Module 5: A Non-Parametric Test

Odds Ratio, RR, GWAS

Agenda:

- **Odds ratio**
- Relative Risk
- Genome-Wide Association Studies

Odds Ratio:

Another type of "Contingency analysis" that **measures the magnitude of association between two categorical variables that each only have two categories:**

– Explanatory and response variables

- the response variable has usually adopts "success" and "failure" as the labels for its two categories
- Used in **case-control** groups
- **Proportion** of success/failure between two groups
- Step 1: Usually testing **Ho: OR=1**

Step 2 (the test statistic)

Odds:

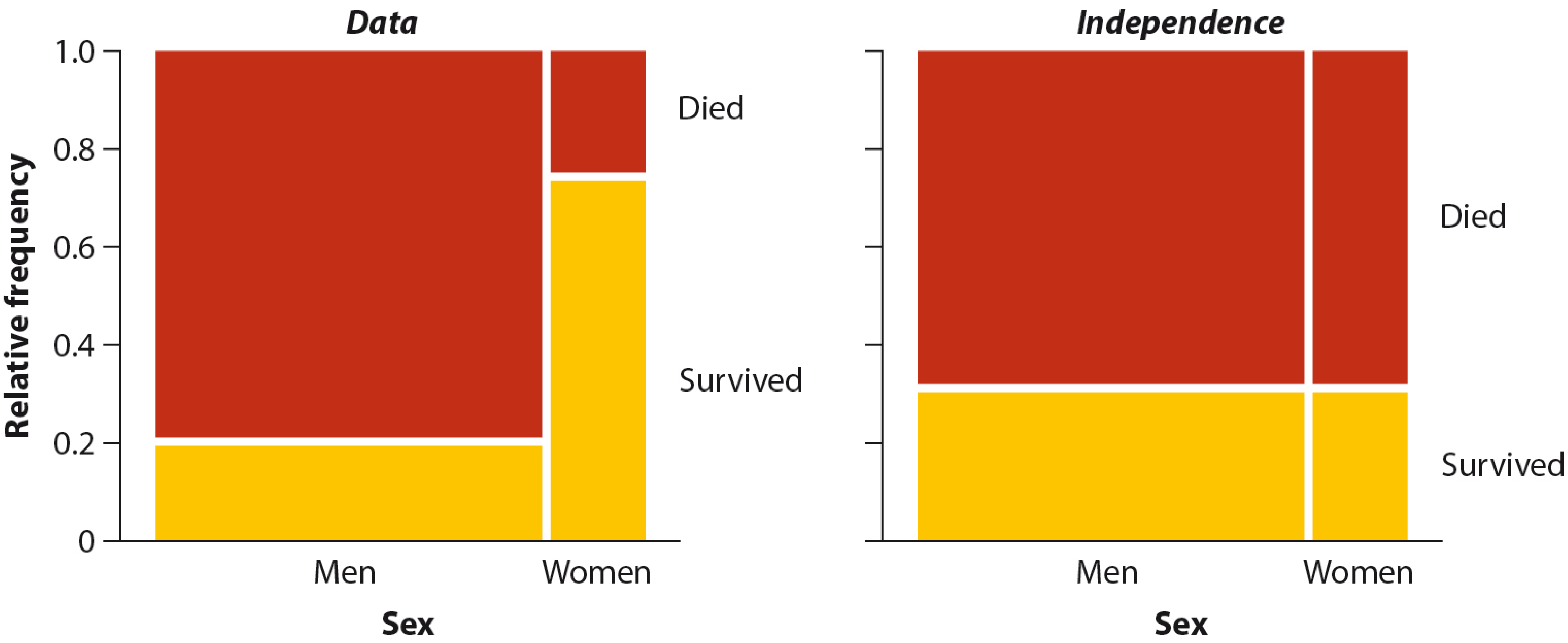
Probability of success divided by the probability of failure

$$O = \frac{p}{1-p}$$

As per usual, we will be using estimates:

$$\hat{O} = \frac{\hat{p}}{1-\hat{p}}$$

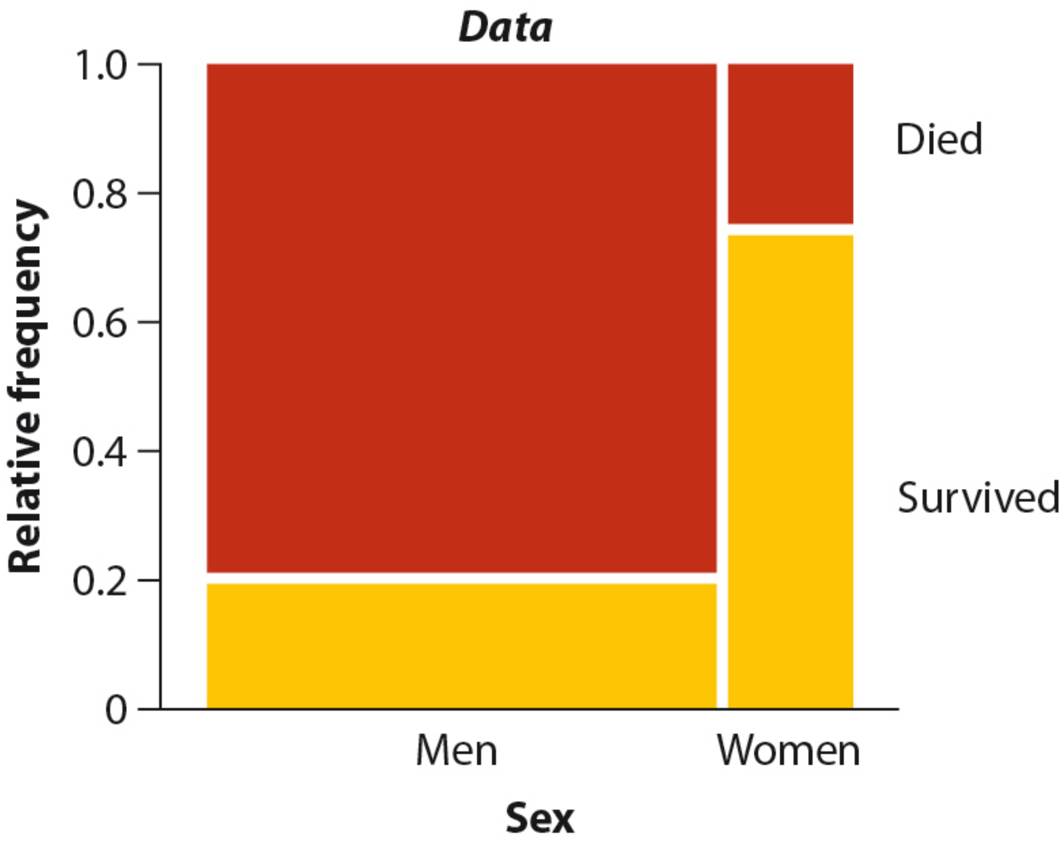
Mosaic plots!



Male versus female passengers on the Titanic

Odds:

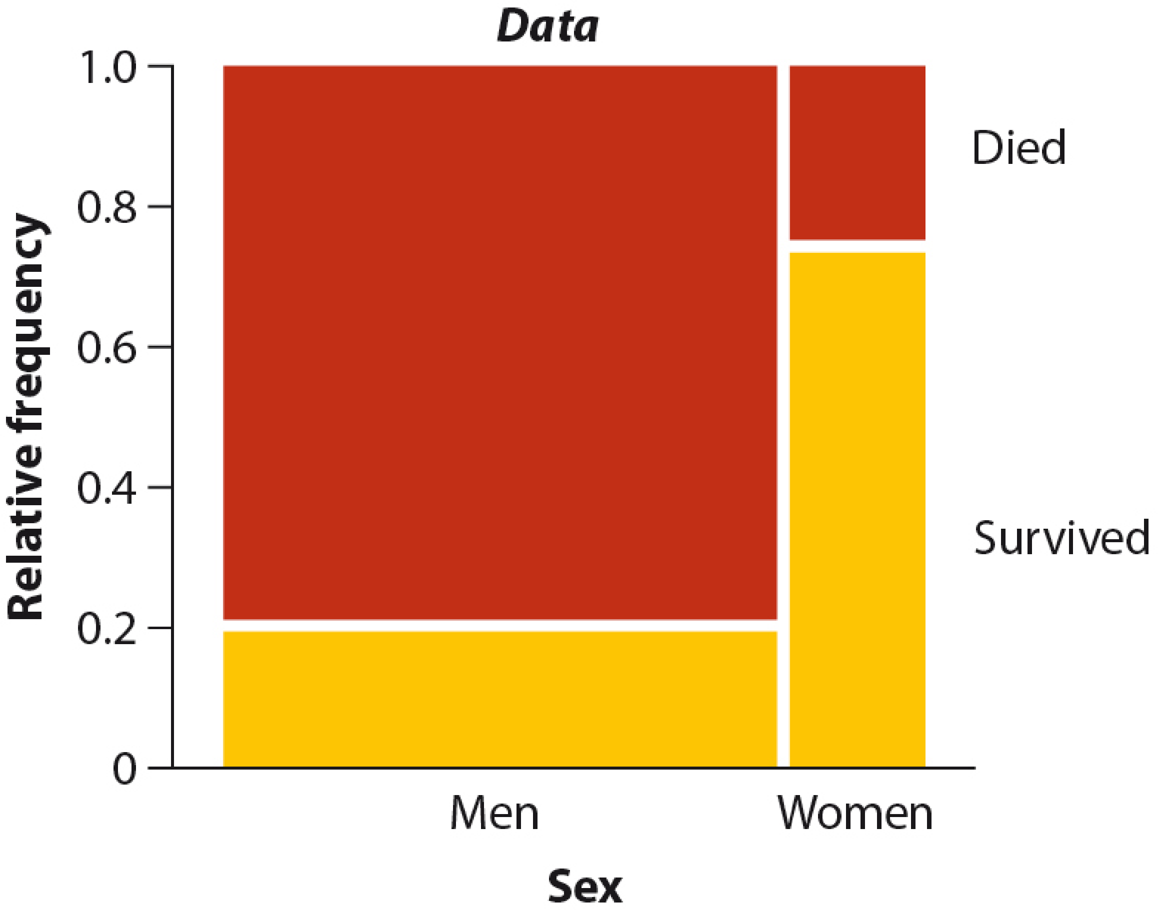
Probability of success divided by the probability of failure



$$O = \frac{p}{1-p}$$

<u>Odds:</u>

Probability of success divided by the probability of failure



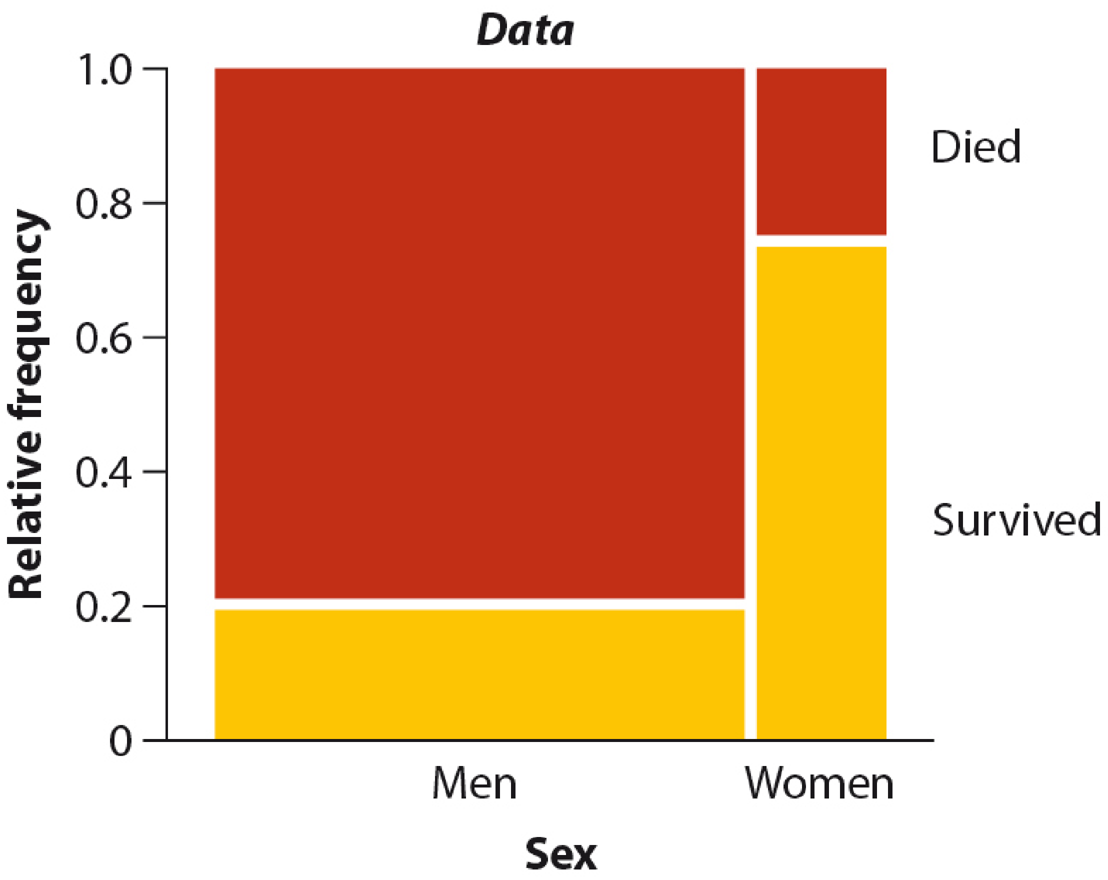
$$O = \frac{p}{1-p}$$

$$O_{men} = \frac{0.20}{1-0.20} = 0.25$$

$$O_{women} = \frac{0.74}{1-0.74} = 2.85$$

Odds:

Probability of success divided by the probability of failure



$$O = \frac{p}{1-p}$$

1 to 4

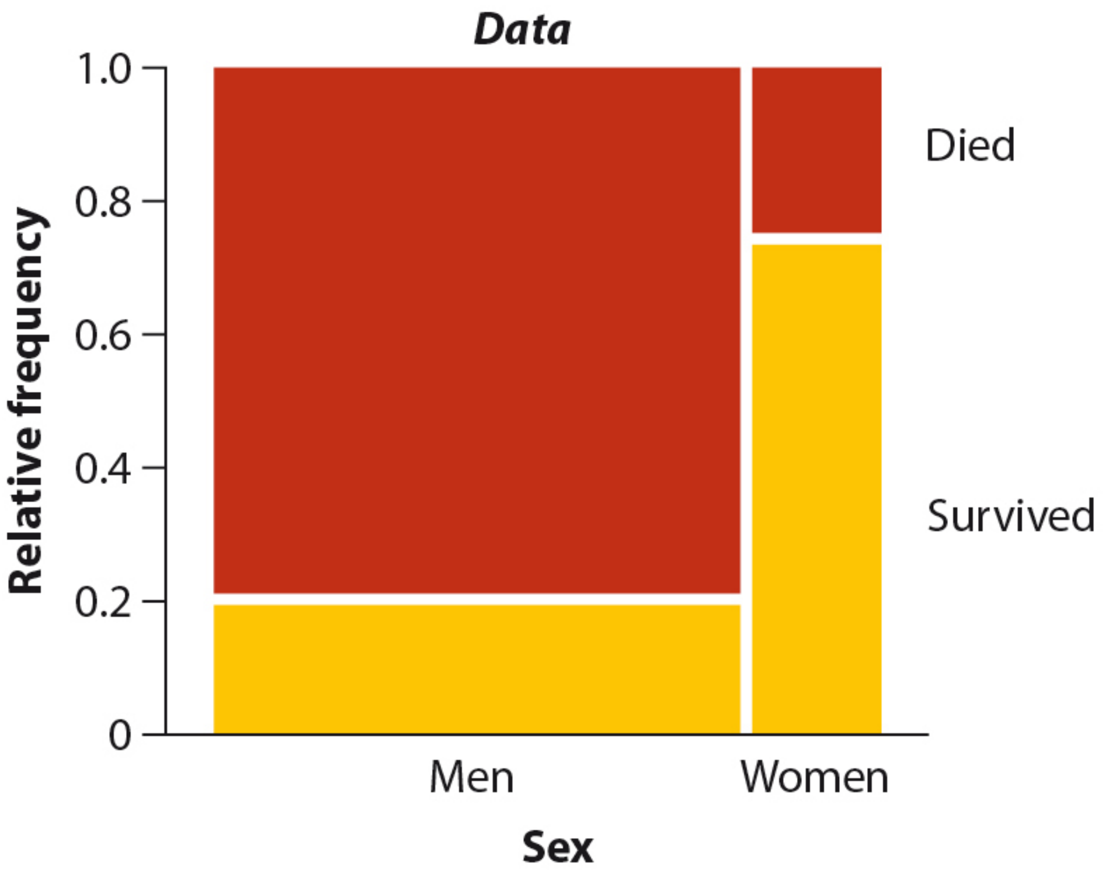
$$O_{men} = \frac{0.20}{1-0.20} = 0.25$$

$$O_{women} = \frac{0.74}{1-0.74} = 2.85$$

3 to 1

Odds Ratio:

The odds of success in one group divided by the odds of success in another group



$$OR = \frac{O_1}{O_2}$$

Odds ratio of female to male survival:

$$OR = \frac{2.85}{0.25} = 11.4$$

Odds Ratio:

The odds of success in one group divided by the odds of success in another group

- *usually asking "Does the treatment/intervention" change the outcome (compared to control)?*

$$\widehat{OR} = \frac{\widehat{O_1}}{\hat{O}_2} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

	Treatment	Control
Success	a	b
Failure	c	d

$$\text{OR} = \frac{\frac{P(Y=1|X=1)}{1-P(Y=1|X=1)}}{\frac{P(Y=1|X=0)}{1-P(Y=1|X=0)}}$$

<u>Odds Ratio:</u>

Measures the <u>magnitude</u> (or strength) **of association between <u>two categorical variables</u> that each only have <u>two categories</u>:**

- Explanatory and response variables
  - the response variable usually adopts "success" and "failure" as the labels for its two categories
  - Used in **case-control** groups
  - **Proportion** of success/failure between two groups
  - Step 1: Usually testing **Ho: OR=1**

<u>The most challenging parts of an odds-ratio:</u>

1. *Keep track of which one is a success, and which one is a failure*
2. *The TRANSFORMATION **necessary** for step 3*

Confidence Interval Odds Ratio:

- Confidence interval is used to determine whether O.R. >>1 or << 1 is statistically significant (Ho: OR =1)

- Same basic idea as confidence intervals:

Point Estimate ± Z*Standard Error

For example, 95% Confidence Interval: $\bar{X} \pm 1.96 * SE_{\bar{x}}$

This corresponds to an interval:

$$\bar{X} - 1.96 * SE_{\bar{x}} < \mu < \bar{X} + 1.96 * SE_{\bar{x}}$$

but… the OR sampling distribution is **right skewed** not Normally distributed!

What do we do?

Confidence Interval Odds Ratio:

Step 3 (determining if it is statistically significant or not):

General approach involves **Transformation (let R handle it!)**:

- ln(OR) ~ Normally distributed
- Confidence Interval boundaries are found
    - Calculate S.E., use Z value corresponding to stated α
- Converted back using **exponential distribution**

Example: ***Step 1:*** *Odds ratio = (a/c)/(b/d) = x.xx*

Step 2: *Calculate ln(OR):*

$$\ln(x.xx) = y.yy$$

Step 3: *The confidence interval for* ***ln (OR)*** *is a normally distributed sampling distribution (unlike the confidence interval for OR). This means that we can use* ***Z****.*

So, for a 95% confidence interval (α = 0.05), we can use 1.96.

$$\ln(\widehat{OR}) - 1.96 * \ln(SE_{OR}) < \ln(OR) < \ln(\widehat{OR}) + 1.96 * \ln(SE_{OR})$$

then you need to transform the lower and upper boundaries back into the original scale:

$$e^{-lowerboundary} < OR < e^{-upper\ boundary}$$

Confidence Interval Odds Ratio:

Step 4:

- **Conclude:**

OR = 1, If the 95% (or 99%) Confidence interval contains 1, the indicates that there is no association at the 5% (or 1%) significant level.

If the 95% (or 99%) Confidence interval does not contain 1 then we can conclude that there is statistically significant (at the 5% or 1% level) association between the variables (i.e.. lack of disease and treatment etc.)