Odds ratios for covariates effects

567 Statistical analysis of social networks

Peter Hoff

Statistics, University of Washington
Statistics for covariate effects

**Descriptive network analysis**: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level **covariate** information.

- Covariate: Node characteristics that “co-vary” with the network.

**Questions:**

- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

**Descriptive network analysis**: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level **covariate** information.

- Covariate: Node characteristics that “co-vary” with the network.

**Questions**:  
- How to describe the relationship between the network and covariates?  
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of
- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.
- Covariate: Node characteristics that “co-vary” with the network.

Questions:
- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.

- Covariate: Node characteristics that “co-vary” with the network.

Questions:

- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.

- Covariate: Node characteristics that “co-vary” with the network.

Questions:

- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of
- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.
- Covariate: Node characteristics that “co-vary” with the network.

Questions:
- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.

- Covariate: Node characteristics that “co-vary” with the network.

Questions:

- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Statistics for covariate effects

Descriptive network analysis: Computation of
- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.
- Covariate: Node characteristics that “co-vary” with the network.

Questions:
- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Descriptive network analysis: Computation of

- graph level statistics: density, degree distribution, centralization
- node level statistics: degrees, centralities

Often we also have node-level covariate information.

- Covariate: Node characteristics that “co-vary” with the network.

Questions:

- How to describe the relationship between the network and covariates?
- Can the covariates explain/predict network behavior?
Example: Girls friendships

mean(Y, na.rm = TRUE)

## [1] 0.04088967

Ce(1*(Y+t(Y) > 0))

## [1] 0.3820283
Covariate effects

We also have data on GPA

- \( \text{hgpa} = \text{indicator of above-average gpa} \);

```r
mean( Y[ hgpa==1, hgpa==1] , na.rm=TRUE)
## [1] 0.04737443

mean( Y[ hgpa==1, hgpa==0] , na.rm=TRUE)
## [1] 0.037623

mean( Y[ hgpa==0, hgpa==1] , na.rm=TRUE)
## [1] 0.03935944

mean( Y[ hgpa==0, hgpa==0] , na.rm=TRUE)
## [1] 0.03903421
```
Covariate effects

We also have data on GPA

- $\text{hgpa} =$ indicator of above-average gpa;

```r
mean( Y[ hgpaz==1, hgpaz==1] ,na.rm=TRUE)
## [1] 0.04737443

mean( Y[ hgpaz==1, hgpaz==0] ,na.rm=TRUE)
## [1] 0.037623

mean( Y[ hgpaz==0, hgpaz==1] ,na.rm=TRUE)
## [1] 0.03935944

mean( Y[ hgpaz==0, hgpaz==0] ,na.rm=TRUE)
## [1] 0.03903421
```
Covariate effects

We also have data on GPA

- \( \text{hgpa} = \) indicator of above-average gpa;

```
mean( Y[ hgpa==1, hgpa==1] ,na.rm=TRUE)
## [1] 0.04737443
mean( Y[ hgpa==1, hgpa==0] ,na.rm=TRUE)
## [1] 0.037623
mean( Y[ hgpa==0, hgpa==1] ,na.rm=TRUE)
## [1] 0.03935944
mean( Y[ hgpa==0, hgpa==0] ,na.rm=TRUE)
## [1] 0.03903421
```
We also have data on smoking behavior:

- \texttt{hsmoke} = indicator of above-average smoking behavior.

\begin{verbatim}
mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04477612

mean( Y[ hsmoke==1, hsmoke==0] ,na.rm=TRUE)
## [1] 0.03062609

mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04400078

mean( Y[ hsmoke==0, hsmoke==0] ,na.rm=TRUE)
## [1] 0.04425837
\end{verbatim}
Covariate effects

We also have data on smoking behavior:

- **hsmoke** = indicator of above-average smoking behavior.

```r
mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04477612

mean( Y[ hsmoke==1, hsmoke==0] ,na.rm=TRUE)
## [1] 0.03062609

mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04400078

mean( Y[ hsmoke==0, hsmoke==0] ,na.rm=TRUE)
## [1] 0.04425837
```
Covariate effects

We also have data on smoking behavior:

- hsmoke = indicator of above-average smoking behavior.

```r
mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04477612
mean( Y[ hsmoke==1, hsmoke==0] ,na.rm=TRUE)
## [1] 0.03062609
mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE)
## [1] 0.04400078
mean( Y[ hsmoke==0, hsmoke==0] ,na.rm=TRUE)
## [1] 0.04425837
```
Summarizing densities of subgraphs

There are a lot of probabilities here (four for each covariate)

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>0.039</td>
<td>0.039</td>
</tr>
<tr>
<td>xi=1</td>
<td>0.038</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Table: gpa

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>xi=1</td>
<td>0.031</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Table: smoking

Note: Such tables correspond to very rudimentary “blockmodels”:
- an observed categorical covariate divides nodes into “blocks”;
- probability of tie between nodes determined by rates between their blocks.

Interpreting probabilities/rates:
How do rates correspond to nodal preferences?
Summarizing densities of subgraphs

There are a lot of probabilities here (four for each covariate)

\[
\begin{array}{cc|cc}
& x_j=0 & x_j=1 \\
\hline
x_i=0 & 0.039 & 0.039 \\
x_i=1 & 0.038 & 0.047 \\
\end{array}
\]

\text{Table : gpa}

\[
\begin{array}{cc|cc}
& x_j=0 & x_j=1 \\
\hline
x_i=0 & 0.044 & 0.044 \\
x_i=1 & 0.031 & 0.045 \\
\end{array}
\]

\text{Table : smoking}

\textbf{Note:} Such tables correspond to very rudimentary “blockmodels”:
- an observed categorical covariate divides nodes into “blocks”;
- probability of tie between nodes determined by rates between their blocks.

\textbf{Interpreting probabilities/rates:}
How do rates correspond to nodal preferences?
Summarizing densities of subgraphs

There are a lot of probabilities here (four for each covariate)

<table>
<thead>
<tr>
<th>xi=0</th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.039</td>
<td>0.039</td>
<td></td>
</tr>
<tr>
<td>0.038</td>
<td>0.047</td>
<td></td>
</tr>
</tbody>
</table>

Table: gpa

<table>
<thead>
<tr>
<th>xi=0</th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.044</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>0.031</td>
<td>0.045</td>
<td></td>
</tr>
</tbody>
</table>

Table: smoking

Note: Such tables correspond to very rudimentary “blockmodels”:
- an observed categorical covariate divides nodes into “blocks”;
- probability of tie between nodes determined by rates between their blocks.

Interpreting probabilities/rates:
How do rates correspond to nodal preferences?
**Odds ratios**

**Odds**: Let $Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{Pr(E)}{1 - Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The “effect” of a variable on a probability is often described via the odds ratio.

**Odds ratio**: Let

- $Pr(E|A)$ = the probability of some event $E$ under condition $A$
- $Pr(E|B)$ = the probability of some event $E$ under condition $B$

The odds ratio is

$$\text{odds}(E : A, B) = \frac{Pr(E|A)}{1 - Pr(E|A)} \frac{1 - Pr(E|B)}{Pr(E|B)}$$

Note that

$$\text{odds}(E : A, B) = 1 \Rightarrow Pr(E|A) = Pr(E|B)$$
Odds ratios

**Odds**: Let $\Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The “effect” of a variable on a probability is often described via the odds ratio.

**Odds ratio**: Let

- $\Pr(E|A) =$ the probability of some event $E$ under condition $A$
- $\Pr(E|B) =$ the probability of some event $E$ under condition $B$

The odds ratio is

$$\text{odds}(E:A,B) = \frac{\Pr(E|A)}{1 - \Pr(E|A)} \frac{1 - \Pr(E|B)}{\Pr(E|B)}$$

Note that

$$\text{odds}(E:A,B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)$$
Odds ratios

**Odds**: Let $\Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The “effect” of a variable on a probability is often described via the odds ratio.

**Odds ratio**: Let

- $\Pr(E|A)$ = the probability of some event $E$ under condition $A$
- $\Pr(E|B)$ = the probability of some event $E$ under condition $B$

The odds ratio is

$$\text{odds}(E : A, B) = \frac{\Pr(E|A) \cdot 1 - \Pr(E|B)}{1 - \Pr(E|A) \cdot \Pr(E|B)}$$

Note that

$$\text{odds}(E : A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)$$
Odds ratios

**Odds:** Let $\Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The “effect” of a variable on a probability is often described via the odds ratio.

**Odds ratio:** Let

- $\Pr(E|A) = \text{the probability of some event } E \text{ under condition } A$
- $\Pr(E|B) = \text{the probability of some event } E \text{ under condition } B$

The odds ratio is

$$\text{odds}(E: A, B) = \frac{\Pr(E|A)}{1 - \Pr(E|A)} \frac{1 - \Pr(E|B)}{\Pr(E|B)}$$

Note that

$$\text{odds}(E: A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)$$
Odds ratios

**Odds**: Let $\Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The “effect” of a variable on a probability is often described via the odds ratio.

**Odds ratio**: Let

- $\Pr(E|A)$ = the probability of some event $E$ under condition $A$
- $\Pr(E|B)$ = the probability of some event $E$ under condition $B$

The odds ratio is

$$\text{odds}(E : A, B) = \frac{\frac{\Pr(E|A)}{1 - \Pr(E|A)}}{\frac{1 - \Pr(E|B)}{\Pr(E|B)}}$$

Note that

$$\text{odds}(E : A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)$$
Odds ratios

**Odds**: Let $\Pr(E)$ be the probability of an event. The odds of $E$ are

$$\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}$$

Probabilities are between 0 and 1, odds are between 0 and $\infty$.

The "effect" of a variable on a probability is often described via the odds ratio.

**Odds ratio**: Let

- $\Pr(E|A) = \text{the probability of some event } E \text{ under condition } A$
- $\Pr(E|B) = \text{the probability of some event } E \text{ under condition } B$

The odds ratio is

$$\text{odds}(E : A, B) = \frac{\Pr(E|A)}{1 - \Pr(E|A)} \cdot \frac{1 - \Pr(E|B)}{\Pr(E|B)}$$

Note that

$$\text{odds}(E : A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)$$
Odds: Let \( \Pr(E) \) be the probability of an event. The odds of \( E \) are

\[
\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}
\]

Probabilities are between 0 and 1, odds are between 0 and \( \infty \).

The “effect” of a variable on a probability is often described via the odds ratio.

Odds ratio: Let

- \( \Pr(E|A) = \) the probability of some event \( E \) under condition \( A \)
- \( \Pr(E|B) = \) the probability of some event \( E \) under condition \( B \)

The odds ratio is

\[
\text{odds}(E : A, B) = \frac{\Pr(E|A)}{1 - \Pr(E|A)} \frac{1 - \Pr(E|B)}{\Pr(E|B)}
\]

Note that

\[
\text{odds}(E : A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)
\]
Odds ratios

Odds: Let \( \Pr(E) \) be the probability of an event. The odds of \( E \) are

\[
\text{odds}(E) = \frac{\Pr(E)}{1 - \Pr(E)}
\]

Probabilities are between 0 and 1, odds are between 0 and \( \infty \).

The “effect” of a variable on a probability is often described via the odds ratio.

Odds ratio: Let

- \( \Pr(E|A) = \) the probability of some event \( E \) under condition \( A \)
- \( \Pr(E|B) = \) the probability of some event \( E \) under condition \( B \)

The odds ratio is

\[
\text{odds}(E : A, B) = \frac{\Pr(E|A)}{1 - \Pr(E|A)} \cdot \frac{1 - \Pr(E|B)}{\Pr(E|B)}
\]

Note that

\[
\text{odds}(E : A, B) = 1 \Rightarrow \Pr(E|A) = \Pr(E|B)
\]
Effect of a covariate on a tie

Let $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$ be a binary variable.

- $x_i$ = indicator of high gpa, or
- $x_i$ = smoking status, or
- $x_i$ = indicator of membership to some group.

Let $\Pr(y_{i,j} = 1|x_i, x_j) = p_{xixj}$

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>xi=1</td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Given a network, we might want to describe the “effect” of $x_i$ and $x_j$ on $y_{i,j}$:

$$\text{odds}(y_{i,j} = 1|\{x_i = 1, x_j = 1\}, \{x_i = 0, x_j = 1\}) = \frac{p_{11}}{1 - p_{11}} \cdot \frac{1 - p_{01}}{p_{01}}$$
Effect of a covariate on a tie

Let $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$ be a binary variable.

- $x_i = \text{indicator of high gpa}$, or
- $x_i = \text{smoking status}$, or
- $x_i = \text{indicator of membership to some group}$.

Let $\Pr(y_{i,j} = 1|x_i, x_j) = p_{x_ix_j}$

\[
\begin{array}{c|cc}
   & x_{j=0} & x_{j=1} \\
\hline
   x_{i=0} & p_{00} & p_{01} \\
   x_{i=1} & p_{10} & p_{11} \\
\end{array}
\]

Given a network, we might want to describe the “effect” of $x_i$ and $x_j$ on $y_{i,j}$:

\[
\text{odds}(y_{i,j} = 1|\{x_i = 1, x_j = 1\}, \{x_i = 0, x_j = 1\}) = \frac{p_{11}}{1 - p_{11}} \frac{1 - p_{01}}{p_{01}}
\]
Effect of a covariate on a tie

Let $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$ be a binary variable.

- $x_i$ = indicator of high gpa, or
- $x_i$ = smoking status, or
- $x_i$ = indicator of membership to some group.

Let $Pr(y_{i,j} = 1|x_i, x_j) = p_{x_ix_j}$

<table>
<thead>
<tr>
<th></th>
<th>$x_j=0$</th>
<th>$x_j=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i=0$</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>$x_i=1$</td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Given a network, we might want to describe the “effect” of $x_i$ and $x_j$ on $y_{i,j}$:

$$\text{odds}(y_{i,j} = 1|\{x_i = 1, x_j = 1\}, \{x_i = 0, x_j = 1\}) = \frac{p_{11}}{1 - p_{11}} \frac{1 - p_{01}}{p_{01}}$$
Effect of a covariate on a tie

Let $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$ be a binary variable.

- $x_i =$ indicator of high gpa, or
- $x_i =$ smoking status, or
- $x_i =$ indicator of membership to some group.

Let $Pr(y_{i,j} = 1|x_i, x_j) = p_{x_i x_j}$

<table>
<thead>
<tr>
<th></th>
<th>$x_j=0$</th>
<th>$x_j=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i=0$</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>$x_i=1$</td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Given a network, we might want to describe the “effect” of $x_i$ and $x_j$ on $y_{i,j}$:

$$
\text{odds}(y_{i,j} = 1|\{x_i = 1, x_j = 1\}, \{x_i = 0, x_j = 1\}) = \frac{p_{11}}{1 - p_{11}} \frac{1 - p_{01}}{p_{01}}
$$
Effect of a covariate on a tie

Let $x_i \in \{0, 1\}$ for $i = 1, \ldots, n$ be a binary variable.

- $x_i$ = indicator of high gpa, or
- $x_i$ = smoking status, or
- $x_i$ = indicator of membership to some group.

Let $\Pr(y_{i,j} = 1|x_i, x_j) = p_{x_i x_j}$

<table>
<thead>
<tr>
<th></th>
<th>$x_j = 0$</th>
<th>$x_j = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i = 0$</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>$x_i = 1$</td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Given a network, we might want to describe the “effect” of $x_i$ and $x_j$ on $y_{i,j}$:

$$\text{odds}(y_{i,j} = 1|\{x_i = 1, x_j = 1\}, \{x_i = 0, x_j = 1\}) = \frac{p_{11}}{1 - p_{11}} \frac{1 - p_{01}}{p_{01}}$$
Odds ratios

\[
p_{11} = \text{mean}(Y \mid hsmoke == 1, hsmoke == 1, \text{na.rm} = \text{TRUE})
\]

\[
p_{01} = \text{mean}(Y \mid hsmoke == 0, hsmoke == 1, \text{na.rm} = \text{TRUE})
\]

\[
\frac{p_{11}/(1-p_{11})}{p_{01}/(1-p_{01})}
\]

## [1] 1.018447

This result says that the odds of a tie are 1.02 times higher under the condition \( x_i = 1, x_j = 1 \) than \( x_i = 0, x_j = 1 \).

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking or
- differential rates of ties among smokers and nonsmokers.
Odds ratios

\[
p_{11} <- \text{mean}( Y[ \text{hsmoke==1, hsmoke==1} ] , \text{na.rm=TRUE})\\
p_{01} <- \text{mean}( Y[ \text{hsmoke==0, hsmoke==1} ] , \text{na.rm=TRUE})\\
\frac{(p_{11}/(1-p_{11}))}{(p_{01}/(1-p_{01}))}\\
\]

## [1] 1.018447

This result says that the odds of a tie are 1.02 times higher under the condition \( x_i = 1, x_j = 1 \) than \( x_i = 0, x_j = 1 \).

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking or
- differential rates of ties among smokers and nonsmokers.
Odds ratios

\[
\frac{\frac{p_{11}}{1-p_{11}}}{\frac{p_{01}}{1-p_{01}}}
\]

\# [1] 1.018447

This result says that the odds of a tie are 1.02 times higher under the condition \( x_i = 1, x_j = 1 \) than \( x_i = 0, x_j = 1 \).

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking or
- differential rates of ties among smokers and nonsmokers.
Odds ratios

```
p11<-mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE)
p01<-mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE)

(p11/(1-p11)) / (p01/(1-p01))
```

## [1] 1.018447

This result says that the odds of a tie are 1.02 times higher under the condition $x_i = 1, x_j = 1$ than $x_i = 0, x_j = 1$.

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking or
- differential rates of ties among smokers and nonsmokers.
This result says that the odds of a tie are 1.02 times higher under the condition $x_i = 1, x_j = 1$ than $x_i = 0, x_j = 1$.

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking or
- differential rates of ties among smokers and nonsmokers.
Odds ratios

This result says that the odds of a tie are 1.02 times higher under the condition $x_i = 1, x_j = 1$ than $x_i = 0, x_j = 1$.

This result seems to suggest that smokers and non-smokers are equally friendly to smokers. However, the result could be due to

- no effect of smoking
- differential rates of ties among smokers and nonsmokers.
A better question to ask might be:

Does a person’s characteristic determine the characteristics of whom they choose as friends?

The probabilities related to this question condition on the existence of a tie:

\[
\Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1 | x_i = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)}
\]

\[
= \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)}
\]

\[
= p_{11} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)}
\]

This probability can be interpreted as, for example,

What is the probability that a friend of a smoker is another smoker?

Such a probability is more descriptive of tie preferences.
A better question to ask might be:

*Does a person’s characteristic determine the characteristics of whom they choose as friends?*

The probabilities related to this question condition on the existence of a tie:

\[
\Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1 | x_i = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)} = \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)} = p_{11} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)}
\]

This probability can be interpreted as, for example,

*What is the probability that a friend of a smoker is another smoker?*

Such a probability is more descriptive of tie preferences.
A better question to ask might be:

*Does a person’s characteristic determine the characteristics of whom they choose as friends?*

The probabilities related to this question condition on the existence of a tie:

\[
\Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1 | x_i = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)} = \frac{\Pr(y_{i,j} = 1 | x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)} = p_{11} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1 | x_i = 1)}
\]

This probability can be interpreted as, for example,

*What is the probability that a friend of a smoker is another smoker?*

Such a probability is more descriptive of tie preferences.
Odds ratios for tie preferences

\[ \Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) = \text{probability that a friend of a smoker is another smoker} \]

However, this probability will mostly reflect the (typically low) overall tie density.

To assess the “effect” of \( x_i \) on choosing another smoker as a friend, we can look at an appropriate odds ratio:

\[
\text{odds}(x_j = 1 : \{y_{i,j} = 1, x_i = 1\}, \{y_{i,j} = 1, x_i = 0\}) = \frac{\Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) \Pr(x_j = 0 | y_{i,j} = 1, x_i = 0)}{\Pr(x_j = 0 | y_{i,j} = 1, x_i = 1) \Pr(x_j = 1 | y_{i,j} = 1, x_i = 0)}
\]
Odds ratios for tie preferences

Pr($x_j = 1 | y_{i,j} = 1, x_i = 1$) = probability that a friend of a smoker is another smoker

However, this probability will mostly reflect the (typically low) overall tie density.

To assess the “effect” of $x_i$ on choosing another smoker as a friend, we can look at an appropriate odds ratio:

$$\text{odds}(x_j = 1 : \{y_{i,j} = 1, x_i = 1\}, \{y_{i,j} = 1, x_i = 0\}) =$$
$$\frac{\Pr(x_j = 1 | y_{i,j} = 1, x_i = 1) \Pr(x_j = 0 | y_{i,j} = 1, x_i = 0)}{\Pr(x_j = 0 | y_{i,j} = 1, x_i = 1) \Pr(x_j = 1 | y_{i,j} = 1, x_i = 0)}$$
Odds ratios for tie preferences

Recall,

\[
\Pr(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1|x_i = 1)}{\Pr(y_{i,j} = 1|x_i = 1)} = \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)} = p_{11} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

Similarly,

\[
\Pr(x_j = 0|y_{i,j} = 1, x_i = 1) = p_{10} \frac{\Pr(x_j = 0)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

and so

\[
\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)/\Pr(y_{i,j} = 1|x_i = 1)}{\Pr(x_j = 0)/\Pr(y_{i,j} = 1|x_i = 1)} = \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)}.
\]
Odds ratios for tie preferences

Recall,

\[
\Pr(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1|x_i = 1)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

\[
= \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

\[
= \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

Similarly,

\[
\Pr(x_j = 0|y_{i,j} = 1, x_i = 1) = p_{10} \frac{\Pr(x_j = 0)}{\Pr(y_{i,j} = 1|x_i = 1)}
\]

and so

\[
\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{\frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)}}{\frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 0)}{\Pr(y_{i,j} = 1|x_i = 1)}}
\]

\[
= \frac{p_{11} \Pr(x_j = 1)}{p_{10} \Pr(x_j = 0)}.
\]
Odds ratios for tie preferences

Recall,

\[ \Pr(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1|x_i = 1)}{\Pr(y_{i,j} = 1|x_i = 1)} \]

\[ = \frac{\Pr(y_{i,j} = 1|x_j = 1, x_i = 1) \Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)} \]

\[ = p_{11} \frac{\Pr(x_j = 1)}{\Pr(y_{i,j} = 1|x_i = 1)} \]

Similarly,

\[ \Pr(x_j = 0|y_{i,j} = 1, x_i = 1) = p_{10} \frac{\Pr(x_j = 0)}{\Pr(y_{i,j} = 1|x_i = 1)} \]

and so

\[ \text{odds}(x_j = 1|y_{i,j} = 1, x_i = 1) = \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)/\Pr(y_{i,j} = 1|x_i = 1)}{\Pr(x_j = 0)/\Pr(y_{i,j} = 1|x_i = 1)} \]

\[ = \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)}. \]
Odds ratios for tie preferences

\[ p_{x_1 x_2} = \Pr(\text{tie } | x_1, x_2) \]
\[ \approx \text{density in the } x_1, x_2 \text{ submatrix} \]

<table>
<thead>
<tr>
<th></th>
<th>x_j=0</th>
<th>x_j=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_i=0</td>
<td>p_{00}</td>
<td>p_{01}</td>
</tr>
<tr>
<td>x_i=1</td>
<td>p_{10}</td>
<td>p_{11}</td>
</tr>
</tbody>
</table>

\[
\text{odds}(x_j = 1 | y_{i,j} = 1, x_i = 1) = \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)}
\]
\[
\text{odds}(x_j = 1 | y_{i,j} = 1, x_i = 0) = \frac{p_{01}}{p_{00}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)}
\]

The odds ratio is therefore

\[
\text{odds ratio}(x_j = 1 | \{y_{i,j} = 1, x_i = 1\} \{y_{i,j} = 1, x_i = 0\}) = \frac{p_{11}/p_{10}}{p_{01}/p_{00}} = \frac{p_{11}p_{00}}{p_{10}p_{01}}
\]
Odds ratios for tie preferences

\[ p_{x_1x_2} = \Pr(\text{tie} \mid x_1, x_2) \]
\[ \approx \text{density in the } x_1, x_2 \text{ submatrix} \]

\[
\begin{array}{c|cc}
  & x_j=0 & x_j=1 \\
  \hline
  x_i=0 & p_{00} & p_{01} \\
  x_i=1 & p_{10} & p_{11} \\
\end{array}
\]

odds \( x_j = 1 \mid y_{i,j} = 1, x_i = 1 \) = \( \frac{p_{11}}{p_{10}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)} \)

odds \( x_j = 1 \mid y_{i,j} = 1, x_i = 0 \) = \( \frac{p_{01}}{p_{00}} \frac{\Pr(x_j = 1)}{\Pr(x_j = 0)} \)

The odds ratio is therefore

\[
odds \text{ ratio}(x_j = 1 \mid \{y_{i,j} = 1, x_i = 1\} \{y_{i,j} = 1, x_i = 0\}) = \frac{p_{11}/p_{10}}{p_{01}/p_{00}} = \frac{p_{11}p_{00}}{p_{10}p_{01}}
\]
Odds ratios for tie preferences

\[ p_{x_1 x_2} = \Pr( \text{tie} | x_1, x_2) \]
\[ \approx \text{density in the } x_1, x_2 \text{ submatrix} \]

<table>
<thead>
<tr>
<th></th>
<th>(x_j=0)</th>
<th>(x_j=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_i=0)</td>
<td>(p_{00})</td>
<td>(p_{01})</td>
</tr>
<tr>
<td>(x_i=1)</td>
<td>(p_{10})</td>
<td>(p_{11})</td>
</tr>
</tbody>
</table>

\[
\text{odds}(x_j = 1| y_{i,j} = 1, x_i = 1) = \frac{p_{11} \Pr(x_j = 1)}{p_{10} \Pr(x_j = 0)}
\]
\[
\text{odds}(x_j = 1| y_{i,j} = 1, x_i = 0) = \frac{p_{01} \Pr(x_j = 1)}{p_{00} \Pr(x_j = 0)}
\]

The odds ratio is therefore

\[
\text{odds ratio}(x_j = 1| \{y_{i,j} = 1, x_i = 1\}\{y_{i,j} = 1, x_i = 0\}) = \frac{p_{11}/p_{10}}{p_{01}/p_{00}} = \frac{p_{11}p_{00}}{p_{10}p_{01}}
\]
Odds ratios for tie preferences

$$\gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}}$$

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

$$\text{odds ratio}(x_i = 1 | \{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.$$ 

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)

to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)

to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}|\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[ \text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma. \]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11} p_{00}}{p_{10} p_{01}} \]

This ratio represents
the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1 | \{y_{i,j} = 1, x_j = 1\} \{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11} p_{00}}{p_{10} p_{01}} = \gamma.
\]

This ratio represents
the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[ \gamma = \frac{p_{11}p_{00}}{p_{10}p_{01}} \]

This ratio represents the relative preference of egos with \( x = 1 \) versus \( x = 0 \) to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1\mid \{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \) to egos with \( x = 1 \).
Odds ratios for tie preferences

\[
\gamma = \frac{p_{11}p_{00}}{p_{01}p_{10}}
\]

This ratio represents

the relative preference of egos with \( x = 1 \) versus \( x = 0 \)
to tie to alters with \( x = 1 \).

Interestingly, one can show (homework?)

\[
\text{odds ratio}(x_i = 1|\{y_{i,j} = 1, x_j = 1\}\{y_{i,j} = 1, x_j = 0\}) = \frac{p_{11}p_{00}}{p_{10}p_{01}} = \gamma.
\]

This ratio represents

the relative attractiveness of alters with \( x = 1 \) versus \( x = 0 \)
to egos with \( x = 1 \).
Odds ratios for tie preferences

\[
\begin{array}{c|cc}
  & x_j=0 & x_j=1 \\
\hline
x_i=0 & p_{00} & p_{01} \\
x_i=1 & p_{10} & p_{11} \\
\end{array}
\]

Are there interesting/useful ways to represent numbers in the table?

- In SNA, more interested in relative rates than absolute rates.
- Absolute rates are derivable from relative rates and a baseline, and vice-versa:

\[
\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, p_{11}/p_{00}\} \\
\sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}
\]
Odds ratios for tie preferences

<table>
<thead>
<tr>
<th></th>
<th>x_j=0</th>
<th>x_j=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_i=0</td>
<td>p_{00}</td>
<td>p_{01}</td>
</tr>
<tr>
<td>x_i=1</td>
<td>p_{10}</td>
<td>p_{11}</td>
</tr>
</tbody>
</table>

Are there interesting/useful ways to represent numbers in the table?

- In SNA, more interested in relative rates than absolute rates.
- Absolute rates are derivable from relative rates and a baseline, and vice-versa:

\[
\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, \frac{p_{01}}{p_{00}}, \frac{p_{10}}{p_{00}}, \frac{p_{11}}{p_{00}}\} \\
\sim \{p_{00}, \frac{p_{01}}{p_{00}}, \frac{p_{10}}{p_{00}}, \frac{(p_{11}p_{00})}{(p_{01}p_{10})}\}\
\]
Odds ratios for tie preferences

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
</tr>
<tr>
<td>xi=1</td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Are there interesting/useful ways to represent numbers in the table?

- In SNA, more interested in relative rates than absolute rates.
- Absolute rates are derivable from relative rates and a baseline, and vice-versa:

$$\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, p_{11}/p_{00}\}$$
$$\sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}$$
Odds ratios for tie preferences

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>(p_{00})</td>
<td>(p_{01})</td>
</tr>
<tr>
<td>xi=1</td>
<td>(p_{10})</td>
<td>(p_{11})</td>
</tr>
</tbody>
</table>

Are there interesting/useful ways to represent numbers in the table?

- In SNA, more interested in relative rates than absolute rates.
- Absolute rates are derivable from relative rates and a baseline, and vice-versa:

\[
\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, p_{11}/p_{00}\}
\]

\[
\sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}
\]
### Odd ratios for tie preferences

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>p_{00}</td>
<td>p_{01}</td>
</tr>
<tr>
<td>xi=1</td>
<td>p_{10}</td>
<td>p_{11}</td>
</tr>
</tbody>
</table>

Are there interesting/useful ways to represent numbers in the table?

- In SNA, more interested in relative rates than absolute rates.
- Absolute rates are derivable from relative rates and a baseline, and vice-versa:

\[
\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, \frac{p_{01}}{p_{00}}, \frac{p_{10}}{p_{00}}, \frac{p_{11}}{p_{00}}\} \\
\sim \{p_{00}, \frac{p_{01}}{p_{00}}, \frac{p_{10}}{p_{00}}, \frac{(p_{11}p_{00})}{(p_{01}p_{10})}\} 
\]
{p_{00}, p_{01}, p_{10}, p_{11}} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}

**Baseline:** $p_{00}$ represents a baseline rate

**Relative rates:** $p_{01}/p_{00}, p_{10}/p_{00}$ represent relative rates
- $p_{01}/p_{00} =$ density of 0 → 1 relative to 0 → 0 (“attractiveness of 1’s”)
- $p_{10}/p_{00} =$ density of 1 → 0 relative to 0 → 0 (“sociability of 1’s”)

**Odds ratio:** $(p_{11}p_{00})/(p_{01}p_{10}) = \gamma$ represents preferences for homophily.
Interpreting probability ratios

\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}

**Baseline:** \(p_{00}\) represents a baseline rate

**Relative rates:** \(p_{01}/p_{00}, p_{10}/p_{00}\) represent relative rates
- \(p_{01}/p_{00}\) = density of \(0 \rightarrow 1\) relative to \(0 \rightarrow 0\) ("attractiveness of 1's")
- \(p_{10}/p_{00}\) = density of \(1 \rightarrow 0\) relative to \(0 \rightarrow 0\) ("sociability of 1's")

**Odds ratio:** \((p_{11}p_{00})/(p_{01}p_{10}) = \gamma\) represents preferences for homophily.
Interpreting probability ratios

\[ \{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\} \]

**Baseline:**  \(p_{00}\) represents a baseline rate

**Relative rates:**  \(p_{01}/p_{00}, p_{10}/p_{00}\) represent relative rates

- \(p_{01}/p_{00}\) = density of 0 → 1 relative to 0 → 0 ("attractiveness of 1’s")
- \(p_{10}/p_{00}\) = density of 1 → 0 relative to 0 → 0 ("sociability of 1’s")

**Odds ratio:** \((p_{11}p_{00})/(p_{01}p_{10}) = \gamma\) represents preferences for homophily.
Interpreting probability ratios

\{p_{00}, p_{01}, p_{10}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{10}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}

**Baseline:** \(p_{00}\) represents a baseline rate

**Relative rates:** \(p_{01}/p_{00}, p_{10}/p_{00}\) represent relative rates
- \(p_{01}/p_{00}\) = density of 0 → 1 relative to 0 → 0 (“attractiveness of 1’s”)
- \(p_{10}/p_{00}\) = density of 1 → 0 relative to 0 → 0 (“sociability of 1’s”)

**Odds ratio:** \((p_{11}p_{00})/(p_{01}p_{10}) = \gamma\) represents preferences for homophily.
Interpreting relative rates

You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \(x_j = 1\) change if you are told that \(j\) has a link from a person \(i\) with \(x_i = 0\).
You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \(x_j = 1\) change if you are told that \(j\) has a link from a person \(i\) with \(x_i = 0\).
Interpreting relative rates

You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \(x_j = 1\) change if you are told that \(j\) has a link from a person \(i\) with \(x_i = 0\).
Interpreting relative rates

You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \(x_j = 1\) change if you are told that \(j\) has a link from a person \(i\) with \(x_i = 0\).
Interpreting relative rates

You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1 | y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \( x_j = 1 \) change if you are told that \( j \) has a link from a person \( i \) with \( x_i = 0 \).
Interpreting relative rates

You can show (for example) that

\[ \frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1 | y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)} \]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \( x_j = 1 \) change if you are told that \( j \) has a link from a person \( i \) with \( x_i = 0 \).
Interpreting relative rates

You can show (for example) that

\[
\frac{p_{01}}{p_{00}} = \frac{\text{odds}(x_j = 1|y_{i,j} = 1, x_i = 0)}{\text{odds}(x_j = 1)}
\]

While this is a ratio of odds, it is not exactly an odds ratio:

- The conditioning events are not complementary.

The ratio still has a reasonable interpretation:

- The ratio can be interpreted as how much the odds of \( x_j = 1 \) change if you are told that \( j \) has a link from a person \( i \) with \( x_i = 0 \).
Friendship example

```
p.smoke<-c(
  mean( Y[ hsmoke==0, hsmoke==0] ,na.rm=TRUE),
  mean( Y[ hsmoke==1, hsmoke==0] ,na.rm=TRUE),
  mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE),
  mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE)  
)

)
## [1] 1.470585

p.gpa<-c(
  mean( Y[ hgpa==0, hgpa==0] ,na.rm=TRUE),
  mean( Y[ hgpa==1, hgpa==0] ,na.rm=TRUE),
  mean( Y[ hgpa==0, hgpa==1] ,na.rm=TRUE),
  mean( Y[ hgpa==1, hgpa==1] ,na.rm=TRUE)  
)

)
## [1] 1.248783
```

Homophily is positive for both smoking and gpa.
Friendship example

```r
p.smoke<-c(
  mean( Y[ hsmoke==0, hsmoke==0] ,na.rm=TRUE),
  mean( Y[ hsmoke==1, hsmoke==0] ,na.rm=TRUE),
  mean( Y[ hsmoke==0, hsmoke==1] ,na.rm=TRUE),
  mean( Y[ hsmoke==1, hsmoke==1] ,na.rm=TRUE) 
)


## [1] 1.470585

p.gpa<-c(
  mean( Y[ hgpa==0, hgpa==0] ,na.rm=TRUE),
  mean( Y[ hgpa==1, hgpa==0] ,na.rm=TRUE),
  mean( Y[ hgpa==0, hgpa==1] ,na.rm=TRUE),
  mean( Y[ hgpa==1, hgpa==1] ,na.rm=TRUE) 
)


## [1] 1.248783
```

Homophily is positive for both smoking and gpa.
Friendship example

p.smoke[1]
## [1] 0.04425837

## [1] 0.6919841

p.smoke[3]/p.smoke[1]
## [1] 0.9941797

## [1] 1.470585
Friendship example

- The baseline rate is low \((p_{00} = 0.04)\)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers \((p_{01}/p_{00} = 0.99)\).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers \((p_{10}/p_{00} = 0.69)\).
- There is strong homophily \((\gamma = 1.47)\)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low ($p_{00} = 0.04$)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers ($p_{01}/p_{00} = 0.99$).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers ($p_{10}/p_{00} = 0.69$).
- There is strong homophily ($\gamma = 1.47$)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker.
Friendship example

- The baseline rate is low ($p_{00} = 0.04$)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers ($p_{01}/p_{00} = .99$).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers ($p_{10}/p_{00} = .69$).
- There is strong homophily ($\gamma = 1.47$)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low \( (p_{00} = 0.04) \)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers \( (p_{01}/p_{00} = .99) \).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers \( (p_{10}/p_{00} = .69) \).
- There is strong homophily \( (\gamma = 1.47) \)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low ($p_{00} = 0.04$)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers ($p_{01}/p_{00} = .99$).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers ($p_{10}/p_{00} = .69$).
- There is strong homophily ($\gamma = 1.47$)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low ($p_{00} = 0.04$)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers ($p_{01}/p_{00} = .99$).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers ($p_{10}/p_{00} = .69$).
- There is strong homophily ($\gamma = 1.47$)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low ($p_{00} = 0.04$)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers ($p_{01}/p_{00} = 0.99$).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers ($p_{10}/p_{00} = 0.69$).
- There is strong homophily ($\gamma = 1.47$)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

- The baseline rate is low \( p_{00} = 0.04 \)
- The rate of ties from nonsmokers to smokers is about the same as that from nonsmokers to nonsmokers \( p_{01}/p_{00} = 0.99 \).
- The rate of ties from smokers to nonsmokers is much lower than that from nonsmokers to nonsmokers \( p_{10}/p_{00} = 0.69 \).
- There is strong homophily \( \gamma = 1.47 \)
  - A tie from a smoker is more likely to be to a smoker than a tie from a nonsmoker is.
  - A tie to a smoker is more likely to be from a smoker than a tie to a nonsmoker is.
Friendship example

\[
\frac{p.gpa[2]}{p.gpa[1]} \\
\frac{p.gpa[3]}{p.gpa[1]} \\
\]

Note: It is possible for \( p_{01}/p_{00} = p_{10}/p_{00} = 1 \), but \( \gamma \) to be large.

- In this case \( \gamma = p_{11}/p_{00} \).
- Deviations from 1 indicate heterogeneity in within-group ties.
- Such deviations indicate within-group preferences, or homophily.
Friendship example

\[
\frac{p.gpa[1]}{p.gpa[2]/p.gpa[1]}
\]
\[
= \frac{0.03903421}{0.9638469}
\]
\[
= 0.0404213
\]

\[
\]
\[
= \frac{1.008332}{1.248783}
\]
\[
= 0.807393
\]

**Note:** It is possible for \( p_{01}/p_{00} = p_{10}/p_{00} = 1 \), but \( \gamma \) to be large.

- In this case \( \gamma = p_{11}/p_{00} \).
- Deviations from 1 indicate heterogeneity in within-group ties.
- Such deviations indicate within group preferences, or homophily.
Friendship example

p.gpa[1]
## [1] 0.03903421

p.gpa[2]/p.gpa[1]
## [1] 0.9638469

p.gpa[3]/p.gpa[1]
## [1] 1.008332

## [1] 1.248783

**Note:** It is possible for \( p_{01}/p_{00} = p_{10}/p_{00} = 1 \), but \( \gamma \) to be large.

- In this case \( \gamma = p_{11}/p_{00} \).
- Deviations from 1 indicate heterogeneity in within-group ties.
- Such deviations indicate within group preferences, or homophily.
Friendship example

\[
\frac{p(\text{gpa}[1])}{p(\text{gpa}[2])/p(\text{gpa}[1])} = \frac{0.03903421}{0.9638469} = 0.040479
\]

\[
\frac{p(\text{gpa}[3])/p(\text{gpa}[1])}{p(\text{gpa}[2])/p(\text{gpa}[1])} = \frac{1.008332}{0.9638469} = 1.047581
\]

\[
\frac{(p(\text{gpa}[1])*p(\text{gpa}[4])) / (p(\text{gpa}[2])*p(\text{gpa}[3]) )}{p(\text{gpa}[2])/p(\text{gpa}[1])} = \frac{1.248783}{0.9638469} = 1.299974
\]

**Note:** It is possible for \( p_{01}/p_{00} = p_{10}/p_{00} = 1 \), but \( \gamma \) to be large.

- **In this case** \( \gamma = p_{11}/p_{00} \).
- **Deviations from 1** indicate heterogeneity in within-group ties.
- **Such deviations** indicate within group preferences, or homophily.
Friendship example

\begin{verbatim}

p.gpa[1]
## [1] 0.03903421
p.gpa[2]/p.gpa[1]
## [1] 0.9638469
p.gpa[3]/p.gpa[1]
## [1] 1.008332
## [1] 1.248783
\end{verbatim}

\textbf{Note:} It is possible for \( p_{01}/p_{00} = p_{10}/p_{00} = 1 \), but \( \gamma \) to be large.

- In this case \( \gamma = p_{11}/p_{00} \).
- Deviations from 1 indicate heterogeneity in within-group ties.
- Such deviations indicate within group preferences, or homophily.
Friendship example

```r
p.gpa[1]
## [1] 0.03903421
p.gpa[2]/p.gpa[1]
## [1] 0.9638469
p.gpa[3]/p.gpa[1]
## [1] 1.008332
## [1] 1.248783
```

**Note:** It is possible for $p_{01}/p_{00} = p_{10}/p_{00} = 1$, but $\gamma$ to be large.

- In this case $\gamma = p_{11}/p_{00}$.
- Deviations from 1 indicate heterogeneity in within-group ties.
- Such deviations indicate within group preferences, or homophily.
A useful tool for describing effects on a binary variable is **logistic regression**

Given

- a binary outcome variable $y$
- binary explanatory variables $x_1, x_2$

A logistic regression model for $y$ in terms of $x_1, x_2$ is

$$
Pr(y = 1|x_1, x_2) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

Based on this, we see that

$$
Pr(y = 0|x_1, x_2) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

$$
\text{odds}(y = 1|x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)
$$

$$
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
$$
A useful tool for describing effects on a binary variable is **logistic regression**

Given

- a binary outcome variable $y$
- binary explanatory variables $x_1$, $x_2$

A logistic regression model for $y$ in terms of $x_1$, $x_2$ is

$$Pr(y = 1|x_1, x_2) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}$$

Based on this, we see that

$$Pr(y = 0|x_1, x_2) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}$$

$$\text{odds}(y = 1|x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)$$

$$\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$
A useful tool for describing effects on a binary variable is **logistic regression**

**Given**

- a binary outcome variable $y$
- binary explanatory variables $x_1$, $x_2$

A logistic regression model for $y$ in terms of $x_1$, $x_2$ is

$$
\text{Pr}(y = 1|x_1, x_2) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

Based on this, we see that

$$
\text{Pr}(y = 0|x_1, x_2) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

$$
\text{odds}(y = 1|x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)
$$

$$
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
$$
A useful tool for describing effects on a binary variable is **logistic regression**

**Given**

- a binary outcome variable $y$
- binary explanatory variables $x_1$, $x_2$

A logistic regression model for $y$ in terms of $x_1$, $x_2$ is

$$
Pr(y = 1|x_1, x_2) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

Based on this, we see that

$$
Pr(y = 0|x_1, x_2) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
$$

$$
\text{odds}(y = 1|x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)
$$

$$
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
$$
A useful tool for describing effects on a binary variable is **logistic regression**

Given

- a binary outcome variable \( y \)
- binary explanatory variables \( x_1, x_2 \)

A logistic regression model for \( y \) in terms of \( x_1, x_2 \) is

\[
\Pr(y = 1 | x_1, x_2) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
\]

Based on this, we see that

\[
\Pr(y = 0 | x_1, x_2) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2}}
\]

\[
\text{odds}(y = 1 | x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)
\]

\[
\log \text{odds}(y = 1 | x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]
Log-odds ratios in logistic regression

For example,

\[
\begin{align*}
\text{odds}(y = 1|0, 0) &= \exp(\beta_0) \\
\text{odds}(y = 1|1, 0) &= \exp(\beta_0 + \beta_1) \\
\text{odds ratio}(y = 1|(1, 0), (0, 0)) &= \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \\
\text{log odds ratio}(y = 1|(1, 0), (0, 0)) &= \beta_1
\end{align*}
\]

In logistic regression

- $\beta_1$, the "effect" of $x_1$, represents the log odds ratio ($y = 1|(1, 0), (0, 0)$)
- $\beta_2$, the "effect" of $x_2$, represents the log odds ratio ($y = 1|(0, 1), (0, 0)$)

What about the interaction?
Log-odds ratios in logistic regression

For example,

\[
\text{odds}(y = 1|0, 0) = \exp(\beta_0)
\]
\[
\text{odds}(y = 1|1, 0) = \exp(\beta_0 + \beta_1)
\]
\[
\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)
\]
\[
\log \text{odds ratio}(y = 1|(1, 0), (0, 0)) = \beta_1
\]

In logistic regression

- \(\beta_1\), the “effect” of \(x_1\), represents the log odds ratio \((y = 1|(1, 0), (0, 0))\)
- \(\beta_2\), the “effect” of \(x_2\), represents the log odds ratio \((y = 1|(0, 1), (0, 0))\)

What about the interaction?
Log-odds ratios in logistic regression

For example,

\[
\begin{align*}
\text{odds}(y = 1|0, 0) &= \exp(\beta_0) \\
\text{odds}(y = 1|1, 0) &= \exp(\beta_0 + \beta_1)
\end{align*}
\]

\[
\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)
\]

\[
\text{log odds ratio}(y = 1|(1, 0), (0, 0)) = \beta_1
\]

In logistic regression

- $\beta_1$, the “effect” of $x_1$, represents the log odds ratio $(y = 1|(1, 0), (0, 0))$
- $\beta_2$, the “effect” of $x_2$, represents the log odds ratio $(y = 1|(0, 1), (0, 0))$

What about the interaction?
Log-odds ratios in logistic regression

For example,

\[
\text{odds}(y = 1|0, 0) = \exp(\beta_0) \\
\text{odds}(y = 1|1, 0) = \exp(\beta_0 + \beta_1)
\]

\[
\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)
\]

\[
\log \text{odds ratio}(y = 1|(1, 0), (0, 0)) = \beta_1
\]

In logistic regression

- \(\beta_1\), the “effect” of \(x_1\), represents the log odds ratio \(y = 1|(1, 0), (0, 0)\)
- \(\beta_2\), the “effect” of \(x_2\), represents the log odds ratio \(y = 1|(0, 1), (0, 0)\)

What about the interaction?
Log-odds ratios in logistic regression

For example,

\[
\text{odds}(y = 1|0, 0) = \exp(\beta_0)
\]
\[
\text{odds}(y = 1|1, 0) = \exp(\beta_0 + \beta_1)
\]

\[
\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)
\]

\[
\text{log odds ratio}(y = 1|(1, 0), (0, 0)) = \beta_1
\]

In logistic regression

- $\beta_1$, the “effect” of $x_1$, represents the log odds ratio $(y = 1|(1, 0), (0, 0))$
- $\beta_2$, the “effect” of $x_2$, represents the log odds ratio $(y = 1|(0, 1), (0, 0))$

What about the interaction?
Log-odds ratios in logistic regression

For example,

\[
\begin{align*}
&\text{odds}(y = 1|0, 0) = \exp(\beta_0) \\
&\text{odds}(y = 1|1, 0) = \exp(\beta_0 + \beta_1) \\
&\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \\
&\log \text{odds ratio}(y = 1|(1, 0), (0, 0)) = \beta_1
\end{align*}
\]

In logistic regression

- \(\beta_1\), the “effect” of \(x_1\), represents the log odds ratio \((y = 1|(1, 0), (0, 0))\)
- \(\beta_2\), the “effect” of \(x_2\), represents the log odds ratio \((y = 1|(0, 1), (0, 0))\)

What about the interaction?
Log-odds ratios in logistic regression

\[
\text{odds}(y = 1| x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)
\]

\[
\text{odds ratio}(y = 1|(1, 1), (0, 1)) = \frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_{12})}{\exp(\beta_0 + \beta_2)} = \exp(\beta_1 + \beta_{12})
\]

\[
\text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1)
\]

Therefore

\[
\log \frac{\text{odds ratio}(y = 1|(1, 1), (0, 1))}{\text{odds ratio}(y = 1|(1, 0), (0, 0))} = \beta_{12}
\]
Log-odds ratios in logistic regression

\[ \text{odds}(y = 1|x_1, x_2) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2) \]

\[ \text{odds ratio}(y = 1|(1, 1), (0, 1)) = \frac{\exp(\beta_0 + \beta_1 + \beta_2 + \beta_{12})}{\exp(\beta_0 + \beta_2)} = \exp(\beta_1 + \beta_{12}) \]

\[ \text{odds ratio}(y = 1|(1, 0), (0, 0)) = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1) \]

Therefore

\[ \frac{\text{odds ratio}(y = 1|(1, 1), (0, 1))}{\text{odds ratio}(y = 1|(1, 0), (0, 0))} = \exp(\beta_{12}) \]

\[ \log \frac{\text{odds ratio}(y = 1|(1, 1), (0, 1))}{\text{odds ratio}(y = 1|(1, 0), (0, 0))} = \beta_{12} \]
Log-odds ratios in logistic regression

\[ o_{x_1 x_2} = \frac{\Pr(y_{1, 2} = 1|x_1, x_2)}{1 - \Pr(y_{1, 2} = 1|x_1, x_2)} = \frac{p_{x_1 x_2}}{1 - p_{x_1 x_2}} \]

<table>
<thead>
<tr>
<th></th>
<th>x2=0</th>
<th>x2=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1=0</td>
<td>( o_{00} )</td>
<td>( o_{01} )</td>
</tr>
<tr>
<td>x1=1</td>
<td>( o_{10} )</td>
<td>( o_{11} )</td>
</tr>
</tbody>
</table>

Under the logistic regression model

\[ \beta_0 = \log o_{00} \]
\[ \beta_1 = \log \frac{o_{10}}{o_{00}} \]
\[ \beta_2 = \log \frac{o_{10}}{o_{00}} \]
\[ \beta_{12} = \log \frac{o_{11}/o_{01}}{o_{10}/o_{00}} = \log \frac{o_{11} o_{00}}{o_{01} o_{10}} \]

How do \( \{\beta_0, \beta_1, \beta_2, \beta_{12}\} \) relate to \( \{p_{00}, p_{10}/p_{00}, p_{01}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\} \) ?
Log-odds ratios in logistic regression

\[ o_{x_1,x_2} = \frac{\Pr(y_{1,2} = 1|x_1, x_2)}{1 - \Pr(y_{1,2} = 1|x_1, x_2)} = \frac{p_{x_1,x_2}}{1 - p_{x_1,x_2}} \]

<table>
<thead>
<tr>
<th></th>
<th>x2=0</th>
<th>x2=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1=0</td>
<td>(o_{00})</td>
<td>(o_{01})</td>
</tr>
<tr>
<td>x1=1</td>
<td>(o_{10})</td>
<td>(o_{11})</td>
</tr>
</tbody>
</table>

Under the logistic regression model

\[
\beta_0 = \log o_{00}
\]

\[
\beta_1 = \log \frac{o_{10}}{o_{00}}
\]

\[
\beta_2 = \log \frac{o_{10}}{o_{00}}
\]

\[
\beta_{12} = \log \frac{o_{11}/o_{01}}{o_{10}/o_{00}} = \log \frac{o_{11}o_{00}}{o_{01}o_{10}}
\]

How do \(\{\beta_0, \beta_1, \beta_2, \beta_{12}\}\) relate to \(\{p_{00}, p_{10}/p_{00}, p_{01}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\}\)?
Friendship example

```r
Xr <- matrix(hsmoke, nrow(Y), ncol(Y))
Xc <- t(Xr)

xr <- c(Xr)
xc <- c(Xc)
y <- c(Y)

fit <- glm(y ~ xr + xc + xr:xc, family = binomial)

exp(fit$coef)
```

Do these numbers look familiar?

```r
p.smoke[1]

## [1] 0.04425837


## [1] 0.6919841

p.smoke[3]/p.smoke[1]

## [1] 0.9941797

(p.smoke[1]*p.smoke[4]) / (p.smoke[2]*p.smoke[3])

## [1] 1.470585
```
Friendship example

```r
Xr<-matrix(hsmoke,nrow(Y),ncol(Y))
Xc<-t(Xr)

xr<-c(Xr)
xc<-c(Xc)
y<-c(Y)

fit<-glm(y~ xr+ xc + xr*xc, family=binomial)

exp(fit$coef)
```

```
## (Intercept)  xr  xc  xr:xc
## 0.04630788  0.68225274  0.99391180  1.49277105
```

Do these numbers look familiar?

```r
p.smoke[1]
```

```
## [1] 0.04425837
```

```r
```

```
## [1] 0.6919841
```

```r
p.smoke[3]/p.smoke[1]
```

```
## [1] 0.9941797
```

```r
```

```
## [1] 1.470585
```
Friendship example

Xr<-matrix(hsmoke,nrow(Y),ncol(Y))
Xc<-t(Xr)

xr<-c(Xr)
xc<-c(Xc)
y<-c(Y)

fit<-glm(y~ xr+ xc + xr*xc, family=binomial)

exp(fit$coef)

## (Intercept) xr xc xr:xc
## 0.04630788 0.68225274 0.99391180 1.49277105

Do these numbers look familiar?

p.smoke[1]

## [1] 0.04425837


## [1] 0.6919841

p.smoke[3]/p.smoke[1]

## [1] 0.9941797

(p.smoke[1]*p.smoke[4]) / (p.smoke[2]*p.smoke[3])

## [1] 1.470585
Comparing summaries

If network density is very low,

- \(1 - p_{x_ix_j} \approx 1\)
- \(o_{x_ix_j} = p_{x_ix_j} / (1 - p_{x_ix_j}) \approx p_{x_ix_j}\)

and so

\[
\begin{array}{c|cc}
& x_i=0 & x_i=1 \\
\hline
x_j=0 & p_{00} & p_{01} \\
x_j=1 & p_{10} & p_{11} \\
\end{array}
\approx
\begin{array}{c|cc}
& x_i=0 & x_i=1 \\
\hline
x_j=0 & o_{00} & o_{01} \\
x_j=1 & o_{10} & o_{11} \\
\end{array}
\]

Therefore

\[
\{p_{00}, p_{10}/p_{00}, p_{01}/p_{00}, p_{11}/p_{00}, p_{11}/p_{01}\} \approx \{o_{00}, o_{10}/o_{00}, o_{01}/o_{00}, o_{11}/o_{00}, (o_{11}/o_{00})/(o_{01}/o_{00})\}
\]

\[
= \{e^{\beta_0}, e^{\beta_1}, e^{\beta_2}, e^{\beta_{12}}\}
\]
Comparing summaries

If network density is very low,

- $1 - p_{x_ix_j} \approx 1$
- $o_{x_ix_j} = p_{x_ix_j}/(1 - p_{x_ix_j}) \approx p_{x_ix_j}$

and so

<table>
<thead>
<tr>
<th>xi=0</th>
<th>xi=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xj=0</td>
<td>$p_{00}$</td>
</tr>
<tr>
<td>xj=1</td>
<td>$p_{10}$</td>
</tr>
</tbody>
</table>

$\approx$

<table>
<thead>
<tr>
<th>xi=0</th>
<th>xi=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xj=0</td>
<td>$o_{00}$</td>
</tr>
<tr>
<td>xj=1</td>
<td>$o_{10}$</td>
</tr>
</tbody>
</table>

Therefore

\[
\{p_{00}, p_{10}/p_{00}, p_{01}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\} \approx \{o_{00}, o_{10}/o_{00}, o_{01}/o_{00}, (o_{11}o_{00})/(o_{01}o_{10})\} \\
= \{e^{\beta_0}, e^{\beta_1}, e^{\beta_2}, e^{\beta_{12}}\}
\]
Comparing summaries

If network density is very low,

- \( 1 - p_{x_i x_j} \approx 1 \)
- \( o_{x_i x_j} = p_{x_i x_j} / (1 - p_{x_i x_j}) \approx p_{x_i x_j} \)

and so

\[
\begin{array}{c|cc}
\text{xi} = 0 & \text{xi} = 1 \\
\hline
\text{xj} = 0 & p_{00} & p_{01} \\
\text{xj} = 1 & p_{10} & p_{11} \\
\end{array}
\approx
\begin{array}{c|cc}
\text{xi} = 0 & \text{xi} = 1 \\
\hline
\text{xj} = 0 & o_{00} & o_{01} \\
\text{xj} = 1 & o_{10} & o_{11} \\
\end{array}
\]

Therefore

\[
\{ p_{00}, p_{10} / p_{00}, p_{01} / p_{00}, (p_{11} p_{00}) / (p_{01} p_{10}) \} \approx \{ o_{00}, o_{10} / o_{00}, o_{01} / o_{00}, (o_{11} o_{00}) / (o_{01} o_{10}) \}
\]

\[
= \{ e^{\beta_0}, e^{\beta_1}, e^{\beta_2}, e^{\beta_{12}} \}
\]
Comparing summaries

If network density is very low,

- \(1 - p_{x_ix_j} \approx 1\)
- \(o_{x_ix_j} = p_{x_ix_j}/(1 - p_{x_ix_j}) \approx p_{x_ix_j}\)

and so

\[
\begin{array}{c|cc}
  & x_i=0 & x_i=1 \\
  \hline 
  x_j=0 & p_{00} & p_{01} \\
  x_j=1 & p_{10} & p_{11} \\
\end{array}
\]

\[
\begin{array}{c|cc}
  & x_i=0 & x_i=1 \\
  \hline 
  x_j=0 & o_{00} & o_{01} \\
  x_j=1 & o_{10} & o_{11} \\
\end{array}
\]

Therefore

\[
\{p_{00}, p_{10}/p_{00}, p_{01}/p_{00}, (p_{11}p_{00})/(p_{01}p_{10})\} \approx \{o_{00}, o_{10}/o_{00}, o_{01}/o_{00}, (o_{11}o_{00})/(o_{01}o_{10})\}
\]

\[
= \{e^{\beta_0}, e^{\beta_1}, e^{\beta_2}, e^{\beta_{12}}\}
\]
Undirected data

Now we have

<table>
<thead>
<tr>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_i=0</td>
<td>p_{00}</td>
</tr>
<tr>
<td>x_i=1</td>
<td>p_{10} = p_{01}</td>
</tr>
</tbody>
</table>

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[ \{ p_{00}, p_{01}, p_{11} \} \sim \{ p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01} \} \]

The interpretation of these is roughly the same as before:

- \( p_{00} \) represents a baseline rate (both x's 0)
- \( p_{01}/p_{00} \) represents a relative rate (one x 0 versus both x's 0)
- \( p_{11}p_{00}/p_{01}^2 \) represents a homophily effect - the preference of like for like.
- the excess within-group density, beyond the effect of one group being more active than another.
Undirected data

Now we have

<table>
<thead>
<tr>
<th></th>
<th>$x_j=0$</th>
<th>$x_j=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i=0$</td>
<td>$p_{00}$</td>
<td>$p_{01} = p_{10}$</td>
</tr>
<tr>
<td>$x_i=1$</td>
<td>$p_{10} = p_{01}$</td>
<td>$p_{11}$</td>
</tr>
</tbody>
</table>

Now there are only three (unique) numbers in the table.

We can express these as follows:

$$\{p_{00}, p_{01}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01}^2\}$$

The interpretation of these is roughly the same as before:

- $p_{00}$ represents a baseline rate (both $x$'s 0)
- $p_{01}/p_{00}$ represents a relative rate (one $x$ 0 versus both $x$'s 0)
- $p_{11}p_{00}/p_{01}^2$ represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Undirected data

Now we have

\[
\begin{array}{c|cc}
 & x_j=0 & x_j=1 \\
\hline
x_i=0 & p_{00} & p_{01} = p_{10} \\
x_i=1 & p_{10} = p_{01} & p_{11} \\
\end{array}
\]

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[
\{p_{00}, p_{01}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01}^2\}
\]

The interpretation of these is roughly the same as before:

- \(p_{00}\) represents a baseline rate (both \(x\)'s 0)
- \(p_{01}/p_{00}\) represents a relative rate (one \(x\) 0 versus both \(x\)'s 0)
- \(p_{11}p_{00}/p_{01}^2\) represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Undirected data

Now we have

<table>
<thead>
<tr>
<th></th>
<th>x_j = 0</th>
<th>x_j = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_i = 0</td>
<td>p_{00}</td>
<td>p_{01}</td>
</tr>
<tr>
<td></td>
<td>p_{10}</td>
<td></td>
</tr>
<tr>
<td>x_i = 1</td>
<td>p_{10}</td>
<td>p_{11}</td>
</tr>
</tbody>
</table>

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[ \{ p_{00}, p_{01}, p_{11} \} \sim \{ p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01}^2 \} \]

The interpretation of these is roughly the same as before:

- \( p_{00} \) represents a baseline rate (both \( x \)'s 0)
- \( p_{01}/p_{00} \) represents a relative rate (one \( x \) 0 versus both \( x \)'s 0)
- \( p_{11}p_{00}/p_{01}^2 \) represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Now we have

\[
\begin{array}{c|cc}
 & x_j=0 & x_j=1 \\
\hline
x_i=0 & p_{00} & p_{01} = p_{10} \\
x_i=1 & p_{10} = p_{01} & p_{11} \\
\end{array}
\]

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[
\{p_{00}, p_{01}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01}^2\}
\]

The interpretation of these is roughly the same as before:

- \(p_{00}\) represents a baseline rate (both \(x\)'s 0)
- \(p_{01}/p_{00}\) represents a relative rate (one \(x\) 0 versus both \(x\)'s 0)
- \(p_{11}p_{00}/p_{01}^2\) represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Undirected data

Now we have

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>(p_{00})</td>
<td>(p_{01} = p_{10})</td>
</tr>
<tr>
<td>xi=1</td>
<td>(p_{10} = p_{01})</td>
<td>(p_{11})</td>
</tr>
</tbody>
</table>

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[
\{p_{00}, p_{01}, p_{11}\} \sim \{p_{00}, p_{01}/p_{00}, p_{11}p_{00}/p_{01}^2\}
\]

The interpretation of these is roughly the same as before:

- \(p_{00}\) represents a baseline rate (both \(x\)’s 0)
- \(p_{01}/p_{00}\) represents a relative rate (one \(x\) 0 versus both \(x\)’s 0)
- \(p_{11}p_{00}/p_{01}^2\) represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Undirected data

Now we have

<table>
<thead>
<tr>
<th></th>
<th>xj=0</th>
<th>xj=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi=0</td>
<td>p₀₀</td>
<td>p₀₁  = p₁₀</td>
</tr>
<tr>
<td>xi=1</td>
<td>p₁₀  = p₀₁</td>
<td>p₁₁</td>
</tr>
</tbody>
</table>

Now there are only three (unique) numbers in the table.

We can express these as follows:

\[
\{p₀₀, p₀₁, p₁₁\} \sim \{p₀₀, p₀₁/p₀₀, p₁₁p₀₀/p₀₁^2\}
\]

The interpretation of these is roughly the same as before:

- \(p₀₀\) represents a baseline rate (both \(x\)'s 0)
- \(p₀₁/p₀₀\) represents a relative rate (one \(x\) 0 versus both \(x\)'s 0)
- \(p₁₁p₀₀/p₀₁^2\) represents a homophily effect - the preference of like for like.
  - the excess within-group density, beyond the effect of one group being more active than another.
Logistic regression for undirected data

\[
\text{log odds}(y = 1| x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]

Here, \(x_1\) and \(x_2\) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \(x_1\) versus that of \(x_2\), we must have \(\beta_1 = \beta_2\), and the model becomes

\[
\text{log odds}(y = 1| x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2
\]

- \(\beta_0\) represents a baseline rate;
- \(\beta_1\) represents the “additive” effect of either \(x_1 = 1\) or \(x_2 = 1\) on the rate;
- \(\beta_{12}\) represents the additional effect of homophily on the rate.
Logistic regression for undirected data

\[
\text{log odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]

Here, \( x_1 \) and \( x_2 \) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \( x_1 \) versus that of \( x_2 \), we must have \( \beta_1 = \beta_2 \), and the model becomes

\[
\text{log odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2
\]

- \( \beta_0 \) represents a baseline rate;
- \( \beta_1 \) represents the “additive” effect of either \( x_1 = 1 \) or \( x_2 = 1 \) on the rate;
- \( \beta_{12} \) represents the additional effect of homophily on the rate.
Logistic regression for undirected data

\[ \log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \]

Here, \( x_1 \) and \( x_2 \) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \( x_1 \) versus that of \( x_2 \), we must have \( \beta_1 = \beta_2 \), and the model becomes

\[ \log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2 \]

- \( \beta_0 \) represents a baseline rate;
- \( \beta_1 \) represents the “additive” effect of either \( x_1 = 1 \) or \( x_2 = 1 \) on the rate;
- \( \beta_{12} \) represents the additional effect of homophily on the rate.
Logistic regression for undirected data

\[
\text{log odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]

Here, \(x_1\) and \(x_2\) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \(x_1\) versus that of \(x_2\), we must have \(\beta_1 = \beta_2\), and the model becomes

\[
\text{log odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2
\]

- \(\beta_0\) represents a baseline rate;
- \(\beta_1\) represents the “additive” effect of either \(x_1 = 1\) or \(x_2 = 1\) on the rate;
- \(\beta_{12}\) represents the additional effect of homophily on the rate.
Logistic regression for undirected data

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]

Here, \(x_1\) and \(x_2\) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \(x_1\) versus that of \(x_2\), we must have \(\beta_1 = \beta_2\), and the model becomes

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2
\]

- \(\beta_0\) represents a baseline rate;
- \(\beta_1\) represents the “additive” effect of either \(x_1 = 1\) or \(x_2 = 1\) on the rate;
- \(\beta_{12}\) represents the additional effect of homophily on the rate.
Logistic regression for undirected data

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2
\]

Here, \(x_1\) and \(x_2\) are not “sender” and “receiver” effects, as there are no senders or receivers.

As there is no way to differentiate the effect of \(x_1\) versus that of \(x_2\), we must have \(\beta_1 = \beta_2\), and the model becomes

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1 (x_1 + x_2) + \beta_2 x_1 x_2
\]

- \(\beta_0\) represents a baseline rate;
- \(\beta_1\) represents the “additive” effect of either \(x_1 = 1\) or \(x_2 = 1\) on the rate;
- \(\beta_{12}\) represents the additional effect of homophily on the rate.
Interpreting coefficients

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_1 x_2
\]

For example, suppose

- \(y_{i,j} = \) indicator of friendship;
- \(x_i = \) indicator of friendliness.

Under no homophily, i.e. \(\beta_{12} = 0\),

\[
\begin{align*}
\log \text{odds}(y = 1|0, 1) &= \beta_0 + \beta_1 \\
\log \text{odds}(y = 1|1, 1) &= \beta_0 + 2\beta_1
\end{align*}
\]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[ \log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2x_1x_2 \]

For example, suppose

- \( y_{i,j} = \) indicator of friendship;
- \( x_i = \) indicator of friendliness.

Under no homophily, i.e. \( \beta_{12} = 0 \),

\[ \log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \]
\[ \log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 \]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_1 x_2
\]

For example, suppose

- \(y_{i,j}\) = indicator of friendship;
- \(x_i\) = indicator of friendliness.

Under no homophily, i.e. \(\beta_{12} = 0\),

\[
\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \\
\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1
\]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[
\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_1 x_2
\]

For example, suppose

- \( y_{i,j} \) = indicator of friendship;
- \( x_i \) = indicator of friendliness.

Under no homophily, i.e. \( \beta_{12} = 0 \),

\[
\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \\
\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1
\]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[
\log \text{ odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2x_1x_2
\]

For example, suppose

- \( y_{i,j} \) = indicator of friendship;
- \( x_i \) = indicator of friendliness.

Under no homophily, i.e. \( \beta_{12} = 0 \),

\[
\log \text{ odds}(y = 1|0, 1) = \beta_0 + \beta_1 \\
\log \text{ odds}(y = 1|1, 1) = \beta_0 + 2\beta_1
\]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[ \log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_1 x_2 \]

For example, suppose

- \( y_{i,j} \) = indicator of friendship;
- \( x_i \) = indicator of friendliness.

Under no homophily, i.e. \( \beta_{12} = 0 \),

\[ \log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \]
\[ \log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 \]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

$$\log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2x_1x_2$$

For example, suppose

- $y_{ij} = \text{indicator of friendship}$;
- $x_i = \text{indicator of friendliness}$.

Under no homophily, i.e. $\beta_{12} = 0$,

$$\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1$$
$$\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1$$

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

\[ \log \text{odds}(y = 1|x_1, x_2) = \beta_0 + \beta_1(x_1 + x_2) + \beta_2 x_1 x_2 \]

For example, suppose

- \( y_{i,j} \) = indicator of friendship;
- \( x_i \) = indicator of friendliness.

Under no homophily, i.e. \( \beta_{12} = 0 \),

\[ \log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \]
\[ \log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 \]

The rate is higher under \((x_i = 1, x_j = 1)\) than \((x_i = 1, x_j = 0)\)

- not because of homophily,
- but because both people are friendly, instead of one.
Interpreting coefficients

Under positive homophily, i.e. $\beta_{12} > 0$,

$$\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1$$
$$\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 + \beta_{12} > \beta_0 + 2\beta_1$$

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$
- both people are friendly,
- additionally, friendly people prefer friendly people.
Under positive homophily, i.e. $\beta_{12} > 0$,

$$\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1$$

$$\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 + \beta_{12} > \beta_0 + 2\beta_1$$

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$

- both people are friendly,
- additionally, friendly people prefer friendly people.
Under positive homophily, i.e. $\beta_{12} > 0$,

\[
\text{log odds}(y = 1|0, 1) = \beta_0 + \beta_1 \\
\text{log odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 + \beta_{12} > \beta_0 + 2\beta_1
\]

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$

- both people are friendly,
- additionally, friendly people prefer friendly people.
Interpreting coefficients

Under positive homophily, i.e. $\beta_{12} > 0$,

\[
\begin{align*}
\log \text{odds}(y = 1|0, 1) &= \beta_0 + \beta_1 \\
\log \text{odds}(y = 1|1, 1) &= \beta_0 + 2\beta_1 + \beta_{12} > \beta_0 + 2\beta_1
\end{align*}
\]

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$

- both people are friendly,
- additionally, friendly people prefer friendly people.
Interpreting coefficients

Under positive homophily, i.e. $\beta_{12} > 0$,

$$
\log \text{odds}(y = 1|0, 1) = \beta_0 + \beta_1 \\
\log \text{odds}(y = 1|1, 1) = \beta_0 + 2\beta_1 + \beta_{12} > \beta_0 + 2\beta_1
$$

The rate is higher under $(x_i = 1, x_j = 1)$ than $(x_i = 1, x_j = 0)$

- both people are friendly,
- additionally, friendly people prefer friendly people.
Computation in R

```r
ys <- c(1*(Y+t(Y)>0))

fit <- glm(ys ~ xr + xc + xr*xc, family=binomial)

exp(fit$coef)
```

```
## (Intercept)    xr    xc    xr:xc
## 0.07455013 0.84579038 0.84579038 1.45293402
```
Summary

- Effects of binary covariates can be described with submatrix densities.
- Submatrix densities can be reparameterized:
  - baseline rate;
  - relative probabilities;
  - homophily.

- These summaries are related to logistic regression coefficients.
• Effects of binary covariates can be described with submatrix densities.

• **Submatrix densities can be reparameterized:**
  - baseline rate;
  - relative probabilities;
  - homophily.

• These summaries are related to logistic regression coefficients.
Summary

- Effects of binary covariates can be described with submatrix densities.
- Submatrix densities can be reparameterized:
  - baseline rate;
  - relative probabilities;
  - homophily.

- These summaries are related to logistic regression coefficients.
Summary

- Effects of binary covariates can be described with submatrix densities.
- Submatrix densities can be reparameterized:
  - baseline rate;
  - relative probabilities;
  - homophily.

- These summaries are related to logistic regression coefficients.
Summary

- Effects of binary covariates can be described with submatrix densities.
- Submatrix densities can be reparameterized:
  - baseline rate;
  - relative probabilities;
  - homophily.

- These summaries are related to logistic regression coefficients.
• Effects of binary covariates can be described with submatrix densities.
• Submatrix densities can be reparameterized:
  • baseline rate;
  • relative probabilities;
  • homophily.
• These summaries are related to logistic regression coefficients.
Summary

• Effects of binary covariates can be described with submatrix densities.
• Submatrix densities can be reparameterized:
  • baseline rate;
  • relative probabilities;
  • homophily.
• These summaries are related to logistic regression coefficients.
Summary

• Effects of binary covariates can be described with submatrix densities.
• Submatrix densities can be reparameterized:
  • baseline rate;
  • relative probabilities;
  • homophily.
• These summaries are related to logistic regression coefficients.