university of groningen

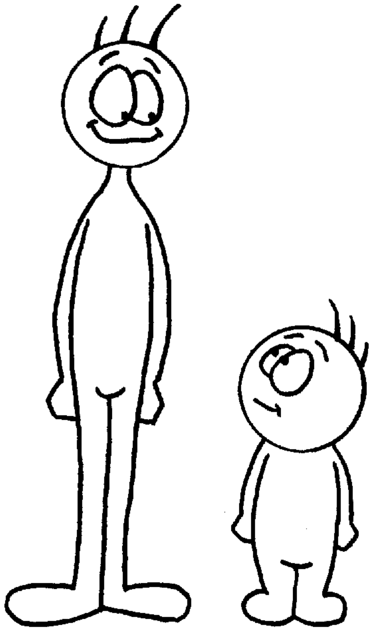


# Introduction to (log) Odds Ratio

Statistics and Methodology

Anna Lobanova: a.lobanova@ai.rug.nl

# Who is more likely to drink beer on Queen's Day - students or teachers?



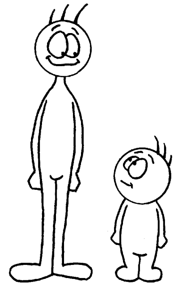
# Example: Who is more likely to drink beer on Queen's Day - students or teachers?

| | Drink | Don't drink | *Total* |
|---|---|---|---|
| Students | 90 | 10 | *100* |
| Teachers | 20 | 80 | *100* |
| *Total* | *110* | *90* | *200* |

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

| | Drink | Don't drink | Total |
|---|---|---|---|
| Students | $n_{11}$ | $n_{12}$ | $n_{1+}$ |
| Teachers | $n_{21}$ | $n_{22}$ | $n_{2+}$ |
| Total | $n_{+1}$ | $n_{+2}$ | $n$ |

# Terminology & Notation:

## A Two-way Contingency Table

response variable

| | Drink | Don't drink | Total |
|---|---|---|---|
| Students | $n_{11} = 90$ | $n_{12} = 10$ | $n_{1+} = 100$ |
| Teachers | $n_{21} = 20$ | $n_{22} = 80$ | $n_{2+} = 100$ |
| Total | $n_{+1} = 110$ | $n_{+2} = 90$ | $n = 200$ |

explanatory variable

joint distribution

marginal distributions

Example 1: Let's use odds ratio to find out!

| | *Drink* | *Don't drink* | *Total* |
|---|---|---|---|
| *Students* | $n_{11} = 90$ | $n_{12} = 10$ | $n_{1+} = 100$ |
| *Teachers* | $n_{21} = 20$ | $n_{22} = 80$ | $n_{2+} = 100$ |
| *Total* | $n_{+1} = 110$ | $n_{+2} = 90$ | $n = 200$ |

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

$\pi_{11} = n_{11}/n_{1+} = 90/100 = 0.9$ $\pi_{12} = n_{12}/n_{1+} = 10/100 = 0.1$

$\pi_{21} = n_{21}/n_{2+} = 20/100 = 0.2$ $\pi_{22} = n_{22}/n_{2+} = 80/100 = 0.8$

Step 3: Odds Ratio (θ) $= \dfrac{0.9/0.1}{0.2/0.8} = \dfrac{0.72}{0.02} = \mathbf{36}$

university of groningen



$$\text{Odds Ratio } (\theta) = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}$$

# Inference from odds ratio:

| If | Then |
|---|---|
| odds ratio = 1 | 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

⇨ 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 | Then |
|---|---|
| odds ratio = 1 | 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(θ)

| Odd Ratio | Log Odds Ratio |
|---|---|
| $\theta = 1$ | $\text{Log}(\theta) = 0$ |
| $\theta = 2$ | $\text{Log}(\theta) = 0.7$ |
| $\theta = 0.5$ | $\text{Log}(\theta) =-0.7$ |
| $\theta = 36$ | $\text{Log}(\theta) =3.6$ |

# Log Odds Ratio log(θ)

The formula for the standard error of log(θ) 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(θ) 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!