Logistic Regression

Comparing Proportions and Odds

Dr. Tackett

11.06.2018
Announcements

- HW 6 due Thursday, 11/8
- Lab 8 due Sunday, 11/11
- Project Proposal due Tuesday, 11/13
  - Example posters in office
- If you are registered, please VOTE! Polls close in NC at 7:30p.
  - Find your polling place at the NC State Board of Elections website
Packages

```r
library(knitr)
library(broom)
library(dplyr)
library(tibble)
library(ggplot2)
```
Modeling Binary Outcomes
FiveThirtyEight March Madness

2018 March Madness Predictions

Live Win Probabilities are "derived using logistic regression analysis, which lets us plug the current state of a game into a model to produce the probability that either team will win the game."

-"How Our March Madness Predictions Work"
FiveThirtyEight Election Forecasts

1 in 5
Chance Democrats win control (18.9%)

4 in 5
Chance Republicans keep control (81.1%)

53
tyco.com Senate forecast

7 in 8
Chance Democrats win control (87.8%)

1 in 8
Chance Republicans keep control (12.2%)

FiveThirtyEight.com House forecast
Our models are probabilistic in nature; we do a lot of thinking about these probabilities, and the goal is to develop \textbf{probabilistic estimates} that hold up well under real-world conditions.

-"How FiveThirtyEight's House, Senate, and Governor Models Work"
Response Variable, $Y$

- $Y$ is a binary response variable
  - $1$: yes
  - $0$: no

- $\text{Mean}(Y) = \pi$
  - $\pi$ is the proportion of "yes" responses in the population

- $\text{Variance}(Y) = \pi(1 - \pi)$
Sampling Distribution of Sample Proportion

- $\hat{\pi}$: average of binary responses in the sample
  - Called the sample proportion
  - This is the statistic, i.e. the estimate of $\pi$

- Given $\hat{\pi}$ is the sample proportion based on a sample of size $n$ from a population with population proportion $\pi$:

$$\hat{\pi} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

...assuming $n$ is "large"
Hypothesis Test for Single Proportion

- Null Hypothesis: $H_0 : \pi = \pi_0$

- Test Statistic:

\[ z = \frac{\hat{\pi} - \pi_0}{SE_0(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \]

- p-value: proportion of $N(0, 1)$ distribution as extreme or more extreme than the test statistic
Confidence Interval for a Single Proportion

- Approximate $100(1 - \alpha)\%$ confidence interval for $\pi$ is

$$\hat{\pi} \pm z^* SE(\hat{\pi})$$

$$= \hat{\pi} \pm z^* \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$
Confidence Interval for a Single Proportion

- Approximate $100(1 - \alpha)\%$ confidence interval for $\pi$ is

$$\hat{\pi} \pm z^* \text{SE}(\hat{\pi})$$

$$= \hat{\pi} \pm z^* \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution.
Confidence Interval for a Single Proportion

- Approximate $100(1 - \alpha)\%$ confidence interval for $\pi$ is

$$\hat{\pi} \pm z^* SE(\hat{\pi})$$

$$= \hat{\pi} \pm z^* \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution

```
# Critical value for 90% CI
qnorm(0.95)
```

```
## [1] 1.644854
```
$n$ "large"

- **Hypothesis Tests**: Test reliable if

$$n\hat{\pi} > 5 \text{ and } n(1 - \hat{\pi}) > 5$$
n "large"

- **Hypothesis Tests**: Test reliable if

  \[ n\hat{\pi} > 5 \text{ and } n(1 - \hat{\pi}) > 5 \]

- **Confidence Intervals**: Confidence interval reliable if

  \[ n\hat{\pi} > 5 \text{ and } n(1 - \hat{\pi}) > 5 \]
Sampling Distribution for Difference in Two Proportions

- Let $\hat{\pi}_1$ and $\hat{\pi}_2$ be sample proportions from independent random samples of size $n_1$ and $n_2$, respectively:

\[
\hat{\pi}_1 - \hat{\pi}_2 \sim N\left(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}\right)
\]

... assuming $n_1$ and $n_2$ are "large"
Standard Errors

- For Confidence Intervals:

\[
SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1}} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}
\]
Standard Errors

- **For Confidence Intervals:**

\[
SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}
\]

- **For Hypothesis Tests:**
  - The null hypothesis is that the proportions are equal
  - We estimate, \( \hat{\pi}_c \), the sample proportion from the combined data

\[
SE_0(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_1} + \frac{\hat{\pi}_c(1 - \hat{\pi}_c)}{n_2}}
\]
Hypothesis Test for Equal Proportions

- We want to test if two groups have equal population proportions

- **Null Hypothesis:** $H_0 : \pi_1 = \pi_2$

- **Test Statistic:**

  $$ z = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - 0}{SE_0(\hat{\pi}_1 - \hat{\pi}_2)} = \frac{(\hat{\pi}_1 - \hat{\pi}_2) - 0}{\sqrt{\frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_1} + \frac{\hat{\pi}_c(1-\hat{\pi}_c)}{n_2}}} $$

- **p-value:** proportion of $N(0, 1)$ distribution as extreme or more extreme than the test statistic
Confidence Interval for Difference in Proportions

- Approximate $100(1 - \alpha)\%$ confidence interval for $\pi_1 - \pi_2$ is

\[
(\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times SE(\hat{\pi}_1 - \hat{\pi}_2)
\]

\[
= (\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}
\]
Confidence Interval for Difference in Proportions

- Approximate $100(1 - \alpha)$% confidence interval for $\pi_1 - \pi_2$ is

\[
(\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times SE(\hat{\pi}_1 - \hat{\pi}_2)
\]

\[
= (\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}
\]

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution
Confidence Interval for Difference in Proportions

- Approximate $100(1 - \alpha)\%$ confidence interval for $\pi_1 - \pi_2$ is

$$
(\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times SE(\hat{\pi}_1 - \hat{\pi}_2)
= (\hat{\pi}_1 - \hat{\pi}_2) \pm z^* \times \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}
$$

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution

# Critical value for 95% CI
qnorm(0.975)

## [1] 1.959964
$n_1$ and $n_2$ "large"

- **Hypothesis Tests**: Test reliable if

\[ n_s \hat{\pi}_c > 5 \quad \text{and} \quad n_s (1 - \hat{\pi}_c) > 5 \]

where $n_s$ is the smaller of $n_1$ and $n_2$, and $\hat{\pi}_c$ is the sample proportion from the combined data
$n_1$ and $n_2$ "large"

- **Hypothesis Tests**: Test reliable if

$$n_s \hat{\pi}_c > 5 \quad \text{and} \quad n_s (1 - \hat{\pi}_c) > 5$$

where $n_s$ is the smaller of $n_1$ and $n_2$, and $\hat{\pi}_c$ is the sample proportion from the combined data.

- **Confidence Intervals**: Confidence interval reliable if

$$n_i \hat{\pi}_i > 5 \quad \text{and} \quad n_i (1 - \hat{\pi}_i) > 5 \quad \text{for} \ i = 1, 2$$
Ex: Eye Witness Identification

- What factors affect identification made in a suspect lineup?
Ex: Eye Witness Identification

- What factors affect identification made in a suspect lineup?

  - Researchers conducted an experiment in which over 2000 participants watched a video of a crime and were tasked with selecting the culprit out of a line up
    - Factors: weapon, features on culprit, lineup procedure
Ex: Eye Witness Identification

- What factors affect identification made in a suspect lineup?

  - Researchers conducted an experiment in which over 2000 participants watched a video of a crime and were tasked with selecting the culprit out of a line up
    - Factors: weapon, features on culprit, lineup procedure

  - We will focus on two main questions today:
    - If the true culprit is not in the lineup, does the presence of a weapon affect the odds of selecting a foil?
    - Does the procedure used to show the lineup affect the odds of selecting the true culprit?
Ex: Eye Witness Identification

- For now, we want to focus on whether or not the presence of a weapon affects the odds of selecting the foil (when the true culprit isn't in the lineup)
  - The foil is a person who looks very similar to the true culprit

- We want to determine the impact of the Weapon Focus Effect

- All participants in this portion of the data saw a simultaneous lineup, i.e. all 6 potential suspects were shown at one time
Ex: Eye Witness Identification

```r
# create two-way table
eye.wit <- frame_matrix(
  ~FoilID, ~Other, ~Total,
  78, 34, 78+34,
  58, 42, 58+42,
  78+58, 34+42, 78+34+58+42
)
rownames(eye.wit) <- c("No Weapon", "Weapon", "Total")
kable(eye.wit, format="html")
```

<table>
<thead>
<tr>
<th></th>
<th>FoilID</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Weapon</td>
<td>78</td>
<td>34</td>
<td>112</td>
</tr>
<tr>
<td>Weapon</td>
<td>58</td>
<td>42</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>136</td>
<td>76</td>
<td>212</td>
</tr>
</tbody>
</table>

**Other:** selected another candidate or selected none
Ex: Eye-Witness Identification

<table>
<thead>
<tr>
<th></th>
<th>FoilID</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Weapon</td>
<td>78</td>
<td>34</td>
<td>112</td>
</tr>
<tr>
<td>Weapon</td>
<td>58</td>
<td>42</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>136</td>
<td>76</td>
<td>212</td>
</tr>
</tbody>
</table>

Is there a significant difference in the proportion of Foil IDs based on whether or not a weapon was present during the crime?

1. Write the null and alternative hypotheses.

2. Calculate the test statistic.
Ex: Hypothesis Test

#Group 1: No Weapon, Group 2: Weapon
#calculate sample proportions
n1 <- eye.wit[1,3]
n2 <- eye.wit[2,3]
p1 <- eye.wit[1,1]/n1
p2 <- eye.wit[2,1]/n2
pc <- (eye.wit[1,1]+eye.wit[2,1])/(n1 + n2)
Ex: Hypothesis Test

#Group 1: No Weapon, Group 2: Weapon
#calculate sample proportions
n1 <- eye.wit[1,3]
n2 <- eye.wit[2,3]
p1 <- eye.wit[1,1]/n1
p2 <- eye.wit[2,1]/n2
pc <- (eye.wit[1,1]+eye.wit[2,1])/(n1 + n2)

#test statistic
(z <- (p1 - p2)/sqrt(pc*(1-pc)/n1 + pc*(1-pc)/n2))

## [1] 1.764657
Ex: Hypothesis Test

#Group 1: No Weapon, Group 2: Weapon
#calculate sample proportions
n1 <- eye.wit[1,3]
n2 <- eye.wit[2,3]
p1 <- eye.wit[1,1]/n1
p2 <- eye.wit[2,1]/n2
pc <- (eye.wit[1,1]+eye.wit[2,1])/(n1 + n2)

#test statistic
(z <- (p1 - p2)/sqrt(pc*(1-pc)/n1 + pc*(1-pc)/n2))

## [1] 1.764657

#p-value
2*(1-pnorm(abs(z)))

## [1] 0.07762146
Ex: 90% Confidence Interval

```r
#calculate SE
se <- sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)
```
Ex: 90% Confidence Interval

```r
#calculate SE
se <- sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)

#calculate critical value
z.star <- qnorm(0.95)
```
Ex: 90% Confidence Interval

```r
#calculate SE
se <- sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)

#calculate critical value
z.star <- qnorm(0.95)

#calculate CI
LB <- (p1 - p2) - z.star * se
UB <- (p1 - p2) + z.star * se
```
Ex: 90% Confidence Interval

```
#calculate SE
se <- sqrt(p1*(1-p1)/n1 + p2*(1-p2)/n2)

#calculate critical value
z.star <- qnorm(0.95)

#calculate CI
LB <- (p1 - p2) - z.star * se
UB <- (p1 - p2) + z.star * se
```

The 90% confidence interval for \((\pi_1 - \pi_2)\) is 0.008 to 0.225.
What are some potential difficulties with reporting results using the difference in proportions? or proportions/percentages in general?
Odds

- Given $\pi$, the population proportion of "yes" responses, the corresponding **odds** of a "yes" response is

  \[ \omega = \frac{\pi}{1 - \pi} \]

- The **sample odds** are $\hat{\omega} = \frac{\hat{\pi}}{1 - \hat{\pi}}$

- **Ex.**
  - sample proportion of **Foil IDs**: 0.642
  - odds of making a **Foil ID**: 1.789 to 1
Properties of the Odds

- Odds $\geq 0$

- If $\hat{p} = 0.5$, then Odds $= 1$

- If odds of "yes" are $\omega$, then $\frac{1}{\omega}$ are the odds for "no"

- If $\omega$ are the odds of "yes", the $\hat{\tau} = \frac{\omega}{(1+\omega)}$
Ratio of Two Odds

- Suppose we have two populations with proportions $\pi_1$ and $\pi_2$ and odds $\omega_1$ and $\omega_2$

- The **odds ratio** is $\phi = \frac{\omega_1}{\omega_2}$
  - *Estimate*: $\hat{\phi} = \frac{\hat{\omega}_1}{\hat{\omega}_2}$

- Good alternative to the difference in proportions

- **Interpretation**: The odds of "yes" in group 1 is $\phi$ times the odds of "yes" in group 2
Why use Odds Ratio?

- In practice, the odds ratio is more consistent across levels of confounding variables
- The odds ratio is more easily interpreted / understood
- The odds ratio can be easily extended to regression analysis
Sampling Distribution of Log Odds Ratio

- Let $\hat{\omega}_1$ and $\hat{\omega}_2$ be sample odds from independent random samples of size $n_1$ and $n_2$, respectively:

\[
\log(\hat{\phi}) = \log \left( \frac{\hat{\omega}_1}{\hat{\omega}_2} \right) \approx N \left( \log \left( \frac{\omega_1}{\omega_2} \right), \frac{1}{n_1 \pi_1 (1 - \pi_1)} + \frac{1}{n_2 \pi_2 (1 - \pi_2)} \right)
\]

... assuming $n_1$ and $n_2$ are "large" based on the thresholds for difference in proportions
Standard Errors

- For Confidence Intervals:

\[
SE[\log(\hat{\phi})] = SE\left[ \log \left( \frac{\hat{\omega}_1}{\hat{\omega}_2} \right) \right] = \sqrt{\frac{1}{n_1\hat{\pi}_1(1 - \hat{\pi}_1)} + \frac{1}{n_2\hat{\pi}_2(1 - \hat{\pi}_2)}}
\]
Standard Errors

- For Confidence Intervals:

\[ SE[\log(\hat{\phi})] = SE \left[ \log \left( \frac{\hat{\omega}_1}{\hat{\omega}_2} \right) \right] = \sqrt{\frac{1}{n_1 \hat{\pi}_1(1 - \hat{\pi}_1)} + \frac{1}{n_2 \hat{\pi}_2(1 - \hat{\pi}_2)}} \]

- For Hypothesis Tests:
  - The null hypothesis is that the proportions are equal
  - We estimate, \( \hat{\pi}_c \), the sample proportion from the combined data

\[ SE_0[\log(\hat{\phi})] = SE_0 \left[ \log \left( \frac{\hat{\omega}_1}{\hat{\omega}_2} \right) \right] = \sqrt{\frac{1}{n_1 \hat{\pi}_c(1 - \hat{\pi}_c)} + \frac{1}{n_2 \hat{\pi}_c(1 - \hat{\pi}_c)}} \]
Hypothesis Test for Odds Ratio

- We want to test whether two groups have equal odds, i.e.
  \[ \phi = \frac{\omega_1}{\omega_2} = 1 \]

- **Null Hypothesis**: \( H_0 : \log(\phi) = 0 \)

- **Test Statistic**:
  \[
  z = \frac{\log(\hat{\phi}) - 0}{\text{SE}_0[\log(\hat{\phi})]} = \frac{\log(\hat{\phi}) - 0}{\sqrt{\frac{1}{n_1\hat{\pi}_c(1-\hat{\pi}_c)} + \frac{1}{n_2\hat{\pi}_c(1-\hat{\pi}_c)}}}
  \]

- **p-value**: proportion of \( N(0, 1) \) distribution as extreme or more extreme than the test statistic
Confidence Interval for Log Odds Ratio

- Approximate $100(1 - \alpha)$% confidence interval for $\log(\phi)$ is

$$\log(\hat{\phi}) \pm z^* \times SE[\log(\hat{\phi})]$$

$$= \log(\hat{\phi}) \pm z^* \times \sqrt{\frac{1}{n_1\hat{\pi}_1(1 - \hat{\pi}_1)} + \frac{1}{n_2\hat{\pi}_2(1 - \hat{\pi}_2)}}$$
Confidence Interval for Log Odds Ratio

- Approximate $100(1 - \alpha)\%$ confidence interval for $\log(\phi)$ is

$$
\log(\hat{\phi}) \pm z^* \times SE[\log(\hat{\phi})]
= \log(\hat{\phi}) \pm z^* \times \sqrt{\frac{1}{n_1 \hat{\pi}_1 (1 - \hat{\pi}_1)} + \frac{1}{n_2 \hat{\pi}_2 (1 - \hat{\pi}_2)}}
$$

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution
Confidence Interval for Log Odds Ratio

- Approximate $100(1 - \alpha)\%$ confidence interval for $\log(\phi)$ is

\[
\log(\hat{\phi}) \pm z^* \times SE[\log(\hat{\phi})] \\
= \log(\hat{\phi}) \pm z^* \times \sqrt{\frac{1}{n_1\hat{\pi}_1(1 - \hat{\pi}_1)} + \frac{1}{n_2\hat{\pi}_2(1 - \hat{\pi}_2)}}
\]

where $z^*$ is the critical value calculated from the $N(0, 1)$ distribution

```r
# Critical value for 95% CI
qnorm(0.975)
```

```r
## [1] 1.959964
```
Confidence Interval for Odds Ratio

- Suppose \( LB \) and \( UB \) are the lower and upper bounds of the 100(1 − \( \alpha \))% confidence interval for \( \log(\phi) \), the log odds ratio.

- The 100(1 − \( \alpha \))% confidence interval for \( \phi \), the odds ratio is

\[
\exp\{LB\} \text{ to } \exp\{UB\}
\]
Ex. Eye Witness Identification

<table>
<thead>
<tr>
<th></th>
<th>FoilID</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Weapon</td>
<td>78</td>
<td>34</td>
<td>112</td>
</tr>
<tr>
<td>Weapon</td>
<td>58</td>
<td>42</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>136</td>
<td>76</td>
<td>212</td>
</tr>
</tbody>
</table>

Do the odds of making a Foil ID differ based on whether or not there was a weapon?
Ex. Eye Witness Identification

#Group 1: No Weapon, Group 2: Weapon

\[
\begin{align*}
\text{omegal} & \leftarrow p_1 / (1-p_1) \\
\text{omega2} & \leftarrow p_2/(1-p_2) \\
\phi & \leftarrow \text{omegal}/\text{omega2}
\end{align*}
\]

- No Weapon: Odds = 0.696
- Weapon: Odds = 0.58
- Odds Ratio = 1.661
Ex. Eye Witness Identification

```r
# Group 1: No Weapon, Group 2: Weapon
omega1 <- p1 / (1-p1)
omega2 <- p2/(1-p2)
phi <- omega1/omega2
```

- No Weapon: Odds = 0.696
- Weapon: Odds = 0.58
- Odds Ratio = 1.661

Is the odds ratio significantly different from 1?
Ex. Eye Witness Identification

\[ H_0 : \log(\phi) = 0 \]
\[ H_a : \log(\phi) \neq 0 \]

```r
# standard error
se.0 <- sqrt((n1*pc*(1-pc))^(-1) + (n2*pc*(1-pc))^(-1))

# test statistic
(z = log(phi)/se.0)

## [1] 1.769219

# p-value
2*(1-pnorm(abs(z)))

## [1] 0.07685732
```
Ex: 90% Confidence Interval

```r
#calculate se
se <- sqrt((n1*p1*(1-p1))^(-1) + (n2*p2*(1-p2))^(-1))
```
Ex: 90% Confidence Interval

```r
# calculate se
se <- sqrt((n1*p1*(1-p1))^(-1) + (n2*p2*(1-p2))^(-1))

# calculate critical value
z.star <- qnorm(0.95)
```
Ex: 90% Confidence Interval

# calculate se
se <- sqrt((n1*p1*(1-p1))^(-1) + (n2*p2*(1-p2))^(-1))

# calculate critical value
z.star <- qnorm(0.95)

# calculate CI for log odds ratio
LB <- log(phi) - z.star * se
UB <- log(phi) + z.star * se
Ex: 90% Confidence Interval

```r
#calculate se
se <- sqrt((n1*p1*(1-p1))^(1) + (n2*p2*(1-p2))^(1))

# calculate critical value
z.star <- qnorm(0.95)

# calculate CI for log odds ratio
LB <- log(phi) - z.star * se
UB <- log(phi) + z.star * se
```

The 90% confidence interval for log(\(\phi\)) is 0.033 to 0.982.
Ex: 90% Confidence Interval

```r
# calculate se
se <- sqrt((n1*p1*(1-p1))^(-1) + (n2*p2*(1-p2))^(-1))

# calculate critical value
z.star <- qnorm(0.95)

# calculate CI for log odds ratio
LB <- log(phi) - z.star * se
UB <- log(phi) + z.star * se
```

The 90% confidence interval for $\log(\phi)$ is 0.033 to 0.982.

The 90% confidence interval for $\phi$ is 1.033 to 2.67
Practice

- When the true culprit is present, does the line up procedure effect the odds of a correct ID?

- The participants in this subset watched a crime video in which there was no weapon

- Two line up procedures:
  - **Simultaneous**: All potential suspects shown at the same time
  - **Sequential**: The potential suspects are shown one at a time

- In the sequential lineup, the true culprit was #5
Practice

- **Goal:** Determine if there is a significant difference in the odds of a correct ID when the lineup is simultaneous vs. when it is sequential

- **Simultaenous:** 40 out of 109 made correct ID

- **Sequential:** 28 out 114 made a correct ID
Practice

- **Goal:** Determine if there is a significant difference in the odds of a correct ID when the lineup is simultaneous vs. when it is sequential.

- **Simultaneous:** 40 out of 109 made correct ID

- **Sequential:** 28 out 114 made a correct ID

1. Calculate the odds ratio.

2. Use the odds ratio to conduct a hypothesis test to determine if the odds significantly differ between the two groups.

3. Calculate a 95% confidence interval for the odds ratio.
Reference