

# Lecture4 Handout

Elizabeth Colantuoni

4/5/2021

## I. Objectives

Upon completion of this session, you will be able to do the following:

- Conduct an analysis to identify which of a possible set of adjustment variables is a “confounder”
- Write a paragraph summarizing your analysis which includes a description of the data and your main findings
- Understand and explain the asymptotic distribution of the mle’s in the GLM family
- Understand and explain the likelihood ratio test for the GLM family
- Understand and explain how to use Wald tests for the GLM family

## II. Assessing confounding in generalized linear models

In Lecture 3, we reviewed how to assess confounding within logistic regression models and demonstrated that older age (age > 65 years) is a confounding variable for *big expenditure* vs. MSCD relationship.

For this part of the lecture, you will be conducting an analysis of the NMES to assess whether education, poverty status, seatbelt use, marital status or geographic region are confounders for the *big expenditure* vs. MSCD relationship.

### A. Analysis

Modify the code below to separately evaluate whether each of the covariates listed above is a confounding variable. NOTE: For this class exercise, we will use ONLY the results from fitting the marginal and conditional logistic regression model. In practice, you may also want to compute the difference between the marginal and conditional coefficient and test statistic using the bootstrap procedure.

```
load('./nmes.rdata')
data = nmes
data[data=="."] = NA

## Create the necessary variables:
data$posexp=ifelse(data$totalexp>0,1,0)
data$mscd=ifelse(data$l5c5+data$chd5>0,1,0)
data1=data[!is.na(data$eversmk),]
data1$older=ifelse(data1$lastage<65,0,1)
data1$bigexp=ifelse(data1$totalexp>1000,1,0)
```

```

# Fit the models, save the log(OR) and 95% CI
model.M = glm(bigexp~mscd,data=data1,family="binomial")
outM = lincom(model.M,c("mscd"))
model.C = glm(bigexp~mscd+older,data=data1,family="binomial")
outC = lincom(model.C,c("mscd"))

# Marginal Model output
outM

##      Estimate    2.5 %   97.5 %   Chisq   Pr(>Chisq)
## mscd 1.825045  1.694177  1.955913  747.095  1.718138e-164

# Conditional Model output
outC

##      Estimate    2.5 %   97.5 %   Chisq   Pr(>Chisq)
## mscd 1.654913  1.521564  1.788262  591.65  1.096494e-130

```

## B. Write-up

Now, summarize your findings as if you were writing a paragraph for a publishable paper.

Use the following structure:

- 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.

## III. Maximum likelihood estimation in logistic regression models

In Lecture 3, we reviewed the Newton-Raphson algorithm for finding the root of a function.

### A. Scalar case

In the scalar case, the goal is to find the value of  $\beta$  that sets the score equation  $U(\beta) = 0$ .

- Step 0: Pick an initial starting value for  $\beta$ , 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  $\hat{\beta}^{(k+1)}$ :

$$\begin{aligned}
U'(\hat{\beta}^{(k)}) &= \frac{U(\hat{\beta}^{(k)}) - 0}{\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)}} \\
[\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)}]U'(\hat{\beta}^{(k)}) &= U(\hat{\beta}^{(k)}) \\
\hat{\beta}^{(k)} - \hat{\beta}^{(k+1)} &= U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)}) \\
\hat{\beta}^{(k+1)} &= \hat{\beta}^{(k)} - U'(\hat{\beta}^{(k)})^{-1}U(\hat{\beta}^{(k)}) \\
&= U'(\hat{\beta}^{(k)})^{-1} \left( U'(\hat{\beta}^{(k)})\hat{\beta}^{(k)} - U(\hat{\beta}^{(k)}) \right)
\end{aligned}$$

- 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.

## B. General case

In general, i.e. when  $\beta$  is a vector, we have:

$$U(\beta) = X'(Y - \mu(\beta))$$

$$U'(\beta) = -X'VX$$

And the Newton-Raphon method is expressed as:

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

where

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

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

## C. Iteratively Re-weighted Least Squares (IRLS)

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  $\hat{\beta}^{(k+1)} = (X'V^{(k)}X)^{-1}(X'V^{(k)}Z^{(k)})$
- 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.

## D. Comparison to weighted least squares

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

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

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,

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

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

## IV. Inference using $\hat{\beta}_{mle}$

Using similar arguments as we did for determining the distribution of  $\hat{\beta}_{mle}$  in linear models, we can show that:

$$\hat{\beta}_{mle} \approx N(\beta, [X^T V X]^{-1})$$

So from this we can derive the following set of tests or estimation relating to  $\beta$ .

### A. Inference for $\beta_j$

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

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

### B. Estimating linear combinations of $\beta$

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

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

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

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

## C. 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  $\beta_{p+j} = 0, for j = 1, \dots, s$ ,

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

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

Alternatively, use a likelihood ratio test!

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

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

$\Delta$  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.

## D. Example: NMES big expenditure - MSCD relationship

```
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 two models we are fitting:

### 1. Testing a single coefficient

In Model0, test the null hypothesis that after adjusting for age, there is no relationship between a big expenditure and a MSCD.

State the null and alternative hypotheses:

```
## In Model 0: Test \beta_{mscd} = 0  
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

## 2. Computing a linear combination of $\beta$

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  $\beta$ ?

```
## 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

## 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
```

### 3. Comparing nested models

Compare Model0 and Model1 using both a Wald test and a likelihood ratio test.

What null and alternative hypotheses are you testing?

```
## 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

## 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
```