

Lab 6: Variance Components In Program Excel

Objectives: The purpose of this exercise is to introduce techniques for variance components that can be applied to any set of demographic parameters, including fecundity rates. The basis for this exercise is the reading by White (2000) that is in the course pack. Download the file *lab06_varcomp.xls* from the course website. This spreadsheet contains the data on mule deer fawn survival from pg 311 of White (2000).

Step 1: Parameter estimation

A first step is to calculate the probability of annual survival for the deer fawns. In cell **D5**, type =C5/B5. Copy and paste this formula into cells **D6 to D11**.

The sampling variance for each annual estimate of survival can be calculated from the binomial distribution as:

$$\text{var}(\hat{S}_i | S_i) = \frac{(\hat{S}_i(1-\hat{S}_i))}{N} \quad (\text{eqn 1})$$

where \hat{S}_i = the probability of survival and N = sample size. In cell **E5**, type =(D5*(1-D5))/B5. Copy and paste this formula into cells **E6 to E11**.

Step 2: Calculating a naïve estimate of process variance with equal sampling variances

Next, calculate a naïve estimate of the overall mean annual survival rate ($\bar{\hat{S}}$). In cell **C15**, type =AVERAGE(D5:D11). The result should be 0.25284. Then, calculate the sample size of years. In cell **A12**, type =COUNT(A5:A11). The result should be 7.

The total variance of which is the sum of the environmental process variance plus the expected sampling variance can be estimated as:

$$\text{var}(\bar{\hat{S}}) = \frac{\sum_{i=1}^N (\hat{S}_i - \bar{\hat{S}})^2}{(N-1)} \quad (\text{eqn 2})$$

where $\bar{\hat{S}}$ is the overall mean survival rate, \hat{S}_i are the annual estimates of annual survival and N = sample size (here, $N = 7$). The next step is to estimate the terms needed for the numerator of this formula. In cell **F5**, type =(D5-\$C\$15)^2. The \$ signs in this formula anchor the cell reference to C15 so that if you cut and paste this formula it will retain the same cell reference. The ^ symbol means 'raise to the power of' in Excel. Cut and paste the formula in cell F5 to cells **F6 to F11**. In cell **F12**, type =SUM(F5:F11), which should total 0.10904. Now, you are ready to calculate the total variance using eqn 2. In cell **B18**, type: =(F12)/(A12*(A12-1)) and the estimate of total variance should be 0.00260.

A naïve estimate of the sampling variance can be calculated as:

$$\hat{\sigma}_{sampling}^2 = \frac{\sum_{i=1}^N \hat{\sigma}_i^2}{N} \quad (\text{eqn 3})$$

To calculate the numerator of this formula, into cell **E12** type: =SUM(E5:E11). Next, enter into cell **C18** the formula =E12/A12, which should be equal 0.00177. A naïve estimate of the process variance can then be calculated by subtraction. In cell **D18** type =B18-C18, which should equal 0.00082. Eqn 3 assumes that the sampling variances are equal which may not be a valid assumption. Type the value observed in cell D18 into cell **C25**, it will be used as a starting point for the next set of calculations.

Step 3: Calculating temporal process variance with unequal sampling variances

Sampling variances are usually not equal and should be weighted to obtain unbiased estimates of process variance. The weights (w_i) can be estimated as:

$$w_i = \frac{1}{\sigma^2 + \hat{v}\hat{\sigma}(\hat{S}_i | S_i)} \quad (\text{eqn 4})$$

where σ^2 = the unknown process variance and $\hat{v}\hat{\sigma}(\hat{S}_i | S_i)$ is the sampling variance calculated in eqn 1. Note that the true weights are unknown because the process variance is unknown. However, as a starting point, we can use the naïve estimate of process variance that you just entered into cell C25. To calculate the weights, enter into cell **G5** the formula =1/(\$C\$25+E5). Cut and paste this formula into cells **G6 to G11** for the remainder of the years. These preliminary estimates of the weights can then be used to obtain a weighted mean survival rate and a weighted estimate of the process variance.

The weighted mean survival rate can be estimated as:

$$\bar{\hat{S}} = \frac{\sum_{i=1}^N w_i \hat{S}_i}{\sum_{i=1}^N w_i} \quad (\text{eqn 5})$$

To estimate the denominator of this formula, enter into cell **G12** the formula =SUM(G5:G11) which should total 3275.79. To estimate the numerator of this formula, enter into cell **H5**: =D5*G5. Copy and paste this formula into cells **H6 to H11**. To sum these values, enter into cell **H12** =SUM(H5:H11) which should total 697.09. The weighted mean survival rate can then be calculated in cell **C21** as =H12/G12 and should equal 0.21280.

The final step in these calculations is a challenging concept. We would like to calculate an estimate of the process variance, which is unknown but is a function of the weights (see above). In order to obtain an estimate of the process variance, it is necessary to calculate the weights that satisfy the following expression:

$$1 = \frac{\sum_{i=1}^N w_i (\hat{S}_i - \bar{\hat{S}})^2}{N-1} \quad (\text{eqn 6})$$

where the term $(\hat{S}_i - \bar{\hat{S}})^2$ is similar to what we have already calculated in cells F5 to F11 of our spreadsheet, except that as an estimate of the overall mean ($\bar{\hat{S}}$) we now want to use our improved weighted mean in cell C21 instead of the naïve mean in cell C15. In cell **I5**, enter the following formula: =G5*(D5-\$C\$21)^2. Copy and paste this formula into cells **I6 to I11**, then sum these values in cell **I12** as =SUM(I5:I11). These values should total 61.02. We can now estimate equation 6 in cell **D25** by typing: =I12/(A12-1). The current total based on our naïve estimate of process variance (0.00082) equals 10.17013, which clearly does not satisfy the requirement that the sum of eqn 6 should equal 1.

The spreadsheet is now set up to calculate the process variance through an iterative process. The values for process variance that are input into cell **C25** are used to calculate weights according to eqn 4 in cells G5 to G11; the weights are then used to calculate eqn 6 in cells I5 to I11 and then cell **D25**. Thus, by adjusting values in C25, you can estimate the value of the process variance that would be required to set the total of eqn 6 to 1. You can try this by trial and error.

A simpler approach is to use the Solver tool of Program Excel. You may need to install this tool if it was not set up at installation. If it is installed, you can select this tool by clicking on Tools | Solver which will produce a dialog box. In **Set Target Cell**, enter \$D\$25 and set Equal To a value of 1. In **By Changing Cells**, enter \$C\$25. Then click on Solve to do the calculations. Next, click on OK to Keep Solver Solution. Your best estimate of the temporal process variance for the survival of mule deer fawns is 0.01706, which exactly matches the estimate reported by White (2000:312).

Step 4: Calculating confidence intervals for the temporal process variance

Confidence intervals for the temporal process variance can be obtained by setting eqn 6 not to one, but to the appropriate chi-square values that correspond to the lower and upper boundaries of the confidence interval:

$$\frac{\sum_{i=1}^N w_i (\hat{S}_i - \bar{\hat{S}})^2}{N-1} = \frac{\chi_{N-1, \alpha_{Lower}}^2}{N-1} \text{ or } \frac{\chi_{N-1, \alpha_{Upper}}^2}{N-1} \quad (\text{eqn 7})$$

In this expression, N = the sample size of years and α = the alpha value for the confidence interval. If 95%CI are desired, then $\alpha_{Lower} = 0.025$ and $\alpha_{Upper} = 0.975$, respectively. The inverse chi-square function, CHIINV(probability, df), returns a chi-square value for a given alpha value and the degrees of freedom. In cell **E26**, enter =CHIINV(0.025,(\$A\$12-1))/(\$A\$12-1) for the lower CI, which should be 2.4082. In cell **E27**, enter =CHIINV(0.975,(\$A\$12-1))/(\$A\$12-1) for the upper CI, which should be 0.2062. Again, you can use Solver to vary the estimate of the process variance in **C25** that is needed to satisfy eqn 7 in cell **D25** to equal the new reference values in cells **E26** and **E27**. The 95%CI for the temporal process variance is: 0.00647 and 0.08700.