

ロジスティック回帰分析における モデルの適合度指標に関する考察と提案

内田 治*

本研究は、データを統計的に解析する方法論のひとつであるロジスティック回帰分析に関するものである。ロジスティック回帰分析において想定したモデルの適合度を評価する際に用いられる Hosmer-Lemeshow 検定の問題点を、人工的に作成した具体的な数値例によって明らかにした。さらに、この問題点を修正した新たな適合度指標を提案した。提案した適合度指標の実用性については、従来から提案されているモデルの対数尤度、Cox & Snell の R^2 値、Nagelkerke の R^2 値と比較して、整合性があることを確認した。また、本研究で提案している適合度指標は、Hosmer-Lemeshow 検定の改善という観点から考え出したものであるが、ロジスティック回帰分析における寄与率 R^2 が通常の回帰分析 (OLS) で得られる寄与率に比べて低い値が得られやすく、モデルの信頼性を評価する際に、実用の場面では使いにくいという問題点の解消にも役立つことを報告する。

キーワード：ロジスティック回帰，適合度，Hosmer-Lemeshow 検定，寄与率，相関係数

Consideration and Proposal Concerning the Measurement of Fitness in Logistic Regression Analysis

Osamu UCHIDA

This study is concerned with logistic regression analysis which is a method of statistical data analysis. In this research, the problem of Hosmer-Lemeshow test used in logistic regression analysis for the purpose of evaluating the fitness of model was clarified by some numerical examples made artificially. In addition, a new measurement of fitness that corrected Hosmer-Lemeshow test was proposed. It was confirmed there was coordination with log-likelihood, Cox & Snell- R^2 , Nagelkerke- R^2 , which are usual measurements in logistic regression analysis. Moreover, it was found that report this new measurement is useful for the cancellation of the problem that contribution rate in logistic regression analysis is of no practical use at evaluating the reliability of model because their values are lower than contribution rate in ordinary least squares.

要約

本研究ではロジスティック回帰分析において想定したモデルの適合度を評価する際に用いられる Hosmer-Lemeshow 検定の問題点を、人工的に作成した具体的

な数値例によって明らかにした。さらに、この問題点を修正した新たな適合度指標を提案した。提案した適合度指標の実用性については、従来から提案されているモデルの対数尤度、Cox & Snell の R^2 値、Nagelkerke の R^2 値と比較して、整合性があることを確

*東京情報大学総合情報学部環境情報学科
Tokyo University of Information Sciences, Faculty of Informatics, Department of Environmental Information

認した。また、本研究で提案している適合度指標は、Hosmer-Lemeshow検定の改善という観点から考え出したものであるが、ロジスティック回帰分析における寄与率 R^2 が通常の回帰分析（OLS）で得られる寄与率に比べて、低い値が得られやすく、モデルの信頼性を評価する際に、実用の場面では使いにくいという問題点の解消にも役立つことを報告する。

1. 緒言

ロジスティック回帰分析は質的変数（二値データ）を目的変数とするときに用いられる回帰分析である。この解析手法は2群の判別問題や発生確率の予測問題に適用される。ロジスティック回帰分析におけるモデルを評価するプロセスは、モデル全体の適合度を評価するプロセスとモデルに含まれる説明変数の有意性を評価するプロセスに分けられる。本研究では、モデル全体の適合度を評価するプロセスに焦点をあてる。モデルの適合度を評価する場合、尤度比検定統計量、寄与率^[1]、判別精度、Hosmer-Lemeshow検定などが用いられるが、この中のHosmer-Lemeshow検定は、その数理的な性質が十分に検討されていない。そこで、この方法の問題点を数値例で提示し、改善案を提唱することを本研究の目的とする。

2. Hosmer-Lemeshow検定

Hosmer-Lemeshow検定はロジスティック回帰モデルの適合度を検定するものである。この方法は、回帰式を用いて計算される発生確率 $p(x_i)$ の値を、その値の大きさに基づいて、事前に決めた複数のクラスに、クラスごとの例数がほぼ同数になるように分割（通常は10個のクラスに分割）して、各クラスでの実際の例数（観測度数）と、推定された発生確率から計算される例数（期待度数）の差によって、モデルの適合度を検討している。分割の方法には、同じ発生確率の標本は同一クラスに分割する方法と、そうしない方法（単純に各クラスの例数が同じになるように分割する方法）とがあるが、本研究においては、多くの統計ソフトウェアで採用されているという理由で、同一の発生確率の標本は同一のクラスに分割するという方法によるHosmer-Lemeshow検定を取り上げることにする。

いま、 k 個のクラスに分割した状況を考えよう。このとき、次のような計算を行う^[1]。

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{N_j \hat{\pi}_j (1 - \hat{\pi}_j)} = \sum_{j=1}^k \frac{(O_j - N_j \hat{\pi}_j)^2}{N_j \hat{\pi}_j (1 - \hat{\pi}_j)} \quad \dots (1)$$

ここに

- N_j = 第 j クラスに含まれる全例数
- O_j = 第 j クラスに含まれる発生例数
- E_j = 第 j クラスに含まれる期待例数
- $\hat{\pi}_j$ = 回帰式により計算される発生確率
 $p(x_i)$ の平均値

式（1）で計算される χ^2 値は自由度（分割したクラス数 - 2）の χ^2 分布に従う。そして、この理論を利用した χ^2 検定をHosmer-Lemeshow検定と呼んでいる。

3. Hosmer-Lemeshow検定の出力例

説明変数をX、目的変数をYとする数値例と、そのデータに適用したロジスティック回帰分析におけるHosmer-Lemeshow検定の結果を示す。

ロジスティック回帰分析における Hosmer-

表1；原データ (n=100)

X	Y	X	Y	X	Y	X	Y
1	0	26	0	51	0	76	1
2	0	27	0	52	1	77	1
3	0	28	0	53	0	78	1
4	0	29	0	54	1	79	1
5	0	30	0	55	1	80	1
6	0	31	0	56	0	81	1
7	0	32	0	57	1	82	1
8	0	33	0	58	0	83	1
9	0	34	0	59	1	84	1
10	0	35	0	60	1	85	1
11	1	36	0	61	1	86	1
12	0	37	0	62	1	87	1
13	0	38	0	63	1	88	1
14	0	39	0	64	1	89	1
15	0	40	0	65	1	90	1
16	0	41	1	66	0	91	1
17	0	42	0	67	1	92	1
18	0	43	0	68	1	93	1
19	0	44	1	69	1	94	1
20	0	45	0	70	1	95	1
21	0	46	1	71	1	96	1
22	0	47	0	72	1	97	1
23	0	48	0	73	1	98	1
24	0	49	1	74	1	99	1
25	0	50	0	75	1	100	1

Lemeshow検定の結果では、以下に示すような検定統計量（ χ^2 値と有意確率）と集計表が出力される。

表 2 ; 検定統計量

χ^2	自由度	有意確率
10.5406	8	0.2291

表 3 ; 検定のための集計表

クラス	Y = 0		Y = 1		合計
	観測値	期待値	観測値	期待値	
1	10	9.97208	0	0.02792	10
2	9	9.89617	1	0.10383	10
3	10	9.62278	0	0.37722	10
4	10	8.73214	0	1.26786	10
5	6	6.54495	4	3.45505	10
6	4	3.45505	6	6.54495	10
7	1	1.26786	9	8.73214	10
8	0	0.37722	10	9.62278	10
9	0	0.10383	10	9.89617	10
10	0	0.02792	10	9.97208	10

この例の場合、有意確率の値が0.2291となっており、有意ではない。このことはモデルが適合している（適合していないとはいえない）ということの意味している。Hosmer-Lemeshow検定は、有意確率の値が大きいほど、すなわち、有意にならないことがモデルの適合度が良いことの根拠となる。

4. Hosmer-Lemeshow検定の問題点

Hosmer-Lemeshow検定は、つぎのような問題を含んでいる。

- ①例数が多いとモデルの適合度がよくても、有意である（適合していない）という結論が得られる。
- ②例数が少ないとモデルの適合度が悪くても、有意でない（適合している）という結論が得られる。
- ③他の適合度合を示す尺度（尤度比、寄与率、正判別率など）と必ずしも整合しない。

上記の①および②はHosmer-Lemeshow検定固有の問題ではなく、実測度数と期待度数の差異にもとづいて行われる適合度の検定すべてに共通するものである。本研究では取り上げず、③についての問題点を提示する。

5. 問題の数値例

Hosmer-Lemeshow検定の問題点を3つの数値例(A、B、C)を用いて示すことにする。

表 4 ; 3つの数値例の原データ (各例ともにn=50)

例A		例B		例C	
X	Y	X	Y	X	Y
1	0	1	0	1	0
2	0	2	0	2	0
3	1	3	0	3	0
4	0	4	0	4	0
5	0	5	0	5	0
6	1	6	0	6	0
7	0	7	0	7	0
8	0	8	0	8	0
9	1	9	1	9	0
10	1	10	0	10	0
11	1	11	0	11	0
12	1	12	0	12	0
13	0	13	0	13	0
14	1	14	0	14	0
15	0	15	0	15	0
16	1	16	0	16	0
17	1	17	1	17	0
18	1	18	1	18	0
19	0	19	1	19	0
20	0	20	0	20	1
21	0	21	0	21	1
22	1	22	0	22	0
23	0	23	0	23	0
24	0	24	0	24	0
25	1	25	1	25	1
26	1	26	0	26	1
27	1	27	1	27	1
28	0	28	1	28	1
29	0	29	1	29	1
30	0	30	1	30	1
31	1	31	1	31	1
32	0	32	0	32	1
33	0	33	1	33	1
34	0	34	1	34	1
35	1	35	1	35	1
36	0	36	1	36	1
37	1	37	1	37	1
38	0	38	1	38	1
39	1	39	1	39	1
40	0	40	0	40	0
41	1	41	0	41	0
42	0	42	1	42	1
43	1	43	1	43	1
44	0	44	1	44	0
45	0	45	1	45	1
46	1	46	1	46	1
47	1	47	1	47	1
48	1	48	1	48	1
49	1	49	0	49	1
50	1	50	1	50	1

3つの例をグラフ化すると、図1から図3のようになる。説明変数は1つ(X)で、1から50までの値をとっている。グラフの横軸は説明変数の値、縦軸は所属群(群0と群1)である。

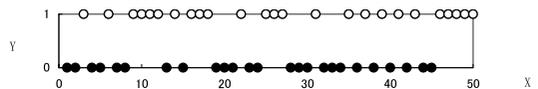


図 1 ; 例Aのドットプロット

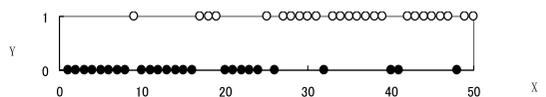


図 2 ; 例Bのドットプロット

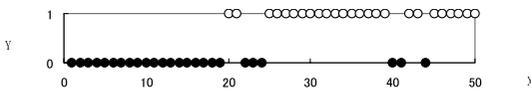


図3：例Cのドットプロット

図1と図2を比較して見ると、例Aより例Bのほうが説明変数Xによって、Y = 0の群とY = 1の群を精度よく判別できることがわかる。また、図2と図3を比較してみると、例Bよりも例CのほうがY = 0の群とY = 1の群を精度よく判別できることがわかる。この3つの例に対して、ロジスティック回帰分析を適用した。その結果を「一般的な適合度を示す統計量」と「Hosmer-Lemeshow検定の結果」に分けて以下に示す。

表5：一般的な適合度を示す統計量

	モデルの -2 対数尤度	Cox & Snell の R ²	Nagelkerke の R ²	正判別率 %
A	68.550	0.015	0.020	52
B	49.389	0.329	0.438	80
C	38.967	0.455	0.607	88

表6：Hosmer-Lemeshow検定の結果

	χ^2	有意確率	自由度
A	7.573	0.476	8
B	7.597	0.474	8
C	16.510	0.036	8

表5を見ると、AのCox & Snell およびNagelkerkのR²値は小さく、正判別率は52%で、モデルの適合度が非常に悪いことがわかる。正判別率が80%を示すBとの間には、どの統計量についても大きな差異が認められる。そして、C、B、Aの順に適合度が良いことがわかる。

一方、表6のHosmer-Lemeshow検定の結果を見ると、AとBの χ^2 値および有意確率は、ほとんど同じ値を示しており、この検定結果だけを見ると、モデルの適合度は同じ程度であるという結論になる。しかも、Aの場合、モデルの適合度が悪いにもかかわらず、有意確率は0.476という値を示しており、モデルは適合しているという結論になってしまっている。また、表3で最も適合度が良い値を示しているCについてのHosmer-Lemeshow検定の結果を見ると、有意確率が0.036という値を示しており、モデルに適合していないという結論になっている。結局、表4では、表3から得られた結論とは逆に、A、B、Cの順に適合度が良いことになる。Hosmer-Lemeshow検定は、このような

問題を含んでおり、この検定の結果を吟味するときには、注意を要することがわかる。

6. 新しい寄与率R_U²の提案

Hosmer-Lemeshow検定には前節で述べたような問題があるとはいえ、発生確率の大小でクラスを作り、クラスごとの度数を吟味するという発想自体は合理的なものであり、全面的に否定されるものではない。そこで、この発想を活かした新しい適合度指標を提案することとした。

Hosmer-Lemeshow検定では、実測度数と期待度数の差で適合度を評価しているが、期待度数を計算するのに用いた発生確率と実際の確率（割合）の相関係数の2乗値で適合度を評価する方法を新たに提案する。本論文では、この値をR_U²と表記し、具体的には、以下に示す手順で計算する。

- ①ロジスティック回帰式を用いてケースごとの発生確率p(x_j)を計算する。
- ②発生確率の大きさに基づいて、事前に決めた複数のクラスに、クラスごとの例数がほぼ同数になるように分割（通常は10個のクラスに分割）する。同じ発生確率となるケースは同一クラスに割り振る。
- ③各クラスにおける一方の群（たとえば、Y = 0の群）の例数の割合p_j（jはクラス番号）を計算する。
- ④クラスごとに予測確率の平均値 $\hat{\pi}_j$ を計算する。
- ⑤p_jと $\hat{\pi}_j$ の相関係数を求め、その値を2乗する。この値をR_U²とする。

7. 数値例による検証

7-1 検証1

表4に示した3つの数値例について、R_U²を計算すると、つぎのような結果が得られる。

$$\begin{aligned} \text{Aにおける } R_U^2 &= 0.125 \\ \text{Bにおける } R_U^2 &= 0.756 \\ \text{Cにおける } R_U^2 &= 0.787 \end{aligned}$$

適合度の悪いAについては、低い値となっており、適合度のよいBとCについては高い値が得られている。

7-2 検証2

10パターンの数値例（各例ともにn=50）を用意して、適合度を示す一般的な統計量、Hosmer-Lemeshow検定の結果、提案する R_U^2 を求めて、比較検討した。この結果から、Hosmer-Lemeshow検定の問題点を再確認するとともに、 R_U^2 の有効性を実証した。なお、Hosmer-Lemeshow検定におけるクラスの数はいは10（ χ^2 検定の自由度は8）とした。

表7：10パターンの数値例

説明変数 X *	目的変数 Y									
	例1	例2	例3	例4	例5	例6	例7	例8	例9	例10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	1
16	0	0	0	0	0	0	0	0	1	1
17	0	0	0	0	0	0	0	1	1	1
18	0	0	0	0	0	0	1	1	1	1
19	0	0	0	0	0	1	1	1	1	1
20	0	0	0	0	1	1	1	1	1	1
21	0	0	0	1	1	1	1	1	1	1
22	0	0	1	1	1	1	1	1	1	1
23	0	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	0	0	0	0	0	0	0	0	0	0
26	1	1	1	1	1	1	1	1	1	1
27	0	0	0	0	0	0	0	0	0	0
28	1	0	0	0	0	0	0	0	0	0
29	1	1	0	0	0	0	0	0	0	0
30	1	1	1	0	0	0	0	0	0	0
31	1	1	1	1	0	0	0	0	0	0
32	1	1	1	1	1	0	0	0	0	0
33	1	1	1	1	1	1	0	0	0	0
34	1	1	1	1	1	1	1	0	0	0
35	1	1	1	1	1	1	1	1	0	0
36	1	1	1	1	1	1	1	1	1	0
37	1	1	1	1	1	1	1	1	1	1
38	1	1	1	1	1	1	1	1	1	1
39	1	1	1	1	1	1	1	1	1	1
40	1	1	1	1	1	1	1	1	1	1
41	1	1	1	1	1	1	1	1	1	1
42	1	1	1	1	1	1	1	1	1	1
43	1	1	1	1	1	1	1	1	1	1
44	1	1	1	1	1	1	1	1	1	1
45	1	1	1	1	1	1	1	1	1	1
46	1	1	1	1	1	1	1	1	1	1
47	1	1	1	1	1	1	1	1	1	1
48	1	1	1	1	1	1	1	1	1	1
49	1	1	1	1	1	1	1	1	1	1
50	1	1	1	1	1	1	1	1	1	1

※ 説明変数 X はすべての例に共通

例1から例10は、この順にモデルの適合度を示す（ $-2 \times$ 対数尤度の値）が大きくなるように、人工的にデータを作成している。

表8：10パターンの数値例に対する適合度

例	モデルの $-2 \times$ 対数尤度	Cox & Snell の R^2	Nagelkerke の R^2	Hosmer-Lemeshow 検定		正判別率 %	提案する R_U^2
				χ^2	有意確率		
1	8.824	0.702	0.936	0.095	1.000	96%	0.999
2	14.472	0.666	0.888	1.442	0.994	92%	0.980
3	19.839	0.628	0.838	4.968	0.761	88%	0.911
4	25.089	0.587	0.783	11.077	0.197	84%	0.795
5	30.246	0.542	0.723	8.927	0.349	80%	0.798
6	35.294	0.494	0.658	10.863	0.210	76%	0.750
7	40.197	0.441	0.589	15.714	0.047	72%	0.646
8	44.911	0.386	0.515	23.111	0.003	68%	0.510
9	49.389	0.329	0.438	32.939	0.000	64%	0.377
10	53.578	0.270	0.360	28.374	0.000	60%	0.343

表8の結果をグラフ化したものが図4である。表8および図4を視察すると、以下のようなことがわかる。

Cox & SnellおよびNagelkerkeの R^2 値、正判別率は適合度の悪化とともに減少してきている。ところが、Hosmer-Lemeshow検定の有意確率は例4と例5で逆転しており、例3と例4の比較では、Cox & SnellおよびNagelkerkeの R^2 値と正判別率がわずかに減少した以外にもかわらず、有意確率は大幅に減少している。さらに、例7を見ると、 R^2 値と正判別率からはモデルの適合度は悪くはないにもかかわらず、有意確率の値は0.047で、モデルは適合していないという結論になる。

一方、本研究で提唱する R_U^2 は例4と例5で逆転していることには変わりはないが、その値の差は無視出来る程度であり、Cox & SnellおよびNagelkerkeの R^2 値および正判別率の値とほぼ同じように減少してきており、適合度を示す指標として信頼できる結果を示している。

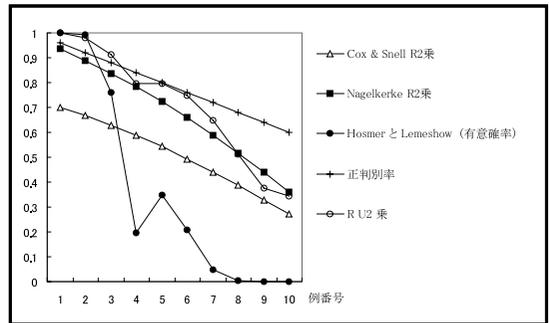


図4：検証結果のグラフ

8. 結語

ロジスティック回帰分析におけるHosmer-Lemeshow検定は、モデルの適合度を忠実に反映しているとは限らず、検定結果を無批判に受けられると誤

った結論を導く可能性がある。本研究では、Hosmer-Lemeshow検定の考え方を活かしつつ、この検定の問題点に対処するために、新たな適合度指標 R_{L}^2 を提案した。この指標はHosmer-Lemeshow検定の問題点を解消するだけでなく、ロジスティック回帰分析において従来から提唱されている様々な寄与率が、通常の回帰分析における寄与率ほど高い値を示さず、直感と合わないという問題点の解消にも役立つものとする。

参考文献

- [1] 丹後俊郎, 山岡和枝, 高木晴良: ロジスティック回帰分析, 朝倉書店 (1996)
- [2] 高橋善弥太: 医師のためのロジスティック・COX回帰入門, 日本医学館 (1995)
- [3] David G. Kleinbaum, etc: Logistic Regression: A Self-Learning Text (Statistics for Biology and Health), Springer Verlag (2002)
- [4] Scott W. Menard: Applied Logistic Regression Analysis (Sage University Papers Series. Quantitative Applications in the Social sciences, No. 106.), Sage Pubns (2001)
- [5] David W., Jr Hosmer, Stanley Lemeshow: Applied Logistic Regression (Wiley Series in Probability and Statistics-Applied Probability and statistics Section), John Wiley & Sons Inc (2001)
- [6] Paul David Allison, Paul Allison: Logistic Regression Using the Sas System Theory & Application, SAS Publishing (1999)
- [7] L Fred C. Pampel: Logistic Regression: A Primer (Sage University Papers Series. Quantitative Applications in the Social Sciences, 132), Sage Pubns (2000)
- [8] James Jaccard: Interaction Effects in Logistic Regression (Sage University Papers Series. Quantitative Applications in the Social Sciences, No. 07-135.), Sage Pubns (2001)
- [9] Logistic Regression Examples Using the Sas System: Version 6, Sas Inst (1995)