When a plot is affected by a disease, the resulting crop loss depends upon two factors, viz. prevalence of the disease as measured by proportion of plants affected by the disease in the plot and the extent of damage caused to the yield by this disease. These two aspects are built into the parameter measuring the crop loss (Chapter 6). Any estimator is thus a function of the estimators of $Q$, $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$ and hence its precision depends upon the precisions of these components. It has been observed that the precision of the crop loss estimator is quite sensitive to the estimator of $P$. Thus the precision of crop loss estimator may be improved by increasing the size of the sample to estimate just $P$, while estimate of $Y$ may still be based on a relatively smaller sample. This leads to a design similar to double sampling but not structurally quite so. In this case $n$ plants are selected at random identifying at each draw the group membership (healthy or diseased) of the plants as well as recording information on yield ($y$). Next an additional number of plants are drawn and information only
on group membership is obtained with no information on yield. Let \((n'-n)\) be the additional number of plants drawn so that in all \(n'\) plants are used for estimating \(P\) and only \(n\) plants for estimating the other parameters. Recording plant yield usually requires harvesting of plants, while identifying a plant as diseased or healthy can be done \textit{in situ} without harvesting. This leads to a cost efficient design in which merely identification of disease is done rapidly on a large number of plants, which in turn contributes to the increased precision for estimation of \(P\) and hence of crop loss estimation.

This aspect is investigated in this chapter and necessary formulae are derived. The results are applied later to simulated data (Chapter 10) as well as to field data (Chapter 11).

### 8.2 Notations

\[ y : \text{Yield of the plant.} \]
\[ x : \text{Indicator variable taking value 1 if the plant is healthy, 0 otherwise.} \]
\[ N_1 : \text{Number of diseased plants in the population.} \]
\[ P = \frac{N_1}{N} \text{Proportion of healthy plants in the population.} \]
\[ Q = 1-P = \frac{N_2}{N} \]
\[ \mu_1 : \text{Mean yield of healthy plants in the plot. (population)} \]

\[ \mu_2 : \text{Mean yield of the diseased plants in the population.} \]

\[ \mu = p\mu_1 + q\mu_2 \text{ Average yield per plant in the plot. (population)} \]

\[ \sigma_1^2 : \text{Variance of the yields of healthy plants in the population.} \]

\[ \sigma_2^2 : \text{Variance of the yields of diseased plants in the population.} \]

\[ n : \text{Number of sample plants on which information on both } x \text{ and } y \text{ is collected.} \]

\[ n' : \text{Total sample size (including the augmented part).} \]

\[ n_1 : \text{Number of sample plants belonging to group 1.} \]

\[ n_2 : \text{Number of sample plants belonging to group 2.} \]

\[ n = n_1 + n_2 \]

\[ \hat{p} = n_1 / n ; \hat{q} = 1 - \hat{p} = n_2 / n \]

\[ n'_1 : \text{Number of sample plants out of } n' \text{ belonging to group 1.} \]

\[ n'_2 : \text{Number of sample plants out of } n' \text{ belonging to group 2.} \]

\[ n' = n'_1 + n'_2 \]

\[ \hat{p} = n'_1 / n' ; \hat{q} = 1 - \hat{p} = n'_2 / n' \]
The sample design consists of first drawing an SRSWOR of size \( n \) one by one. For each plant the yield is recorded and it is also noted whether the plant is healthy or diseased. The outcome of each draw about group membership may be associated with a Bernoulli random variable \( x \) which takes the value 1 (healthy group) with probability \( P \) and zero (diseased group) with probability \( Q = 1 - P \). After the size \( n \) is reached draws are continued to provide an additional \((n' - n)\) units but information on \( x \) alone is noted from these units. Out of the first \( n \) units let \( n_1 \) be the number of plants belonging to the first group and \( n_2 \) the number belonging to the second. And out of the total \( n' \) sample units let \( n'_1 \) be the number belonging to the first group and \( n'_2 \) be the number belonging to the second. It may be noted that even though there are \( n'_1 \) sample units from the first group, information on yield is available only from \( n_1 \) of these.

The scheme is termed augmented sampling to indicate the fact that the draws are continued beyond \( n \) units to obtain a larger sample for estimating \( P \).
8.4 The Model

Consider the population of \( N \) plants to be a mixture of two subpopulations mixed up intimately in the proportion \( P = \frac{N_1}{N} \) of healthy plants and \( Q = (1-P) = \frac{N_2}{N} \) of diseased plants. Further assume the yield \((y)\) in the subpopulation of healthy plants to be normally distributed with mean \( \mu_1 \) and variance \( \sigma_1^2 \); and that in the subpopulation of diseased plants to be normally distributed with mean \( \mu_2 \) and variance \( \sigma_2^2 \). Then the appropriate model is a mixture of these two normal densities in the ratio \( P : Q \) as under:

\[
f(x, y) = [P^x(1-P)^{(1-x)}][xf_1(y)+(1-x)f_2(y)]; \quad x=0,1
\]

\[
= g(x) \ h(y|x)
\]

[ see (6.1) to (6.4) ]

in which \( g(x) = P^x(1-P)^{1-x} \) may be readily seen to be the (marginal) Bernoulli distribution for \( x \) and \( h(y|x) = [xf_1(y)+(1-x)f_2(y)] \) is the conditional density function for \( y \) given \( x \).
For the present design the data format will be as follows:

<table>
<thead>
<tr>
<th>Character</th>
<th>Initial set</th>
<th>Augmented set</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>x_1 x_2 x_3 \cdots x_n x_{n+1} x_{n+2} \cdots x_n'</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>y_1 y_2 y_3 \cdots y_n</td>
<td></td>
</tr>
</tbody>
</table>

(8.1)

As defined in Chapter 1, crop loss per diseased plant is measured by $\mu_1 - \mu_2$ and total loss by $N_2(\mu_1 - \mu_2)$. This is shown to be equivalent to $N(\mu - \mu)$ in which $\mu$ is the average yield per plant in the plot. The estimation of the crop loss requires the estimation of its defining parameters viz., $P$, $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$. In what follows the maximum likelihood estimators will be obtained. This will be followed by the estimator for crop loss in section (8.7).

8.5 The Maximum Likelihood Estimator:

for $\Theta = (P, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$

The likelihood function $L(\Theta)$ for the total sample set of observations, we note, is the product of the two likelihood functions viz.

$$\prod_{i=1}^{n} [f(x_i, y_i)]'$$

(8.2)

associated with the first
n independent observation vectors \([(x_1, y_1), \ldots, (x_n, y_n)]\) on which information on both \(x\) and \(y\) is recorded and

\[
\pi \left[ P^{i(1-P)^{1-x_x}} \right]_{i=n+1}^{n'}
\]

the augmented set of \((n'-n)\) independent observations recording \(x - \text{information only}\). Thus

\[
L(\Theta) = \prod_{i=1}^{n} [f(x_i, y_i)] \cdot \prod_{i=n+1}^{n'} \left[ P^{x_i(1-P)^{1-x_x}} \right]
\]

\[
= \prod_{i=1}^{n} [g(x_i) \cdot h(y_i|x_i)] \cdot \prod_{i=n+1}^{n'} \left[ g(x_i) \right]
\]

\[
= \prod_{i=1}^{n'} \left[ g(x_i) \right] \cdot \prod_{i=1}^{n} \left[ h(y_i|x_i) \right]
\]

\[
\text{(8.4)}
\]

I.e.

\[
L(\Theta) = [P^{1-x_x}(1-P)^{1-x_x}] \cdot \prod_{i=1}^{n} \left[ x_i f_1(y_i) + (1-x_x)f_2(y_i) \right]
\]

\[
\text{(8.5)}
\]

\[
= L_1(P) \cdot L_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2), \text{ (say)}
\]

\[
= \prod_{i=1}^{n'} \left[ x_i \right] \cdot \prod_{i=1}^{n} \left[ x_i f_1(y_i) + (1-x_x)f_2(y_i) \right]
\]

with \(L_1(P) = P^{1-x_x}(1-P)^{1-x_x}\),

\[
\text{and } L_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \prod_{i=1}^{n} \left[ x_i f_1(y_i) + (1-x_x)f_2(y_i) \right]
\]
Since $L(\Theta)$ is the product of two explicit likelihood functions $L_1(P)$ and $L_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, the maximum likelihood estimators of $P$, $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$ are those which maximise $L_1(P)$ with respect to $P$ and $L_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ with respect to $\mu_1$, $\mu_2$, $\sigma_1^2$ and $\sigma_2^2$.

### 8.5.1 Maximum Likelihood Estimator of $P$

Consider $L_1(P) = \prod_{i=1}^{n'} g(x_i)$,

$$
\sum_{i=1}^{n'} x_i - \sum_{i=1}^{n'} x_i = p^1 \cdot (1-P)^{1-n_i} \quad \text{ (8.6)}
$$

Therefore \( \ln L_1(P) = \sum_{i=1}^{n'} x_i \cdot \ln P + (n' - \sum_{i=1}^{n'} x_i) \cdot \ln (1-P) \),

and the log likelihood equation is

$$
\frac{d\ln L_1(P)}{dP} = \frac{n' \cdot \bar{x}}{P} - \frac{n' \cdot (1-\bar{x})}{1-P} = 0 \quad \text{ (8.7)}
$$

in which $\bar{x} = \frac{\sum_{i=1}^{n'} x_i}{n'}$.

The solution of this for $P$ is

$$
P = \bar{x} \quad \text{(based on all } n' \text{ observations).} \quad \text{ (8.8)}
$$

I.e. 

$$
P = \frac{\sum_{i=1}^{n_i} x_i}{n'} = \frac{n_i}{n'}
$$

where $n_i$ is the total number of sample plants belonging to the healthy group and has $x_i = 1 \ (i=1,2,\ldots,n_i)$. 
Further \( \frac{d^2 \ln L_1(P)}{dP^2} \) is given by

\[
\frac{n' \bar{x} (1-\bar{x})}{(1-P)^2}, \text{ and hence}
\]

we get

\[
E[- \frac{d^2 \ln L_1}{dP^2}] = \frac{E[ \frac{n' \bar{x}}{P} + \frac{n'(1-\bar{x})}{(1-P)^2} ]}{P^2}.
\]

\[
= \frac{n' \bar{x} + n'(1-P)}{(1-P)^2}.
\]

\[
= \frac{n'}{P} + \frac{n'}{(1-P)}.
\]

\[
= \frac{n'}{P(1-P)}.
\]

Hence \( V(\hat{P}) = \frac{P(1-P)}{n'} \).

8.5.2 Maximum Likelihood Estimators of \( \mu_1, \mu_2, \sigma_1^2 \) and \( \sigma_2^2 \)

Consider \( L_2 = \prod_{i=1}^{n} h(y_i|x_i) \). Note that \( x_i \) takes value 0 or 1. Thus the n sample observations consist of \( n_1 \) pairs of the type \((1, y_i)\) and \( n_2 \) pairs of the type \((0, y_i)\), with \( n = n_1 + n_2 \). Then an obvious break up of \( L_2 \) is

\[
L_2 = \prod_{i=1}^{n_1} h(y_i|x_i=1) \prod_{i=1}^{n_2} h(y_i|x_i=0)
\]
Thus the log likelihood becomes

\[ \ln L_2 = \sum_{1}^{n_1} \ln f_1(y_1) + \sum_{1}^{n_2} \ln f_2(y_1) , \]  

(8.10)

\[ = \sum_{1}^{n_1} \left[ -\ln \sigma_1 - \ln \sqrt{2\pi} - \frac{1}{2\sigma_1^2} (y_1 - \mu_1)^2 \right] + \]

\[ + \sum_{1}^{n_2} \left[ -\ln \sigma_2 - \ln \sqrt{2\pi} - \frac{1}{2\sigma_2^2} (y_1 - \mu_2)^2 \right] . \]  

(8.11)

Differentiating \( \ln L_2 \) with respect to \( \mu_1, \mu_2, \sigma_1^2, \sigma_2^2 \) and setting the derivative \( \frac{\partial \ln L_1}{\partial \mu_1} \) etc. equal to zero, we get the mle, say, \( \mu_1^*, \mu_2^*, \sigma_1^*, \sigma_2^* \) as the customary estimators defined by

\[ \mu_1^* = \overline{y}_1^* , \quad \sigma_1^* = \sum_{1}^{n_1} \frac{(y_1 - \overline{y}_1^*)^2}{n_1} , \]  

(8.12)

\[ \mu_2^* = \overline{y}_2^* , \quad \sigma_2^* = \sum_{1}^{n_2} \frac{(y_1 - \overline{y}_2^*)^2}{n_2} . \]

Here \( n_1 \) and \( n_2 \) are random quantities. The star (*) indicates that the estimators are conditional maximum likelihood estimators, conditional on \( n_1 \) and \( n_2 \) being fixed.
For notational convenience let us use $z_{1i}$ to indicate observations coming from $f_1(y)$ population and $z_{2i}$ for those from $f_2(y)$ population. Then the estimators $\hat{\mu}_1^*$ and $\hat{\mu}_2^*$ may be redefined as

$$\hat{\mu}_1^* = \sum_{i=1}^{n_1} z_{1i} / n_1 ; \quad \hat{\mu}_2^* = \sum_{i=1}^{n_2} z_{2i} / n_2 . \quad (8.13)$$

We can now have two estimators for $\mu$ because of the involvement of $P$.

8.6 Estimators of $\mu$

8.6.1 Case I: $P$ estimated from only the initial set.

As discussed in Chapter 6, $\hat{\mu} = \hat{P}\hat{\mu}_1^* + Q\hat{\mu}_2^*$ is an unbiased estimator of $\mu = P\mu_1 + Q\mu_2$. Note that $\hat{\mu}$ does not make use of the additional information $(x_{n+1}, x_{n+2}, \ldots, x_n)$ on $x$.

From Theorem 6.2 we have

$$V(\hat{\mu}) = \frac{1}{n} \left[ P\sigma_1^2 + Q\sigma_2^2 + PQ(\mu_1 - \mu_2)^2 \right] . \quad (8.13a)$$

8.6.2 Case II: $P$ estimated from all the observations.

Theorem 8.1

The estimator $\hat{\mu} = \hat{P}\hat{\mu}_1^* + (1-P)\hat{\mu}_2^* , \quad (8.14)$
where \( \hat{p} = \frac{n_1}{n'} \), is an unbiased estimator of \( \mu \) with sampling variance

\[
V(\hat{\mu}) = \frac{1}{n'^2} \left[ \sigma_1^2 E \left( \frac{n_1^2}{n_1} \right) + \sigma_2^2 E \left( \frac{n_2^2}{n_2} \right) + n'PQ(\mu_1 - \mu_2)^2 \right]. \tag{8.15}
\]

Proof:

Consider \( \hat{\mu} = \hat{p}\mu_1 + (1-\hat{p})\mu_2 \),

\[
= \hat{p} \left( \sum_{i=1}^{n_1} z_{1i}/n_1 \right) + (1-\hat{p}) \left( \sum_{i=1}^{n_2} z_{2i}/n_2 \right), \tag{8.16}
\]

where \( \hat{p} = \frac{n_1}{n'} \) and \( \hat{q} = (1-\hat{p}) = \frac{n_2}{n'} \).

In order to evaluate this expression, first consider

\[
E[\hat{p}(\sum_{i=1}^{n_1} z_{1i}/n_1)] = E[\frac{n_1}{n'} (\sum_{i=1}^{n_1} z_{1i}/n_1)],
\]

\[
E \left[ \frac{n_1}{n_1' n_1} \sum_{i=1}^{n_1} z_{1i} \mid n_1', n_1 \right] .
\]

We note that \( n_1' \) and \( n_1 \) are not independent since the set of \( n_1 \) observations is the subset of \( n_1' \) observations and that \( z_{1i} \) is for the observation \( y_i \) from the first group only.

Hence \( E[\hat{p}(\sum_{i=1}^{n_1} z_{1i}/n_1)] = E \left[ \frac{n_1}{n_1' n_1} \mu_1 \right] \),

\[
\frac{\mu_1}{n'} E \left( \frac{n_1'}{n_1} \right), \frac{\mu}{n', n_1'}
\]

\[
= P \mu_1 . \tag{8.17}
\]
Likewise \( E[(1-\hat{P})(\sum_{i=2}^{n_1} z_{1i})] = Q\mu_2 \) \hspace{1cm} (8.18)

Thus substituting (8.17) and (8.18) in (8.16) we have

\[ E(\hat{\mu}) = P\mu_1 + Q\mu_2 = \mu. \]

The variance of \( \hat{\mu} \)

From (8.14): \( V(\hat{\mu}) = V(\hat{\mu}_1^\prime) + V((1-\hat{P})\hat{\mu}_2^\prime) + 2\text{Cov}[(\hat{\mu}_1^\prime), (1-\hat{P})\hat{\mu}_2^\prime] \)

\( \hspace{1cm} (8.19) \)

Consider first \( V(\hat{\mu}_1^\prime) = V\left[ \frac{n_i^1 \sum_{i=1}^{n_1} z_{1i}}{n_1} \right] \).

It is to be recognised first \( \hat{\mu}_1^\prime \) is in reality an expression with three random variables \( n_i^1, n_1 \) and \( z_{1i} \). And it may also be noted that while \( n_i^1 \) has a range from 0 to \( n' \) (which includes \( n_1 \)), \( n_1 \) ranges from 0 to \( n(<n') \).

\[ V(\hat{\mu}_1^\prime) = E \sum_{i=1}^{n_i^1} \left[ \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right]^2 |n_1, n_i^1] + \]

\[ + E \sum_{i=1}^{n_i^1} \sum_{i=1}^{n_i^1} \left[ \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right]^2 |n_1, n_i^1] + \]

\[ + V \sum_{i=1}^{n_i^1} \sum_{i=1}^{n_i^1} \left[ \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right]^2 |n_1, n_i^1] \]

\[ = E \left[ E \left[ \frac{n_i^1}{n_1} \frac{\sigma_i^2}{n_1} \right] |n_1, n_i^1] \right] + E \left[ V \left[ \frac{n_i^1}{n_1} \mu_1 |n_i^1] \right] \right] + \]

\[ + V \left[ E \left[ \frac{n_i^1}{n_1} \mu_1 |n_i^1] \right] \right]. \]
\[
\begin{align*}
E_{n_1', n_1} & \left( \frac{n_1^2}{n_1} \right) \frac{\sigma_1^2}{n_1^2} + O + V_{n_1'} \left( \frac{n_1}{n_1'} \mu_1 \right), \\
E_{n_1', n_1} & \left( \frac{n_1^2}{n_1} \right) \frac{\mu_1^2}{n_1^2} V(n_1'), \\
\frac{\sigma_1^2}{n_1^2} E_{n_1} \left( \frac{n_1^2}{n_1} \right) + \frac{\mu_1^2}{n_1^2} V(n_1'), \\
\frac{\sigma_1^2}{n_1^2} E_{n_1} \left( \frac{n_1^2}{n_1} \right) + \frac{\mu_1^2}{n_1^2} V(n_1').
\end{align*}
\]

(8.20)

Similarly \( V\left((1-P)\hat{\mu}_2\right) = \frac{\sigma_2^2}{n_1^2} E_{n_2} \left( \frac{n_2^2}{n_2} \right) + \frac{P\mu_2^2}{n_1} \).

(8.21)

Consider \( \text{Cov} \left[ \hat{\mu}_1', (1-P)\hat{\mu}_2 \right] \)

\[
\begin{align*}
= \text{Cov} \left[ \frac{n_1}{n_1'} \frac{\Sigma z_{1i}}{n_1}, \frac{n_2}{n_2'} \frac{\Sigma z_{2i}}{n_2} \right], \\
= E \left[ \text{Cov} \left[ \frac{n_1}{n_1'} \frac{\Sigma z_{1i}}{n_1}, \frac{n_2}{n_2'} \frac{\Sigma z_{2i}}{n_2} \right] \right] | n_1', n_1 \] + \\
\text{Cov} \left[ \left[ E \left( \frac{n_1}{n_1'} \frac{\Sigma z_{1i}}{n_1} \right), E \left( \frac{n_2}{n_2'} \frac{\Sigma z_{2i}}{n_2} \right) \right] \right] | n_1', n_1 \],
\end{align*}
\]

[we note that if \( n_1 \) is held fixed, so is \( n_2 \) since \( n_1 + n_2 = n' \)]

\[
\begin{align*}
E \left[ \frac{n_1 n_2}{n_1'} \text{Cov}(\Sigma z_{1i}, \Sigma z_{2i}) \right] | n_1', n_1 \] + \\
\text{Cov} \left[ \left( \frac{n_1}{n_1'} \mu_1 \right), \left( \frac{n_2}{n_2'} \mu_2 \right) \right] | n_1', n_1 \]
\end{align*}
\]

\[
\frac{n_1 n_2}{n_1^2} \text{Cov}(n_1', n_2')
\]

\[
= \frac{PQ}{\mu_1 \mu_2}. \]

Substituting (8.20), (8.21) and (8.22) in (8.19) we have

\[ V(\hat{\mu}) = \frac{\sigma_1^2}{n_1^2} E\left(\frac{n_1^2}{n_1}\right) + \frac{\mu_1^2}{n_1} PQ + \frac{\sigma_2^2}{n_2^2} E\left(\frac{n_2^2}{n_2}\right) + \frac{\mu_2^2}{n_2} PQ - \frac{2PQ}{n_1} \mu_1 \mu_2, \]

\[ \therefore V(\hat{\mu}) = \frac{1}{n_1^2} \left[ \sigma_1^2 E\left(\frac{n_1^2}{n_1}\right) + \sigma_2^2 E\left(\frac{n_2^2}{n_2}\right) + n' PQ(\mu_1 - \mu_2)^2 \right], \]

in which \( E\left(\frac{n_1^2}{n_1}\right) \) and \( E\left(\frac{n_2^2}{n_2}\right) \) may be evaluated using Lemma (5.2) of Chapter 5.

Thus we have two estimates of \( \mu \) namely

1. \( \hat{\mu} = \hat{P}\hat{\mu}_1 + \hat{Q}\hat{\mu}_2 \), where \( \hat{P} = \frac{n_1}{n} \) from (8.13a)

with \( V(\hat{\mu}) = \frac{1}{n} \left[ P\sigma_1^2 + Q\sigma_2^2 + PQ(\mu_1 - \mu_2)^2 \right] \) and

2. \( \bar{\mu} = \hat{P}\hat{\mu}_1 + \hat{Q}\hat{\mu}_2 \), where \( \hat{P} = \frac{n_1}{n} \) and

\[ V(\bar{\mu}) = \frac{1}{n_1^2} \left[ \sigma_1^2 E\left(\frac{n_1^2}{n_1}\right) + \sigma_2^2 E\left(\frac{n_2^2}{n_2}\right) + n' PQ(\mu_1 - \mu_2)^2 \right], \]

from (8.15).

Both the variance expressions above are exact. However only \( V(\hat{\mu}) \) has a closed form while \( V(\bar{\mu}) \) does not have this since it involves \( E(\frac{n_1^2}{n_1}) \) and \( E(\frac{n_2^2}{n_2}) \) which can be evaluated at various orders of approximation. It may be readily seen that \( V(\bar{\mu}) \) reduces to \( V(\hat{\mu}) \) when \( n' = n \).
Comparison of the variances of two estimates

\[ V(\hat{\mu}) = \frac{1}{n} [P \sigma_1^2 + Q \sigma_2^2 + PQ(\mu_1 - \mu_2)^2] \quad \text{and} \]

\[ V(\bar{\mu}) = \frac{1}{n_1} [\sigma_1^2 + Q_1 \sigma_2^2 + P \sigma_1 \sigma_2 (\bar{r}_1 - \bar{r}_2)]^2 + \frac{1}{n_2} [\sigma_2^2 + Q_2 \sigma_1^2 + P \sigma_1 \sigma_2 (\bar{r}_2 - \bar{r}_1)]^2. \]

Hence

\[ V(\hat{\mu}) - V(\bar{\mu}) = \sigma_1^2 \left[ \frac{P}{n} - \frac{1}{n_1^2} E(\frac{n_1^2}{n_1}) \right] + \sigma_2^2 \left[ Q - \frac{1}{n_2^2} E(\frac{n_2^2}{n_2}) \right] + \]

\[ + PQ(\mu_1 - \mu_2)^2 \left[ \frac{1}{n} - \frac{1}{n_1} \right]. \]

Substituting for \( E(\frac{n_1^2}{n_1}) \) and \( E(\frac{n_2^2}{n_2}) \) from Lemma (n.2)

\[ V(\hat{\mu}) - V(\bar{\mu}) = \sigma_1^2 \left[ \frac{P}{n} - \frac{1}{n_1^2} \left( \frac{n_1^2 PQ + n_1^2 Q^2}{n} \right) \right] + \]

\[ + \sigma_2^2 \left[ Q - \frac{1}{n_2^2} \left( \frac{n_2^2 PQ + n_2^2 Q^2}{n} \right) \right] + PQ(\mu_1 - \mu_2)^2 \frac{(n_1 - n)}{n_1 n} + \]

\[ V(\hat{\mu}) - V(\bar{\mu}) = - \frac{Q_1^2}{n_1 n} - \frac{Q_2^2}{n_2 n} + \frac{P \sigma_1 \sigma_2}{n_1 n} + \frac{P \sigma_1 \sigma_2}{n_2 n} + PQ(\mu_1 - \mu_2)^2 \frac{(n_1 - n)}{n_1 n}. \]

\[ V(\hat{\mu}) - V(\bar{\mu}) \]

is positive if

\[ (n_1 - n) PQ(\mu_1 - \mu_2)^2 > Q_1^2 + Q_2^2, \]

i.e.

\[ (n_1 - n) > \frac{Q_1^2 + Q_2^2}{PQ(\mu_1 - \mu_2)^2}. \] (8.22a)

Hence the number of additional units to be drawn should satisfy this condition in order to have a reduction in sampling variance.
**THEOREM 8.2**

The estimator \( \hat{N}(\mu_1 - \mu) \) is unbiased for the total crop loss with

\[
V[N(\hat{\mu}_1 - \mu)] = N^2 \left[ \sigma_1^2 \left( \frac{1}{n_1} \right) - 2E\left( \frac{n_1}{n_1} \right) + \frac{1}{n_1^2} E\left( \frac{n_1^2}{n_1} \right) \right] +
\]

\[
+ \sigma_2^2 \frac{1}{n_2} E\left( \frac{n_2^2}{n_2} \right) + \frac{pq}{n} \left( \mu_1 - \mu_2 \right)^2 .
\]  

\[
(8.24)
\]

**Proof:**

\[
E[N(\hat{\mu}_1 - \mu)] = N[E(\hat{\mu}_1) - E(\mu)] ,
\]

\[
= N(\mu_1 - \mu) .
\]

\[
V[N(\hat{\mu}_1 - \mu)] = N^2[V(\hat{\mu}_1) + V(\mu) - 2Cov(\hat{\mu}_1, \mu)] .
\]  

\[
(8.25)
\]

In order to evaluate (8.25) consider first

\[
V(\hat{\mu}_1) = V\left( \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right) ,
\]

\[
+ \frac{E \left[ V \left( \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right) \right]}{n_1} + V \left[ \frac{E \left( \frac{\sum_{i=1}^{n_1} z_{1i}}{n_1} \right)}{n_1} \right]
\]

\[
+ \frac{E \left[ \frac{\sigma_1^2}{n_1} \right]}{n_1} + V \left[ \frac{\mu_1}{n_1} \right] ,
\]

\[
= \sigma_1^2 E\left( \frac{1}{n_1} \right) .
\]  

\[
(8.26)
\]

The other variance \( V(\mu) \) is given in (8.15).
Next consider $Cov(\hat{\mu}_1, \bar{\mu})$

$$= E[\hat{\mu}_1 (\hat{\mu}_1 + Q \hat{\mu}_2)] - E(\hat{\mu}_1) E(\bar{\mu}),$$

$$= E(\hat{\mu}_1^2) + E(Q \hat{\mu}_1 \hat{\mu}_2) - (E \mu_1 + Q \mu_2),$$

$$= E \left[ E \left( \hat{\mu}_1^2 | n_1, n_1 \right) \right] + E \left[ E \left( Q \hat{\mu}_1 \hat{\mu}_2 \right) \right] - P \mu_1^2 - Q \mu_1 \mu_2,$$

$$= E \left[ \hat{\mu}_1^2 \right] + E \left( Q \mu_1 \mu_2 \right) - P \mu_1^2 - Q \mu_1 \mu_2,$$

$$= E \left[ \frac{\sigma^2}{n_1} + \mu_1^2 \right] + Q \mu_1 \mu_2 - P \mu_1^2 - Q \mu_1 \mu_2,$$

$$= \frac{\sigma^2}{n_1} E(\frac{n_1}{n_1}) + P \mu_1^2 - P \mu_1^2,$$

$$= \frac{\sigma^2}{n_1} E(\frac{n_1}{n_1}).$$

Substituting (8.26), (8.15) and (8.27) in (8.25) we have

$$V[N(\hat{\mu}_1 - \bar{\mu})] = N^2 \left[ \sigma^2 E(\frac{1}{n_1}) + \frac{1}{n_1} \sigma^2 E(\frac{n_1}{n_1}) + \sigma^2 E(\frac{n_2}{n_2}) + \sigma^2 E(\frac{n_1}{n_1}) \right] + n'PQ(\mu_1 - \bar{\mu})^2 - \frac{2\sigma^2}{n_1} E(\frac{n_1}{n_1}).$$
\[ N^2 \left[ \sigma_1^2 \left( \frac{1}{n_1} - \frac{2E}{n_1} \frac{n_1^1}{n_1} \right) + \frac{1}{n_1^2} E\left( \frac{n_1^1}{n_1} \right) \right] \\
+ \sigma_2^2 \frac{1}{n_1^2} E\left( \frac{n_2^1}{n_2} \right) + \frac{PQ}{n_1} (\mu_1 - \mu_2)^2 \].

in which \( E\left( \frac{n_2^1}{n_2} \right) \) may be evaluated using (5.1).

**Remark**: Expression (8.24) is for the variance of the crop loss estimator under the model discussed in section 8.4. While (6.32) is the corresponding expression when only a sample of fixed size \( n \) is drawn.