CHAPTER 5

ESTIMATION OF CROP LOSS UNDER MULTIPLE CLASSIFICATION OF THE DISEASE

5.1 Introduction

In the foregoing chapters the discussion was directed towards the construction of estimator of crop loss when the plants were classified into one of two classes viz., Healthy and Diseased. However there are situations [James et al (1968), James et al (1972), James and Shih (1973)] where the diseased group of plants is further classified into many categories depending on the severity of the attack. The crop loss due to a disease may also depend heavily on the severity. Thus we may desire to estimate the crop loss based on the twin criteria of incidence of the disease and its severity. This then becomes a problem of multiple classes, since the manifested disease is in several classes of severity.

Let us suppose that the crop is affected by a disease with \( k \) classes (including the healthy group) of severity. When a random sample of fixed size \( n \) is drawn, let \( n_1, n_2, n_3, \ldots, n_k \) be the number of observations falling into these \( k \) classes, \( \sum_{i=1}^{k} n_i = n \). The objectives
are to estimate (i) the overall crop loss and (ii) the loss due to a particular degree of severity of disease, and (iii) comparison of crop losses between any two levels of severity.

5.2 Notations

\(N_j\) : Number of plants in the \(j^{th}\) class in the population.

\(\Sigma N_j = N\) : Total number of plants in the plot.

\(P_j = N_j/N\) : Proportion of plants in the \(j^{th}\) class in the population \((j = 1, 2, \ldots, k)\).

\(Y_i\) : Yield of the \(i^{th}\) plant.

\(Y_j = \frac{\Sigma N_j Y_j}{N}\) : Mean yield of a plant in the \(j^{th}\) class \((j = 1, 2, \ldots, k)\).

\(\bar{Y} = \frac{\Sigma N_j Y_j}{N}\) : Overall mean yield per plant in the plot.

\(S_j^2 = \frac{\Sigma N_j (Y_i - Y_j)^2}{(N_j - 1)}\) : Population mean sum of squares for the \(j^{th}\) class.

\(S^2 = \frac{N \Sigma (Y_i - \bar{Y})^2}{(N-1)}\) : Overall mean square.
\( n_j \): Number of plants in the \( j^{th} \) class in the sample.

\( \Sigma n_j = n \) Total sample size.

\( \hat{P}_j = n_j/n \) Estimator of \( P_j \) (when \( N_j \)’s are unknown, \( j = 1,2,\ldots,k \)).

\( \bar{y}_j \): Sample mean yield of the \( j^{th} \) class.

\( \bar{y} = \frac{\Sigma n_j \bar{y}_j}{\Sigma n_j} \) Overall sample mean and is an estimator of \( \bar{Y} \) when \( N_j \)’s are unknown.

\( \bar{y}_w = \frac{\Sigma n_j \bar{y}_j}{\Sigma n_j} \) Weighted sample mean when \( N_j \) is known.

\( s_j^2 = \frac{\Sigma_j (y_j - \bar{y}_j)^2}{(n_j - 1)} \) Sample mean sum of squares for the \( j^{th} \) class.

\( s^2 = \frac{\Sigma (y_j - \bar{y})^2}{(n-1)} \) Overall sample mean square.

\( \hat{N}_j \): Estimator of \( N_j \) when \( N_j \) is unknown.

It is assumed that there is atleast one observation from each class in the sample to allow construction of point estimators. However for estimating the precision, atleast two observations are assumed to be available in the sample from each class.
5.3 Estimation of Overall Crop Loss

Relevant formulae for crop loss estimation are derived when the subpopulation sizes $N_j$ are unknown and also when they are known although the latter case is less realistic.

5.3.1 Case of $N_j$'s not known

As defined earlier the overall crop loss in the plot is $N(\overline{y}_1-\overline{y})$ and an estimator of this is $N(\overline{y}_1-\overline{y})$. (5.1)

Here $\overline{y}_1$ and $\overline{y}$ are as defined in section 5.2.

**Theorem 5.1**

Under SRSWOR of size $n$, $N(\overline{y}_1-\overline{y})$ is an unbiased estimator of the total crop loss with the variance given by

\[
V[N(\overline{y}_1-\overline{y})] = N^2 \left[ s_1^2 \left( \frac{1}{N_1} - \frac{1}{n} \right) + \frac{N-n}{N} \frac{s^2}{n} \right]. \quad (5.2)
\]

**Proof:**

\[
E[N(\overline{y}_1-\overline{y})] = N \left[ E(\overline{y}_1) - E(\overline{y}) \right],
\]

\[= N(\overline{y}_1-\overline{y}).\]

Consider $V(\overline{y}_1-\overline{y}) = V(\overline{y}_1)+V(\overline{y})-2\text{Cov}(\overline{y}_1, \overline{y}). \quad (5.3)$
From Lemma 4.5 we have

\[ V(\overline{y}_1) = S_1^2 \left[ E\left(1/n_1\right) - 1/N_1 \right]. \]  

(5.4)

Since \( \overline{y} \) is the mean of the \( n \) observations under SRSWOR,

\[ V(\overline{y}) = \frac{N-n}{N} \frac{S^2}{n}. \] 

(5.5)

From Lemma 4.7

\[ \text{Cov}(\overline{y}_1, \overline{y}) = \frac{S_1^2}{n} \left( 1 - \frac{E(n_1)}{N_1} \right), \]

\[ = \frac{S_1^2}{n} \left( 1 - \frac{n}{N} \right), \text{ since } E(n_1) = nP_1. \] 

(5.6)

Substituting (5.4), (5.5) and (5.6) in (5.3) we have

\[ V(\overline{y}_1 - \overline{y}) = S_1^2 \left[ E\left(\frac{1}{n_1}\right) - \frac{1}{N_1} \right] + \frac{N-n}{N} \frac{S^2}{n} - \frac{2S_1^2}{n} \left( 1 - \frac{n}{N} \right). \]

Accordingly

\[ V[N(\overline{y}_1 - \overline{y})] = N^2 \left[ S_1^2 \left[ E\left(\frac{1}{n_1}\right) - \frac{1}{N_1} - \frac{2}{n} + \frac{2}{N} \right] + \frac{N-n}{N} \frac{S^2}{n} \right]. \]

It may be seen that this expression is the same as that used for two classes, as it should be.

This expression can be evaluated after substituting \( (4.9) \) for \( E(1/n_1) \).

5.3.2 Case of \( N_i \)'s known

In practice \( N_j \)'s are usually unknown. But on any occasion if they are known then:
\[ \overline{y}_w = \frac{k}{N} \sum_j \overline{y}_j / \sum_j \]  

(5.7)

is an unbiased estimator of \( \overline{Y} \) (Cochran, 1977).

**Theorem 5.2**

When \( N_j \)'s are known, \( N(\overline{y}_1 - y_w) \) is an unbiased estimator of the total crop loss when the sampling is SRSWOR and its variance is given by

\[ V[N(\overline{y}_1 - y_w)] = N^2 \left[ \frac{S_1^2}{N_1} \left( 1 - \frac{1}{N_1} \right) \left( 1 - 2 \frac{N_1}{N} \right) + \right. \]

\[ + \frac{N - N}{N} \left[ \sum_j \frac{N_j S_j^2}{N} + \frac{N}{n(N-1)} \left( \frac{k}{N} \right) \right] \]  

(5.8)

**Proof:**

\[ E[N(\overline{y}_1 - y_w)] = N E(\overline{y}_1) - N E(\overline{y}_w), \]

\[ = N(\overline{y}_1 - \overline{Y}). \]

Thus \( N(\overline{y}_1 - y_w) \) is an unbiased estimator of the total crop loss.

Now \( V(\overline{y}_1 - y_w) = V(\overline{y}_1) + V(\overline{y}_w) - 2 \text{Cov}(\overline{y}_1, \overline{y}_w) \).  

(5.9)

From Lemma 4.5

\[ V(\overline{y}_1) = S_1^2 \left[ E(1/n_1) - \frac{1}{N_1} \right]. \]

(5.10)
Since \( N_j \)'s (hence \( P_j \)'s) are known the problem is treated as that of post stratification with known results (see Cochran, 1977).

The expression for \( V(\overline{y}_w) \) is then given by

\[
V(\overline{y}_w) = \frac{N-n}{Nn} \sum \frac{k N_j S_i^2}{N} + \frac{k}{n^3} \sum \frac{NS_j^2}{N_j} \text{ V}(n_j).
\]  

(5.11)

Now \( \text{Cov}(\overline{y}_1, \overline{y}_w) = E[\overline{y}_1 \overline{y}_w] - E(\overline{y}_1) E(\overline{y}_w) \),

\[
= E[\overline{y}_1(P_1 \overline{y}_1 + P_2 \overline{y}_2 + \cdots + P_k \overline{y}_k)] - \overline{y}_1(P_1 \overline{y}_1 + P_2 \overline{y}_2 + \cdots + P_k \overline{y}_k),
\]

\[
= P_1 E(\overline{y}_1^2) + P_2 E(\overline{y}_1 \overline{y}_2) + \cdots + P_k E(\overline{y}_1 \overline{y}_k) - P_1 \overline{y}_1^2 - P_2 \overline{y}_2^2 - \cdots - P_k \overline{y}_k^2.
\]

\[
= P_1 E(\overline{y}_1^2) + P_2 \overline{y}_1 \overline{y}_2 + \cdots + P_k \overline{y}_1 \overline{y}_k - P_1 \overline{y}_1^2 - P_2 \overline{y}_2^2 - \cdots - P_k \overline{y}_k^2.
\]

(5.12)

Substituting (5.10), (5.11) and (5.12) in (5.9) we have

\[
V(\overline{y}_1 - \overline{y}_w) = s_1^2[E(1/n_1) - \frac{1}{N_1}](1-2P_1) + \frac{N-n}{Nn} \sum \frac{N_j S_j^2}{N} + \frac{NS_j^2}{N_n^3} \text{ V}(n_j).
\]
Since the sampling is SRSWOR

\[ V(n_j) = \frac{(N-n) \cdot np_j \cdot Q_j}{(N-1)} = \frac{(N-n) \cdot nN_j(N-N_j)}{N^2}. \]

\[ \therefore V(\overline{Y}_1-\overline{Y}_w) = s_I^2[E(1/n_1) - \frac{1}{N_1}] \left[ 1 - \frac{2N_1}{N} \right] + \]

\[ \frac{N-n}{Nn} \left[ \sum_{i=1}^{k} \frac{N_i S_i^2}{N} + \frac{N}{n(N-1)} \frac{k}{1} \frac{(1 - \frac{N_1}{N}) S_j^2}{1} \right] \]

From this (5.8) follows.

It can be evaluated after substituting (4.9) for \( E(1/n_1) \). \( E(1/n_1) \) itself may be evaluated to different degrees of approximation.

5.4 Loss from a particular level of severity

\( \overline{Y}_1 \) is the mean yield of healthy plants and \( \overline{Y}_j \) the mean yield of the \( j^{th} \) class in the population with a specified severity of the disease. Hence \( \overline{Y}_1-\overline{Y}_j \) is the average loss per diseased plant for the \( j^{th} \) level of severity. Since only \( N_j \) plants belonging to this group have suffered this loss, the total loss due to this severity class \( j \) \( (j = 1, 2, \ldots, k) \) is given by \( N_j(\overline{Y}_1-\overline{Y}_j). \)

The two cases when \( N_j \)'s are unknown and when they are known are considered here.
5.4.1 Case of $N_j$'s not known

The total number $N_j$ is generally not known in advance.

**Lemma 5.1**

For any two random variables $x_1$ and $x_2$ an approximation for the expectation and variance of the ratio $x_1/x_2$ (Kendall and Stuart, 1958, p. 232) are

$$E(x_1/x_2) \approx \frac{E(x_1)}{E(x_2)}.$$  

$$V(x_1/x_2) \approx \left[ \frac{E(x_1)}{E(x_2)} \right]^2 \left[ \frac{V(x_1)}{(E(x_1))^2} + \frac{V(x_2)}{(E(x_2))^2} - \frac{2\text{Cov}(x_1, x_2)}{E(x_1)E(x_2)} \right].$$

**Lemma 5.2**

$$\frac{n_j^2}{E(n_j)} = \frac{E(n_j^2)}{E(n_j)} = \frac{n_jp_j(1-p_j) + n^2p_j^2}{nP_1}$$

**Theorem 5.3**

When $N_j$'s are not known and when the scheme is SRSWOR

$$\hat{N}_j(\bar{y}_j - \bar{y}_j) \text{ with } \hat{N}_j = N \frac{n_j}{n}$$

(5.13)

is an unbiased estimator of the total crop loss from $j$th level of severity of disease with the variance given by
\[
\text{\textbf{Proof : }}
\]

Taking \( \hat{N}_j = N \frac{n_1}{n} \) as an estimator of \( N_j \) (the choice being obvious),

\[
\hat{N}_j(\overline{y}_1 - \overline{y}_j) = N \frac{n_1}{n} (\overline{y}_1 - \overline{y}_j) = \hat{N}_j(\overline{y}_1 - \overline{y}_j) \quad \text{and}
\]

\[
E[N\hat{P}_j(\overline{y}_1 - \overline{y}_j)] = N[E(\hat{P}_j\overline{y}_1) - E(\hat{P}_j\overline{y}_j)] ,
\]

\[
= N[ E [\hat{P}_j(E(\overline{y}_1 | n_j, n_1))] - E [\hat{P}_j(E(\overline{y}_j | n_j, n_1))] ] ,
\]

\[
= N[ E (\hat{P}_j\overline{y}_1) - E (\hat{P}_j\overline{y}_j) ] ,
\]

\[
= N[ \overline{y}_1 E(\hat{P}_j) - \overline{y}_j E(\hat{P}_j) ] ,
\]

\[
= NP_j(\overline{y}_1 - \overline{y}_j) = N_j(\overline{y}_1 - \overline{y}_j) .
\]

Hence \( \hat{N}_j(\overline{y}_1 - \overline{y}_j) \) is an unbiased estimator of the total loss from a particular class of disease. Now consider
\[ V[\hat{N}_j(\bar{Y}_1-\bar{Y}_j)] = N^2V[\hat{P}_j(\bar{Y}_1-\bar{Y}_j)] , \]

\[ = N^2[V(\hat{P}_j\bar{Y}_1)+V(\hat{P}_j\bar{Y}_j)-2\text{Cov}(\hat{P}_j\bar{Y}_1, \hat{P}_j\bar{Y}_j)].(5.15) \]

We may evaluate (5.15) term by term. Consider first \( V(\hat{P}_j\bar{Y}_1) \). This involves three sets of selected random variables \( n_1, n_j \) and \( y \). Since in a multinomial distribution \( n_1, n_2, \ldots, n_k \) are all not independent, the evaluation of \( V[\hat{P}_j\bar{Y}_1] \) therefore requires first its evaluation conditionally holding all the \( n_1 \)'s (excepting \( n_1, n_j \)) constant and then averaging the resulting expression by evaluating its expectation over all \( n_1 \neq n_1 \) and \( n_j \).

\[ V[\hat{P}_j\bar{Y}_1] = \sum_{\text{all } n_i} E_{\text{all } n_i} \left[ V(\hat{P}_j\bar{Y}_1) \right] , \]

\[ = \sum_{\text{all } n_i} E_{n_i \neq n_1, n_j} \left[ V\left( \frac{n_i}{n} \left( \frac{1}{n_1} \Sigma \bar{Y}_1 \right) \right) \right] . \]

Since there are three sets of random variables \( n_1, n_j, y \) in the expression inside the square bracket, this expression reduces to \( \text{[see Cochran 1977, p. 276]} \)
\[ V[\hat{p}_j \bar{y}_1] = \sum_{n_1, n_j \neq n_1, n_j} E \left[ E \left[ V(1 \frac{n_1}{n} \frac{1}{1} \mid n_1) \right] \right] + \]

+ \[ V \left[ E(\frac{n_1}{n} \frac{1}{1} \mid n_1) \right] \]

+ \[ V \left[ E \left< \frac{n_1}{n} \frac{1}{1} \mid n_1 \right> \right] \],

\[ E \left[ E \left[ V(\frac{n_1}{n} \frac{1}{1} \mid n_1) \right] \right] + \]

+ \[ E \left[ V \left[ E(\frac{n_1}{n} \frac{1}{1} \mid n_1) \right] \right] + \]

+ \[ E \left[ V \left< E(\frac{n_1}{n} \frac{1}{1} \mid n_1) \right> \right] \right] \] \text{. (5.16)}

Let us evaluate (5.16) term by term.

First

\[ E \left[ E \left[ V(\frac{n_1}{n} \frac{1}{1} \mid n_j, n_1) \right] \right] = \]
\[ S^2 = \frac{1}{n} \sum_{n_j, n_1} \left( \frac{n^2}{n_1} - \frac{E(n_1^2)}{N_1} \right) \],

\[ = \frac{1}{n} \sum_{n_j, n_1} \left( \frac{n^2}{n_1} - \frac{V(n_1) + (E(n_1))^2}{N_1} \right) \] (5.17)

and we merely note here \( E(n_j^2/n_1) \) is a mixed moment of \( n_j, n_1 \).

The second term is

\[ E \left[ V \left[ E(n_j/n)(\sum_{n_1} n_1 y_{1}/n_1) \right] \right] = \]

all \( n_1, n_j \neq n_1, n_j \)

= \[ E \left[ V \left( \frac{n_1}{n} \sum_{j} y_{1} \right) \right] = 0 \] (5.18)

since the expression inside the bracket is constant w.r.t. \( n_1 \).

The third term is

\[ E \left[ V \left[ E \left( \frac{n_1}{n} \sum_{n_1} y_{1}/n_1 \right) \right] \right] = \]

all \( n_1, n_j \neq n_1, n_j \)
\[
E \left[ \prod_{n_1, n_j} \left( \frac{n_1}{n} \prod_{i=1}^{n_1} \prod_{j=1}^{n_j} \right) \right],
\]

\[
= E \left[ \prod_{n_1, n_j} \left( \frac{n_1}{n} \prod_{i=1}^{n_1} \prod_{j=1}^{n_j} \right) \right],
\]

\[
= E \left[ \prod_{n_1, n_j} \left( \frac{n_1}{n} \prod_{i=1}^{n_1} \prod_{j=1}^{n_j} \right) \right],
\]

Substituting (5.17), (5.18) and (5.19) in (5.16),

\[
V(\hat{p}_j y_j) = \frac{s_j^2}{n} \left[ E \left( \frac{n_1}{n} \right) + \frac{V(n_1) + (E(n_1))^2}{N_1} \right] + \frac{V^2}{n^2} V(n_j).
\]

Next consider \(V(\hat{p}_j y_j)\) in (5.15) which involves only the random variables \(n_j\) and \(y\) and hence

\[
E \left[ \prod_{n_1} \left( \frac{n_1}{n} \prod_{i=1}^{n_1} \prod_{j=1}^{n_j} \right) \right],
\]

\[
= \frac{s_j^2}{n} \left[ E \left( \frac{n_1}{n} \right) + \frac{V(n_1) + (E(n_1))^2}{N_1} \right] + \frac{V^2}{n^2} V(n_j).
\]
\[
\begin{align*}
E \left[ \frac{1}{n} \sum_{j \neq n_j} \left( \frac{S^2}{N_j} \left[ \frac{1}{n} \sum_{j \neq n_j} [E(n_j)(N_j-E(n_j)) - V(n_j)] + \gamma_j^2 V(n_j) \right] \right) \right] \\
= \frac{1}{n^2} \sum_{j \neq n_j} \left( \frac{S^2}{N_j} \left[ \frac{1}{n} \sum_{j \neq n_j} [E(n_j)(N_j-E(n_j)) - V(n_j)] + \gamma_j^2 V(n_j) \right] \right)
\end{align*}
\]

(5.21)

Consider \( \text{Cov}(\hat{P}_j \bar{y}_1, \hat{P}_j \bar{y}_j) \) in (5.15).

This will be equal to

\[
\begin{align*}
E \left[ \sum_{n_j \neq n_1, n_j} \left[ \text{Cov}[\hat{P}_j \bar{y}_1, \hat{P}_j \bar{y}_j] | n_1, n_j] + \right. \\
\left. \text{Cov}[E(\hat{P}_j \bar{y}_1 | n_1, n_j), E(\hat{P}_j \bar{y}_j | n_1, n_j)] \right] \right]
\end{align*}
\]

= \[
\begin{align*}
E \left[ \sum_{n_j \neq n_1, n_j} \left[ \hat{P}_j^2 \text{Cov}(\bar{y}_1, \bar{y}_j) + \gamma_1 \gamma_2 \text{Cov}(\hat{P}_j, \hat{P}_j) \right] \right]
\end{align*}
\]

(5.22)

Substituting (5.20), (5.21) and (5.22) in (5.15), we have
N^2 V[\bar{N}_j(\bar{y}_1-\bar{y}_j)] = N^2 \left[ \frac{S^2_1}{n^2_{j,n_1}} \sum_{j=1}^{n_1} E (\frac{n_1^2}{n_1}) - \frac{1}{N^2_1} [V(n_j)+(E(n_j))^2] \right] +

\frac{V^2_1 V(n_1)}{N^2_1} + \frac{1}{N^2_1} [E(n_j)(N_j-E(n_j))-V(n_j)] +
\frac{V^2_2}{n^2} V(n_j) - \frac{2V_1 V_2}{n^2} V(n_j) \right],

\frac{N^2}{n^2} \left[ S^2_1\left[E(\frac{n_1^2}{n_1}) - \frac{1}{N^2_1} (V(n_j)+(E(n_j))^2)\right] +
\frac{S^2_1}{N^2_1} [E(n_j)(N_j-E(n_j))-V(n_j)] + V(n_j)(\bar{y}_1-\bar{y}_2)^2 \right]

with \( E(n_j) = nP_j = n \frac{N_1}{N} \) and \( V(n_j) = \frac{N-n}{N-1} \frac{nN_1(N-N_1)}{N^2} \).

An explicit expression for the variance can be obtained by substituting (4.9) in the above expressions.

5.4.2 Case of \( N_j \)'s known

**Theorem 5.4**

When \( N_j \)'s are known and the sampling is SRSWOR

\[ N_j(\bar{y}_1-\bar{y}_j) \] \hspace{1cm} (5.23)

is an unbiased estimator of the total crop loss from a severity of class \( j \) of disease with the variance given by
\[ V(N_j(\bar{y}_1 - \bar{y}_j)) = N_j^2 \left[ S_1^2 \left( E \left( \frac{1}{n_1} \right) - \frac{1}{N_1} \right) + S_j^2 \left( E \left( \frac{1}{n_j} \right) - \frac{1}{N_j} \right) \right]. \] 

**Proof:**

\[ V(N_j(\bar{y}_1 - \bar{y}_j)) = N_j^2 \left[ V(\bar{y}_1) + V(\bar{y}_j) \right]. \]

Since \( \text{Cov}(\bar{y}_1, \bar{y}_j) = 0 \) [see (4.19)].

But from (4.10), \( V(\bar{y}_1) = S_1^2 \left[ E \left( \frac{1}{n_1} \right) - \frac{1}{N_1} \right] \), and similarily \( V(\bar{y}_j) = S_j^2 \left[ E \left( \frac{1}{n_j} \right) - \frac{1}{N_j} \right]. \) Hence

\[ V(N_j(\bar{y}_1 - \bar{y}_j)) = N_j^2 \left[ S_1^2 \left( E \left( \frac{1}{n_1} \right) - \frac{1}{N_1} \right) + S_j^2 \left( E \left( \frac{1}{n_j} \right) - \frac{1}{N_j} \right) \right]. \]