gofbinreg: Goodness-of-fit (GOF) statistics in binary regression models

Xuelu Sun, Stephen Quinn & Sunil Bhar

School of Health Science

Swinburne University of Technology, Australia

Overview

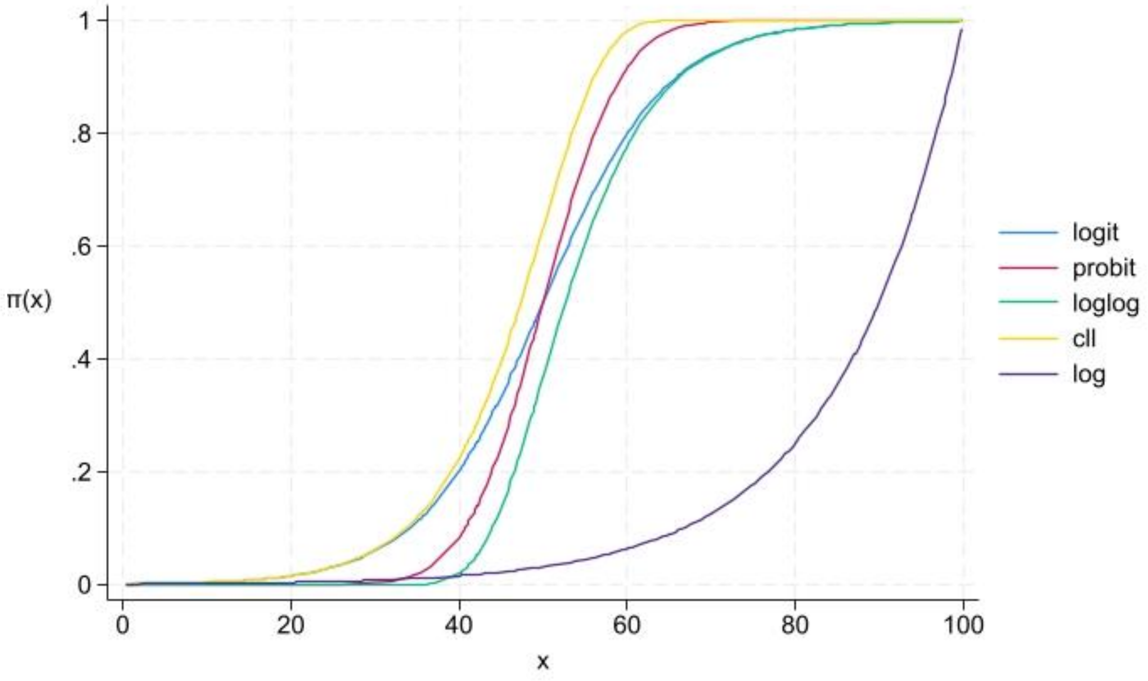
1. Introduction
2. The Key GOF statistics
3. Performance of the GOF statistics
   - Rejection rates under the null
   - Powers under the alternative
4. `gofbinreg` : GOF statistics for binary regression models
   - Ties problem
   - Syntax, options & stored results
   - Real data example

1. Introduction - Binary regression model

- Binary regression model: the response variable is binary (Yes / No)
- Link function g(x): connects Y and x (covariates)
- Most widely used – Logit $\Pr(Y=1\,|\,\mathbf{x})=\pi(\mathbf{x})=\dfrac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}$

1. Introduction – Other links of binary regression models

Link	Typical use
Log	• Direct estimation of risk ratios • Public health and clinical studies
Probit	• Economics (binary choice models) • Biostatistics (dose–response, latent risk)
Log-log	• Rare events accumulating with exposure • Reliability analysis and environmental studies
Complementary log-log(cloglog)	• Time -to- event or hazard-based binary outcomes • Survival analysis and epidemiology



2. Key GOF Statistics

1) Hosmer-Lemeshow statistic (*HL*) :

- Introduced by Hosmer and Lemeshow in 1980
- Widely reported and almost statistical packages

2) Unweighted sum of squared statistic (*USOS*) :

- Introduced by Copas in 1989

3) Hjort-Hosmer statistic (*HH*):

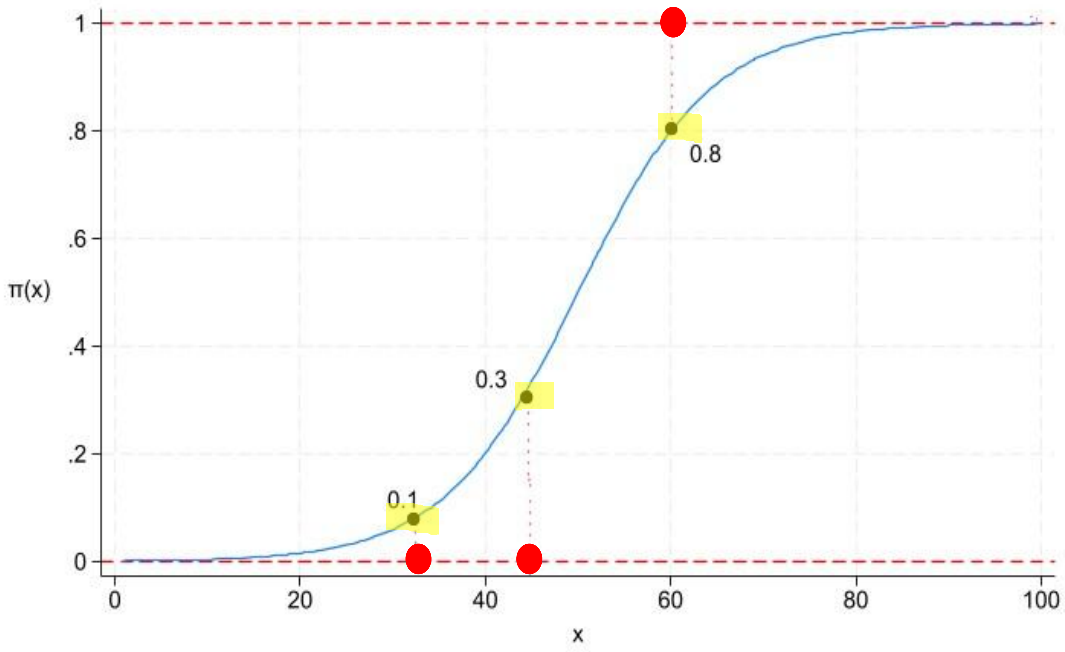
- Introduced by Hosmer and Hjort in 2002

SWINBURNE UNIVERSITY OF TECHNOLOGY

2. GOF Statistics - Hosmer-Lemeshow Statistic (*HL*)

Basic idea: group the fitted probabilities into g (usually10) groups

$$\widehat{C} = \sum_{k=1}^{10} \frac{(o_{1k} - n_k^{'}\overline{\pi}_k)^2}{n_k^{'}\overline{\pi}_k(1-\overline{\pi}_k)} \sim \chi^2(g-2)$$



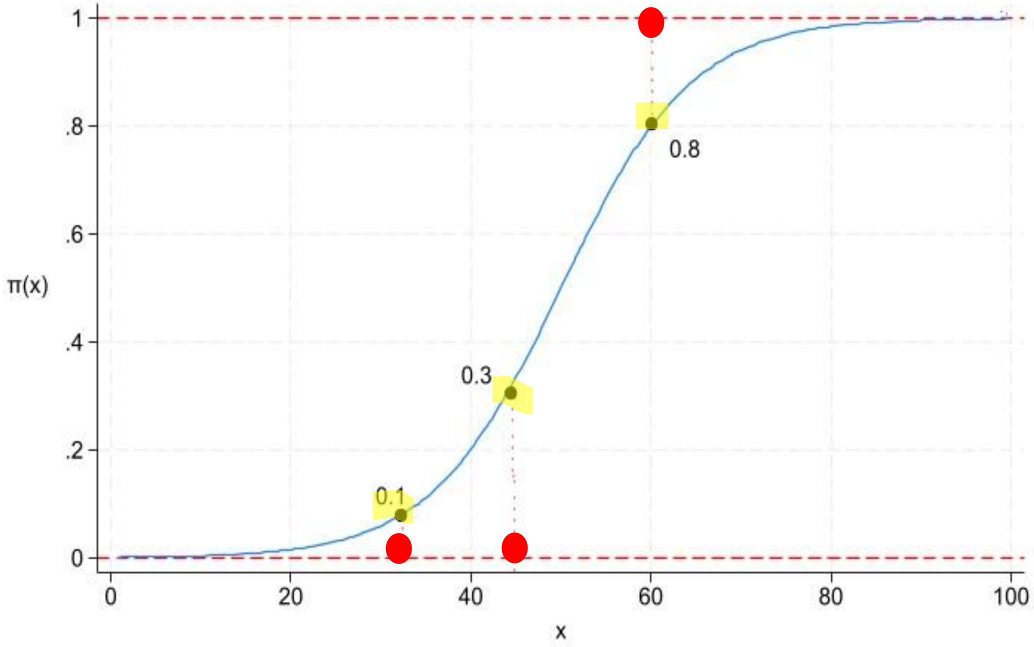
E.g. 3 points comprise in one decide

$$\widehat{C}_k = \frac{(o_{1k} - n_k^{'}\overline{\pi}_k)^2}{n_k^{'}\overline{\pi}_k(1-\overline{\pi}_k)}$$

$$= \frac{(1-3*0.4)^2}{3*0.4(1-0.6)}$$

2. GOF Statistics - Unweighted sum of squared Statistic (*USOS*)

- Residuals = y - $\hat{\pi}(\mathbf{x})$



$$\hat{S} = \sum_{i=1}^{n} (y_i - \hat{\pi}(\mathbf{x}_i))^2$$

$$= (0-0.1)^2 + (0-0.3)^2 + (1-0.8)^2 = 0.14$$

$$\hat{Z}_{\hat{S}} = \frac{\hat{S} - \sum_{i=1}^{n} \hat{\pi}(\mathbf{x}_i)(1-\hat{\pi}(\mathbf{x}_i)}{\hat{\sigma}_{\hat{S}}} \sim N(0,1)$$

$\hat{\sigma}_{\hat{S}}^{\ 2}$ = residual sum of squares of the regression

2. USOS– Calculation on $\hat{\sigma}_{\hat{S}}^{2}$

Link	$\pi(\mathbf{x})$	Weight $=\frac{G'(\eta)^2}{\pi(\mathbf{x})(1-\pi(\mathbf{x}))}$	Regressand $=\frac{\pi(\mathbf{x})(1-2\pi(\mathbf{x}))(1-\pi(\mathbf{x}))}{G'(\eta)}$
Logit	$\frac{e^{g(\mathbf{x})}}{1+e^{g(\mathbf{x})}}$	$\pi(\mathbf{x})(1-\pi(\mathbf{x}))$	$1-2\pi(\mathbf{x})$
Log	$e^{g(x)}$	$\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}$	$(1-2\pi(\mathbf{x}))(1-\pi(\mathbf{x}))$
Probit	$\Phi(g(x))$	$\frac{\phi^2}{\pi(\mathbf{x})(1-\pi(\mathbf{x}))}$	$\frac{\pi(\mathbf{x})(1-\pi(\mathbf{x}))(1-2\pi(\mathbf{x}))}{\phi}$
Log log	$e^{-e^{-g(x)}}$	$\frac{\pi(\mathbf{x})\ln^2\pi(\mathbf{x})}{1-\pi(\mathbf{x})}$	$-\frac{(1-\pi(\mathbf{x}))(1-2\pi(\mathbf{x}))}{\ln\pi(\mathbf{x})}$
CLL	$1-e^{-e^{g(x)}}$	$-\frac{(1-\pi(\mathbf{x}))\ln^2(1-\pi(\mathbf{x}))}{\pi(\mathbf{x})}$	$-\frac{\pi(\mathbf{x})(1-2\pi(\mathbf{x}))}{\ln(1-\pi(\mathbf{x}))}$

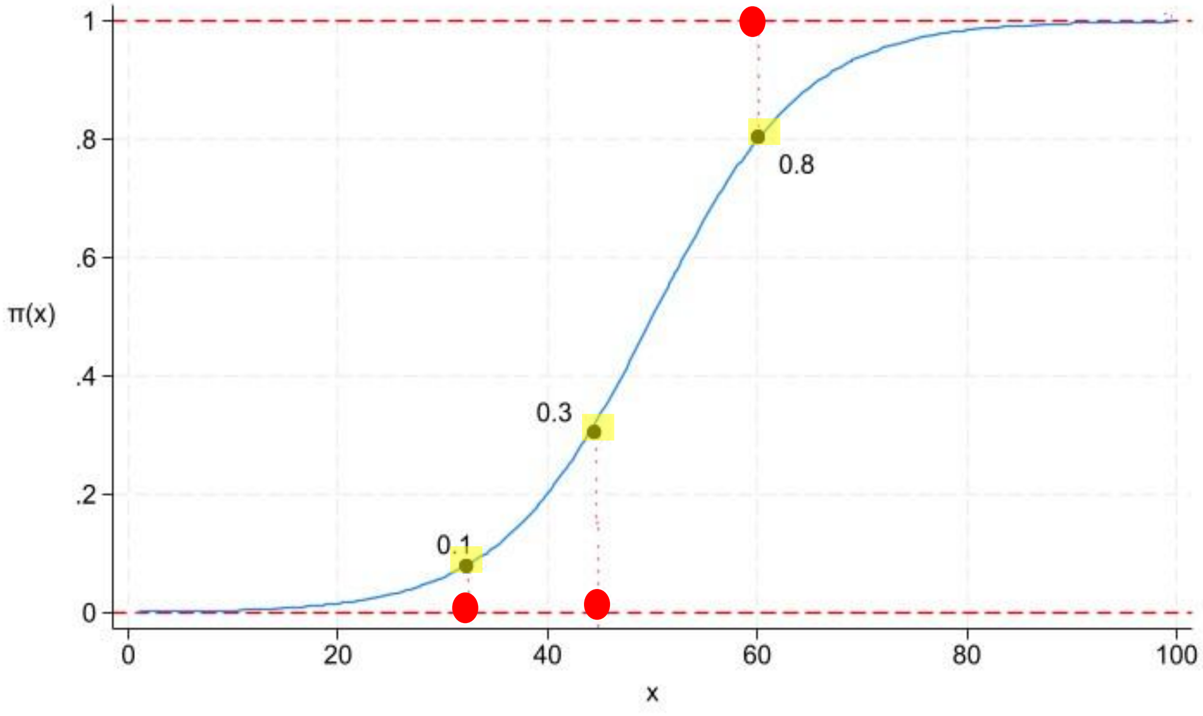
$G'(\eta)$ is the first derivative of the inverse link function

E.g. logistic y x
regress $1-2\pi(\mathbf{x})$ x [aweight = $\pi(\mathbf{x})(1-\pi(\mathbf{x}))$]

SWINBURNE UNIVERSITY OF TECHNOLOGY

2. GOF Statistics - Hjort-Hosmer Statistic (*HH*)

- Based on residuals and partial sums



Residuals	Partial sums
-0.1	-0.1
-0.3	**-0.4**
0.2	-0.2

|M| = 0.4

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

2. GOF Statistics - Hjort-Hosmer Statistic (*HH*)

$|M|$ is compared to 100 secondary simulated maximal partial sum $|M_i|$, each from the correctly specified models:

a) Use covariates (x) to generate the outcome Y

b) Use the same covariates (x) to fit the model

Well-fit → $|M| \approx |M_i|$

Poor-fit → $|M| > |M_i|$

$$\text{p-value} = \frac{\sum_{i}^{100} I_i(|M_i| - |M| \geq 0)}{100}$$

3. Performance of the GOF statistics

- Method - Simulation

  - 1000 replications

  - Sample size of each replication: 200, 400, 800 and 1600

  - Consider
    - Correctly specified model
    - Incorrectly specified model

3. Performance of the GOF statistics

E.g. Suppose we have a correctly specified model with Y and single x

- Generate $\pi(\mathbf{x}) = \dfrac{e^{b0+b1x}}{1+e^{b0+b1x}}$ (e.g. logistic)
- Outcomes: Y= $\begin{cases} 1 & \text{if} \quad \pi(\mathbf{x}) < u \\ 0 & \text{if} \quad \pi(\mathbf{x}) > u, u \in U(0,1) \end{cases}$
- Fit the regression model (logistic Y x)
- Run GOF statistics: how often reject the model (Type I error 5%)?

SWINBURNE UNIVERSITY OF TECHNOLOGY

3. Performance of the GOF statistics - Rejection rates under the null

E.g. **cloglog link**

Distribution of the continuous covariate	Goodness-of-fit statistics											
	N = 200			N = 400			N = 800			N = 1600		
	HL	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*
$U(0,100)$	7.3	4.0	3.8	7.4	3.9	5.1	6.6	4.9	5.1	4.0	4.5	4.3
$N(50,15)$	5.0	4.7	5.6	4.1	4.9	4.1	3.9	5.2	5.7	4.2	4.7	4.9
$\chi^2(4)*6$	4.3	4.1	4.9	6.0	5.8	4.3	6.5	5.5	5.7	6.1	4.7	4.4
$100-\chi^2(4)*6$	4.3	4.1	4.1	7.4	5.1	5.2	6.3	5.7	5.4	5.2	4.8	5.2
Average rejection rate	5.2	4.2	4.8	6.2	4.9	4.7	5.8	5.3	5.5	4.9	4.7	4.7

Other links: All average rejection rates around the expected 5%

SWINBURNE UNIVERSITY OF TECHNOLOGY

3. Performance of the GOF statistics - Powers under the alternative

Incorrectly specified models

- 1 continuous + 1 dichotomous (e.g. logistic y x d)
- 1 continuous + 1 dichotomous + interaction
- 1 continuous + 1 continuous + interaction + 1 continuous + 1 dichotomous (in turn)
- Fit incorrect model by omitting term

- Run GOF: larger rejection rate (more often reject), better performance of the statistic

3. Performance of the GOF statistics - Powers under the alternative

Logit – *Dichotomous term (y x d)*

Goodness-of-fit statistics												
Distribution of the continuous covariate	N = 200			N = 400			N = 800			N = 1600		
	HL	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*
$U(0,100)$	12.6	57.0	40.6	39.1	90.8	71.4	85.1	99.7	95.8	99.9	100.0	100.0
$N(50,15)$	9.3	31.0	19.7	23.6	61.1	37.6	53.7	91.7	65.9	92.5	99.6	95.4
$\chi^2(4)*6$	6.9	3.2	10.9	11.2	14.8	14.0	23.6	60.1	33.1	51.2	95.5	61.9
$100-\chi^2(4)*6$	0.9	21.8	12.7	1.1	43.3	23.5	2.1	75.2	52.5	6.5	95.8	82.1
Average rejection rate	**7.4**	**28.3**	**21.0**	18.8	52.5	36.6	41.1	81.7	61.8	**62.5**	**97.7**	**84.9**

Average power: *USOS* > *HH* > *HL*

3. Performance of the GOF statistics - Powers under the alternative

Logit - *Interaction term (y x d x*d)*

Distribution of the continuous covariate	N = 200			N = 400			N = 800			N = 1600		
	HL	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*
$U(0,100)$	70.6	67.3	23.6	72.9	69.9	31.8	80.1	81.3	47.4	91.3	93.0	72.2
$N(50,15)$	79.7	76.0	23.7	78.9	77.1	26.5	78.9	82.3	34.2	87.7	92.3	52.0
$\chi^2(4)*6$	71.0	45.4	15.4	73.6	48.8	19.0	78.9	50.2	24.0	86.8	63.0	33.4
$100-\chi^2(4)*6$	53.5	43.9	35.3	55.3	51.0	45.3	65.3	59.6	61.7	87.7	81.8	88.2
Average power	68.7	58.2	24.5	70.2	61.7	30.7	75.8	68.4	41.9	**88.4**	**82.5**	**61.5**

Goodness-of-fit statistics

Average power: *HL > USOS > HH*

3. Performance of the GOF statistics - Powers under the alternative

*Logit - Multivariable model (y x d x1 d1 x*d)*

Omission	x1			d1			x*d		
	HL	*USOS*	*HH*	*HL*	*USOS*	*HH*	*HL*	*USOS*	*HH*
N=200									
$U(0,100)$	10.6	10.8	23.8	12.4	14.6	18.6	7.7	5.9	6.6
$N(50,15)$	22.4	23.5	24.3	23.4	28.2	30.6	7.9	6.2	7.8
$\chi^2(4)*6$	4.2	4.7	7.3	5.3	5.2	7.3	3.2	5.8	5.9
$100-\chi^2(4)*6$	65.4	48.2	72.6	21.4	46.8	52.0	25.6	51.0	69.7
N=400									
$U(0,100)$	10.5	10.1	14.1	47.8	45.9	74.3	3.5	3.8	5.5
$N(50,15)$	23.5	23.9	25.6	32.6	33.3	29.2	4.6	7.1	8.4
$\chi^2(4)*6$	6.9	6.2	5.3	5.4	8.1	7.8	5.6	4.1	4.9
$100-\chi^2(4)*6$	57.0	50.9	61.6	26.7	50.3	56.7	28.8	54.3	72.6
N=800									
$U(0,100)$	10.6	16.0	15.7	26.8	64.8	71.2	6.8	6.7	8.8
$N(50,15)$	22.5	24.9	35.3	22.8	25.4	24.7	10.7	10.2	9.4
$\chi^2(4)*6$	5.3	5.8	6.2	4.6	5.3	6.8	5.4	8.6	7.3
$100-\chi^2(4)*6$	60.6	57.8	65.3	23.8	67.9	65.4	30.9	59.7	77.8

HH > USOS > HL

SWINBURNE UNIVERSITY OF TECHNOLOGY

3. Performance of the GOF statistics - Powers under the alternative

- Simulation Results
  - All statistics play a role in model check
  - Overall, the *HH* and *USOS* showed higher powers than *HL* to detect an incorrectly specified model.
  - Sometimes, all statistics performed badly, e.g. when one continuous covariate omitted.

4.Gofbinreg

5.1 Syntax

```
gofbinreg [ if ] [ in ] [ weight ] [ , options ]
```

5.2 Description

The **gofbinreg** reports the Hosmer-Lemeshow statistic, the normalised unweighted sum of squares statistic and the Hjort-Hosmer statistic for a binary regression model. The results of the Hosmer-Lemeshow statistic are identical to those obtained using postestimation commands involving **estat gof**. The **gofbinreg** command can be used for the major binary links and their convenience commands listed in Table 2. For the Hjort-Hosmer statistic, **gofbinreg** will randomly sort the observations within ties a number of times (the default is 100) and report the range of possible *p*-values, the median *p*-value, and draw a histogram of these *p*-values as an option.

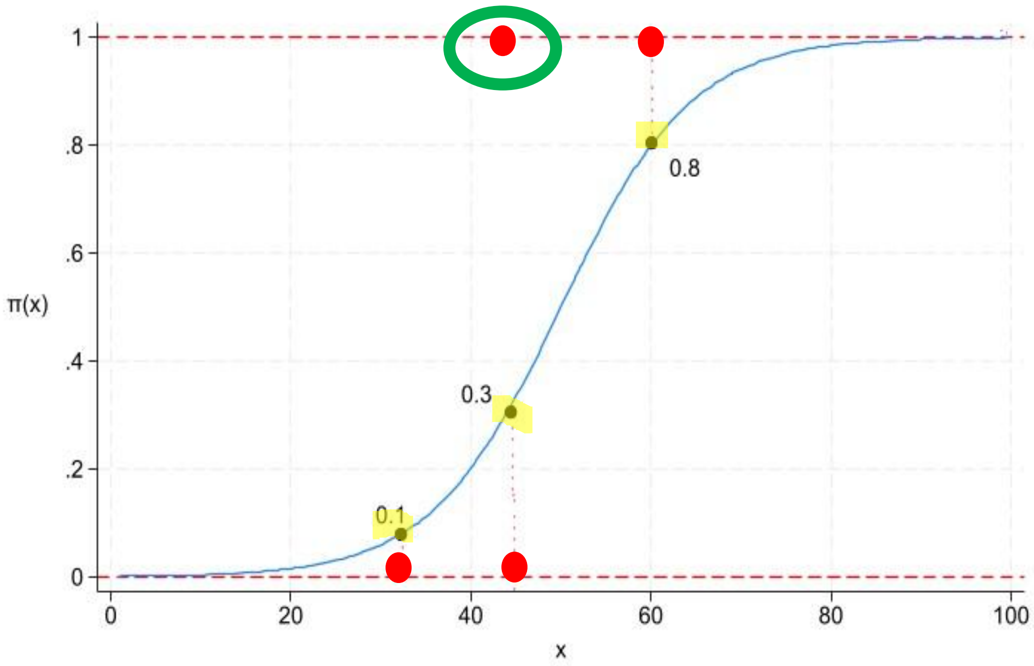
Estat gof

Table 2: Stata model regression codes used before the **gofbinreg** command.

Binary model	Stata command
Logit	`logistic...` `logit...` `glm..., family(binomial) link(logit)` `binreg..., or`
Log Binomial	`glm..., family(binomial) link(log)` `binreg..., rr`
Probit	`probit...` `glm..., family(binomial) link(probit)`
Log-log	`glm..., family(binomial) link(loglog)`
Complementary log-log	`cloglog...` `glm..., family(binomial) link(cloglog)`

4. Gofbinreg - Ties problem with *HH*

- Common problem in real world data
- Two/more observations share same covariate patterns & predicted probabilities



Residuals	|M|
-0.1	-0.1
-0.3	-0.4
0.7	**0.3**
0.2	0.5

→

Residuals	|M|
-0.1	-0.1
0.7	**0.6**
-0.3	0.3
0.2	0.5

4. `Gofbinreg` - Ties problem

HL: Aavoided this problem in **`estat gof`** by placing ties in the same group

USOS: Not affect as based on the sum of squares of residuals

***HH*: how to report???**

Calculate 100 maximal partial sums by randomly sorting observations within ties 100 times

→ Run *HH* statistic 100 times, get 100 possible p-values

→ Output the range, median and histogram of the p-values

4. Gofbinreg - - Ties problem with *HH*

Ties

The Hjort-Hosmer statistic

There exists ties in the dataset

Range of possible p-values = [0.100,0.790]

Median of the maximal partial sums = 3.114

Median p-value = 0.675

No Ties

The Hjort-Hosmer statistic

Maximal partial sums = 3.656

Prob > chi2= 0.09

4. Gofbinreg - Options

reps(#) specifies the number of times the maximal partial sums for the model in question is calculated for the Hjort-Hosmer test. If this option is not given, the default is **reps**(*100*).

hist draws the histogram of 100 possible p-values for Hjort-Hosmer test when the dataset exists ties.

quick reports the quick result of the Hjort-Hosmer test that compares each maximal partial sum of the model in question to one set of 100 secondary simulated maximal partial sums.

thorough reports the thorough result of the Hjort-Hosmer test that compares each maximal partial sum of the model in question to 100 secondary simulated maximal partial sums, that are different for each secondary simulation. If this option is not given, the default option is **quick**.

4. Gofbinreg– Stored Results

gofbinreg saves the following in r()

Scalars

	r(N)	number of observations; Hosmer-Lemeshow test
	r(m)	number of covariates patterns or groups; Hosmer-Lemeshow test
HL	r(df)	degrees of freedom; Hosmer-Lemeshow test
HL	r(HL_chi2)	chi-squared statistic; Hosmer-Lemeshow test
HL	r(HL_p)	probability>chi-squared; Hosmer-Lemeshow test
USOS	r(USOS_z)	normalised unweighted sum of squares test statistic
USOS	r(USOS_p)	probability>chi-squared; normalised unweighted sum of squares test
HH	r(HH_reps)	number of simulations; Hjort-Hosmer test
HH	r(med_MPS)	median of the maximal partial sum; Hjort-Hosmer test
HH	r(HH_min)	minimal possible p-value; Hjort-Hosmer test
HH	r(HH_med)	median possible p-value; Hjort-Hosmer test
HH	r(HH_max)	maximal possible p-value; Hjort-Hosmer test

4. Gofbinreg - Real data example

- Low Birth Weight (LBW) - a large study at Baystate Medical, which includes 189 births to women seen in the obstetrics clinic
- Goal: determine the risk factors of low birth weight

Variable	Description (Name)	Codes/Values
	Identification code (ID)	1-189
Outcome variable	Low birth weight (LOW)	
Exposure variables	Age of mother (AGE)	Years
	Weight of mother at last menstrual period (LWT)	Pounds
	Race (RACE)	1 = White 2 = Black 3 = Other
	Smoking status during pregnancy (SMOKE)	0 = No 1 = Yes
	History of premature labour (PTL)	0 = None 1 = One 2 = Two, etc
	History of hypertension (HT)	0 = No 1 = Yes
	Presence of uterine irritability (UI)	0 = No 1 = Yes

SWINBURNE UNIVERSITY OF TECHNOLOGY

4. Gofbinreg - Real data example

Fit the regression model

```
. logistic low i.race i.smoke ptl ht ui

Logistic regression                                     Number of obs =    189
                                                        LR chi2(6)    =  26.56
                                                        Prob > chi2   = 0.0002
Log likelihood = -104.0559                              Pseudo R2     = 0.1132

------------------------------------------------------------------------------
         low | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
      Black  |   2.967744   1.485446     2.17   0.030      1.11269    7.915503
      Other  |   2.882775   1.204357     2.53   0.011     1.271154    6.537673
             |
       smoke |
     Smoker  |   2.693815   1.041508     2.56   0.010     1.262607    5.747346
         ptl |   1.778619   .5937932     1.72   0.085      .924504     3.42182
          ht |   3.912475   2.477848     2.15   0.031     1.130755    13.53738
          ui |   2.350317   1.060721     1.89   0.058     .9704446    5.692224
       _cons |   .1169809   .0451435    -5.56   0.000     .0549076    .2492284
------------------------------------------------------------------------------
```

4. Gofbinreg - Real data example

```
. gofbinreg, hist

The Hosmer-Lemeshow statistic

 Number of observations =   189

       Number of groups =     8

Hosmer-Lemeshow chi2(6) =  6.33

            Prob > chi2 = 0.387

Warning: There are only 8 distinct quantiles because of ties.

The Unweighted sum of squared statistic

Normalised USOS statistic = 0.911

              Prob > chi2 = 0.362
```

4. Gofbinreg - Real data example

The Hjort-Hosmer statistic

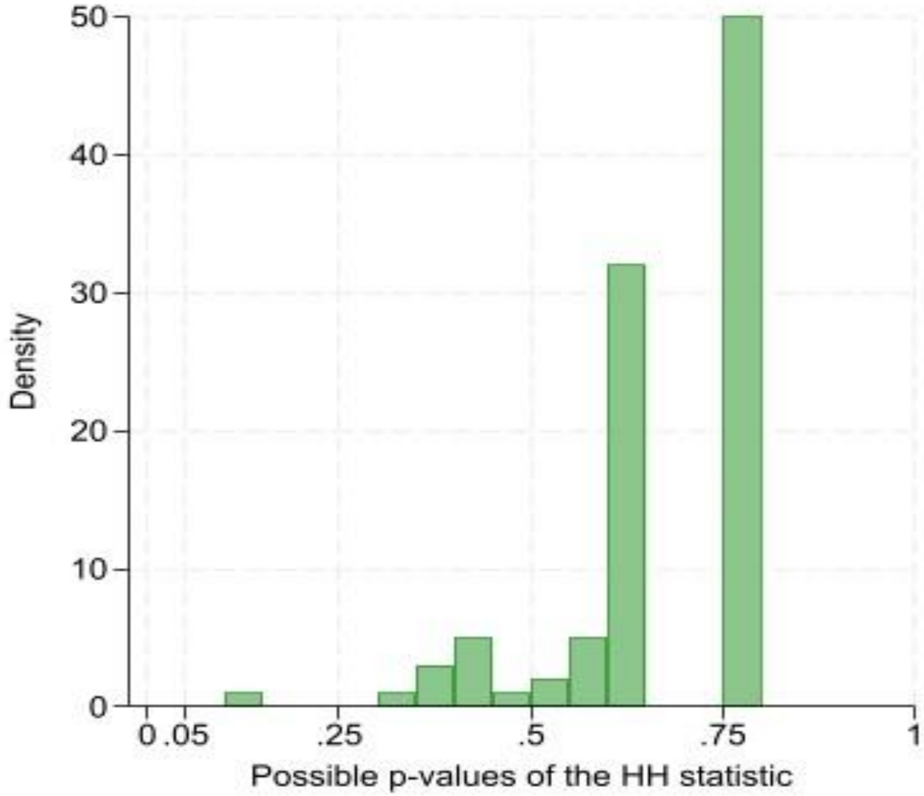
There exists ties in the dataset

Number of simulations = 100

Range of possible p-values = [0.100,0.790]

Median of the maximal partial sums = 3.114

Median p-value = 0.675



Thank for listening!

Questions or Comments?