東海管理評論
民國八十九年十一月，第二卷，第一期，63-74

Tests for Response Probability in Multiple Logistic Regression Models

Wen-Ta Kuo[*] *Wei-Hsiung Shen*[**]

Abstract

In this paper we propose several test procedures, such as the **likelihood ratio test**, **uniformly most powerful unbiased test** and the **Wald test**, for testing the response probability in a multiple **logistic regression** set up when the observations are independent binomial variables. An application of the tests is provided.

Keywords : Binomial Distribution, Likelihood Ratio Test, Uniformly Most Powerful Unbiased Test, Wald Test, Logistic Regression.

1. Introduction

The multiple logistic regression model can be briefly described as follows. Denote a set of predictors for a binary response variable Y by $X_1, X_2, \cdots, X_k$ Let $\underline{x}_i = (x_{i0}, \cdots, x_{ik})'$ be the i th setting of values of k explanatory variables, $i = 1, \ldots, I$, where $x_{i0} = 1$. Model for the logit of the probability π that $Y = 1$ is

$$\operatorname{logit}(\pi(\underline{x}_i)) = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} \qquad \cdots\cdots (1)$$

which, by definition, yields

$$\pi_i = \pi(\underline{x}_i) = \frac{\exp(\sum_{j=0}^{k} \beta_j x_{ij})}{1 + \exp(\sum_{j=0}^{k} \beta_j x_{ij})} .$$

Here the parameter β_j refers to the effect of X_j on the log odds that $Y = 1$, controlling the other X's.

* Department of Statistics, Tunghai University.
** Department of Statistics, Tunghai University.

Agresti(1990) mentioned the following in his book (1990, p.112)：When more than one observation on Y occurs at a fixed x_i value, it is sufficient to record the number of observations n_i and the number of '1' outcomes. Thus we let Y_i refer to this success count rather than to individual binary response. The $\{Y_i : i = 1, \cdots, I\}$ are independent binomial random variables with $E[Y_i] = n_i \pi(\underline{x}_i)$, where $n_1 + n_2 + \cdots + n_I = N$.

Accordingly, for $Y_i \sim Binomial(n_i, \pi(\underline{x}_i))$, the likelihood function is

$$L(\beta_0, \beta_1, \cdots, \beta_k; y_1, \cdots, y_I) = \prod_{i=1}^{I} \binom{n_i}{y_i} [\pi(\underline{x}_i)]^{y_i} [1 - \pi(\underline{x}_i)]^{n_i - y_i}$$

$$= \prod_{i=1}^{I} \binom{n_i}{y_i} \{1 + \exp(\sum_{j=0}^{k} \beta_j x_{ij})\}^{-n_i} \{\exp(\sum_{j=0}^{k} \beta_j x_{ij})\}^{y_i}$$

and the log-likelihood function equals

$$l(\beta_0, \beta, \cdots, \beta_k; y_1, \cdots, y_I)$$

$$= \sum_{j=1}^{I} \log \binom{n_i}{y_i} - \sum_{i=1}^{I} n_i \log\{1 + \exp(\sum_{i=1}^{I} \beta_j x_{ij})\} + \sum_{i=1}^{I} y_i (\sum_{j=0}^{k} \beta_j x_{ij})$$

Obviously, the maximum likelihood estimates (MLE) of β's are obtained by solving the likelihood equations：

$$\sum_{i=1}^{I} n_i x_{ij} \frac{\exp(\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2})} = \sum_{i=1}^{I} y_i x_{ij} , \; j = 0, \ldots, k \qquad \cdots\cdots\cdots (2)$$

For the univariate $(k = 1)$ binomial response logistic regression model, Agresti (1990) used the Newton-Raphson method to solve the likelihood equations and determine approximate estimates of β_0 and β_1. Dobson(1990) derived some criteria such as the likelihood ratio test (LRT) and large sample test for goodness of fit of a model in this context. For the bivariate case, Shen(2000) proposed several test procedures for testing the significance of regression coefficients in a multiple regression set up.

In this paper we propose several test procedures for testing

$$H_0 : \pi(\underline{x}) = \pi_0 \text{ vs. } H_1 : \pi(\underline{x}) \neq \pi_0$$

for a specified set of doses $\underline{x}$, where π_0 is a given constant, $0 < \pi_0 < 1$. To motivate the proposed tests, we have taken $k = 2$ throughout the paper. The generalization to $k > 2$ is obvious, though the computations become quite messy. When $k = 2$, the problem is equivalent to testing

$$H_0 : \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \theta_0 \text{ vs. } H_1 : \beta_0 + \beta_1 x_1 + \beta_2 x_2 \neq \theta_0$$

where $\theta_0 = \ln \frac{\pi_0}{1-\pi_0}$. Write $\theta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 - \theta_0$ so that we have

$$H_0 : \theta = 0 \text{ vs. } H_1 : \theta \neq 0.$$

In Section 2 we discuss the LRT for the sake of completeness. In Section 3, using the standard theory of exponential families (Lehmann, 1986), the uniformly most powerful unbiased (UMPU) test for H_0 versus H_1 has been derived. An additional large sample test has been derived in Section 4. An application of the use of the above tests is provided in the context of the analysis of a low birth weight problem (Hosmer and Lemeshow, 1989).

2. Likelihood Ratio Test

In this section we discuss the LRT for testing $H_0 : \pi(\underline{x}) = \pi_0$ vs. $H_1 : \pi(\underline{x}) \neq \pi_0$. The unrestricted MLE $\hat{\beta}_{MLE} = (\hat{\beta}_{0,MLE}, \hat{\beta}_{1,MLE}, \hat{\beta}_{2,MLE})'$ is the solution of the equations (2) for $k = 2$. Under the null hypothesis H_0, the restricted MLEs of β_1 and β_2 are obtained from the last two equations, $j = 1, 2$ in (2) for $k = 2$, after setting $\beta_0 = \theta_0 - \beta_1 x_1 - \beta_2 x_2$ which are given by

$$\sum_{i=1}^{I} \frac{n_i (x_{ij} - x_j) \exp(\theta_0 + \beta_1 (x_{i1} - x_1) + \beta_2 (x_{i2} - x_2))}{1 + \exp(\theta_0 + \beta_1 (x_{i1} - x_1) + \beta_2 (x_{i2} - x_2))} = \sum_{i=1}^{I} y_i (x_{ij} - x_j) \qquad j = 1, 2 \quad \cdots\cdots\cdots\cdots (3)$$

Denote the solutions of (3) by $\widetilde{\beta}_{MLE} = (\widetilde{\beta}_{1,MLE}, \widetilde{\beta}_{2,MLE})'$. Then the LRT of H_0 vs. H_1 rejects H_0 whenever $\lambda = L(\widetilde{\beta}_{MLE}) / L(\hat{\beta}_{MLE})$ is small, or equivalently,

$$D = -2\log\lambda = -2[l(\widetilde{\beta}_{MLE}) - l(\hat{\beta}_{MLE})] > \chi^2_{1;\alpha}$$

where $\chi^2_{1;\alpha}$ is upper α level cut-off point of χ^2_1 distribution. This is an approximate test, which holds in large samples.

3. Uniformly Most Powerful Unbiased Test

Writing $\beta_0 = \theta - \beta_1 x_1 - \beta_2 x_2 + \theta_0$ the joint probability density function of $(Y_1, \ldots, Y_I)$ can be written as

$$\begin{aligned} &P(Y_1 = y_1, \ldots, Y_I = y_I \mid n_1, \ldots, n_I, \underline{x}_1, \underline{x}_2; \theta, \beta_1, \beta_2) \\ &= \prod_{i=1}^{I} \binom{n_i}{y_i} \prod_{i=1}^{I} \{1 + \exp((\theta + \theta_0) + \beta_1(x_{i1} - x_1) + \beta_2(x_{i2} - x_2))\}^{-n_i} \\ &\exp\left\{(\theta + \theta_0)\sum_{i=1}^{I} y_i + \beta_1 \sum_{i=1}^{I} y_i(x_{i1} - x_1) + \beta_2 \sum_{i=1}^{I} y_i(x_{i2} - x_2)\right\} \end{aligned} \quad \cdots\cdots (4)$$

Therefore, by Neyman-Fisher factorization theorem, it follows from (4) that

$$S_0 = \sum_{i=1}^{I} y_i,\ S_1 = \sum_{i=1}^{I} y_i(x_{i1} - x_1), \text{ and } S_2 = \sum_{i=1}^{I} y_i(x_{i2} - x_2)$$

are jointly sufficient statistics for θ, β_1 and β_2. It also follows from (4) that the joint distribution of S_0, S_1 and S_2 is given by

$$\begin{aligned} &P(S_0 = s_0, S_1 = s_1, S_2 = s_2 \mid \theta, \beta_1, \beta_2) \\ &= C_{\theta,\beta_1,\beta_2}(s_0, s_1, s_2)\exp\{(\theta_0 + \theta)s_0 + \beta_1 s_1 + \beta_2 s_2\} \end{aligned} \quad \cdots\cdots (5)$$

where $C_{\theta,\beta_1,\beta_2}(s_0, s_1, s_2)$ is an appropriate function of θ, β_1, β_2 and (s_0, s_1, s_2) whose actual determination is not necessary at this point. However, (5) shows that the joint distribution of (S_0, S_1, S_2) belongs to a three-parameter exponential family (discrete), and this distribution is complete whenever $(\theta, \beta_1, \beta_2)' \in w$ which contains a three-dimensional cube in R^3, which is assumed to hold.

To derive an "optimum" test of $H_0 : \pi(\underline{x}) = \pi_0$ it is clear from (5) that, under H_0, S_1 and S_2 are jointly sufficient for (β_1, β_2). Again, the distribution, which belongs to a two-exponential family, is obviously complete.

Now, using the standard results on exponential families (Lehmann (1986)), we conclude the following. The UMPU test rejects H_0 whenever, for given $S_1 = s_1$ and $S_2 = s_2$, s_0 is either too large or too small. In other words, the UMPU test function at level α is given by

$$\phi(s_0) = \begin{cases} 1, & \text{if } s_0 < c_1(s_1, s_2) \text{ or } s_0 > c_2(s_1, s_2) \\ \gamma_i, & \text{if } s_0 = c_i(s_1, s_2) \;, \quad i = 1,2 \\ 0, & \text{if } c_1(s_1, s_2) < s_0 < c_2(s_1, s_2) \end{cases}$$

where $c_1(s_1, s_2)$, $c_2(s_1, s_2)$, γ_1 and γ_2 satisfy

$$\begin{aligned} &P(S_0 < c_1(s_1, s_2) \mid S_1 = s_1, S_2 = s_2; H_0) \\ &\quad + P(S_0 > c_2(s_1, s_2) \mid S_1 = s_1, S_2 = s_2; H_0) \\ &\quad + \sum_{i=1}^{2} \gamma_i P(S_0 = c_i(s_1, s_2) \mid S_1 = s_1, S_2 = s_2; H_0) = \alpha \end{aligned} \qquad \cdots\cdots\cdots \quad (6)$$

and

$$\begin{aligned} &\sum_{A} s_0 P(S_0 = s_0 \mid S_1 = s_1, S_2 = s_2; H_0) \\ &\quad = E[S_0 \mid S_1 = s_1, S_2 = s_2; H_0](1 - \alpha) \end{aligned} \qquad \cdots\cdots\cdots \quad (7)$$

where $A = \{s_0 : c_1(s_1, s_2) < s_0 < c_2(s_1, s_2)\}$. It may be noted that the former is the size condition while the latter is the unbiasedness condition.

To compute $c_1(s_1, s_2)$ and $c_2(s_1, s_2)$ for given values of s_1 and s_2, we observe from (3.1) that the conditional probability distribution of S_0, given $S_1 = s_1$, $S_2 = s_2$, and H_0, is given by

$$P(S_0 = s_0 \mid S_1 = s_1, S_2 = s_2; H_0) = \frac{P(S_0 = s_0 \mid S_1 = s_1, S_2 = s_2; H_0)}{P(S_1 = s_1, S_2 = s_2 \mid H_0)} = \frac{\exp(\theta_0 s_0)\sum_{A_1}\left\{\prod_{i=1}^{I}\binom{n_i}{y_i}\right\}}{\sum_{A_2}\left\{\prod_{i=1}^{I}\binom{n_i}{y_i}\exp(\theta_0\sum_{i=1}^{I} y_i)\right\}} \quad \cdots\cdots\cdots\cdots\cdots (8)$$

where

$$A_1 = \left\{\underline{y} : \sum_{i=1}^{I} y_i = s_0, \sum_{i=1}^{I} y_i(x_{i1} - x_1) = s_1, \sum_{i=1}^{I} y_i(x_{i2} - x_2) = s_2\right\}$$

and

$$A_2 = \left\{\underline{y} : \sum_{i=1}^{I} y_i(x_{i1} - x_1) = s_1, \sum_{i=1}^{I} y_i(x_{i2} - x_2) = s_2\right\}$$

In actual practice, we proceed as follows : Given s_1, s_2 (the observed values of $\sum_{i=1}^{I} y_i(x_{i1} - x_1)$ and $\sum_{i=1}^{I} y_i(x_{i2} - x_2)$, respectively) and α (the level of significance), we first enumerate all possible combinations $\underline{y}$ ($0 \le y_i \le n_i, i = 1,\dots, I$) such $\underline{y} \in A_2$ that so that the denominator of (8) is evaluated. Next, for all these combinations of $\underline{y} \in A_2$, the numerator is evaluated for values of to given by $\sum_{i=1}^{I} y_i$. Once the conditional probabilities in (8) are evaluated, $c_1(s_1, s_2)$ and $c_2(s_1, s_2)$ are obtained from (6) and (7) by trial and error.

4. The Wald Test

Here we essentially follow an old idea of Berkson(1955). See also Sinha(1988). Let $p_i = y_i / n_i$, $i = 1,\dots, I$ be the sample proportion of 1's at the ith setting and $q_i = 1 - p_i$. Writing $w_i = \log(p_i / q_i), i = 1,\dots, I$ it follows from (1) that, in large samples, w_i satisfies :

$$E(w_i) \approx \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

and

$$Var(w_i) \approx Var(p_i)\left\{\frac{\partial}{\partial p_i}\ln\frac{p_i}{q_i}\bigg|_{p_i=\pi_i}\right\}^2 = \frac{1}{n_i\pi_i(1-\pi_i)}$$

We propose to first estimate β_0, β_1 and β_2 by a weighted least squares method, and then use the resultant estimates to test $H_0 : \pi(\underline{x}) = \pi_0$. The appropriate weights to be used here according to the weighted least squares theory (Neter, Wasserman and Kutner, 1990) are the reciprocals of the estimated variance of w_i, *i.e.*, $n_i p_i q_i$.

Let $U = \sum_{i=1}^{I} n_i p_i q_i (w_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2$. Minimizing U with respect to $\beta = (\beta_0, \beta_1, \beta_2)'$, we readily arrive at the following normal equation：

$$A\beta = \underline{Z}$$

where

$$A = \begin{bmatrix} \sum_{i=1}^{I} n_i p_i q_i & \sum_{i=1}^{I} x_{i1} n_i p_i q_i & \sum_{i=1}^{I} x_{i2} n_i p_i q_i \\ \sum_{i=1}^{I} x_{i1} n_i p_i q_i & \sum_{i=1}^{I} x_{i1}^2 n_i p_i q_i & \sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i \\ \sum_{i=1}^{I} x_{i2} n_i p_i q_i & \sum_{i=1}^{I} x_{i1} x_{i2} n_i p_i q_i & \sum_{i=1}^{I} x_{i2}^2 n_i p_i q_i \end{bmatrix}, \quad \underline{Z} = \begin{bmatrix} \sum_{i=1}^{I} w_i n_i p_i q_i \\ \sum_{i=1}^{I} w_i x_{i1} n_i p_i q_i \\ \sum_{i=1}^{I} w_i x_{i2} n_i p_i q_i \end{bmatrix}.$$

The weighted least squares estimate (WLSE) $\hat{\beta}_{WLSE} = (\hat{\beta}_{0,WLSE}, \hat{\beta}_{1,WLSE}, \hat{\beta}_{2,WLSE})'$ from the above equations is given by：

$$\hat{\beta}_{WLSE} = A^{-1}\underline{Z}$$

Using the asymptotic normality of p_i's, it is not difficult to show that in large samples

$$\hat{\beta}_{WLSE} \sim N(\beta, \Sigma(\beta))$$

where

$$[\Sigma(\beta)]^{-1} = \begin{bmatrix} \sum_{i=1}^{I} n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i1}n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i2}n_i\pi_i(1-\pi_i) \\ \sum_{i=1}^{I} x_{i1}n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i1}^2 n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i1}x_{i2}n_i\pi_i(1-\pi_i) \\ \sum_{i=1}^{I} x_{i2}n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i1}x_{i2}n_i\pi_i(1-\pi_i) & \sum_{i=1}^{I} x_{i2}^2 n_i\pi_i(1-\pi_i) \end{bmatrix}$$

Define

$$\hat{\Sigma}(\hat{\beta}) = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix} = \left\{ \left[\Sigma(\beta)^{-1} \Big|_{\pi_i = p_i} \right] \right\}^{-1}$$

We are now in a position to describe the Wald test for $H_0 : \pi(\underline{x}) = \pi_0$ versus $H_1 : \pi(\underline{x}) \neq \pi_0$. Here we propose to use the test statistic

$$T = \frac{\left|\hat{\beta}_{0,WLSE} + \hat{\beta}_{1,WLSE}x_1 + \hat{\beta}_{2,WLSE}x_2 - \theta_0\right|}{\sqrt{v_{11} + x_1^2 v_{22} + x_2^2 v_{33} + 2x_1 v_{12} + 2x_2 v_{13} + 2x_1 x_2 v_{23}}}$$

and reject H_0 if $T > z_{\alpha/2}$.

5. An Application

A woman's behavior during pregnancy may greatly alter the chance of delivering a baby of normal birth weight. We use as our example a subset of the variables from the data for a study of risk factor associated with low infant birth weight reported in Hosmer and Lemeshow (1989; data were collected at Baystate medical Center, Springfield, MA, 1986). Relevant data appear in 《Table 1》. The variables used are smoking status during pregnancy (SMOKE : 1 = Yes, 0 = No) and presence of uterine irritability (UI : 1 = Yes, 0 = No).

The LRT statistics (D) and P-values for testing H_0 versus H_1, for four combinations of SMOKE and UI are given in 《Table 2》. We used the Newton-Raphson method to find the iterative MLE $\hat{\pi}$ of π For each combination, we take four values of π_0, around the estimated value $\hat{\pi}$, for testing the hypothesis at 5% level. For the UMPU test, 《Table 3》 shows $c_1(s_1, s_2)$, $c_2(s_1, s_2)$, γ_1, γ_2 and s_0 for H_0 vs. H_1. The values of test statistics of the Wald test and their P-values are given in 《Table 4》.

《Tab. 1》 Relevant data for a study of risk factorsassociated with Low Infant Birth Weight.

Factor		Observation	
SMOKE	UI	n_i	y_i
0	0	100	22
0	1	15	7
1	0	61	23
1	1	13	7

《Tab. 2》 Likelihood ratio test statistics (D) and P-values for four combinations of SMOKE and UI.

SMOKE	UI	$\hat{\pi}$	π_0	D	P-value	Significant
0	0	0.227	0.4	14.5340	0.0001	*
			0.3	2.9202	0.0875	
			0.2	1.5639	0.2111	
			0.1	15.3940	0.0001	*
0	1	0.422	0.6	3.0890	0.0788	
			0.5	1.1880	0.4442	
			0.4	0.8388	0.8235	
			0.3	0.4389	0.1992	
1	0	0.366	0.5	4.8880	0.0270	*
			0.4	0.3306	0.5653	
			0.3	1.3606	0.2434	
			0.2	10.1070	0.0015	*
1	1	0.590	0.7	1.2990	0.2544	
			0.6	0.0103	0.9191	
			0.5	0.7494	0.3867	
			0.4	3.3435	0.0675	

《Tab. 3》 Uniformly most powerful unbiased test for four combinations of SMOKE and UI. ($s_0 = 59$)

SMOKE	UI	π_0	c_1	c_2	γ_1	γ_2	Significant
0	0	0.4	68	87	0.0117	0.1536	*
		0.3	57	75	0.0169	0.3496	
		0.2	49	65	0.1154	0.0078	
		0.1	40	53	0.0010	0.4166	*
0	1	0.6	58	67	0.0264	0.1747	
		0.5	57	66	0.0104	0.2410	
		0.4	55	66	0.0080	0.0059	
		0.3	52	61	0.0009	0.8641	
1	0	0.5	61	77	0.0041	0.2578	*
		0.4	54	69	0.0017	0.2220	
		0.3	47	62	0.0357	0.0173	
		0.2	42	55	0.0205	0.0273	*
1	1	0.7	55	65	0.0239	0.0642	
		0.6	54	63	0.09947	0.0008	
		0.5	53	62	0.0011	0.5864	
		0.4	51	60	0.1914	0.0577	

《Tab. 4》 The Wald tests T for four combinations of SMOKE and UI.

SMOKE	UI	π_0	T	P-value	Significant
0	0	0.4	3.5503	0.0004	*
		0.3	1.6349	0.1012	
		0.2	0.6917	0.4892	
		0.1	4.1989	0.00003	*
0	1	0.6	1.7656	0.0775	
		0.5	0.7709	0.4408	
		0.4	0.2238	0.8228	
		0.3	1.3078	0.1910	
1	0	0.5	2.1899	0.0286	*
		0.4	0.5731	0.5666	
		0.3	1.1888	0.2346	
		0.2	3.3381	0.0008	*
1	1	0.7	1.1621	0.2452	
		0.6	0.1020	0.9188	
		0.5	0.8708	0.3838	
		0.4	1.8436	0.0652	

6. Discussion

The results by using the test procedures discussed in the previous sections are all consistent. Although the UMPU test is to be preferred, we recommend using the Wald test proposed in Section 4 due to its obvious simplicity of calculation even for a general k.

Acknowledgements

Our sincere thanks are due to a referee for some excellent constructive comments which led to an improved version.

References

Agresti, A. (1990), *Categorical Data Analysis*, John Wiley and Sons, New York.

Berkson, J. (1955), "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function" , *Journal of the American Statistical Association*, 50 : 130~162.

Dobson, A.J. (1990), *An Introduction to Generalized Linear Models*, Chapman and Hall, London.

Hosmer, D.W. and S. Lemeshow (1989), *Applied Logistic Regression*, John Wiley and Sons, New York.

Lehmann, E.L. (1986), *Testing Statistical Hypotheses*, 2nd ed., John Wiley and Sons, New York.

Neter, J., Wasserman W. and M. Kutner (1990), *Applied Linear Statistical Models*, Toppan Co., Tokyo.

Shen, W.H. (2000), "Exact and Approximate Tests in Multiple Logistic Regression Models" , *Journal of Applied Statistical Science*, 9 : 183~192.

Sinha, B.K. (1988), "Berkson's Bioassay Problem : Issues and Controversies" , *Methods of Operations Research : Symposium on Operations Research*, 60 : 91~102.

應變數具二項分佈之羅吉斯複迴歸模型之反應機率檢定

郭文達* 沈維雄**

摘要

本研究針對反應變數具二項分佈之羅吉斯迴歸模型之反應機率的假設檢定提出三個檢定式：概度比檢定式，一致最強力不偏檢定式和華德檢定，並以一個實例應用說明。

關鍵詞：二項分佈、羅吉斯迴歸、概度比檢定、一致最強力不偏檢定、華德檢定。

*　東海大學統計學系。

**　東海大學統計學系。