

# Lecture 6: Model Comparison

Lin Chen

Department of Public Health Sciences  
The University of Chicago

# Comparing Logistic Regression Models

- We introduced the deviance statistic as a way to evaluate model fit in general, relative to the full model.
- Deviance can also be used to compare two linear logistic models for binomial data, and in particular, one model that is nested within another.
- *Nested Models*: all model terms (e.g., covariates) of a smaller model (one with fewer terms) occur in a larger model. This is a necessary condition for using most model comparison tests.
- The difference in deviance between two models measures the extent to which the additional terms improve the fit of the model, and is often used for nested model comparison.

# Comparing Logistic Regression Models: Set-up

## Defining nested models

- Model (1): <sup>current</sup>  $\text{logit}(p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_h X_h$ ;  
Model (2): <sup>full</sup>  $\text{logit}(p) = \beta_0 + \beta_1 X_1 + \cdots + \beta_h X_h + \beta_{h+1} X_{h+1} + \cdots + \beta_k X_k$ .
- # parameter: Model (1)  $h + 1$ ; Model (2)  $k + 1$
- Note: all covariates in Model (1) are in Model (2). Contrasting these models provides a natural way to assess model value vs. complexity

We can set up a formal test via the deviance statistic, testing

$$H_0 : \beta_{h+1} = \cdots = \beta_k = 0$$

vs.

$$H_1 : \text{at least one } \beta_{h+1} = \cdots = \beta_k \neq 0$$

# Comparing Logistic Regression Models: Set-up (cont.)

- Under  $H_0 : \beta_{h+1} = \dots = \beta_k = 0$ , i.e., the reduced Model (1) is the true model, the difference in deviance  $D_1 - D_2 \sim \chi^2_{k-h}$  approximately.

- Note that

$$D_1 = -2\{\log \hat{L}_1 - \log \hat{L}_f\} \sim \chi^2_{n-h-1}$$

$$D_2 = -2\{\log \hat{L}_2 - \log \hat{L}_f\} \sim \chi^2_{n-k-1}$$

$$D_1 - D_2 = -2\{\log \hat{L}_1 - \log \hat{L}_2\} \sim \chi^2_{k-h}$$

The deviance test comparing two nested models is a **likelihood ratio test**. The **likelihood ratio test statistic**

$$\Lambda = -2(\log \hat{L}_1 - \log \hat{L}_2) \sim \chi^2_{\text{df}} \text{ under } H_0$$

- Since the term involving  $\hat{L}_f$  disappears in  $D_1 - D_2$ , **the  $\chi^2$  approximation for  $D_1 - D_2$  can be used to compare nested models for any type of data including ungrouped binary data, sparse data, and grouped binary data.**

# Example 1: *Propagation of plum root-stocks* – nested model comparison

This data examined whether handling of plant cuttings influences probability of survival upon planting

Table 1: Survival rate of plum root-stock cuttings *Grouped Data (4)*

Length of cutting	Time of planting	Number surviving out of 240	Proportion surviving
Short	At once	107	0.45
	In spring	31	0.13
Long	At once	156	0.65
	In spring	84	0.35

# Plum-root: possible models

- Define `length` as an indicator for length of cutting: 0 for short length of cutting, 1 for long length of cutting. Define `time` as an indicator for time of planting: 0 for planting at once, 1 for planting in spring.
- There are five possible models for the survival probabilities of plants:
  - Model 1:  $\text{logit}(p) = \beta_0$   
This model says the survival probability is completely independent of both the length of the cutting and the time of planting.
  - Model 2:  $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length}$   
This model says the survival depends on only the length of the cutting.
  - Model 3:  $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{time}$   
This model says the survival depends on only the time of planting.

# Plum-root: possible models (continued)

- Fourth possible model for the survival probabilities of plants:

- Model 4:  $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time}$

This model indicates the outcome depends on both length of cutting and time of planting, where these two factors act (more or less) independently on outcome.

To interpret  $\beta_1$ , fix the value for variable *time*:

For *length* = 1 : log odds for survival =  $\beta_0 + \beta_1 + \beta_2 \cdot \text{time}$

For *length* = 0 : log odds for survival =  $\beta_0 + \beta_2 \cdot \text{time}$

Thus,  $\beta_1 = \text{log odds (when } \textit{length} = 1) - \text{log odds (when } \textit{length} = 0)$

Hence  $\beta_1$  is the log odds ratio of survival with long cutting compared to short cutting, holding time of planting fixed or sometimes stated as adjusting for time of planting. In this model, it is assumed that for both *time*, the log OR comparing long versus short cutting is the same.

# Plum-root: possible models (continued)

- The fifth (most complex) possible model for the survival probabilities of plants:
  - Model 5:  $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time} + \beta_3 \cdot (\text{length} \times \text{time})$

This model suggests the outcome depends on both length of cutting and time of planting, and there is an interaction between length and time, so that the effect of length of cutting depends on the actual time of planting (or equivalently here, the effect of the time of planting depends on the actual length of cutting).

When there is an interaction effect, there is no longer a single unique effect of length on the log odds of survival, because it depend on whether one is referring to the effect of length when planted at once or when planted in spring. Similarly, the effect of time on the survival depends on the length of cutting.



# Plum-root: possible models (continued)

- Model 5:  $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time} + \beta_3 \cdot (\text{length} \times \text{time})$

Consider the effect of *length*: The question of “what is the effect of length on the log odds of survival” depends on time of planting:

- 1 When planted at once ( $\text{time} = 0$ ):

$$\text{logit}(p_{l=1,t=0}) = \beta_0 + \beta_1$$

$$\text{logit}(p_{l=0,t=0}) = \beta_0$$

$$\text{logit}(p_{l=1,t=0}) - \text{logit}(p_{l=0,t=0}) = \beta_1$$

Hence the log odds ratio of survival with long cutting compared to short cutting when planted at once is  $\beta_1$ .

- 2 When planted in spring ( $\text{time} = 1$ ):

$$\text{logit}(p_{l=1,t=1}) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

$$\text{logit}(p_{l=0,t=1}) = \beta_0 + \beta_2$$

$$\text{logit}(p_{l=1,t=1}) - \text{logit}(p_{l=0,t=1}) = \beta_1 + \beta_3$$

Hence the log odds ratio of survival with long cutting compared to short cutting when planted in spring is  $\beta_1 + \beta_3$ .

*Note:* Model 5 is *full model*. There are four odds (and four  $\hat{p}_i$ 's) for four groups of roots.

# Plum-root: possible models (continued)

- Hierarchical principle: higher-order terms may only be included in a model when the corresponding lower-order terms are present to make the higher-order terms interpretable. This is a conceptual rather than mathematical principle generally. The interaction ( $length \times time$ ) is only included in a model that both  $length$  and  $time$  are included.
- We will use `glm` to fit a series of models, since it outputs the deviance directly.
- The function `glm` also outputs AIC/BIC. AIC (Akaike information criterion) and BIC (Bayesian information criterion) are widely used to compare models, including non-nested ones. A smaller AIC/BIC implies a better fit. AIC/BIC does not provide a formal test. When comparing nested models, we prefer *difference in deviance*.
  - $AIC = -2\log\hat{L} + 2p$
  - $BIC = -2\log\hat{L} + \log(n) \cdot p$ , where  $n$  is sample size and  $p$  is the number of parameters

# Plum-root: data

We change the variables to make  $\beta_0$  (baseline log odds) and the coefficient for interaction term interpretable.

```
. use plumroot.dta
. list
```

	length	time	number	total
1.	1	1	107	240
2.	1	2	31	240
3.	2	1	156	240
4.	2	2	84	240

```
. replace length = length-1
. replace time = time-1
. generate tl = time*length
. list
```

	length	time	number	total	tl
1.	0	0	107	240	0
2.	0	1	31	240	0
3.	1	0	156	240	0
4.	1	1	84	240	1

# Plum-root model 1: $\text{logit}(p) = \beta_0$

```
. glm number, family(binomial total) nolog
```

Generalized linear models

Optimization : ML

Deviance = 151.0193159

Pearson = 141.0527464

No. of obs = 4

Residual df = 3

Scale parameter = 1

(1/df) Deviance = 50.33977

(1/df) Pearson = 47.01758

Variance function:  $V(u) = u*(1-u/\text{total})$

Link function :  $g(u) = \ln(u/(\text{total}-u))$

[ Binomial ]

[ Logit ]

Log likelihood = -86.88071255

AIC = 43.94036

BIC = 146.8604

number	OIM					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	-.4315763	.0660584	-6.53	0.000	-.5610484	-.3021041

This so-called ‘null’ model predicts one probability for all records in the dataset.

# Plum-root model 2: $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length}$

```
. glm number length , family(binomial total) nolog
```

Generalized linear models

Optimization : ML

Deviance = 105.1824146

Pearson = 101.9439614

No. of obs = 4

Residual df = 2

Scale parameter = 1

(1/df) Deviance = 52.59121

(1/df) Pearson = 50.97198

Variance function:  $V(u) = u \cdot (1 - u / \text{total})$

Link function :  $g(u) = \ln(u / (\text{total} - u))$

[ Binomial ]

[ Logit ]

Log likelihood = -63.9622619

AIC = 32.98113

BIC = 102.4098

number	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
length	.9075571	.1360283	6.67	0.000	.6409466	1.174168
_cons	-.9075571	.1008482	-9.00	0.000	-1.105216	-.7098983

This result indicates that cutting length is important

# Plum-root model 3: $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{time}$

```
. glm number time , family(binomial total) nolog
```

Generalized linear models

Optimization : ML

Deviance = 53.44041229

Pearson = 52.31577106

No. of obs = 4

Residual df = 2

Scale parameter = 1

(1/df) Deviance = 26.72021

(1/df) Pearson = 26.15789

Variance function:  $V(u) = u \cdot (1 - u / \text{total})$

Link function :  $g(u) = \ln(u / (\text{total} - u))$

[ Binomial ]

[ Logit ]

Log likelihood = -38.09126073

AIC = 20.04563

BIC = 50.66782

number	OIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
time	-1.347222	.1408757	-9.56	0.000	-1.623333	-1.071111
_cons	.1922567	.0917092	2.10	0.036	.01251	.3720034

Time of planting is also important

# Plum-root model 4: $\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time}$

```
. glm number length time, family(binomial total) nolog
```

Generalized linear models		No. of obs	=	4
Optimization : ML		Residual df	=	1
		Scale parameter	=	1
Deviance	= 2.293839315	(1/df) Deviance	=	2.293839
Pearson	= 2.270478953	(1/df) Pearson	=	2.270479
Variance function:	$V(u) = u \cdot (1 - u / \text{total})$	[ Binomial ]		
Link function :	$g(u) = \ln(u / (\text{total} - u))$	[ Logit ]		
Log likelihood	= -12.51797425	AIC	=	7.758987
		BIC	=	.907545

number	OIM		z	P> z	[95% Conf. Interval]	
	Coef.	Std. Err.				
length	1.017691	.14548	7.00	0.000	.7325559	1.302827
time	-1.427542	.1464624	-9.75	0.000	-1.714603	-1.140481
_cons	-.3039203	.1171647	-2.59	0.009	-.533559	-.0742816

When both `length` and `time` are considered together, each predictor is still associated with log of odds of survival, i.e. affects survival probability.

# Plum-root model 5: The full model –

$$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time} + \beta_3 \cdot (\text{length} \times \text{time})$$

```
. glm number length time tl , family(binomial total) nolog
```

Generalized linear models		No. of obs	=	4
Optimization	: ML	Residual df	=	0
		Scale parameter	=	1
Deviance	= 8.95647e-14	(1/df) Deviance	=	.
Pearson	= 8.06851e-17	(1/df) Pearson	=	.
Variance function:	V(u) = u*(1-u/total)	[ Binomial]		
Link function	: g(u) = ln(u/(total-u))	[ Logit]		
		AIC	=	7.685527
Log likelihood	= -11.37105459	BIC	=	8.96e-14

number	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
length	.8365595	.1875623	4.46	0.000	.4689442	1.204175
time	-1.690827	.2321796	-7.28	0.000	-2.14589	-1.235763
tl	.4527483	.3008943	1.50	0.132	-.1369937	1.04249
_cons	-.2175203	.1298637	-1.67	0.094	-.4720486	.037008

The interaction effect does not appear to be needed, based on the test on the  $\beta$  coefficient. We also can test via the  $D$  statistic.



# Plum-root: summary of model fit results

\*Midterm\*

Table 2: Values of deviances for the plum root-stock data.

Model	Terms fitted in model	Deviance	d.f.
(1)	$\text{logit}(p) = \beta_0$	151.02	3
(2)	$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length}$	105.18	2
(3)	$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{time}$	53.44	2
(4)	$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time}$	2.29	1
(5)	$\text{logit}(p) = \beta_0 + \beta_1 \cdot \text{length} + \beta_2 \cdot \text{time} + \beta_3 \cdot (\text{length} \times \text{time})$	0	0

Table 3: Analysis of deviance table for the plum root-stock data.

Source	Models	$\Delta D$	$\Delta$ d.f.
length	(1) vs (2)	45.84	1
time length	(2) vs (4)	102.89	1
time+length	(1) vs (4)	148.73	2

Source	Models	$\Delta D$	$\Delta$ d.f.
time	(1) vs (3)	97.58	1
length time	(3) vs (4)	51.15	1
LT time,length	(4) vs (5)	2.29	1

- Conclusion: Model (4) is best. The test statistic 2.29 with 1 d.f. has a p-value of 0.13, no significant difference in fit comparing with full model, thus interaction term can be omitted.

# Comparing Related Non-nested Models

- Nested models select in hierarchical fashion among all candidate models. Assessing difference in deviance provides a formal test for comparing nested models.
- Though **generally** non-nested models *cannot* be directly compared in a formal test, the deviance statistic can be helpful in choosing among alternate models in an informal way (not a formal test)
- One example is when two models only differ by the functional form of a single predictor

## Example 2: *Evaluation of an Anti-pneumococcus Serum* – Non-nested models of different forms of a predictor

This dataset summarizes effects of an anti-pneumococcus serum over increasing doses, exposing batches of infected mice

**Table 4:** Number of deaths from pneumonia amongst batches of 40 mice exposed to different doses of a serum

Dose of serum	Number of deaths out of 40
0.0028	35
0.0056	21
0.0112	9
0.0225	6
0.0450	1

# Example: *Evaluation of an Anti-pneumococcus Serum*

```
. use Anti-pneumococcus_serum.dta  
. list
```

	dose	y	n
1.	.0028	35	40
2.	.0056	21	40
3.	.0112	9	40
4.	.0225	6	40
5.	.045	1	40

# Anti-pneumococcus serum: $\text{logit}(p) = \beta_0 + \beta_1 d$

```
. glm y dose, fam(binomial n) nolog
```

Generalized linear models  
Optimization : ML

Deviance = 15.90001864  
Pearson = 16.34155504

Variance function:  $V(u) = u \cdot (1 - u/n)$   
Link function :  $g(u) = \ln(u/(n-u))$

Log likelihood = -16.33425515

No. of obs = 5  
Residual df = 3  
Scale parameter = 1  
(1/df) Deviance = 5.300006  
(1/df) Pearson = 5.447185

[ Binomial ]  
[ Logit ]

AIC = 7.333702  
BIC = 11.0717

-----							
	y	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
-----							
	dose	-146.6927	26.36298	-5.56	0.000	-198.3632	-95.02222
	_cons	1.217898	.2929699	4.16	0.000	.6436879	1.792109
-----							

Often in dose-response processes, failures/deaths occur as a multiplicative function of dose. We next predict death with  $\log_e(dose)$

# Anti-pneumococcus serum: $\text{logit}(p) = \beta_0 + \beta_1 \log(d)$

```
. gen ldose = log(dose)
. glm y ldose, fam(binomial n) nolog
```

Generalized linear models	No. of obs	=	5
Optimization : ML	Residual df	=	3
	Scale parameter	=	1
Deviance = 2.80889589	(1/df) Deviance	=	.9362986
Pearson = 2.917404799	(1/df) Pearson	=	.9724683
Variance function: $V(u) = u*(1-u/n)$	[Binomial]		
Link function : $g(u) = \ln(u/(n-u))$	[Logit]		
	AIC	=	4.715478
Log likelihood = -9.788693777	BIC	=	-2.019418

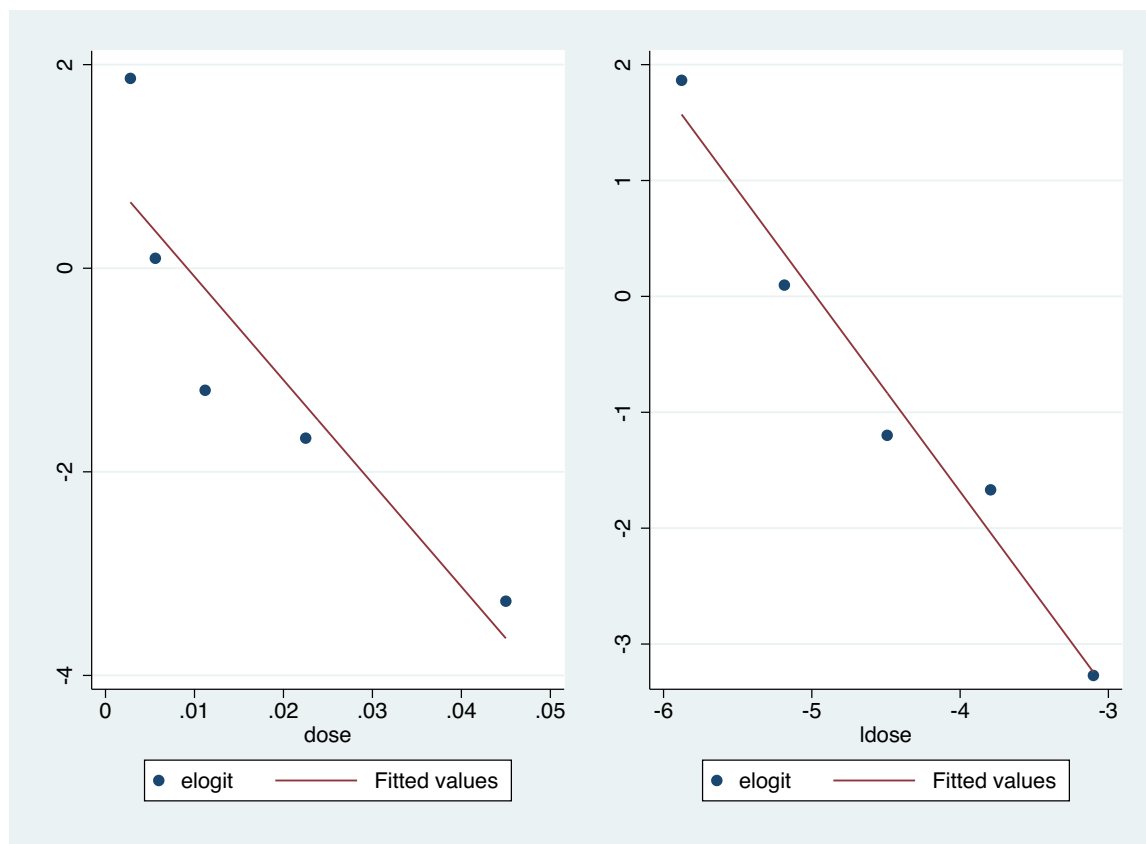
		OIM					
	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	ldose	-1.829621	.2545469	-7.19	0.000	-2.328524	-1.330719
	_cons	-9.189392	1.25511	-7.32	0.000	-11.64936	-6.729422

Next, evaluate the fit. The log(dose) model has a good fit. It also has a smaller AIC and a smaller BIC comparing with the dose model.

```
. display chi2tail(3,15.9000186)
. .00118878
. display chi2tail(3,2.80889589)
. .42203759
```

# Empirical logit against dose and logdose

- . twoway (scatter elogit dose) (lfit elogit dose), name(A,replace)
- . twoway (scatter elogit ldose) (lfit elogit ldose), name(B,replace)
- . graph combine A B



This plot shows why model with  $\log(\text{dose})$  is superior with respect to fit

# Example 3: Toxicity of cypermethrin – Model fit versus model complexity

We may consider powers of predictor variable  $X$  in the linear model.

**Table 5:** *Toxicity of cypermethrin:* Mortality/Disability of tobacco budworm moths 72 hours after exposure to cypermethrin

Sex of moth	Dose of cypermethrin	Number affected out of 20
Male	1.0	1
	2.0	4
	4.0	9
	8.0	13
	16.0	18
	32.0	20
Female	1.0	0
	2.0	2
	4.0	6
	8.0	10
	16.0	12
	32.0	16



# *Ex./Toxicity of cypermethrin to moths*

```
. list
```

	sex	dose	y	n
1.	1	1	1	20
2.	1	2	4	20
3.	1	4	9	20
4.	1	8	13	20
5.	1	16	18	20
6.	1	32	20	20
7.	2	1	0	20
8.	2	2	2	20
9.	2	4	6	20
10.	2	8	10	20
11.	2	16	12	20
12.	2	32	16	20

```
. gen female = sex-1
```

# Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{dose}$$

```
. glm y female dose, fam(binomial n) nolog
```

Generalized linear models

Optimization : ML

Deviance = 27.96797246

Pearson = 27.9327425

Variance function:  $V(u) = u*(1-u/n)$

Link function :  $g(u) = \ln(u/(n-u))$

No. of obs = 12

Residual df = 9

Scale parameter = 1

(1/df) Deviance = 3.107552

(1/df) Pearson = 3.103638

[ Binomial ]

[ Logit ]

Log likelihood = -29.03918673

AIC = 5.339864

BIC = 5.603813

y	OIM					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	-.9685483	.3295427	-2.94	0.003	-1.61444	-.3226565
dose	.1599557	.0234136	6.83	0.000	.1140658	.2058455
_cons	-1.166068	.2615487	-4.46	0.000	-1.678694	-.6534417

```
. di chi2tail(9, 27.96797246)
.00096569
```

A large deviance comparing with full model. Not a good fit. *Reject H<sub>0</sub>.*

*grouped data*

# Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{dose} + \beta_3 \text{dose}^2$$

```
. gen dose2 = dose^2
. glm y female dose dose2, fam(binomial n) nolog
```

Generalized linear models  
Optimization : ML

Deviance = 14.83516788  
Pearson = 12.31865123

No. of obs = 12  
Residual df = 8  
Scale parameter = 1  
(1/df) Deviance = 1.854396  
(1/df) Pearson = 1.539831

Variance function:  $V(u) = u \cdot (1 - u/n)$   
Link function :  $g(u) = \ln(u/(n-u))$

[Binomial]  
[Logit]

Log likelihood = -22.47278444

AIC = 4.412131  
BIC = -5.044085

	y	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
female		-1.071221	.3515239	-3.05	0.002	-1.760195	-.3822465
dose		.3735387	.0677652	5.51	0.000	.2407214	.5063561
dose2		-.0071597	.0019935	-3.59	0.000	-.0110668	-.0032526
_cons		-1.932169	.3596517	-5.37	0.000	-2.637073	-1.227265

```
. display chi2tail(1, 13.132805)
.00029017
```

$\Delta D = 27.96797246 - 14.83516788 = 13.132805$  and it is significant with  $p = 0.0029$ . By adding quadratic power of dose after adjusting for sex and dose. It implies a better fit for the bigger model (with  $\text{dose}^2$ ).

# Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{dose} + \beta_3 \text{dose}^2 + \beta_4 \text{dose}^3$$

```
. gen dose3 = dose^3
. glm y female dose dose2 dose3, fam(binomial n) nolog
```

Generalized linear models	No. of obs	=	12
Optimization : ML	Residual df	=	7
	Scale parameter	=	1
Deviance	=	8.180102569	(1/df) Deviance = 1.168586
Pearson	=	6.505512529	(1/df) Pearson = .9293589
Variance function: V(u) = u*(1-u/n)	[ Binomial]		
Link function : g(u) = ln(u/(n-u))	[ Logit]		
Log likelihood	=	-19.14525179	AIC = 4.024209
			BIC = -9.214244

	y	Coef.	OIM Std. Err.	z	P> z	[95% Conf. Interval]	
female		-1.097325	.3557186	-3.08	0.002	-1.794521	-.4001293
dose		.8527524	.2080075	4.10	0.000	.4450652	1.26044
dose2		-.0494398	.016956	-2.92	0.004	-.082673	-.0162067
dose3		.0008912	.0003504	2.54	0.011	.0002043	.001578
_cons		-2.989358	.5939998	-5.03	0.000	-4.153576	-1.82514

```
. di chi2tail(1, 6.6550653)
.00988743
```

$\Delta D = 14.83516788 - 8.180102569 = 6.6550653$  with  $p = 0.0099$ . It implies a better fit for the model with dose cubed.

# Toxicity of cypermethrin to moths: $\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{dose} + \beta_3 \text{dose}^2 + \beta_4 \text{dose}^3 + \beta_5 \text{dose}^4$

```
. gen dose4 = dose^4
. glm y female dose dose2 dose3 dose4, fam(binomial n) nolog
```

Generalized linear models	No. of obs	=	12
Optimization : ML	Residual df	=	6
	Scale parameter	=	1
Deviance	(1/df) Deviance	=	.919201
Pearson	(1/df) Pearson	=	.6809983
Variance function: $V(u) = u(1-u/n)$			
Link function : $g(u) = \ln(u/(n-u))$			
[Binomial]			
[Logit]			
Log likelihood	AIC	=	3.968801
	BIC	=	-9.394234

		OIM				
y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female	-1.092711	.3536025	-3.09	0.002	-1.785759	-.3996628
dose	1.962322	.7458292	2.63	0.009	.5005237	3.42412
dose2	-.2511731	.1286264	-1.95	0.051	-.5032762	.00093
dose3	.0129631	.0075796	1.71	0.087	-.0018926	.0278188
dose4	-.0002126	.0001329	-1.60	0.110	-.0004732	.0000479
_cons	-4.535646	1.222658	-3.71	0.000	-6.932011	-2.13928

```
. di chi2tail(1, 2.66489655)
.10258449
```

$\Delta D = 8.180102569 - 5.515206088 = 2.6648965$  with  $p = 0.103$  by adding quartic power of dose after adjusting for sex, dose, dose<sup>2</sup> and dose<sup>3</sup>. *Fail to Reject H<sub>0</sub>*

The 4th power of dose term do not add much to the model fit. The log likelihood increases (deviance decreases) when adding more predictors.

# Modeling Summary: Toxicity of cypermethrin to moths

Table 6: Model comparison using deviance.

Source of variation	$\Delta D$	$\Delta$ d.f.	p-value
$dose^2$   <i>sex, dose</i>	13.132805	1	0.00029017
$dose^3$   <i>sex, dose, dose<sup>2</sup></i>	6.6550653	1	0.00988743
$dose^4$   <i>sex, dose, dose<sup>2</sup>, dose<sup>3</sup></i>	2.6648965	1	.10258449

The deviance for the model

$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \text{dose} + \beta_3 \text{dose}^2 + \beta_4 \text{dose}^3$  is 8.180102569, with d.f. 7.

```
. di chi2tail(7, 8.180102569)  
.316981
```

This is a fair fit. Analysis suggests that there is evidence of linear relationship between the log odds of the response probability and the cubic function (or non-linear function) of dose of cypermethrin.

# Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \log(\text{dose})$$

## An alternate model

```
. gen ldose = log(dose)
```

```
. glm y female ldose, fam(binomial n) nolog
```

Generalized linear models

Optimization : ML

Deviance = 6.757064232

Pearson = 5.306017059

No. of obs = 12

Residual df = 9

Scale parameter = 1

(1/df) Deviance = .7507849

(1/df) Pearson = .5895575

Variance function:  $V(u) = u*(1-u/n)$

Link function :  $g(u) = \ln(u/(n-u))$

[ Binomial ]

[ Logit ]

Log likelihood = -18.43373262

AIC = 3.572289

BIC = -15.6071

		OIM					
y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
female	-1.100743	.3558271	-3.09	0.002	-1.798152	-.403335	
ldose	1.535336	.1891048	8.12	0.000	1.164698	1.905975	
_cons	-2.372412	.3855109	-6.15	0.000	-3.127999	-1.616825	

```
. display chi2tail(9,6.757064232)
```

.66239575 "Insignificant difference to 'perfect'"

# Toxicity of cypermethrin to moths: Sex-logDose interaction

## What about Sex-logDose interaction?

```
. glm y ldose if female==1, fam(binomial n) nolog  
...
```

Female

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ldose	1.307134	.2410767	5.42	0.000	.8346325	1.779636
_cons	-2.993542	.5526998	-5.42	0.000	-4.076813	-1.91027

```
. glm y ldose if female==0, fam(binomial n) nolog
```

Male

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ldose	1.81628	.3059458	5.94	0.000	1.216637	2.415923
_cons	-2.818555	.5479868	-5.14	0.000	-3.892589	-1.744521

The slopes for female and male are somewhat different but not by a lot.



# Toxicity of cypermethrin: Sex-logDose interaction

```
. gen female_ldose=female*ldose
```

```
. glm y ldose female female_ldose, fam(binomial n) nolog
```

Generalized linear models

Number of obs = 12

Deviance = 4.993727308

(1/df) Deviance = .6242159

Log likelihood = -17.55206415

y	OIM					
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
ldose	1.81628	.3059458	5.94	0.000	1.216637	2.415923
female	-.1749868	.7783101	-0.22	0.822	-1.700447	1.350473
female_ldose	-.5091458	.3895136	-1.31	0.191	-1.272578	.2542868
_cons	-2.818555	.5479868	-5.14	0.000	-3.892589	-1.744521

Interaction

V.S.

```
. glm y ldose female, fam(binomial n) nolog
```

Deviance = 6.757064232

(1/df) Deviance = .7507849

No  
Interaction.

...  
Difference in deviance for 2 nested  
. di ( 6.757064232-4.993727308 ) → No need to  $\chi^2$ : Already calculated  
1.7633369

```
. di chi2tail(1, 1.7633369)
```

.18420884

Reject  $H_0$ : No statistical significance between the 2 nested models.

$\Delta D = 1.76$  with a  $p$ -value of 0.18. The interaction effect is not significant.

## Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \log(\text{dose})$$

When comparing (non-nested) models with different number of parameters, we have to consider both deviance (model fit) and the number of parameters (model complexity). The model with less complexity AND better fit is preferred.

In this example, the log dose model has a deviance of  $D_{\log(\text{dose})} = 6.76$  and a deviance  $P$ -value of 0.66 (a good fit comparing with full model), using only 3 parameters. Comparing with the deviance of the cubic dose model ( $D_{\text{dose}^3} = 8.18$ ) with 5 parameters, the log dose model fits the data very well.

The good fit suggests that there is a linear relationship between the log odds of the response and the logarithm of dose

# Toxicity of cypermethrin to moths:

$$\text{logit}(p) = \beta_0 + \beta_1 \text{female} + \beta_2 \log(\text{dose})$$

The estimated model is  $\text{logit}(\hat{p}) = \hat{\beta}_0 + \hat{\beta}_1 \text{female} + \hat{\beta}_2 \log(\text{dose})$ . What are the interpretations of the estimated coefficients?

- The estimate  $\hat{\beta}_0$  is the estimated *log of odds* of being affected for male with dose=1 (i.e., *log of dose being zero*). We may also exponentiate it and interpret it as “the *estimated odds of being affected for male with dose=1 is  $\exp(\hat{\beta}_0)$* .”
- The estimate  $\hat{\beta}_1$  is the estimated *log odds ratio* of being affected comparing female versus male at any given dose. The *odds ratio* is  $\exp(\hat{\beta}_1)$  when comparing female versus male at any given dose.
- The estimate  $\hat{\beta}_2$  is the estimated *log odds ratio* for *one unit increase in log(dose)* for either sex. The odds ratio is  $2^{\hat{\beta}_2}$  for each *doubling of dose* for either sex (see Slide 35 in Lecture 4 for details).

*Interpret based on original meaning*

# Example 4: ESR data – comparing nested models with ungrouped binary data

**Determination of the ESR:** the data (ungrouped binary data with a continuous predictor) were obtained in order to study the extent to which the disease state of an individual, reflected in the ESR (the erythrocyte sedimentation rate) reading, is related to the level of a plasma protein, fibrinogen. The outcome is whether each individual has an ESR reading greater than 20 (implying inflammation). There are 32 observations each with a unique fib value.

. list in 8/15

	indivi~l	fib	y
8.	8	2.21	0
9.	9	3.15	0
10.	10	2.6	0
11.	11	2.29	0
12.	12	2.35	0
13.	13	5.06	1
14.	14	3.34	1
15.	15	2.38	1

For ungrouped binary data, the deviance itself would not be approximately  $\chi^2_{n-p}$  distributed but  $\Delta D$  can be used to compare nested models.

Here we compare M1:  $\text{logit}(p) = \beta_0 + \beta_1 \times \text{fib}$  versus M0:  $\text{logit}(p) = \beta_0$ .

```
. glm y fib , family (binomial) nolog
```

Generalized linear models		Number of obs	=	32
Optimization	: ML	Residual df	=	30
Deviance	= 24.84035674	Scale parameter	=	1
...		(1/df) Deviance	=	.8280119
Log likelihood	= -12.42017837	AIC	=	.9012611
		BIC	=	-79.13172

*ungrouped data:  
can't use deviance.*

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fib	1.827081	.9008558	2.03	0.043	.0614359	3.592726
_cons	-6.845075	2.770287	-2.47	0.013	-12.27474	-1.415412

```
. glm y, family (binomial) nolog
```

Generalized linear models		Number of obs	=	32
Optimization	: ML	Residual df	=	31
Deviance	= 30.88496417	Scale parameter	=	1
...		(1/df) Deviance	=	.9962892
Log likelihood	= -15.44248209	AIC	=	1.027655
.....(omitted)		BIC	=	-76.55285

```
. di chi2tail(1, 30.88496417 - 24.84035674 )  
.01394883
```

*✓ reduced      ✓ full*  
*\* Difference in deviance applies for 2 ungrouped nested*

# The likelihood ratio test

The likelihood ratio test for all predictors is also one output of the `logit` function.

```
. logit y fib , nolog
```

Logistic regression

Number of obs = 32  
LR chi2(1) = 6.04  
Prob > chi2 = 0.0139  
Pseudo R2 = 0.1957

Log likelihood = -12.420178

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
fib	1.827081	.9008558	2.03	0.043	.0614358	3.592726
_cons	-6.845075	2.770287	-2.47	0.013	-12.27474	-1.415412

```
. di -2*(-15.442482+12.420178)  
6.044608
```

*Remember to x2 to calculate deviance*

## Model Comparison

- The *difference in deviance statistic*,  $\Delta D$ , is used to compare **nested models**, and guide the choice of which covariates/predictors.
- Testing difference in deviance for two nested models is a **likelihood ratio** test. It can be used for any binary data including grouped, ungrouped, sparse binary data.
- As a general fit statistic, deviance statistic  $D$  can also be used to compare alternate predictor forms in an informal way.
- AIC and BIC are used for informal model comparison. A lower AIC/BIC implies a better fit.

$$\log(C-f) = \frac{\log C}{\log f} \} \text{Ratio}$$