Sequential Defect Removal Sampling

Douglas G. Bonett • J. Arthur Woodward
College of Business, University of Wyoming, Department of Management and Marketing, P.O. Box 3275, Room 228, Laramie, Wyoming 82071-3275
University of California, Los Angeles, California 90024

Standard inspection methods underestimate the true number of defects or nonconformities in a complex product (e.g., automobile, mobile home, airplane, circuit board, computer program) when an inspector is unable to identify every defect with certainty. A nonlinear statistical model with a nonlinear constraint is developed for estimating the unknown number of defects in a product when inspection is imperfect. A sequential defect removal sampling plan is defined in which two or more inspectors examine in sequence a product or sample of products and then mark or correct any observed defects prior to the next inspection. The number of defects identified by each inspector provides the information needed to estimate the number of defects in the product in addition to the number of defects that have eluded all inspectors. A goodness-of-fit test of model assumptions is presented. A test of hypothesis regarding the unknown number of defects in quality improvement experiments also is described.
(*Imperfect Inspection; Nonlinear Statistical Model; Quality Management*)

1. Introduction

For some complex products, inspection will be imperfect and every defect or nonconformity will not be detected. When inspection is imperfect, the quality of the product will be overestimated. Various aspects of the problem of imperfect inspection have been extensively discussed (see Bonett 1988).

A sequential defect removal sampling plan is proposed here in which two or more inspections are performed and in each inspection the number of defects is recorded and then marked or corrected prior to the next inspection. Independence of inspections is not assumed. A nonlinear statistical model of the number of identified defects is developed. The parameters of the model correspond to the unknown total number of defects in the product prior to inspection and the unknown probability of detecting a defect during an inspection. The method of minimum chi-square is used to obtain efficient estimators of the model parameters. A Pearson statistic for evaluating model goodness of fit is presented. A Wald statistic is defined for testing hypotheses in quality improvement experiments.

2. Statistical Model

A product is inspected on $t > 1$ occasions. On each occasion, f_i defects are observed and corrected. Let $\mathbf{F}$ denote a $t \times 1$ observable random vector of frequency counts with typical element F_i. Note that f_i is a sample realization of F_i. The expected value of F_i is denoted as μ_i. It is assumed that the elements in $\mathbf{F}$ are multinomial random variables with expected values $\mu_i = N\pi(1 - \pi)^{i-1}$ where N is the unknown number of defects in a product prior to inspection and π is the unknown probability of detecting a defect common to all inspectors. Since the unknown parameters N and π are multiplicative functions of μ_i, a log-linear model of $\mathbf{F}$ may be defined as

$$\mathbf{F} = \exp(\mathbf{X}\beta) + \epsilon \tag{1}$$

where $\mathbf{F}' = [F_1 F_2 \cdots F_t]$, β is a 3×1 vector of unknown parameters with $\beta_1 = \ln(N)$, $\beta_2 = \ln(\pi)$, $\beta_3 = \ln(1 - \pi)$, and $\mathbf{X}$ is a $t \times 3$ matrix of known constants with rows $[1\ 1\ 0]$ to $[1\ 1\ t-1]$. Note that β_2 and β_3 are functionally related with $\exp(\beta_3) = 1 - \exp(\beta_2)$.

0025-1909/94/4007/0898$01.25

3. Parameter Estimation

The method of minimum modified transformed chi-square (Ferguson 1958) is applied to obtain an estimator of β. The minimum chi-square estimator may be obtained from the scoring iteration

$$\beta_{j+1} = \beta_j + \mathbf{P}_j\mathbf{g}_j \quad (2)$$

where

$$\mathbf{P}_j = (\mathbf{X}'\mathbf{D}_1\mathbf{X} + \mathbf{D}_2\mathbf{v}\mathbf{v}'\mathbf{D}_2)^{-1}, \quad (3)$$

$$\mathbf{g}_j = (\mathbf{X}'\mathbf{D}_1(\ln(\mathbf{f}) - \mathbf{X}\beta_j) + \mathbf{D}_2\mathbf{v}(1 - \mathbf{v}'\mathbf{\Theta}_j), \quad (4)$$

$\mathbf{v}' = [0\ 1\ 1]$, $\mathbf{D}_1$ is a diagonal matrix with the elements of the vector $\mathbf{f}$ along the principal diagonal, $\mathbf{\Theta} = \exp(\beta)$, $\mathbf{f}$ is a $t \times 1$ sample realization of the random vector $\mathbf{F}$ with typical element f_i, and $\mathbf{D}_2$ is a diagonal matrix with the elements of the vector $\mathbf{\Theta}_j = \exp(\beta_j)$ along the principal diagonal. Sampling zeros in $\mathbf{f}$ are replaced with small non-zero values such as $1/t$ (Grizzle, Starmer, and Koch 1969). Iteration of (2) terminates when the value of $(\beta_{j+1} - \beta_j)'(\beta_{j+1} - \beta_j)$ is sufficiently small. Convergence is extremely fast. Let $n = \mathbf{1}'\mathbf{f}$. A good starting vector in (2) is the 3×1 vector $\beta_1 = [(nf_1/n) \times (1 - f_1/n)]$.

The final iteration of (2) gives the minimum modified transformed chi-square estimate of β subject to the nonlinear constraint $\exp(\beta_1) = 1 - \exp(\beta_2)$. This estimate will be denoted as $\bar{\beta}$. To be precise, the estimate $\bar{\beta}$ is a sample realization of the estimator $\hat{\beta}$.

A consistent estimate of the covariance matrix of $\hat{\beta}$ is defined as

$$\bar{\mathbf{\Sigma}}_{\hat{\beta}} = \mathbf{P}\mathbf{X}'\bar{\mathbf{\Sigma}}_{\mathbf{F}}\mathbf{X}\mathbf{P} \quad (5)$$

where $\mathbf{P}$ is the final iteration of (3), $\bar{\mathbf{\Sigma}}_{\mathbf{F}} = \mathbf{D}_1 - \mathbf{f}\mathbf{f}'/\bar{N}$, and $\bar{N}$ is defined in (6). The square root of the i^{th} diagonal element of $\bar{\mathbf{\Sigma}}_{\hat{\beta}}$ is denoted as $\overline{se}(\hat{\beta}_i)$ and represents the standard error of the estimator of β_i.

The estimates of N and π are defined as

$$\bar{N} = \exp(\bar{\beta}_1), \quad (6)$$

$$\bar{\pi} = \exp(\bar{\beta}_2), \quad (7)$$

and the estimate of the $t \times 1$ vector μ is defined as

$$\bar{\mu} = \exp(\mathbf{X}\bar{\beta}). \quad (8)$$

Standard errors of the estimators are obtained by application of the delta method (see Bishop, Fienberg, and Holland 1975, chapter 14). The asymptotic standard error of the estimator of N is estimated as

$$\overline{se}(\hat{N}) = \overline{se}(\hat{\beta}_1)\bar{N} \quad (9)$$

and the asymptotic standard error of the estimator of π is estimated as

$$\overline{se}(\hat{\pi}) = \overline{se}(\hat{\beta}_2)\bar{\pi}. \quad (10)$$

In the special case of $t = 2$ inspections, closed form estimates of N and π can be derived. Assuming $f_1 > f_2$, these estimates are defined as

$$\bar{N} = f_1/(f_1 - f_2) \quad (11)$$

$$\bar{\pi} = 1 - f_2/f_1 \quad (12)$$

Using a Poisson approximation to the multinomial and assuming $f_1 > f_2$, the estimated asymptotic standard errors of $\hat{N}$ and $\hat{\pi}$ in the special case of $t = 2$ are defined as

$$\overline{se}(\hat{N}) = [f_1^3(f_1 - 2f_2)^2 + f_1^4 f_2)/(f_1 - f_2)^4]^{1/2} \quad (13)$$

$$\overline{se}(\hat{\pi}) = [(f_2 + f_2^2/f_1)/f_1]^{1/2} \quad (14)$$

In multiple group designs, N and π are estimated in group k and are denoted as $\bar{N}_k$ and $\bar{\pi}_k$. Similarly, (9) and (10) are computed in group k and are denoted as $\overline{se}(\hat{N}_k)$ and $\overline{se}(\hat{\pi}_k)$.

4. Hypothesis Testing

Pearson and Wald tests are presented in this section. The Pearson test is useful for assessing the appropriateness of the model and the Wald test is useful in quality improvement studies.

The Pearson goodness-of-fit test statistic is defined as

$$c = \Sigma(f_i - \bar{\mu}_i)^2/\bar{\mu}_i \quad (15)$$

where summation is over $i = 1, 2, \cdots t$ inspections and is evaluated with $t - 2$ degrees of freedom. If c exceeds the critical chi-square value, then we reject the null hypothesis that $\mu_i = N\pi(1 - \pi)^{i-1}$.

In quality improvement experiments with $r > 1$ treatments, $\bar{N}_k$ and $\overline{se}(\hat{N}_k)$ are computed for each treatment. Let $\bar{\mathbf{N}}$ denote an $r \times 1$ vector with typical element $\bar{N}_k$ and let $\bar{\Sigma}_{\mathrm{N}}$ denote a diagonal matrix with $\overline{se}(\hat{N}_k)^2$ in the k^{th} row. A Wald test of general linear null hypothesis H_0: $\mathbf{HN} = \mathbf{h}$ is based on a chi-square statistic

$$w = (\mathbf{H}\bar{\mathbf{N}} - \mathbf{h})'(\mathbf{H}\bar{\Sigma}_{\bar{N}}\mathbf{H}')^{-1}(\mathbf{H}\bar{\mathbf{N}} - \mathbf{h}), \quad (16)$$

that is evaluated with s degrees of freedom where $\mathbf{H}$ is an $s \times r$ matrix of known constants, $\mathbf{h}$ is an $s \times 1$ vector of known constants.

5. Model Assumptions

The log-linear model defined in §2 is based on three assumptions: 1) common π across all inspectors, 2) constant N across inspections, and 3) common π across all defects (homogeneity). The Pearson test presented in the previous section provides a test of the first two assumptions in a $t > 2$ sampling plan. Stratification is used to deal with a violation of the third assumption.

In any statistical data analysis application, it is important to know the assumptions underlying the analysis and methods for testing the assumptions. Also, it is important to know the effects of violating the assumptions, since the effects may or may not be trivial.

Assumption of Common π Across Inspections. The assumption of a common π for all t inspections may be violated, if for instance, the first inspection is performed by an inexperienced employee and all other inspections are performed by highly trained employees. Then π may differ across inspections. In Table 1, N is estimated from a three-sample inspection plan where π differs across the three inspections and the true value of N equals 100. We note that in the three-sample plan, the estimates of N are positively biased and range from 100.2 to 111.4 depending on the magnitude of inequalities of π across inspections. One may conclude that for moderate violations of the first assumption, for instance, $\pi(\max) - \pi(\min) < .3$, the estimate of N is not severely biased. Table 1 also gives the Pearson noncentrality parameter values and power for $\alpha = .05$. One may observe that for a moderate violation of the first assumption, the Pearson test appears to have good power for $N = 100$. With smaller N or π, the observed frequency counts will be smaller and the Pearson test will be less powerful. We conclude that the log-linear model is fairly robust to a violation of the first assumption and that the Pearson test is able to detect a moderate violation of this assumption in a sample of moderate size.

Table 1 Effect of Violating Assumption of Common π Across Inspections

π_1	π_2	π_3	N	$\bar{N}$	Noncentrality Parameter	Power for $\alpha = 0.05$
0.8	0.8	0.8	100	100.0	0	0.050
0.8	0.7	0.8	100	100.2	0.89	0.157
0.8	0.6	0.8	100	101.4	3.25	0.438
0.8	0.5	0.8	100	103.6	6.63	0.731
0.8	0.4	0.8	100	106.8	10.75	0.906
0.8	0.3	0.8	100	111.4	15.48	0.997

Note. π_i is the probability of inspector i detecting a defect.

Table 2 Effect of Violating Constant N Assumption

N_1	N_2	N_3	$\bar{N}$	Noncentrality Parameter	Power for $\alpha = 0.05$
100	100	100	100.0	0	0.050
100	110	110	111.2	0.43	0.101
100	120	120	123.9	1.65	0.250
100	130	130	138.3	3.58	0.473
100	140	140	154.6	6.10	0.695
100	100	110	113.6	4.52	0.566
100	100	120	132.5	10.57	0.902
100	100	130	157.2	16.19	0.984
100	100	140	190.0	21.33	0.996
100	110	120	123.8	1.29	0.206
100	120	140	155.1	1.72	0.258
100	130	160	198.6	1.54	0.237
100	140	180	264.1	1.13	0.186

Note. $\pi = 0.8$ in all conditions; N_i is the number of defects in the product during inspection i.

Assumption of Constant N Across Inspections. The second assumption states that N must be constant across inspections. In applications where a finished good is inspected sequentially by two or more inspectors at a common location, this assumption should be easily satisfied. However, if inspection occurs at different locations (e.g., factory, warehouse, dealer) defects may be introduced during shipment and N may be larger at inspection $i + 1$ than at inspection i. Table 2 presents estimates of N in a three-sample inspection for three classes of assumption violations. In the first case, N increases only after the first inspection. In the second case, N increases only after the second inspection. In the third case, N increases after both the first and second inspections.

From the results in Table 2, it appears that the estimate of N is actually estimating the final number of defects in a product (rather than the number of defects prior to inspection) when the constant N assumption has been violated. It is reassuring to observe that the estimate of N (under this interpretation) has a moderate positive bias if N has been increased only after the first inspection. However, the bias is more severe if N is increased after the second inspection. Also, we note that the Pearson test appears to have adequate power under a moderate violation of the constant N assumption.

For the case where N increases after both the first and second inspection, the estimate of N has a much greater bias and the Pearson test is relatively insensitive to a violation of the constant N assumption. The case where N increases after two or more inspections could occur in applications where a product is inspected after each of two or more shipments. However, by not counting shipment related defects (specific type of scratches and dents, breakage due to vibration, etc.) at each inspection, the sequential defect removal sampling plan may still be effectively applied to estimate the number of defects prior to inspection.

Assumption of Equal π Across Defects. The third assumption, which we will also refer to as the homogeneity assumption, states that each defect must have the same probability of detection. As in other statistical sampling plans, one solution to a violation of this assumption involves stratification (Cochran 1977). Homogenous classes of defect types are identified so that the probability of defect detection within each class is approximately equal for each defect.

Table 3 presents estimates of N from a three-sample inspection plan where defects have one of two different detection probabilities and the true number of defects is 100. In the second row of the table, one type of defect has a detection probability of .3 while the other has a detection probability of .9. The estimate of N is 86.5 and is negatively biased by almost 14%. Rows 3 through 5 in the table show the amount of bias when the detection probabilities differ by .2 at three different levels (.3 and .5, .5 and .7, .7 and .9). The negative bias for theses cases ranges from 5.9% to 1.5%.

The results of Table 3 also illustrate the usefulness of stratification. Suppose that detection probabilities ranged from .3 to .9. Defect types may be classified into three distinct categories (A, B, and C) based on previous inspection information where type A defects have typical detection probabilities in the .3 to .5 range, type B in the .5 to .7 range, and type C in the .7 to .9 range. Then, the estimated number of defects within each stratum will have a much smaller negative bias. The bias can be reduced further by using four or five strata.

Table 3 Effect of Violating Assumption of Common π Across Defects

π_1	π_2	$\bar{N}$
0.6	0.6	100.0
0.3	0.9	86.5
0.3	0.5	94.1
0.5	0.7	97.3
0.7	0.9	98.5

Note. π_i is the probability of detecting defect type i; $N = 100$ in all conditions.

6. Examples

Consider the problem of inspecting a mobile home in the final assembly bay, by the dealer upon delivery, and by the buyer at the time of purchase. Assume N and π are constant across inspections.

Suppose that 25, 15, and 8 defects are detected and corrected in the first, second and third inspections respectively so that $\mathbf{f}' = [25\ 15\ 8]$. The estimate of N is $\exp(\bar{\beta}_1) = 59.1$ (or 59) and an estimate of π is $\exp(\bar{\beta}_2) = .428$ with estimated standard errors of 8.70 and .110 respectively. We estimate that the mobile home has $59 - (25 + 15 + 8) = 11$ defects that have not been detected. The Pearson statistic (15) equals .032 suggesting that the model assumptions have been satisfied.

A second example, suggested by a referee, involves a circuit board test in which the board is tested until two successive tests yield zero defects. When two successive zeros are observed, the manufacturer then concludes that all defects have been identified. Suppose $\mathbf{f}' = [12\ 6\ 0\ 0]$. Our method predicts that there are two undetected defects in the board and the manufacturer should consider making an additional inspection.

7. Summary

Imperfect inspection is a serious problem. When inspection is imperfect, standard quality assessment analyses yield negatively biased estimates of the total number of defects per product. As a result, the producer will overestimate the quality of the product.

A byproduct of sequential defect removal sampling is that a greater number of defects will be removed from a product compared to single inspection plans. Although sequential defect removal sampling is particularly well suited to applications where a product is inspected more than once, some manufacturers would benefit by requiring additional inspections of their finished goods. In doing so, they would obtain more accurate quality estimates while improving outgoing quality.

References

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis*, MIT Press, Cambridge, MA, 1975.

Bonett, D. G., "Estimating the Number of Defects Under Imperfect Inspection," *J. Applied Statistics*, 15 (1988), 63–67.

Cochran, W. G., *Sampling Techniques*, John Wiley & Sons, New York, 1977.

Ferguson, T. S., "A Method of Generating Best Asymptotically Normal Estimates with an Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 29 (1958), 1046–1062.

Grizzle, J. E., C. F. Starmer and G. G. Koch, "Analysis of Categorical Data by Linear Models," *Biometrics*, 28 (1969), 137–156.

Accepted by L. Joseph Thomas, former Departmental Editor; received January 15, 1990. *This paper has been with the authors* 14 *months for* 1 *revision.*