A Continuous Sampling Plan for Monitoring Survey Form Quality

Paul Novotny, Bureau of Labor Statistics

A multi-level continuous sampling plan was designed to monitor the quality of survey forms submitted to the Bureau of Labor Statistics International Price Program (IPP). The quality was measured in terms of the form's adherence to the standards in a collection manual. The possible errors ranged from miscoding the date to interviewing the wrong company. Markov chain theory was used to look at the long run implications of different review policies. The design was field tested and results were presented analyzing the error rates to determine practical levels of average outgoing quality and a mechanism for changing audit rates for survey forms.

Key Words : Quality Control, Markov Chains, Rectifying Inspection, Average Outgoing Quality Limit

Introduction

The IPP produces price indexes for U.S. imports and exports. Each year, a sample of importers and exporters is chosen to report product prices. These sampled companies are contacted by IPP field workers who collect information on the company and the products which will be reported. These products are determined in the field by the company and the field workers using a statistical selection procedure. All of the collected information and the documentation for product selection are reported on forms to the national office in Washington, D.C. The specifications for filling out the forms and for product selection are documented in a collection manual.

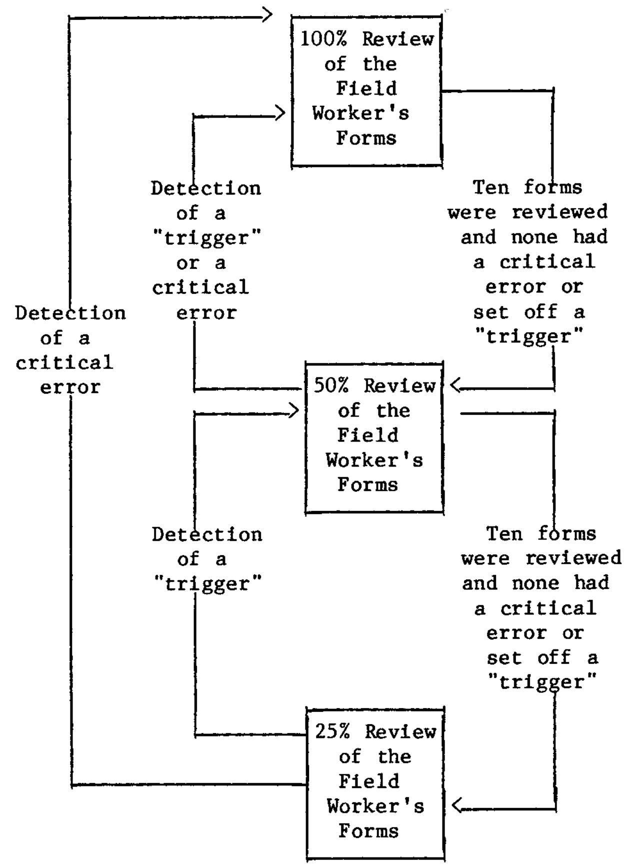
The IPP is developing a system for monitoring and improving the quality of these survey forms. The quality of the forms is measured by their adherence to the standards descibed in the collection manual. From the standards, 170 different potential errors were identified. These errors ranged from miscoding the date to interviewing the wrong company. This list of errors includes five errors that were considered critical errors. These five errors were labelled as critical since any of these errors would make the data unusable to the IPP.

The purpose of this quality project is to provide feedback to the field workers concerning any recurring form problems, to identify areas in the collection manual or forms which are causing undue problems so that these problems can be corrected, and to maintain a good level of usable data while keeping the effort of reviewing forms at a minimum.

The Sampling Design

The sampling plan, for reviewing the forms submitted by each field worker, consists of three levels of review. These levels are a review of 100% of a field worker's forms, a 50% review and a 25% review. The level of review is determined by the level of quality observed on the most recently reviewed forms. The level of review increases, or is "triggered", whenever more than a fixed number of errors are found on a form. The detection of a critical error automatically increases the review rate to 100%. If ten consecutive forms are reviewed and none of the ten forms contains a critical error or sets off a "trigger", the level of review decreases by one level. A diagram of the sampling plan is shown below.

Sampling Plan
for each Field Worker



All field workers start at the 100% review level. The less error prone workers will then trickle down into the lower levels of review.

The sampling plan will involve a rectifying inspection. This means that if an error is detected in the review of a form, that error will be corrected. Thus, at 100% review, no errors should theoretically be missed. At 50% review about half of the errors will be missed, and at 25% review three-fourths of the errors will be missed.

Development of the Sampling Plan

The development of the sampling plan involved analyzing a wide range of possible interviewer error rates, determining the asymptotic probabilities of the field worker being at each level of review, calculating the average amount of inspection and the average outgoing quality for the field worker, determining the sensitivity to critical errors, and selecting a "trigger" mechanism which results in practical levels of review and quality.

The following specifications were used in the analysis:

Let T be the "trigger" value used in the sampling design, and R be the average number of errors made per form by a field worker. Let

P1 = Pr(no critical errors and not over T errors on any of the next ten forms)

P2 = Pr(at least one critical error in the next ten forms)

P3 = Pr(no critical errors but at least one of the next ten forms has over T errors).

The transition probabilities for the sampling plan are

	A	B	C
A	1-P1	P1	0
B	1-P1	0	P1
C	P2	P3	P1

which is represented by the matrix

$$P = \begin{pmatrix} 1-P1 & P1 & 0 \\ 1-P1 & 0 & P1 \\ P2 & P3 & P1 \end{pmatrix}.$$

Here A corresponds to the state of 100% review, B corresponds to the state of 50% review, C corresponds to the state of 25% review, and the ijth element of the transition matrix is the probability that an interviewer at review level i will change to review level j within the next ten reviewed forms.

The sampling design forms a Markov chain, which means that the probability of changing to any given level of review is dependent only on the current level of review. The transition probabilities do not depend on any other previous levels of review. Therefore, a convergence theorem of Markov chains can be applied to obtain the asymptotic probability of being at each review level.

In this theorem, P^k is the k step transition matrix. The ij^{th} element of this matrix is the probability that an interviewer starting at review level i will be in review level j after k realizations of the sampling design. The theorem is (see Cinlar, p. 378):

Theorem: Let P be an irreducible aperiodic Markov matrix. Then for all i,j,

$$\lim_{k\to\infty} P^k(i,j) = \pi(j) > 0$$

independent of i. The row vector π is the unique solution of

$$\pi \cdot P = \pi , \quad \pi \cdot 1 = 1.$$

Moreover, the convergence is geometric: there exist constants $\alpha > 0$ and $0 \leq \beta < 1$ such that $\left|P^k(i,j) - \pi(j)\right| \leq \alpha\beta^k$, $k = 1, 2, \ldots$ for all i,j.

The first graph gives an indication of how far off the Markov approximations will be to the actual probabilities after a given number of steps through the sampling plan. A step through the sampling plan is the process of reviewing ten forms and adjusting the review level if necessary. The actual probabilities were calculated for each possible error rate using the nth step transition matrix, P^n. The maximum difference between the asymptotic probabilities and P^n was plotted for the various error rates.

This graph shows that even after only two steps through the sampling plan, the actual and asymptotic probabilities differ by at most 0.2. After only four steps, the probabilities differ by at most 0.05. These are again the worst cases, and the differences for most error rates are much smaller.

Since all interviewers will start at 100% review, the starting quality and percent of review will be higher than the estimates based on the asymptotic probabilities. These levels of review and quality will then converge down to the asymptotic levels. The asymptotic probabilities do not depend on the initial level of review. Therefore, if different starting levels are used, the approximations will still hold.

The asymptotic probabilities for this sampling plan are:

$$\pi_A = 1 - \frac{P1}{1-P1P3}$$

$$\pi_B = \frac{P1 \cdot (1 - P1)}{1 - P1P3}$$

$$\pi_C = \frac{P1^2}{1 - P1P3}$$

where π_i is the asymptotic probability of being in review state i.

The quantities to be compared in the analysis are the average outgoing quality and the average amount of review at each possible field worker error rate. (See Duncan, p. 297-333) The average outgoing quality is the average number of errors missed for each level of review weighted by the asymptotic probability of being in each of these review levels. The average amount of review is the sum of the review rates weighted by the asymptotic probabilities of being at those review levels. Let L_i be the proportion of forms reviewed at review level i. Then $L_A = 1.00$, $L_B = 0.50$, $L_C = 0.25$. Algebraically, we then have:

$$\text{Average Outgoing Quality} = (1 - L_A) \cdot \pi_A \cdot R + (1 - L_B) \cdot \pi_B \cdot R + (1 - L_C) \cdot \pi_C \cdot R$$

and

$$\text{Average Amount of Review} = (L_A \cdot \pi_A + L_B \cdot \pi_B + L_C \cdot \pi_C) \cdot 100\%.$$

The sensitivity of the sampling plan to critical errors can be evaluated using a worse case scenario. A critical error will be hardest to detect if it is occurring only by itself. In this case, the level of review is not influenced by the "trigger" mechanism. Plotting the average outgoing quality by the rate of this one critical error results in graph number two. The worst outgoing quality occurs when this error is being made an average of once every ten forms. In this case, the sampling plan should reduce the outgoing rate of this error to about once every fifty forms. So in the worst case, the sampling plan will miss only one out of every five occurrences of this critical error. However, the actual rate should be much less, since the outgoing error rates are much smaller for other incoming error rates. Also, including other critical errors and other regular errors will increase the review rate and hence the outgoing quality.

Now, if P1, P2, and P3 are estimated, the corresponding amounts of review and quality can be estimated. A reasonable set of assumptions are that the number of errors made by each field worker is normally distributed and that the variability of the number of errors is the same for each field worker. The critical error rates were estimated from the field test. Let "r" be the estimated error rate for each of the five critical errors.

From a field test, the standard deviation of field worker errors was estimated to be 2.80 errors, and the field test presented no evidence that the assumptions were invalid. This allows estimates of P1, P2 and P3 by:

$$\hat{P}_1 = (Pr(X \leq T) \cdot (1 - r)^5)^{10}$$

$$\hat{P}_2 = 1 - ((1-r)^5)^{10}$$

$$\hat{P}_3 = 1 - \hat{P}_1 - \hat{P}_2$$

where $Pr(X \leq T)$ is the probability that a normal distribution with standard deviation 2.80 and mean equal to the field worker's average error rate will not have over T errors. The value of $(1-r)^5$ is the probability that none of the five critical errors occurs. The values of the average outgoing quality and the average amount of review were calculated for error rates ranging from 0 to 10 errors per form and for possible "trigger" values of from 3 to 10 errors. The results are shown in the third and fourth graphs.

Graph number three shows the average amount of review over the range of possible error rates. It shows that all "triggers" converge to 25% review for the lower error rates and 100% review for the larger error rates. The differences are in how quickly the curves go from 25% to 100%. A "trigger" of 3, for example, will result in a sharp jump between 25% and 100% review.

Graph number four shows the average outgoing quality of the forms over the range of possible error rates. To determine the quality of the resulting forms, consider the average outgoing quality limit (AOQL) for each of the "triggers". The AOQL is the largest value of the average outgoing quality over all error rates. For error rates of one to six errors and "triggers" of four to ten, as the "trigger" decreases by one the AOQL decreases by about 0.15.

The results of the field test were used to determine the error rates that will be encountered in practice. These actual error rates were used with the asymptotic results to obtain practical levels of review and average outgoing quality. These results are shown in the following table for the selected "trigger" of six errors.

Region	Approximate Percent of Forms Reviewed	Average Incoming Error Rate	Average Outgoing Error Rate
1	72%	7.06	0.19
2	69%	2.83	0.13
3	75%	4.71	0.11
4	74%	5.00	0.06
5	66%	4.60	0.48
6	59%	3.60	0.68
Total	70%	5.00	0.25

All of these values were weighted by the estimated level of review of each field worker and the average number of company visits for each field worker.

When this quality project has been in place for a reasonable length of time, the average number of errors per form should decrease. There should also be a corresponding decrease in the amount of review needed in each region.

Conclusion

The results of this paper show that the proposed sampling plan can be used to obtain levels of review and quality which are acceptable to the International Price Program. These levels of review and quality were adjusted by varying the "trigger" mechanism, and the field test results showed that six was a workable "trigger" value. This "trigger" value should result in an initial overall review rate of about 70% and an outgoing error rate of about 0.25 errors per form.

This quality project is scheduled to be implemented in January of 1987. A national office committee will be set up to monitor the results of this project. In particular, the committee will monitor the results of the reviewers to make sure the reviewers are checking the forms uniformly and accurately. This committee will also monitor the quality of the forms which are not reviewed. They can then make adjustments to the design to account for any major changes in the field.

References

Cinlar, Erhan, Introduction to Stochastic Processes, Prentice-Hall, Inc., 1975.

Duncan, Acheson, Quality Control and Industrial Statistics, Richard D. Irwin, Inc., 1965.

Table I

Deviations from the Asymptotic Probabilities

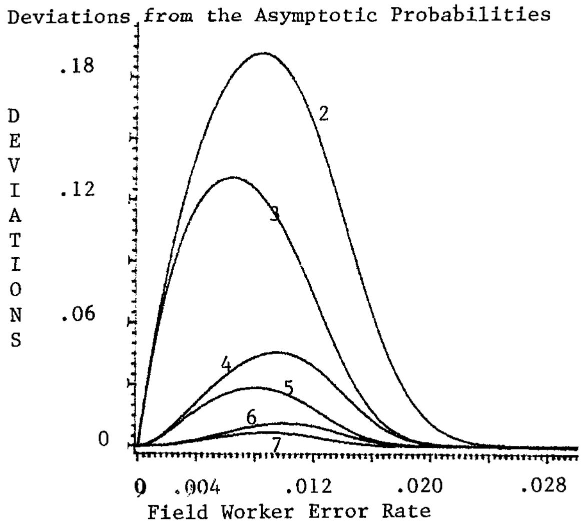


Table II
Plot of Average Outgoing Quality
by Rate of the Critical Error

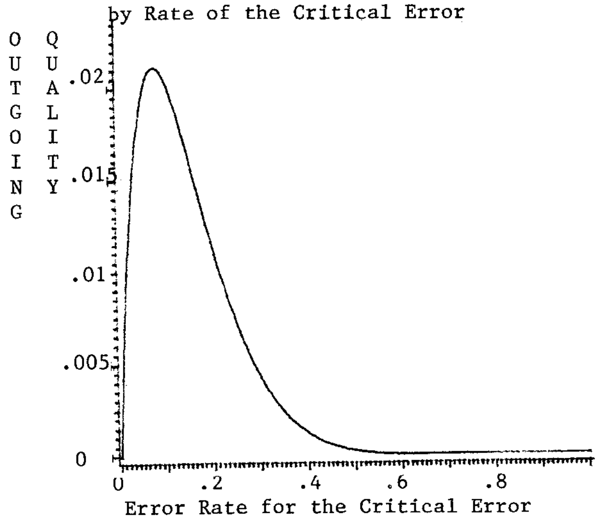


Table III
Plot of Average Amount of Review
by Field Worker Error Rate
for Selected Trigger Values

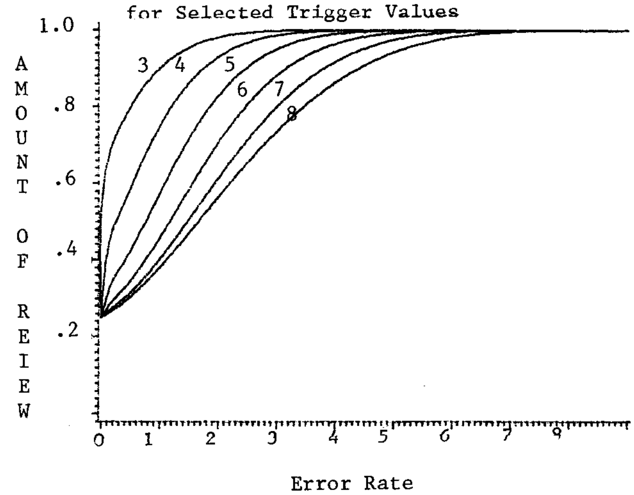


Table IV
Plot of Average Outgoing Quality
by Field Worker Error Rate
for Selected Trigger Values