Solution to Exercise 5.4 (Version 1, 26/09/14)

from Statistical Methods in Biology: Design & Analysis of Experiments and Regression (2014) S.J. Welham, S.A. Gezan, S.J. Clark & A. Mead. Chapman & Hall/CRC Press, Boca Raton, Florida. ISBN: 978-1-4398-0878-8

Exercise 5.4 (Data: courtesy J. Baverstock, Rothamsted Research)

Compare the unbiased sample variances for each treatment group from Exercise 4.3 using Bartlett's test. Is there any evidence of variance heterogeneity?

Solution 5.4

The data set of Exercise 4.3 (data in file FUNGUS.DAT) consists of observations from t = 3 treatments (1 = *B. bassiana*, 2 = *P. neoaphidis*, 3 = Control) with unequal replication ($n_1 = 9$, $n_2 = n_3 = 12$). There is of a total of N = 33 observations. We represent the kth observation of number of nymphs within the jth treatment group as $Nymphs_{jk}$. The treatment sample means, $\overline{Nymphs}_{j\bullet}$, for j = 1, 2, 3, are equal to 66.6, 53.8 and 57.3, respectively. The unbiased sample variance for the jth treatment is calculated as

$$s_j^2 = \frac{1}{\left(n_j - 1\right)} \sum_{k=1}^{n_j} \left(Nymphs_{jk} - \overline{Nymphs}_{j\bullet}\right)^2.$$

For example, the unbiased sample variance for the *B. bassiana* (first) treatment is

$$\begin{aligned} s_1^2 &= \frac{1}{9} \sum_{k=1}^{9} \left(Nymphs_{1k} - 66.6\right)^2 \\ &= \frac{1}{9}[(64-66.6)^2 + \ldots + (68-66.6)^2 + \ldots + (72-66.6)^2] = 331.78. \end{aligned}$$

The unbiased sample variances for the other two treatments are calculated similarly as s_2^2 = 233.61, and s_3^2 = 162.79. The pooled variance is

$$s^2_{\text{pooled}} = \frac{(n_1 - 1)\times s_1^2 + (n_2 - 1)\times s_2^2 + (n_3 - 1)\times s_3^2}{n_1 + n_2 + n_3 - 3} = \frac{(8\times 331.78) + (11\times 233.61) + (11\times 162.79)}{30} = 233.82.$$

The scaling factor c required to calculate Bartlett's test statistic, X^2, is

$$c = \frac{1}{3(t-1)}\left[\left(\sum_{j=1}^{t} \frac{1}{(n_j - 1)}\right) - \frac{1}{N-t}\right] = \frac{1}{6}\left[\left(\frac{1}{8} + \frac{2}{11}\right) - \frac{1}{30}\right] = 0.046.$$

We can calculate Bartlett's test statistic as

$$
\begin{aligned}
X^2 &= \frac{1}{(1+c)}\left[\ (N-t)\log_e(s^2_{\text{pooled}}) - \sum_{j=1}^{t}(n_j-1)\log_e(s_j^2)\ \right] \\
&= \frac{1}{1.046}\left[\ 30\times\log_e(233.88) - (8\times\log_e(331.78) + 11\times\log_e(233.61) + 11\times\log_e(162.79))\ \right] \\
&= 1.142.
\end{aligned}
$$

The 95th percentile of the chi-squared distribution on $t - 1 = 2$ df is 5.991. The observed test statistic is less than this value, and so there is no evidence of variance heterogeneity. The observed significance level of the test is 0.565.