Goodness of fit in binary regression models

nusos.ado and binfit ado

Steve Quinn,[1] David W Hosmer[2]

1. Department of Statistics, Data Science and Epidemiology, Swinburne University of Technology, Melbourne, Australia
2. Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst MA, USA

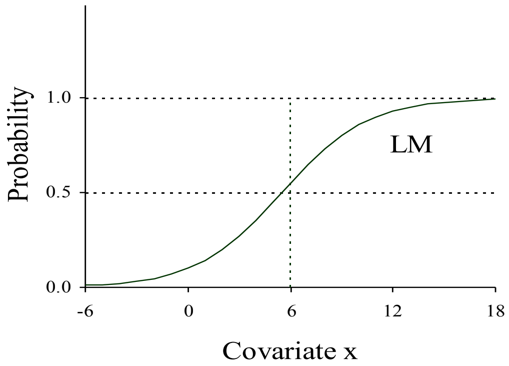
SWINBURNE UNIVERSITY OF TECHNOLOGY

Structure of the talk

- Background
  - logistic regression
    - The Hosmer-Lemeshow statistic
- Motivation - other forms of binary regression
  - Log binominal regression
    - The Hjort-Hosmer statistic
  - Complementary log-log regression
    - The unweighted sum of squares statistic

Background – The logistic model

Logistic regression has long been the workhorse of statistical analysis of binary outcome (yes/no) data.



$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \pi(\mathbf{x}_i) = \frac{e^{x_i^{'}\boldsymbol{\beta}}}{1 + e^{x_i^{'}\boldsymbol{\beta}}}$$

- Outputs Odds Ratios ≈ RR
- Symmetric around y = 0.5

$$\text{If } \boldsymbol{Z_i} = 1 - \boldsymbol{Y_i} \text{ then}$$

$$\Pr(\boldsymbol{Y_i} = 1 \mid \boldsymbol{x_i}) = 1 - \Pr(\boldsymbol{Z_i} = 1 \mid \boldsymbol{x_i})$$

Hosmer-Lemeshow statistic

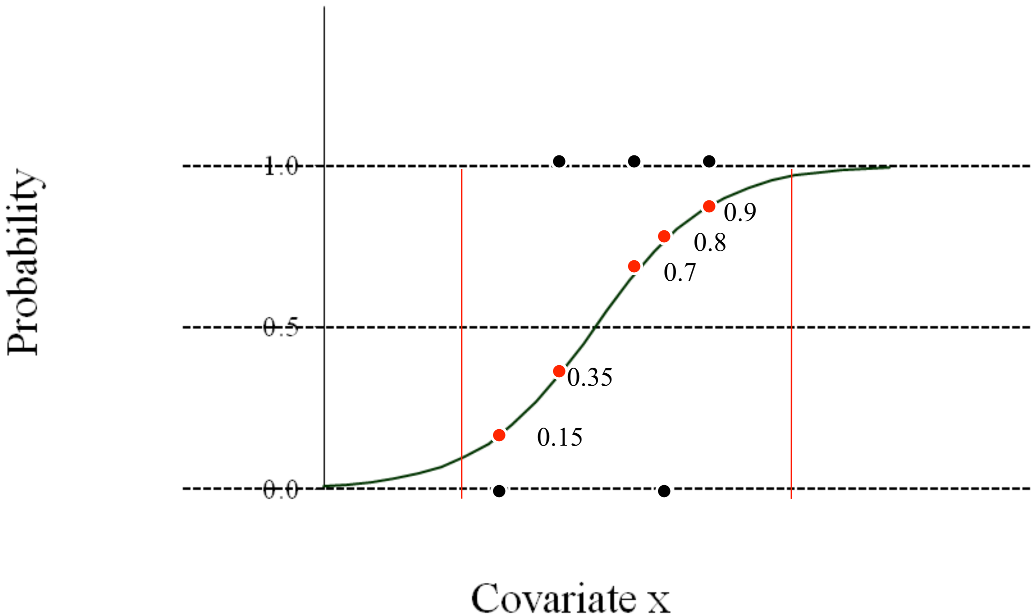
- Hosmer-Lemeshow "deciles-of-risk" test,

Hosmer, D. W. and S. Lemeshow (1980). "A goodness-of-fit test for the multiple logistic regression model." Communications in statistics **A10**: 1043-1069.

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} \qquad \hat{C} : \chi^2_{g-2}$$

Normally, 10 groups

Hosmer-Lemeshow statistic



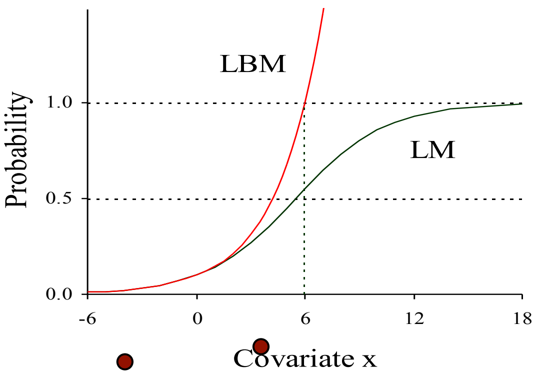
$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n_k\bar{\pi}_k)^2}{n_k\bar{\pi}_k(1-\bar{\pi}_k)}$$

$$\hat{C}_i = \frac{(3-5*0.5)^2}{5*0.5*(1-0.5)} = 0.2$$

The log binomial model (the log-linear model)

Log link

$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \pi(\mathbf{x}_i) = e^{x_i'\boldsymbol{\beta}}$$



- Not symmetric
- Estimation algorithm can fail to converge
- Can produce inadmissible solutions

- Outputs RR

Hjort–Hosmer recommended GOF statistic to assess log binomial regression

Hjort-Hosmer statistic

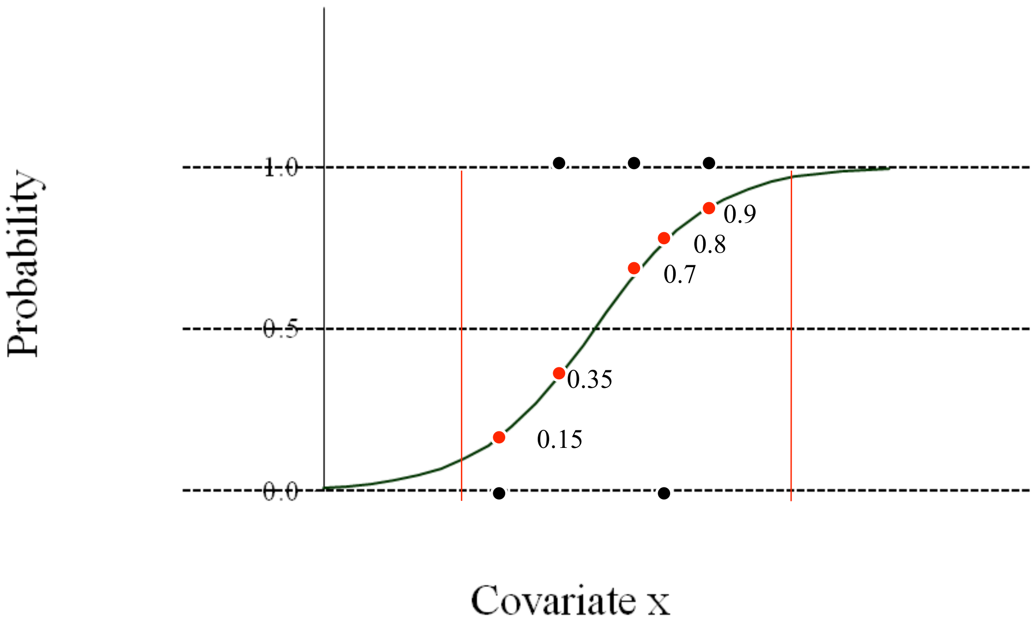
Hosmer DW, Hjort NL, (2002). “Goodness-of-fit processes for logistic regression: simulation results.” Statistics in medicine. 21(18), 2723-2738.

Quinn SJ, Hosmer DW, Blizzard L, Goodness-of-fit statistics for log-link regression models. J Stat Comp Sim. 85(12) (2014), 2533-2545

Hjort–Hosmer example

Based on partial sums of residuals, sorted by their fitted values. Absolute maximal partial sum |M| are calculated.

Rationale: If the model is well-fit, then |M| is small.



Residuals	Partial sums
–0.15	–0.15
0.65	0.50
0.30	0.80
–0.85	–0.05
0.10	0.05

|M|= 0.8

What is a small |M|?

$|M|$ is compared to n secondary partial sums $|M_j|$, each from a "correct" model:

a) comprises the same vector of covariates

b) outcomes simulated using that vector of covariates.

$$\text{P-value} = \sum_j \mathbf{I_j}(|M_j| - |M|)/\mathbf{n}.$$

SWINBURNE UNIVERSITY OF TECHNOLOGY

Performance of HH vs. HL

- The correct model
  - rejection rates of both HH and HL ≈ 5%
- An incorrectly specified model
  - HH > HL by ≈ 10%
  - rejection rates of both HH and HL ≈ 5%

- SUGM 2015
  - An ado file - **hh.ado**

What about other forms of binary regression?

Probit

Complementary log-log (CLL)

Log-log

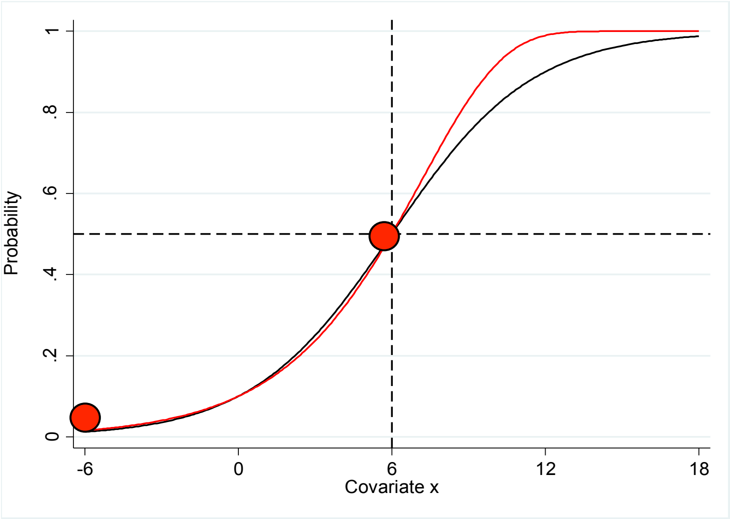
Arc-sin

A corresponding study to that published in 2014 has been carried out for CLL

- Not symmetric
- Still used today

Complementary log-log model

$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i\right) = \pi(\mathbf{x}_i) = 1 - e^{-e^{\mathbf{x}_i'\boldsymbol{\beta}}}$$



- Complementary log-log link
- Not symmetric
- Coefficients not interpretable.

Why bother?

- It has been used to calculate prevalence ratios (vs. prevalence odds ratios)

Bhattacharya R, Shen C, Sambamoorthi U, *Excess risk of chronic physical conditions associated with depression and anxiety*. BMC psychiatry. 14(2014), pp. 10.

- It has been used based on a biological expectation of an asymmetrical relationship between the systematic and random components

Gyimah SO, Adjei JK, Takyi BK, Religion, contraception, and method choice of married women in Ghana. Journal of religion and health. 51(4) (2012), pp. 1359-1374.

Recommended GOF statistic to assess complementary log-log regression?

The normalized unweighted sum of squares statistic.

Unweighted sum of squares

Copas JB (1989). "Unweighted sum of squares test for proportions." Appl. Statist. 38(1), 71-80.

$$US\hat{OS} = \sum_{j=1}^{J} \left(y_j - m_j \hat{\pi}(\mathbf{x}_j) \right)^2;$$

Unfortunately this formula does not follow a known distribution in general.

SWINBURNE UNIVERSITY OF TECHNOLOGY

The normalised unweighted sum of squares

Osius, G. Rojek, D. (1992) Normal Goodness-of-fit tests for multinomial models with large degrees of freedom. J. Amer. Stat. Ass. 87(42) 1145-52.

$$z_{\hat{S}} = \frac{USO\hat{S} - \sum_{j=1}^{J} \hat{V}_j}{\hat{\sigma}_S} \sim N(0,1)$$

numerator: $\hat{V}_j = m_j \hat{\pi}(\mathbf{x}_j)(1 - \hat{\pi}(\mathbf{x}_j))$

denominator : $\hat{\sigma}_S$ = RSS from a linear regression.

SWINBURNE UNIVERSITY OF TECHNOLOGY

The normalised unweighted sum of squares

Dependent variable = $(1-2\hat{\pi}(\mathbf{x}_j))\hat{\pi}(\mathbf{x}_j)(1-\hat{\pi}(\mathbf{x}_j))\big/G'(\eta)$

Independent variables = model covariates

Weights = $G'(\eta)^2\big/((1-\hat{\pi}(\mathbf{x}_j))\,\hat{\pi}(\mathbf{x}_j))$,

where $G'(\eta)$ is the first derivative of the inverse link function.

Logistic $G'(\eta) = \hat{\pi}(\mathbf{x}_j)(1-\hat{\pi}(\mathbf{x}_j))$

CLL $G'(\eta) = (1-\hat{\pi}(\mathbf{x}_j))\ln(1\text{-}\hat{\pi}(\mathbf{x}_j))$_

SWINBURNE UNIVERSITY OF TECHNOLOGY

Performance of the statistics- simulations

- Specify the vector of covariates in the model and take 1000 draws from the vector space e.g. $\boldsymbol{x} \in \boldsymbol{U}(0,10),\ \boldsymbol{d} = 0,1$

- Specify the distribution function

$$\Pr\left(Y_i = 1 \mid \mathbf{x}_i, \beta_0, \beta_1, \beta_2\right) = \pi(\mathbf{x}_i) = 1 - e^{-e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}}$$

- Derive outcomes

$$Y_i = \begin{cases} 1 \text{ if } 1 - e^{-e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}} > u \\ 0 \text{ if } 1 - e^{-e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}} < u \end{cases}$$

Three scenarios considered

1. The correct model – CLL regress $\boldsymbol{Y}$ on $\boldsymbol{x}, \boldsymbol{d}$

2. Power (by omitting terms) – CLL regress $\boldsymbol{Y}$ on $\boldsymbol{x}$

3. Power (wrong link)

determine outcomes by

$$Y_i = \begin{cases} 1 \text{ if } \dfrac{e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}}{1 + e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}} > u \\ 0 \text{ if } \dfrac{e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}}{1 + e^{\beta_0 + \mathrm{x}_i'\beta_1 + d_i'\beta_2}} < u \end{cases}$$

CLL regress $\boldsymbol{Y}$ on $\boldsymbol{x}, \boldsymbol{d}$

SWINBURNE UNIVERSITY OF TECHNOLOGY

Power under the null – the correct model

Table 1. simulated per cent rejection at the level using sample sizes of 200 with 600 replications

1 continuous covariate		Goodness-of-fit statistics‡		
P(Y=1|x=10)*	Distribution	HL	NUSOS	HH
0.9	*U(0,10)*	7.4	5	5.5
0.1	*U(0,10)*	1.2	1.5	2.2
0.999	*N(5,3)*	6.4	3.6	6.4
0.5	$\chi(1)$	1.9	7.8	0.4
0.9	*U(0,10)*	6.8	4.8	5.1
0.1	*U(0,10)*	3.2	4	3.7
0.999	*N(5,3)*	7.2	3.6	5.3
0.5	$\chi(1)$	8.1	3.9	5.8
		5.3	4.3	4.3

*The curve also passes through P(Y=1|x0) = 0.001

SWINBURNE UNIVERSITY OF TECHNOLOGY

Power under the null – the correct model

Table 2. simulated per cent rejection at the level using sample sizes 200 with 600 replications

1 continuous covariate + 1 dichotomous			Goodness-of-fit statistics‡		
P(Y=1|x=10,d=0)	P(Y=1|x=10,d=1)	Distribution	HL	NUSOS	HH
0.999	0.5	*U(0,10)*	6.6	3.8	5.0
0.999	0.5	*N(5,3)*	9.0	4.1	5.5
0.5	0.25	$\chi(1)$	2.7	8.3	6.1
0.5	0.25	$\chi(5)$	1.0	4.9	4.6
0.999	0.5	*U(0,10)*	8.0	4.5	5.4
0.999	0.5	*N(5,3)*	5.8	3.5	5.7
0.5	0.25	$\chi(1)$	7.7	6.0	5.5
0.5	0.25	$\chi(5)$	7.9	3.3	3.7
			6.1	4.8	5.2

*The curve also passes through P(Y=1|x=0,d=0) = 0.001

Power under the alternative – incorrect models

Table 3. simulated per cent rejection at the level using sample size 200 with 60 replications

1 continuous + 1 continuous² covariate			Goodness-of-fit statistics‡		
P(Y=1|x=5)	P(Y=1|x=10)	Distribution	HL	NUSOS	HH
0.5	0.999	*U(0,10)*	15.2	22.5	17.1
0.3	0.5	*U(0,10)*	57.2	42.6	85.3
0.75	0.999	*N(5,3)*	13.1	20.2	15.3
0.75	0.999	$\chi(1)$	6.3	12.1	13.4
0.5	0.999	*U(0,10)*	38.7	50.5	40.5
0.3	0.5	*U(0,10)*	99.1	76.7	100
0.75	0.999	*N(5,3)*	5.0	50.5	35.3
0.75	0.999	$\chi(1)$	15.5	22.6	29.9
			31.3	37.2	42.1

*The curve also passes through P(Y=1|x=0) = 0.001

Power under the alternative – incorrect models

Table 4. simulated per cent rejection at the level using sample sizes Of 200 with 600 replications

1 continuous + 1 dichotomous + interaction covariate			Goodness-of-fit statistics‡		
P(Y=1|x=10,d=0)	P(Y=1|x=10,d=1)	Distribution	HL	NUSOS	HH
0.999	0.25	*U(0,10)*	19.3	8.2	5.9
0.999	0.5	*N(5,3)*	12.1	40	33.2
0.999	0.5	$\chi(3)$	13.2	5.8	6
0.5	0.25	$\chi(5)$	3.8	5.5	21.1
0.999	0.25	*U(0,10)*	28.5	14.3	12.9
0.999	0.5	*N(5,3)*	52.7	91.8	83.1
0.999	0.5	$\chi(3)$	22.4	8.3	5.1
0.5	0.25	$\chi(5)$	8.9	4.8	17.2
			20.1	22.3	23.1

*The curve also passes through P(Y=1|x=0,d=0) = 0.001

SWINBURNE UNIVERSITY OF TECHNOLOGY

Power under the alternative – incorrect models

Table 5. simulated per cent rejection at the level using sample sizes of 200 with 600 replications

1 continuous + 1 dichotomous + interaction covariate			Goodness-of-fit statistics‡		
P(Y=1|x=10,d=0)	P(Y=1|x=10,d=1)	Distribution	HL	NUSOS	HH
0.999	0.25	*U(0,10)*	2.7	7.7	13.3
0.999	0.5	*N(5,3)*	7.2	4.9	5.5
0.999	0.5	$\chi(3)$	3.6	6.8	6.9
0.5	0.25	$\chi(5)$	6.5	3.7	5.2
0.999	0.25	*U(0,10)*	6.3	29.3	32.2
0.999	0.5	*N(5,3)*	7.4	4.5	5.6
0.999	0.5	$\chi(3)$	3.7	12.1	12.9
0.5	0.25	$\chi(5)$	6.2	4.0	5.6
			5.5	9.1	10.9

*The curve also passes through P(Y=1|x=0,d=0) = 0.001

SWINBURNE UNIVERSITY OF TECHNOLOGY

Power under the alternative – incorrect links

Table 6. simulated per cent rejection at the level using sample size 200 with 600 replications

1 continuous covariate		Goodness-of-fit statistics‡		
P(Y=1|x=10,d=0)	Distribution	HL	NUSOS	HH
0.999	*U(0,10)*	22.7	33.7	27.1
0.9	*U(0,10)*	1.6	5.0	8.9
0.999	$\chi(1)$	5.0	5.3	5.4
0.999	*U(0,10)*	61.4	76	69
0.9	*U(0,10)*	6.2	5.2	20
0.999	$\chi(1)$	4.3	8.7	11.6
		16.9	22.3	23.7

*The curve also passes through P(Y=1|x=0,d=0) = 0.001

Power under the alternative – incorrect links

Table 7. simulated per cent rejection at the level using sample sizes of 200 with 600 replications

1 continuous + 1 dichotomous covariate			Goodness-of-fit statistics‡		
P(Y=1|x=10,d=0)	P(Y=1|x=10,d=0)	Distribution	HL	NUSOS	HH
0.999	*0.5*	*U(0,10)*	7.1	15.9	13.3
0.9	*0.5*	*U(0,10)*	2.4	3.8	6.6
0.999	*0.5*	*N(5,5)*	3.3	21.1	46.6
0.999	*0.5*	*N(5,1)*	7.5	14.0	12.5
0.999	*0.5*	*U(0,10)*	4.7	48.2	70.4
0.9	*0.5*	*U(0,10)*	2.7	5.7	13.3
0.999	*0.5*	*N(5,5)*	3.1	27.5	31.5
0.999	*0.5*	*N(5,1)*	21.8	41.7	37.1
			6.6	22.2	28.9

*The curve also passes through P(Y=1|x=0,d=0) = 0.001

Positives of each statistic

	HL	NUSOS	HH
Easy to understand	Yes	No	Yes
Always produces a p-value	No	Yes	Yes
In the packages today	Logistic only	No	No
Quick	Yes	Yes	No
Link Invariant	Yes	No	Yes
Well-defined	No	Yes	No

Well-defined

Both HH and HL need to deal with ties

Case	π	Residual	Partial sum		Case	π	Residual	Partial Sum
0	0.2	-.2	-.2		1	0.2	0.8	0.8
1	0.2	0.8	0.6		0	0.2	0.8	0.6
		|M|	0.6				|M|	0.8

For HL the size of each decile is varied so that all ties are in the same grouping

For HH ties can be randomly sorted.

SWINBURNE UNIVERSITY OF TECHNOLOGY

Example

```
. cloglog foreign headroom

Iteration 0:   log likelihood = -43.291693
Iteration 1:   log likelihood = -42.033306
Iteration 2:   log likelihood = -42.027844
Iteration 3:   log likelihood = -42.027843

Complementary log-log regression               Number of obs     =        74
                                               Zero outcomes     =        52
                                               Nonzero outcomes  =        22

                                               LR chi2(1)        =      6.01
Log likelihood = -42.027843                    Prob > chi2       =    0.0142
```

foreign	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
headroom	-.6205448	.2619884	-2.37	0.018	-1.134033	-.107057
_cons	.7196626	.718866	1.00	0.317	-.689289	2.128614

SWINBURNE UNIVERSITY OF TECHNOLOGY

Example

```
.
. binfit

*****************************************************
*   The Hosmer_Lemeshow p-value is 0.070            *
*   The Hjort-Hosmer p-value is 0.000               *
*   The normalised unweighted SOS p-value is 0.004   *
*****************************************************

.
```

- It Assumes that the Hosmer-Lemeshow partitions in deciles of risk.
- Runs 100 secondary simulations in the Hjort-Hosmer statistic

Questions or comments ?