Proportions

- Consider a *bernoulli trial* in which we have a binary response: \( y=1 \) indicates "yes" and \( y=0 \) indicates "no"
- \( \pi \) = true proportion of "yes"; also probability of "yes" for a hypothetical pop. or randomized exp.
- Mean of \( y \) is \( \pi \); Variance of \( y \) is \( \pi (1-\pi) \)
- Sample proportion is \( \hat{\pi} \)

Odds

- How to compare 2 proportions
  - differences? \( \pi_1 - \pi_2 \)
  - odds of "yes": \( \omega = \frac{\pi}{(1-\pi)} \)
  - odds ratio:
    \[
    \frac{\omega_2}{\omega_1} = \frac{\frac{\pi_2}{(1-\pi_2)}}{\frac{\pi_1}{(1-\pi_1)}}
    \]

Example, p. 526

<table>
<thead>
<tr>
<th>Placebo</th>
<th>Cold</th>
<th>No Cold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>335</td>
<td>76</td>
</tr>
<tr>
<td>Vitamin C</td>
<td>302</td>
<td>105</td>
</tr>
</tbody>
</table>

"The odds of a cold on placebo are 1.53 times the odds of a cold on Vitamin C."

Regression with a binary response

- Model considered to date is inappropriate for binary data: \( \mu[Y|X] = \alpha + \beta X \)
  - require fitted values between 0, 1
  - changes in \( X \) have less impact for values of response near 0 or 1 than near 0.5
  - variance of response depends on the mean!
- *Curvilinear* relationship between \( X \) and \( \pi \):
  \[
  \logit(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X
  \]
- For every unit increase in \( X \), the odds of "yes" increase multiplicatively by...
Spotted Owl Habitat

- Nesting locations of Northern Spotted Owl and availability of mature forests.
- 30 nest sites + 30 sites selected at random.
- Response variable: presence/absence of owl nest
- Explanatory variable: Percentage of mature forest (>80 years) in 7 rings of different radii (.91 km, 1.18 km, ..., 3.38 km)
- Case-control study; retrospective sampling

Univariate logistic regression

- Consider the model relating owl presence (1 or 0) to percentage of mature forest in ring 1 (0.91 km)
  \[ \text{logit}(\pi) = -4.425 + 0.0617 \text{ percentage} \]
- For each additional 10% of old forest in ring 1, what is the estimated increase in the odds that the site is a nest site?
- Odds of nest site for 50% mature forest?
- Comparing 50% to 90% mature forest, what is the odds ratio?

Maximum likelihood estimation

- Simple example: Bernoulli Distribution
- Model a trial outcome Y as Bernoulli(\(\pi\))
  \[ P(Y = y) = \pi^y (1 - \pi)^{1-y} \]
- For \(n\) responses, and each outcome \(Y_i\) having unknown probability \(\pi_i\), we can write the probability for the set of outcomes in terms of the unknown parameters. This is known as the likelihood or probability of the data.
- Maximum likelihood estimation can be used to estimate the \(\pi_i\)'s.
### Sploplus Output

*** Generalized Linear Model ***

Call: glm(formula = owl ~ pctstring1, family = binomial(link = logit), data = Ex2015, na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))

Deviance Residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.950618</td>
<td>-1.080485</td>
<td>0.1622955</td>
<td>0.9529993</td>
<td>1.746862</td>
</tr>
</tbody>
</table>

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-4.42531863</td>
<td>1.58291121</td>
<td>-2.795683</td>
</tr>
<tr>
<td>pctstring1</td>
<td>0.06166194</td>
<td>0.02129754</td>
<td>2.895261</td>
</tr>
</tbody>
</table>

(Dispersion Parameter for Binomial family taken to be 1)

Null Deviance: 83.17766 on 59 degrees of freedom

Residual Deviance: 71.09694 on 58 degrees of freedom

Number of Fisher Scoring Iterations: 4