1 Background

2 GLMs

3 Logistic Regression

4 Additional Example
Regression so far ...

At this point we have covered:

- Simple linear regression
  - Relationship between numerical response and a numerical or categorical predictor
Regression so far...

At this point we have covered:

- **Simple linear regression**
  - Relationship between numerical response and a numerical or categorical predictor

- **Multiple regression**
  - Relationship between numerical response and multiple numerical and/or categorical predictors
Regression so far ...

At this point we have covered:

- Simple linear regression
  - Relationship between numerical response and a numerical or categorical predictor
- Multiple regression
  - Relationship between numerical response and multiple numerical and/or categorical predictors

What we haven’t seen is what to do when the predictors are weird (nonlinear, complicated dependence structure, etc.) or when the response is weird (categorical, count data, etc.)
Recap of what you should know how to do ...

- Model parameter interpretation
- Hypothesis tests for slope and intercept parameters
- Hypothesis tests for all regression parameters
- Confidence intervals for regression parameters
- Confidence and prediction intervals for predicted means and values (SLR only)
- Model diagnostics, residuals plots, outliers
- $R^2$, Adjusted $R^2$
- Model selection (MLR only)
- Simple transformations
Odds

Odds are another way of quantifying the probability of an event, commonly used in gambling (and logistic regression).

For some event $E$,

$$\text{odds}(E) = \frac{P(E)}{P(E^c)} = \frac{P(E)}{1 - P(E)}$$

Similarly, if we are told the odds of $E$ are $x$ to $y$ then

$$\text{odds}(E) = \frac{x}{y} = \frac{x}{x+y}$$

which implies

$$P(E) = \frac{x}{x+y}, \quad P(E^c) = \frac{y}{x+y}$$
1 Background

2 GLMs

3 Logistic Regression

4 Additional Example
Example - Donner Party

In 1846 the Donner and Reed families left Springfield, Illinois, for California by covered wagon. In July, the Donner Party, as it became known, reached Fort Bridger, Wyoming. There its leaders decided to attempt a new and untested route to the Sacramento Valley. Having reached its full size of 87 people and 20 wagons, the party was delayed by a difficult crossing of the Wasatch Range and again in the crossing of the desert west of the Great Salt Lake. The group became stranded in the eastern Sierra Nevada mountains when the region was hit by heavy snows in late October. By the time the last survivor was rescued on April 21, 1847, 40 of the 87 members had died from famine and exposure to extreme cold.

### Example - Donner Party - Data

<table>
<thead>
<tr>
<th>Age</th>
<th>Sex</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.00</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>40.00</td>
<td>Female</td>
</tr>
<tr>
<td>3</td>
<td>40.00</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>30.00</td>
<td>Male</td>
</tr>
<tr>
<td>5</td>
<td>28.00</td>
<td>Male</td>
</tr>
<tr>
<td>43</td>
<td>23.00</td>
<td>Male</td>
</tr>
<tr>
<td>44</td>
<td>24.00</td>
<td>Male</td>
</tr>
<tr>
<td>45</td>
<td>25.00</td>
<td>Female</td>
</tr>
</tbody>
</table>
### Example - Donner Party - EDA

#### Status vs. Gender:

<table>
<thead>
<tr>
<th>Status</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Survived</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Example - Donner Party - EDA

Status vs. Gender:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Died</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td>Survived</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Status vs. Age:

![Box plot showing age distribution for died and survived individuals.]

Statistics 102 (Colin Rundel)  
Lec 20  
April 15, 2013  
7 / 30
Example - Donner Party - ???

It seems clear that both age and gender have an effect on someone’s survival, how do we come up with a model that will let us explore this relationship?
Example - Donner Party - ???

It seems clear that both age and gender have an effect on someone’s survival, how do we come up with a model that will let us explore this relationship?

Even if we set Died to 0 and Survived to 1, this isn’t something we can transform our way out of - we need something more.
Example - Donner Party - ???

It seems clear that both age and gender have an effect on someone’s survival, how do we come up with a model that will let us explore this relationship?

Even if we set Died to 0 and Survived to 1, this isn’t something we can transform our way out of - we need something more.

One way to think about the problem - we can treat Survived and Died as successes and failures arising from a binomial distribution where the probability of a success is given by a transformation of a linear model of the predictors.
It turns out that this is a very general way of addressing this type of problem in regression, and the resulting models are called generalized linear models (GLMs). Logistic regression is just one example of this type of model.
Generalized linear models

It turns out that this is a very general way of addressing this type of problem in regression, and the resulting models are called generalized linear models (GLMs). Logistic regression is just one example of this type of model.

All generalized linear models have the following three characteristics:

1. A probability distribution describing the outcome variable
2. A linear model
   \[ \eta = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n \]
3. A link function that relates the linear model to the parameter of the outcome distribution
   \[ g(p) = \eta \text{ or } p = g^{-1}(\eta) \]
1 Background

2 GLMs

3 Logistic Regression

4 Additional Example
Logistic regression is a GLM used to model a binary categorical variable using numerical and categorical predictors.

We assume a binomial distribution produced the outcome variable and we therefore want to model $p$ the probability of success for a given set of predictors.
Logistic regression is a GLM used to model a binary categorical variable using numerical and categorical predictors.

We assume a binomial distribution produced the outcome variable and we therefore want to model $p$ the probability of success for a given set of predictors.

To finish specifying the Logistic model we just need to establish a reasonable link function that connects $\eta$ to $p$. There are a variety of options but the most commonly used is the logit function.

Logit function

$$\text{logit}(p) = \log \left( \frac{p}{1 - p} \right), \text{ for } 0 \leq p \leq 1$$
Properties of the Logit

The logit function takes a value between 0 and 1 and maps it to a value between $-\infty$ and $\infty$.

Inverse logit (logistic) function

$$g^{-1}(x) = \frac{\exp(x)}{1 + \exp(x)} = \frac{1}{1 + \exp(-x)}$$

The inverse logit function takes a value between $-\infty$ and $\infty$ and maps it to a value between 0 and 1.

This formulation also has some use when it comes to interpreting the model as logit can be interpreted as the log odds of a success, more on this later.
The logistic regression model

The three GLM criteria give us:

\[ y_i \sim \text{Binom}(p_i) \]

\[ \eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \]

\[ \text{logit}(p) = \eta \]

From which we arrive at,

\[ p_i = \frac{\exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i})}{1 + \exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_n x_{n,i})} \]
Example - Donner Party - Model

In R we fit a GLM in the same was as a linear model except using `glm` instead of `lm` and we must also specify the type of GLM to fit using the `family` argument.

```r
summary(glm(Status ~ Age, data=donner, family=binomial))
```

```r
## Call:
## glm(formula = Status ~ Age, family = binomial, data = donner)
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.81852 0.99937 1.820 0.0688 .
## Age -0.06647 0.03222 -2.063 0.0391 *
##
## Null deviance: 61.827 on 44 degrees of freedom
## Residual deviance: 56.291 on 43 degrees of freedom
## AIC: 60.291
##
## Number of Fisher Scoring iterations: 4
```
### Logistic Regression

**Example - Donner Party - Prediction**

|                      | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 1.8185   | 0.9994     | 1.82    | 0.0688   |
| Age                  | -0.0665  | 0.0322     | -2.06   | 0.0391   |

Model:

\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age}
\]
Example - Donner Party - Prediction

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 1.8185 | 0.9994 | 1.82 | 0.0688 |
| Age | -0.0665 | 0.0322 | -2.06 | 0.0391 |

Model:

$$\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age}$$

Odds / Probability of survival for a newborn (Age=0):
Example - Donner Party - Prediction

|            | Estimate | Std. Error | z value | Pr(>|z|) |
|------------|----------|------------|---------|----------|
| (Intercept)| 1.8185   | 0.9994     | 1.82    | 0.0688   |
| Age        | -0.0665  | 0.0322     | -2.06   | 0.0391   |

Model:

\[
\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times \text{Age}
\]

Odds / Probability of survival for a newborn (Age=0):

\[
\log\left(\frac{p}{1-p}\right) = 1.8185 - 0.0665 \times 0
\]

\[
\frac{p}{1-p} = \exp(1.8185) = 6.16
\]

\[
p = 6.16 / 7.16 = 0.86
\]
Model:

\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age}
\]

Odds / Probability of survival for a 25 year old:

\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 25
\]

\[
p = \text{exp}(0.156) = 1.17
\]

\[
\frac{p}{1 - p} = 0.539
\]

Odds / Probability of survival for a 50 year old:

\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 0
\]

\[
p = \text{exp}(-1.5065) = 0.222
\]

\[
\frac{p}{1 - p} = 0.181
\]
Example - Donner Party - Prediction (cont.)

Model:
\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age}
\]

Odds / Probability of survival for a 25 year old:
\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 25
\]
\[
\frac{p}{1 - p} = \exp(0.156) = 1.17
\]
\[
p = 1.17/2.17 = 0.539
\]
Example - Donner Party - Prediction (cont.)

Model:
\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age}
\]

Odds / Probability of survival for a 25 year old:
\[
\log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 25
\]
\[
\frac{p}{1 - p} = \exp(0.156) = 1.17
\]
\[
p = 1.17/2.17 = 0.539
\]

Odds / Probability of survival for a 50 year old:
Model:

\[ \log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times \text{Age} \]

Odds / Probability of survival for a 25 year old:

\[ \log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 25 \]

\[ \frac{p}{1 - p} = \exp(0.156) = 1.17 \]

\[ p = 1.17/2.17 = 0.539 \]

Odds / Probability of survival for a 50 year old:

\[ \log \left( \frac{p}{1 - p} \right) = 1.8185 - 0.0665 \times 0 \]

\[ \frac{p}{1 - p} = \exp(-1.5065) = 0.222 \]

\[ p = 0.222/1.222 = 0.181 \]
\[
\log \left( \frac{p}{1-p} \right) = 1.8185 - 0.0665 \times \text{Age}
\]
log \left( \frac{p}{1-p} \right) = 1.8185 - 0.0665 \times \text{Age}
## Example - Donner Party - Interpretation

|                  | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | 1.8185   | 0.9994     | 1.82    | 0.0688   |
| Age              | -0.0665  | 0.0322     | -2.06   | 0.0391   |

Simple interpretation is only possible in terms of log odds and log odds ratios for intercept and slope terms.

**Intercept**: The log odds of survival for a party member with an age of 0. From this we can calculate the odds or probability, but additional calculations are necessary.

**Slope**: For a unit increase in age (being 1 year older) how much will the log odds ratio change, not particularly intuitive. More often then not we care only about sign and relative magnitude.
Logistic Regression

Example - Donner Party - Interpretation - Slope

\[
\log \left( \frac{p_1}{1 - p_1} \right) = 1.8185 - 0.0665(x + 1)
\]
\[
= 1.8185 - 0.0665x - 0.0665
\]

\[
\log \left( \frac{p_2}{1 - p_2} \right) = 1.8185 - 0.0665x
\]

\[
\log \left( \frac{p_1}{1 - p_1} \right) - \log \left( \frac{p_2}{1 - p_2} \right) = -0.0665
\]

\[
\log \left( \frac{p_1}{1 - p_1} / \frac{p_2}{1 - p_2} \right) = -0.0665
\]

\[
\frac{p_1}{1 - p_1} / \frac{p_2}{1 - p_2} = \exp(-0.0665) = 0.94
\]
**Example - Donner Party - Age and Gender**

summary(glm(Status ~ Age + Sex, data=donner, family=binomial))

```r
## Call:
## glm(formula = Status ~ Age + Sex, family = binomial, data = donner)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.63312 1.11018  1.471 0.1413
## Age         -0.07820 0.03728 -2.097 0.0359 *
## SexFemale 1.59729 0.75547  2.114 0.0345 *
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 61.827 on 44 degrees of freedom
## Residual deviance: 51.256 on 42 degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
```

**Gender slope**: When the other predictors are held constant this is the log odds ratio between the given level (Female) and the reference level (Male).
Example - Donner Party - Gender Models

Just like MLR we can plug in gender to arrive at two status vs age models for men and women respectively.

**General model:**

$$\log \left( \frac{p_1}{1 - p_1} \right) = 1.63312 - 0.07820 \times \text{Age} + 1.59729 \times \text{Sex}$$

**Male model:**

$$\log \left( \frac{p_1}{1 - p_1} \right) = 1.63312 - 0.07820 \times \text{Age} + 1.59729 \times 0$$

$$= 1.63312 - 0.07820 \times \text{Age}$$

**Female model:**

$$\log \left( \frac{p_1}{1 - p_1} \right) = 1.63312 - 0.07820 \times \text{Age} + 1.59729 \times 1$$

$$= 3.23041 - 0.07820 \times \text{Age}$$
Logistic Regression

Example - Donner Party - Gender Models (cont.)

![Graph showing logistic regression model for status vs. age for males and females.](image-url)
Example - Donner Party - Gender Models (cont.)

![Graph showing logistic regression for gender models.](image)
Hypothesis test for the whole model

summary(glm(Status ~ Age + Sex, data=donner, family=binomial))

## Call:
## glm(formula = Status ~ Age + Sex, family = binomial, data = donner)
##
## ## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.63312  1.11018  1.471 0.1413
## Age -0.07820  0.03728 -2.097 0.0359 *
## SexFemale 1.59729  0.75547  2.114 0.0345 *
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 61.827 on 44 degrees of freedom
## Residual deviance: 51.256 on 42 degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4
Hypothesis test for the whole model

summary(glm(Status ~ Age + Sex, data=donner, family=binomial))

## Call:
## glm(formula = Status ~ Age + Sex, family = binomial, data = donner)
##
## ## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.63312 1.11018 1.471 0.1413
## Age -0.07820 0.03728 -2.097 0.0359 *
## SexFemale 1.59729 0.75547 2.114 0.0345 *
## ---
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 61.827 on 44 degrees of freedom
## Residual deviance: 51.256 on 42 degrees of freedom
## AIC: 57.256
##
## Number of Fisher Scoring iterations: 4

Note that the model output does not include any F-statistic, as a general rule there are not single model hypothesis tests for GLM models.
### Hypothesis tests for a coefficient

|              | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------|----------|------------|---------|----------|
| (Intercept)  | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age          | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale    | 1.5973   | 0.7555     | 2.11    | 0.0345   |

We are however still able to perform inference on individual coefficients, the basic setup is exactly the same as what we’ve seen before except we use a Z test.

Note the only tricky bit, which is way beyond the scope of this course, is how the standard error is calculated.
Testing for the slope of Age

|          | Estimate | Std. Error | z value | Pr(>|z|) |
|----------|----------|------------|---------|----------|
| (Intercept) | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age       | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale | 1.5973   | 0.7555     | 2.11    | 0.0345   |
## Testing for the slope of Age

|                | Estimate | Std. Error | z value | Pr(>|z|)  |
|----------------|----------|------------|---------|-----------|
| (Intercept)    | 1.6331   | 1.1102     | 1.47    | 0.1413    |
| Age            | -0.0782  | 0.0373     | -2.10   | 0.0359    |
| SexFemale      | 1.5973   | 0.7555     | 2.11    | 0.0345    |

\[ H_0 : \beta_{age} = 0 \]

\[ H_A : \beta_{age} \neq 0 \]
### Testing for the slope of Age

|               | Estimate | Std. Error | z value | Pr(>|z|) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age           | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale     | 1.5973   | 0.7555     | 2.11    | 0.0345   |

\[
H_0 : \beta_{age} = 0 \\
H_A : \beta_{age} \neq 0
\]

\[
Z = \frac{\hat{\beta}_{age} - \beta_{age}}{SE_{age}} = \frac{-0.0782 - 0}{0.0373} = -2.10
\]

\[
p\text{-value} = P(|Z| > 2.10) = P(Z > 2.10) + P(Z < -2.10) \\
= 2 \times 0.0178 = 0.0359
\]
### Confidence interval for age slope coefficient

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age            | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale      | 1.5973   | 0.7555     | 2.11    | 0.0345   |

Remember, the interpretation for a slope is the change in log odds ratio per unit change in the predictor.
Confidence interval for age slope coefficient

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age            | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale      | 1.5973   | 0.7555     | 2.11    | 0.0345   |

Remember, the interpretation for a slope is the change in log odds ratio per unit change in the predictor.

Log odds ratio:

\[
CI = PE \pm CV \times SE = -0.0782 \pm 1.96 \times 0.0373 = (-0.1513, -0.0051)
\]
Confidence interval for age slope coefficient

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | 1.6331   | 1.1102     | 1.47    | 0.1413   |
| Age            | -0.0782  | 0.0373     | -2.10   | 0.0359   |
| SexFemale      | 1.5973   | 0.7555     | 2.11    | 0.0345   |

Remember, the interpretation for a slope is the change in log odds ratio per unit change in the predictor.

Log odds ratio:

\[
CI = PE \pm CV \times SE = -0.0782 \pm 1.96 \times 0.0373 = (-0.1513, -0.0051)
\]

Odds ratio:

\[
\exp(CI) = (\exp(-0.1513), \exp(-0.0051)) = (0.85960, 0.9949)
\]
1 Background

2 GLMs

3 Logistic Regression

4 Additional Example
Example - Birdkeeping and Lung Cancer

A 1972 - 1981 health survey in The Hague, Netherlands, discovered an association between keeping pet birds and increased risk of lung cancer. To investigate birdkeeping as a risk factor, researchers conducted a case-control study of patients in 1985 at four hospitals in The Hague (population 450,000). They identified 49 cases of lung cancer among the patients who were registered with a general practice, who were age 65 or younger and who had resided in the city since 1965. They also selected 98 controls from a population of residents having the same general age structure.

### Example - Birdkeeping and Lung Cancer - Data

<table>
<thead>
<tr>
<th></th>
<th>LC</th>
<th>FM</th>
<th>SS</th>
<th>BK</th>
<th>AG</th>
<th>YR</th>
<th>CD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LungCancer</td>
<td>Male</td>
<td>Low</td>
<td>Bird</td>
<td>37.00</td>
<td>19.00</td>
<td>12.00</td>
</tr>
<tr>
<td>2</td>
<td>LungCancer</td>
<td>Male</td>
<td>Low</td>
<td>Bird</td>
<td>41.00</td>
<td>22.00</td>
<td>15.00</td>
</tr>
<tr>
<td>3</td>
<td>LungCancer</td>
<td>Male</td>
<td>High</td>
<td>NoBird</td>
<td>43.00</td>
<td>19.00</td>
<td>15.00</td>
</tr>
<tr>
<td>147</td>
<td>NoCancer</td>
<td>Female</td>
<td>Low</td>
<td>NoBird</td>
<td>65.00</td>
<td>7.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>

**LC** - Whether subject has lung cancer
**FM** - Sex of subject
**SS** - Socioeconomic status
**BK** - Indicator for birdkeeping
**AG** - Age of subject (years)
**YR** - Years of smoking prior to diagnosis or examination
**CD** - Average rate of smoking (cigarettes per day)

*Note* - NoCancer is the reference response (0 or failure), LungCancer is the non-reference response (1 or success) - this matters for interpretation.
Example - Birdkeeping and Lung Cancer - EDA

![Graph showing the relationship between Years of Smoking and Age (Years) for Bird and No Bird categories with Lung Cancer and No Lung Cancer.]

<table>
<thead>
<tr>
<th></th>
<th>Bird</th>
<th>No Bird</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Cancer</td>
<td>▲</td>
<td>●</td>
</tr>
<tr>
<td>No Lung Cancer</td>
<td>△</td>
<td>○</td>
</tr>
</tbody>
</table>
Example - Birdkeeping and Lung Cancer - Model

summary(glm(LC ~ FM + SS + BK + AG + YR + CD, data=bird, family=binomial))

## Call:
## glm(formula = LC ~ FM + SS + BK + AG + YR + CD, family = binomial, 
##     data = bird)

## Coefficients:
##                  Estimate Std. Error   z value  Pr(>|z|)       
## (Intercept)    -1.93736   1.80425  -1.074     0.282924
## FMFemale        0.56127   0.53116   1.057     0.290653
## SSHigh          0.10545   0.46885   0.225     0.822050
## BKBird          1.36259   0.41128   3.313     0.000923 ***
## AG              -0.03976   0.03548  -1.120     0.262503
## YR              0.07287   0.02649   2.751     0.005940 **
## CD              0.02602   0.02552   1.019     0.308055

## (Dispersion parameter for binomial family taken to be 1)

## Null deviance: 187.14  on 146  degrees of freedom
## Residual deviance: 154.20 on 140  degrees of freedom
## AIC: 168.2

## Number of Fisher Scoring iterations: 5
### Example - Birdkeeping and Lung Cancer - Interpretation

|                  | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | -1.9374  | 1.8043     | -1.07   | 0.2829   |
| FMFemale         | 0.5613   | 0.5312     | 1.06    | 0.2907   |
| SSHigh           | 0.1054   | 0.4688     | 0.22    | 0.8221   |
| BKBird           | 1.3626   | 0.4113     | 3.31    | 0.0009   |
| AG               | -0.0398  | 0.0355     | -1.12   | 0.2625   |
| YR               | 0.0729   | 0.0265     | 2.75    | 0.0059   |
| CD               | 0.0260   | 0.0255     | 1.02    | 0.3081   |

Keeping all other predictors constant, the odds ratio of getting lung cancer for bird keepers vs non-bird keepers is $\exp(1.3626) = 3.91$. The odds ratio of getting lung cancer for an additional year of smoking is $\exp(0.0729) = 1.08$. 
### Example - Birdkeeping and Lung Cancer - Interpretation

|                  | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------|----------|------------|---------|----------|
| (Intercept)      | -1.9374  | 1.8043     | -1.07   | 0.2829   |
| FMFemale         | 0.5613   | 0.5312     | 1.06    | 0.2907   |
| SSHigh           | 0.1054   | 0.4688     | 0.22    | 0.8221   |
| **BKBird**       | **1.3626**| **0.4113** | **3.31**| **0.0009**|
| AG               | -0.0398  | 0.0355     | -1.12   | 0.2625   |
| **YR**           | **0.0729**| **0.0265** | **2.75**| **0.0059**|
| CD               | 0.0260   | 0.0255     | 1.02    | 0.3081   |

Keeping all other predictors constant then,
Keeping all other predictors constant then,

- The odds ratio of getting lung cancer for bird keepers vs non-bird keepers is \( \exp(1.3626) = 3.91 \).
Example - Birdkeeping and Lung Cancer - Interpretation

|                        | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------------|----------|------------|---------|---------|
| (Intercept)            | -1.9374  | 1.8043     | -1.07   | 0.2829  |
| FMFemale               | 0.5613   | 0.5312     | 1.06    | 0.2907  |
| SSHigh                | 0.1054   | 0.4688     | 0.22    | 0.8221  |
| BKBird                | 1.3626   | 0.4113     | 3.31    | 0.0009  |
| AG                    | -0.0398  | 0.0355     | -1.12   | 0.2625  |
| YR                    | 0.0729   | 0.0265     | 2.75    | 0.0059  |
| CD                    | 0.0260   | 0.0255     | 1.02    | 0.3081  |

Keeping all other predictors constant then,

- The odds ratio of getting lung cancer for bird keepers vs non-bird keepers is \( \exp(1.3626) = 3.91 \).
- The odds ratio of getting lung cancer for an additional year of smoking is \( \exp(0.0729) = 1.08 \).