Introduction to (log) Odds Ratio

Statistics and Methodology
Anna Lobanova: a.lobanova@ai.rug.nl
Who is more likely to drink beer on Queen’s Day - students or teachers?
Example: Who is more likely to drink beer on Queen’s Day - students or teachers?

<table>
<thead>
<tr>
<th></th>
<th>Drink</th>
<th>Don’t drink</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>90</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Teachers</td>
<td>20</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>90</td>
<td>200</td>
</tr>
</tbody>
</table>

**Group 1** = students, **group 2** = teachers

**Event** – drinking beer at Queen’s Day

**Question:** Is one group more likely to drink beer on Queen’s Day than the other group? Or is this event independent of professional status?
## Terminology & Notation:

### A Two-way Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Drink</th>
<th>Don’t drink</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>$n_{1+}$</td>
</tr>
<tr>
<td>Teachers</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$n_{2+}$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_{+1}$</td>
<td>$n_{+2}$</td>
<td>$n$</td>
</tr>
</tbody>
</table>
## Terminology & Notation:

### A Two-way Contingency Table

<table>
<thead>
<tr>
<th></th>
<th>Drink</th>
<th>Don’t drink</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>( n_{11} = 90 )</td>
<td>( n_{12} = 10 )</td>
<td>( n_{1+} = 100 )</td>
</tr>
<tr>
<td>Teachers</td>
<td>( n_{21} = 20 )</td>
<td>( n_{22} = 80 )</td>
<td>( n_{2+} = 100 )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_{+1} = 110 )</td>
<td>( n_{+2} = 90 )</td>
<td>( n = 200 )</td>
</tr>
</tbody>
</table>

- **Explanatory variable**
- **Response variable**

**Joint distribution**

**Marginal distributions**
Example 1: Let’s use **odds ratio** to find out!

<table>
<thead>
<tr>
<th></th>
<th>Drink</th>
<th>Don’t drink</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>$n_{11} = 90$</td>
<td>$n_{12} = 10$</td>
<td>$n_{1+} = 100$</td>
</tr>
<tr>
<td>Teachers</td>
<td>$n_{21} = 20$</td>
<td>$n_{22} = 80$</td>
<td>$n_{2+} = 100$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_{+1} = 110$</td>
<td>$n_{+2} = 90$</td>
<td>$n = 200$</td>
</tr>
</tbody>
</table>

**Step 1:** the odds of a student drinking beer is 90 to 10 or $9/1$ and the odds of a teacher drinking beer is 20 to 80 or $1/4 = 0.25:1$

**Step 2:** the probability of success for every cell is

\[
\pi_{11} = \frac{n_{11}}{n_{1+}} = \frac{90}{100} = 0.9 \quad \pi_{12} = \frac{n_{12}}{n_{1+}} = \frac{10}{100} = 0.1
\]

\[
\pi_{21} = \frac{n_{21}}{n_{2+}} = \frac{20}{100} = 0.2 \quad \pi_{22} = \frac{n_{22}}{n_{2+}} = \frac{80}{100} = 0.8
\]

**Step 3:** Odds Ratio ($\theta$) = $\frac{0.9/0.1}{0.2/0.8} = \frac{0.72}{0.02} = 36$
Odds Ratio (\( \theta \)) = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}
Inference from odds ratio:

If

<table>
<thead>
<tr>
<th>Condition</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>odds ratio = 1</td>
<td>the event is equally likely in both groups</td>
</tr>
<tr>
<td>odds ratio &gt; 1</td>
<td>the event is more likely in Group 1</td>
</tr>
<tr>
<td>odds ratio &lt; 1</td>
<td>the event is more likely in Group 2</td>
</tr>
</tbody>
</table>

⇒ the greater the number, the stronger the association

⇒ is never a negative number

In example 1:

odds ratio = 36

students are much more likely to drink beer than teachers!
Inference from odds ratio:

If

odds ratio = 1

Then

the event is equally likely in both groups

odds ratio > 1

the event is more likely in Group 1

odds ratio < 1

the event is more likely in Group 2

⇒

the greater the number

the stronger the association

In example 1:

odds ratio = 36

students are much more likely to drink beer than teachers!
Odds Ratio:

› is suitable for categorical data;

› usually deals with associations between 2 categorical variables;

› a change of values (in rows with columns) does not play a role;

› unlike chi-square, odds ratio gives us a direction of association!
BUT

- for small to moderate sample sizes, the sampling distribution of the odds ratio is highly skewed!

WHY?
BUT

› for small to moderate sample sizes, the sampling distribution of the odds ratio is highly skewed!

› To overcome this problem, one can use an alternative but equivalent measure – Log Odds Ratio
## Log Odds Ratio $\log(\theta)$

<table>
<thead>
<tr>
<th>Odd Ratio</th>
<th>Log Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 1$</td>
<td>$\log(\theta) = 0$</td>
</tr>
<tr>
<td>$\theta = 2$</td>
<td>$\log(\theta) = 0.7$</td>
</tr>
<tr>
<td>$\theta = 0.5$</td>
<td>$\log(\theta) = -0.7$</td>
</tr>
<tr>
<td>$\theta = 36$</td>
<td>$\log(\theta) = 3.6$</td>
</tr>
</tbody>
</table>
Log Odds Ratio $\log(\theta)$

The formula for the **standard error** of $\log(\theta)$ is very simple:

(1)  $SE(\log(\theta)) = \text{square-root}(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22})$.

Knowing this standard error, one can test (2) the **significance** of $\log(\theta)$ and/or construct (3) **confidence intervals**:

(2)  $z = \log(\theta)/SE\log(\theta)$

(3)  $\log(\theta) \pm z_{\alpha/2} \times SE\log(\theta)$

$z_{\alpha/2}$ is the $z$ value defining the confidence limits
Summary:

› (Log) Odds Ratio is meant for categorical data;

› Mostly used in two by two tables;

› Unlike chi-square it provides info about the direction of association;

› When the sample is small/moderate, it is better to use Log Odds Ratio;

› It is a good tool for finding associations between variables!