The simplest method for estimating population size is the Lincoln-Petersen method. The basic method involves a single marking and a single recapture episode. If we mark a sample of animals, $M$ and capture a sample of animals, $C$, containing $m$ marked individuals, we estimate the size of the population at the time of the original marking, $N$, as:

$$\hat{N} = \frac{MC}{m}$$
To understand how this formula works we reorganize to get:

\[
\frac{\hat{N}}{M} = \frac{C}{m}
\]

This formula says that the ratio of marked individuals in the original sample to total population size is the same as the ratio of marked individuals to total sample size in the second sample.
Note, if we estimate capture probability, $p$, on the second capture as:

$$\hat{p} = \frac{m}{M}$$

then we can estimate $N$ as:

$$\hat{N} = \frac{C}{\hat{p}}$$
The Lincoln-Petersen method requires several assumptions.

1. The population is closed, so that \( N \) is constant.

2. All individuals have the same probability of being captured in the first sample.

3. Marking animals does not affect catchability.

4. Individuals do not lose marks between the two sampling periods.
We can relax the first assumption somewhat, if we are only interested in the size of the population at the time the original marks were applied and marked and unmarked individuals leave the population at the same rate.

The original L-P estimator produces a biased estimate of population size. An approximately unbiased estimator (Seber 1982) is:

\[
\hat{N} = \frac{(M + 1)(C + 1)}{(m + 1)} - 1
\]
We can calculate two forms of confidence intervals for the ratio \( \frac{m}{C} \). The first is a normal approximation to the binomial:

\[
\frac{m}{C} \pm \left\{ z_\alpha \left[ \sqrt{\frac{(1 - f)\left(\frac{m}{C}\right)\left(1 - \frac{m}{C}\right)}{C - 1}} \right] + \frac{1}{2C} \right\}
\]

where \( z_\alpha \) is a standard normal deviate for a (1-\( \alpha \)) level of confidence (1.96 for 95% confidence limits) and \( f \) is the fraction of the total population (\( \frac{m}{M} \)) sampled in the second sample.
For example, if $M = 1800$ and CI for $m/C$ ranged from 0.07 to 0.11, we can calculate the CI for $\hat{N}$ as below:

$$\hat{N} = \frac{CM}{m}$$

$$\Rightarrow 1800/0.07 \rightarrow 1800/0.11$$

$$\Rightarrow 16364 < \hat{N} < 25714$$
If we relax the assumption of a closed population and allow for births and deaths we need a method of estimation that accounts for this additional complexity. The Jolly-Seber method was developed for situations in which multiple capture occasions were used to estimate a number of parameters, including population size.
### Mark-recapture data for a series of samples of field voles (Krebs 1999)

<table>
<thead>
<tr>
<th>Time of last capture</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>61</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>77</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- **Total marked (mt)**: 0, 15, 16, 37, 64, 79, 81, 76, 8, 15, 19
- **Total unmarked (u)**: 22, 26, 32, 45, 25, 22, 26, 15, 11, 12, 3
- **Total caught (nt)**: 22, 41, 48, 82, 89, 101, 107, 91, 19, 27, 22
- **Total released (s)**: 21, 41, 46, 82, 88, 99, 106, 90, 19, 26, 22
The parameters we need for our calculations are:

- $m_t$ Number of marked animals caught in sample $t$
- $u_t$ Number of unmarked animals caught in sample $t$
- $n_t$ total number caught in sample $t = m_t + u_t$
- $s_t$ total number released after sample $t$
  - $(n_t – \text{accidental deaths or removals})$
- $m_{rt}$ Number marked caught in sample $t$ that were last caught in sample $r$
- $R_t$ Number of the $s_t$ inds. released in sample $t$ that were caught again in a later sample
- $Z_t$ number marked before sample $t$, not caught at $t$, but caught in some sample after $t$
Our estimate of population size is:

\[ \text{Popsize} = \frac{\text{Size...marked...pop}}{\text{proportion...marked}} \]

\[ \text{Prop...marked} = \hat{\alpha} = \frac{m_t + 1}{n_t + 1} \] (+1 corrects for bias in small samples)

\[ \hat{M} = \frac{(s_t + 1)Z_t}{R_t + 1} + m_t \]

\[ \hat{M} \] is estimated size of the marked population just before \( t \).

The number of marked animals present at \( t \) is the sum of the number marked that were caught at \( t \) plus the number of marked animals that were alive but not caught at \( t \).
Our estimate of population size is therefore:

\[ \hat{N}_t = \frac{\hat{M}_t}{\hat{\alpha}_t} \]

Note that the ratio, \((R_t+1)/(s_t+1)\) is like a crude survival estimate, \(\). In fact J-S methods were the first statistically appropriate survival estimates for capture-recapture data.
One estimate of population size for $t = 6$ would be:

\[ \hat{\alpha}_6 = \frac{79 + 1}{101 + 1} = 0.784 \]

\[ \hat{M}_6 = \frac{(99 + 1)(7)}{81 + 1} + 79 = 87.54 \]

\[ \hat{N}_6 = \frac{87.54}{0.784} = 111.6 \]
An important assumption is that all individuals are equally catchable and there are several tests of this assumption. We will explore one such test, using the Poisson distribution. If all individuals are equally catchable, the distribution of numbers of individuals caught \( n \) times should be distributed as a Poisson distribution. We can then use a goodness of fit (this will be important later) to compare the observed distribution of numbers of captures against that predicted by the Poisson. A significant test statistic tells us that individuals were not all equally catchable.

The Poisson has a distribution:

\[
p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}
\]

where \( \lambda \) is the mean (and the variance) of the distribution.
We are interested in the mean of the Poisson truncated to exclude 0 captures (we don’t care about those). The mean of this truncated distribution is (from Krebs 1999):

\[
\bar{x} = \frac{\lambda}{1 - e^{-\lambda}}
\]

We can use the mean number of captures of each individual in our sample to estimate the mean of the truncated distribution.

Consider the sample of captures on the next table.
<table>
<thead>
<tr>
<th>Number of captures</th>
<th>Number caught</th>
<th>Expected frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>184</td>
<td>174.6</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>16.7</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.7</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

These are capture patterns for hares reported by Keith and Meslow (1968) (in Krebs 1999).

We can use this estimate to estimate the mean of the underlying Poisson distribution.

\[
x = 1.4253 = \frac{\lambda}{1 - e^{-\lambda}}
\]

\[\Rightarrow \lambda = 0.7563\]
Expected number of hares caught once =

\[ f_1\left(\frac{e^{-\lambda} \lambda^1}{(1 - e^{-\lambda})1!}\right) = 261\left(\frac{e^{-0.7563} \ 0.7563}{1 - e^{-0.7563} \ 1}\right) = 174.6 \]

Caught twice

\[ f_1\left(\frac{e^{-\lambda} \lambda^2}{(1 - e^{-\lambda})2!}\right) = 261\left(\frac{e^{-0.7563} \ 0.7563^2}{1 - e^{-0.7563} \ 2}\right) = 66.0 \]

Using the same method, the numbers expected to be caught 3 and > 3 times are 16.7 and 3.7, respectively.
We use a $\chi^2$ goodness-of-fit test to compare the observed distribution of captures against that expected under a truncated Poisson distribution.

$$
\chi^2 = \sum \left[ \frac{(O - E)^2}{E} \right]
$$

$$
= \frac{(184 - 174.6)^2}{174.6} + \frac{(55 - 66.0)^2}{66.0} + \frac{(14 - 16.7)^2}{16.7} + \frac{(8 - 3.7)^2}{3.7}
$$

$$
= 7.77
$$

$$
df = (n - 2) = 4 - 2 = 2
$$

$P < 0.05$, that a value this large was generated by chance. Thus, we reject the hypothesis of equal catchability.