Data Set 2: Parasite Eggs on Stinkbugs

Background

*Nezara viridula* (the southern green stinkbug, or green vegetable bug) is a major crop pest world-wide. It was unintentionally introduced to Hawai‘i in the early 1960s. Soon after it was discovered here the Hawai‘i Department of Agriculture began a biological control program against it, introducing several species of parasitoid insects, which unfortunately also attack native bugs such as the koa bug, *Coleotichus blackburniae*.

I have been studying the interaction of one of these introduced biological control agents, the feather-legged fly *Trichopoda pilipes*, with both *N. viridula* and *C. blackburniae*. One part of this research has involved simply collecting stinkbugs from natural populations and counting *T. pilipes* eggs on them. The distribution of eggs over hosts is informative about the behavior of the fly (e.g. does it detect and avoid stinkbugs which are already parasitized), and can strongly affect parasite reproduction and therefore host-parasite population dynamics.

The question:

*What is the shape of the distribution of eggs per host, and what is its spread?*

Note: Characterizing the distribution in the ways relevant to behavioral and population ecology involves specialized measures outside the material for this course. This handout therefore includes no statistical inference.

The data

21 adult stinkbugs were collected from weeds near Waipahu, on two mornings in June 1991. All eggs on each bug were counted. The following are the egg numbers:

```
  0   0   0   0   1   1   1
  2   2   2   3   3   3   5
  5   6   8   8   8   11  34
```
Data exploration

Displays

Histograms

The histogram to the right is what Minitab produces with its default settings. It incorrectly implies a frequency of 10 for the interval from -2.5 to 2.5, when obviously negative counts are not possible. The histogram to the left below fixes this problem by specifying “cut-points” starting at 0. Finally, the histogram below to the right has a bar for each discrete count, thus completely preserving the information in the data.

Stem-and-leaf plots

with outlier
0  000111222333
0  556888
1  1
1
2
2
3  4

without outlier
0  000011
0  222333
0  55
0  6
0  888
1  1

Boxplots

with outlier

without outlier

Data Set 2: Parasite Eggs on Stinkbugs
Normal quantile-quantile plots

Summary statistics

with outlier

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>Variance</th>
<th>CoefVar</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>4.90</td>
<td>7.39</td>
<td>54.59</td>
<td>150.64</td>
<td>0.00</td>
<td>1.00</td>
<td>3.00</td>
<td>7.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Maximum</th>
<th>Range</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>34</td>
<td>34</td>
<td>6.00</td>
</tr>
</tbody>
</table>

without outlier

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>StDev</th>
<th>Variance</th>
<th>CoefVar</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>3.45</td>
<td>3.27</td>
<td>10.68</td>
<td>94.73</td>
<td>0.00</td>
<td>1.00</td>
<td>2.50</td>
<td>5.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Maximum</th>
<th>Range</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>11</td>
<td>4.75</td>
</tr>
</tbody>
</table>

Comparison with Poisson distribution

If eggs were distributed randomly over the bugs the eggs would have a Poisson distribution (see handout on binomial and related distributions). In the data used in this handout, and indeed in almost all our collections, there are more bugs with extreme — either very low or very high — numbers of eggs than would be expected in a Poisson distribution. As the comparison to the right shows, the observed distribution is much more “aggregated” (a few bugs having most of the eggs) than random. Note that this deviation from randomness is the opposite of the even distribution that would occur if female flies preferred to oviposit on bugs with the fewest eggs already on them.
Conclusions

The distribution is basically unimodal; the second peak at eggs = 8 is the sort of random feature that is not surprising in such a small sample.

The distribution is strongly skewed, with most bugs having no or very few eggs, and only a few bugs having several to many eggs. The peak of the distribution actually is at 0 eggs; there is no left tail at all. The distribution is strongly non-normal.

Because of the skewed distribution and the outlier, the mean and median are quite different. (Note the greater effect of removing the outlier on the mean than on the median.) The median is a better measure of the “typical” egg load: fully half the bugs had no more than two eggs on them.

For the same reasons, the IQR is a better measure of spread than is the standard deviation. The middle half of the observations spans a range of 6 eggs (from 1 to 7).

The distribution clearly is not remotely like a normal distribution, and indeed is more “aggregated” than a Poisson distribution, indicating non-random distribution of the eggs with some bugs getting far more than their share of eggs.