A SAS/IML® Macro for Goodness-of-Fit Testing in Logistic Regression Models with Sparse Data

Oliver Kuss

Institute of Medical Epidemiology, Biometry and Informatics,

Halle/Saale, Germany

SUGI 26
SAS® Users Group International
Long Beach, California • April 22-25, 2001

Overview

- Introduction
- Goodness-of-Fit Testing
- %GOFLOGIT
- Conclusion

Introduction (I): The Logistic Regression Model

- **Logistic Regression is the standard analyzing tool for binary responses**
- **Reasons: Interpretation, Prognosis, Software**
- **Logistic Regression in SAS®: Use LOGISTIC, GENMOD, CATMOD, PROBIT procedure**

Introduction (II): Assessing Goodness-of-Fit

- **Residual Analysis*:***
  Check individual observations
  (`PROC LOGISTIC: INFLUENCE-` **and** `IPLOTS-`**Option in** `MODEL` **statement)**
- **Goodness-of-Fit Testing:**
  Combine evidence on lack-of-fit in a single number and apply statistical test

Introduction (III): Notation

N grouped observations (y_i, x_i),

y_i response, $y_i \sim \text{binomial}(m_i, \pi_i)$,

x_i vector of covariates $(1, x_{i1}, \ldots, x_{ip})$

Model equation:

$$\text{logit}(\pi_i) = x_i \beta$$

`PROC LOGISTIC` **syntax:**

```
MODEL y/m=x1...xp;
```

Goodness-of-Fit Testing (I): Standard Tests

Pearson statistic

$$X^2 = \sum_{i=1}^{N} \frac{\left(y_i - m_i\hat{\pi}_i\right)^2}{m_i\hat{\pi}_i\left(1-\hat{\pi}_i\right)}$$

Deviance

$$D = 2\sum_{i=1}^{N} y_i \ln\left(\frac{y_i}{\hat{\pi}_i}\right) + \left(m_i - y_i\right)\ln\left(\frac{m_i - y_i}{m_i - \hat{\pi}_i}\right)$$

Statistical test:

Compare X^2 and D to a χ^2_{N-p-1}-distribution

Goodness-of-Fit Testing (II): Standard Tests and Sparse Data

Validity of the χ^2_{N-p-1}- distribution relies on the assumption of large m_i !!!

Unrealistic with continuous covariates

See what happens with $m_i \equiv 1$:

Pearson statistic $X^2 \approx N$

Deviance $$D = 2\sum_{i=1}^{N} \hat{\pi}_i \ln\left(\frac{\hat{\pi}_i}{1-\hat{\pi}_i}\right) + \ln(1-\hat{\pi}_i)$$

Goodness-of-Fit Testing (III): Solutions in PROC LOGISTIC

- **Group individual observations in new groups depending on the estimated probability from the model**

  (→ Hosmer-Lemeshow test)

  `MODEL` **statement** `LACKFIT`**-Option**

- **Build groups of individual observations on your own**

  `MODEL` **statement**

  `AGGREGATE=(`*`variable-list`*`)`

Goodness-of-Fit Testing (IV): Solutions in PROC LOGISTIC

However.....

- **The Hosmer-Lemeshow test depends heavily on the calculating algorithm**
- **Building new groups on your own is kind of arbitrary and open for manipulation**

Goodness-of-Fit Testing (V): Solutions in the Literature

- **X^2 and D are normally distributed with sparse data (Osius/Rojek X_O^2, McCullagh X_{McC}^2)**
- **Farrington test**

$$X_F^2 = X^2 + \sum_{i=1}^{N} \frac{-\left(1-2\hat{\pi}_i\right)}{m_i\hat{\pi}_i\left(1-\hat{\pi}_i\right)}\left(y_i - m_i\hat{\pi}_i\right)$$

Goodness-of-Fit Testing (VI): Solutions in the Literature

- **IM test (White, Orme IM_{DIAG})**
  Compare two different estimators of the information matrix which should give equivalent results under a satisfactory model fit
- **RSS test**

$$RSS = \sum_{i=1}^{M} \left(y_i - \hat{\pi}_i \right)^2$$

Goodness-of-Fit Testing (VII): Which Test works ???

Limited evidence in the literature

Hosmer et al. find:

X^2_{McC} **and** RSS **perform best, but:**

- **did not include all available tests**
- **only considered $m_i \equiv 1$**

Goodness-of-Fit Testing (VIII): Which Test works ???

Own simulation study with SAS/IML

Results:

- **X^2 and D don't work**
- X_F^2 **outperforms** X_{McC}^2 **and** X_O^2
- IM_{DIAG} **outperforms** RSS
- **In every situation there is a test that outperforms the Hosmer-Lemeshow test**

%GOFLOGIT (I)

= SAS/IML Macro that performs the prescibed tests

Syntax:

```
%goflogit(data=,y=,m=,xlist=,logistic=);
```

where

- **y= variable with number of observed events**
- **m= variable with number of observed trials**
- **xlist= list of covariates**
- **logistic= optional running of** `PROC LOGISTIC`

%GOFLOGIT (II): An Example

Risk factors for occupational hand eczema in hairdressers, M=574 (340 events), N=334, 6 covariates

Syntax:

```
%goflogit(data=hairdresser,y=y,m=m,xlist=
         wet_work permanent_wave atopy
     dyshidrosis center
     skin_protection,
         logistic=ON);
```

%GOFLOGIT (III): Results

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
7.5735	8	0.4762

	Value	p-Value
Standard Pearson Test	369.246	0.054
Standard Deviance	387.080	0.012
Osius-Test	1.702	0.044
McCullagh-Test	1.861	0.031
Farrington-Test	0.232	0.408
IM-Test	6.426	0.491
RSS-Test	108.862	0.063

%GOFLOGIT (IV): Results*

*** = after removal of two outliers**

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
9.2703	8	0.3200

	Value	p-Value
Standard Pearson Test	331.402	0.391
Standard Deviance	373.570	0.033
Osius-Test	-0.027	0.511
McCullagh-Test	0.106	0.458
Farrington-Test	0.185	0.427
IM-Test	3.097	0.876
RSS-Test	106.923	0.734

Conclusion

- Goodness-of-fit testing in logistic regression is an important, but non-trivial task
- Don’t trust the standard methods if you have sparse data
- Use the %GOFLOGIT Macro instead

About the Speaker

Speaker	Oliver Kuss
Location of company	Magdeburger Str. 27 06097 Halle/Saale, Germany
Telephone	+49-345-557-3582
Fax	+49-345-557-3580
E-Mail	Oliver.Kuss@medizin.uni-halle.de

Literature

- Farrington, C.P., (1996), ”On Assessing Goodness of Fit of Generalized Linear Models to Sparse Data,” *Journal of the Royal Statistical Society, B,* 58, 349-360.
- Hosmer, D.W., Hosmer, T., Le Cessie, S. and Lemeshow, S. (1997), ”A comparison of goodness-of-fit tests for the logistic regression model,” *Statistics in Medicine*, 16, 965-980.
- Hosmer, D.W. and Lemeshow, S. (1980), ”Goodness of fit tests for the multiple logistic regression model,” *Communications in Statistics – Theory and Methods*, 9, 1043-1069.
- McCullagh, P. (1985), ”On the Asymptotic Distribution of Pearson´s Statistic in Linear Exponential-Family Models,” *International Statistical Review*, 53, 61-67.
- Orme, C. (1988), ”The calculation of the information matrix test for binary data models,” *The Manchester School,* 54, 370-376.
- Osius, G. and Rojek, D. (1992), ”Normal Goodness-of-Fit Tests for Multinomial Models With Large Degrees of Freedom,” *Journal of the American Statistical Association,* 87, 1145-1152.
- White, H. (1982), ”Maximum Likelihood Estimation of Misspecified Models,” *Econometrica,* 50, 1-25.