JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Lecture 2

Review Generalized Linear Models
More on Logistic Regression:
Regression adjustment and continuous covariates

Review of Lecture 1: GLMs

- Generalized Linear models
  - Defines a class of regression models for outcomes from the exponential family of distributions
  - Exponential family includes: Normal, Bernoulli/Binomial, Poisson, Gamma, Beta, among others

- Requires specification of three components:

Review of Lecture 1: GLMs

- Linear Model

- Logistic Model

Review of Lecture 1: Key quantities for simple logistic regression

- Assume your outcome is Y taking values 0 vs. 1 and your primary exposure variable X is also binary taking values 0 vs. 1.

- Mean:

- Odds:

- Logit:

- Odds ratio

Review of Lecture 1 + additional models

- In this lecture we will consider 4 logistic regression models:
- Model A:
- Model B:
- Model C:
- Model D:
- We will fit models B through D and a few additional models as well.

Revisit Model B

```
## Create the necessary variables:
data$posexp=ifelse(data$totalexp>0,1,0)
data$mscd=ifelse(data$lc5+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)

## Model B
modelB = glm(bigexp~mscd,data=data1,family="binomial")
lincom(modelB,c("(Intercept)","mscd"))
```

```
##               Estimate   2.5 %      97.5 %     Chisq    Pr(>Chisq)
## (Intercept) -0.7395315 -0.7806967 -0.6983663 1239.792 1.372226e-271
## mscd         1.825045   1.694177   1.955913   747.095  1.718138e-164
```

```
lincom(modelB,c("(Intercept)","mscd"),eform=TRUE)
```

```
##              Estimate  2.5 %     97.5 %    Chisq    Pr(>Chisq)
## (Intercept) 0.4773375 0.4580868 0.4973973 1239.792 1.372226e-271
## mscd        6.203076  5.442166  7.070374  747.095  1.718138e-164
```

Revisit Model B

```
lincom(modelB,c("(Intercept)","mscd"))

##              Estimate   2.5 %      97.5 %     Chisq    Pr(>Chisq)
## (Intercept) -0.7395315 -0.7806967 -0.6983663 1239.792 1.372226e-271
## mscd         1.825045   1.694177   1.955913   747.095  1.718138e-164

lincom(modelB,c("(Intercept)","mscd"),eform=TRUE)

##              Estimate  2.5 %     97.5 %    Chisq    Pr(>Chisq)
## (Intercept) 0.4773375 0.4580868 0.4973973 1239.792 1.372226e-271
## mscd        6.203076  5.442166  7.070374  747.095  1.718138e-164
```

- Interpret beta_0 and exp(beta_0)

- Interpret beta_1 and exp(beta_1)

Interpretation of $\exp(\hat{\beta}_{mscd})$

- Two interpretations:

Revisit Model C

```
## Model C
modelC = glm(bigexp~mscd+older+mscd:older,data=data1,family="binomial")
lincom(modelC,c("mscd","mscd+mscd:older","mscd:older"))
```

```
##                 Estimate   2.5 %      97.5 %    Chisq    Pr(>Chisq)
## mscd            1.969895   1.735287   2.204503  270.8301 7.481555e-61
## mscd+mscd:older 1.491115   1.329415   1.652815  326.6619 5.126712e-73
## mscd:older      -0.4787796 -0.7637143 -0.193845 10.84618 0.0009899951
```

```
lincom(modelC,c("mscd","mscd+mscd:older","mscd:older"),eform=TRUE)
```

```
##                 Estimate 2.5 %     97.5 %    Chisq    Pr(>Chisq)
## mscd            7.169921 5.670554  9.065741  270.8301 7.481555e-61
## mscd+mscd:older 4.442046 3.778832  5.221658  326.6619 5.126712e-73
## mscd:older      0.619539 0.4659326 0.8237856 10.84618 0.0009899951
```

Revisit Model C

```r
## Model C
modelC = glm(bigexp~mscd+older+mscd:older,data=data1,family="binomial")
lincom(modelC,c("mscd","mscd+mscd:older","mscd:older"))
```

```
##                  Estimate   2.5 %      97.5 %    Chisq    Pr(>Chisq)
## mscd             1.969895   1.735287   2.204503  270.8301 7.481555e-61
## mscd+mscd:older  1.491115   1.329415   1.652815  326.6619 5.126712e-73
## mscd:older       -0.4787796 -0.7637143 -0.193845 10.84618 0.0009899951
```

```r
lincom(modelC,c("mscd","mscd+mscd:older","mscd:older"),eform=TRUE)
```

```
##                  Estimate 2.5 %     97.5 %    Chisq    Pr(>Chisq)
## mscd             7.169921 5.670554  9.065741  270.8301 7.481555e-61
## mscd+mscd:older  4.442046 3.778832  5.221658  326.6619 5.126712e-73
## mscd:older       0.619539 0.4659326 0.8237856 10.84618 0.0009899951
```

Model D: Parameter interpretation and estimation

- Model D:
- How do you interpret coefficient for MSCD?
- How do we estimate this coefficient?
  - Inverse-variance weighting estimation!
  - Same as linear regression!
  - Need to consider the age (young vs. old) specific 2x2 tables.

	Young		Old	
	MSCD = 1	MSCD = 0	MSCD = 1	MSCD = 0
Bigexp = 1	273	1802	713	1547
Bigexp = 0	101	4780	232	2236

Model D: Parameter interpretation and estimation

Age group	$log\hat{OR}$	$se(log\hat{OR})$	$var(log\hat{OR})$	$\frac{1}{var(log\hat{OR})}$	$w = \frac{\frac{1}{var(log\hat{OR})}}{\Sigma(\frac{1}{var(log\hat{OR})})}$
Younger	1.97	0.12	0.0144	69.4	0.32
Older	1.49	0.083	0.0069	144.9	0.68

$$\hat{\beta}_1 = 1.97 \times 0.32 + 1.49 \times 0.68 = 1.64$$

$$se(\hat{\beta}_1) = \frac{1}{\sqrt{\sum(\frac{1}{var(log\hat{OR})})}} = \frac{1}{\sqrt{214.3}} = 0.068$$

Model D: Parameter interpretation and estimation

```r
modelD = glm(bigexp~mscd+older,data=data1,family="binomial")
summary(modelD)$coeff
```

```
##                Estimate Std. Error   z value      Pr(>|z|)
## (Intercept) -0.9577826 0.02700779 -35.46321 1.815505e-275
## mscd         1.6549130 0.06803662  24.32386 1.096494e-130
## older        0.5638298 0.04104938  13.73540  6.230701e-43
```

```r
lincom(modelD,c("mscd","older"),eform=TRUE)
```

```
##        Estimate 2.5 %    97.5 %   Chisq    Pr(>Chisq)
## mscd  5.232625 4.57938  5.979054 591.65   1.096494e-130
## older 1.75739  1.621537 1.904625 188.6613 6.230701e-43
```

You practice: Use the output above, interpret $exp(\hat{\beta}_2)$.

Model D: Adjustment for continuous covariates

- Now, imagine Model D but where we allow age to be a continuous variable
- Model D with continuous age:
- Can you draw a picture of this model?

Model D: Adjustment for continuous covariates

```
modelDagecont = glm(bigexp~mscd+lastage,data=data1,family="binomial")
summary(modelDagecont)$coeff
```

```
##               Estimate  Std. Error   z value      Pr(>|z|)
## (Intercept) -2.27990966 0.099135981 -22.99780 4.903428e-117
## mscd         1.60502065 0.068269770  23.50998 3.224831e-122
## lastage      0.02574057 0.001599682  16.09105  2.947835e-58
```

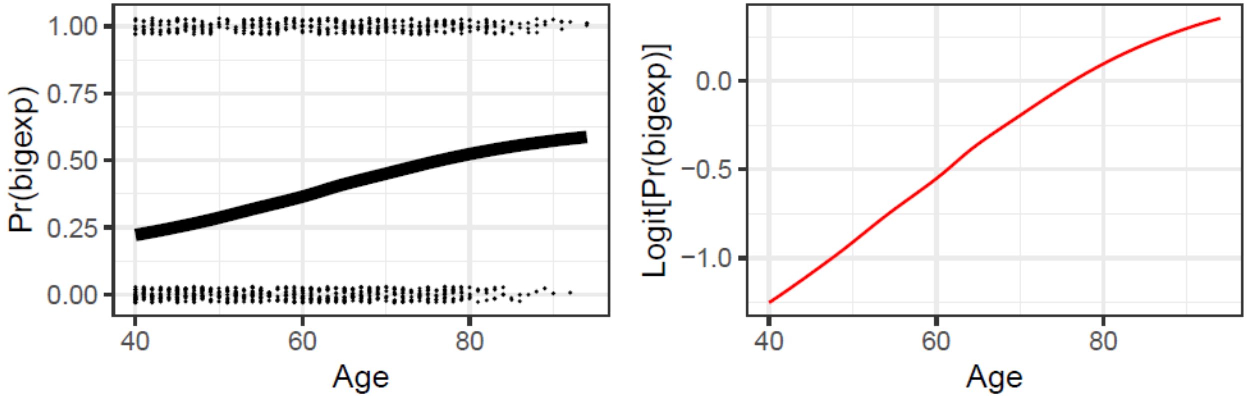
```
lincom(modelDagecont,c("mscd","lastage"),eform=TRUE)
```

```
##         Estimate 2.5 %    97.5 %   Chisq   Pr(>Chisq)
## mscd    4.977962 4.35452  5.690664 552.719 3.224831e-122
## lastage 1.026075 1.022863 1.029297 258.922 2.947835e-58
```

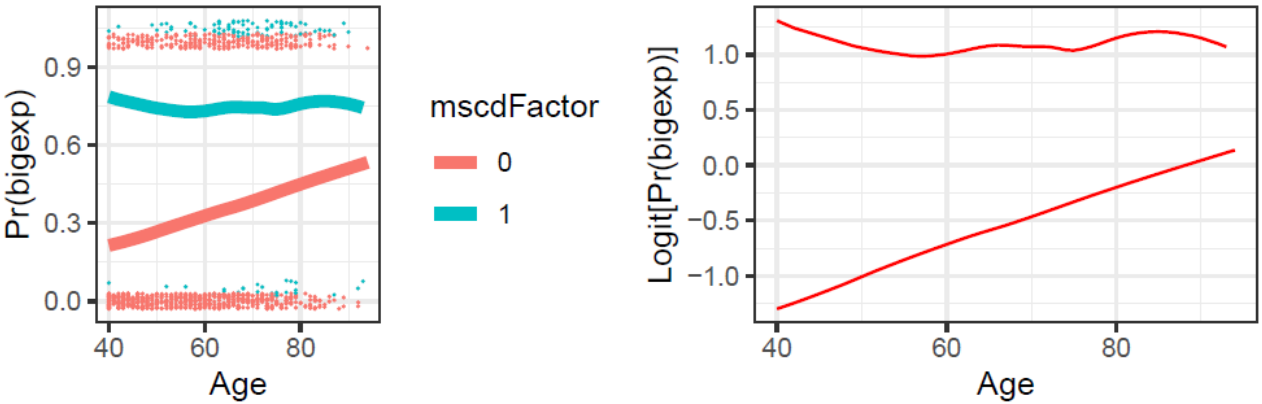
- Interpret both of the coefficients:

Assessing functional form for continuous covariates

- How do we know if the relationship between the logit of a big expenditure and age is linear?



Revisit interaction Model C with continuous age



- What do you think about the MSCD-specific relationship between a big expenditure and age?
  - Linear? Non-linear?

Revision interaction Model C with continuous age

- Assuming the linear assumption is okay!

```r
data1$age_c = data1$lastage - 60

modelCcont = glm(bigexp~mscd+age_c+mscd:age_c,data=data1,family="binomial")
lincom(modelCcont,c("mscd","mscd+mscd:age_c","mscd+20*mscd:age_c","mscd:age_c"))
```

```
##                    Estimate   2.5 %       97.5 %      Chisq    Pr(>Chisq)
## mscd               1.792367   1.625218    1.959516    441.7169 4.579093e-98
## mscd+mscd:age_c    1.768144   1.607967    1.928321    468.0905 8.342527e-104
## mscd+20*mscd:age_c 1.307903   1.113342    1.502464    173.595  1.213445e-39
## mscd:age_c         -0.0242232 -0.03631431 -0.01213209 15.41797 8.616514e-05
```

```r
lincom(modelCcont,c("mscd","mscd+mscd:age_c","mscd+20*mscd:age_c","mscd:age_c"),eform=TRUE)
```

```
##                    Estimate  2.5 %     97.5 %    Chisq    Pr(>Chisq)
## mscd               6.003646  5.079528  7.09589   441.7169 4.579093e-98
## mscd+mscd:age_c    5.859966  4.992649  6.877953  468.0905 8.342527e-104
## mscd+20*mscd:age_c 3.69841   3.044517  4.492744  173.595  1.213445e-39
## mscd:age_c         0.9760678 0.9643371 0.9879412 15.41797 8.616514e-05
```

Revision interaction Model C with continuous age

- Assuming the linear assumption is okay!

```
data1$age_c = data1$lastage - 60

modelCcont = glm(bigexp~mscd+age_c+mscd:age_c,data=data1,family="binomial")
lincom(modelCcont,c("mscd","mscd+mscd:age_c","mscd+20*mscd:age_c","mscd:age_c"))
```

```
##                    Estimate   2.5 %       97.5 %      Chisq    Pr(>Chisq)
## mscd               1.792367   1.625218    1.959516    441.7169 4.579093e-98
## mscd+mscd:age_c    1.768144   1.607967    1.928321    468.0905 8.342527e-104
## mscd+20*mscd:age_c 1.307903   1.113342    1.502464    173.595  1.213445e-39
## mscd:age_c         -0.0242232 -0.03631431 -0.01213209 15.41797 8.616514e-05
```

```
lincom(modelCcont,c("mscd","mscd+mscd:age_c","mscd+20*mscd:age_c","mscd:age_c"),eform=TRUE)
```

```
##                    Estimate  2.5 %     97.5 %    Chisq    Pr(>Chisq)
## mscd               6.003646  5.079528  7.09589   441.7169 4.579093e-98
## mscd+mscd:age_c    5.859966  4.992649  6.877953  468.0905 8.342527e-104
## mscd+20*mscd:age_c 3.69841   3.044517  4.492744  173.595  1.213445e-39
## mscd:age_c         0.9760678 0.9643371 0.9879412 15.41797 8.616514e-05
```

Revision interaction Model C with continuous age

- How would you rewrite the lincom commands to get estimates of the relationship between having a big expenditure and age, separately for those with and without a MSCD?

```
modelCcont = glm(bigexp~mscd+age_c+mscd:age_c,data=data1,family="binomial")
lincom(modelCcont,c("mscd","mscd+mscd:age_c","mscd+20*mscd:age_c","mscd:age_c"))
```

Where to next?

- Assessing for confounding in logistic regression models
  - See “Note on confounding and effect modification 2019” by Scott Zeger
  - Additional references are provided in Lecture 2 Handout

- Statistical inference in logistic regression models
  - Maximum likelihood estimation
  - Iteratively reweighted least squares