



JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Lecture 4

Assessing confounding, MLE and inference in logistic regression models

Lecture 3 Review

- ▶ Non-linearity effect in logistic regression models
 - ▶ Assume X and Z are independent, i.e. no confounding
 - ▶ $|\beta_{1c}| > |\beta_{1m}|$, difference depends on association between X – Y and Z – Y and Var(Z)
- ▶ Impact on assessing confounding:
 - ▶ $|\beta_{1c}| < |\beta_{1m}|$, “positive confounding” despite the non-linearity effect
 - ▶ $|\beta_{1c}| > |\beta_{1m}|$ and $Z_c > Z_m$ “negative confounding”
 - ▶ β_{1c} and β_{1m} have different signs! Same for Z statistics. “qualitative confounding”
 - ▶ In cases where $|\beta_{1c}| > |\beta_{1m}|$ and $Z_c = Z_m$ “non-linearity effect”



Exercise

- ▶ Open Lecture4-Handout.Rmd
- ▶ Determine which, if any, of the following variables are confounders for the “big expenditure” vs. MSCD relationship
 - ▶ Education, marital status, poverty status, seatbelt use, geographic region or **ever smoker**
- ▶ Write a paragraph summarizing your analysis
 - ▶ Start by summarizing the outcome, exposure and potential confounding variables
 - ▶ Present the findings from the marginal model
 - ▶ Present the findings from the conditional models with quantitative support for or against each variable’s confounder status.



Summary of analysis

- ▶ Evidence of confounding



Summary of analysis

- ▶ Written summary



MLE in logistic models

Assume the following model:

- $Y_i \sim \text{Bernoulli}(\mu_i)$ for $i = 1, \dots, n$ independent observations.
- Define the vector of covariates for subject i as $x_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})$.
- Define the vector of association parameters $\beta = (\beta_0, \beta_1, \dots, \beta_p)$.
- Assume the logit link such that:

$$\log\left(\frac{\mu_i}{1 - \mu_i}\right) = x_i^t \beta \rightarrow \mu_i = \frac{e^{x_i^t \beta}}{1 + e^{x_i^t \beta}}$$

NOTE: We should really write $\mu_i(x_i, \beta)$ i.e. μ_i is a function of x_i and β . In this handout, I will simplify this to $\mu_i(\beta)$.



MLE in logistic models

We can express the likelihood function as:

$$\begin{aligned}L(\beta|y) &= Pr(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n|\beta) \\&= \prod_{i=1}^n Pr(Y_i = y_i|\beta) \\&= \prod_{i=1}^n \mu_i(\beta)^{y_i} [1 - \mu_i(\beta)]^{1-y_i}\end{aligned}$$

The log-likelihood function is:

$$\log[L(\beta|y)] = \sum_{i=1}^n y_i \log[\mu_i(\beta)] + (1 - y_i) \log[1 - \mu_i(\beta)]$$



MLE in logistic models

The score equation, $U(\beta)$ is the derivative of the log-likelihood function with respect to β .

$$\begin{aligned}U(\beta) &= \frac{\partial \log[L(\beta|y)]}{\partial \beta} \\&= \sum_{i=1}^n y_i \frac{\partial \log[\mu_i(\beta)]}{\partial \beta} + (1 - y_i) \frac{\partial \log[1 - \mu_i(\beta)]}{\partial \beta} \\&= \sum_{i=1}^n y_i (x_i [1 - \mu_i(\beta)]) + (1 - y_i) [-\mu_i(\beta) x_i] \\&= \sum_{i=1}^n x_i (y_i - y_i \mu_i(\beta) + (-\mu_i(\beta)) + y_i \mu_i(\beta)) \\&= \sum_{i=1}^n x_i (y_i - \mu_i(\beta)) \\&= X^t(Y - \mu(\beta))\end{aligned}$$

MLE in logistic models

NOTE: We will also need to know $U'(\beta) = \frac{\partial U(\beta)}{\partial \beta}$

$$\begin{aligned}U'(\beta) &= \frac{\partial U(\beta)}{\partial \beta} \\&= \frac{\partial}{\partial \beta} X' (Y - \mu(\beta)) \\&= -X' \frac{\partial \mu_i(\beta)}{\partial \beta} \\&= -X' V X\end{aligned}$$

where we already showed that:

$$\frac{\partial \mu_i(\beta)}{\partial \beta} = \mu_i(\beta) \frac{\partial \log[\mu_i(\beta)]}{\partial \beta} = \mu_i(\beta)(1 - \mu_i(\beta))x_i$$

and $V_{n \times n} = \text{diag}(\mu_i(\beta)[1 - \mu_i(\beta)])$.



Newton-Raphson Method to find “beta”

- Step 0: Pick an initial starting value for β , call this $\hat{\beta}^{(k)}$.
- Step 1: Compute the slope of $U(\beta)$ at $\hat{\beta}^{(k)}$, i.e. compute $U'(\hat{\beta}^{(k)})$.
- Step 2: Construct the tangent line, which is a line that passes through the points $(\hat{\beta}^{(k)}, U(\hat{\beta}^{(k)}))$ and $(\hat{\beta}^{(k+1)}, 0)$ and has slope $U'(\hat{\beta}^{(k)})$.
- Step 3: Solve the following for $\hat{\beta}^{(k+1)}$:

$$U'(\hat{\beta}^{(k)}) = \frac{U(\hat{\beta}^{(k)}) - 0}{\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)}}$$

$$[\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)}]U'(\hat{\beta}^{(k)}) = U(\hat{\beta}^{(k)})$$

$$\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)} = U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)})$$

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} - U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)})$$

$$= U'(\hat{\beta}^{(k)})^{-1} \left(U'(\hat{\beta}^{(k)})\hat{\beta}^{(k)} - U(\hat{\beta}^{(k)}) \right)$$

- Step 4: Stop if $|\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}|$ is small. If not, let $k = k + 1$ and repeat Steps 2 through 4.

Newton-Raphson Method to find “beta”

- In general, when “beta” is a vector:

$$\begin{aligned}\hat{\beta}^{(k+1)} &= U'(\hat{\beta}^{(k)})^{-1} \left(U'(\hat{\beta}^{(k)})\hat{\beta}^{(k)} - U(\hat{\beta}^{(k)}) \right) \\ &= -(X'V^{(k)}X)^{-1} \left[-(X'V^{(k)}X)\hat{\beta}^{(k)} - X'(Y - \mu(\hat{\beta}^{(k)})) \right] \\ &= (X'V^{(k)}X)^{-1} \left[X'V^{(k)} \left(X\hat{\beta}^{(k)} + V^{-1(k)}(Y - \mu(\hat{\beta}^{(k)})) \right) \right] \\ &= (X'V^{(k)}X)^{-1} (X'V^{(k)}Z^{(k)})\end{aligned}$$

where

$$V^{(k)} = \text{diag}(\mu_i(\beta^{(k)})[1 - \mu_i(\beta^{(k)})])$$

$$Z^{(k)} = X\hat{\beta}^{(k)} + V^{-1(k)} \left(Y - \mu(\hat{\beta}^{(k)}) \right) = \text{a surrogate response.}$$

Iteratively Re-weighted Least Squares

The general procedure is:

- Step 0: Set an initial value for $\hat{\beta}^{(k)}$, $k = 0$.
- Step 1: Calculate: $V^{(k)}$, $\hat{\mu}(\hat{\beta}^{(k)})$, $Z^{(k)}$.
- Step 2: Update $\hat{\beta}^{(k+1)} = (X^T V^{(k)} X)^{-1} (X^T V^{(k)} Z^{(k)})$
- Step 3: Stop if $\sum_{j=1}^{p+1} \left(\hat{\beta}_j^{(k+1)} - \hat{\beta}_j^{(k)} \right)^2 < \epsilon$; if not, let $k = k + 1$ and repeat Steps 2 and 3.



IRLS vs weighted least squares

Compare the IRLS to the weighted least squares solution we derived last term:

$$\hat{\beta}_{WLS} = (X' \hat{V}^{-1} X)^{-1} (X' \hat{V}^{-1} Y)$$

These are different! \hat{V} vs. \hat{V}^{-1} .

Recall that we derived: $\frac{\partial \mu(\beta)}{\partial \beta} = V X = \text{diag} [\mu(\beta)(1 - \mu(\beta))] X$

So that,

$$\begin{aligned} \hat{\beta}^{(k+1)} &= (X' V^{(k)} X)^{-1} (X' V^{(k)} Z^{(k)}) \\ &= \left(\frac{\partial \hat{\mu}(\beta^{(k)})}{\partial \beta} \hat{V}^{(k)-1} \frac{\partial \hat{\mu}(\beta^{(k)})}{\partial \beta} \right)^{-1} \left(\frac{\partial \hat{\mu}(\beta^{(k)})}{\partial \beta} \hat{V}^{(k)-1} Z^{*(k)} \right) \end{aligned}$$

where $Z^{*(k)} = \frac{\partial \hat{\mu}(\beta^{(k)})}{\partial \beta} \hat{\beta}^{(k)} + (Y - \mu(\hat{\beta}^{(k)}))$.

Inference in logistic regression models

- ▶ Using similar arguments as we did for linear models: $\hat{\beta}_{mle} \approx N(\beta, [X'VX]^{-1})$

- ▶ Inference for a single coefficient:

$$\text{Test } H_0 : \beta_j = b \text{ via } Z = \frac{\hat{\beta}_j - b}{\sqrt{[X'VX]_{jj}^{-1}}}$$

Confidence intervals can be derived as: $\hat{\beta}_j \pm 1.96 \sqrt{[X'VX]_{jj}^{-1}}$

- ▶ Inference for a linear combination of coefficients:

Define $d = w'\beta$ where w is a $(p+1) \times 1$ vector of scalars to create the relevant linear combination of β .

Estimate d via $w'\hat{\beta}$ and $se(\hat{d}) = \sqrt{w'[X'VX]^{-1}w}$

Confidence interval for d : $\hat{d} \pm 1.96 se_{\hat{d}}$.

Test $H_0 : d = \delta$ via $Z = \frac{\hat{d} - \delta}{se_{\hat{d}}}$.

Inference in logistic regression models: Nested models

Here we assume we have a model with $\beta = (\beta_0, \beta_1, \dots, \beta_p, \beta_{p+1}, \dots, \beta_{p+s})$ and define $\beta^+ = (\beta_{p+1}, \dots, \beta_{p+s})$.

To conduct a Wald test of H_0 : all $\beta_{p+j} = 0, \text{ for } j = 1, \dots, s$,

$$W = \hat{\beta}^{+T} \left[(X^+ V X^+)^{-1}_{(+,+)} \right]^{-1} \hat{\beta}^+ \approx \sum_{j=1}^s Z_j^2 \sim \chi_s^2$$

reject H_0 if $W > \chi_{s, 1-0.05/2}^2$.

When the null hypothesis is true and sample size is large enough:

$$\Delta = -2 \left[\log \text{Like}_N(y, \hat{\beta}_N) - \log \text{Like}_E(y, \hat{\beta}_E) \right] \sim \chi_s^2$$

Δ represents the “change in deviance” where

$$\text{deviance} = -2 \left[\log \text{Like}_N(y, \hat{\beta}_N) - \log \text{Like}_E(y, y) \right] \sim \chi_s^2$$

where $\log \text{Like}_E(y, y)$ is the biggest possible value.

The deviance is a measure of fidelity of the model to the data, like the residual sum of squares for linear regression.

Examples

```
data1$agec = data1$lastage - 60
data1$agesp1 = ifelse(data1$lastage>65,data1$lastage-65,0)
data1$agesp2 = ifelse(data1$lastage>80,data1$lastage-80,0)

fit0 = glm(bigexp~mscd+agec+agesp1+agesp2,data=data1,family="binomial")
fit1 = glm(bigexp~mscd*(agec+agesp1+agesp2),data=data1,family="binomial")
```

- Write out the model you are fitting in “fit0” and “fit1”.

Example: Testing a single coefficient

- ▶ Test the null hypothesis that after adjusting for age, there is no relationship between a big expenditure and a MSCD.

```
summary(fit0)$coefficients
```

##	Estimate	Std. Error	z value	Pr(> z)
## (Intercept)	-0.716235408	0.030036992	-23.8451109	1.138097e-125
## mscd	1.603178804	0.068286173	23.4773561	6.949175e-122
## agec	0.028079056	0.002891139	9.7121075	2.677428e-22
## agesp1	-0.005830743	0.007465457	-0.7810296	4.347851e-01
## agesp2	-0.002128496	0.019276490	-0.1104193	9.120769e-01

Example: Linear combination of coefficients

- ▶ Using Model1, estimate the log odds ratio of a big expenditure comparing persons with and without a MSCD whom are 70 years old.
- ▶ What is the appropriate linear combination of β ?

Confirm using lincom command

```
lincom(fit1,c("mscd+10*mscd:agec+5*mscd:agesp1"))
```

```
##
```

	Estimate	2.5 %	97.5 %	Chisq	Pr(>Chisq)
## mscd+10*mscd:agec+5*mscd:agesp1	1.513507	1.351594	1.67542	335.6613	5.620428e-75

Example: Linear combination of coefficients

```
## In Model 1: Compute the OR for big expenditure vs. mscd for 70 year olds
```

```
w = c(0,1,0,0,0,10,5,0)
```

```
var.cov = summary(fit1)$cov.scaled
```

```
beta = fit1$coefficients
```

```
# estimate
```

```
t(w) %*% beta
```

```
##           [,1]
```

```
## [1,] 1.513507
```

```
# standard error
```

```
t(w) %*% var.cov %*% w
```

```
##           [,1]
```

```
## [1,] 0.006824451
```

```
# test statistic
```

```
t(w) %*% beta / sqrt(t(w) %*% var.cov %*% w)
```

```
##           [,1]
```

```
## [1,] 18.32106
```

```
# Square test statistic ~ chi-square 1
```

```
(t(w) %*% beta / sqrt(t(w) %*% var.cov %*% w))^2
```

```
##           [,1]
```

```
## [1,] 335.6613
```

Example: Nested models

- ▶ Model0 is nested within Model1.
- ▶ What null and alternative hypothesis are you testing if you compare Model1 and Model 0?
- ▶ Wald test

```
## Nested model: Wald test for interaction
index = 6:8
# Compute the wald test
w = t(fit1$coeff[index]) %*% solve(var.cov[index,index]) %*% fit1$coeff[index]
w
```

```
##           [,1]
## [1,] 14.53997
```

```
pchisq(w,lower.tail=FALSE,df=3)
```

```
##           [,1]
## [1,] 0.002255128
```



Example: Nested models

► Likelihood ratio test

```
## Nested model: likelihood ratio test  
lrtest(fit1,fit0)
```

```
## Likelihood ratio test  
##  
## Model 1: bigexp ~ mscd * (agec + agesp1 + agesp2)  
## Model 2: bigexp ~ mscd + agec + agesp1 + agesp2  
##   #Df  LogLik Df  Chisq Pr(>Chisq)  
## 1    8 -7126.9  
## 2    5 -7134.5 -3 15.185   0.001665 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Where to next?

▶ Prediction!

