Precept 4 - More GLMs:
Models of Binary and Lognormal Outcomes

Soc 504: Advanced Social Statistics

Ian Lundberg

Princeton University

March 2, 2017

¹These slides owe an enormous debt to generations of TFs in Gov 2001 at Harvard. Many slides are directly adapted from those by Brandon Stewart and Stephen Pettigrew.
GLMs

Complementary log-log

Quantities of Interest

Outline

1. GLMs
   - General Structure of GLMs
   - Procedure for Running a GLM

2. Complementary log-log

3. Quantities of Interest
Replication Paper

Any thoughts or issues to discuss?
Generalized Linear Models
All of the models we’ve talked about belong to a class called generalized linear models (GLM).
Generalized Linear Models

All of the models we’ve talked about belong to a class called \textit{generalized linear models (GLM)}.

Three elements of a GLM:
Generalized Linear Models

All of the models we’ve talked about belong to a class called **generalized linear models (GLM)**.

Three elements of a GLM:
- A distribution for $Y$ (stochastic component)
Generalized Linear Models

All of the models we’ve talked about belong to a class called \textit{generalized linear models (GLM)}.

Three elements of a GLM:

- A distribution for $Y$ (stochastic component)
- A linear predictor $X\beta$ (systematic component)
Generalized Linear Models

All of the models we’ve talked about belong to a class called **generalized linear models (GLM)**.

Three elements of a GLM:

- A distribution for $Y$ (stochastic component)
- A linear predictor $X\beta$ (systematic component)
- A link function that relates the linear predictor to a parameter of the distribution. (systematic component)
1. Specify a distribution for $Y$
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
- Binary: Bernoulli
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
- Ordered Categories:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:

- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
- Ordered Categories: Normal with observation mechanism
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
- Ordered Categories: Normal with observation mechanism
- Unordered Categories:
1. Specify a distribution for $Y$

Assume our data was generated from some distribution.

Examples:
- Continuous and Unbounded: Normal
- Binary: Bernoulli
- Event Count: Poisson
- Duration: Exponential
- Ordered Categories: Normal with observation mechanism
- Unordered Categories: Multinomial
2. Specify a linear predictor

We are interested in allowing some parameter of the distribution $\theta$ to vary as a (linear) function of covariates. So we specify a linear predictor.
2. Specify a linear predictor

We are interested in allowing some parameter of the distribution $\theta$ to vary as a (linear) function of covariates. So we specify a linear predictor.

$$X\beta = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$$
3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean). Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of the distribution for $Y$.

$g(\theta) = X\beta$ $\theta = g^{-1}(X\beta)$

This is the systematic component that we've been talking about all along.
3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean).
3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean).

Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of distribution for $Y$. 
3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean).

Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of distribution for $Y$.

$$g(\theta) = X\beta$$
### 3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean).

Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of distribution for $Y$.

$$g(\theta) = X\beta$$

$$\theta = g^{-1}(X\beta)$$
3. Specify a link function

The link function relates the linear predictor to some parameter $\theta$ of the distribution for $Y$ (almost always the mean).

Let $g(\cdot)$ be the link function and let $E(Y) = \theta$ be the mean of distribution for $Y$.

$$g(\theta) = X\beta$$

$$\theta = g^{-1}(X\beta)$$

This is the systematic component that we’ve been talking about all along.
Example Link Functions

Identity:
\[ \text{Link: } \mu = X\beta \]

Inverse:
\[ \text{Link: } \lambda^{-1} = X\beta \]

Logit:
\[ \text{Link: } \ln(\pi_1 - \pi) = X\beta \]

Inverse Link:
\[ \pi = \frac{1}{1 + e^{-X\beta}} \]

Probit:
\[ \text{Link: } \Phi^{-1}(\pi) = X\beta \]

Inverse Link:
\[ \pi = \Phi(X\beta) \]

Log:
\[ \text{Link: } \ln(\lambda) = X\beta \]

Inverse Link:
\[ \lambda = \exp(X\beta) \]
Example Link Functions

Identity:
Example Link Functions

Identity:

- Link: $\mu = X\beta$
Example Link Functions

Identity:

- Link: $\mu = X\beta$

Inverse:

Logistic:

- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$

Inverse Link: $\pi = \frac{1}{1 + e^{-X\beta}}$

Probit:

- Link: $\Phi^{-1}(\pi) = X\beta$

Inverse Link: $\pi = \Phi(X\beta)$

Log:

- Link: $\ln(\lambda) = X\beta$

Inverse Link: $\lambda = e^{X\beta}$
Example Link Functions

Identity:
- Link: $\mu = X\beta$

Inverse:
- Link: $\lambda^{-1} = X\beta$
Example Link Functions

Identity:
- Link: \( \mu = X\beta \)

Inverse:
- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)
Example Link Functions

Identity:
- Link: \( \mu = X\beta \)

Inverse:
- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)

Logit:
Example Link Functions

Identity:

- Link: \( \mu = X\beta \)

Inverse:

- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)

Logit:

- Link: \( \ln\left(\frac{\pi}{1-\pi}\right) = X\beta \)
Example Link Functions

Identity:
- Link: $\mu = X\beta$

Inverse:
- Link: $\lambda^{-1} = X\beta$
- Inverse Link: $\lambda = (X\beta)^{-1}$

Logit:
- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$
- Inverse Link: $\pi = \frac{1}{1+e^{-X\beta}}$
Example Link Functions

Identity:
- Link: $\mu = X\beta$

Inverse:
- Link: $\lambda^{-1} = X\beta$
- Inverse Link: $\lambda = (X\beta)^{-1}$

Logit:
- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$
- Inverse Link: $\pi = \frac{1}{1+e^{-X\beta}}$

Probit:
Example Link Functions

Identity:
- Link: $\mu = X\beta$

Inverse:
- Link: $\lambda^{-1} = X\beta$
- Inverse Link: $\lambda = (X\beta)^{-1}$

Logit:
- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$
- Inverse Link: $\pi = \frac{1}{1+e^{-X\beta}}$

Probit:
- Link: $\Phi^{-1}(\pi) = X\beta$
Example Link Functions

Identity:
- Link: \( \mu = X\beta \)

Inverse:
- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)

Logit:
- Link: \( \ln\left(\frac{\pi}{1-\pi}\right) = X\beta \)
- Inverse Link: \( \pi = \frac{1}{1+e^{-X\beta}} \)

Probit:
- Link: \( \Phi^{-1}(\pi) = X\beta \)
- Inverse Link: \( \pi = \Phi(X\beta) \)
Example Link Functions

Identity:

- Link: \( \mu = X\beta \)

Inverse:

- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)

Logit:

- Link: \( \ln \left( \frac{\pi}{1-\pi} \right) = X\beta \)
- Inverse Link: \( \pi = \frac{1}{1+e^{-X\beta}} \)

Probit:

- Link: \( \Phi^{-1}(\pi) = X\beta \)
- Inverse Link: \( \pi = \Phi(X\beta) \)

Log:
Example Link Functions

Identity:
- Link: $\mu = X\beta$

Inverse:
- Link: $\lambda^{-1} = X\beta$
- Inverse Link: $\lambda = (X\beta)^{-1}$

Logit:
- Link: $\ln\left(\frac{\pi}{1-\pi}\right) = X\beta$
- Inverse Link: $\pi = \frac{1}{1+e^{-X\beta}}$

Probit:
- Link: $\Phi^{-1}(\pi) = X\beta$
- Inverse Link: $\pi = \Phi(X\beta)$

Log:
- Link: $\ln(\lambda) = X\beta$
Example Link Functions

Identity:

- Link: \( \mu = X\beta \)

Inverse:

- Link: \( \lambda^{-1} = X\beta \)
- Inverse Link: \( \lambda = (X\beta)^{-1} \)

Logit:

- Link: \( \ln\left(\frac{\pi}{1-\pi}\right) = X\beta \)
- Inverse Link: \( \pi = \frac{1}{1+e^{-X\beta}} \)

Probit:

- Link: \( \Phi^{-1}(\pi) = X\beta \)
- Inverse Link: \( \pi = \Phi(X\beta) \)

Log:

- Link: \( \ln(\lambda) = X\beta \)
- Inverse Link: \( \lambda = \exp(X\beta) \)
4. Estimate Parameters via ML
4. Estimate Parameters via ML

Use `optim` to estimate the parameters just like we have all along.
5. Quantities of Interest

Simulate parameters from multivariate normal.

Run $X\beta$ through inverse link function to get expected values.

Draw from distribution of $Y$ for predicted values.
5. Quantities of Interest

1. Simulate parameters from multivariate normal.
5. Quantities of Interest

1. Simulate parameters from multivariate normal.
2. Run $X\beta$ through inverse link function to get expected values.
5. Quantities of Interest

1. Simulate parameters from multivariate normal.
2. Run $X\beta$ through inverse link function to get expected values.
3. Draw from distribution of $Y$ for predicted values.
Outline

1. GLMs
   - General Structure of GLMs
   - Procedure for Running a GLM

2. Complementary log-log

3. Quantities of Interest
GLMs

Complementary log-log Quantities of Interest

EVICTED

POVERTY AND PROFIT IN THE AMERICAN CITY

MATTHEW DESMOND
We will use data from the Fragile Families and Child Wellbeing Study to study the cumulative risk of eviction over child for children born in large American cities.
Research question and data

What is the probability of eviction in a given year for a child with a given set of covariates?
Research question and data

What is the probability of eviction in a given year for a child with a given set of covariates?

ffEviction.csv is the data that we use.
Fragile Families data
Fragile Families data

What’s in the data?
### Fragile Families data

#### What’s in the data?

```r
> head(d)
   idnum income married cm1ethrace ev
1   0001   1.5       0     Hispanic    0
2   0002   1.6       0       Black    0
3   0003   2.7       0      White      0
4   0004   1.0       0     Hispanic    0
5   0006   0.2       0       Black    0
6   0007   1.3       0     Hispanic    0
```

```r
> summary(d)
   idnum income married cm1ethrace ev
 Length:12298 Min. :0.000 Min. :0.0000 White :2709 Min. :0.0000
 Class :character 1st Qu.:0.500 1st Qu.:0.0000 Black :5911 1st Qu.:0.0000
 Mode :character Median :1.200 Median :0.0000 Hispanic:3225 Median :0.0000
       Mean :1.666 Mean :0.2502 Other : 453 Mean :0.02301
       3rd Qu.:2.400 3rd Qu.:1.0000 3rd Qu.:0.0000
       Max. :5.000 Max. :1.0000 Max. :1.0000
```
Fragile Families data

ev:

income: family income / poverty line at age 1
married: were the parents married at the birth?
cm1ethrace: mother's race/ethnicity
Fragile Families data

\( ev \): dependent variable; was this child evicted in a given year?
Fragile Families data

ev: dependent variable; was this child evicted in a given year?
income:
Fragile Families data

$ev$: dependent variable; was this child evicted in a given year?

$\text{income}$: family income / poverty line at age 1
Fragile Families data

ev: dependent variable; was this child evicted in a given year?
income: family income / poverty line at age 1
married:
Fragile Families data

ev: dependent variable; was this child evicted in a given year?
income: family income / poverty line at age 1
married: were the parents married at the birth?
Fragile Families data

ev: dependent variable; was this child evicted in a given year?
income: family income / poverty line at age 1
married: were the parents married at the birth?
cm1lethrace:
Fragile Families data

ev: dependent variable; was this child evicted in a given year?
income: family income / poverty line at age 1
married: were the parents married at the birth?
cm1ethrace: mother's race/ethnicity
Outline

1. GLMs
   - General Structure of GLMs
   - Procedure for Running a GLM

2. Complementary log-log

3. Quantities of Interest
Our outcome variable is whether or not a child was evicted.

What's the first question we should ask ourselves when we start to model this dependent variable?
Our outcome variable is whether or not a child was evicted
Our outcome variable is whether or not a child was evicted.

What’s the first question we should ask ourselves when we start to model this dependent variable?
1. Specify a distribution for $Y_i \sim \text{Bernoulli}(\pi_i)$

\[
p(y_i | \pi_i) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1 - y_i}
\]

2. Specify a linear predictor: $X_i \beta$
1. Specify a distribution for $Y$
1. Specify a distribution for $Y$

$$Y_i \sim \text{Bernoulli}(\pi_i)$$
1. Specify a distribution for $Y$

$$Y_i \sim \text{Bernoulli} (\pi_i)$$

$$p(y|\pi) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$
1. Specify a distribution for $Y$

$Y_i \sim \text{Bernoulli}(\pi_i)$

$p(y|\pi) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$

2. Specify a linear predictor:
1. Specify a distribution for $Y$

$$Y_i \sim \text{Bernoulli}(\pi_i)$$

$$p(y|\pi) = \prod_{i=1}^{n} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

2. Specify a linear predictor:

$$X_i \beta$$
3. Specify a link (or inverse link) function.
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

\[
\log(-\log(1 - \pi_i)) = X_i\beta
\]

\[
\pi_i = 1 - \exp(-\exp(X_i\beta))
\]
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

\[
\log(-\log(1 - \pi_i)) = X_i \beta
\]

\[
\pi_i = 1 - \exp(-\exp(X_i \beta))
\]

We could also have chosen several other options:
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

\[
\log(-\log(1 - \pi_i)) = X_i\beta
\]

\[
\pi_i = 1 - \exp(-\exp(X_i\beta))
\]

We could also have chosen several other options:

- Probit: \( \pi_i = \Phi(x_i/\beta) \)
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

\[ \log(-\log(1 - \pi_i)) = \mathbf{X}_i\beta \]

\[ \pi_i = 1 - \exp(-\exp(\mathbf{X}_i\beta)) \]

We could also have chosen several other options:

- Probit: \( \pi_i = \Phi(\mathbf{x}_i/\beta) \)
- Logit:
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

$$\log(-\log(1 - \pi_i)) = X_i \beta$$

$$\pi_i = 1 - \exp(-\exp(X_i \beta))$$

We could also have chosen several other options:

- Probit: $$\pi_i = \Phi(x_i \beta)$$
- Logit:
  $$\pi_i = \frac{1}{1 + e^{-x_i \beta}}$$
3. Specify a link (or inverse link) function.

Complementary Log-log (cloglog):

\[
\log(-\log(1 - \pi_i)) = X_i \beta
\]

\[
\pi_i = 1 - \exp(-\exp(X_i \beta))
\]

We could also have chosen several other options:

- Probit: \( \pi_i = \Phi(x_i \beta) \)
- Logit:
  \[
  \pi_i = \frac{1}{1 + e^{-x_i \beta}}
  \]
- Scobit: \( \pi_i = (1 + e^{-x_i \beta})^{-\alpha} \)
Our model

In the notation of Unifying Political Methodology, this is the model we've just defined:

\[ Y_i \sim \text{Bernoulli}(\pi_i) \]

\[ \pi_i = 1 - \exp(-\exp(X_i \beta)) \]
In the notation of *Unifying Political Methodology*, this is the model we’ve just defined:
In the notation of *Unifying Political Methodology*, this is the model we’ve just defined:

\[ Y_i \sim \text{Bernoulli}(\pi_i) \]
\[ \pi_i = 1 - \exp(-\exp(X_i \beta)) \]
Log-likelihood of the c-loglog
Log-likelihood of the c-loglog

\[ \ell(\beta \mid Y) = \log(L(\beta \mid Y)) \]
Log-likelihood of the c-loglog

\[ \ell(\beta \mid Y) = \log(L(\beta \mid Y)) \]
\[ = \log(p(Y \mid \beta)) \]
Log-likelihood of the c-loglog

\[
\ell(\beta \mid Y) = \log(L(\beta \mid Y)) \\
= \log(p(Y \mid \beta)) \\
= \log \left( \prod_{i=1}^{n} p(Y_i \mid \beta) \right)
\]
Log-likelihood of the c-loglog

\[
\ell(\beta \mid Y) = \log(L(\beta \mid Y)) \\
= \log(p(Y \mid \beta)) \\
= \log\left(\prod_{i=1}^{n} p(Y_i \mid \beta)\right) \\
= \log\left(\prod_{i=1}^{n} [1 - \exp(-\exp(X_i\beta))]^{Y_i} \exp(-\exp(X_i\beta))^{(1-Y_i)}\right)
\]
Log-likelihood of the c-loglog

\[ \ell(\beta \mid Y) = \log(L(\beta \mid Y)) = \log(p(Y \mid \beta)) = \log \left( \prod_{i=1}^{n} p(Y_i \mid \beta) \right) \]

\[ = \log \left( \prod_{i=1}^{n} \left[ 1 - \exp(-\exp(X_i\beta)) \right]^{Y_i} \left[ \exp(-\exp(X_i\beta)) \right]^{(1-Y_i)} \right) \]

\[ = \sum_{i=1}^{n} \left( Y_i \log(1 - \exp(-\exp(X_i\beta))) + (1 - Y_i) \log[\exp(-\exp(X_i\beta))] \right) \]
Log-likelihood of the c-loglog

\[ \ell(\beta \mid Y) = \log(L(\beta \mid Y)) \]
\[ = \log(p(Y \mid \beta)) \]
\[ = \log \left( \prod_{i=1}^{n} p(Y_i \mid \beta) \right) \]
\[ = \log \left( \prod_{i=1}^{n} [1 - \exp(-\exp(X_i \beta))]^{Y_i} [\exp(-\exp(X_i \beta))]^{(1-Y_i)} \right) \]
\[ = \sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i \beta)))) + (1 - Y_i) \log[\exp(-\exp(X_i \beta))] \]
\[ = \sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i \beta))) - (1 - Y_i) \exp(X_i \beta)) \]
Coding our log likelihood function

\[ \sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i \beta)))) - (1 - Y_i) \exp(X_i \beta) \]
Coding our log likelihood function

\[ \sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i \beta)))) - (1 - Y_i) \exp(X_i \beta)) \]

cloglog.loglik <- function(par, X, y) {
}
Coding our log likelihood function

\[ \sum_{i=1}^{n} (Y_i \log(1 - \exp(- \exp(X_i \beta)))) - (1 - Y_i) \exp(X_i \beta)) \]

cloglog.loglik <- function(par, X, y) {
    beta <- par
Coding our log likelihood function

\[
\sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i \beta)))) - (1 - Y_i) \exp(X_i \beta))
\]

cloglog.loglik <- function(par, X, y) {
  beta <- par
  log.lik <- sum(y * log(1 - exp(-exp(X %*% beta))) -
                 (1 - y) * exp(X %*% beta))
  return(log.lik)
}
Coding our log likelihood function

\[ \sum_{i=1}^{n} (Y_i \log(1 - \exp(-\exp(X_i\beta)))) - (1 - Y_i) \exp(X_i\beta)) \]

cloglog.loglik <- function(par, X, y) {
  beta <- par

  log.lik <- sum(y * log(1 - exp(-exp(X %*% beta))) - (1 - y) * exp(X %*% beta))

  return(log.lik)
}
Finding the MLE

\[
X <- \text{model.matrix}(\sim \text{married} + \text{cm1ethrace} + \text{income}, \text{data} = \text{d})
\]

\[
\text{opt} <- \text{optim}(\text{par} = \text{rep}(0, \text{ncol}(X)), \text{fn} = \text{cloglog.loglik}, \text{X} = X, \text{y} = \text{d$ev}}, \text{control} = \text{list(fnscale = -1), hessian = T, method = "BFGS"})
\]

Point estimate of the MLE:

\[
\text{opt$par}[1] = -2.704, -1.211, -0.526, -0.620, -0.272, -0.348
\]
Finding the MLE

\[
X <- \text{model.matrix}(\sim \text{married} + \text{cm1lethrace} + \text{income}, \\
data = d)
\]

\[
\text{opt <- optim(par = rep(0, ncol(X))},
\]

Point estimate of the MLE:

\[
\text{opt$par[1]}: -2.704
\]

\[
\text{opt$par[2]}: -1.211
\]

\[
\text{opt$par[3]}: -0.526
\]

\[
\text{opt$par[4]}: -0.620
\]

\[
\text{opt$par[5]}: -0.272
\]

\[
\text{opt$par[6]}: -0.348
\]
Finding the MLE

```
X <- model.matrix(~married + cm1ethrace + income, data = d)
opt <- optim(par = rep(0, ncol(X)),
             fn = cloglog.loglik,
             X = X, y = d$ev,
             control = list(fnscale = -1), hessian = T, method = "BFGS")
```

Point estimate of the MLE: `opt$par`
Finding the MLE

```r
X <- model.matrix(~married + cm1lethrace + income, 
                 data = d)

opt <- optim(par = rep(0, ncol(X)), 
             fn = cloglog.loglik,
             X = X, 
             y = d$ev,
             control = list(fnscale = -1),
             hessian = T,
             method = "BFGS")

Point estimate of the MLE:
opt$par

[1] -2.704 -1.211 -0.526 -0.620 -0.272 -0.348
```
Finding the MLE

\[
X \leftarrow \text{model.matrix}(\text{~married + cm1ethrace + income, } \text{data = d})
\]

\[
\text{opt} \leftarrow \text{optim}(\text{par = rep(0, ncol(X)), }
\hspace{1cm} \text{fn = cloglog.loglik,}
\hspace{1cm} \text{X = X,}
\hspace{1cm} \text{y = d$ev,}
\hspace{1cm} \text{control = list(fnscale = -1)},
\]

Point estimate of the MLE:

\[
\text{opt$par [1]} = -2.704 \hspace{0.5cm} -1.211 \hspace{0.5cm} -0.526 \hspace{0.5cm} -0.620 \hspace{0.5cm} -0.272 \hspace{0.5cm} -0.348
\]
Finding the MLE

```r
X <- model.matrix(~married + cm1ethrace + income, data = d)
opt <- optim(par = rep(0, ncol(X)),
            fn = cloglog.loglik,
            X = X,
            y = d$ev,
            control = list(fnscale = -1),
            hessian = T,
            method = "BFGS")
Point estimate of the MLE:
opt$par
[1] -2.704 -1.211 -0.526 -0.620 -0.272 -0.348
```
Finding the MLE

\[
X <- \text{model.matrix}(~\text{married} + \text{cm1ethrace} + \text{income},
\quad \text{data} = d)
\]
\[
\text{opt} <- \text{optim}(\text{par} = \text{rep}(0, \text{ncol}(X)),
\quad \text{fn} = \text{cloglog.loglik},
\quad X = X,
\quad y = \text{d$ev},
\quad \text{control} = \text{list(fnscale} = -1),
\quad \text{hessian} = T,
\quad \text{method} = "\text{BFGS}")
\]
Finding the MLE

```r
X <- model.matrix(~married + cm1ethrace + income,
                 data = d)
opt <- optim(par = rep(0, ncol(X)),
             fn = cloglog.loglik,
             X = X,
             y = d$ev,
             control = list(fnscale = -1),
             hessian = T,
             method = "BFGS")

Point estimate of the MLE:

opt$par
[1] -2.704 -1.211 -0.526 -0.620 -0.272 -0.348
```
Standard errors of the MLE

Credit to Stephen Pettigrew for including this figure in slides.
Standard errors of the MLE

Recall that the standard errors are defined as the diagonal of:

$$\sqrt{-\left[ \frac{\partial^2 \ell}{\partial \beta^2} \right]^{-1}}$$

\(^2\text{Credit to Stephen Pettigrew for including this figure in slides.}\)
Standard errors of the MLE

Recall that the standard errors are defined as the diagonal of:

$$\sqrt{-\left[ \frac{\partial^2 \ell}{\partial \beta^2} \right]^{-1}}$$

where $\frac{\partial^2 \ell}{\partial \beta^2}$ is the

\(^2\)Credit to Stephen Pettigrew for including this figure in slides.
Standard errors of the MLE

Recall that the standard errors are defined as the diagonal of:

$$
\sqrt{-\left[ \frac{\partial^2 \ell}{\partial \beta^2} \right]^{-1}}
$$

where $\frac{\partial^2 \ell}{\partial \beta^2}$ is the

\[^2\text{Credit to Stephen Pettigrew for including this figure in slides.}\]
Standard errors of the MLE
Standard errors of the MLE

Variance-covariance matrix:

```
-solve(opt$hessian)
[1,]  0.026  -0.002  -0.021  -0.021  -0.019  -0.006
[2,] -0.002   0.063   0.004   0.002  -0.001  -0.004
[3,] -0.021   0.004   0.026   0.019   0.018   0.002
[4,] -0.021   0.002   0.019   0.033   0.018   0.002
[5,] -0.019  -0.001   0.018   0.018   0.128   0.002
[6,] -0.006  -0.004   0.002   0.002   0.002   0.004
```
Standard errors of the MLE

**Variance-covariance matrix:**

$$-\text{solve}(\text{opt}\$\text{hessian})$$

$$\begin{bmatrix}
[1,] & 0.026 & -0.002 & -0.021 & -0.021 & -0.019 & -0.006 \\
[2,] & -0.002 & 0.063 & 0.004 & 0.002 & -0.001 & -0.004 \\
[3,] & -0.021 & 0.004 & 0.026 & 0.019 & 0.018 & 0.002 \\
[4,] & -0.021 & 0.002 & 0.019 & 0.033 & 0.018 & 0.002 \\
[5,] & -0.019 & -0.001 & 0.018 & 0.018 & 0.128 & 0.002 \\
[6,] & -0.006 & -0.004 & 0.002 & 0.002 & 0.002 & 0.004 \\
\end{bmatrix}$$

**Standard errors:**

$$\text{sqrt(diag}(-\text{solve}(\text{opt}\$\text{hessian})))$$

$$\begin{bmatrix}
[1] & 0.162 & 0.252 & 0.160 & 0.183 & 0.358 & 0.064 \\
\end{bmatrix}$$
Outline

1. GLMs
   - General Structure of GLMs
   - Procedure for Running a GLM

2. Complementary log-log

3. Quantities of Interest
Interpreting c-loglog coefficients

Here's a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does this table even mean?
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does this
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does this table
Interpreting c-loglog coefficients

Here’s a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does this table even
Interpreting c-loglog coefficients

Here's a nicely formatted table with your regression results from our model:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.70</td>
<td>0.16</td>
</tr>
<tr>
<td>Married</td>
<td>-1.21</td>
<td>0.25</td>
</tr>
<tr>
<td>Black</td>
<td>-0.53</td>
<td>0.16</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.62</td>
<td>0.18</td>
</tr>
<tr>
<td>Other</td>
<td>-0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Income / poverty line</td>
<td>-0.35</td>
<td>0.06</td>
</tr>
</tbody>
</table>

But what does this table even mean?
Interpreting c-loglog results

What does it even mean for the coefficient for married to be -1.21?

All else constant, children of married parents have -1.21 points lower log rate of eviction.

And what are log rates?

Nobody thinks in terms of log odds, or probit coefficients, or exponential rates.
Interpreting c-loglog results

What does it even mean for the coefficient for married to be -1.21?
Interpreting c-loglog results

What does it even mean for the coefficient for married to be -1.21?

All else constant, children of married parents have -1.21 points lower log rate of eviction.
Interpreting c-loglog results

What does it even mean for the coefficient for married to be -1.21?

All else constant, children of married parents have -1.21 points lower log rate of eviction.

And what are log rates?
Interpreting c-loglog results

What does it even mean for the coefficient for married to be -1.21?

All else constant, children of married parents have -1.21 points lower log rate of eviction.

And what are log rates? Nobody thinks in terms of log odds, or probit coefficients, or exponential rates.
If there’s one thing you take away from this class, it should be this:
If there’s one thing you take away from this class, it should be this:

When you present results, **always** present your findings in terms of something that has substantive meaning to the reader.
If there’s one thing you take away from this class, it should be this:

**When you present results, always present your findings in terms of something that has substantive meaning to the reader.**

For binary outcome models that often means turning your results into predicted probabilities, which is what we’ll do now.
If there’s one thing you take away from this class, it should be this:

**When you present results, **always** present your findings in terms of something that has substantive meaning to the reader.**

For binary outcome models that often means turning your results into predicted probabilities, which is what we’ll do now.

If there’s a second thing you should take away, it’s this:
If there’s one thing you take away from this class, it should be this:

When you present results, **always** present your findings in terms of something that has substantive meaning to the reader.

For binary outcome models that often means turning your results into predicted probabilities, which is what we’ll do now.

If there’s a second thing you should take away, it’s this:

**Always account for all types of uncertainty when you present your results**
If there’s one thing you take away from this class, it should be this:

**When you present results, always present your findings in terms of something that has substantive meaning to the reader.**

For binary outcome models that often means turning your results into predicted probabilities, which is what we’ll do now.

If there’s a second thing you should take away, it’s this:

**Always account for all types of uncertainty when you present your results**

We’ll spend the rest of today looking at how to do that.
Getting Quantities of Interest

1. Write out your model and estimate $\hat{\beta}_{MLE}$ and the Hessian.
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty.
3. Multiply these simulated $\tilde{\beta}$'s by some covariates in the model to get $\tilde{X}\beta$.
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function.
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty.
6. Store the mean of these simulations, $E[y|X]$.
7. Repeat steps 2 through 6 thousands of times.
8. Use the results to make fancy graphs and informative tables.
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian.
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty.
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$.
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function.
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty.
6. Store the mean of these simulations, $E[y|X]$.
7. Repeat steps 2 through 6 thousands of times.

Use the results to make fancy graphs and informative tables.
Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
7. Repeat steps 2 through 6 thousands of times
8. Use the results to make fancy graphs and informative tables
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

By the central limit theorem, we assume that $\hat{\beta}_{MLE} \sim \text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

By the central limit theorem, we assume that $\hat{\beta}_{MLE} \sim \text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$

$\hat{\beta}$ is the vector of our estimates for the parameters, opt$\text{par}$
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

By the central limit theorem, we assume that

$$\hat{\beta}_{MLE} \sim \text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$$

$\hat{\beta}$ is the vector of our estimates for the parameters, opt$par$

$\hat{V}(\hat{\beta})$ is the variance-covariance matrix, -solve(opt$hessian$)
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

By the central limit theorem, we assume that
$\hat{\beta}_{MLE} \sim \text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$

$\hat{\beta}$ is the vector of our estimates for the parameters, opt$\cdot$par

$\hat{V}(\hat{\beta})$ is the variance-covariance matrix, -solve(opt$\cdot$hessian)

We hope that the $\hat{\beta}$s we estimated are good estimates of the true $\beta$s, but we know that they aren’t exactly perfect because of estimation uncertainty.
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

By the central limit theorem, we assume that

$\hat{\beta}_{MLE} \sim \text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$

$\hat{\beta}$ is the vector of our estimates for the parameters, opt$\text{par}$

$\hat{V}(\hat{\beta})$ is the variance-covariance matrix, -solve(opt$hessian)'

We hope that the $\hat{\beta}$s we estimated are good estimates of the true $\beta$s, but we know that they aren’t exactly perfect because of estimation uncertainty.

So we account for this uncertainty by simulating $\beta$s from the multivariate normal distribution defined above
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

Simulate one draw from $\text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$
Simulate from the sampling distribution of $\hat{\beta}_{MLE}$

Simulate one draw from $\text{mvnorm}(\hat{\beta}, \hat{V}(\hat{\beta}))$
Install the mvtnorm package if you need to

```r
install.packages("mvtnorm")
require(mvtnorm)
```

```r
sim.betas <- rmvnorm(n = 1,
                      mean = opt$par,
                      sigma = -solve(opt$hessian))
```

```
[1,]  -2.825 -1.424 -0.478 -0.358 -0.123 -0.237
```
Now we need to choose some values of the covariates that we want predictions about. Let's make predictions for one white child born to married parents with family income at the poverty line. Recall that our predictors (in order) are:

```r
> colnames(X)
[1] "(Intercept)" "married" "cm1ethraceBlack" "cm1ethraceHispanic"
[5] "cm1ethraceOther" "income"
```

We can set the values of X as:

```r
setX <- c(1, 1, 0, 0, 0, 1)
```
Now we need to choose some values of the covariates that we want predictions about.
Now we need to choose some values of the covariates that we want predictions about.

Let’s make predictions for one white child born to married parents with family income at the poverty line. Recall that our predictors (in order) are:

```r
> colnames(X)
[1] "(Intercept)"           "married"           "cm1ethraceBlack"  "cm1ethraceHispanic"
[5] "cm1ethraceOther"       "income"
```

We can set the values of X as:
Untransform $X^\beta$

Now we need to choose some values of the covariates that we want predictions about.

Let’s make predictions for one white child born to married parents with family income at the poverty line Recall that our predictors (in order) are:

```r
> colnames(X)
[1] "(Intercept)" "married" "cm1ethraceBlack" "cm1ethraceHispanic"
[5] "cm1ethraceOther" "income"
```

We can set the values of $X$ as:

```r
setX <- c(1, 1, 0, 0, 0, 1)
```
Untransform $X\beta$
Untransform $X\beta$

Now we multiply our covariates of interest by our simulated parameters:

```r
setX %*% t(sim.betas)
  [,1]
[1,] -4.486287
```

If we stopped right here we'd be making two mistakes:

1. $-4.486287$ is not the predicted probability (obviously - it's negative!), it's the predicted log rate.
2. We haven't done a very good job of accounting for the uncertainty in the model.
Untransform $X\beta$

Now we multiply our covariates of interest by our simulated parameters:

```r
setX %*% t(sim.betas)
[,1]
[1,] -4.486287
```

If we stopped right here we’d be making two mistakes.
Untransform $X\beta$

Now we multiply our covariates of interest by our simulated parameters:

```
setX %*% t(sim.betas)
[,1]
[1,] -4.486287
```

If we stopped right here we’d be making two mistakes.

1. -4.486287 is not the predicted probability (obviously - it’s negative!), it’s the predicted log rate
Untransform $X\beta$

Now we multiply our covariates of interest by our simulated parameters:

```r
setX %*% t(sim.betas)
[,1]
[1,] -4.486287
```

If we stopped right here we’d be making two mistakes.

1. -4.486287 is not the predicted probability (obviously - it’s negative!), it’s the predicted log rate
2. We haven’t done a very good job of accounting for the uncertainty in the model
Untransform $X\beta$
To turn $\tilde{X} \tilde{\beta}$ into a predicted probability we need to plug it back into our link function, which was $1 - \exp(-\exp(X_i\beta))$. 

```r
> sim.p <- 1 - exp(-exp(setX %*% t(sim.betas)))
> sim.p[,1]
[,1] 0.0111992
```
Untransform $X_\beta$

To turn $\tilde{X}\tilde{\beta}$ into a predicted probability we need to plug it back into our link function, which was $1 - \exp(-\exp(X_i\beta))$

```r
> sim.p <- 1 - exp(-exp(setX %*% t(sim.betas)))
> sim.p
[,1]
[1,] 0.0111992
```
Simulate from the stochastic function
Simulate from the stochastic function

Now we have to account for fundamental uncertainty by simulating from the original stochastic function, Bernoulli
Simulate from the stochastic function

Now we have to account for fundamental uncertainty by simulating from the original stochastic function, Bernoulli

```r
> draws <- rbinom(n = 10000, size = 1, prob = sim.p)
> mean(draws)
[1] 0.0117
```
Simulate from the stochastic function

Now we have to account for fundamental uncertainty by simulating from the original stochastic function, Bernoulli

```r
> draws <- rbinom(n = 10000, size = 1, prob = sim.p)
> mean(draws)
[1] 0.0117
```

Is 0.0117 our best guess at the predicted probability of eviction?
Simulate from the stochastic function

Now we have to account for fundamental uncertainty by simulating from the original stochastic function, Bernoulli

```r
> draws <- rbinom(n = 10000, size = 1, prob = sim.p)
> mean(draws)
[1] 0.0117
```

Is 0.0117 our best guess at the predicted probability of eviction? Nope
Store and repeat
Store and repeat

Remember, we only took 1 draw of our $\beta$s from the multivariate normal distribution.
Store and repeat

Remember, we only took 1 draw of our $\beta$s from the multivariate normal distribution.

To fully account for estimation uncertainty, we need to take tons of draws of $\tilde{\beta}$.
Store and repeat

Remember, we only took 1 draw of our $\beta$s from the multivariate normal distribution.

To fully account for estimation uncertainty, we need to take tons of draws of $\tilde{\beta}$.

To do this we’d need to loop over all the steps I just went through and get the full distribution of predicted probabilities for this case.
Speeding up the process
Speeding up the process

Or, instead of using a loop, let’s just vectorize our code:
Speeding up the process

Or, instead of using a loop, let’s just vectorize our code:

```r
sim.betas <- rmvnorm(n = 10000,
                     mean = opt$par,
                     sigma = -solve(opt$hessian))
```
Speeding up the process

Or, instead of using a loop, let’s just vectorize our code:

```r
sim.betas <- rmvnorm(n = 10000,
                      mean = opt$par,
                      sigma = -solve(opt$hessian))

head(sim.betas)
[1,] -2.800 -0.955 -0.545 -0.531 -0.217 -0.319
[2,] -2.410 -0.879 -0.772 -0.896 -0.301 -0.485
[3,] -2.715 -1.672 -0.483 -0.741 -0.365 -0.333
[4,] -2.718 -0.993 -0.588 -0.545 -0.094 -0.389
[5,] -2.479 -1.161 -0.721 -0.794 -0.601 -0.413
[6,] -2.625 -1.227 -0.654 -0.586 -0.496 -0.351
```
Speeding up the process

Or, instead of using a loop, let’s just vectorize our code:

```r
sim.betas <- rmvnorm(n = 10000, 
                      mean = opt$par, 
                      sigma = -solve(opt$hessian))

head(sim.betas)
```

```
[1,] -2.800 -0.955 -0.545 -0.531 -0.217 -0.319
[2,] -2.410 -0.879 -0.772 -0.896 -0.301 -0.485
[3,] -2.715 -1.672 -0.483 -0.741 -0.365 -0.333
[4,] -2.718 -0.993 -0.588 -0.545 -0.094 -0.389
[5,] -2.479 -1.161 -0.721 -0.794 -0.601 -0.413
[6,] -2.625 -1.227 -0.654 -0.586 -0.496 -0.351
```

```r
dim(sim.betas)
```

```
[1] 10000   6
```
Speeding up the process
Speeding up the process

Now multiply the $10,000 \times 6$ $\tilde{\beta}$ matrix by your $1 \times 6$ vector of $\tilde{X}$ of interest

```r
pred.xb <- setX %*% t(sim.betas)
> pred.prob <- 1 - exp(-exp(pred.xb))
> pred.prob[,1:5]
```

```
[1] 0.016858874 0.022674923 0.008876607 0.016426627 0.017223131
```
Speeding up the process

Now multiply the $10,000 \times 6 \tilde{\beta}$ matrix by your $1 \times 6$ vector of $\tilde{X}$ of interest

```r
pred.xb <- setX %*% t(sim.betas)
```
Speeding up the process

Now multiply the 10,000 x 6 \( \tilde{\beta} \) matrix by your 1 x 6 vector of \( \tilde{X} \) of interest

\[ \text{pred.xb} <- \text{setX} \times \text{t(sim.betas)} \]

And untransform them

\[ \text{pred.prob} \left[ \right] \text{[,1:5]} \]
\[ \text{[1]} 0.016858874 0.022674923 0.008876607 0.016426627 0.017223131 \]
Speeding up the process

Now multiply the $10,000 \times 6 \bar{\beta}$ matrix by your $1 \times 6$ vector of $\bar{X}$ of interest

```
pred.xb <- setX %*% t(sim.betas)
```

And untransform them

```
> pred.prob <- 1 - exp(-exp(pred.xb))
> pred.prob[,1:5]
[1] 0.016858874 0.022674923 0.008876607 0.016426627 0.017223131
```
Look at Our Results in Tabular Form
Look at Our Results in Tabular Form
Look at Our Results

```r
> mean(pred.prob)
[1] 0.01446521

> quantile(pred.prob, prob = c(.025, .975))
2.5% 97.5%
0.00844883 0.02295435

> mean(pred.prob > .02)
[1] 0.0828
```
Look at Our Results

```r
> mean(pred.prob)
[1] 0.01446521
> quantile(pred.prob, prob = c(.025, .975))
    2.5%       97.5%
0.00844883  0.02295435

> mean(pred.prob > .02)
[1] 0.0828
```
Different QOI

What if our QOI was the chance of any eviction from birth to age 9?

$$P(\text{Ever evicted}) = 1 - P(\text{Never evicted}) = 1 - 9 \prod_{i=1}^{9} (1 - P(\text{Evicted at age } i))$$

All that will change is the very last step.

Note: We assume independence between eviction in each year, and a constant risk over time. This corresponds to an Exponential survival model.
Different QOI

What if our QOI was the chance of any eviction from birth to age 9?

\[ P(\text{Ever evicted}) = 1 - P(\text{Never evicted}) = 1 - \prod_{i=1}^{9} (1 - P(\text{Evicted at age } i)) = 1 - (1 - p)^9 \]

All that will change is the very last step.

\[ ^3 \text{Note: We assume independence between eviction in each year, and a constant risk over time. This corresponds to an Exponential survival model.} \]
What if our QOI was the chance of any eviction from birth to age 9?

\[ P(\text{Ever evicted}) = 1 - P(\text{Never evicted}) \]
\[ = 1 - \prod_{i=1}^{9} (1 - P(\text{Evicted at age } i)) \]
\[ = 1 - (1 - p)^9 \]

All that will change is the very last step.³

³Note: We assume independence between eviction in each year, and a constant risk over time. This corresponds to an Exponential survival model.
Different QOI

We already estimated the sampling distribution of $p$ and stored samples from this distribution in the vector `predprob`. Now we can just transform them!

$$P(\text{Ever evicted}_i) = 1 - (1 - p_i)^9$$

```
pred.prob <- 1 - (1 - pred.prob) ^ 9
mean(pred.prob)
[1] 0.1224441
```

The probability of eviction looks much higher than we had thought!
Different QOI

We already estimated the sampling distribution of $p$ and stored samples from this distribution in the vector `predprob`. Now we can just transform them!

$$P(\text{Ever evicted}_i) = 1 - (1 - p_i)^9$$

```r
> cum.prob <- 1 - (1 - pred.prob) ^ 9
> mean(cum.prob)
[1] 0.1224441
```

12%! The probability of eviction looks much higher than we had thought!
Look at Our Results For Both QOIs
Look at Our Results For Both QOIs
Quantities of interest matter in continuous cases as well

Example: Modeling log income
Modeling log income

We want to model the effect of college \((D)\) on earnings \((Y)\), net of age \((X)\).
Let's get some data! http://cps.ipums.org
Causal identification

We state an **ignorability assumption**: 

\[ \{ Y(0), Y(1) \} \perp \perp D | X \]

\[ D \]

\[ Y \] = Age

This is our identification strategy but it says nothing about estimation.
Causal identification

We state an **ignorability assumption**:

\[
\{ Y(0), Y(1) \} \perp \!
\!
\perp D \mid X
\]
Causal identification

We state an **ignorability assumption**:\[
\{ Y(0), Y(1) \} \perp D \mid X
\]

\[ X = \text{Age} \]
Causal identification

We state an **ignorability assumption**:

\[
\{ Y(0), Y(1) \} \perp \perp D \mid X
\]

\[ X = \text{Age} \]

This is our **identification strategy**
Causal identification

We state an **ignorability assumption**:

\[ \{ Y(0), Y(1) \} \perp \!\!\!\!\!\!\!\!\!\!\!\!\perp D \mid X \]

This is our **identification strategy** but it says nothing about **estimation**.
Estimation via GLMs

1. Specify a distribution for $Y$...
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

$$Y \sim \text{LogNormal}(\mu, \sigma^2)$$
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

   $Y \sim \text{LogNormal}(\mu, \sigma^2)$

2. Specify a linear predictor
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

$$Y \sim \text{LogNormal}(\mu, \sigma^2)$$

2. Specify a linear predictor

$$X \beta$$
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

   $Y \sim \text{LogNormal}(\mu, \sigma^2)$

2. Specify a linear predictor

   $X \beta$

3. Specify a link function
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

\[ Y \sim \text{LogNormal}(\mu, \sigma^2) \]

2. Specify a linear predictor

\[ X\beta \]

3. Specify a link function

\[ \log(\mu) = X\beta \]
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!
   \[ Y \sim \text{LogNormal}(\mu, \sigma^2) \]

2. Specify a linear predictor
   \[ X \beta \]

3. Specify a link function
   \[ \log(\mu) = X \beta \]

4. Estimate parameters via maximum likelihood
Estimation via GLMs

1. Specify a distribution for $Y$... LogNormal!

\[ Y \sim \text{LogNormal}(\mu, \sigma^2) \]

2. Specify a linear predictor

\[ X\beta \]

3. Specify a link function

\[ \log(\mu) = X\beta \]

4. Estimate parameters via maximum likelihood

5. Simulate quantities of interest
Likelihood

\[ L(\beta, \sigma^2 \mid Y) \propto f(Y \mid \beta, \sigma^2) \]

\[ = \prod_{i=1}^{n} f(Y_i \mid \beta, \sigma^2) \]

\[ = \prod_{i=1}^{n} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - \mu)^2}{2\sigma^2} \right) \]
Log likelihood

\[ \ell(\beta, \sigma^2 \mid \mathbf{Y}) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right] \]
Log likelihood

\[ \ell(\beta, \sigma^2 \mid Y) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right] \]

\[ = \ln \left[ \prod_{i=1}^{N} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \]
Log likelihood

\[
\ell(\beta, \sigma^2 \mid \mathbf{Y}) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right] \\
= \ln \left[ \prod_{i=1}^{N} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \\
= \sum_{i=1}^{N} \ln \left[ \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right]
\]
Log likelihood

\[ \ell(\beta, \sigma^2 \mid Y) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right] \]

\[ = \ln \left[ \prod_{i=1}^{N} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \]

\[ = \sum_{i=1}^{N} \ln \left[ \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \]

\[ = \sum_{i=1}^{N} \left( -\ln(Y_i) - \ln(\sigma) - \ln(\sqrt{2\pi}) + \ln \left[ \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \right) \]
Log likelihood

\[
\ell(\beta, \sigma^2 \mid Y) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right]
\]

\[
= \ln \left[ \prod_{i=1}^{N} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right]
\]

\[
= \sum_{i=1}^{N} \ln \left[ \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right]
\]

\[
= \sum_{i=1}^{N} \left( -\ln(Y_i) - \ln(\sigma) - \ln(\sqrt{2\pi}) + \ln \left[ \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \right)
\]

\[
= \sum_{i=1}^{N} \left( -\ln(Y_i) - \ln(\sigma) - \ln(\sqrt{2\pi}) - \frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right)
\]
Log likelihood

\[ \ell(\beta, \sigma^2 \mid \mathbf{Y}) = \ln \left[ \prod_{i=1}^{N} f(Y_i \mid \beta, \sigma^2) \right] \]

\[ = \ln \left[ \prod_{i=1}^{N} \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \]

\[ = \sum_{i=1}^{N} \ln \left[ \frac{1}{Y_i \sigma \sqrt{2\pi}} \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \]

\[ = \sum_{i=1}^{N} \left( -\ln(Y_i) - \ln(\sigma) - \ln(\sqrt{2\pi}) + \ln \left[ \exp \left( -\frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \right] \right) \]

\[ = \sum_{i=1}^{N} \left( -\ln(Y_i) - \ln(\sigma) - \ln(\sqrt{2\pi}) - \frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \]

\[ = \sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i \beta)^2}{2\sigma^2} \right) \]
Coding our log likelihood function

\[
\sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right)
\]
Coding our log likelihood function

\[ \sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right) \]

```r
logNormal.log.lik <- function(par, X, y) {
  beta <- par[-length(par)]
  sigma2 <- exp(par[length(par)])
  log.lik <- sum(-log(sqrt(sigma2)) - ((log(y) - X %*% beta) ^ 2) / (2*sigma2))
  return(log.lik)
}
```
Coding our log likelihood function

\[ \sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right) \]

\[
\text{logNormal.log.lik} <- \text{function}(\text{par}, X, y) \{
\text{beta} <- \text{par}[-\text{length}(\text{par})]
\}
\]
Coding our log likelihood function

\[ \sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right) \]

\text{logNormal.log.lik} <- function(par, X, y) {
  beta <- par[-length(par)]
  sigma2 <- exp(par[length(par)])
  return(log.lik)
}
Coding our log likelihood function

\[
\sum_{i=1}^{N} \left( - \ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right)
\]

\[
\text{logNormal.log.lik} \left<\text{ function}(\text{par, X, y}) \right> \{
\begin{align*}
\beta & \left<\text{ par[-length(par)]} \right. \\
\text{sigma2} & \left<\text{ exp(par[length(par)])} \right. \\
\text{log.lik} & \left<\text{ sum(-log(sqrt(sigma2)) - ((log(y) - X \%\% beta) ^ 2) / (2*sigma2))} \right.
\end{align*}
\]
Coding our log likelihood function

\[ \sum_{i=1}^{N} \left( -\ln(\sigma) - \frac{(\ln(Y_i) - X_i\beta)^2}{2\sigma^2} \right) \]

\[
\text{logNormal.log.lik} <- \text{function(par, X, y) } \{ \\
\beta <- \text{par}[-\text{length(par)}] \\
\sigma^2 <- \exp(\text{par}[\text{length(par)}]) \\
\text{log.lik} <- \text{sum}(-\log(\sqrt{\sigma^2}) - \sqrt{\sigma^2} - \left( (\log(y) - X \%\% beta)^2 \right) / (2*\sigma^2)) \\
\text{return(log.lik)} \\
\} 
\]
Finding the MLE

\[
X <- \text{model.matrix(}~ \text{college + age,} \\
    \text{data = d})
\]

\[
y <- d\$\text{incwage}
\]

\[
\text{opt} <- \text{optim(}par = \text{rep(0, ncol(X) + 1)}, \\
    \text{fn = logNormal.log.lik,} \\
    \text{y = y,} \\
    \text{X = X,} \\
    \text{control = list(fnscale = -1),} \\
    \text{method = "BFGS",} \\
    \text{hessian = TRUE})
\]
Extract the MLE

```r
> opt$par
[1] 8.84550213 0.72904517 0.03270657 -0.03216774
```
Extract the MLE

> opt$par
[1] 8.84550213 0.72904517 0.03270657 -0.03216774

See that it matches what we get with LM

> lm.fit <- lm(log(incwage) ~ college + age, +             data = d)
> coef(lm.fit)
  (Intercept) collegeTRUE age
  8.84550213 0.72904517 0.03270657
Extract the MLE

> opt$par
[1]  8.84550213  0.72904517  0.03270657 -0.03216774

See that it matches what we get with LM

> lm.fit <- lm(log(incwage) ~ college + age,
+     data = d)
> coef(lm.fit)
(Intercept) collegeTRUE age
  8.84550213  0.72904517  0.03270657

Why is that last term negative?
Extract the MLE

> opt$par
[1]  8.84550213  0.72904517  0.03270657 -0.03216774

See that it matches what we get with LM

> lm.fit <- lm(log(incwage) ~ college + age,
+     data = d)
> coef(lm.fit)
(Intercept) collegeTRUE       age
   8.84550213   0.72904517   0.03270657

Why is that last term negative? Because it’s the \( \log \) of \( \sigma^2 \)!
See how $\sigma^2$ matches

```r
> summary(lm.fit)

Call:
lm(formula = log(incwage) ~ college + age, data = d)
.......other output....
Residual standard error: 0.9841 on 68932 degrees of freedom

We estimated
```
See how $\sigma^2$ matches

```r
> summary(lm.fit)

Call:
  lm(formula = log(incwage) ~ college + age, data = d)
  ........other output....
Residual standard error: 0.9841 on 68932 degrees of freedom

We estimated $\sigma^2 = e^\gamma = e^{-0.03216774}$

> exp(opt$par[4])
[1] 0.9683441
See how $\sigma^2$ matches

```r
> summary(lm.fit)

Call:
  lm(formula = log(incwage) ~ college + age, data = d)

....other output....
Residual standard error: 0.9841 on 68932 degrees of freedom

We estimated $\sigma^2 = e^\gamma = e^{-0.03216774}$

> exp(opt$par[4])
[1] 0.9683441

It matches!
We know how to calculate the variance-covariance matrix - the inverse of the negative Hessian!

\[
\text{vcov.optim <- -solve(opt$hessian)}
\]

\[
\text{vcov.optim}
\]

\[
\begin{bmatrix}
[1,,] & 1.938105e-04 & -6.974235e-06 & -4.762674e-06 & 1.171351e-13 \\
[2,,] & -6.974235e-06 & 6.311970e-05 & -4.031325e-07 & 1.714362e-14 \\
[3,,] & -4.762674e-06 & -4.031325e-07 & 1.316824e-07 & 2.028694e-14 \\
[4,,] & 1.171351e-13 & 1.714362e-14 & 2.028694e-14 & 2.901283e-05
\end{bmatrix}
\]

And we can compare that to the canned version...

\[
\text{vcov(lm.fit)}
\]

\[
\begin{bmatrix}
(Intercept) & collegeTRUE & age \\
(Intercept) & 1.938189e-04 & -6.974537e-06 & -4.762880e-06 \\
collegeTRUE & -6.974537e-06 & 6.312244e-05 & -4.031500e-07 \\
age & -4.762880e-06 & -4.031500e-07 & 1.316881e-07
\end{bmatrix}
\]

...which matches!
Variance-covariance matrix matches

We know how to calculate the variance-covariance matrix - the inverse of the negative Hessian!

```r
> vcov.optim <- -solve(opt$hessian)
> vcov.optim

[1,]  1.938105e-04 -6.974235e-06 -4.762674e-06  1.171351e-13
[2,] -6.974235e-06  6.311970e-05 -4.031325e-07  1.714362e-14
[3,] -4.762674e-06 -4.031325e-07  1.316824e-07  2.028694e-14
[4,]  1.171351e-13  1.714362e-14  2.028694e-14  2.901283e-05
```

And we can compare that to the canned version...
Variance-covariance matrix matches

We know how to calculate the variance-covariance matrix - the inverse of the negative Hessian!

```r
> vcov.optim <- -solve(opt$hessian)
> vcov.optim

[1,] 1.938105e-04 -6.974235e-06 -4.762674e-06 1.171351e-13
[2,] -6.974235e-06 6.311970e-05 -4.031325e-07 1.714362e-14
[3,] -4.762674e-06 -4.031325e-07 1.316824e-07 2.028694e-14
[4,] 1.171351e-13 1.714362e-14 2.028694e-14 2.901283e-05
```

And we can compare that to the canned version...

```r
> vcov(lm.fit)

(Intercept) collegeTRUE age
(Intercept) 1.938189e-04 -6.974537e-06 -4.762880e-06
collegeTRUE -6.974537e-06 6.312244e-05 -4.031500e-07
age -4.762880e-06 -4.031500e-07 1.316881e-07
```

....which matches!
What is the effect of college on earnings?

Our model can easily estimate the effect of college on log earnings.
What is the effect of college on earnings?

Our model can easily estimate the effect of college on log earnings. But we want the effect of college on earnings.
What is the effect of college on earnings?

Our model can easily estimate the effect of college on log earnings. But we want the effect of college on earnings.

One could try to interpret the lognormal regression coefficients. We estimated $\beta_{\text{College}} = 0.729$. 

What is the effect of college on earnings?

Our model can easily estimate the effect of college on log earnings. But we want the effect of college on earnings.

One could try to interpret the lognormal regression coefficients. We estimated $\beta_{\text{College}} = 0.729$.

This means a unit increase in age increases earnings by a factor of $e^{0.729} = 2.208$. 
What is the effect of college on earnings?

Our model can easily estimate the effect of college on log earnings. But we want the effect of college on earnings.

One could try to interpret the lognormal regression coefficients. We estimated $\beta_{\text{College}} = 0.729$.

This means a unit increase in age increases earnings by a factor of $e^{0.729} = 2.208$. 
**Student:** Can’t we just exponentiate things?
Student: Can’t we just exponentiate things?

No! We run into Jensen’s inequality.

\[ E(Y) = E(e^{\log(Y)}) \geq e^{E(\log(Y))} \]

So we can’t just exponentiate predicted values. What can we do instead?
**Student:** Can’t we just exponentiate things?

No! We run into Jensen’s inequality.

\[ E(Y) = E(e^{\log(Y)}) \geq e^{E(\log(Y))} \]

So we can’t just exponentiate predicted values. What can we do instead? **Simulation!**
How wrong can you be?

[If time allows, we can walk through this in R]
How wrong can you be?

[If time allows, we can walk through this in R]
Using the naive approach of exponentiating things, we would find an effect of college on earnings of \$25,181.27.
How wrong can you be?

[If time allows, we can walk through this in R]
Using the naive approach of exponentiating things, we would find an effect of college on earnings of $25,181.27.$

Using simulation, the actual average treatment effect is $43,266.67!$
How wrong can you be?

[If time allows, we can walk through this in R]
Using the naive approach of exponentiating things, we would find an effect of college on earnings of $25,181.27

Using simulation, the actual average treatment effect is $43,266.67!

Why the discrepancy? (Draw on the board).
**Student:** Ok. But since there were no interactions in the model, we don’t have to average over the population, right?
Student: Ok. But since there were no interactions in the model, we don’t have to average over the population, right?

No! Effect sizes depend on the values chosen for the rest of the covariates.
Different effect sizes for different groups!

<table>
<thead>
<tr>
<th></th>
<th>Effect</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-year-olds</td>
<td>23,276.88</td>
<td>19,466.92</td>
<td>27,494.30</td>
</tr>
<tr>
<td>50-year-olds</td>
<td>62,319.15</td>
<td>51,498.74</td>
<td>73,129.50</td>
</tr>
</tbody>
</table>

Since 50-year-olds have higher predicted earnings to begin with, multiplying by a factor of 2.208 increases their earnings by more dollars.
Student: Ok, so now I see that I need to carefully specify and simulate my quantity of interest, involving both estimation uncertainty and fundamental uncertainty.
**Student:** Ok, so now I see that I need to carefully specify and simulate my quantity of interest, involving both estimation uncertainty and fundamental uncertainty.

Which of those goes away as the sample size grows?
Estimation uncertainty disappears in large samples

Each row section below shows the difference in earnings (college - noncollege), for 50 year olds.

$'N = 100'$

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in any two 50-year-olds</td>
<td>-186146.3 890301.0</td>
<td></td>
</tr>
<tr>
<td>Average difference</td>
<td>53387.9 249905.7</td>
<td></td>
</tr>
</tbody>
</table>

$'N = 1,000'$

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in any two 50-year-olds</td>
<td>-169635.63 479623.19</td>
<td></td>
</tr>
<tr>
<td>Average difference</td>
<td>51505.21 85801.36</td>
<td></td>
</tr>
</tbody>
</table>

$'N = 50,000'$

<table>
<thead>
<tr>
<th></th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference in any two 50-year-olds</td>
<td>-177369.77 459907.68</td>
<td></td>
</tr>
<tr>
<td>Average difference</td>
<td>52190.08 73289.31</td>
<td></td>
</tr>
</tbody>
</table>

The expected difference is more precisely estimated with larger sample sizes, but fundamental uncertainty makes it always difficult to make precise predictions about the difference between any two actual individuals.
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write down your model and estimate $\hat{\beta}$, the MLE.
2. Simulate from the sampling distribution of $\hat{\beta}$ to incorporate estimation uncertainty.
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$.
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function.
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty.
6. Store the mean of these simulations, $\mathbb{E}[y|X]$.
7. Repeat steps 2 through 6 thousands of times.
8. Use the results to make fancy graphs and informative tables.
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
7. Repeat steps 2 through 6 thousands of times
8. Use the results to make fancy graphs and informative tables
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write out your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\tilde{\beta}$
4. Plug $\tilde{X}\tilde{\beta}$ into your link function, $g^{-1}(\tilde{X}\tilde{\beta})$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\tilde{\beta})$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write our your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
7. Repeat steps 2 through 6 thousands of times

Use the results to make fancy graphs and informative tables
Conclusion: Getting Quantities of Interest

How to present results in a better format than just coefficients and standard errors:

1. Write out your model and estimate $\hat{\beta}_{MLE}$ and the Hessian
2. Simulate from the sampling distribution of $\hat{\beta}_{MLE}$ to incorporate estimation uncertainty
3. Multiply these simulated $\tilde{\beta}$s by some covariates in the model to get $\tilde{X}\beta$
4. Plug $\tilde{X}\beta$ into your link function, $g^{-1}(\tilde{X}\beta)$, to put it on the same scale as the parameter(s) in your stochastic function
5. Use the transformed $g^{-1}(\tilde{X}\beta)$ to take thousands of draws from your stochastic function and incorporate fundamental uncertainty
6. Store the mean of these simulations, $E[y|X]$
7. Repeat steps 2 through 6 thousands of times
8. Use the results to make fancy graphs and informative tables