequencies. Even though it is not feasible to count the number of observations correctly classified in a prediction success table, the table can be used to evaluate the percentage of correct predictions for each possible outcome. The histogram of predicted probabilities is based on predicted probabilities and outcomes (i.e., event vs. nonevent). It is available in SPSS LOGISTIC REGRESSION with the CLASSPLOT command, or BMDP LR with the PRINT HISTOGRAM command. Researchers can inspect such a histogram for overlapping probabilities between outcome categories. Different cutoffs on the predicted probability impact the percentage of false positive and false negative classifications.

The third classification table is a 2 × 2 table in which rows represent the two possible outcomes and columns are high and low probabilities, based on a cutoff point. The cutoff point is specified by researchers or set at 0.5 by statistical packages. SAS prepares this classification table by using a reduced-bias algorithm, whereas SPSS, SYSTAT, BMDP, and STATA do not. The algorithm minimizes the bias of using same observations both for model fitting and for predicting probabilities (SAS, 1995, 1999). Consequently, sensitivity, specificity, percentage of correct classification, false positive, and false negative presented in SAS classification tables are less biased than those computed from other packages. In case that the event outcome is overrepresented in the sample, researchers may specify the PEVENT= option in SAS LOGISTIC to denote a prior probability that is the population proportion of events. The PEVENT specification corrects the calculation of sensitivity, specificity, false positives, false negatives, and percentage of correct classifications because these values are considered posterior probabilities by Bayes's theorem. However, the specification of prior probabilities has no impact on the estimation of parameters or the evaluation of the model.

In the opinion of Hosmer and Lemeshow (2000), "the classification table is most appropriate when classification is a stated goal of the analysis; otherwise it should only supplement more rigorous methods of assessment of fit" (p. 160).

Graphing Prediction Accuracy

One primary goal of performing logistic regression is to generate an equation that can reliably classify observations into one of two outcomes. The degree to which predictions agree with the data may be shown graphically by either a receiver operating characteristic (ROC) curve or an overlay plot of sensitivity and specificity versus predicted probabilities (Afifi & Clark, 1990; Hosmer & Lemeshow, 2000, pp. 160–164).

The ROC curve is a plot of sensitivity versus 1_minus_specificity. Sensitivity is defined as the proportion of observations correctly classified as an event. It is also called the true positive fraction. Specificity is defined as the proportion of observations correctly classified as a nonevent. Hence, 1_minus_specificity is the proportion of observations misclassified as an event; it is also called the false positive fraction. Both sensitivity and 1_minus_specificity change as a function of the cutoff used on estimated probabilities. Because each cutoff yields one sensitivity

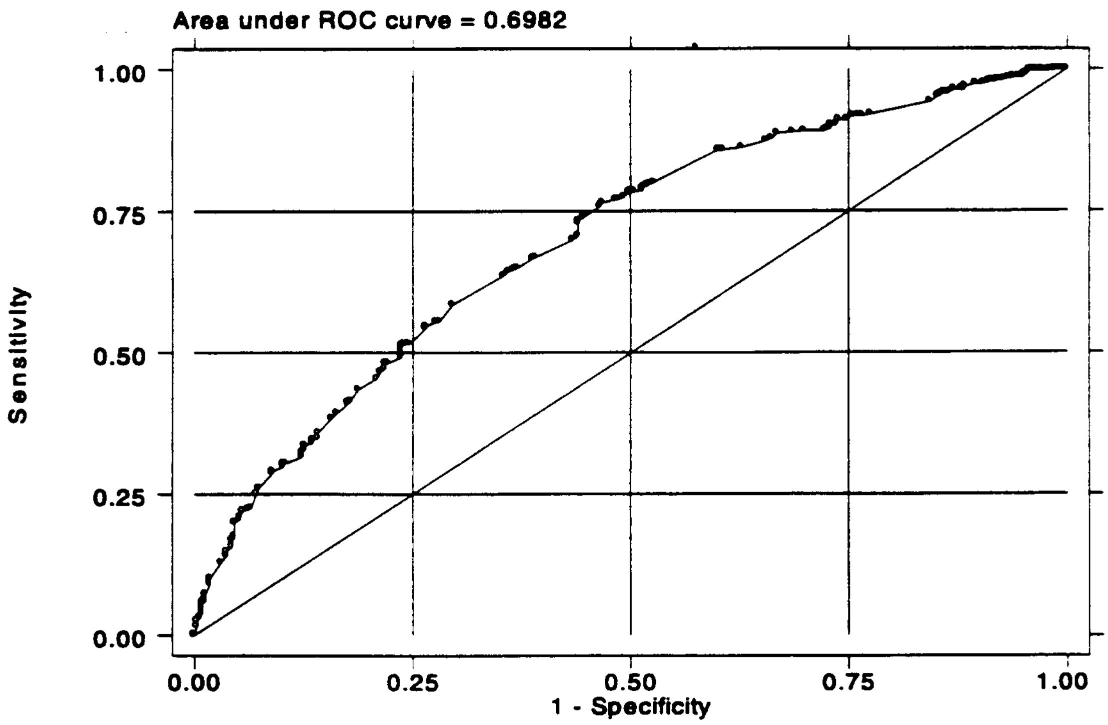

FIGURE 2 Receiver operating characteristic curve generated by STATA LOGISTIC. Area under the ROC curve = 0.6982.

and one specificity, the ROC curve is a plot based on multiple cutoffs (Figure 2). This figure suggests that in order for the logistic model to correctly classify a large proportion (say 75%) of married women who worked in 1975, it has to misclassify approximately 45% of nonworking women as working women. Researchers contemplating competing models for their data can rely on a ROC curve for an informed decision. The model with a larger area below the ROC curve (i.e., a larger *c* statistic) is considered a better model (Afifi & Clark, 1990). Alternatively, the one with the greatest height on the ROC curve at a desirable probability cutoff should be chosen (Afifi & Clark, 1990). In other words, the best model is the one associated with the greatest sensitivity and the lowest 1_minus_specificity.

Three packages are capable of plotting the ROC curve: BMDP, STATA, and SAS. BMDP LR produces the curve with the PRINT command and PLOT/COST options, whereas the curve is requested in STATA LOGISTIC with the LROC command. In SAS LOGISTIC, researchers first establish a data set with the OUTROC= option, then submit this data set to the GPLOT procedure to plot the ROC curve (Peng & So, 1998). It is worth noting that sensitivity and specificity used in plotting the ROC curve are not corrected for bias by these three packages.

The second plot—an overlay plot of sensitivity and specificity against probability cutoffs—is useful for determining an appropriate cutoff for future classifications (Figure 3 based on the logistic model). The point at which two curves intersect is an optimal cutoff. The intersecting point treats two groups (i.e., working women and nonworking women) equally in terms of the proportion of correct classifications.

For this model, the optimal cutoff of 0.5848 yielded approximately 64% of correct classifications for both groups. Although Figure 3 is drawn by STATA with the LSENS command, BMDP can generate this plot by the PRINT command with COST and PLOT options. SAS too is capable of plotting Figure 3 with a simple data manipulation step and the GPLOT procedure (Peng & So, 1998).

OUTLIERS AND DIAGNOSTIC STATISTICS

In addition to the multiple indexes discussed so far, researchers should also include diagnostic analyses of any logistic regression model. The purposes of diagnostic analyses are two fold: (a) to identify potential outliers and (b) to understand the model's poor fit to certain observations. These analyses are carried out by a series of diagnostic statistics proposed by Pregibon in 1981. These statistics should be computed from covariate patterns. Only when the number of unique covariate patterns is much smaller than the number of observations is the examination of diagnostic statistics meaningful. For this logistic model, 160 unique covariate patterns were identified with 14 predictors. Hence, it was appropriate to carry out diagnostic analysis of the result.

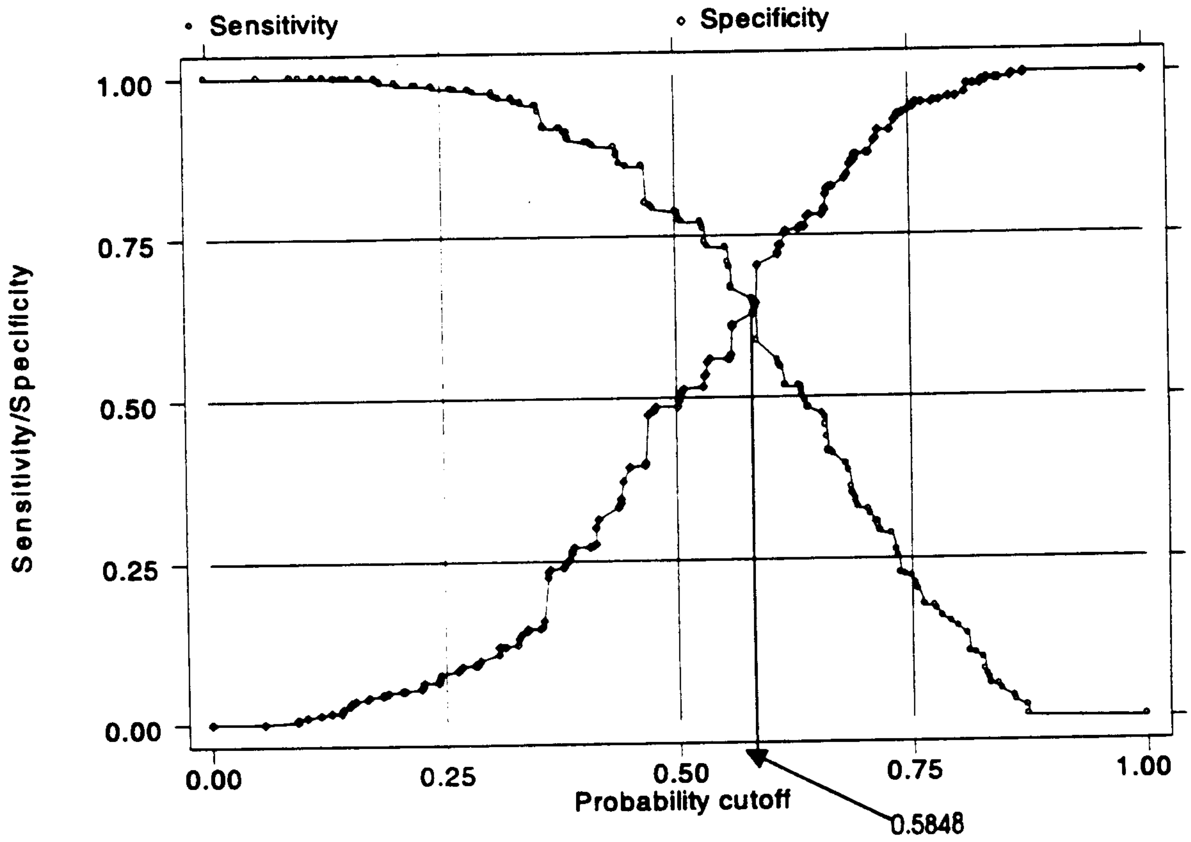

FIGURE 3 Overlay plot of sensitivity and specificity versus various probability cutoffs generated by STATA LOGISTIC. The two plots intersect at $p = 0.5848$ that represents the most optimal cutoff on probabilities.

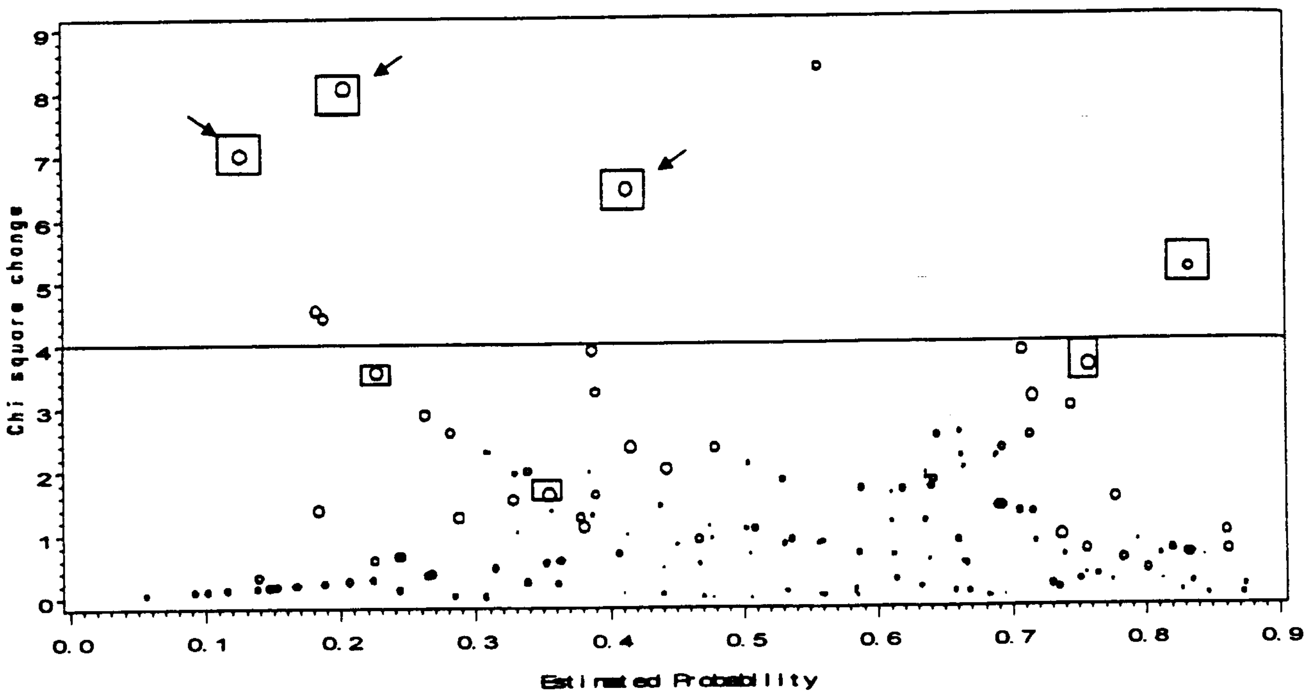

FIGURE 4 Plot of change in Pearson chi-square versus estimated probabilities generated by MINITAB BLOGISTIC.

Diagnostic statistics available from statistical software include Pearson and deviance residuals, change in the Pearson chi-square statistic and change in the deviance, the change in parameter estimates due to a particular covariate pattern deleted, the hat matrix diagonal, and confidence interval displacement diagnostics (see Appendix B Results and Diagnostic statistics).

The Pearson residual and the deviance residual are components of the Pearson chi-square statistic and of the deviance respectively. A "large" value of either statistic is indicative of a poorly explained covariate pattern. Because *large* is a relative term, researchers must rely on their own judgement in deciding if a particular residual is large, compared to other residuals. For this reason, the Pearson residual and the deviance residual are not as informative as their respective changes. The change in Pearson chi-square is defined as the difference in Pearson chi-square due to the deletion of a particular covariate pattern. A large value once again indicates that the corresponding covariate pattern contributes heavily to the disagreement between data and predicted probabilities. In fact, a change exceeding four signals an ill-fit covariate pattern. The selection of four as a criterion is based on the critical value (i.e., 3.841) of the chi-square distribution with one degree of freedom and an alpha level of 0.05. Using this criterion, we identified seven covariate patterns, or 23 observations, that were poorly fit by the logistic model. These are shown in Figure 4. Similarly, the change in deviance due to the deletion of a particular covariate pattern signals the possibility of an outlier, if the change is large relative to other changes. Unlike the change in Pearson chi-square, there is no recommended criterion by which one can judge changes in deviance. We chose four for the same reason as four was used to assess the change in Pearson chi-square. Our decision was justified by the fact that the chi-square approximation for differences in deviance is adequate, although it is inadequate for deviancies themselves (Ryan, 1997, p.270). Figure 5, based on change in deviance, identifies five covariate patterns, or 30 observations, that were poorly explained by the present model.

The change in standardized parameter estimates ($\Delta\beta$), due to the omission of a covariate pattern, is another useful way of locating ill-fit covariate patterns because a "large" change signals instability in estimates. Figure 6 is a display of changes in the Pearson chi-square statistic versus estimated probabilities using $2.5 \times |\Delta\beta|$ as the size of the plotting symbol. This plot reveals that seven covariate patterns (marked by a square), or 36 observations, were associated with 0.20 or greater standardized changes in the *k5* coefficient estimate. On the basis of Figures 4, 5, and 6, we identified three covariate patterns or 11 observations to be outliers. They are highlighted by arrows in Figure 6. These potential outliers need to be investigated further to determine if they are incorrectly coded or the model is misspecified for them. Whatever the cause may be, outliers should not be discarded solely to improve the fit of a model (Long, 1997, p. 99).

Hat matrix diagonals refer to diagonal elements in the hat matrix (Hosmer & Lemeshow, 2000). A large hat matrix diagonal discloses those covariate patterns that are unusually influential in the covariate pattern space. When a predicted probability is extreme, say less than 0.1 or greater than 0.9, its hat matrix diagonal may be greatly reduced; hence, it is not a valid indicator of the model's fit (Hosmer & Lemeshow, 2000, pp. 171–172).

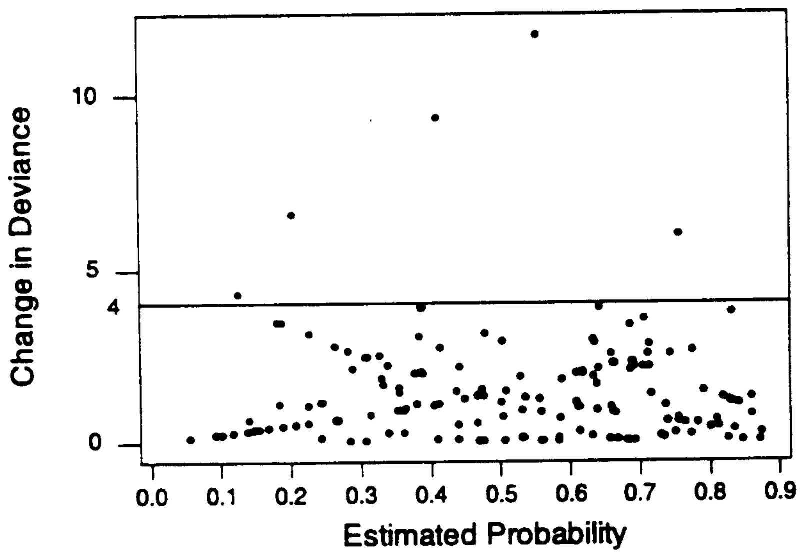

FIGURE 5 Plot of change in deviance versus estimated probabilities generated by MINITAB BLOGISTIC.

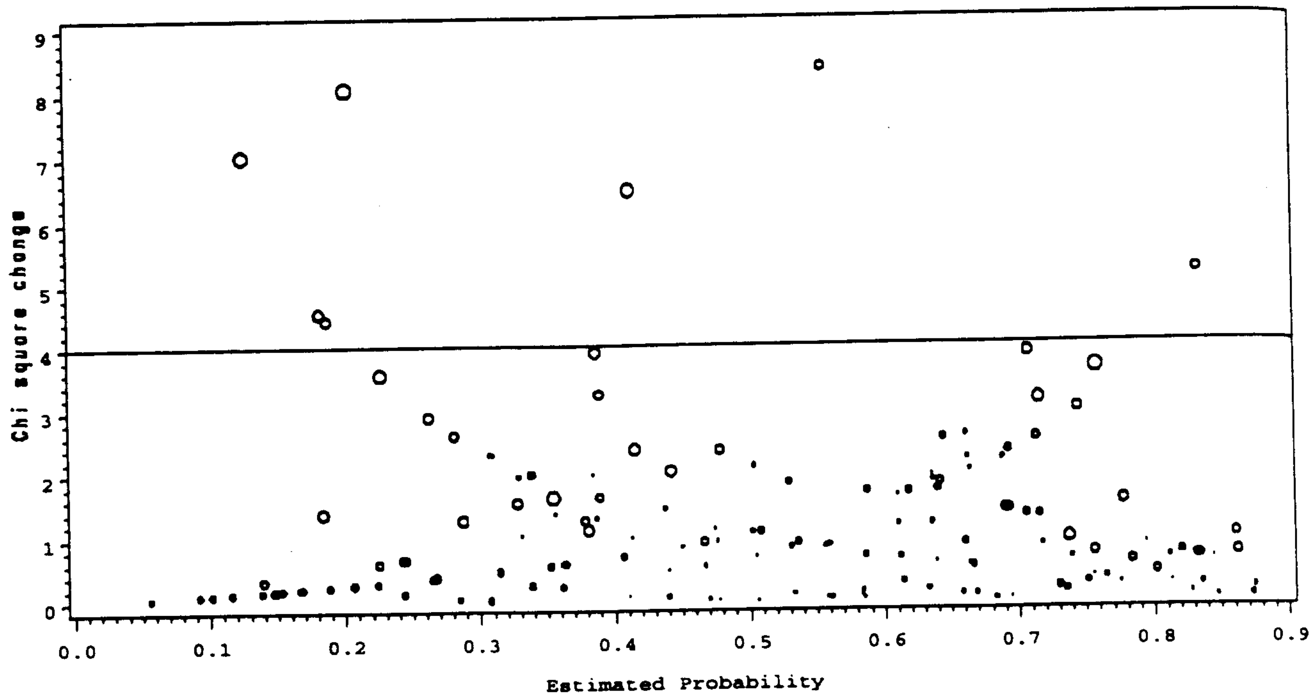

FIGURE 6 Plot of change in Pearson chi-square versus estimated probabilities with the size of circles representing the absolute, standardized change in parameter estimates of *k5*, multiplied with 2.5, generated by SAS GPLOT.

The confidence interval displacement diagnostics provide scalar measure of the influence of individual observations on the parameter estimates. These confidence interval displacement diagnostics are based on the same ideas as the Cook distance in linear regression diagnostics. A plot of these diagnostics against observations reveals observations that exercise a large influence over parameter estimates. Hence, these observations are candidates for potential outliers.

A summary of diagnostic statistics is given in Appendix B under Results and Diagnostics statistics. A note of caution is issued here: The calculation of diagnostic statistics by SPSS and SYSTAT is always based on individual observations. It is therefore inappropriate to evaluate these diagnostic statistics when the number of covariate patterns is much smaller than the number of observations. The computation of diagnostic statistics by SAS depends on the data format. BMDP, MINITAB, and STATA always base the computation of diagnostic statistics on covariate patterns. Appendix C summarizes capabilities of the six packages for graphing diagnostic statistics. Besides Figures 4, 5, and 6 discussed so far, Long (1997) and SAS (1999) suggested that index plots also be used to search for poorly fit observations. SAS is the only software that has an option (i.e., IPLOT) for generating index plots. The other five packages use two procedures—the logistic regression and the plot—to draw index plots. Index plots are examined in the same manner as Figures 4 to 6.

The standardized Pearson residual index plot, suggested by Long (1997, p. 99), is a plot of standardized residuals. Hence, any observation outside the bounds of

±2 may be considered a potential outlier (Figure 7). The IPLOT option in SAS LOGISTIC does not generate such a plot. Researchers can circumvent this limitation by first calculating standardized Pearson residuals by dividing the Pearson residuals with leverages (i.e., diagonals of the hat matrix). Then input the standardized Pearson residuals into the GPLOT procedure to draw this plot.

MODELING STRATEGY AND REPORTING

An ideal modeling approach in logistic regression is to consider and contrast all models that are theoretically significant and practically important. This course of action is not feasible for most researchers. As an alternative, we recommend the following steps:

1. Perform a descriptive analysis of each predictor and its relation with the outcome variable. Results from this initial analysis provide much insight into potentially viable models for the data.
2. Properly transform categorical predictors by a set of design variables and include the design variables in logistic models, in lieu of categorical predictors.
3. Correctly identify the event of the outcome and model its probability by a series of univariate logistic regressions, each based on a single predictor.

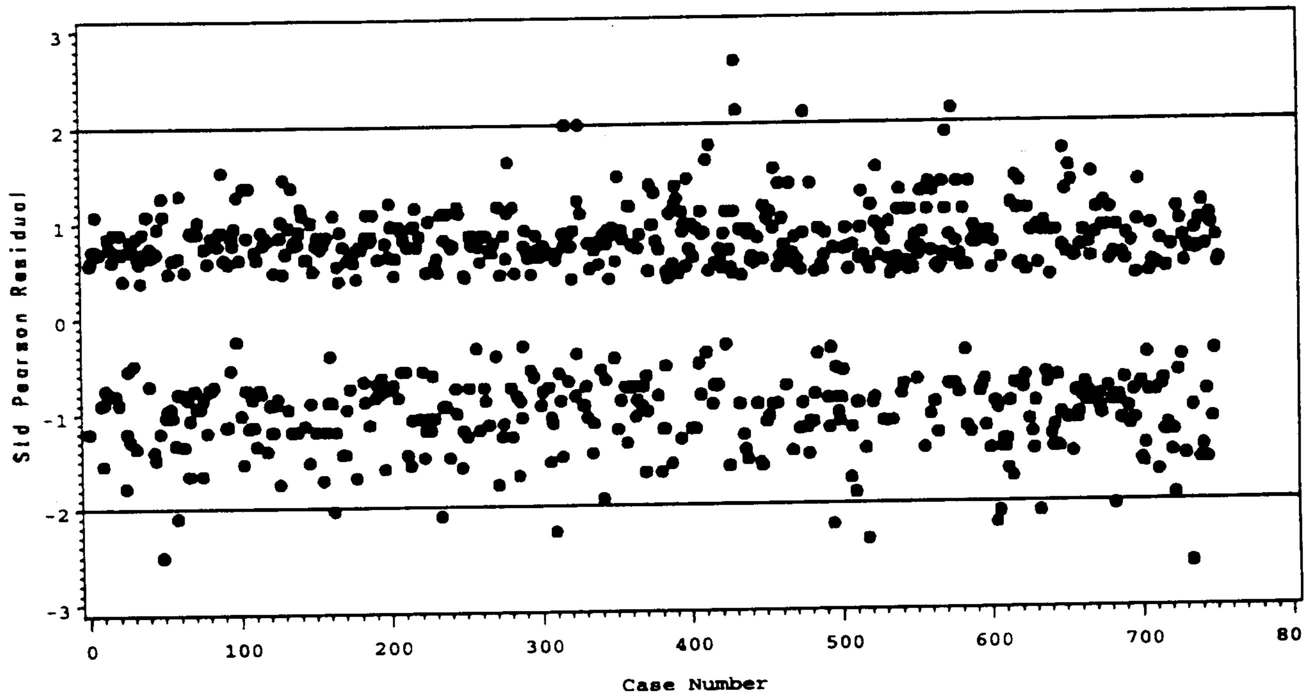

FIGURE 7 Index plot of standardized Pearson residual versus case numbers generated by SAS GPLOT.

4. Based on results from univariate analyses (as described in number 3), fit a preliminary multivariate logistic model using all predictors that are of importance or of interest to the researcher.

5. Fit alternative models to data. Alternative models may be derived from the preliminary multivariate model by (a) adding two-way, three-way, and so on, interactions of significant main predictors; (b) creating polynomial trends from continuous predictors; (c) performing nonlinear transformations (e.g., log, square root, etc.) of continuous predictors; or (d) removing unimportant or statistically insignificant predictors.

6. Compare the performance of alternative models with that of the preliminary multivariate model in terms of models' overall test of all parameters, interpretability and statistical significance of each predictor, goodness-of-fit statistics, predictive power, accuracy of prediction, and diagnostic results. A good model surpasses competing models in more areas than one. Thus, a researcher needs to gather as much information as possible before accepting a model as the best model for the data.

In presenting logistic regression results, researchers should include comprehensive information similar to those portrayed in Table 3, Figures 2, 3, 6, and 7. The decision to accept one model to be the "best fit" among its competitors should be justified by multiple indicators, including the model's overall test of all parameters, interpretability and statistical significance of each predictor, goodness-of-fit statistics, predictive power, accuracy of prediction, and diagnostic results. Cross-validating the best-fit model with other samples increases the generalizability of findings, although exploratory models cannot be replicated exactly. Last but not the least, researchers should pay attention to mathematical definitions of statistics generated by any statistical package. Among the six packages we reviewed, none was found to be error-free. A reference to the software should inform readers of programming mistakes and limitations and help researchers verify results with another package.

EVALUATIONS OF SIX PACKAGES FOR LOGISTIC REGRESSION

Even though, for most studies, commercially available statistical packages yield similar estimates of parameters and standard errors (Long, 1997), they differ in flexibility of data formats, model building strategies, treatment of interaction terms, goodness-of-fit and diagnostic statistics calculated, and graphic capabilities. In this section, attention is focused on computing algorithms, function, and relative strengths and weaknesses of each statistical package.

Computing Algorithms

The Newton–Raphson algorithm is used by SAS, SPSS, BMDP, and STATA, the iterative-reweighted least squares algorithm is used by MINITAB, and the Gauss–Newton algorithm is used by SYSTAT. The default algorithm in SAS is actually the Fisher-scoring algorithm; the Newton–Raphson algorithm is used only when the Fisher-scoring algorithm fails to converge. In rare cases when one of these algorithms does not converge, researchers should turn to alternative algorithms to circumvent this problem (Long, 1997).

Function

Six logistic regression procedures provided helpful statistics. All features common or unique to these procedures are summarized in Appendix B. For building and selecting the best logistic model, only five procedures were compared. MINITAB was excluded because it did not provide a selection method. Readers should pay special attention to statistics computed at each stage of the selection process, as they directly affect the kind of final "best" model identified by each package.

Relative Strengths and Weaknesses

An ideal statistical package for logistic regression should be user-friendly and comprehensive in its options and output. Each package we examined possesses certain features of this "ideal" package. We recommend the versatile SAS LOGISTIC and BMDP LR for researchers experienced with logistic regression techniques and programming. Diagnostic statistics in SAS LOGISTIC are based either on observations or covariate patterns. Several unique goodness-of-fit indexes and selection methods are provided in SAS. Its ability to fit a broad class of binary response models, plus its provision to correct for oversampling, overdispersion, and bias introduced into predicted probabilities, sets it apart from the other five. BMDP LR performs logistic regression on covariate patterns. It is a stepwise procedure that provides the greatest flexibility in selecting the best set of predictors, under the hierarchical modeling restriction. Unfortunately, diagnostic analyses suggested by Hosmer and Lemeshow (2000) cannot be performed in BMDP as it does not compute such statistics as: change in Pearson chi-square, change in deviance, or change in parameter estimates.

If either SPSS LOGISTIC REGRESSION or SYSTAT LOGIT is the only package available, researchers must be aware that both compute the goodness-of-fit and diagnostic statistics from individual observations. Consequently, these statistics are inappropriate for statistical tests. With dazzling graphic interfaces, both packages are user-friendly. They provide several goodness-of-fit indexes and selection methods. However, in both SPSS LOGISTIC REGRESSION and SYSTAT LOGIT, Pearson or deviance goodness-of-fit statistics, and diagnostic statistics are calculated from individual observations, rather than covariate patterns. They should not be readily interpreted as chi-square values.

MINITAB BLOGISTIC is the simplest to use. It adopts the hierarchical modeling restriction in direct modeling (see Appendix A). Diagnostic statistics are calculated correctly from covariate patterns. A substantial number of goodness-of-fit indexes are available including the unique Brown statistic. Diagnostic graphics suggested by Hosmer and Lemeshow (2000) are programmed as subcommands. However, the absence of predictor selection methods may make it less appealing to some researchers.

STATA LOGISTIC provides the most detailed information on parameter estimates, yet its goodness-of-fit indexes are limited. Its command language is easy to learn. It generates high quality graphics with a single command. Diagnostic statistics are calculated correctly. Model selections are carried out in two procedures: SW for stepwise selection and LOGISTIC for logistic regression modeling. Multicollinearity among predictors is examined automatically during stepwise modeling. When checking multicollinearity, STATA defines predictors broadly as terms that can refer to a single predictor, an interaction between predictors, or a series of dummy variables grouped by parenthesis. STATA examines multicollinearity in predictors both within and between terms. We recommend MINITAB and STATA for beginners, although experienced researchers may also employ them for logistic regression.

SUMMARY

Logistic regression has been gaining popularity among social sciences researchers with the wide availability of sophisticated statistical software that performs comprehensive analyses of this technique. Yet a recent review of 52 articles, published between 1988 and 1999 in three higher educational journals, revealed lack of standards in the analysis and reporting of logistic regression (Peng et al., in press). The level of completeness and accuracy of supplementary analyses was uneven across studies. Thus, we feel that there is a need to provide a primer on logistic regression for researchers, editors, and journal readers. Specifically, this article was written to illustrate the implementation of direct logistic regression modeling and its supplementary evaluations. A real-world data set was analyzed by a complex logistic model to explain the likelihood of women entering the paid labor force based on their educational and demographical information. Results were evaluated and diagnosed in terms of the overall test of all parameters, interpretability, and statistical significance of each predictor, goodness-of-fit statistics, predictive power, accuracy of prediction, and identification of potential outliers. Guidelines are offered for modeling strategies and reporting standards in logistic regression. Furthermore, six statistical packages were employed to perform logistic regression. Their strengths and weaknesses were noted in terms of flexibility, accuracy, completeness, and usefulness.

ACKNOWLEDGMENT

We wish to thank Gary M. Ingersoll, Lisa Kurz, Dan J. Mueller, Edward St. John, Dale Weigel, Jin Zhu, and Larry Hoezee for their helpful comments on earlier drafts of this article.

REFERENCES

Afifi, A. A., & Clark, V. (1990). *Computer-aided multivariate analysis* (2nd ed.). New York: Van Nostrand Reinhold.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. *Higher Education: Handbook of Theory and Research, 8,* 379–410.

BMDP Statistical Software. (1992). *BMDP statistical software manual* (Vol. 2). Los Angeles: Author.

Brown, C. C. (1982). On a goodness of fit test for the logistic model based on score statistics. *Communication in Statistics: Theory and Method, 11,* 1087–1105.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. *Higher Education: Handbook of Theory and Research, X,* 225–256.

Chuang, H. L. (1997). High school youth's dropout and re-enrollment behavior. *Economics of Education Review, 16,* 171–186.

Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. *Journal of Health and Social Behavior, 25,* 334–348.

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). London: Chapman & Hall.

Demaris, A. (1992). "Logit modeling: Practical applications," in *Sage University Paper Series on Quantitative Applications in the Social Sciences,* 07–086. Newbury Park, CA: Sage.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association, 70,* 892–898.

Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: Wiley.

Janik, J., & Kravitz, H. M. (1994). Linking work and domestic problem with police suicide. *Suicide and Life Threatening Behavior, 24,* 267–274.

Jennings, D. E. (1986). Judging inference adequacy in logistic regression. *Journal of the American Statistical Association, 81,* 471–476.

Kleinbaum, D. G. (1994). *Logistic regression: A self-learning text.* New York: Springer-Verlag.

Lei, P.-W., & Koehly, L. M. (April, 2000). *Linear discriminant analysis versus logistic regression: A comparison of classification errors.* Paper presented at the 2000 annual meeting of American Educational Research Association, New Orleans, LA.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105–142). New York: Academic.

Menard, S. (1995). Applied logistic regression analysis. in *Sage University Paper Series on Quantitative Applications in the Social Sciences,* 07–106. Thousand Oaks, CA: Sage.

Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician, 54,* 17–24.

Minitab. (2000). *MINITAB user's guide 2: Data analysis and quality tools, Release 13.* State College, PA: Author.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica, 55,* 765–799.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691–692.

Okun, M. A., Benin, M., & Brandt-Williams, A. (1996). Stay in college: Moderators of the relation between intention and institutional departure. *Journal of Higher Education, 67,* 577–596.

Peng, C. Y., & So, T. S. (1998). If there is a will, there is a way: Getting around defaults of PROC LOGISTIC in SAS. *Proceedings of the MidWest SAS Users Group 1998 Conference,* 243–252. Retrieved July 23, 2001 from http://php.indiana.edu/~peng/articles/mwsug98.pdf

Peng, C. Y., & So, T. S. H. (in press). Modeling strategies in logistic regression. *Journal of Modern Applied Statistical Methods.*

Peng, C. Y., So, T. S., Stage, F. K., & St. John, E. P. (in press). The use and interpretation of logistic regression in higher education journals: 1988–1999. *Research in Higher Education.*

Peterson, T. (1984). A comment on presenting results from logit and probit models. *American Sociological Review, 50,* 130–131.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics, 9,* 705–724.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics, 32,* 761–768.

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association, 73,* 699–705.

Ryan, T. P. (1997). *Modern regression methods.* New York: Wiley.

SAS Institute. (1995). *Logistic regression examples using the SAS® system, Version 6.* Cary, NC: Author.

SAS Institute. (1999). *SAS/STAT® user's guide, Version 8, Vol. 2.* Cary, NC: Author.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral science* (2nd ed.). New York: McGraw-Hill.

SPSS. (1999a). *SPSS® regression models 10.* Chicago: Author.

SPSS. (1999b). *SYSTAT® 9.0 Statistics 1.* Chicago: Author.

StataCorp. (1999). *STATA® reference manual release 6* (Vols. 1–4). College Station, TX: Author.

St. John, E. P., Paulsen, M. B., & Starkey, J. B. (1996). The nexus between college choice and persistence. *Research in Higher Education, 37,* 175–220.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Tolman, R. M., & Weisz, A. (1995). Coordinated community intervention for domestic violence: The effects of arrest and prosecution on recidivism of woman abuse perpetrators. *Crime and Delinquency, 41,* 481–495.

Yarandi, H. N., & Simpson, S. H. (1991). The logistic regression model and the odds of testing HIV positive. *Nursing Research, 40,* 372–373.

Appendix A
Glossary of Terms Used in Logistic Regression

Term	*Definition*
A nonevent	A negative outcome or outcome of no interest (e.g., diagnosed as HIV negative, normal in learning, staying in college, being rejected, etc.).
An event	A positive outcome or outcome of interest (e.g., diagnosed as HIV positive, learning disabled, dropping out from college, being admitted, etc.).
Concordant pair	A pair of observations is said to be concordant if the observation associated with the event outcome has a higher predicted probability derived from the logistic model.
Confounder	A predictor in the logistic regression model that is related both to the outcome variable and a risk factor (also a predictor).
delta-*p*	It measures the change in predicted probability as one unit change in a predictor (X_j) while holding other predictors at a constant. Delta-*p* is not a constant over the range of X_j; it also depends on values of other predictors held at a constant.
Deviance (G^2)	−2 log likelihood of a particular model.
Direct modeling	The modeling technique that permits researchers to include main effects and interactions into a regression model according to a theory-based proposition.
Dummy coding (reference cell or indicator coding)	A coding scheme in which one category of a nominal variable is coded as 0 on all dummy variables, others are coded as 1 on one of the dummy variables and 0 on others. Such a coding scheme is particularly useful/relevant when the slope coefficient (or e^{β}) is directly interpreted as the odds ratio of the current category at risk, compared to the category coded as 0 on all dummy variables.
Effects coding (deviations from the average/mean coding)	A coding scheme in which one category of a nominal variable is coded as (−1) on all dummy variables, and other categories are coded as 1 on one dummy and 0 on all others. This coding strategy is derived from ANOVA framework. Slope coefficients based on this coding scheme are deviations of odds ratio deviating from the average odds ratio
Effect modifier	A predictor in the logistic regression model that interacts with a risk factor (also a predictor variable).
False negative	The proportion of observations incorrectly classified as associated with the nonevent outcome, among all classified as nonevents.
False positive	The proportion of observations incorrectly classified as associated with the event outcome, among all classified as events.
Independent variable	Predictors in the logistic regression model, also called covariates, explanatory, or predictor variables.
log-likelihood	The value of the log-likelihood function of a logistic regression model. When model parameters are substituted by their maximum likelihood estimates, the log-likelihood achieves its maximal value.

(continued)

Appendix A (Continued)

Term	*Definition*
Logit	Natural log of odds = $\ln\left(\frac{p}{1-p}\right) = \log_e(\text{odds}) = \text{logit}(p)$ or the regression model that linearly links the logit transformation of predicted probabilities with a set of parameters.
Marginal probability (or marginal effect or partial effect or partial change)	The partial derivative of the logistic regression density function, with respect to X_j. It is the slope of the probability curve relating x_j [a specific value of predictor X_j] to $Pr(Y = 1 \mid X_1 = x_1, \ldots, X_k = x_k)$, holding all other predictors constant. It is conditioned on the logistic regression model being realized on all predictors.
Normit	The regression model that linearly links the inverse of cumulative normal predicted probabilities with a set of parameters.
Odds	$\frac{p}{1-p} \neq$ probability (p) or likelihood
Odds ratio	A measure of association which equals $\frac{p_1}{1-p_1} \Big/ \frac{p_0}{1-p_0}$, where p_1 = probability of an event, given the membership in Group 1, p_0 = probability of an event, given the membership in Group 0. An odds ratio greater than 1 implies an increased likelihood; conversely, an odds ratio less than 1 implies a decreased likelihood. Invariant to the change in sample size, marginal shifts, interchange of rows (or columns), or row (or columns) multiplications. Can be used in multiple-way tables.
Orthogonal polynomial coding (trend analysis coding)	A coding scheme in which linear, quadratic, cubic, quintic, and so on, trends are built into the coefficients (or weights) of all dummy variables. These coefficients, after being squared, will sum to 1.
Predicted probability	Estimated probability derived from a logistic regression model.
Probit	An alternative name for normit.
R^2	Numerous formulae have been proposed for this concept within the context of logistic regression modeling. In Menard's (2000) empirical study, five were defined and investigated. They are: $R^2 =$ the ordinary least square R^2 $= 1 - \sum(y-\hat{y})^2 / \sum(y-\bar{y})^2$. $R_L^2 =$ the logilihoodration R^2 $= 1 - [\ln(L_M / \ln(L_0))]$.

(continued)

APPENDIX A (Continued)

Term	Definition
	R^2_M = the geometric mean squared improvement per observation R^2 $= 1 - (L_0 / L_M)^{2/n}$.
	R^2_N = the adjusted geometric mean squared improvement R^2 $= [1 - (L_0 / L_M)^{2/n} / [1 - (L_0)^{2/n}]$.
	R^2_C = the contigency coefficient R^2 $= G_M / (G_M + n)$.
	G_M = the model chi-square statistic $= -2[\ln(L_0 - \ln(L_M)]$,
	ln = the natural log transformation,
	L_M = the likelihood function for the model containing all predictors,
	L_0 = the likelihood function for the intercept-only model.
	n = the sample size,
	y = the observed outcome value, coded as an interger,
	$\bar{y}$ = the mean of the outcome values,
	$\hat{y}$ = the predicted probability of the outcome variable, ranging from 0 to 1.
Relative risk	A generic term that has been used interchangeably with either odds ratio or risk ratio. Since odds ratio and risk ratio are two distinct concepts, the use of this term should be restricted to only one of these two.
Risk factor	A predictor in the logistic regression model that is of primary interest to researchers. This term stems from epidemiology in which a risk factor is always related to health.
Sensitivity	A proportion of observations correctly classified as associated with the event outcome, among all event observations.
Somer's D statistic	$D = (nc - nd)/t$, where nc = number of concordant pairs, nd = number of discordant pairs, and t = total number of pairs with different outcomes. D_{yx} = correct measure of association between data and the logistic model, where y = event or nonevent status, x = the predicted probability derived from the model. SAS and MINITAB are two statistical software that compute Somer's D statistic. Unfortunately, they both calculate the incorrect D_{xy}.
Specificity	A proportion of observations correctly classified as associated with the nonevent outcome, among all nonevent observations.
Stepwise modeling	The modeling technique that yields the "best" regression models according to predetermined criteria and/or statistical software's restrictions. Its approach is atheoretical.
Tobit	The normit (or probit) model applied to censored data.

Appendix B
Options and Features Available in Six Statistical Packages for Logistic Regression

Features	*Diagnostic Statistics Calculation Depends on Data Formats*	*Diagnostic Statistics Calculation Based on Observations*		*Diagnostic Statistics Calculation Based on Covariate Patterns*		
	SAS LOGISTIC	*SPSS LOGISTIC REGRESSION*	*SYSTAT LOGIT*	*BMDP LR*	*MINITAB BLOGISTIC*	*STATA LOGISTIC*
Data requirements						
Data format(s) accepted: Raw data format	•	•	•	•	•	•
Frequency data format	•	•	•	•	•	•
Covariate pattern (events/trials) format	•	Use PROBIT		•	•	Use GLOGIT
Default category under the dependent variable to be modeled if coded 0 and 1	0	1	1	1	1	1
Can rearrange the value order of outcome variable	•		•	•	•	
Coding schemes for creating a set of design variables for discrete predictor	Effect, GLM, Orthpoly, Polynomial, Reference	Deviation, Simple, Difference, Helmert, Repeated, Polynomial, Indicator	Effect, Dummy	Marginal, Partial, Orthogonal	Dummy	Dummy
Can specify case weights	•	•				•

Specification of model						
Available link functions for outcome probabilities	Logit, probit (or normit), complementary log-log	Logit	Logit	Logit	Logit, normit, gompit	Logit
Can specify the initial estimates for all the parameters in the model	•				•	
Interaction can be specified as A × B term	•	•	•	•	•	•
Main effects, A and B, must be included along with their interaction, A × B				No, only if (RULE = NONE)	•	•
Correction for over-sampling	•					
Correction for over-dispersion	•					•
Can handle other regression designs:						
1:1 case-control design (use CLOGIT)	•		•	•	•	
1:M case-control design	(use PHREG)		•			(use CLOGIT)
N:M case-control design	(use PHREG)					(use CLOGIT)
Discrete choice model	(use PHREG)		•			(use CLOGIT)
Poisson regression	•	(use LOGLINEAR)				(use POISSON)
Multiple ordered outcome categories	•	(use PLUM)		(use PR)	(use OLOGISTIC)	(use OLOGIT)
Multiple unordered outcome categories	(use CATMOD)	(use NOMREG)	•	(use PR)	(use NLOGISTIC)	(use MLOGIT)

(continued)

APPENDIX B (Continued)

Features	Diagnostic Statistics Calculation Depends on Data Formats	Diagnostic Statistics Calculation Based on Observations		Diagnostic Statistics Calculation Based on Covariate Patterns		
	SAS LOGISTIC	*SPSS LOGISTIC REGRESSION*	*SYSTAT LOGIT*	*BMDP LR*	*MINITAB BLOGISTIC*	*STATA LOGISTIC*
Selection of predictors						
Selection methods						
Forward	•			•		•
Backward	•			•		•
Forward stepwise (Fstep)	•	•	•	•		•
Backward stepwise (Bstep)		•	•	•		•
Other modeling methods						
Force entry of selected predictors	•	•	•	•		•
The best k-predictors model	•					
Sequential modeling	•	•				• (HIER)
Hierarchical modeling	•			•		
Starting and stopping with k-predictors in selection	•					
Selection criteria						
Stepwise based on						
Conditional statistic	•	•				
Likelihood ratio		•				•
Wald statistic		•	For Fstep			•
Score statistic			For Bstep			
Maximum likelihood ratio				•		
ACE (Estimate asymptotic covariance matrix of β)				•		

Default probability for variable entry	0.05	0.05	0.05	0.10		No default
Default probability for variable removal	0.05	0.10	0.10	0.15		No default
Results						
Evaluations of the model						
Log-likelihood or −2Log-likelihood for intercept-only model		•	•	•		
Log-likelihood −2Log-likelihood for intercept with predictors model	•	•	•	•	•	•
Akaike Information Criterion	•					
Score statistic	•	•				
Schwartz Criterion	•					
Wald test	•					
Residual chi-square performed for selection methods only	•	•				
Test for variable not in the model performed for selection methods only	Score test	Score test	Score test	Approximate χ^2 or F		z test
Test of variable in the model performed for the full model and stepwise models	Wald test	Wald test	t ratio	Coefficient/SE		z test
Statistics related to regression coefficient estimates						
Standard error of the regression coefficient	•	•	•	•	•	•
Robust estimate of variance for the coefficients						•
Goodness-of-fit chi-square test for individual predictor in the specified model		• (LR)		•		

(continued)

APPENDIX B (Continued)

Features	Diagnostic Statistics Calculation Depends on Data Formats	Diagnostic Statistics Calculation Based on Observations		Diagnostic Statistics Calculation Based on Covariate Patterns		
	SAS LOGISTIC	*SPSS LOGISTIC REGRESSION*	*SYSTAT LOGIT*	*BMDP LR*	*MINITAB BLOGISTIC*	*STATA LOGISTIC*
Can perform a combined test on design variables	•	•	•	•	•	
Can perform a combined test on two or more predictors	•		•			
Regression coefficient divided by standard error	Wald chi-square	Wald chi-square	*t* ratio	Coefficient/SE	*z* ratio	*z* ratio
Probability value of coefficient divided by standard error	•	•	••		•	•
Confidence interval of the regression coefficient	• (CLPARM=)					• •
Odds ratio or exp(β)	Odds ratio	exp(β)	Odds ratio	exp(β)	Odds ratio	Odds ratio
Confidence interval of the odds ratio or exp(β)	• (CLODDS=)	•	•	•	•	•
Can set the significant level for the confidence intervals	• (ALPHA=)	• (CI(#))		• (CONF=#)		• (LEVEL(#))
Partial correlation between outcome and each predictor		•				
Correlations among regression coefficients	• (CORRB)	•		•		• (VCE, CORR)
Covariances among regression coefficients	• (COVB)			•	•	• (VCE)

Goodness-of-fit statistics						
Hosmer–Lemeshow statistic	•	•	•	•	•	•
Can change the number of groups used in Hosmer–Lemeshow statistic			•		•	•
Brown goodness-of-fit statistic				•	• (2 tests)	
Goodness-of-fit chi-square test based on observed versus expected frequency (deviance)	•	•	•	•	•	•
McFadden's rho-squared for the model			•			
Cox & Snell R^2	•	•				
Nagelkerke R^2	•	•				
Pseudo R^2						•
Validation of predicted probabilities						
Predicted probability of event for each observation or covariate pattern	•	•	•	•	•	•
Measures of association of predicted probabilities and observed outcome	•					
Concordant pairs	•				•	
Discordant pairs	•				•	
Somer's D statistic	•				•	
Gamma statistic	•				•	
Kendall Tau a statistic	•				•	
c statistic	•			•		•
Classification table	•	•	•	•		•
Can specify the cutoff point for classification	•	•		•		•

(continued)

APPENDIX B (Continued)

Features	Diagnostic Statistics Calculation Depends on Data Formats	Diagnostic Statistics Calculation Based on Observations		Diagnostic Statistics Calculation Based on Covariate Patterns		
	SAS LOGISTIC	*SPSS LOGISTIC REGRESSION*	*SYSTAT LOGIT*	*BMDP LR*	*MINITAB BLOGISTIC*	*STATA LOGISTIC*
Diagnostic statistics						
Diagonal of the hat matrix (leverage) for each observation or covariate pattern	•	•	•	•	•	•
Logit residual for each observation		•				
Likelihood residual for each observation or covariate pattern						
Different between observed and predicted probabilities (residual)		•				
Standard error of the predicted probability for each covariate pattern				•		
Studentized residual for each observation		•				
Pearson residual (standardized residual) for each observation or covariate pattern	•	•	•	•	•	•
Change in Pearson chi-square statistic due to deleting each observation or covariate pattern	•		•		•	•

Deviance (deviance residual) for each observation or covariate pattern	•	•	•	•	•	•
Change in deviance due to deleting the individual observation or covariate pattern	•		•		•	•
Standardized difference in the parameter estimate due to deleting each observation or covariate pattern	•	•	•		•	•
Cook distance (or similar statistic) for each observation or covariate pattern	•	•		•		
Confidence interval displacement diagnostics for each observation	•					

Appendix C
Graphic Capabilities of Six Statistical Packages for Logistic Regression

SAS LOGISTIC	*SPSS LOGISTIC REGRESSION*	*SYSTAT LOGIT*	*BMDP LR*	*MINITAB BLOGISTIC*	*STATA LOGISTIC*
Scatterplot of Pearson residual versus observation number	Histogram of predicted probabilities for each group	NA	Histogram of predicted probabilities for each group	Scatterplot of Delta Pearson χ^2 versus estimated event probability	Receiver operating characteristic curve
Scatterplot of deviance residual versus observation number			Scatterplot of observed proportions versus predicted log-odds	Scatterplot of delta deviance versus estimated event probability	Sensitivity/ specificity versus cut-off points
Scatterplot of diagonal elements of the hat matrix versus observation number			Scatterplot of observed proportions versus predicted log-odds	Scatterplots of delta β (based on standardized Pearson residual) versus estimated event probability	
Scatterplots of delta β (standardized) versus observation number			Scatterplot of the percentage of correct classification of observations as a function of the cut-off points	Scatterplots of delta β versus estimated event probability	
Scatterplot of confidence interval displacement diagnostics C versus observation number			Receiver operating characteristic plot	Scatterplot of delta Pearson χ^2 versus leverage	
Scatterplot of confidence interval displacement diagnostics CBAR versus observation number				Scatterplot of delta deviance versus leverage	
Scatterplot of delta deviance versus observation number				Scatterplots of Delta β (based on standardized Pearson residual) versus leverage	
Scatterplot of delta Pearson χ^2 versus observation number				Scatterplots of delta β versus leverage	

UNDERSTANDING STATISTICS, 1(2), 137

ERRATUM

Correction to "Logistic Regression Analysis and Reporting: A Primer"

Chao-Ying Joanne Peng and Tak-Shing Harry So
Volume 1, pp. 31–70

Figures 4 and 6 were printed incorrectly. The correct figures are below. We apologize to Chao-Ying Joanne Peng and Tak-Shing Harry So for the error.

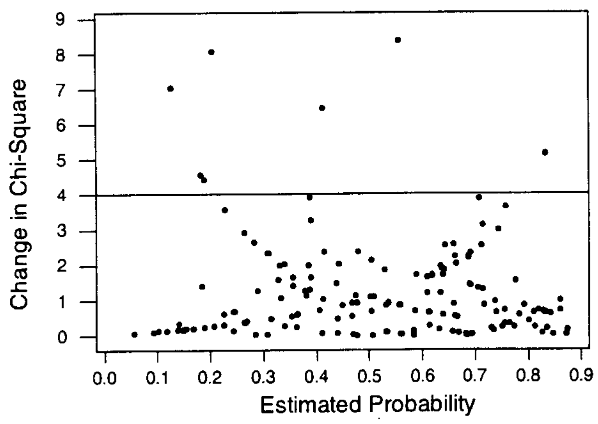

FIGURE 4 Plot of change in Pearson chi-square versus estimated probabilities generated by MINITAB BLOGISTIC.

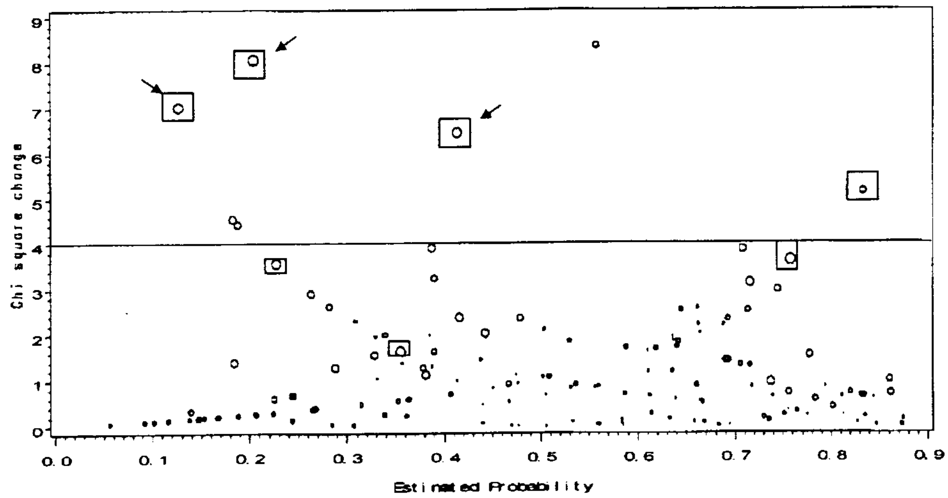

FIGURE 6 Plot of change in Pearson chi-square versus estimated probabilities with the size of circles representing the absolute, standardized change in parameter estimates of *k5*, multiplied with 2.5, generated by SAS GPLOT.