We have discussed various sources of variance that impact the dynamics of a population: demographic, environmental and spatial (process), individual heterogeneity, and genetic variances. In addition, the concept of sampling variance, or the uncertainty of our estimates of population parameters, has been frequently mentioned. This chapter covers the statistical methodology to estimate the different variance components from data.

Consider the example situation of estimating survival rates each year for 10 years from a deer population. Each year, the survival rate is different from the overall mean, because of snow depth, cold weather, etc. Let the true, but unknown, overall mean be $S$. Then the survival rate for each year can be considered to be $S +$ some deviation:

<table>
<thead>
<tr>
<th>Year $i$</th>
<th>Environmental Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$S + e_1$</td>
</tr>
<tr>
<td>2</td>
<td>$S + e_2$</td>
</tr>
<tr>
<td>3</td>
<td>$S + e_3$</td>
</tr>
<tr>
<td>4</td>
<td>$S + e_4$</td>
</tr>
<tr>
<td>5</td>
<td>$S + e_5$</td>
</tr>
<tr>
<td>6</td>
<td>$S + e_6$</td>
</tr>
<tr>
<td>7</td>
<td>$S + e_7$</td>
</tr>
<tr>
<td>8</td>
<td>$S + e_8$</td>
</tr>
<tr>
<td>9</td>
<td>$S + e_9$</td>
</tr>
<tr>
<td>10</td>
<td>$S + e_{10}$</td>
</tr>
<tr>
<td>Mean</td>
<td>$\bar{S}$</td>
</tr>
</tbody>
</table>

The estimator of $S$ is $\bar{S}$:
with the variances of the $S_i$:
\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{10} (S_i - \bar{S})^2}{10}
\]

where the random variables $e_i$ are selected from a distribution with mean 0 and variance $\sigma^2$. In reality, we are never able to observe the annual rates because of sampling variation or demographic variation. For example, even if we observed all the members of a population, we would still not be able to say the observed survival rate was $S_j$ because of demographic variation. Consider flipping 10 pennies. We know that the true probability of a head is 0.5, but we will not always observe that value exactly. The same process operates in a population as demographic variation. Even though the true probability of survival is 0.5, we would not necessarily see exactly $\frac{1}{2}$ of the population survive on any given year.

Hence, what we actually observe are the quantities:

<table>
<thead>
<tr>
<th>$i$</th>
<th>Mean</th>
<th>Year $i$</th>
<th>Year $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$S$</td>
<td>$S + e_1 + f_1$</td>
<td>$\hat{S}_1$</td>
</tr>
<tr>
<td>2</td>
<td>$S$</td>
<td>$S + e_2 + f_2$</td>
<td>$\hat{S}_2$</td>
</tr>
<tr>
<td>3</td>
<td>$S$</td>
<td>$S + e_3 + f_3$</td>
<td>$\hat{S}_3$</td>
</tr>
<tr>
<td>4</td>
<td>$S$</td>
<td>$S + e_4 + f_4$</td>
<td>$\hat{S}_4$</td>
</tr>
<tr>
<td>5</td>
<td>$S$</td>
<td>$S + e_5 + f_5$</td>
<td>$\hat{S}_5$</td>
</tr>
<tr>
<td>6</td>
<td>$S$</td>
<td>$S + e_6 + f_6$</td>
<td>$\hat{S}_6$</td>
</tr>
<tr>
<td>7</td>
<td>$S$</td>
<td>$S + e_7 + f_7$</td>
<td>$\hat{S}_7$</td>
</tr>
<tr>
<td>8</td>
<td>$S$</td>
<td>$S + e_8 + f_8$</td>
<td>$\hat{S}_8$</td>
</tr>
<tr>
<td>9</td>
<td>$S$</td>
<td>$S + e_9 + f_9$</td>
<td>$\hat{S}_9$</td>
</tr>
<tr>
<td>10</td>
<td>$S$</td>
<td>$S + e_{10} + f_{10}$</td>
<td>$\hat{S}_{10}$</td>
</tr>
<tr>
<td>Mean</td>
<td>$S$</td>
<td>$\bar{S}$</td>
<td>$\bar{S}$</td>
</tr>
</tbody>
</table>

where the $e_i$ are as before, but we also have additional variation from sampling (or demographic
variation), $f_i$.

The standard approach to estimating the sampling variance separately from the environmental variance is to take replicate observations within each year in this example, so that the within cell replicates can be used to estimate the sampling variance, whereas the between cell variance is used to estimate the environmental variation. Years are assumed to be a random effect, and mixed model analysis of variance procedures are used (e.g., Bennington and Thayne 1994). This approach assumes that each cell has the same sampling variance. An example of the application of a random effects model is Koenig et al. (1994). They considered year effects, species effects, and individual tree effects.

The assumption of constant variance within cells across a variety of treatment effects is often not true, i.e., the sampling variance of a binomial distribution is a function of the parameter estimate. Another common violation of this assumption is caused by the variable of interest being distributed lognormally, so that the coefficient of variation is constant across cells, so that the cell variance is a function of the cell mean. Further, the empirical estimation of the variance from replicate measurements may not be the most efficient procedure. Again, the binomial estimator of a survival rate is a good example of this. Therefore, the remainder of this chapter describes methods which can be viewed as extensions of the usual variance component analysis based on replicate measurements within cells. We will be examining estimators for the situation where the within cell variance is estimated by some other estimator than from the moment estimator based on replicate observations.

Assume that we can estimate the sampling variance for each year, given a value of $S_i$ for the year. For example, the sampling variation for a binomial is

$$\text{Var}(\tilde{S} | S_j) = \frac{\tilde{S}_i(1 - \tilde{S}_i)}{n_i}$$

where $n_i$ is the number of animals monitored to see if they survived. Then, can we estimate the variance term due to environmental variation, given that we have estimates of the sampling variance for each year?

If we assume all the sampling variances are equal, the estimate of the overall mean is still just the mean of the 10 estimates:

$$\overline{\tilde{S}} = \frac{\sum_{i=1}^{10} \tilde{S}_i}{10}$$

with the theoretical variance being
\[
\text{var}(\bar{S}) = \sigma^2 + \frac{E[\text{var}(\hat{S}|S)]}{10}
\]

i.e., the total variance is the sum of the environmental variance plus the expected sampling variance. This total variance can be estimated as

\[
\hat{\text{var}}(\bar{S}) = \frac{\sum_{i=1}^{10} (\hat{S}_i - \bar{S})^2}{10(10 - 1)}
\]

We can estimate the expected sampling variance as the mean of the sampling variances

\[
\hat{E}(\text{var}(\hat{S}|S)) = \frac{\sum_{i=1}^{10} \text{var}(\hat{S}_i|S)}{10}
\]

so that the estimate of the environmental variance obtained by solving for \( \sigma^2 \)

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{10} (\hat{S}_i - \bar{S})^2}{(10 - 1)} - \frac{\sum_{i=1}^{10} \text{var}(\hat{S}_i|S)}{10}
\]

However, normally, the sampling variances are not all equal, so that we have to weight them to obtain an unbiased estimate of \( \sigma^2 \). The general theory says to use a weight, \( w_i \)

\[
w_i = \frac{1}{\sigma^2 + \text{var}(\hat{S}_i|S)}
\]

so that the estimator of the weighted mean is

\[
\bar{S} = \frac{\sum_{i=1}^{10} w_i \hat{S}_i}{\sum_{i=1}^{10} w_i}
\]

with theoretical variance (see Box 1 for a derivation of this result)

\[
\text{var}(\bar{S}) = \frac{1}{\sum_{i=1}^{10} w_i}
\]
and empirical variance estimator

\[
\text{vâr}(\bar{S}) = \frac{\sum_{i=1}^{10} w_i (\hat{S}_i - \bar{S})^2}{\left[\sum_{i=1}^{10} w_i \right] (10 - 1)}
\]

When the \( w_i \) are the true (but unknown) weights, we have

\[
\frac{1}{\sum_{i=1}^{10} w_i} = \frac{\sum_{i=1}^{10} w_i (\hat{S}_i - \bar{S})^2}{\left[\sum_{i=1}^{10} w_i \right] (10 - 1)}
\]

giving the following

\[
1 = \frac{\sum_{i=1}^{10} w_i (\hat{S}_i - \bar{S})^2}{(10 - 1)}
\]

Hence, all we have to do is manipulate this equation with a value of \( \sigma^2 \) (that is imbedded in the \( w_i \) term) to obtain an estimator of \( \sigma^2 \).

To obtain a confidence on the estimator of \( \sigma^2 \), we can substitute the appropriate chi-square values in the above relationship. To find the upper confidence interval value, \( \hat{\sigma}^2_U \), solve the equation

\[
\frac{\sum_{i=1}^{10} w_i (\hat{S}_i - \bar{S})^2}{(10 - 1)} = \frac{\chi^2_{10, 1, \alpha_U}}{10 - 1}
\]

and for the lower confidence interval value, \( \hat{\sigma}^2_L \), solve the equation

\[
\frac{\sum_{i=1}^{10} w_i (\hat{S}_i - \bar{S})^2}{(10 - 1)} = \frac{\chi^2_{10, 1, \alpha_L}}{10 - 1}
\]

As an example, consider the following fawn survival data from Little Hills.
The survival rates are the number of collared animals that lived divided by the total number of collared animals. For 1981, $\hat{S}_{1981} = 15/46 = 0.326087$. The sampling variance associated with this estimate computed as

$$\text{var}(\hat{S}_{1981}) = \frac{\hat{S}_{1981}(1 - \hat{S}_{1981})}{46}$$

which equals 0.0047773. The spreadsheet VARCOMP.WB2 computes $\hat{\sigma}^2$ as 0.0170632 ($\hat{\sigma} = 0.1306262$), with a 95% confidence interval of (0.0064669, 0.0869938) for $\sigma^2$, and (0.0804167, 0.2949472) for $\sigma$.

Sensitivity of the Leslie matrix elements is inversely related to the process variance of the life history traits (Pfister 1998). That is,
because the process variance of $\lambda$ is a function of the process variance of the parameter 

$$\text{var}(\lambda) = \sum_{i,j} \left( \frac{\delta \lambda}{\delta a_{ij}} \right)^2 \text{var}(a_{ij}),$$

assuming no covariances between the $a_{ij}$ elements. In order to persist, a population must have a limited amount of variation. Thus, natural selection will select against high process variance in a parameter that $\lambda$ is very sensitive too.

Literature Cited


Box 1. Derivation of the result that

\[
\text{var}(\bar{S}) = \frac{1}{\sum_{i=1}^{10} w_i}
\]

Starting with the result that

\[
w_i = \frac{1}{\sigma^2 + \text{var}(\hat{S}_i|S_j)} = \frac{1}{\text{var}(\hat{S}_i)}
\]

the derivation is as follows.

\[
\text{var}(\bar{S}) = \text{var}\left(\frac{\sum_{i=1}^{10} w_i \hat{S}_i}{\sum_{i=1}^{10} w_i}\right)
\]

\[
= \frac{1}{\left(\sum_{i=1}^{10} w_i\right)^2} \text{var}\left(\sum_{i=1}^{10} w_i \hat{S}_i\right)
\]

\[
= \frac{1}{\left(\sum_{i=1}^{10} w_i\right)^2} \sum_{i=1}^{10} \text{var}(w_i \hat{S}_i)
\]

\[
= \frac{1}{\left(\sum_{i=1}^{10} w_i\right)^2} \sum_{i=1}^{10} w_i^2 \text{var}(\hat{S}_i)
\]

\[
= \frac{1}{\left(\sum_{i=1}^{10} w_i\right)^2} \sum_{i=1}^{10} w_i
\]

\[
= \frac{1}{\sum_{i=1}^{10} w_i}
\]