Paired data

Gather 100 rats and determine whether they are infected with viruses A and B.

<table>
<thead>
<tr>
<th>I-B</th>
<th>NI-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-A</td>
<td>9</td>
</tr>
<tr>
<td>NI-A</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$p_{00}$</td>
<td>$p_{01}$</td>
<td>$p_{0+}$</td>
</tr>
<tr>
<td></td>
<td>$p_{10}$</td>
<td>$p_{11}$</td>
<td>$p_{1+}$</td>
</tr>
<tr>
<td>B</td>
<td>$p_{+0}$</td>
<td>$p_{+1}$</td>
<td>1</td>
</tr>
</tbody>
</table>

Is the rate of infection of virus A the same as that of virus B?

In other words: Is $p_{1+} = p_{+1}$? Equivalently, is $p_{10} = p_{01}$?
McNemar’s test

$H_0: p_{01} = p_{10}$

Under $H_0$, e.g. if $p_{01} = p_{10}$, the expected counts for cells 01 and 10 are both equal to $(n_{01} + n_{10})/2$.

The $\chi^2$ test statistic reduces to $X^2 = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}}$

For large sample sizes, this statistic has null distribution that is approximately a $\chi^2(df = 1)$.

For the example: $X^2 = (20 - 9)^2 / 29 = 4.17 \rightarrow P = 4.1\%$. 
An exact test

Condition on $n_{01} + n_{10}$.

Under $H_0$, $n_{01} \mid n_{01} + n_{10} \sim \text{Binomial}(n_{01} + n_{10}, 1/2)$.

In R, use the function `binom.test`.

---

For the example, $P = 6.1\%$. 

### Paired data

<table>
<thead>
<tr>
<th></th>
<th>I-B</th>
<th>NI-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-A</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>NI-A</td>
<td>20</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>NI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>18</td>
<td>82</td>
</tr>
<tr>
<td>B</td>
<td>29</td>
<td>71</td>
</tr>
</tbody>
</table>

- $P = 6.1\%$
- $P = 9.5\%$

---

→ Taking appropriate account of the “pairing” is important!
Deviations from Random Coil Behaviour

Are there site-specific deviations from random coil dimensions?

Förster Resonance Energy Transfer enables us to measure the distance between two dye molecules within a certain range. This can be used to study site-specific deviations from random coil dimensions in highly denatured peptides.
Deviations from Random Coil Behaviour
We have two underlying distributions for the green and red photons:

- One stemming from a peptide only having a donor dye.
- One stemming from a peptide being properly tagged with a donor and an acceptor dye.

Assume a photon has probability $p_0$ of being red in the former situation, and $p_1$ in the latter.
Deviations from Random Coil Behaviour
Deviation from Random Coil Behaviour

Assume we observe $n_i$ photons at time point $i$. Then the number of red photons is simply Bernoulli$(n_i, p_i)$, where $p_i$ is either $p_0$ or $p_1$. Assume that the probability of observing photons from a peptide without an acceptor dye at any time is $p$, independent of the total number of photons observed. Let $X$ be the number of red photons. Then

$$P(X = x_i|n_i) = P(X = x_i|n_i, p_0) \times p + P(X = x_i|n_i, p_1) \times (1 - p)$$

$$= \binom{n_i}{x_i} p_0^{x_i} (1 - p_0)^{n_i-x_i} \times p + \binom{n_i}{x_i} p_1^{x_i} (1 - p_1)^{n_i-x_i} \times (1 - p),$$

and hence

$$L(p, p_0, p_1) = \prod_{i=1}^{N} \left[ \binom{n_i}{x_i} p_0^{x_i} (1 - p_0)^{n_i-x_i} \times p + \binom{n_i}{x_i} p_1^{x_i} (1 - p_1)^{n_i-x_i} \times (1 - p) \right].$$
Deviations from Random Coil Behaviour

![Graph showing deviations from random coil behaviour. The x-axis represents the total number of photons, ranging from 50 to 100, while the y-axis represents the number of red photons, ranging from 0 to 80. The graph displays a scatter plot with points dispersed across the chart, indicating the distribution of red photons relative to the total number of photons.](image-url)
Deviations from Random Coil Behaviour

![Graph showing the relationship between total number of photons and number of red photons. The graph includes a scatter plot with two trend lines, one in red and one in green. The x-axis represents the total number of photons, ranging from 50 to 100, and the y-axis represents the number of red photons, ranging from 0 to 80. The trend lines indicate a positive correlation between the two variables.]
Deviations from Random Coil Behaviour

$\hat{\rho}_1 = 0.431$