Spring 2014
Stat 8004: Statistical Methods II
Lecture 10

1 Review of Last Lectures

Logistic Model:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha_0 + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} \tag{1}$$

Here $\pi_i = \mathbb{P}(Y_i = 1)$ is the probability that the binary outcome of interest taking the value of one.

The interpretation of the parameters:

- α_0 is the log odds of the reference group.
- β_j is the increase in the log odds caused by one unit increase in the jth covariate X_j, $j = 1, \ldots, p$. Let $\boldsymbol{x}^{(1)} = (x_1, \ldots, x_j, \ldots, x_p)$ and $\boldsymbol{x}^{(2)} = (x_1, \ldots, x_j + 1, \ldots, x_p)$. Then

$$\begin{aligned}\beta_j = \log\left(\frac{\pi(\boldsymbol{x}^{(2)})}{1-\pi(\boldsymbol{x}^{(2)})}\right) - \log\left(\frac{\pi(\boldsymbol{x}^{(1)})}{1-\pi(\boldsymbol{x}^{1})}\right) &= \log\left(\frac{\pi(\boldsymbol{x}^{(2)})/\{1-\pi(\boldsymbol{x}^{(2)})\}}{\pi(\boldsymbol{x}^{(1)})/\{1-\pi(\boldsymbol{x}^{(1)})\}}\right)\\ &= \log\left(\frac{\pi(\boldsymbol{x}^{(2)})}{1-\pi(\boldsymbol{x}^{(2)})}\right) - \log\left(\frac{\pi(\boldsymbol{x}^{(1)})}{1-\pi(\boldsymbol{x}^{(1)})}\right).\end{aligned}$$

Likelihood function for logistic model:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\frac{\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{1+\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right\}^{y_i}\left\{\frac{1}{1+\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right\}^{1-y_i} \tag{2}$$

Score function $U(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$.

$$
\begin{aligned}
U(\boldsymbol{\theta})_\alpha &= \frac{\partial l}{\partial \alpha} = \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{1+\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})} \\
&= \sum_{i=1}^n (y_i - \pi_i) = \sum_{i=1}^n [y_i - \mathbb{E}\,(y_i \mid \boldsymbol{x}_i)] = m_1 - \mathbb{E}\,(m_1 \mid \boldsymbol{\theta}), \\
U(\boldsymbol{\theta})_{\beta_j} &= \frac{\partial l}{\partial \beta_j} = \sum_{i=1}^n x_{ij} \left(y_i - \frac{\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{1+\exp(\alpha + \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})} \right) \\
&= \sum_{i=1}^n x_{ij}(y_i - \pi_i) = \sum_{i=1}^n x_{ij}\{y_i - \mathbb{E}\,(y_i \mid \boldsymbol{x}_i)\},
\end{aligned}
$$

To obtain MLE, we set $U(\hat{\boldsymbol{\theta}}) = 0$, and solve for $\hat{\boldsymbol{\theta}}$. The close-form solutions do not exist. The solution can be obtained by the iterated reweighted least squares algorithm.

2 Inference of Logistic Model for 2×2 Unconditional Contingency Table

the analysis of the 2x2 table from a study involving two independent groups of subjects, either from a cross-sectional, prospective, or retrospective study, unmatched.

Contingency Table:

	Group		
Response	1	2	
+	a	b	m_1
−	c	d	m_2
	n_1	n_2	N

or

	Group		
Response	1	2	
+	x_1	x_2	m_1
−	$n_1 - x_1$	$n_2 - x_2$	m_2
	n_1	n_2	N

2.1 Model, Likelihood, Score function and Information Matrix

WLOG, suppose this is a prospective study with n_1 and n_2 fixed. Under logit model, assume

$$
\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta, \quad \log\left(\frac{\pi_2}{1-\pi_2}\right) = \alpha.
$$

The inverse functions:

$$
\begin{aligned}
\pi_1 &= \frac{e^{\alpha+\beta}}{1+e^{\alpha+\beta}}, \quad 1-\pi_1 = \frac{1}{1+e^{\alpha+\beta}} \\
\pi_2 &= \frac{e^{\alpha}}{1+e^{\alpha}}, \quad 1-\pi_2 = \frac{1}{1+e^{\alpha}}
\end{aligned}
$$

and the odds ratio is

$$\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = e^{\beta}.$$

For prospective unconditional 2×2 contingency table, we can model the data by the produce binomial distribution.

$$L(\pi_1, \pi_2) \propto \pi_1^a(1-\pi_1)^c\pi_2^b(1-\pi_2)^d,$$

and the log likelihood is

$$l(\pi_1, \pi_2) = a\log(\pi_1) + c\log(1-\pi_1) + b\log(\pi_2) + d\log(1-\pi_2).$$

Exprssing the probabilities in terms of the parmater $\boldsymbol{\theta} = (\alpha, \beta)^{\mathrm{T}}$. Then,

$$l(\boldsymbol{\theta}) = m_1\alpha + a\beta - n_1\log(1+e^{\alpha+\beta}) - n_2\log(1+e^{\alpha}).$$

The score function $U(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ is

$$\begin{aligned} U(\boldsymbol{\theta})_\alpha &= \partial l/\partial\alpha = m_1 - n_1\pi_1 - n_2\pi_2, \\ U(\boldsymbol{\theta})_\beta &= \partial l/\partial\beta = a - n_1\pi_1. \end{aligned}$$

To get MLE, set $U(\hat{\boldsymbol{\theta}}) = 0$. Then

$$\hat{\alpha} = \log(b/d), \quad \hat{\beta} = \log(ad/bc).$$

For this simple case, we don't need to use iterated reweighted least squares and the close-form exists.

Now let's calculate the Hessian matrix $H(\boldsymbol{\theta}) = \partial^2 l/\partial\boldsymbol{\theta}^2$.

$$\begin{aligned} H(\boldsymbol{\theta})_\alpha &= \partial^2 l/\partial\alpha^2 = -n_1\pi_1(1-\pi_1) - n_2\pi_2(1-\pi_2) \\ H(\boldsymbol{\theta})_\beta &= \partial^2 l/\partial\beta^2 = -n_1\pi_1(1-\pi_1) \\ H(\boldsymbol{\theta})_{\alpha\beta} &= \partial^2 l/[\partial\alpha\partial\beta] = -n_1\pi_1(1-\pi_1). \end{aligned}$$

Thus, the expected information matrix $I(\boldsymbol{\theta}) = -H(\boldsymbol{\theta})$ is

$$I(\boldsymbol{\theta}) = \begin{pmatrix} n_1\psi_1 + n_2\psi_2 & n_1\psi_1 \\ n_1\psi_1 & n_1\psi_1 \end{pmatrix},$$

where $\psi_i = \pi_i(1-\pi_i)$ is the variance of the Bernoulli r.v. with mean parameter π_i.

Let $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^{\mathrm{T}}$. We know that $\mathrm{Var}(\hat{\boldsymbol{\theta}}) = I(\boldsymbol{\theta})^{-1}$. For a 2×2 invertible matrix A,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

the inverse A^{-1} is

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix},$$

where the determinant $|A| = a_{11}a_{22} - a_{12}a_{21}$. Therefore,

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}) = I(\boldsymbol{\theta})^{-1} = \begin{pmatrix} \frac{1}{n_2\psi_2} & -\frac{1}{n_2\psi_2} \\ -\frac{1}{n_2\psi_2} & \frac{n_1\psi_1+n_2\psi_2}{n_1\psi_1 n_2\psi_2} \end{pmatrix}.$$

Note that the estimates are correlated even thought the underlying likelihood is the product of two independent likelihoods. That is because each of these two bionoial likelihoods involves the intercept α.

2.2 Tests of Significance

Usually, we are interested to know whether $\pi_1 = \pi_2$. Equivalently, we are testing $\mathrm{H}_0: \ \beta = 0$.

2.2.1 Wald Test

Wald test statistics:

$$X_W^2 = \frac{\hat{\beta}^2}{\widehat{\mathrm{Var}}(\hat{\beta})}.$$

For 2×2 unconditional contingency table:

$$X_W^2 = \frac{[\log(ad/bc)]^2}{\left[\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right]}.$$

Since the Wald test uses the variace estimated under the alternative $\widehat{\mathrm{Var}}(\hat{\beta} \mid \mathrm{H}_1)$, this test does not equal to the Wald test we mentioned before for testing $\mathrm{H}_0: \ OR = 0$.

Under H_0, $X_W^2 \sim \chi^2(1)$.

2.2.2 Likelihood Ratio Test

For 2×2 table, under a logit model, the likeohood ratio test for testing $\mathrm{H}_0: \ \beta = 0$ is provided by

$$X_{LR}^2 = 2\log[L(\hat{\alpha}, \hat{\beta})] - 2\log[L(\hat{\alpha}_{|\beta=0})].$$

The null likelihood is

$$l(\alpha_{|\beta=0}) = \log[L(\alpha_{|\beta=0})] - N\log(1 + e^{\alpha}),$$

and the score equation is

$$U(\boldsymbol{\theta})_{\alpha_{|\beta=0}} = m_1 - N\frac{e^a}{1 + e^a},$$

which, when set to zero, yields the estimate

$$\hat{\alpha}_0 = \hat{\alpha}_{|\beta=0} = \log(m_1/m_2).$$

Thus, the likelihood ratio test statistic is

$$X^2_{LR}/2 = m_1\hat{\alpha} + a\hat{\beta} - n_1 \log\left(1 + e^{\hat{\alpha}+\hat{\beta}}\right) - n_2 \log\left(1 + e^{\hat{\alpha}}\right) - m_1\hat{\alpha}_0 + N \log\left(1 + e^{\hat{\alpha}_0}\right).$$

Under H_0, $X^2_{LR} \sim \chi^2(1)$.

2.3 Efficient Sore Test

Under $\mathrm{H}_0 : \ \beta = 0$, we set $U(\boldsymbol{\theta})_{\hat{\alpha}|\beta=0}$. Let $\hat{\boldsymbol{\theta}}_0 = (\hat{\alpha}_0, \beta_0)^{\mathrm{T}}$. Then

$$U(\hat{\boldsymbol{\theta}}_0)_\beta = a - \frac{n_1 e^{\hat{\alpha}_0}}{1 + e^{\hat{\alpha}_0}} = a - n_1 m_2/N = a - \hat{\mathbb{E}}\,(a \mid \mathrm{H}_0).$$

We also need $I(\boldsymbol{\theta})^{-1}$ under $\mathrm{H}_0 : \beta = 0$. Under H_0, $\pi_1 = \pi_2 = \pi$, and it follows that

$$\hat{\pi} = \frac{e^{\hat{\alpha}_0}}{1 + e^{\hat{\alpha}_0}} = \frac{m_1}{N}.$$

Therefore

$$I(\hat{\boldsymbol{\theta}}_0)^{-1} = \frac{1}{\hat{\pi}(1-\hat{\pi})} \cdot \frac{1}{N n_1 - n_1^2} \cdot \begin{pmatrix} n_1 & -n_1 \\ -n_1 & N \end{pmatrix}.$$

The efficient score test is $X^2 = U(\hat{\boldsymbol{\theta}}_0)^{\mathrm{T}} \left[I(\hat{\boldsymbol{\theta}}_0)^{-1}\right] U(\hat{\boldsymbol{\theta}}_0)$. For this case, the resulting efficient score test is

$$X^2 = \frac{[a - \hat{\mathbb{E}}\,(a \mid \mathrm{H}_0)]^2}{\hat{V}_u(a)},$$

where $\hat{V}_u(a) = m_1 m_2 n_1 n_2 / N^3$. It is equivalent to the Cochran's test.

3 Inference of Logistic Model under General Cases

Under general cases, the logistic model is as model (1). We can also define the Wald test, the likelihood ratio test and the efficient score test. But there might not exist the close-form expressions.

Example: Passengers on the Titanic. Please read the R code.

Test	H_0	Test Statistic	Null distribution
Wald	$\boldsymbol{\theta} = \boldsymbol{\theta}_0$	$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^{\mathrm{T}} I(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$	$\chi^2(r)$
LR	$\boldsymbol{\theta}_{r\|q} = 0$	$2 \log L(\hat{\boldsymbol{\theta}}) - 2 \log L(\hat{\boldsymbol{\theta}}_{0,q}, 0)$	$\chi^2(r)$
Score	$\boldsymbol{\theta}_{r\|q} = 0$	$U(\hat{\boldsymbol{\theta}}_{0,q})^{\mathrm{T}} [I(\boldsymbol{\theta}_{0,q})^{-1}] U(\hat{\boldsymbol{\theta}}_{0,q})$	$\chi^2(r)$

Table 1: Tests of Significant for Logistic Model. In the table, r = d.f. differences between null and alternative

4 Logistic Model with and without Interactions

In Lecture 9, we discussed using logistic model without interactions to analyze data for 2×2 stratified contigency tables. Here we extend the results to logistic mode with interactions.

Example: Knee Surgery.

Direct Injuries

	Response		
Exposure	Success	Partially Success	
New	40	30	70
Old	15	15	30
	55	45	100

Twist Injuries

	Response		
Exposure	Success	Partially Success	
New	15	5	20
Old	55	25	80
	70	30	100

Table 2: Stratiefied Data of Knee Injuries and Operations

For the ith patient ($i = 1, \ldots, n$), let

$$Y_i = \begin{cases} 1 & \text{if the } i\text{th patient had completely successful surgery} \\ 0 & \text{otherwise} \end{cases}$$

Let

$$x_i = \begin{cases} 1 & \text{if the } i\text{th patient had the new surgery} \\ 0 & \text{otherwise} \end{cases},$$

and

$$z_i = \begin{cases} 1 & \text{if the } i\text{th patient had direct knee injury} \\ 0 & \text{otherwise} \end{cases}.$$

Let $\pi_i = \mathbb{E}\,(Y_i)$.

4.1 Review: The logistic model without interaction

If we use the logistic model without interaction:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 x_i + \beta_2 z_i. \tag{3}$$

- α: the log odds of successful operation in the reference group, which is a group of twist injuries and taking the old treatment.
- β_1: the log odds difference between the new treatment and the old treatment for both injury groups (common odds ratio).
- β_2: the log odds difference between the direct injuries and the twist injuries.

We can further summerized the results in Table 3.

Category	x_i	z_i	Odds	π_i
1	0	0	e^{α}	$\frac{e^{\alpha}}{1+e^{\alpha}}$
2	1	0	$e^{\alpha+\beta_1}$	$\frac{e^{\alpha+\beta_1}}{1+e^{\alpha+\beta_1}}$
3	0	1	$e^{\alpha+\beta_2}$	$\frac{e^{\alpha+\beta_2}}{1+e^{\alpha+\beta_2}}$
4	1	1	$e^{\alpha+\beta_1+\beta_2}$	$\frac{e^{\alpha+\beta_1+\beta_2}}{1+e^{\alpha+\beta_1+\beta_2}}$

Table 3: Odds and probabilities for the model with no interaction

4.2 The logistic mode with interaction

Consider the following model with interaction:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + x_i\beta_1 + z_i\beta_2 + x_i z_i\beta_3. \tag{4}$$

- α: the log odds of successful operation in the reference group, which is a group of twist injuries and taking the old treatment.
- β_1: the log odds difference between the new treatment and the old treatment in the twist injury group.
- β_2: the log odds difference of the old treatment between the direct injuries and the twist injuries
- β_3: the extra log odds difference between the new and old treatment in the direct injury group.

Category	x_i	z_i	Odds	π_i
1	0	0	e^{α}	$\frac{e^{\alpha}}{1+e^{\alpha}}$
2	1	0	$e^{\alpha+\beta_1}$	$\frac{e^{\alpha+\beta_1}}{1+e^{\alpha+\beta_1}}$
3	0	1	$e^{\alpha+\beta_2}$	$\frac{e^{\alpha+\beta_2}}{1+e^{\alpha+\beta_2}}$
4	1	1	$e^{\alpha+\beta_1+\beta_2+\beta_3}$	$\frac{e^{\alpha+\beta_1+\beta_2+\beta_3}}{1+e^{\alpha+\beta_1+\beta_2+\beta_3}}$

Table 4: Odds and probabilities for the model with no interaction

In this model, the odds ratio in two injury groups are different. The odds ratio between two treatments in the twist injury group is $\exp(\beta_1)$, which can be estimated by $\exp(\hat{\beta}_1)$; and the odds ratio in the injury group is $\exp(\beta_1 + \beta_3)$, which can be estimated by $\exp(\hat{\beta}_1 + \hat{\beta}_3)$.

```
> fit.int <- glm(success ~ trt + injury + trt*injury, data =
  knee, family = binomial)
> summary(fit.int)

Call:
glm(formula = success ~ trt + injury + trt * injury, family =
  binomial,
    data = knee)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6651  -1.3018   0.8657   1.0579   1.1774

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.78846    0.24121   3.269  0.00108 **
trt1          0.31015    0.56995   0.544  0.58632
injury1      -0.78846    0.43762  -1.802  0.07160 .
trt1:injury1 -0.02247    0.71869  -0.031  0.97505
---
Signif. codes:  0   ***   0.001   **   0.01   *   0.05
   .   0.1      1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 264.63  on 199  degrees of freedom
Residual deviance: 259.06  on 196  degrees of freedom
AIC: 267.06

Number of Fisher Scoring iterations: 4

> exp(fit.int$coef[2]) # odds ratio in twist injury group
```

```
    trt1
1.363636
> exp(fit.int$coef[2] + fit.int$coef[4]) # odds ratio in
  direct injury group
    trt1
1.333333
```

If we assume the odds ratios are different in two injury type groups, then the estimates of the odds ratio in the twist group is 1.36, and in the direct injury group is 1.33.

On the other hand, the interaction term is not significant in the model, indicating that the odds ratios might be the same across two injury groups. So for this data set, it is better to use Model (3).

5 Inference for 2×2 Contidional Contingency Table

For 2×2 conditional contingency table, we should use hypergeometric distribution to model the data.

$$L_c(\varphi) = P(a \mid n_1, m_1, N, \phi) = \frac{\binom{n_1}{a}\binom{N-n_1}{m_1-a}\varphi^a}{\sum_{i=a_l}^{a_u}\binom{n_1}{i}\binom{N-n_1}{m_1-i}\varphi^i},$$

where $a_l = \max(0, m_1 - n_2)$ and $a_u = \min(m_1, n_1)$.

Let $\beta = \log(\varphi)$. The log likelihood function is

$$l(\beta) = a\beta - \log\left[\sum_{i=a_l}^{a_u}\binom{n_1}{i}\binom{N-n_1}{m_1-i}e^{i\beta}\right].$$

And the score equation is

$$U(\beta) = l'(\beta) = a - \frac{\sum_{i=a_l}^{a_u}\binom{n_1}{i}\binom{N-n_1}{m_1-i}ie^{i\beta}}{\sum_{i=a_l}^{a_u}\binom{n_1}{i}\binom{N-n_1}{m_1-i}e^{i\beta}} = a - E(a \mid \beta).$$

Set $U(\hat{\beta}) = 0$ and there is no close-form solution. owever, the solution is readily obtained by an iterative procedure such as Newton-Raphson iteration described subsequently. To do so requires the expressions for the Hessian or observed information. Using a summary notation such that

$$C_i = \binom{n_1}{i}\binom{N-n_1}{m_1-i}.$$

Then the Fisher's information is

$$I(\beta) = -l''(\beta) = \frac{\left[\sum_i C_i e^{i\beta}\right]\left[\sum_i i^2 C_i e^{i\beta}\right] - \left[\sum_i iC_i e^{i\beta}\right]^2}{\left[\sum_i C_i e^{i\beta}\right]^2} = \frac{\sum_i i^2 C_i e^{i\beta}}{\sum_i C_i e^{i\beta}} - \left[\frac{\sum_i iC_i e^{i\beta}}{\sum_i C_i e^{i\beta}}\right]^2.$$

Thus the expected information is

$$I(\beta) = \mathbb{E}\,(a^2 \mid \beta) - \mathbb{E}\,(a \mid \beta)^2 = V(a \mid \beta).$$

Although the MLE must be solved for iteratively, the score test $\mathrm{H}_0 : \beta = \beta_0 = 0$ is readily obtained directly.

$$\begin{aligned} U(\beta_0) &= a - E(a \mid \mathrm{H}_0) = a - \frac{m_1 n_1}{N} \\ I(\beta_0) &= V(a \mid \beta_0) = V_c(a) = \frac{n_1 n_2 m_1 m_2}{N^2(N-1)}, \end{aligned}$$

and thus the score test is

$$X^2 = \frac{U(\beta_0)^2}{I(\beta_0)} = \frac{[a - \mathbb{E}\,(a)]^2}{V_c(a)},$$

which is the Mantel-Haenszel test.