Odds Ratio:

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:

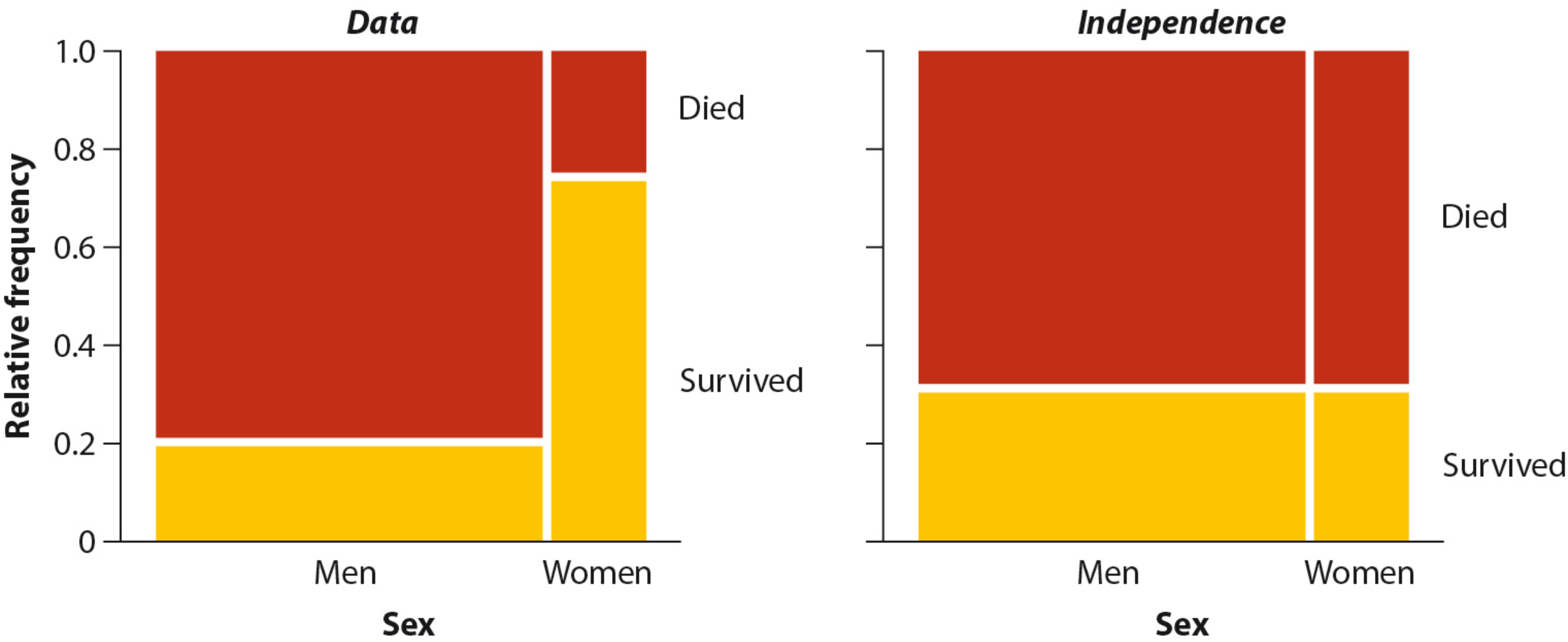
Probability of success divided by the probability of failure

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

As per usual, we will be using estimates:

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

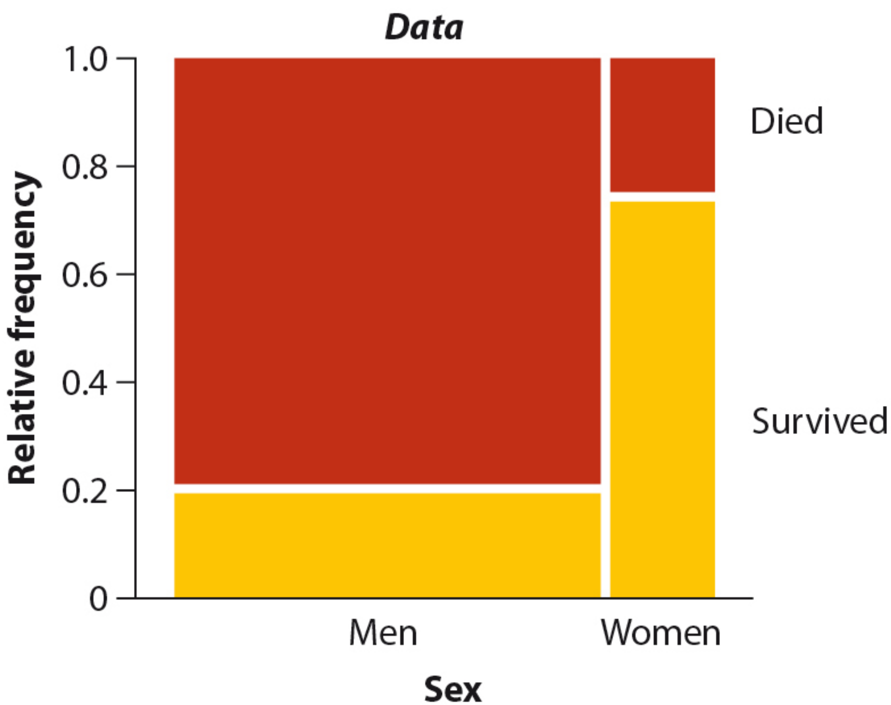
Contingency Analysis



Odds:

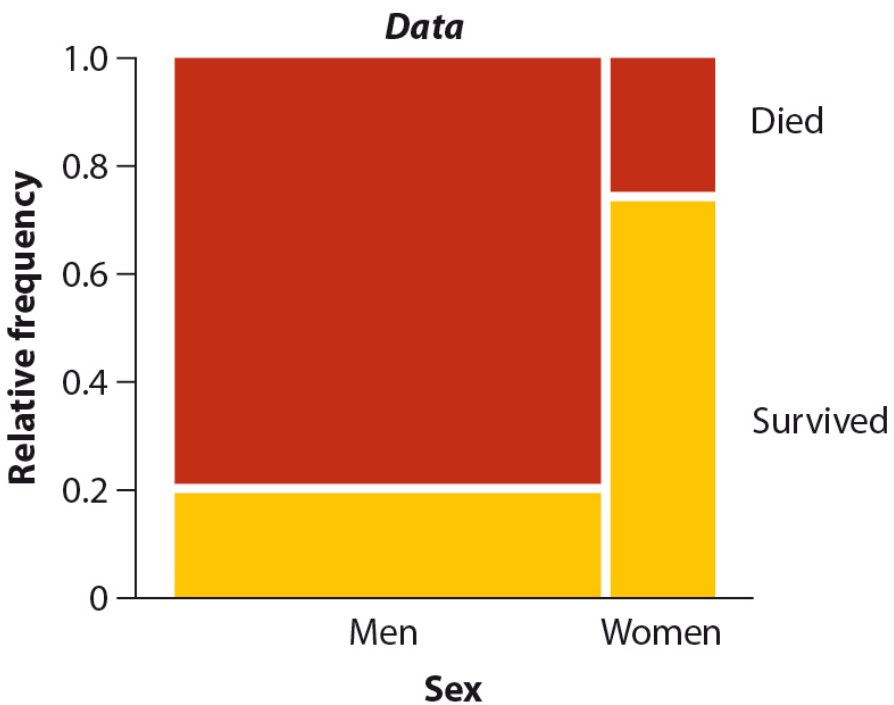
Probability of success divided by the probability of failure

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



Odds:

Probability of success divided by the probability of failure



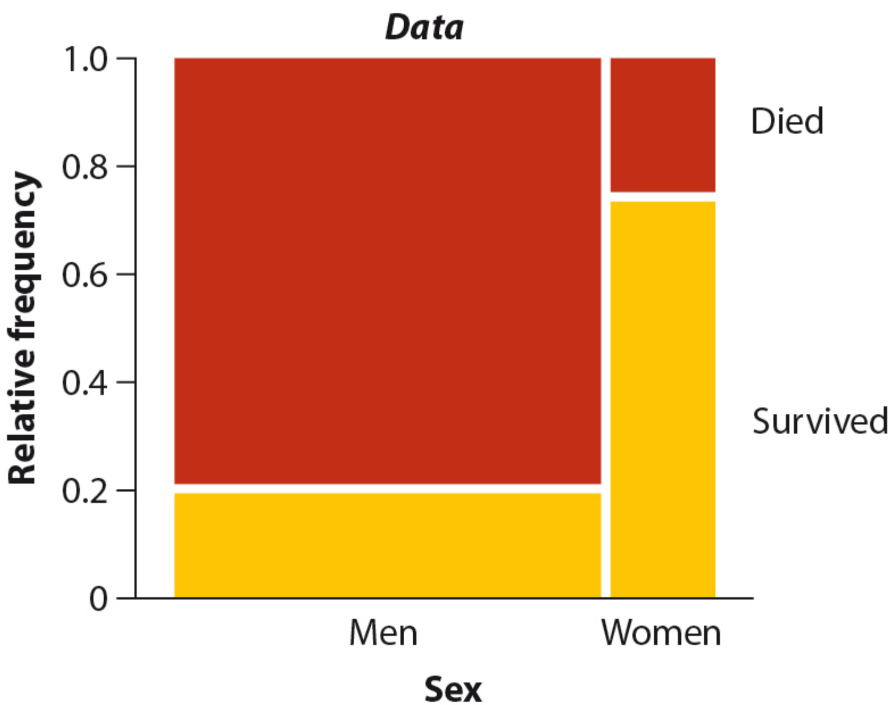
$$O = \frac{p}{1-p}$$

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

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

Odds:

Probability of success divided by the probability of failure



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

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

1 to 4

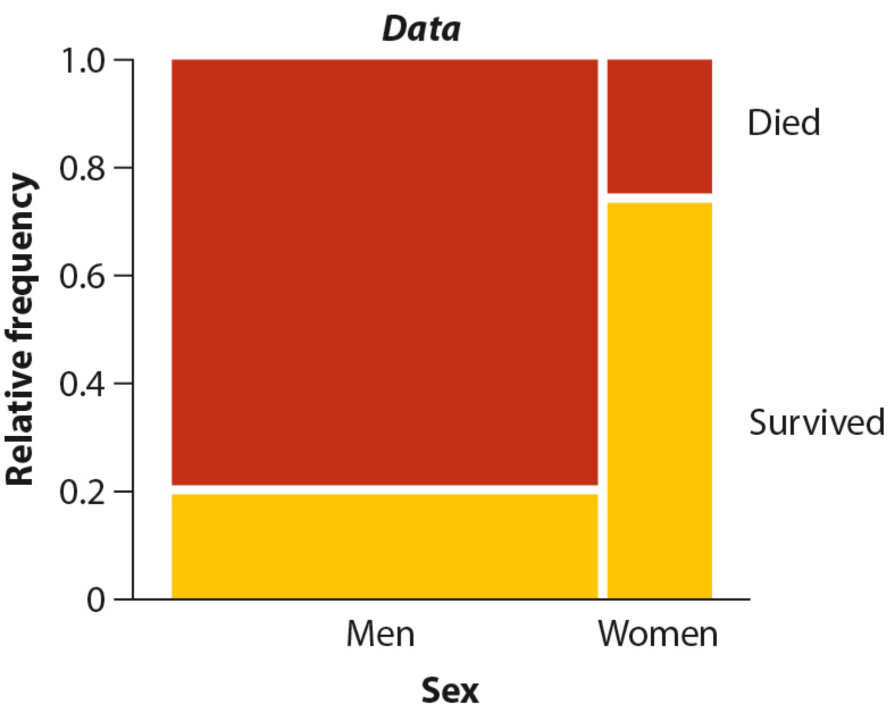
$$O_{women} = \frac{0.74}{1-0.74} = 2.85$$

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)?*

$$\hat{O}R = \frac{\hat{O}_1}{\hat{O}_2} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

	Treatment	Control
Success	a	b
Failure	c	d

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 we have discussed so far:

Point Estimate ± Z*Standard Error

but… the OR sampling distribution is right skewed

What do we do?

Confidence Interval Odds Ratio:

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

- ln(OR) ~ Normally distributed
- Confidence Interval boundaries are found
  - Calculate S.E.
- Converted back using **exponential distribution**

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 (ie. lack of disease and treatment etc)

Revisit this Example:

The influence of SES on preterm delivery rates

Socio-Economic status	Cases	Controls
Upper	11	40
Upper-middle	14	45
Middle	33	64
Lower-middle	59	91
Lower	53	58
Unknown	5	5

You could work through this data set (used in Module_4_video_12B). This time, focus on the two extreme categories, Upper and Lower, so you can use Odds Ratio instead and compare your answer to the X^2 Contingency Analysis result.

Odds Ratio:

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**

The most challenging parts of an odds-ratio:

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

Confidence Interval Odds Ratio:

Example: The influence of SES on preterm delivery rates

Step 1: Odds ratio = (53/58)/(11/40) = 3.32

Step 2: Calculate ln(OR):

ln(3.32) = 1.20

Step 3: The confidence interval for the **ln (OR)** is a normally distributed sampling distribution, we can use **Z*** . So for a 95% confidence interval ($\alpha = 0.05$), we can use 1.96.

<u>Confidence Interval Odds Ratio:</u>

<u>Example:</u> The influence of SES on preterm delivery rates

Step 1: Odds ratio = (53/58)/(11/40) = 3.32

Step 2: Calculate ln(OR):

ln(3.32) = 1.20

Step 3: The confidence interval for the ln (OR) is a normally distributed sampling distribution, we can use **Z** . So for a 95% confidence interval ($\alpha = 0.05$), we can use 1.96.

Step 4: Calculate SE

SE[ln(OR)] = sqrt(1/53 + 1/58+1/11+1/40) = 0.39

Confidence Interval Odds Ratio:

Example: The influence of SES on preterm delivery rates

Step 1: Odds ratio = (53/58)/(11/40) = 3.32
Step 2: Calculate ln(OR):

$$\ln(3.32) = 1.20$$

Step 3: The confidence interval for the ln (OR) is a normally distributed sampling distribution, we can use **z** . So for a 95% confidence interval ($\alpha = 0.05$), we can use 1.96.
Step 4: Calculate SE

$$SE[\ln(OR)] = \mathrm{sqrt}(1/53 + 1/58+1/11+1/40) = 0.39$$

Step 5:

$$1.20 \pm 1.96*0.39 = (0.44, 1.97)$$

Step 6: CONVERT

$$(e^{0.44}, e^{1.97}) = (1.55, 7.14)$$

Confidence Interval Odds Ratio:

Example: The influence of SES on preterm delivery rates

Step 1: Odds ratio = (53/58)/(11/40) = 3.32

Step 2: Calculate ln(OR):

$$\ln(3.32) = 1.20$$

Step 3: The confidence interval for the ln (OR) is a normally distributed sampling distribution, we can use **z** . So for a 95% confidence interval ($\alpha = 0.05$), we can use 1.96.

Step 4: Calculate SE

$$SE[\ln(OR)] = \text{sqrt}(1/53 + 1/58+1/11+1/40) = 0.39$$

Step 5:

$$1.20 \pm 1.96*0.39 = (0.44, 1.97)$$

Step 6: CONVERT

$$(e^{0.44}, e^{1.97}) = (1.55, 7.14)$$

Step 7: Interpret/conclude

The odds ratio for pre-term delivery for Lower SES compared to the Upper SES is 3.32 indicate increased odds of pre-term delivery for mothers in Lower SES. The 95% Confidence Interval of the Odds Ratio (1.55 , 7.14) indicates that odds of pre-term delivery are significantly higher for the Lower SES group compared to the Upper SES group (at 0.05 significance level) because the CI does not contain 1.

Relative Risk

- Another commonly used measure of association between two categorical variables (when both have two categories)

- Especially appropriate for comparing risk of rare and undesirable outcome ie. SID syndrome in children who sleep facing upward and those who sleep on their stomach (as long as babies are randomly sampled)

- Similar to odds ratio but perhaps more intuitive

- Should give similar answer as OR for rare events

$$\hat{R}R = \frac{\hat{p}_1}{\hat{p}_2} = \frac{'treatment'}{'control'}$$

Relative Risk

Example:

Given, event B = diagnosis of breast cancer w/i 2 years

event A = positive mammogram at present

event A^C = negative mammogram at present

P(diagnosis of breast cancer w/i 2 years|negative mammogram) = 20/100,000

P(diagnosis of breast cancer w/i 2 years|positive mammogram)=1/10

$$R\hat{R} = \frac{P(B \mid A)}{P(B \mid A^C)} = \frac{1/10}{2/10,000} = 500$$

Individuals with a positive mammogram at present have a relative risk of developing breast cancer within two years that is 500 times those of individuals with a negative mammogram.

Fisher's Exact Test:

- 2 x 2 contingency analysis
  - based on **hypergeometric** distribution with four classes

  - Answers the question: given two-way tables with the same fixed margin totals as the observed one, what is the chance of obtaining the observed cell frequencies *a,b,c* and *d and all cell frequencies that represent a greater deviation from expectation?* ***NOTE:*** This is a similar definition to the P-value!

- No assumptions about size of expectations

- cumbersome to do it by hand (use R)

```
> fisher.test(matrix-data)
```

The total number of ways in which a two way table with fixed marginal totals can be obtained is:

$$\binom{n}{(a+b)}\binom{n}{(a+c)} = \frac{n!}{(a+b)!(c+d)!} \times \frac{n!}{(a+c)!(b+d)!}$$

Leads to the probability of obtaining a 2x2 table with the frequencies a,b,c and d:

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

Example: All but 28 trees of two species of acacia, species A and B, were cleared from an area in Central America. These trees were un-infested (no ant colonies). Next 16 different colonies of ants from species X from an area nearby, were brought in and placed equidistant from the 28 acacia trees. The ant colonies had been harvested from cut-down trees of species A.

Contingency Analysis

<u>Example:</u> All but **28** trees of two species of acacia, species **A** and **B**, were cleared from an area in Central America. These trees were uninfested (no ant colonies). Next **16** different colonies of ants from species **X** from an area nearby, were brought in and placed equidistant from the **28** acacia trees. The ant colonies had been harvested from cut-down trees of species **A**.

a	**b**	**a+b**
<u>c</u>	**<u>d</u>**	**<u>c+d</u>**
a+c	**b+d**	**a+b+c+d**

Contingency Analysis

<u>Example:</u> All but **28** trees of two species of acacia, species **A** and **B**, were cleared from an area in Central America. These trees were uninfested (no ant colonies). Next **16** different colonies of ants from species **X** from an area nearby, were brought in and placed equidistant from the **28** acacia trees. The ant colonies had been harvested from cut-down trees of species **A**.

a	b	a+b
<u>c</u>	<u>d</u>	<u>c+d</u>
a+c	b+d	a+b+c+d

Species	Not Invaded	Invaded	Total
A	2	13	**15**
B	10	3	**13**
Totals	**12**	**16**	**28**

Contingency Analysis

<u>Example:</u> All but **28** trees of two species of acacia, species **A** and **B**, were cleared from an area in Central America. These trees were uninfested (no ant colonies). Next **16** different colonies of ants from species **X** from an area nearby, were brought in and placed equidistant from the **28** acacia trees. The ant colonies had been harvested from cut-down trees of species **A**.

1	14	15
<u>11</u>	<u>2</u>	<u>13</u>
12	16	28

0	15	15
<u>12</u>	<u>1</u>	<u>13</u>
12	16	28

Species	Not Invaded	Invaded	Total
A	2	13	**15**
B	10	3	**13**
Totals	**12**	**16**	**28**

Contingency Analysis

<u>Example:</u> All but **28** trees of two species of acacia, species **A** and **B**, were cleared from an area in Central America. These trees were uninfested (no ant colonies). Next **16** different colonies of ants from species **X** from an area nearby, were brought in and placed equidistant from the **28** acacia trees. The ant colonies had been harvested from cut-down trees of species **A**.

Species	Not Invaded	Invaded	Total
A	2	13	**15**
B	10	3	**13**
Totals	**12**	**16**	**28**

```
> fisher.test(acacia)
      Fisher's Exact Test for Count Data
data:  acacia
p-value = 0.001624
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.003786123 0.425475250
sample estimates:
odds ratio
0.05401494
```

G-Tests:

- Similar to χ^2 test
- Uses a longer formula (involving **ln**)
- G-test statistic uses χ^2 distribution with *df* = (r-1)(c-1)
- Applied over wider range of circumstances