

# Categorical Data Analysis

## Lecture 5 & 6

Graduate School of Advanced Science and Engineering  
Rei Monden

2024.12.19

# Contents of this lecture

## Today's contents

- ▶ Generalized Linear Models (GLM)
  - ▶ Components of a GLM
  - ▶ GLMs for binary data
  - ▶ GLMs for counts and rates
- ▶ Statistical inference and model checking
  - ▶ The deviance of a GLM
  - ▶ (Nested) Models comparison using the deviance

# Generalized linear models

# Generalized linear models

Generalized linear models (GLMs) encompass a large family of statistical models, including:

- ▶ Multiple linear regression
- ▶ ANOVA
- ▶ Logistic (also nominal, ordinal) regression
- ▶ Poisson regression
- ▶ Loglinear models
- ▶ ...

Today we go through some basics of GLMs.

Of course, we will focus on GLMs for *categorical* response variables.

## Components of a GLM

# Components of a GLM

GLMs have 3 components:

1. **Random component:**

Probability distribution of  $Y$ , the response variable.

For example:

- ▶  $Y = \#$  successes in a fixed number of trials: *Binomial* dist.
- ▶  $Y =$  counts per time period: *Poisson* or *negative binomial* dists.

2. **Linear predictor:**

Specifies the  $p$  predictors linearly:

$$\alpha + \beta_1 x_1 + \cdots + \beta_p x_p.$$

# Components of a GLM

## 3. Link function:

Function relating  $\mu = E(Y)$  to the linear predictor:

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_p x_p.$$

For example:

- ▶ Ordinary linear regression:  $g(\mu) = \mu$  (identity link function).
- ▶ Logistic regression:  $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$  (logit link function).
- ▶ Poisson regression:  $g(\mu) = \log(\mu)$  (log link function).

## GLMs for binary data



## GLMs for binary data

- ▶ Assume  $Y$  is dichotomous:  $1 = \text{success}$ ,  $0 = \text{failure}$ .
- ▶ Denote:
  - ▶  $\pi = P(Y = 1)$ : Probability of success
  - ▶  $1 - \pi = P(Y = 0)$ : Probability of failure
- ▶  $\mu = E(Y) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi$ .

Let's entertain two possible GLMs:

1. **Linear probability model**: Using the **identity** link function:

$$\underbrace{\pi}_{\mu} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p.$$

2. **Logistic regression model**: Using the **logit** linear function:

$$\underbrace{\log\left(\frac{\pi}{1 - \pi}\right)}_{\text{logit}(\mu)} = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p.$$

## Linear probability model

$$\pi = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p.$$

This is a GLM with *binomial* random component and *identity* link function.

Interpretation of regression coefficient  $\beta_i$ :

*For each unit increase in  $x_i$ ,  $\pi$  is expected to change by  $\beta_i$ , while controlling for the remaining predictors.*

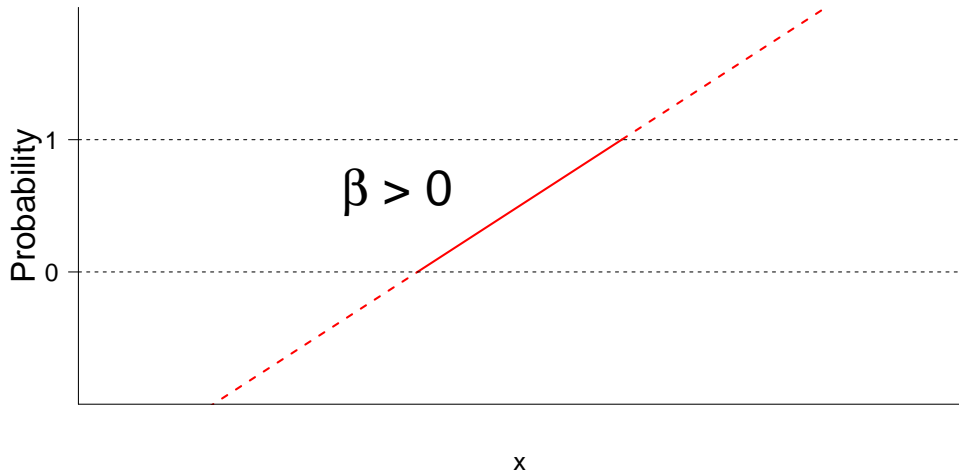
Major problem of this model:

*Depending on the values of the predictors, it can predict success probabilities  $\pi$  below 0 or above 1.*

Such predictions are uninterpretable.

## Linear probability model

$$\pi = \alpha + \beta x.$$



## Logistic regression model

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p.$$

This is a GLM with *binomial* random component and *logit* link function.

Solving the equation above with respect to  $\pi$  gives

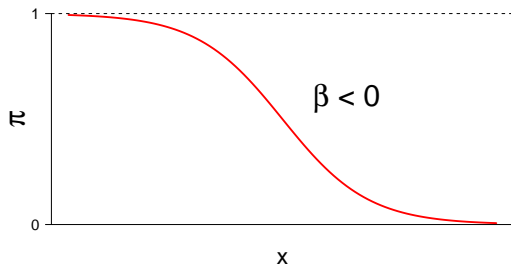
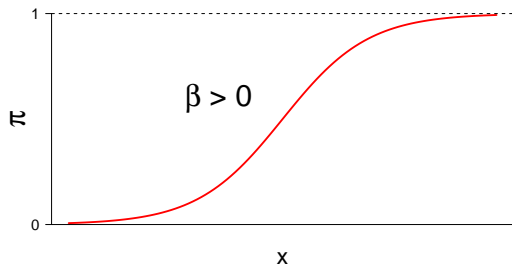
$$\pi = \frac{\exp(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p)}{1 + \exp(\alpha + \beta_1 x_1 + \cdots + \beta_p x_p)}$$

Clearly, no predictions outside the  $[0, 1]$  interval are now possible.  
In this respect, this model outperforms the linear probability model.

# Logistic regression model

$$\pi = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

This is a non-linear, S-shaped, model.



## GLMs for binary data – Example in R

Snoring	Heart Disease		Total
	Yes	No	
Never	24	1355	1379
Occasional	35	603	638
Nearly every night	21	192	213
Every night	30	224	254
			2484

Predictor: *Snoring*, categorical variable with 4 levels.

Outcome: *Heart Disease*, binary variable.

## GLMs for binary data – Example in R

```
# Create data frame:
HD.df <- data.frame(Yes = c(24, 35, 21, 30),
                    No  = c(1355, 603, 192, 224),
                    x   = c(0, 2, 4, 5) # 'continuous' Snoring
                    )
rownames(HD.df) <- c("Never", "Occasional",
                    "Nearly every night", "Every night"
                    )

HD.df
```

Output:

	Yes	No	x
Never	24	1355	0
Occasional	35	603	2
Nearly every night	21	192	4
Every night	30	224	5

## GLMs for binary data – Example in R

```
# Fit the linear probability model:  
glm(cbind(Yes, No) ~ x,  
    family = binomial(link = "identity"),  
    data    = HD.df  
)
```

Output:

```
-----  
Coefficients:  
(Intercept)          x  
    0.01725      0.01978  
-----
```



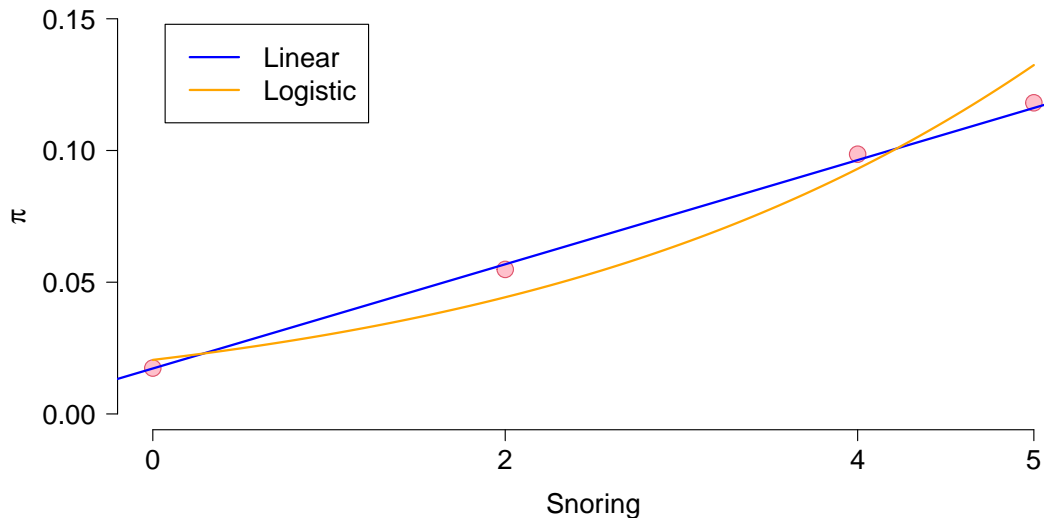
## GLMs for binary data – Example in R

```
# Fit logistic regression model:  
glm(cbind(Yes, No) ~ x,  
    family = binomial(link = "logit"),  
    data    = HD.df  
)
```

Output:

```
-----  
Coefficients:  
(Intercept)          x  
    -3.8662      0.3973  
-----
```

## GLMs for binary data – Example in R



## Exercise 5-1

Describe the purpose of the link function of a GLM.

Define the identity link and explain why it is not often used with a binomial parameter.

## Exercise 5-2

Use the table presented on the slide 14.

- (a) Re-fit the logistic regression model using the scores (i) (0, 2, 4, 6), (ii) (0, 1, 2, 3), (iii). (1, 2, 3, 4). Compare the model parameter estimates under the three choices. What can you conclude about the effect of linear transformations of scores that preserve the relative size of the spacing between scores?
- (b) Fit the logistic regression model using the scores (0, 2, 6, 7), approximating the number of days in a week that the subject snores. Compare the fitted values to those with the scores (0, 2, 4, 5) used in the text example. Do results seem to be sensitive to the choice of scores?

## GLMs for counts and rates

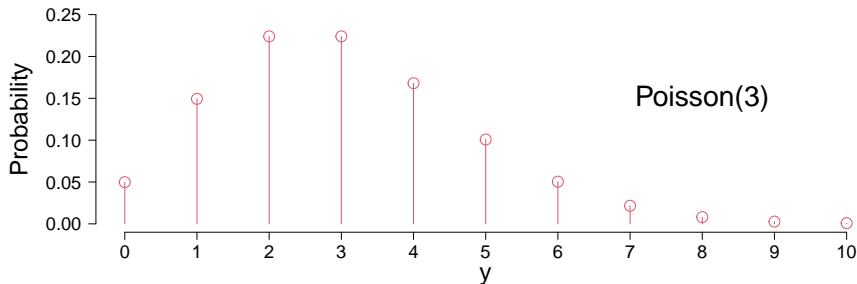
## GLMs for counts and rates

For counts, a commonly used random component is the **Poisson** distribution:

$$P(y) = \frac{e^{-\mu} \mu^y}{y!}, \text{ for } y = 0, 1, 2, \dots$$

**Property:**

$$E(Y) = \text{var}(Y) = \mu.$$



# Poisson loglinear model

The **Poisson loglinear model** combines the Poisson random component with the *log* link. For example, using only one predictor:

$$\log(\mu) = \alpha + \beta x.$$

Solving the equation with respect to  $\mu$  we have that

$$\mu = \exp(\alpha + \beta x) = e^{\alpha} (e^{\beta})^x.$$

## *Interpretation*

A one-unit increase in  $x$  either:

► *Changes* the logarithm of the expected value by  $\beta$

or

► *Multiplies* the expected value by  $e^{\beta}$ .

## GLMs for counts and rates – Example in R

Width	Satellites
28.3	8
22.5	0
26.0	9
⋮	⋮
28.0	0
27.0	0
24.5	0

173 rows in total.

Predictor: *Width*, shell width (continuous variable).

Outcome: *Satellites*, count variable.

```
# Import data frame from file:  
crab.df <- read.table("Crabs.dat", header = TRUE)
```



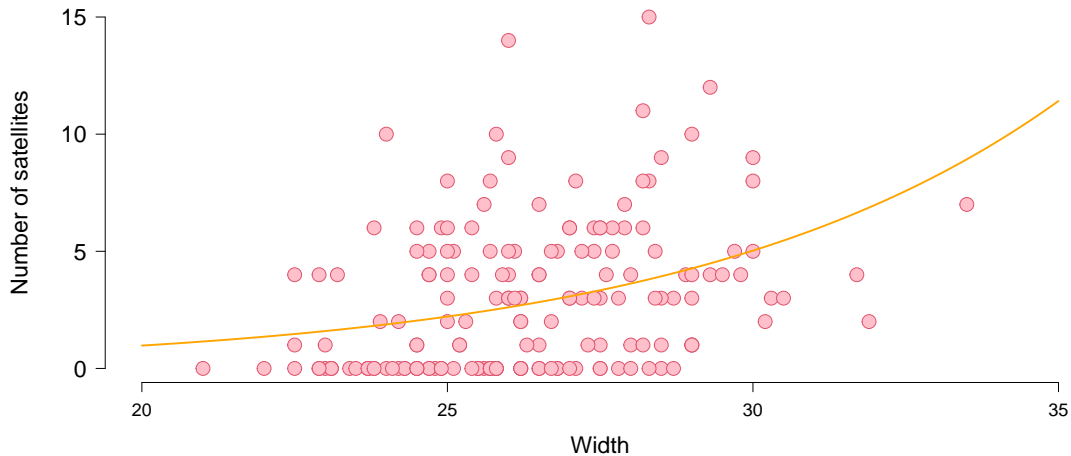
## GLMs for counts and rates – Example in R

```
# Fit poisson loglinear model:  
glm(sat ~ width,  
    family = poisson(link = "log"),  
    data   = crab.df  
)
```

Output:

```
-----  
Coefficients:  
(Intercept)      width  
      -3.305       0.164  
-----
```

## GLMs for counts and rates – Example in R



# Statistical inference and model checking

# Statistical inference and model checking

*Maximum likelihood* (ML) estimation is the method used to fit a GLM to data.

Here's one cool property:

*The same algorithm applies to all GLMs!!*

What we now discuss is various ways of performing **inference** and **model checking**.

And, because of the unified algorithm alluded to above, these techniques apply to *all* GLMs we will learn.

# Likelihood function

The secret for all tests available – Wald, score, and likelihood ratio – is the (log-)likelihood function.

Assuming a random component  $f(y; \beta)$  and assuming that the data  $\{y_1, \dots, y_n\}$  are independent, then the likelihood function is

$$l(\beta; y) = \prod_{i=1}^n f(y; \beta).$$

The log-likelihood is

$$L(\beta; y) = \log(l(\beta; y)) = \sum_{i=1}^n \log(f(y; \beta)).$$

For simplicity, let's think of  $\beta$  as a single parameter to be estimated.

# Likelihood function

Test:

$$\mathcal{H}_0 : \beta = 0 \quad \text{versus} \quad \mathcal{H}_1 : \beta \neq 0.$$

The various tests statistics – Wald, score, and likelihood ratio – use the following elements:

Score function	Hessian function	Information function
$U = \frac{\partial L}{\partial \beta}$	$H = \frac{\partial^2 L}{\partial \beta^2}$	$I = -\mathbb{E}(H)$

►  $\hat{\beta}$  = **MLE** of  $\beta$ : The value of  $\beta$  that maximizes  $L(\beta; y)$ .

►  $\text{var}(\hat{\beta}) = I(\hat{\beta})^{-1}$ .

$$SE(\hat{\beta}) = \sqrt{\text{var}(\hat{\beta})} = \sqrt{I(\hat{\beta})^{-1}}.$$

► Maximized log-likelihoods under  $\mathcal{H}_0$  and  $\mathcal{H}_1$ :

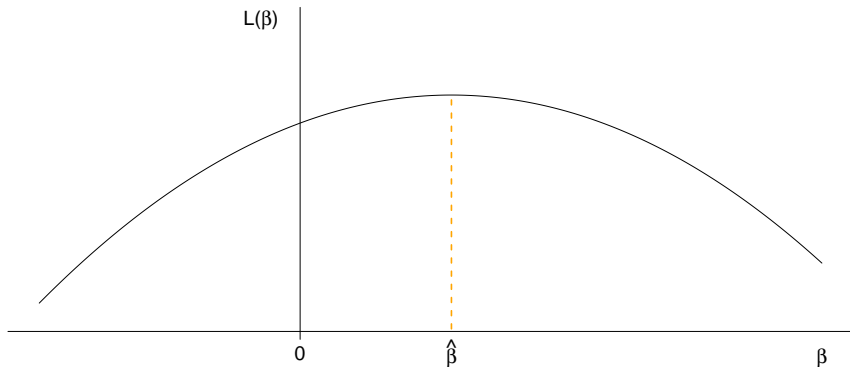
$$L_0 = L(0; y) \quad , \quad L_1 = L(\hat{\beta}; y)$$

# Likelihood function - Wald statistic

Wald test statistic:

$$z_W = \frac{\hat{\beta}}{SE(\hat{\beta})} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0, 1)$$

Based on the log-likelihood function evaluated only at the MLE,  $\hat{\beta}$ .

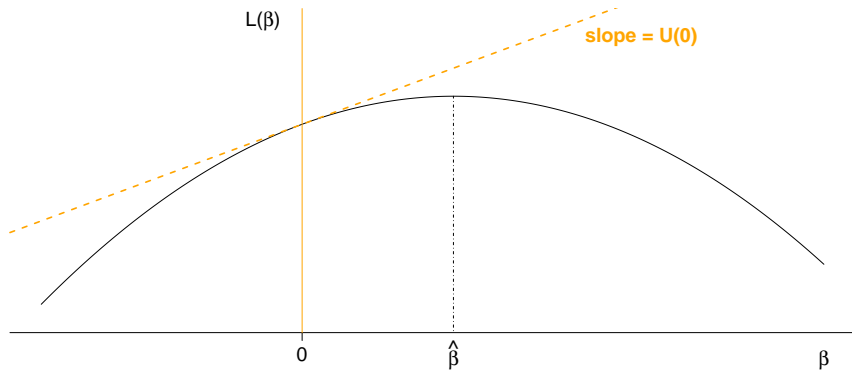


# Likelihood function - Score statistic

Score test statistic:

$$z_S = \frac{U(0)}{\sqrt{I(0)}} \underset{\mathcal{H}_0}{\sim} \mathcal{N}(0, 1)$$

Based on the *slope* of the tangent to  $L(\beta = 0; y)$  (i.e., under  $\mathcal{H}_0$ ).



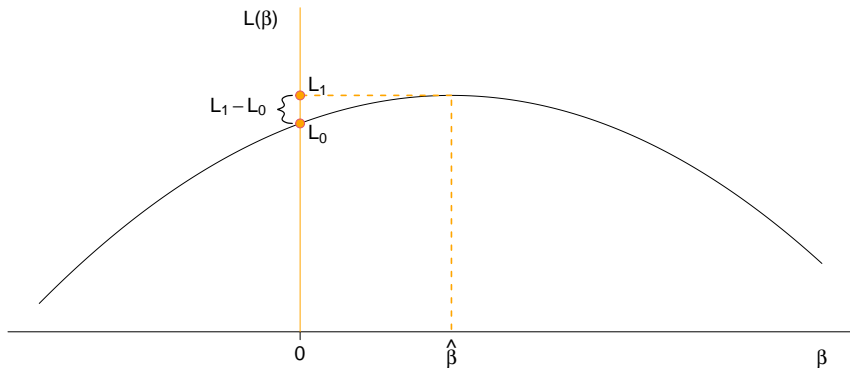


## Likelihood function - Likelihood ratio statistic

Likelihood ratio test statistic (asymptotic sampling distribution):

$$z_{LR}^2 = 2 \log \left( \frac{l_1}{l_0} \right) = 2 [\log(l_1) - \log(l_0)] = 2(L_1 - L_0) \underset{\mathcal{H}_0}{\sim} \chi^2(1)$$

Based on the log-likelihood function evaluated both at 0 and the MLE,  $\hat{\beta}$ .



## Wald, score, and likelihood ratio tests

$$z_W = \frac{\hat{\beta}}{SE(\hat{\beta})}, \quad z_S = \frac{U(0)}{\sqrt{I(0)}}, \quad z_{LR} = \sqrt{2(L_1 - L_0)}.$$

All test statistics are **asymptotically**  $\mathcal{N}(0, 1)$  distributed.

Equivalently,

$$z_W^2 = \left( \frac{\hat{\beta}}{SE(\hat{\beta})} \right)^2, \quad z_S^2 = \frac{U(0)^2}{I(0)}, \quad z_{LR}^2 = 2(L_1 - L_0)$$

are all asymptotically  $\chi^2(1)$  distributed.

## Confidence intervals – Wald, score, and likelihood ratio

Each test statistic has an associated  $100(1 - \alpha)\%$  CI.

The simplest to present is Wald's  $100(1 - \alpha)\%$  CI:

$$\hat{\beta} \pm z_{\alpha/2} \times SE(\hat{\beta}).$$

For all CIs we will rely on R for the computations.

*Note:*

The likelihood-ratio test-based CI for  $\beta$  is called a **profile likelihood CI**.

## Wald, score, and likelihood ratio methods

Which of the three methods is preferable?

- ▶ For large samples, they are all similar.
- ▶ Wald test is very simple and widely available in software.
- ▶ But, likelihood-ratio tests are better for small sample of very large effects.
- ▶ Score tests are not widely available, but work rather well too.

## Wald, score, and likelihood ratio – Example in R

Ideology	Evolution		Total
	Yes	No	
1 = Extremely conservative	11	37	48
2 = Conservative	46	104	150
3 = Slightly conservative	70	72	142
4 = Moderate	241	214	455
5 = Slightly liberal	78	36	114
6 = Liberal	89	24	113
7 = Extremely liberal	36	6	42
			1064

Predictor: (Political) *Ideology*, categorical variable with 7 levels, here treated as continuous.

Outcome: *Evolution*, binary variable.

## Wald, score, and likelihood ratio – Example in R

```
# Create data frame:
evol.df <- data.frame(
  Ideology = 1:7,
  Yes      = c(11, 46, 70, 241, 78, 89, 36),
  No       = c(37, 104, 72, 214, 36, 24, 6)
)

rownames(evol.df) <- c("Extremely conservative", "Conservative",
  "Slightly conservative", "Moderate",
  "Slightly liberal", "Liberal",
  "Extremely liberal"
)

evol.df
```

Output:

	Ideology	Yes	No
Extremely conservative	1	11	37
Conservative	2	46	104
-----			
Extremely liberal	7	36	6

## Wald, score, and likelihood ratio – Example in R

```
# Fit logistic regression model:  
evol.fit <- glm(cbind(Yes, No) ~ Ideology,  
               family = binomial(link = "logit"),  
               data    = evol.df  
               )  
summary(evol.fit)
```

Output:

```
-----  
Coefficients:  
             Estimate Std. Error z value Pr(>|z|)  
(Intercept) -1.75658    0.20500  -8.569  <2e-16 ***  
Ideology      0.49422    0.05092   9.706  <2e-16 *** <--  
-----
```

## Wald, score, and likelihood ratio – Example in R

Output:

-----  
Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.75658	0.20500	-8.569	<2e-16	***
Ideology	0.49422	0.05092	9.706	<2e-16	*** <--

-----

Some results:

►  $\hat{\beta}_{\text{Ideology}} = 0.49$

►  $SE(\hat{\beta}_{\text{Ideology}}) = 0.05$



## Wald, score, and likelihood ratio – Example in R

-----  
Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.75658	0.20500	-8.569	<2e-16 ***
Ideology	0.49422	0.05092	9.706	<2e-16 *** <--

-----

Wald test:

$$z_W = \frac{\hat{\beta}_{\text{Ideology}}}{SE(\hat{\beta}_{\text{Ideology}})} = \frac{0.49422}{0.05092} = 9.71.$$

Since  $p < .05$ , at significance level  $\alpha = .05$ , we reject the null hypothesis that  $\beta_{\text{Ideology}} = 0$ .

## Wald, score, and likelihood ratio – Example in R

Wald 95% CI:

```
confint.default(evol.fit)
```

Output:

	2.5 %	97.5 %
(Intercept)	-2.1583654	-1.3547930
Ideology	0.3944219	0.5940161

## Wald, score, and likelihood ratio – Example in R

Score test:

```
library(statmod)

# Fit the null model, i.e., with no predictors:
null.fit <- glm(cbind(Yes, No) ~ 1,
               family = binomial(link = "logit"),
               data = evol.df
             )

# Score 5% test:
z.S      <- glm.scoretest(null.fit, evol.df$Ideology); z.S
pvalue.S <- 2 * (1 - pnorm(z.S)); pvalue.S
```

Output:

```
10.20299 # z.S      = the score test statistic
0        # pvalue.S = the corresponding p-value (< .001)
```

Since  $p < .05$ , at significance level  $\alpha = .05$ , we reject the null hypothesis that  $\beta_{\text{Ideology}} = 0$ .

## Wald, score, and likelihood ratio – Example in R

Score CI:

```
library(glmtoolbox)

confint2(evol.fit, test = "score")
```

Output:

	Lower limit	Upper limit
(Intercept)	-2.15771	-1.35551
Ideology	0.39459	0.59386

## Wald, score, and likelihood ratio – Example in R

Likelihood-ratio test:

```
drop1(evol.fit, test = "LRT")
```

Output:

```
-----  
              Df Deviance      AIC    LRT  Pr(>Chi)  
Ideology    1    113.20 149.815 109.48 < 2.2e-16 ***  
-----
```

$\chi^2(1) = 109.48, p < .001$ :

At significance level  $\alpha = .05$ , we reject the null hypothesis that  $\beta_{\text{Ideology}} = 0$ .

## Wald, score, and likelihood ratio – Example in R

Likelihood-ratio CI, i.e., profile likelihood CI:

```
confint(evol.fit)
```

Output:

	2.5 %	97.5 %
(Intercept)	-2.165294	-1.3609733
Ideology	0.396166	0.5959414

## Exercise 5-3

An experiment analyzes imperfection rates for two processes used to fabricate silicon wafers for computer chips. For treatment A applied to 10 wafers, the numbers of imperfections are 8, 7, 6, 6, 3, 4, 7, 2, 3, 4. Treatment B applied to 10 other wafers has 9, 9, 8, 14, 8, 13, 11, 5, 7, 6 imperfections. Treat the counts as independent Poisson variates having means  $\mu_A$  and  $\mu_B$ . Consider the model  $\log\mu = \alpha + \beta x$ , where  $x = 1$  for treatment B and  $x = 0$  for treatment A.

- (a) Show that  $\beta = \log\mu_B - \log\mu_A = \log\left(\frac{\mu_B}{\mu_A}\right)$  and  $e^\beta = \frac{\mu_B}{\mu_A}$ .
- (b) Fit the model (recall slide 25). Report the prediction equation and interpret  $\hat{\beta}$ .
- (c) Use Wald's test to compare the hypotheses  $H_0 : \mu_A = \mu_B$  versus  $H_1 : \mu_A \neq \mu_B$ . Show how to compute Wald's test statistic using the fitted model from question (b). Interpret the result of the test (use significance level  $\alpha = 5\%$ ).

## The deviance of a GLM



# The deviance of a GLM

Notation:

- ▶  $M$  = our model of interest.  
From the previous example, it could be

$$\text{logit}(\pi) = \alpha + \beta \times \text{Ideology}.$$

- ▶  $S$  = **saturated** model:  
This model has one parameter per Ideology group (7 in total), so it fits data *perfectly*.  
In fact, this is an *over* fit.

From the previous example, that is

$$\text{logit}(\pi) = \beta_1 \times (\text{Ideology}=1) + \dots + \beta_7 \times (\text{Ideology}=7).$$

# The deviance of a GLM

The **deviance**,  $D$ , of model  $M$  is

$$D = 2(L_S - L_M),$$

where  $L$  refers to the log-likelihood.

*Result:*

If model  $M$  fits well enough (i.e., similarly to the saturated model  $S$ ), then

$$D \sim \chi^2(n - k),$$

where

- ▶  $n$  = number of parameters in model  $S$ ;
- ▶  $k$  = number of parameters in model  $M$ .

The deviance can therefore be used for comparing our model to the saturated model.

# The deviance of a GLM – Example in R

Model  $M$ :

$$\text{logit}(\pi) = \alpha + \beta \times \text{Ideology}$$

```
evol.fit <- glm(cbind(Yes, No) ~ Ideology,  
               family = binomial(link = "logit"),  
               data = evol.df  
             )  
logLik(evol.fit) #  $L_M$ 
```

Output:

```
'log Lik.' -19.16594 (df=2)
```

Therefore:

- ▶  $L_M = -19.166$ .
- ▶  $k = 2$ , corresponding to the two regression coefficients  $(\alpha, \beta)$ .

# The deviance of a GLM – Example in R

Model  $S$ :

$$\text{logit}(\pi) = \beta_1 \times (\text{Ideology}=1) + \dots + \beta_7 \times (\text{Ideology}=7)$$

```
sat.fit <- glm(cbind(Yes, No) ~ diag(7),  
              family = binomial(link = "logit"),  
              data    = evol.df  
            )  
logLik(sat.fit) #  $L_S$ 
```

Output:

```
'log Lik.' -17.30594 (df=7)
```

Therefore:

- ▶  $L_S = -17.306$ .
- ▶  $n = 7$ , corresponding to the total number of Ideology groups.

## The deviance of a GLM – Example in R

$$\begin{aligned} D &= 2(L_S - L_M) \\ &= 2[-17.306 - (-19.166)] \\ &= 3.72 \end{aligned}$$

- ▶ Assuming that model  $M$  fits as well as the saturated model, then

$$D \sim \chi^2(7 - 2) = \chi^2(5).$$

- ▶  $p\text{-value} = P(D > 3.72) = .59$ .

### Conclusion:

There is **no evidence** that model  $M$  fits worse than the saturated model.  
We therefore retain model  $M$  (= desirable outcome).

(Nested) Models comparison using the deviance

## (Nested) Models comparison using the deviance

The deviance can be used to compare any model of interest to the saturated model.

But we are often interested in directly comparing *two* models of interest, say,  $M_0$  and  $M_1$ .

We can use deviances for this, as long as  $M_0$  and  $M_1$  are **nested** (say,  $M_0$  is nested within  $M_1$ ).

This means that the models have:

- ▶ the same random component;
- ▶ the same link function;
- ▶ the linear component of  $M_0$  is a constrained version of that of  $M_1$ .

## (Nested) Models comparison using the deviance

For example:

- ▶  $M_0 : \text{logit}(\pi) = \alpha$   
(i.e., the intercept-only model, with no predictors).
- ▶  $M_1 : \text{logit}(\pi) = \alpha + \beta \times \text{Ideology}$

The linear component of  $M_0$ ,  $\alpha$ , is equal to  $\alpha + 0 \times \text{Ideology}$ , which is a constrained version of the linear component of  $M_1$ .

In this case,  $M_0$  is **nested** within  $M_1$ .



## (Nested) Models comparison using the deviance

Assume that the more complex model  $M_1$  fits the data *well enough*.

Q: Why?

A: If  $M_1$  does not fit well, then  $M_0$  (a constrained version of  $M_1$ ) will also not fit well, and we have nothing much to talk about.

Therefore... Assuming  $M_1$  fits well, then we know that

$$D_1 = 2(L_S - L_1) \sim \chi^2(n - k_1),$$

where  $k_1$  is the number of parameters in model  $M_1$ .

## (Nested) Models comparison using the deviance

The question is now whether  $M_0$ , a constrained (simpler!) version of  $M_1$ , also fits the data *well enough*.

► If yes, then

$$D_0 = 2(L_S - L_0) \sim \chi^2(n - k_0)$$

where  $k_0$  is the number of parameters of  $M_0$ , and

$$\begin{aligned} D_0 - D_1 &= 2(L_S - L_0) - 2(L_S - L_1) \\ &= 2(L_1 - L_0) \sim \chi^2(k_1 - k_0). \end{aligned}$$

► If *not*, then  $D_0 - D_1$  will be bigger than expected from  $\chi^2(k_1 - k_0)$ .

## (Nested) Models comparison using the deviance

### In summary:

To compare  $M_0$  (nested within  $M_1$ ) to  $M_1$  we use null hypothesis significance testing.

Assume a Type I error rate  $\alpha$  (e.g., 5%).

Procedure:

- ▶ Compute the test statistic:

$$D_0 - D_1 = 2(L_1 - L_0).$$

- ▶ Assume that  $M_0$  fits well (the null hypothesis).

In this case,  $D_0 - D_1 \sim \chi^2(k_1 - k_0)$ .

- ▶ Compute the  $p$ -value:  $p = p(\chi^2_{k_1 - k_0} > (D_0 - D_1))$ .

- ▶ If  $p < \alpha$  then we reject  $M_0$  and accept  $M_1$ .

- ▶ If  $p > \alpha$  then we fail to reject  $M_0$ .

This means that there is not evidence against the simpler model  $M_0$  and we retain it.

## (Nested) Models comparison using the deviance – In R

Model  $M_1$ :

$$\text{logit}(\pi) = \alpha + \beta \times \text{Ideology}$$

```
evol.fit <- glm(cbind(Yes, No) ~ Ideology,  
               family = binomial(link = "logit"),  
               data = evol.df  
             )  
logLik(evol.fit) # L_M1
```

Output:

```
'log Lik.' -19.16594 (df=2)
```

Therefore:

- ▶  $L_1 = -19.166$ .
- ▶  $k_1 = 2$ , corresponding to the two regression coefficients  $(\alpha, \beta)$ .

## (Nested) Models comparison using the deviance – In R

Model  $M_0$ :

$$\text{logit}(\pi) = \alpha$$

```
intercept.fit <- glm(cbind(Yes, No) ~ 1,  
                     family = binomial(link = "logit"),  
                     data = evol.df  
                     )  
logLik(intercept.fit) # L_M0
```

Output:

```
'log Lik.' -73.90733 (df=1)
```

Therefore:

- ▶  $L_0 = -73.907$ .
- ▶  $k_0 = 1$ , corresponding to intercept  $\alpha$ .

## (Nested) Models comparison using the deviance – In R

►  $M_0 : \text{logit}(\pi) = \alpha$

►  $M_1 : \text{logit}(\pi) = \alpha + \beta \times \text{Ideology}$

Assume that  $M_0$  fits the data well enough (i.e., as well as  $M_1$ ).

Then,

$$\begin{aligned} D_0 - D_1 &= 2(L_1 - L_0) \\ &= 2[-19.166 - (-73.907)] \\ &= 109.48 \sim \chi^2(2 - 1) = \chi^2(1) \end{aligned}$$

and

$$p\text{-value} = P(D_0 - D_1 > 109.48) < .001.$$

### Conclusion:

$M_0$ , albeit simpler, fits significantly worse than  $M_1$ .

We therefore reject  $M_0$  and accept  $M_1$ .

## (Nested) Models comparison using the deviance – In R

We actually had already seen this test result before today...

Recall:

*Likelihood-ratio test:*

```
drop1(evol.fit, test = "LRT")
```

*Output:*

```
-----  
              Df Deviance      AIC      LRT Pr(>Chi)  
Ideology    1    113.20 149.815 109.48 < 2.2e-16 ***  
-----
```

$\chi^2(1) = 109.48, p < .001$ :

At significance level  $\alpha = .05$ , we reject the null hypothesis that  $\beta_{\text{Ideology}} = 0$ .

This relation between likelihood ratio tests and comparison of nested models is quite general.

## Exercise 5-4

Use the same data from Exercise 5-3.

- (a) Fit the model  $\log \mu = \alpha$ . That is, fit a Poisson model with no predictors.
- (b) Observe that the intercept-only model from question (a) is *nested* within the model that you fitted in Exercise 5-3(b). Using R, compare both models using the deviance. Interpret the test result (use  $\alpha = 5\%$ ).



## In the next lecture

We are mainly going through Chapter 4, but the following sections will not be covered.

4.1.4, 4.3.3, 4.4.3, 4.6.1, 4.6.3.