CHAPTER - IV

MISCLASSIFIED DATA IN CURTAILED SAMPLING PLANS

4.1 In this chapter we have obtained the maximum likelihood estimate of the fraction defective when data from the curtailed sampling plans are subject to misclassification. The maximum likelihood estimate of the probability of misclassification is also obtained. The asymptotic variances and covariances of these estimates are derived.

4.2 Description of Misclassification :

Rejection of a lot involves sometimes a botheration. An immediate consequence of the rejection of a lot leads to inspection of all the articles of a lot, when screening is prevailing. Rejection of a lot may raise undue doubt, whether the quality of the production has deteriorated. The party concerned, therefore, may avoid the rejection of a lot by practising deliberative errors in inspection. In Plan 1, we know that a lot is rejected if the number of defectives in n inspected articles is c+1 (=k) or more. Therefore, an inspector is inclined to classify a defective as an nondefective, when he finds exactly c+1 defectives in the articles inspected;

for this will lead to acceptance of a lot and avoidance of the further bothersome consequences. Cohen [10,11] has considered this type of misclassification. He has obtained the maximum likelihood estimate of the fraction defective. when data of Plan 1 are subject to this type of misclassification.

We consider here the same type of misclassification, namely, the misclassification which leads to acceptance of a lot when it is being inspected under the Curtailed Sampling Plans. Let us try to determine the stage at which the inspector may classify a defective as a nondefective which will lead to success in his manipulation. Recalling the statements of the Curtailed Sampling Plans, Plans 1 and 2, it is evident that if the inspector classifies a defective as a nondefective when the kth defective appears at the nth inspection, it will lead to acceptance of the lot. We assume that the inspector does so with probability θ.

However, it should be noted that in Plan 3 misclassification at an earlier stage of the inspection could have resulted in both the acceptance of the lot and the curtailing of the inspection. For instance, let us consider the case, n=30 and k=5. Say, the inspector has noted not a single defective in 25 inspected articles. Further he finds that the

26th article is defective. He is then supposed to continue inspection in accordance with the statement of the plan. However, had he classified the 26th article as nondefective he could have stopped the inspection, resulting in the acceptance of the lot. It may be very unusual to find too much disloyalty in an inspector. It is, therefore, assumed that he does not misclassify at that stage. We assume that he classifies a defective as a nondefective when his misclassification will not lead to curtailment of the inspection but will merely lead to the acceptance of a lot which otherwise would have been rejected, had the correct classification been rigidly followed. The results of the estimation are obtained when the data related to only this type of misclassification are presented.

4.3 Furthermore, it is assumed that the inspector gives complete information of the inspection. Thus, we have considered only the results belonging to Situation A under misclassification. The other assumptions related to p, the lot size etc. stated in Section 2.3 are valid here also.

4.4 Probability Functions :

Recalling the definitions of the random variables x,y,z, i, and s given in Section 2.3 we have the following probability functions of the random variable s subject to the misclassification discribed above, in the respective plans:

Plan 2:

$$m_2(s)=\begin{cases}\binom{n}{x}p^x q^{n-x} & s=x=0,1,\ldots,k-2,\\ \binom{n}{k-1}p^{k-1}q^{n-k+1}\left[1+(n-k+1)p\theta/nq\right] & s=x=k-1,\\ \binom{y-1}{k-1}p^k q^{y-k} & s=y=k,k+1,\ldots,n-1,\\ (1-\theta)\binom{n-1}{k-1}p^k q^{n-k} & s=y=n,\end{cases} \qquad \ldots(4.1)$$

Plan 3 :

$$m_3(s)=\begin{cases}\binom{n-k+i}{n-k}q^{n-k+1}p^i & s=i=0,1,\ldots,k-2,\\ \binom{n-1}{n-k}q^{n-k+1}p^{k-1}(1+\theta p/q) & s=i=k-1,\\ \binom{y-1}{k-1}p^k q^{y-k} & s=y=k,k+1,\ldots,n-1,\\ (1-\theta)\binom{n-1}{k-1}p^k q^{n-k} & s=y=n,\end{cases} \qquad \ldots(4.2)$$

where $q=1-p$, $0 \leq p \leq 1$ and $0 \leq \theta \leq 1$.

4.5 MLE of p and θ :

We obtain the maximum likelihood estimates of the fraction defective and the probability of misclassification when T lots have undergone the inspection under the curtailed Sampling Plans subject to the misclassification described above. Each accepted or rejected lot will give rise to one observation associated with x,y,z, or i which belongs to one of the populations defined by (4.1) and (4.2). The observed frequencies associated with these variables are $a_{x,2}$, $r_{y,2}$, $a_{i,3}$(or $a_{z,3}$) and $r_{y,3}$, whose exact definitions are given in Section 2.6.1. However, it should be noted that the observed frequency $a_{k-1,2}$ is the mixture of the true frequency for x=k-1 and the frequency arising due to misclassification. The same applies to $a_{k-1,3}$. Similarly the observed frequency $r_{n,j}$ for j=2 and 3 is the reduced frequency due to misclassification.

Then in the notations of the classical theory of estimation based on the fixed sample size, we have T observations ($T=T_{a,j}+T_{r,j}$) on the random variable s from the populations $m_j(s)$ (j=2,3). The likelihood function based on T such observations will be as follows:

Plan 2:

$$L_2 = \prod_{x=0}^{k-1} \left[\binom{n}{x} p^x q^{n-x}\right]^{a_{x,2}} \prod_{y=k}^{n} \left[\binom{y-1}{k-1} p^k q^{y-k}\right]^{r_{y,2}}$$

$$\cdot \left[1+(n-k+1)\theta p/nq\right]^{a_{k-1,2}} (1-\theta)^{r_{n,2}}$$

Plan 3:

$$L_3 = \prod_{i=0}^{k-1} \left[\binom{n-k+i}{n-k} q^{n-k+1} p^i\right]^{a_{i,3}} \prod_{y=k}^{n} \left[\binom{y-1}{k-1} p^k q^{y-k}\right]^{r_{y,3}}$$

$$\cdot (1+p\theta/q)^{a_{k-1,3}} (1-\theta)^{r_{n,3}}$$

Equating the first partial derivatives of the likelihood functions with respect to p and θ to zero, we get the following equations for estimating p and θ.

Plan 2:

$$p^2(k-1)(n\,T_{a,2}+D_{r,2})$$

$$+\,p\left[(n-k+1)\,a_{k-1,2}+nr_{n,2}-n(nT_{a,2}+D_{r,2})\right.$$

$$\left.-(k-1)(D_{a,2}+kT_{r,2})\right]$$

$$+\,n(D_{a,2}+kT_{r,2})-nr_{n,2}=0 \qquad \ldots(4.3)$$

and $\theta = \left[(n-k+1)pa_{k-1,2}-nqr_{n,2}\right]/\,p(n-k+1)(a_{k-1,2}+r_{n,2})$ ...(4.4)

Plan 3:

$$p=\frac{D_{a,3}+kT_{r,3}-r_{n,3}}{D_{a,3}+(n-k+1)T_{a,3}+D_{r,3}-r_{n,3}-a_{k-1,3}}$$

and $\theta = (pa_{k-1,3}-qr_{n,3})/p(a_{k-1,3}+r_{n,3})$ ...(4.5)

where

$$D_{a,2}=\sum_{x=0}^{k-1}xa_{x,2}\,, \qquad D_{r,j}=\sum_{y=k}^{n}yr_{y,j}\,(j=2,3),$$

$$D_{a,3}=\sum_{i=0}^{k-1}ia_{i,3}$$

4.6 Variance Covariance Matrix of the Estimate ($\hat{p}$ and $\hat{\theta}$):

We need the following expectations to compute the variances and covariances of the estimates,

$$E\,(D_{a,2}+kT_{r,2})/T = J_1 - p\,m_2(k-1)/A_2 \qquad \ldots(4.6)$$

$$pE\,(nT_{a,2}+D_{r,2})/T = J_2 \qquad \ldots(4.7)$$

$$E\,(D_{a,3}+kT_{r,3})/T = J_3 - p\,m_3(k-1)/A_3 \qquad \ldots(4.8)$$

and

$$pE\,[D_{a,3}+(n-k+1)\,T_{a,3}+D_{r,3}]/T = J_4 \qquad \ldots(4.9)$$

where (i) $A_2 = p+nq/(n-k+1)\theta$, $A_3 = p+q/\theta$,

(ii) $m_2(k-1)$ and $m_3(k-1)$ are respectively the values of $m_2(s)$ and $m_3(s)$ at $s=k-1$,

and (iii) J_1, J_2, J_3 and J_4 are as defined in (3.7), (3.8), (3.9) and (3.10) of Chapter III.

Further, noting that $J_1=J_2$ and $J_3=J_4$ and using the results (4.6) through (4.9) the expectations of the second derivatives of the likelihood functions are found to be :

<u>Plan 2</u>:

$$-E(\partial^2\log L_2/\partial p^2)/T = J_1/p^2q+(1/q^2A_2)(1/A_2-1/p)m_2(k-1)=\phi_{11} \qquad \ldots(4.10)$$

$$-E(\partial^2\log L_2/\partial p\,\partial\theta)/T = -nm_2(k-1)/(n-k+1)A_2^2\theta^2=\phi_{12}=\phi_{21} \qquad \ldots(4.11)$$

and

$$-E(\partial^2\log L_2/\partial\theta^2)/T = m_2(n)/(1-\theta)^2+p^2m_2(k-1)/A_2^2\theta^2=\phi_{22} \qquad \ldots(4.12)$$

Plan 3:

$$- E(\partial^2 \log L_3/\partial p^2)/T = J_3/p^2 q + (1/q^2 A_3)(1/A_3 - 1/p)m_3(k-1) = \phi_{11} \quad \ldots(4.13)$$

$$- E(\partial^2 \log L_3/\partial p \partial \theta)/T = -m_3(k-1)/A_3^2\theta^2 = \phi_{12} = \phi_{21} \quad \ldots(4.14)$$

and

$$- E(\partial^2 \log L_3/\partial \theta^2)/T = m_3(n)/(1-\theta)^2 + p^2 m_3(k-1)/A_3^2\theta^2 = \phi_{22} \quad \ldots(4.15)$$

Then the variance covariance matrix of $\hat{p}$ and $\hat{\theta}$ is given by $[\phi_{ij}]^{-1}/T$ i.e. the asymptotic variances and co-variance are :

$$V(\hat{p}) = \phi_{22}/(\phi_{11}\phi_{22} - \phi_{12}^2)T \quad \ldots(4.16)$$

$$V(\hat{\theta}) = \phi_{11}/T(\phi_{11}\phi_{22} - \phi_{12}^2) \quad \ldots(4.17)$$

and

$$\text{Cov}(\hat{p},\hat{\theta}) = -\phi_{12}/T(\phi_{11}\phi_{22} - \phi_{12}^2) \quad \ldots(4.18)$$

4.7 Particular Cases $\theta = 1$ and $\theta = 0$:

4.7.1 Case $\theta = 1$. When the misclassification is carried with certainty the observed frequency $r_{n,j}$ $(j=2,3)$ will be zero. The probability functions can be easily obtained by substituting $\theta=1$ in the expressions of $m_j(s)$ given by (4.1) and (4.2). The case then reduces to estimation from a population with one parameter. The maximum likelihood equation for estimating p and variance of the estimate may be derived in the usual way. They are :

Plan 2 :

$$p^2(k-1)(nT_{a,2}+D_{r,2})$$
$$+\ p\left[(n-k+1)a_{k-1,2} - n(nT_{a,2}+D_{r,2}) - (k-1)(D_{a,2}+kT_{r,2})\right]$$
$$+\ n(D_{a,2}+kT_{r,2}) = 0 \qquad \ldots(4.19)$$

and

$$\left[TV(\hat{p})\right]^{-1} = J_1/p^2q + (1/q^2A_2')(1/A_2'-1/p)m_2'(k-1) \qquad \ldots(4.20)$$

where

$$A_2' = p+nq/(n-k+1),\quad m_2'(k-1) = \binom{n}{k-1}p^{k-1}q^{n-k+1}$$
$$.\ \left[1+(n-k+1)p/nq\right]$$

Plan 3 :

$$p = \frac{D_{a,3} + kT_{r,3}}{D_{a,3}+(n-k+1)T_{a,3}+D_{r,3}-a_{k-1,3}} \qquad \ldots(4.21)$$

and

$$\left[TV(\hat{p})\right]^{-1} = J_3/p^2q + (1/q^2)(1-1/p)m_3'(k-1) \qquad \ldots(4.22)$$

where

$$m_3'(k-1) = \binom{n-1}{n-k}q^{n-k+1}p^{k-1}(1+p/q)$$

4.7.2 Case $\theta=0$. It is the case of correct classification i.e. the case where there is no misclassification. The case then reduces to the case dealt in Section 3.2 .

4.8 A Numerical Example :

We illustrate the results of Plan 2 by an example. Table 4.1 gives the tabulation of 200 observations associated with the inspection of 200 lots under Plan 2 with n=15 and k=3. It is assumed that the data are subject to misclassification of the type considered here.

Now from the table below we find $T_{a,2}=117$, $T_{r,2}=83$, $D_{a,2}=155$, $D_{r,2}=796$, $a_{k-1,2}=54$ and $r_{n,2}=3$. Substituting these values in equations (4.3) and (4.4) we get p=0.1604 and θ=0.629. Now the hypothetical values of p and θ used for obtaining the above random sample of 200 observations are 0.15 and 2/3 respectively. Substituting these hyopthetical values of p and θ in (4.10), (4.11) and (4.12) we find $\phi_{11}=102.417$, $\phi_{12}=0.31093$, $\phi_{22}=0.13710$. Then we have $V(\hat{p}) \doteq 0.00004916$, $V(\hat{\theta}) \doteq 0.036725$ for T=200. Therefore, $S.E.(\hat{p}) \doteq 0.00701$ and $S.E.(\hat{\theta}) \doteq 0.1916$. The absolute difference between $\hat{p}$ and p, and $\hat{\theta}$ and θ may be attributed due to sampling fluctuations since we observe that

$$|\hat{p} - p| = 1.48 \; S.E.(\hat{p}) \quad \text{and} \quad |\hat{\theta} - \theta| = 0.20 \; S.E.(\hat{\theta})$$

In practice, one may use $\hat{p}$ and $\hat{\theta}$ to compute the estimates of the asymptotic variances and covariance when one does not know the true values of p and θ. The use of the Tables of Binomial Proabability Distribution [55] has been made for the above computations.

Table 4.1

Number of defectives	Number of accepted lots	Number of articles inspected	Number of rejected lots	Number of articles inspected	Number of rejected lots
(1)	(2)	(3)	(4)	(3)	(4)
0	16	3	1	10	14
1	47	4	3	11	6
2	54	5	9	12	9
-	-	6	7	13	13
-	-	7	3	14	3
-	-	8	5	15	3
-	-	9	7	-	-
Total	117	Total	-	-	83