ERRORS IN INSPECTION AND GRADING: DISTRIBUTIONAL ASPECTS OF SCREENING AND HIERARCHAL SCREENING

by

Samuel Kotz
University of Maryland

and

Norman L. Johnson
University of North Carolina at Chapel Hill

ABSTRACT

Continuing the studies of Johnson *et al* (1980) and Johnson and Kotz (1981), further distributions arising from models of errors in inspection and grading of samples from finite, possibly stratified lots are obtained. Screening, and hierarchal screening forms of inspection are also considered, and the effects of errors on the advantages of these techniques assessed.

NOTATION SUMMARY

Single Sampling

p Probability that a defective item is classified as defective

p' Probability that a nondefective item is classified as defective

$\bar{p} = N^{-1}\{Dp + N\text{-}D)p'\}$ = Probability that an individual chosen at random is classified as defective.

Hierarchal Sampling

n_j Number of items tested *en bloc* at j-th stage

p_j Probability that a group containing at least one defective is classified as defective, at j-th stage.

p_j' Probability that a group containing no defectives is classified as defective, at j-th stage

$P_0(n_j)$ = Probability that a random sample of n_j items contains no defective items (j = 1, 2).

$P_0^*(n_j)$ = Probability that a random sample of n_j items contains no defective items, given that it contains at least one nondefective.

$$P_{(j)} = (p_j - p_j')P_0(n_j)$$

$$P_{(j)}^* = (p_j - p_j')P_0^*(n_j)$$

1. INTRODUCTION

For convenience, we first summarize some results obtained by Johnson et al. (1980) and Johnson & Kotz (1981).

A random sample of size n is chosen (without replacement) from a lot of size N, which contains D defective (or nonconforming) members. Suppose that on inspection of items in the sample, the probability that a defective item will be classified as defective is p, while the probability that a non-defective item will (erroneously) be classified as "defective" is p'.

If Y denotes the actual unknown number of defective items in the sample, the distribution of Y is hypergeometric with parameters (n, D, N). If X denote the number (among these Y) which are (correctly) classified as defective, and X' the number (among the (n-Y) nondefective items in the sample) which are (incorrectly) classified as "defective" then, conditionally on Y, the variables X and X' are independent binomial with parameters (Y,p) and (n-Y, p') respectively. Averaging over the distribution of Y, the distribution of

$$Z = X + x',$$

the total number of items described as "defective" after inspection of the sample, can be formally expressed as

$$\text{Binomial } (Y,p) * \text{Binomial } (n-Y,\ p') \underset{\underset{\sim}{Y}}{\wedge} \text{Hypergeometric } (n,D,N) \tag{1}$$

The symbol * stands for convolution, and the symbol $\underset{\underset{\sim}{Y}}{\wedge}$ indicates the "corresponding" operation with Y distributed as "(see, e.g. Johnson & Kotz (1969, Chapter 8)). Distribution (1) is a mixture of convolutions of two binomial distributions. The r-th descending factorial moment of Z is

$$\mu_{(r)}(Z) = E(Z^{(r)}) = \frac{n^{(r)}}{N^{(r)}} \sum_{j=0}^{r} \binom{r}{j} p^j p'^{r-j} D^{(j)} (N-D)^{(r-j)} \tag{2}$$

where $Z^{(r)} = Z(Z-1)\ \ldots\ (Z-r+1)$. In particular

$$E(Z) = n\bar{p}/N \tag{3.1}$$

$$\mathrm{var}(Z) = n\bar{p}(1-\bar{p}) - \frac{n(n-1)}{N-1} \cdot \frac{D}{N}\left(1-\frac{D}{N}\right)(p-p')^2 \tag{3.2}$$

where $\bar{p} = \{Dp + (N-D)p'\}N^{-1}$ is the probability that a single individual, chosen at random from the lot, will be described as "defective" after inspection (whether it is really defective or not). Tables of the distribution of Z for n=10 with p = 0.75(0.05)0.95; p' = 0(0.025)0.1; N = 100 and D = 5, 10, 20; N = 200 and D = 10, 20, 40 are presented in Johnson and Kotz (1981). If p' = 0, we have the situation described in Johnson et al. (1980) and (1) becomes a hypergeometric-binomial distribution. For fixed values of $D/N = \lambda$ say, the distribution (1) is quite sensitive to changes in p and p', but not to changes in D and N unless n/N (the sampling fraction) is large. It is easy to see that as D, N $\to \infty$ (with $D/N = \lambda$) the distribution of Z tends to binomial with parameters $(n,\lambda p+1(1-\lambda)p')$.

Johnson and Kotz (1981) also, *inter alia*, consider lots divided into k strata $\Pi_1, \ldots, \Pi_k$ of sizes $N_1, \ldots, N_k$ (with $\sum_{j=1}^{k} N_j=N$) and suppose that the probability of an individual from the j-th stratum being classified as defective is p_j. (The situation considered at the beginning of this section corresponds to k=2, $N_1=D$, $N_2=N-D$, $p_1 = p$, $p_2 = p'$.) The distribution of the total number (Z) classified as defective is (in an obvious notation)

$$\mathop{*}_{j=1}^{k} \text{Binomial } (Y_j,P_j) \bigwedge_{\underset{\sim}{Y}} \text{multivariate Hypergeometric } (n,\underset{\sim}{N},N)$$

($\underset{\sim}{Y}=(Y_1, \ldots Y_k)$ is the vector of numbers from $\Pi_1, \ldots, \Pi_k$ in the sample - $Y_1 + \ldots + Y_k = n$) (4)

The r-th factorial moment of Z is

$$\frac{n^{(r)}}{N^{(r)}} \sum_r{}' \binom{r}{r_1, r_2, \ldots, r_k} \prod_{j=1}^{k} \left(p_j^{r_j} N_j^{(r_j)}\right) \tag{5}$$

where $\sum_r'$ denotes summation subject to $r_1 + \ldots + r_k = r$ and $\binom{r}{r_1, r_2, \ldots, r_k}$ is the multinomial coefficient $r!/\left(\prod_{j=1}^{k} r_j!\right)$.

In particular, with $\bar{p} = N^{-1} \sum_{j=1}^{k} N_j p_j$ (again, denoting the probability that an individual chosen at random is classified as "defective"

$$E(Z) = n\bar{p} \tag{6.1}$$

$$\text{var}(Z) = n\bar{p}(1-\bar{p}) - \frac{n(n-1)}{n(N-1)} \sum_{j=1}^{k} N_j (p_j - \bar{p})^2 \tag{6.2}$$

(Note that var(Z) is usually less than the "binomial" value $n\bar{p}(1-\bar{p})$.)

2. GRADING

Formulae (4)-(6) can be regarded as generalizations of (1)-(3). Further useful generalization is obtained by considering judgement not to be restricted to "defective" or "non-defective" but to include assignment to one of several categories - as would be the case, if product were being graded in terms of quality and/or size with regard to marketing. (One aspect of this situation is the multivariate topic of disiminent analysis, but here we are concerned with the consequences rather than the methods of assignment.)

We will analyze a situation in which the aim of judgement is assign an individual to one of s classes $C_1, \ldots, C_s$. We denote the probabilith that an individual, who really belongs to C_j will be assigned to C_i by P_{ij}, with, of course $\sum_{i=1}^{s} P_{ij} = 1$. Still further generalization is possible by introducing stratification within each class. This leads to straightforward, but notationally complicated elaboration, and will not be pursued here.

Let Y_j denote the number of individuals belonging to C_j, in a random sample of size n for a lot of size N containing N_j individuals in C_j (j=1, ..., s; $\sum_{i=1}^{s} N_i = N$); and let Z_{ij} denote the number, among these Y_j, assigned to

C_i (i=1, ..., s). The $\underset{\sim}{Y}$ has a multivariate hypergeometric distribution with parameters $(n;\underset{\sim}{N};N)$, so that

$$P(\underset{\sim}{Y}) = \left\{ \prod_{j=1}^{s} \binom{N_j}{Y_j} \right\} / \binom{N}{n} \qquad \left(\sum_{j=1}^{s} Y_j = n \right) \tag{7}$$

Also, given $\underset{\sim}{Y}$, $\underset{\sim}{Z}_j = (Z_{ij}, \ldots, Z_{sj})$ has a multinomial distribution with parameters $(Y_j, \underset{\sim}{P}_j)$ where $\underset{\sim}{P}_j = (P_{ij}, \ldots, P_{sj})$, and

$$P(\underset{\sim}{Z}_j | \underset{\sim}{Y}) = Y_j! \prod_{i=1}^{s} (P_{ij}^{Z_{ij}} / Z_{ij}!) \qquad \left(\sum_{i=1}^{s} Z_{ij} = Y_j \right) \tag{8}$$

and $\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_s$ are mutually independent. It follows that the joint distribution of all the Z's, $\underset{\sim}{Z} = (\underset{\sim}{Z}_1, \ldots, \underset{\sim}{Z}_s)$ is

$$P(\underset{\sim}{Z}) = P(\underset{\sim}{Z}|\underset{\sim}{Y})P(\underset{\sim}{Y}) = \binom{N}{n}^{-1} \prod_{j=1}^{s} \left\{ N_j^{(z_{ij})} \prod_{i=1}^{s} (P_{ij}^{Z_{ij}} / Z_{ij}!) \right\} \tag{9}$$

where $Z_{ij} = \sum_{i=1}^{s} Z_{ij}$. (Note that the Y_j's are *determined* by the Z_{ij}'s.)

Formally we can write the distribution of $\underset{\sim}{Z}$ as

$$\mathop{*}_{j=1}^{s} \text{Multinomial } (Y_j, \underset{\sim}{P}_j) \bigwedge_{\underset{\sim}{Y}} \text{ Multivariate Hypergeometric}(n, \underset{\sim}{N}, N) \tag{10}$$

and it might be called a "multivariate hypergeometric-multinomial distribution.

The joint factorial moments of $\underset{\sim}{Z}$ can be obtained from

$$E\left(\prod_{i=1}^{s} \prod_{j=1}^{s} Z_{ij}^{(r_{ij})} \right) = E_{\underset{\sim}{Y}} \left[E\left(\prod_{j=1}^{s} \prod_{i=1}^{s} Z_{ij}^{(r_{ij})} \Big| \underset{\sim}{Y} \right) \right] = E \left(\prod_{j=1}^{s} Y_j^{(r_{ij})} \prod_{i=1}^{s} P_{ij}^{r_{ij}} \right)$$

$$= \frac{n^{(r..)}}{N^{(r..)}} \prod_{j=1}^{s} \left\{ N_j^{(r._j)} \prod_{i=1}^{s} P_{ij}^{r_{ij}} \right\} \tag{11}$$

where $r._j = \sum_{i=1}^{s} r._j$; $r.. = \sum_{i=1}^{s} \sum_{j=1}^{s} r_{ij}$. In particular, with $i \neq i_1' j \neq j'$

$$E(Z_{ij}) = nN^{-1}N_jP_{ij} \tag{12.1}$$

$$E(Z_{ij}Z_{i'j}) = n^{(2)}N_j^{(2)}P_{ij}P_{i'j}/N^{(2)} \tag{12.2}$$

$$E(Z_{ij}Z_{i'j'}) = n^{(2)}N_jN_{j'}P_{ij}P_{i'j'}/N^{(2)} \tag{12.3}$$

whence

$$\text{cov}\ (Z_{ij},Z_{i'j}) = -nN_j\{(N-n)N_j+N(n-1)\}\ P_{ij}P_{i'j}N^{-2}(N-1)^{-1} \tag{13.1}$$

$$\text{cov}\ (Z_{ij},Z_{i'j'}) = -nN_jN_{j'}(N-n)P_{ij}P_{i'j'}N^{-2}(N-1)^{-1} \tag{13.2}$$

Using these results we can find the covariance between $Z_{i\cdot} = \sum_{j=1}^{s} Z_{ij}$ and $Z_{i'\cdot} = \sum_{j=1}^{s} Z_{i'j}$ - that is the total number of individuals assigned to C_i, C_j, respectively,

$$\text{cov}\ (Z_{i\cdot},Z_{i'\cdot}) = -n\{(N-n)\overline{P}_i\overline{P}_{i'} + (n-1)(\overline{P_i P_{i'}})\} \tag{14}$$

where

$$\overline{P}_i = N^{-1}\sum_{j=1}^{s} N_jP_{ij};\quad (\overline{P_iP_{i'}}) = N^{-1}\sum_{j=1}^{s} N_jP_{ij}P_{i'j}$$

($\overline{P}_i$ is the probability that an individual chosen at random is assigned to C_i, $(\overline{P_iP_{i'}})$ is the probability that an individual chosen at random would be assigned to C_i on one judgement and $C_{i'}$ on another.). The marginal distribution of $Z_{i\cdot}$ is like (4) with k replaced by s and p_j by P_{ij}., so we can use the corresponding formulae for moments.

3. GROUP SCREENING

Further interesting distributions arise in connection with "group screening" (Dorfman (1943)), in which groups of units can be tested for the existence of one or more defective units among them. This can be practicable, for example, when testing liquids for presence of contaminants, and is then suggested as a possible way of reducing the average total amount of testing.

Suppose that material from n_1 units is mixed and tested for presence of "defective" material. If a negative result ("no defectives") is obtained, no further action is taken, but if there is a positive result, each unit is tested separately.

Let p_1, p_1' denote the probabilities of obtaining correct or incorrect positive results, respectively, at the first test. As before, p, p' denote the probabilities of correct or incorrect positive results, respectively, when units are tested individually; D, N denote the number of defective units and the total number of units in the population respectively, and Y denotes the actual number of defective units among the n_1 tested.

The overall probability of obtaining a positive result on the first test is

$$\{1-P_o(n_1)\}p_1 + P_o(n_1)p_1' = p_1 - (p_1-p_1')P_o(n_1) \qquad (15)$$

where $P_o(n_1) = (N-D)^{(n_1)}/N^{(n_1)}$ is the probability that the sample contains no defective items.

The distribution of Z can be represented as

$$(\text{Binomial }(WY,p) * \text{Binomial}(W(n_1-Y),p')) \underset{\underset{\sim}{Y}}{\wedge} \text{Hypergeometric }(n_1,D,N) \qquad (16)$$

where W is an indicator variable, defined by

$$W = \begin{cases} 1 & \text{if the first test gives a positive result} \\ 0 & \text{otherwise} \end{cases}$$

Denote that

$$P(W=1)|Y>0) = p_1 \ ; \quad P(W=1)|Y=0) = p_1' \qquad (17)$$

An explicit formula for P(Z=z) is

$$P(Z-z) = E_{W,Y}\left[\sum_x{}' \binom{WY}{x} p^x (1-p)^{WY-x} \binom{W(n_1-Y)}{z-x} p'^{3-x}(1-p')^{W(n_1-Y)-z+x}\right]$$

where $\sum_x'$ denotes summation over $(0,z-W(n_1-Y)) \le x \le \min(WY,z)$.

We have

$$E(Z^{(r)}|W,Y) = \sum_{i=0}^{r} \binom{r}{i} (WY)^{(i)} \{W(n_1-Y)\}^{(r-i)} p^i p'^{r-i}$$

Noting (17) and

$$E(Y^{(i)}(n_1-Y)^{(r-i)}|Y>0)P(Y>0) = \begin{cases} n_1^{(r)} D^{(i)} (N-D)^{(r-i)}/N^{(r)} & \text{for } i>0 \\ n_1^{(r)} [(N-D)^{(r)}\{N^{(r)}\}^{-1} - P_0(n_1)] & \text{for } i = 0 \end{cases} \qquad (18.1)$$

$$E(Y^{(i)}(n_1-Y)^{(r-i)}|Y=0)P(Y=0) \begin{cases} 0 & \text{for } i > 0 \\ n_1^{(r)} P_0(n_1) & \text{for } i = 0 \end{cases} \qquad (18.2)$$

we find

$$E(Z^{(r)}) = n_1^{(r)} \{p_1 (N^{(r)})^{-1} \sum_{i=0}^{r} \binom{r}{i} D^{(i)} (N-D)^{(r-i)} p^i p'^{r-i} - (p_1-p_1')p'^r P_0(n_1)\} \qquad (19.1)$$

In particular

$$E(Z) = n_1\{p_1\bar{p} - (p_1-p_1')p'P_0(n_1)\} \qquad (19.2)$$

$$= n_1(p_1\bar{p} - P_{(1)}p')$$

where $P_{(1)} = (p_1-p_1')P\ (n_1)$

and

$$\text{var}(Z) = n_1^{(2)}\left[p_1\left\{\bar{p}^2 - \frac{D(p^2-\bar{p}^2)+(N-D)(p'^2-\bar{p}^2)}{N\ (N-1)}\right\} - P_{(1)}p'^2\right]$$

$$+ n_1(p_1\bar{p} - P_{(1)}p') - n_1^2(p_1\bar{p} - P_{(1)}p')^2 \qquad (19.3)$$

If there is no possibility of a "false positive" (so that $p_1'p' = 0$) we find

$$E(Z) = n_1 p_1 pD/N$$

$$\text{Var}(Z) = \frac{n_1 p_1 pD}{N}\left(1 - \frac{p_1 pD}{N}\right) + \frac{n_1(n_1-1)pD}{N^2(N-1)}\{N(D-1)-(N-1)Dp_1\}$$

In general, it is to be expected (and hoped) that $p_1 > p_1'$ just as $p > p'$, since we would expect (hope) that the probability of correct decision would exceed that of incorrect decision. It may well happen that $p_1 < p$ since detection of a defective may be more difficult with the mixture of material from separate units. More complicated distributions will be obtained if it is supposed that p_1 depends on the value of Y (the number of defective units). It does not seem unreasonable to suppose that p_1 might increase with Y.

The effectiveness of the screening procedure is measured by the three quantities

(i) Probability of correct classification for defectives = $p_1 p$ (20.1)

(ii) Probability of correct classification for nondefective =

$$P_0^*(n_1)(1-p_1'p') + (1-P_0^*(n_1))(1-p_1p') = 1-(p_1-P_{(1)}^*)p'$$

where

$$P_{(1)}^* = (P_1-p_1')P_0^*(n_1); \text{ and } P_0^*(n_1) = (N-D-1)^{(n_1-1)}/(N-1)^{(n_1-1)} \qquad (20.2)$$

is the probability that the sample contains no defectives, given that one member of the sample is nondefective.

(iii) Expected number of tests = $1 + n_1[P_0(n_1)p_1' + \{1-P_0(n_1)\}p_1]$ (20.3

$$= 1 + n_1(p_1 - P_{(1)})$$

of course, the larger the values of (i) and (ii), and the smaller the value of (iii), the better.

From (20.1) it is clear that the probability of correct classification of a defective is *decreased* by the screening process. (since $p_1p < p$). The value of screening must therefore come from increased correct classification of nondefectives or reduction in the expected number of tests. Table 1 contains some relevant numerical information. In the absence of screening, n_1 tests would be necessary, so

$$1-\{1+n_1)p_1-P_{(1)}\}/n, = 1-n_1^{-1} -p_1+ P_{(1)}$$

indicates the proportionate saving from screening, and this is given (as a percentage) in the last column of Table 1. We have taken $N = \infty$, so that $\omega = D/N$ is to be interpreted as proportion defective, because the values do not depend greatly on N. As N decreases, $P_{(1)}$ and $P^*_{(1)}$ decrease so that both the proportionate saving in number of tests from screening and the probability of correct identification of nondefectives decrease, (though not substantially, unless N is quite small). It is to be noted that screening *increases* the probability of correct identification of nondefectives.

4. HIERARCHAL GROUP SCREENING

Sometimes additional saving in the expected number of tests, and improved accuracy in classification, can be attained by using two-or more-stage screening - that is, hierarchal screening. For simplicity, we will consider two-stage procedures, with a first-stage sample of size $n_1 = hn_2$. (Generalization to more than two stages follows similar lines). If a positive results is obtained for the combined sample, it is split into h subsamples, each of size n_2 and each is then treated as in Section 3. Letting p_2, p_2' denote the respective probabilities of correct and incorrect positive results when testing each of the second stage (size n_2) subsamples, the three quantities measuring effectiveness are:

(i) Probability of correct classification for defectives = p_1p_2p (21.1)

(ii) Probability of correct classification for nondefectives

$$= P_0^*(n_1)(1-p_1'p_2'p) + (1-P_0^*(n_1))(1-p_1) + (P_0^*(n_2)-P_0^*(n_1))p_1(1-p_2p') + (1-P_0^*(n_2))p_1(1-p_2p')$$

$$= 1-p_1p_2p' + P_0^*(n_1)(p_1-p_1')p_2'p' + P_0^*(n_2)p_1(p_2-p_2')p'$$

$$= 1-p_1p_2p' + (P_{(1)}^*p_2' + P_{(2)}^*p_1)p' \qquad (21.2)$$

where $P_{(2)}^* = P_0^*(n_2)(p_2-p_2')$

(Note that $P_0^*(n_2) - P_0^*(n_1)$ is the probability that a random sample of n_1, known to contain at least one nondefective, also contains at least one defective, but a randomly chosen subsample of size n, containing at least one nondefective, in fact contains no defectives.)

(iii) Expected number of tests

$$= 1 + h(p_1-P_{(1)}) + n_1[P_0(n_1)p_1'p_2' + (1-P_0(n_2))p_1p_2 + (P_0(n_2)-P_0(n_1))p_1p_2']$$

$$= 1 + h(p_1-P_{(1)}) + n_1(p_1p_2 - P_{(1)}p_2' - P_{(2)}p_1) \qquad (21.3)$$

where $P_{(2)} = P_0(n_2)(p_2-p_2')$.

The proportional reduction in expected number of tests is

$$1 - p_1p_2 - n_2^{-1}(p_1-P_{(1)}) + P_{(1)}p_2' + P_{(2)}p_1 \qquad (21.4)$$

Values given by (21.2) and (21.4) are shown in Table 2.

From (21.1) we see that the probability of correct classification of a defective is decreased more than for simple screening (cf. (20.1)). As some compensation, the probability of correct classification of nondefectives is higher.

As in the case of simple screening the advantages of a screening procedure are greater when the proportion (ω) of defectives is smaller. The effect of finite lot size (N) is to decrease $P_{(1)}$, $P_{(2)}$, $P_{(1)}^*$, $P_{(2)}^*$. From equations (21) it can be seen that this will

(i) not affect the probability of correct classification of defectives

(ii) decrease the probability of correct classification of nondefectives

(iii) decrease the expected number of tests.

Some analysis of the distribution of the total number (Z) of items classified as "defectives" by this hierarchal screening procedure is given in the Appendix. Here we just give the expected value

$$E(Z) = n_1[p_1p_2(p-p')DN^{-1} + (p_1p_2 - P_{(2)}p_1 - P_{(1)}p_2')p'] \qquad (22)$$

ACKNOWLEDGEMENT

Samuel Kotz's work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0301. Norman L. Johnson's work was supported by the National Science Foundation under Grant MCS-8021704.

REFERENCES

Dorfman, R. (1943). The detection of defective members of large populations. *Ann. Math. Statist. 14*, 436-440.

Johnson, N.L. and Kotz, S. (1969). *Distributions in Statistics--Discrete Distributions.* New York: John Wiley and Sons.

Johnson, N.L. and Kotz, S. (1981). Faulty inspection distributions -- some generalizations. (To be published in *Proc. of ONR/ARO Reliability Workshop*, April 1981) *Institute of Statistics Mimeo Series #1335*, University of North Carolina at Chapel Hill.

Johnson, N.L., Kotz, S. and Sorkin, H.L. (1980). Faulty inspection distributions. *Commun. Statist. A9(9)*, 917-922.

Johnson, N.L. and Leone, F.C. (1977). *Statistics and Experimental Design: In Engineering and the Physical Sciences*, 2nd Edition. New York: John Wiley and Sons. (pp 914-915).

Table 1: Simple Screening

(i) Probability of correct classification of defective is $p_1 p$

(ii) These tables correspond to $N = \infty$. Values of $\omega(=D/N)$ are proportions of defectives in the lot. $P_o(n_1) = (1-\omega)^{n_1}$; $P_o^*(n_1) = (1-\omega)^{n_1-1}$. An N decreases, P_o and P_o^* decrease; other quantities are not changes.

(iii) Note the values shown do not depend on p.

p_1	p_1'	p	p'	ω	n_1	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVE	EXPECTED PERCENT REDUCTION IN TESTS
0.98	0.05	0.98	0.05	0.05	6	0.9870	53.7
					8	0.9835	51.2
					10	0.9803	47.7
					12	0.9774	43.9
				0.1	6	0.9785	34.8
					8	0.9732	29.5
					10	0.9690	24.4
					12	0.9656	20.9
				0.2	6	0.9662	9.7
					8	0.9608	5.1
					10	0.9572	2.0
					12	0.9550	0.0

Table 1: Simple Screening (cont'd)

p_1	p_1'	p	p'	ω	n_1	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVE	EXPECTED PERCENT REDUCTION IN TESTS
0.95	0.05	0.95	0.05	0.05	6	0.9873	54.5
					8	0.9839	52.2
					10	0.9804	48.9
					12	0.9781	45.3
				0.1	6	0.9791	36.2
					8	0.9740	31.2
					10	0.9699	26.4
					12	0.9666	22.1
				0.2	6	0.9672	11.9
					8	0.9619	7.6
					10	0.9589	4.7
					12	0.9564	2.9

Table 1: Simple Screening (cont'd)

p_1	p_1'	p	p'	ω	n_1	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVES	EXPECTED PERCENT REDUCTION IN TESTS %
0.90	0.05	0.95	0.05	0.05	6	0.9878	55.8
					8	0.9847	53.9
					10	0.9818	50.9
					12	0.9792	47.6
				0.1	6	0.9801	38.5
					8	0.9753	34.1
					10	0.9715	29.6
					12	0.9683	25.7
				0.2	6	0.9689	15.5
					8	0.9639	11.8
					10	0.9607	9.1
					12	0.9587	7.5

Table 1: Simple Screening (cont'd)

p_1	p_1'	p	p'	ω	n_1	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVE	EXPECTED PERCENT REDUCTION IN TESTS
0.95	0.10	0.90	0.10	0.05	6	0.9718	60.8
					8	0.9644	58.9
					10	0.9586	55.9
					12	0.9533	52.6
				0.1	6	0.9552	43.5
					8	0.9457	39.1
					10	0.9379	34.6
					12	0.9317	30.7
				0.2	6	0.9329	20.5
					8	0.9228	16.8
					10	0.9164	14.1
					12	0.9123	12.5

Note that in the last set, the expected percent reduction in expected number of tests is always 5% greater than for the corresponding case in the penultimate set. This is because the value of $p_1 - p_1'$ is the same (0.85) in the two sets while the value of p_1 is 0.05 greater in the last set.

Table 2: Two-Stage Hierarchal Screening

p_1	p_1'	p_2	p_2'	p'	ω	n_1	n_2	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVES	EXPECTED PERCENT REDUCTION IN TESTS %
0.98	0.05	0.98	0.05	0.05	0.05	6	2	0.9971	65.2
						6	3	0.9949	63.7
						12	2	0.9966	56.5
						12	3	0.9944	60.3
						12	4	0.9924	60.4
						12	6	0.9886	57.1
					0.1	6	2	0.9943	39.3
						6	3	0.9902	40.0
						12	2	0.9937	34.9
						12	3	0.9896	39.5
						12	4	0.9859	38.8
						12	6	0.9796	33.4
					0.2	6	2	0.9892	10.1
						6	3	0.9819	10.6
						12	2	0.9886	8.5
						12	3	0.9813	12.1
						12	4	0.9755	10.4
						12	6	0.9671	4.6

Table 2: Two-Stage Hierarchal Screening (cont'd)

p_1	p_1'	p_2	p_2'	p'	ω	n_1	n_2	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVES	EXPECTED PERCENT REDUCTION IN TESTS(%)
0.98	0.05	0.90	0.05	0.05	0.05	6	2	0.9973	58.9
						6	3	0.9959	60.1
						12	2	0.9968	57.3
						12	3	0.9948	61.5
					0.1	6	2	0.9948	40.8
						6	3	0.9910	42.1
						12	2	0.9941	36.4
						12	3	0.9904	41.6
					0.2	6	2	0.9900	12.9
						6	3	0.9833	14.5
						12	2	0.9894	11.3
						12	3	0.9828	15.9

Table 2: Two-Stage Hierarchal Screening (cont'd)

p_1	p_1'	p_2	p_2'	p'	ω	n_1	n_2	PROBABILITY OF CORRECT CLASSIFICATION FOR NONDEFECTIVES	EXPECTED PERCENT REDUCTION IN TESTS(%)
0.95	0.05	0.95	0.05	0.05	0.05	6	2	0.9972	59.2
						6	3	0.9954	60.2
						12	2	0.9968	57.8
						12	3	0.9947	61.7
					0.1	6	2	0.9947	40.9
						6	3	0.9908	42.0
						12	2	0.9941	37.2
						12	3	0.9902	41.8
					0.2	6	2	0.9898	13.4
						6	3	0.9830	14.3
						12	2	0.9893	11.9
						12	3	0.9825	15.9

Table 2: Two-Stage Hierarchal Screening (cont'd)

p_1	p_1'	p_2	p_2'	p'	ω	n_1	n_2	PROBABILITY OF CORRECT CLASSIFICATION OF NONDEFECTIVES	EXPECTED PERCENT REDUCTION IN TESTS(%)
0.95	0.10	0.95	0.05	0.05	0.25	6	2	0.9971	57.2
						6	3	0.9953	58.8
						12	2	0.9967	56.3
						12	3	0.9946	60.7
					0.1	6	2	0.9946	39.4
						6	3	0.9907	41.0
						12	2	0.9941	36.4
						12	3	0.9902	41.3
					0.2	6	2	0.9898	12.7
						6	3	0.9830	13.8
						12	2	0.9893	11.7
						12	3	0.9825	15.8

Note that in the last set the probabilities of correct classification of nondefectives are very slightly greater (no more than 0.0001) that the corresponding probabilities for the penultimate set, while the expected reduction in number of tests is slightly (c. 1-2%) less. (This comparison reflects the effect of changing p_1' from 0.05 to 0.10).

APPENDIX

In order to study the distribution of the total number (Z) of items declared "defective" we introduce the auxiliary indicator variables:

$$W_j = \begin{cases} 1 & \text{if items in the j-th subsample are tested individually} \\ 0 & \text{otherwise} \end{cases} \tag{A1}$$

for j = 1, ..., h. Conditional on $W_1, \ldots, W_h$ and the *actual* numbers $Y_1, \ldots, Y_h$ of defectives in the corresponding subsamples, Z is distributed as

$$\text{Binomial } \left(\sum_{j=1}^{h} W_j Y_j, p\right) * \text{Binomial } \left(\sum_{j=1}^{h} W_j (n_2 - Y_j), p'\right) \tag{A2}$$

The conditional r-th factorial moment of Z is

$$E(Z^{(r)} | \underset{\sim}{W}, \underset{\sim}{Y}) = \sum_{i=0}^{r} \binom{r}{i} \left(\sum_{j=1}^{h} W_j Y_j\right)^{(i)} \left\{\sum_{j=1}^{h} W_j (n_2 - Y_j)\right\}^{(r-i)} p^i p'^{r-i} \tag{A3}$$

Conditional on $\underset{\sim}{Y}$ for any α, $\alpha' > 0$; and $j \neq j'$

$$E(W_j^{\alpha} | \underset{\sim}{Y}) = P(W_j = 1 | \underset{\sim}{Y}) = \begin{cases} p_1 p_2 & \text{if } Y_j > 0 \\ p_1 p_2' & \text{if } Y_j = 0, \ \sum_1^h Y_i > 0 \\ p_1' p_2' & \text{if } Y_j = 0, \ \sum_1^h Y_i \end{cases} \tag{A4}$$

$$E(W_j^{\alpha} W_{j'}^{\alpha'} | \underset{\sim}{Y}) = P(W_j = 1, W_{j'} = 1 | \underset{\sim}{Y}) = \begin{cases} p_1 p_2^2 & \text{if } Y_j > 0, \ Y_{j'} > 0 \\ p_1 p_2 p_2' & \text{if } Y_j > 0, \ Y_{j'} = 0 \text{ or } Y_j = 0, \ Y_{j'} > 0 \\ p_1 p_2'^2 & \text{if } Y_j = Y_{j'} = 0, \ \sum_1^h Y_i > 0 \\ p_1' p_2'^2 & \text{if } Y_j = Y_{j'} = 0 = \sum_1^h Y_i \end{cases} \tag{A5}$$

We have

$$P(Y_j = 0) = P_o(n_2); \; P(Y_j>0) = 1-P_0(n_2) \tag{A6.1}$$

$$P(Y_j = 0, \; Y_j' = 0) = P_0(2n_2) \tag{A6.2}$$

$$P(Y_j = 0, \; Y_j' > 0) = P(Y_j' > 0, \; Y_j = 0) = P_0(n_2)-P_0(2n_2) \tag{A6.3}$$

$$P(Y_j > 0, \; Y_j' > 0) = 1-2P_0(n_2) + P_0(2n_2) \tag{A6.4}$$

$$P(Y_j = Y_j' = 0, \; \sum_{i=1}^{h} Y_i > 0) = P_0(2n_2) - P_0(n_1) \tag{A6.5}$$

$$P(Y_j = 0 = \sum_{i=1}^{h} Y_i) = P(Y_j = Y_j' = 0 = \sum_{i=1}^{h} Y_i) = P_0(n_1) \tag{A6.6}$$

whence

$$E(W_j^{\alpha}) = p_1p_2 - P_{(2)}p_1 - P_{(1)}p_2' \tag{A7.1}$$

$$E(W_j^{\alpha} \; W_{j'}^{\alpha'}) = p_1p_2^2 - 2p_1p_2P_{(2)} + p_1(p_2-p_2')^2P_0(2n_2) - p_2'^2P_{(1)} \tag{A7.2}$$

Also, from (A4) and (A5), with $\beta, \; \beta' > 0$

$$E(W_j^{\alpha}Y_j^{\beta}) = p_1p_2E(Y_j^{\beta}) \tag{A8.1}$$

$$E(W_j^{\alpha}W_{j'}^{\alpha'}Y_j^{\beta}Y_j^{\beta'}) = p_1p_2^2E(Y_j^{\beta}Y_j^{\beta'}) \tag{A8.2}$$

$$E(W_j^{\alpha}W_{j'}^{\alpha'}Y_j^{\beta}) = p_1p_2^2\{1-P_0(n_2)|- E(Y_j^{\beta}|Y_{j'} > 0) + p_1p_2p_2'P_0(n_2)E(Y_j^{\beta}|Y_{j'}=0) \tag{A8.3}$$

The joint distribution function of $\underset{\sim}{Y}$ is

$P(\underset{\sim}{Y}) =$ (A9)

$$\binom{D}{Y_1, \ldots, Y_h, D - \sum_1^h Y_i} \binom{N-D}{n_2-Y_1, \ldots, n_2-Y_h, N-D-n_1+\sum_1^h Y_i} \Bigg/ \binom{N}{n_2, \ldots, n_2, N-n_1}$$

and the $\underset{\sim}{r}$-th joint factorial moment of $\underset{\sim}{Y}$ is

$$E\left(\prod_{j=1}^{h} Y_j^{(r_j)}\right) = \left(\prod_{j=1}^{h} n_2^{(r_j)}\right) D^{(r)}/N^{(r)} \tag{A10}$$

where $r = \sum_{i=1}^{h} r_i$. In particular for all $j \neq j'$

$$E(Y_j) = n_2 D/N \tag{A11.1}$$

$$E(Y_j^2) = n_2 D\{(n_2-1)D+N-n_2\}/N^{(2)} \tag{A11.2}$$

$$E(Y_j Y_j') = n_2^2 D^{(2)}/N^{(2)} \tag{A11.3}$$

$$E(Y_j | Y_j' = 0) = n_2 D/(N-n_2); \; E(Y_j | Y_j' > 0) = n_2 D\{1-n_2 N^{-1} - P_0(n_2)\}\{1-P_0(n_2)\}^{-1} \times (Nn_2)^{-1} \tag{A11.4}$$

Applying formulae (A3) - (A11) we get after some algebraic rearrangement the formula (22) for E(Z). In order to calculate the variance of Z, we have to calculate

$$E(Z^{(2)}) = E[p^2 (\sum_{j=1}^{h} W_j Y_j)(\sum_{j=1}^{h} W_j Y_j - 1) + 2pp' (\sum_{j=1}^{h} W_j Y_j)\{\sum_{j=1}^{h} W_j (n_2 - Y_j)\} + p'^2 \{\sum_{j=1}^{h} W_j (n_2-Y_j)\}\{\sum_{j=1}^{h} W_j (n_2-Y_j) - 1\}]$$

$$= (p-p')^2 E[(\sum_{j=1}^{h} W_jY_j)^2] - (p^2-p'^2)E(\sum_{j=1}^{h} W_jY_j)$$

$$+ 2n_2p'(p-p')E[(\sum_{j=1}^{h} W_jY_j)(\sum_{j=1}^{h} W_j)] - n_2p'^2 E(\sum W_j)$$

$$+ n_2^2p'^2 E[(\sum W_j)^2]$$

$$= (p-p')^2[hE(W_jY_j^2) + h(h-1)E(W_jW_j'Y_jY_j')] - p^2-p'^2)hE(W_jY_j)$$

$$+ 2n_2p'(p-p')[hE(W_jY_j) + h(h-1)E(W_jW_j'Y_j)] - n_2p'^2 hE(W_j)$$

$$+ n_2^2p'^2 [hE(W_j) + h(h-1)E(W_jW_j')] \quad \text{(A12)}$$

where $E(W_j)$ and $E(W_jW_j')$ are given by (A7), and

$$E(W_jY_j) = n_2p_1p_2DN^{-1} \quad \text{(A13.1)}$$

$$E(W_jY_j^2) = n_2p_1p_2D\{(n_2-1)D+N-n_2\}N^{-1}(N-1)^{-1} \quad \text{(A13.2)}$$

$$E(W_jW_j'Y_j) = n_2p_1p_2D\{N^{-1}p_2-(N-n_2)^{-1}P_{(2)}\} \quad \text{(A13.3)}$$

The resulting formulas for $E(Z^{(2)})$ and

$$var(Z) = E(Z^{(2)}) + E(Z) - \{E(Z)\}^2$$

are complicated, but numerical calculation is straightforward. If $p' = p_1' = p_2' = 0$ (so that there are no false positives), then

$$E(Z^{(2)}) = p^2[hE(W_jY_j^2) + h(h-1)E(W_jW_j'Y_jY_j') - hE(W_jY_j)] =$$

$$\frac{n_1p_1p_2p^2D}{N}\left[\frac{(n_2-1)(D-1)}{N-1} + (h-1)\{p_2-(1-n_2N^{-1})P_{(2)}\}\right]$$

and also

$$E(Z) = \frac{n_1 p_1 p_2 p^D}{N} \tag{A14}$$

so that

$$\mathrm{Var}(Z) = \frac{n_1 p_1 p_2 p^D}{N} \left| \left\{ \frac{(n_2-1)(D-1)}{N-1} + (h-1)[p_2-(1-n_2N^{-1})P_{(2)}]\right\}p \right.$$

$$\left. + 1 - \frac{n_1 p_1 p_2 p^D}{N} \right] \tag{A15}$$