Wald's Test as Applied to Hypotheses in Logit Analysis

WALTER W. HAUCK, JR. and ALLAN DONNER*

For tests of a single parameter in the binomial logit model, Wald's test is shown to behave in an aberrant manner. In particular, the test statistic decreases to zero as the distance between the parameter estimate and null value increases, and the power of the test, based on its large-sample distribution, decreases to the significance level for alternatives sufficiently far from the null value.

KEY WORDS: Wald's test; Large-sample tests; Binomial logit model.

1. INTRODUCTION

In testing hypotheses about parameters in a logit model, one generally uses large-sample tests. The choice is between the likelihood-ratio test and other consistent tests which are asymptotically equivalent to the likelihood-ratio test under the null hypothesis. (See, e.g., Rao 1965, pp. 347–352.) One of these alternatives, due to Wald (1943), is particularly convenient, since it requires fitting the model only under the alternative hypothesis (in contrast to the likelihood-ratio test, which requires fitting the model twice). Due to the iterative nature of maximum likelihood estimation when applied to logit analysis, this is a definite advantage for Wald's test over the likelihood-ratio test. However, as Rao points out (1965, p. 350), little is known about the comparative power functions of these tests.

In this article we show, for testing hypotheses regarding a single parameter in a binomial logit model, that Wald's test has the following undesirable features: (1) for any sample size, Wald's test statistic decreases to zero as the distance between the parameter estimate and null value increases; (2) the power of Wald's test, based on its large-sample distribution, decreases to the significance level for alternatives far from the null value.

2. BEHAVIOR OF WALD'S TEST

Let $Y_1, \ldots, Y_N$ be independent, binary-response variables whose probability functions satisfy

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j X_{ij} = f_\beta(\mathbf{X}_i) \quad , \tag{2.1}$$

where

$$P_i = P(Y_i = 1) = [1 + \exp(-f_\beta(\mathbf{X}_i))]^{-1} \ ,$$

and $X_{i1}, \ldots, X_{ik}$ are observations on k independent variables.

Without loss of generality, we select β_k as the parameter of interest. Suppose the hypotheses to be tested are:

$$H_0\colon \beta_k = \beta_{k0} \text{ vs } H_a\colon \beta_k \neq \beta_{k0} \ . \tag{2.2}$$

Let $\hat{\beta}_k$ be the maximum likelihood estimate of β_k, and let $\mathbf{H}$ be the inverse of the sample information matrix (the inverse of the negative second-derivative matrix of the log likelihood evaluated at the maximum likelihood estimates under H_a). Wald's test statistic for (2.2) is

$$W = (\hat{\beta}_k - \beta_{k0})^2 / H_{kk} \ , \tag{2.3}$$

where H_{kk} is the estimated variance of $\hat{\beta}_k$. Under H_0, W has the same asymptotic χ_1^2 distribution as the likelihood ratio statistic $(-2 \ln \lambda)$.

The data in Table 1 illustrate the aberrant behavior of W. The problem is to test the equality of two proportions. Written as a logit model, there is a single independent variable X_1 defined by $X_1 = 0$ for group one and $X_1 = 1$ for group two. The equality of the population proportions in the two groups corresponds to $\beta_1 = 0$, where β_1 is the natural log of the odds ratio. The cases presented in Table 1 are for $N = 200$. The proportion of $Y_1, \ldots, Y_{100}$ (group one) that are observed as 1's is p_1 and the corresponding proportion for $Y_{101}, \ldots, Y_{200}$ (group two) is p_2.

As can be seen, for both values of p_1 considered, W increases with $|p_2 - p_1|$, for p_2 near p_1, but then eventually starts decreasing. The likelihood ratio statistic, in contrast, is monotone increasing. Wald's test behaves this way because, for sufficiently large p_2, H_{11} increases faster than $\hat{\beta}_1^2$.

This behavior also occurs in the general case (2.1), as can be seen in two ways. First, for any N,

$$\lim_{|\hat{\beta}_k| \to \infty} W = 0 \tag{2.4}$$

for fixed $\hat{\beta}_0, \ldots, \hat{\beta}_{k-1}$. The proof of (2.4) is essentially identical to part of the following derivation ((2.12) to (2.15) with $\mathbf{H}$ replacing $\mathbf{I}^{-1}$) and so is omitted.

* Walter W. Hauck, Jr. is Assistant Professor, Dept. of Mathematics, Worcester Polytechnic Institute, Worcester, MA 01609 and Instructor in Preventive and Social Medicine (biostatistics), Harvard Medical School, Boston, MA 02115. Allan Donner is Assistant Professor, Dept. of Epidemiology and Preventive Medicine, University of Western Ontario, London, Ontario N68 5B7 Canada. Work on this article was done when Allan Donner was Biostatistician, Channing Laboratory and Instructor in Preventive and Social Medicine, Harvard Medical School, Boston, MA 02115. The authors thank Sharon Anderson and Sandra West for their helpful comments and criticisms and William M. McCormack, M.D., for permission to cite his data.

© Journal of the American Statistical Association
December 1977, Volume 72, Number 360
Theory and Methods Section

1. Comparison of Wald's and Likelihood-Ratio Tests

p_2	$\hat{\beta}_1$	W	$-2\ln\lambda$
	a. $p_1 = 0.5$		
.60	0.405	2.01	2.02
.70	0.847	8.19	8.40
.80	1.386	18.75	20.27
.90	2.197	31.95	40.70
.93	2.587	34.56	49.69
.94	2.752	34.85	53.16
.95	2.944	34.61	56.94
.99	4.595	20.11	77.27
	b. $p_1 = 0.25$		
.01	−3.497	11.50	30.89
.02	−2.793	13.85	26.24
.03	−2.377	14.24	22.57
.04	−2.079	13.78	19.52
.05	−1.846	12.91	16.91
.10	−1.099	7.34	8.01
.15	−0.636	3.07	3.15
.35	0.480	2.37	2.39
.45	0.898	8.60	8.88
.55	1.299	18.01	19.11
.65	1.718	30.33	33.30
.75	2.197	45.27	52.32
.85	2.833	60.92	78.25
.90	3.296	66.05	95.26
.91	3.412	66.37	99.14
.92	3.541	66.28	103.23
.95	4.043	61.95	117.03
.99	5.694	30.48	141.96

The second approach is to look at the large-sample distribution of W. This is a noncentral chi-square distribution whose noncentrality parameter ξ we now derive; we then show that

$$\lim_{|\beta_k|\to\infty} \xi = 0 \ . \tag{2.5}$$

Let $\mathbf{I}$ be the population information matrix for $(Y_1, \ldots, Y_N)$. From maximum likelihood theory, we know that

$$\frac{\hat{\beta}_k - \beta_k}{(I^{kk})^{\frac{1}{2}}} \xrightarrow[N\to\infty]{d} N(0, 1) \ , \tag{2.6}$$

where I^{kk} is the (k, k)th element of $\mathbf{I}^{-1}$ and $\xrightarrow{d}$ denotes convergence in distribution. For sufficiently large samples, we then have

$$Z = (\hat{\beta}_k - \beta_{k0})/(I^{kk})^{\frac{1}{2}} \mathrel{\dot{\sim}} N((\beta_k - \beta_{k0})/(I^{kk})^{\frac{1}{2}}, 1) \ , \tag{2.7}$$

where $\dot{\sim}$ means "is approximately distributed as." Hence

$$Z^2 = (\hat{\beta}_k - \beta_{k0})^2/I^{kk} \mathrel{\dot{\sim}} \chi_1^2(\xi) \ , \tag{2.8}$$

the noncentral chi-square distribution with noncentrality parameter

$$\xi = (\beta_k - \beta_{k0})^2/I^{kk} \ . \tag{2.9}$$

Since H_{kk} converges in probability to I^{kk}, and

$$W = (\hat{\beta}_k - \beta_{k0})^2/H_{kk} = (I^{kk}/H_{kk})Z^2 \ , \tag{2.10}$$

W and Z^2 will have the same large-sample distribution, i.e.,

$$W \mathrel{\dot{\sim}} \chi_1^2(\xi) \ . \tag{2.11}$$

The noncentrality parameter for the distribution of W thus has the same form as W, with population values replacing sample values.

Using the formula for the inverse of a partitioned matrix with

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{1k} \\ \mathbf{I}_{1k}' & I_{kk} \end{bmatrix}, \qquad I^{kk} = (I_{kk} - \mathbf{I}_{1k}'\mathbf{I}_{11}^{-1}\mathbf{I}_{1k})^{-1} \ . \tag{2.12}$$

So,

$$\xi = (\beta_k - \beta_{k0})^2(I_{kk} - \mathbf{I}_{1k}'\mathbf{I}_{11}^{-1}\mathbf{I}_{1k}) \le (\beta_k - \beta_{k0})^2 I_{kk} \ , \tag{2.13}$$

since $\mathbf{I}_{11}^{-1}$ is positive definite. Now,

$$I_{kk} = \sum_{i=1}^{N} P_i Q_i X_{ik}^2 \ ,$$

where $Q_i = 1 - P_i$. Applying L'Hôpital's rule,

$$\lim_{|\beta_k|\to\infty} (\beta_k - \beta_{k0})^2 P_i Q_i = 0 \tag{2.14}$$

for any i such that $X_{ik} \neq 0$. (If $X_{ik} = 0$, then

2. Results of Logit Analysis with T. Vaginalis as Dependent Variable (455 cases)

Independent variable	Coefficient	Estimated S.E.	Wald's test: Statistic	Wald's test: P value	Likelihood ratio test: Statistic	Likelihood ratio test: P value
1) White race	−4.71	1.66	8.05	.005	14.48	.000
2) Vaginal discharge during last week	3.09	1.31	5.56	.018	8.24	.004
3) Sexually inexperienced or partners usually use condoms	−10.42	29.81	0.12	.729	5.89	.015
4) Smokes cigarettes	2.62	1.34	3.82	.051	4.91	.027
5) Engages in cunnilingus	4.45	1.86	5.72	.017	9.99	.002
6) Has ever douched	−3.99	1.73	5.32	.021	8.90	.003
7) Colonized with *Mycoplasma hominis*	3.98	1.60	6.19	.013	11.19	.001
8) Vaginal discharge noted during physical examination	6.84	2.81	5.93	.015	18.23	.000
9) History of gonorrhea	−14.46	107.49	0.02	.888	7.59	.006

NOTE: All independent variables are dichotomous (1 = yes, 0 = no).

$X_{ik}^2 P_i Q_i = 0$.) Therefore, from (2.13) and (2.14),

$$0 \leq \lim_{|\beta_k| \to \infty} \xi \leq \lim_{|\beta_k| \to \infty} (\beta_k - \beta_{k0})^2 I_{kk} = 0 \ . \quad (2.15)$$

Since ξ decreases to 0 for fixed (large) N, the distribution of W tends to its null (central χ^2) distribution. Hence the power of W decreases to the significance level.

3. EXAMPLE

To illustrate how Wald's test can be misleading, we use data collected by McCormack (1976) in a study of women college students. On the basis of laboratory tests, the organism *T. vaginalis* can be isolated in some of these women. A logit analysis based on (2.1) with $k = 9$ was performed to determine whether the presence of *T. vaginalis* was related to specified characteristics of the women. The Wald statistic, the likelihood-ratio statistic, and their corresponding p values for testing H_0: $\beta_i = 0$ vs H_a: $\beta_i \neq 0$, $i = 1, \ldots, 9$, are listed in Table 2. The p values for Wald's test are consistently larger than those for the likelihood-ratio test. Of particular note are variables three and nine which, using Wald's test, do not attain any conventional level of significance but, using the likelihood-ratio test, are statistically significant at the two percent level.

Clearly, the use of Wald's test for hypotheses of the type considered here can leave the user in a quandry. Does a small value of the statistic mean that the actual value of the parameter is near or very far from the null value? Until more work has been done on this problem, and conditions under which Wald's test will be misleading are more precisely delineated, we would recommend use of the likelihood-ratio test (or some other large-sample test) instead.

[*Received November 1976. Revised June 1977.*]

REFERENCES

McCormack, William M. (1976), Unpublished Observations, Channing Laboratory, Harvard Medical School, Boston, Mass.

Rao, C. Radhakrishna (1965), *Linear Statistical Inference and its Applications*, New York: John Wiley & Sons.

Wald, Abraham (1943), "Tests of Statistical Hypotheses Concerning Several Parameters when the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482.