Inference on Arthropod Demographic Parameters: Computational Advances Using R

ALINE DE HOLANDA NUNES MAIA,1,2 RICARDO ANTONIO DE ALMEIDA PAZIANOTTO,1 ALFREDO JOSÉ BARRETO LUIZ,4 JEANNE SCARDINI MARINHO-PRADO,1 AND AHMAD PERVEZ3

J. Econ. Entomol. 107(1): 432–439 (2014); DOI: http://dx.doi.org/10.1603/EC13222

ABSTRACT We developed a computer program for life table analysis using the open source, free software programming environment R. It is useful to quantify chronic nonlethal effects of treatments on arthropod populations by summarizing information on their survival and fertility in key population parameters referred to as fertility life table parameters. Statistical inference on fertility life table parameters is not trivial because it requires the use of computationally intensive methods for variance estimation. Our codes present some advantages with respect to a previous program developed in Statistical Analysis System. Additional multiple comparison tests were incorporated for the analysis of qualitative factors; a module for regression analysis was implemented, thus, allowing analysis of quantitative factors such as temperature or agrochemical doses; availability is granted for users, once it was developed using an open source, free software programming environment. To illustrate the descriptive and inferential analysis implemented in lifetable.R, we present and discuss two examples: 1) a study quantifying the influence of the proteinase inhibitor berenil on the eucalyptus defoliator Thyrinteina arnobia (Stoll) and 2) a study investigating the influence of temperature on demographic parameters of a predaceous ladybird, Hippodamia variegata (Goeze).

KEY WORDS fertility life table, intrinsic rate of increase, uncertainty assessment, linear regression

Life table parameters are important quantitative indicators of interactions between arthropod populations and the environment. By summarizing information on both fertility and survivorship, they can assess chronic sublethal effects that are undetected by acute survival assays. For example, by using life table analysis, Marinho-Prado (2011) quantified effects of the proteinase inhibitor berenil on the eucalyptus defoliator Thyrinteina arnobia (Stoll); Nascimento et al. (1998) and Nardo et al. (2001) assessed the effect of Bacillus thuringiensis Berliner formulation on Podisus nigrispinus (Dallas), predator of the soybean caterpillar; Liu et al. (2005) investigated the influence of Bt cotton plants on the nontarget pest Aphis gossypii (Glover); and Lumbierres et al. (2004) evaluated the performance of Rhopalosiphum padi (L.) (Hemiptera: Aphididae) on Bt maize.

In life table studies, oviposition and survival data are collected over time (usually daily) and summarized into fertility life tables (FLT) for later estimation of the following parameters for each treatment evaluated: net reproductive rate (Ro), intrinsic rate of increase (Rm), doubling time (DT), mean generation time (TMG), and finite rate of increase (λ). For detailed descriptions of FLT parameters see Maia et al. (2000). Studies for FLT parameter estimation are developed in laboratory conditions, usually following a completely randomized design, for which the experimental unit is an insect or mite female. The units are followed from their onset of oviposition until their death to record longevity and the number of eggs laid per female in each time interval (usually a day).

As FLT parameters summarize data from experimental units into a single estimate for each group (treatment), the information on within treatment variance is not readily available, thus requiring the use of computationally intensive methods for its estimation. Among them, the jackknife method as proposed by Meyer et al. (1986) is the most widely used method worldwide for variance estimation in FLT analysis (Maia et al. 2000). Variance estimates are required for any inferential procedure such as constructing confidence bands, performing multiple comparison tests for qualitative treatments, or fitting regression models for describing relationships between FLT parameters and quantitative factors (e.g., temperature, pesticide level). Hulting et al. (1990) developed a Pascal code to obtain jackknife estimates of all FLT parameters, confidence bands for Rm, and tests for comparing Rm among groups. That computer program, however, presents some limitations as pointed out by Maia et al. (2000): 1) the preimaginal survivorship is set as common to all groups; 2) the algorithm for estimating sex ratio is not adequate when some females laid no eggs.

1 Embrapa Meio Ambiente, P.O. Box 69, Jaguariúna, SP, Brazil.
2 Corresponding author, e-mail: aline.maia@embrapa.br.
3 Biocontrol Laboratory, Department of Zoology, Govt. Degree College, Talwari-Tharali, Chamoli, Uttarakhand, India.

0022-0493/14/0432-0439$04.00/0 © 2014 Entomological Society of America
Table 1. Information required for construction of fertility life tables for arthropods

<table>
<thead>
<tr>
<th>Information</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of females per group ($n_g$)</td>
<td>No. of mated females in each group $g$ ($g = 1, 2, \ldots, G$).</td>
</tr>
<tr>
<td>Time to onset of adulthood ($t^{(0)}_{gi}$)</td>
<td>Time span between the day of oviposition that originated a female $i$ ($i = 1, 2, \ldots, n_g$), of a group $g$ and the day when that female becomes an adult.</td>
</tr>
<tr>
<td>Longevity ($t^{(2)}_{gi}$)</td>
<td>Time span between oviposition date and death for female $i$ of group $g$.</td>
</tr>
<tr>
<td>No. of eggs ($y_{g,at}$)</td>
<td>No. of eggs laid per female $i$ in the group $g$ at each time interval $t$ between ($t^{(2)}_{gi}$).</td>
</tr>
<tr>
<td>Sex ratio ($SR_g$)</td>
<td>Proportion of female offspring in the group $g$. It is estimated by egg sampling for sex identification at adult stage.</td>
</tr>
<tr>
<td>Immature stage survivorship ($ISS_g$)</td>
<td>Percent of female offspring which survive until mating in group $g$, also estimated by sampling.</td>
</tr>
</tbody>
</table>

Materials and Methods

In this article, we consider analyses of life table data coming from completely randomized designs for which $n$ experimental units (females) are randomly allocated to $G$ groups (treatments). Females are mated and then followed until their death. Survivorship and oviposition data are recorded at regular time intervals (days). For detailed description see Maia et al. (2000).

Fertility Life Tables. The R computer program lifeTable.r reads an Excel spreadsheet (information summarized in Table 1) and uses algorithms for life table calculation corresponding to each treatment in two different ways: using complete data or data coming from each jackknife iteration ($i$).

An FLT is composed of the following columns: 1) treatment or group identification ($G$); 2) female pivotal age ($t$); 3) cumulative survivorship at age $t$ ($L_t$); 4) mean number of female offspring produced per female at each oviposition date ($M_t$); 5) net number of female offspring produced per female at each oviposition date ($L_t M_t$); and 6) the product $t^* L_t^* M_t$ that is used to obtain an approximate estimate for mean generation time ($MGT$).

We obtain estimates for each FLT parameter ($Ro$, $Rm$, $DT$, $T$, and $\lambda$) in each treatment by summarizing information coming from the corresponding life tables. Initially, we use iterative methods for estimating $Rm$ and subsequently $DT$, $T$, and $\lambda$, which are derived from that rate. $Ro$ is not affected by the method adopted for $Rm$ estimation (approximate or iterative). Details on FLT parameter estimation using the iterative process are presented in Maia et al. (2000).

In lifetable.sas, we obtain $Rm$ iteratively from equation 1. The optimal value of $r$ corresponds to the one that minimizes the difference between the two sides of equation 1. As initial value, of $r$, we use the $Rm$ estimate obtained via the approximated method; the search is made in the interval between $0.80$ and $1.20$ $Rm$.

\[
\sum_{i=0}^{s(n)} e^{-s(n)} \times Ro_{g,at} = 1 \tag{1}
\]

In which $t^{(0)}_{gi}$ and $t^{(2)}_{gi}$ are the time to onset of adulthood and the time to death, respectively, for female $i$ ($i = 1, 2, \ldots, n_g$) in group $g$; $r_{g,at}$ is the root of equation 1 for group $g$ to be estimated via the iterative process;
and Ro is the net number of females produced by female i in group g at the pivotal age t.

In the R program, we obtained Ro estimates using the \textit{uniroot} function based on Brent’s iterative method (Brent 1973), for which discretization of the searching interval is not required. At this stage, we construct graphs showing the oviposition patterns across time and box and whiskers plots for total number of eggs laid per female in each group.

\textbf{Calculation of Pseudovalues.} We calculate pseudo-values for each FLT parameter in each jackknife step i by subtracting the estimate for step zero from the respective estimate for step i, using the number of females considered in each step as weights. For example, the Ro pseudo-value for group g step i (PSVRo \(_{g}^{(i)}\)) is given by:

\[
PSVRo_{g}^{(i)} = n_{g} \times Ro_{g}^{(0)} - (n_{g} - 1) \times Ro_{g}^{(0)} \quad [2]
\]

The calculation of pseudovalues is performed in a similar way for the remaining FLT parameters, thus resulting in \(n_{p}\) pseudovalues for each parameter after completion of the jackknife procedure. Using pseudovalues, we obtain estimates of uncertainty measures such as variance or SE for each parameter estimated in each group.

At this stage, we construct box-and-whiskers-plots to illustrate the dispersion of pseudovalues within each group, which is directly related to the uncertainty of the respective estimates. The program generates a table with jackknife estimates of FLT parameters and respective SEs for each group.

\textbf{Qualitative Factors: Comparing Parameters Between Groups.} For qualitative factors, for example, diet types (Nascimento et al.1998, Nardo et al. 2001), we implemented the following statistical tests: 1) t-test for pair-wise comparisons, using Siddak’s adjustment for the significance level in cases with more than two treatments (\(G > 2\)) as available in lifetable.sas; 2) Tukey and Duncan multiple comparison tests; 3) Dunnett test to compare treatments with a control; and 4) F test for contrasts established a priori. Notice that cases 2 to 4 require that homoscedasticity is assumed.

\textbf{Quantitative Factors: Fitting Polynomial Regression Models.} For quantitative factors, for example, temperature levels (Laumann et al. 2008) or protease inhibitor concentration (Marinho-Prado 2011), we developed codes for linear regression fitting, including diagnostics of residuals and influence analysis.

After model fitting, the R program generates: 1) graphs showing pseudovalues, fitted models and respective 95% confidence bands for each FLT parameter and 2) a table presenting estimates of linear model parameters, respective SEs, and nominal significance levels associated to respective hypothesis tests.

\section*{Results and Discussion}

The R computer program presents some advantages: beyond being easily available to users, it includes modules for analysis that are not available in lifetable.sas. The use of Brent’s iterative method for obtaining \(Rm\) from equation 1 makes the processing of root searching more efficient when compared with the one used in lifetable.sas. The program lifetable.r and the data file of case study 1 presented here are available for download at http://www.cnpm.ebraspa.br/forms/Rlifetable.php3.

To illustrate the use of methods implemented into lifetable.R, plus to interpret and discuss outputs, we

\begin{table}[h]
\centering
\begin{tabular}{lcccc}
\hline
\textbf{Parameter} & \textbf{Effect} & \textbf{Estimate} & \textbf{SE} & \textbf{t-statistic} & \textbf{P-value} \\
\hline
Net reproductive rate & \(b_0\) & 223.95 & 15.59 & 12.05 & <0.0001 \\
& \(b_1\) & -1010.48 & 265.28 & -3.77 & 0.0009 \\
Mean generation time & \(b_0\) & 55.32 & 1.91 & 28.94 & <0.0001 \\
& \(b_1\) & 69.17 & 27.59 & 2.51 & 0.0190 \\
Doubling time (DT) & \(b_0\) & 7.63 & 0.33 & 21.12 & <0.0001 \\
& \(b_1\) & -19.52 & 4.8 & 4.06 & 0.0004 \\
Finite rate of increase & \(b_0\) & 0.074 & 0.004 & 283.32 & <0.0001 \\
& \(b_1\) & -0.192 & 0.051 & -3.71 & 0.0010 \\
Intrinsic rate of increase (Rm) & \(b_0\) & 1.102 & 0.004 & 27.33 & <0.0001 \\
& \(b_1\) & -0.208 & 0.056 & -3.72 & 0.0010 \\
\hline
\end{tabular}
\caption{Estimates of the linear model coefficients (\(\beta_0\), intercept; \(\beta_1\), slope) used to represent the influence of berenil concentration (0, 0.06, and 0.12%) on FLT parameters of \textit{T. arnobia}.}
\end{table}

\section*{Appendix A}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Oviposition patterns of \textit{T. arnobia} for each berenil concentration (0, 0.06, and 0.12%) across time.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2.png}
\caption{Box and whisker plots showing total number of eggs laid per female \textit{T. arnobia} for each berenil concentration (0, 0.06, and 0.12%). The bottom and top edges of the box are located at the 25th and 75th percentiles. The center horizontal line is drawn at the median. The vertical dashed lines extend from the box as far as the data extend to a distance at most 1.5 interquartile ranges. Circles represent outliers.}
\end{figure}
Fig. 3. Fitted linear regression models used to describe berenil effect on life table parameters of the eucalyptus defoliator caterpillar, *T. arnobia*: (a) net reproductive rate (*Ro*); (b) mean generation time (*MGT*); (c) doubling time (*DT*); (d) finite rate of increase (*λ*); and (e) intrinsic rate of increase (*Rm*). Dashed lines represent 95% confidence bands for predicted values.

Fig. 4. Oviposition patterns of *H. variegata* for each temperature (20, 23, 25, 27, and 30°C) across time.
The influence of berenil in *T. arnobia* fertility is observed in the delay in the oviposition period. Oviposition did not clearly show the expected triangular pattern (Dixon and Agarwala 2002). High variability in fecundity among females or oviposition dates was observed (Figs. 1 and 2). Fecundity data are presented in a box-plot for visualization of mean, median, variability, and asymmetry of fertility distributions (Fig. 2).

As we have quantitative treatments, the appropriate way of exploring treatment effects on FLT parameters is via regression models instead of multiple comparison tests. In the current example, because we have only three concentrations under test, we have proposed a simple linear model. In Table 2, we present parameter estimates for \( \beta_0 \) (intercept) and \( \beta_1 \) (slope) of the fitted linear models for \( R_0, MGT, DT, \lambda, \) and \( R_m \). We observe positive slope estimates for \( MGT \) and \( DT \) and negative ones for the remaining parameters (Fig. 3); corresponding nominal significance levels (\( P \) values) < 0.001 indicate a reduction of the potential population growth as the berenil concentration increases. Such results are consistent with the expected effect for a protease inhibitor and they show that sublethal concentrations of berenil effectively harmed population parameters of *T. arnobia*.

Thus, sublethal concentrations of berenil could be used for pest control, interfering with the population growth rates of insect pests but without causing high insect mortality, which may minimize the risk of development of resistant insect biotypes (Wolfson and Murdock 1995).

**Case Study II: Effect of Temperature on *H. variegata*.** *H. variegata* is an important aphidophagous ladybird with Palearctic origin and cosmopolitan distribution (Krafsur et al. 1996, Franzmann 2002, Omkar and Pervez 2004). It could potentially be used as a biocontrol agent for a variety of aphids (Nedved 1999); hence, information on its ecology and life history could present two case studies: 1) a study to evaluate the effect of berenil (a protease inhibitor) on FLT parameters of the eucalyptus defoliator *T. Arnobia* and 2) a study on the influence of temperature on the life table parameters of a predaceous ladybird, *H. variegata*.

**Case Study I: Effect of Berenil on *T. arnobia*.** Caterpillars of *T. arnobia* have been reported in most areas where *Eucalyptus* spp. is cultivated in Brazil (Holtz et al. 2003). This species’ outbreaks are believed to be favored by the extensive areas of eucalyptus plantations established in the country, where caterpillars find abundant and stable shelter and food supply (Santos et al. 1996). These conditions have allowed *T. arnobia* to cause serious damage to these plants. Berenil is an example of a synthetic trypsin inhibitor that may have potential for use as an insecticide (Oliveira et al. 1993).

![Fig. 5. Box and whisker plots showing total number of eggs laid per female *H. variegata* at constant temperatures (20, 23, 25, 27, and 30°C). The bottom and top edges of the box are located at the 25th and 75th percentiles. The center horizontal line is drawn at the median. The vertical dashed lines extend from the box as far as the data extend to a distance at most 1.5 interquartile ranges. Circles represent outliers.](image)

Table 3. Estimates of polynomial model coefficients (\( \beta_0 \), intercept; \( \beta_1 \), linear effect; \( \beta_2 \), quadratic effect; and \( \beta_3 \), cubic effect) used to represent the influence of temperature on *H. variegata* FLT parameters

<table>
<thead>
<tr>
<th>FLT parameter</th>
<th>Effect</th>
<th>Estimate</th>
<th>SE</th>
<th>t-statistic</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net reproductive rate (<em>R</em>&lt;sub&gt;0&lt;/sub&gt;)</td>
<td>( \beta_0 )</td>
<td>27332.73</td>
<td>3626.12</td>
<td>7.54</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>-3003.60</td>
<td>444.71</td>
<td>-7.86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>148.33</td>
<td>17.98</td>
<td>8.42</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>-2.05</td>
<td>0.2396</td>
<td>-8.85</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mean generation time (<em>MGT</em>)</td>
<td>( \beta_0 )</td>
<td>-719.39</td>
<td>208.25</td>
<td>-3.45</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>101.73</td>
<td>25.54</td>
<td>3.98</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>4.33</td>
<td>1.0025</td>
<td>-4.20</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>0.059</td>
<td>0.0138</td>
<td>4.31</td>
<td>0.0001</td>
</tr>
<tr>
<td>Doubling time (<em>DT</em>)</td>
<td>( \beta_0 )</td>
<td>-119.1</td>
<td>30.40</td>
<td>-3.92</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>16.98</td>
<td>3.73</td>
<td>4.56</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>-0.743</td>
<td>0.1507</td>
<td>-4.91</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>0.0013</td>
<td>0.0020</td>
<td>5.14</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Finite rate of increase (<em>( \lambda )</em>)</td>
<td>( \beta_0 )</td>
<td>7.46</td>
<td>0.9790</td>
<td>7.64</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>-0.8136</td>
<td>0.1201</td>
<td>-6.78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>0.0339</td>
<td>0.0048</td>
<td>6.99</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>-0.00046</td>
<td>0.00006</td>
<td>-7.11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intrinsic rate of increase (<em>R</em>&lt;sub&gt;m&lt;/sub&gt;)</td>
<td>( \beta_0 )</td>
<td>5.59</td>
<td>0.8390</td>
<td>6.59</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_1 )</td>
<td>-0.6041</td>
<td>0.1029</td>
<td>-6.75</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_2 )</td>
<td>0.029</td>
<td>0.0042</td>
<td>6.97</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>( \beta_3 )</td>
<td>-0.00039</td>
<td>0.00006</td>
<td>-7.10</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* Nominal significance level.
be of great potential value. Temperature has a major impact on the demographic attributes of arthropods, especially ladybirds (Pervez and Omkar 2004).

In this case study, we explore the analysis of temperature influence (20, 23, 25, 27, and 30°C) on FLT parameters of H. variegata. Temperature-dependent oviposition patterns (Fig. 4) in H. variegata revealed a triangular fecundity shape (Dixon and Agarwala 2002). The age-specific fecundity was temperature driven and shifted to the younger side with the increase in temperature. Box and whisker plots (Fig. 5) reveal that fecundity increases with the increase in temperature up to 27°C and then declines with further increase in temperature, also showing a temperature dependent between-female variability.

As in case study I, we used regression methods to quantify the effect of a quantitative factor (temperature) on demographic parameters. Because we have a high number of levels of the quantitative factor, we proposed a cubic polynomial model. Estimates of the coefficients of the polynomial model proposed to represent the influence of temperature on H. variegata FLT parameters are shown in Table 3. Intrinsic and finite rates of increase along with the net reproductive rate of H. variegata increased with temperature up to 27°C and then decreased (Fig. 6). Such a pattern
corresponds to negative estimates for $\beta_2$, positive ones for $\beta_3$, and negative ones for $\beta_4$. The reverse was observed in the case of “generation time” and “doubling time” (Fig. 6; Table 3).

We are using polynomial models as an approximation to nonlinear models, which are more adequate to describe temperature effects. Procedures for parameter estimation are simpler and computational requirements are lower in polynomial models. However, they have the limitation of being valid only within the range of temperature levels used in the experiment: predicted patterns derived from polynomial models may deviate from the actual behavior of temperature influence for values out of that range. We are currently developing an R code for nonlinear regression fitting by combining the jackknife method with nonlinear procedures available in the nlme package.

Despite being widely used, the jackknife method itself presents limitations in cases for which the survival patterns or the distribution of fertility are highly asymmetrical. Whenever influence analysis shows critical outliers, the use of alternative methods such as parametric or nonparametric bootstrap are needed. These methods applied to life-table analysis are still not readily available for users, thus requiring research for both methodology and software development.

In this article, we present a computer program for the analysis of FLT that was developed using the R language and free software environment. It facilitates the objective quantification of the influence of treatments on FLT parameters that are important indicators of sublethal effects. We have developed the tools to perform inferential analysis for either quantitative or qualitative factors, an advance as compared with a similar SAS program in which only methods for qualitative factors were available. We also developed models to generate statistical reports containing life tables, parameter estimates, results of model fitting, or multiple comparison procedures as well as high quality graphs for most relevant outputs.

Acknowledgments
A. P. is thankful to University Grants Commission, New Delhi, India for financial assistance in the form Major Research Project (No. F. 41-18/2012 [SR]).

References Cited


Received 11 May 2013; accepted 14 October 2013.