CONTROL CONCEPTS IN SYNERGISTIC MAN-MACHINE INFORMATION SYSTEMS

Victor Selman
Charles I. Bartfeld
Jerry Selman
Jesse Gershberg

ABSTRACT

This paper describes the coupling of man-machine capabilities/limitations for controlling the efficiency (error rate) of large-scale Management Information Systems (MIS). Error reduction and prevention results are continually assessed by a Pareto-type analysis, and resources are allocated based upon an agreed prioritized listing of the seriousness of the consequences of error.
Statistical Quality Control (SQC) techniques and Sampling Inspection and Acceptance Plans were used both separately and in combination to control a large-scale MIS for public health.

INTRODUCTION

This paper deals with the essence of an innovative approach to ensure and maintain the quality of a large-scale operational MIS, namely applying standard quality control and acceptance sampling techniques to reduce and control error rates.

This approach is based upon a control system actually developed to monitor and control the quality of the man-machine subsystems used in processing data for a large-scale MIS for public health. Included in the system are data editing procedures intended to eliminate as many errors as feasible from the input data, and to control the various man and machine related error rates.

Computers are used to discriminate precisely, track changing edit criteria, rearrange data from diverse files and perform detailed manipulations and calculations. In the design of control procedures detailed simulations using GPSS were suggested in order to test out the operational validity of the concept with reference to delays within and between subsystems. The synergistic power of man-machine coordination provides insight into levels of risk to bridge the gap of real-world problems---between the precision and completeness of the machine and the imprecision and incompleteness of human thinking.

The rationale for this approach in using simple straightforward computerized control charts is their ease of understanding and comprehension, as contrasted with complex mathematical formulas and techniques. Quality control and statistical sampling techniques are used separately as well as in combination.

QUALITY CONTROL

Statistical Quality Control (SQC) techniques establish a sampling procedure to monitor the error rate (or quality) of various operations--some manual, some computer generated in the MIS. Decision criteria were developed to determine whether the specific aspects of data handling were in control. The SQC goal is to identify for management unusual variability--indicating that the process has undergone a change from the natural variability expected. This means that as yet unrecognized exogenous influences are acting upon the MIS.

For example, if the error rate at a certain subsystem operating under normal conditions in an MIS is about 1 percent, there would not necessarily be exactly 100 errors in a batch (lot) of 10,000 items/documents/transactions. However, on the average, the proportion of errors would be about 1 percent. If a certain batch contained 10 percent errors, quality control techniques (as well as common sense) would indicate that something had happened to the process which had been averaging 1 percent. The conclusion is that the process average had changed. Such a process is out of control, and would plot outside the quality control limits on Shewhart Control Charts, which are the so-called "three-sigma limits," including 99.73 percent of the population of interest. SQC procedures contain rules and tables to determine when variability from an error rate average is unusual. The rules are a function of the batch size, N, and the error rate, p, when the process is operating in a state of control.

To continually monitor the error rate, a table and chart (in time sequence) of sequential batches are maintained for important subsystems. Interpretation of the analysis of the data would detect trends and values out of control. As the general system included many subsystems acting as inputs into other subsystems, the relevant charts were modified for variable batch sizes both by changing the control limits for each specific batch size, or transforming the ordinate.

ACCEPTANCE SAMPLING

Acceptance Sampling techniques offer guidance on selecting the sample randomly and using the results to monitor and maintain an Acceptable Quality Level (AQL) for the subsystems and the overall MIS. Procedures too costly to be carried out on a 100 percent basis can be done on a sampling basis to reveal the process characteristics.

Sampling plans depend upon the characteristics desired using single-stage, double- or multi-stage plans. Each sampling plan, through its Operating Characteristic (OC) curve describes how often it will accept batches of a given error (quality) level. The objective is to find a plan that provides "acceptable" outgoing quality of data at minimum inspection cost for the MIS subsystems.

In general, larger sample sizes up to a point minimize the risk of sampling variability. Sampling plans are also used to allocate inspection efforts: batches accepted by the plan are not subject to further inspection; but batches rejected by the plan receive 100 percent inspection, and all defective batches are "rectified."

The Average Outgoing Quality Limit (AOQL), in effect, guarantees that the average output quality will be below a specific maximum. For batches with high error rates, such plans reject-reinspect-rectify many batches to achieve the required AOQL, but for batches with low error rates, these plans do not reject often. There exist specific plans which were developed by Dodge and Romig and Sampling Inspection Tables.

This AOQL concept was used in the large-scale MIS to track the quality as the documents/data flow through the process from one station to the next. The output from one station is the input quality to its sequential station. Table 1 shows the quality inspection levels implied by various error/defect rate requirements.

*Victor Selman - Consultant to Computer Sciences Corp.

TABLE 1

Insection Levels Implied By Defect Rate Requirements

GUARANTEED QUALITY LEVEL AFTER INSPECTION AOQL	PROCESS CAPABILITY (POSTULATED) p	MINIMUM SAMPLING FOR PRESCRIBED AOQL (n,c)	AVERAGE BATCH PROPORTION INSPECTED
.001	.001	(1210,2)	22.9%
.001	.002	(2028,4)	50.7%
.001	.0005	(775,1)	13.1%
.001	.0001	(355,0)	6.9%

LEGEND:

N = Batch Size = 10,000
n = Sample Size
c = Acceptance Number
p = Defect Rate characteristic of Claims Processing Subsystem

REMARK:

The AOQL Concept represents a highly conservative form of "consumer" protection which is, by design, entirely independent of "producer" claims of quality. This explains why inspection and screening are generally required even where the AOQL and p-levels coincide (see above). In this application the consumer may be identified with an auditor and the producer with an individual key punch operator.

COMBINED ACCEPTANCE SAMPLING & STATISTICAL QUALITY CONTROL

To maintain and control error rates and data quality as the information flows through a large MIS, a coordinated "hybrid" procedure was utilized. Quality Control procedures supplemented acceptance sampling plans for coding, keypunch-verification, etc., to prevent errors from entering the machine part of the process, rather than just yielding an AOQL. For example, an acceptance sampling procedure may reject batches containing the work of several different keypunchers. The individual quality control chart gives information indicating the source of error, and location within the rejected batches. In this case, the control chart indicates when to retrain individual keypunchers, etc.

TECHNIQUES USEFUL FOR CONTROL

Due to length limitations of this paper only selected techniques useful in the control of major subsystems can be identified, however, control problems relating to minor subsystems and their linkage and lower level operations cannot be addressed adequately.

MULTISTAGE SAMPLES

These are samples of samples of samples, as sampling is performed at more than one level (stage). Sub-samples are selected from each higher level sample.

SAMPLE PLANS

A sample plan provides the reviewer an organized method for defining the problem, choosing the most appropriate of several sampling methods and sample selection techniques, specifying sample reliability, evaluating the results, and documenting the program, as follows:

- Define the objective of the MIS;
- Determine the appropriate sampling method for each subsystem;
- Define the population of each subsystem of the MIS precisely, including stratification;
- Define the organization of items within the subsystem population, noting any periodic characteristics of the MIS;
- Specify the sample selection technique;
- State the population size;
- Estimate the maximum number (or dollar amount) of errors that the reviewer considers not to be material for the MIS;
- Specify the statistical assurance in terms of the reliability level and the precision desired;
- Specify the sample size;
- Record, evaluate and report the results for each MIS subsystem;
- Recommend changes in the MIS.

Statistical sampling in auditing deals with variables sampling and attributes sampling. Variables sampling is useful for estimating the average age of certain classes of recipients, or to answer questions concerning the amount of the population.

Attributes sampling (estimation sampling for attributes) is concerned with the number of occurrences that exist in a population-- the answer to the question, "How many?" For attributes sampling the value sought is an error rate stated as a percentage in absolute terms.

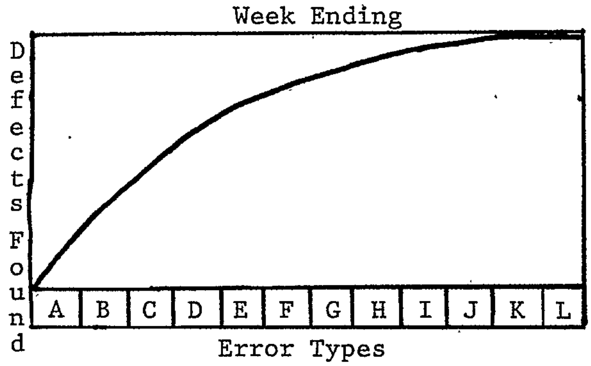
ERROR SELECTION PROCEDURES

The guiding principle in deciding on errors that could be improved with least effort is a Pareto-type analysis. This tool enables the identification of those "vital few" errors which contain the bulk of the improvement potential.

The graph in Figure 1 constitutes a Pareto Curve which has general applicability. It indicates that most of the defective work comes from a small fraction of the total possible causes. In this graph errors A-D constitute about 80% of all errors, thus identifying those errors where efforts to reduce the proportion of defectives should be concentrated. This curve is sometimes called the "80/20" Curve, because traditionally, about 80% of the defective work stems from 20% of the causes. This 80/20 rule-of-thumb, which had been used by accountants, paraphases a mathematical curve better known as the exponential distribution.

FIGURE 1

PARETO CURVE FOR ERROR ANALYSIS



MONITORING OF DATA QUALITY

The purpose of monitoring data quality for any MIS is to insure that the data error rate is acceptable, and is maintained for incoming data. There steps are involved in instituting and maintaining a MIS for monitoring data quality: (1) Defining what is meant by MIS error, including definition of the measurable parameters, (2) Establishing and maintaining an acceptable quality level (AQL), (3) Identifying causes of unacceptable error rates for each MIS subsystem and implementing corrective action.

- Error Definition. The term error may refer to a record containing one or more incorrect characters, a field containing one or more incorrect characters, or a specific incorrect character. Error rate is given as the Number incorrect divided by the Total Number (incorrect plus correct).
- Acceptable Quality Level (AQL). The AQL is defined as the maximum error rate that can be tolerated in the incoming data for each MIS subsystem. If the actual error rate exceeds this designated maximum value, the usefulness of the report is significantly diminished. The designation of a particular AQL can be made on the basis of historical experience or data, or subjectively. An AQL based on data is the error rate computed from a sample of that data.
- Unacceptable error rates. Once an error rate is out of control, the assignable causes are sought. The error rates of the sources can be computed, and compared to determine if if one or more sources are the major error sources. The chi-square (χ^2) is useful in making this comparison.

An important advantage of statistical sampling is the potential savings in costs through reduction in sample sizes when large populations are involved, in particular is sample sizes in the past have been determined as a percent of the population.

STATISTICAL ASSURANCE

The specific measure of the reliance that the reviewer can place on the inferences drawn from the characteristics of the sample is the statistical assurance. This measure is based upon the laws and theorie ries of probability which require that sample elements be randomly selected from the entire population. Assurance is stated in terms of reliability and precision.

RELIABILITY

Reliability is the probability that the statistic is within a stated range of the parameter in the population; it is also called the confidence level, and stated as a percent.

The selection of a reliability level depends upon how much risk the reviewer is willing to accept that the sample result lies outside the range chosen. Risk is the complement of reliability.

The reviewer's choice is between the penalty to be paid for being wrong and the cost of getting a more reliable sample.

PRECISION

The extent to which the population parameter may differ from the sample statistic and still be acceptable to the reviewer is known as precision; it is also referred to as a confidence limit or interval. Precision is a measure, frequently stated as a percent, extending on either side of the sample result or statistic, and within which range the value of the same characteristic for the population will lie at the stated reliability level.

QUALITY AUDITS

Acceptance Sampling Plans and Quality Control charting was also utilized for Quality Audits, and for using and understanding computerized audit programs, such as Haskins and Sell's Audi-tape Program. Sampling requirements are determined by the precision of the estimate required, the cost involved, and the risk acceptable to the auditor for making an error.

Computer Software Programs* available for QC use include

1. Random Number Generation
2. Histograms and Data Plotting
3. Control Charts
4. Limit Testing
5. Distribution Testing
6. Sampling Plan
7. Correlation Analysis
8. Analysis of Variance
9. Data Comparison
10. Probability Calculation (including Poisson, Binomial, Exponential, Normal, Hypergeometric, Chi-Square/Degrees of Freedom, Weibull Distributions)

SUMMARY

In the past statistical quality control techniques were primarily used to implement specific activities, but not complete MIS systems.

It has been demonstrated that a classical statistical quality control approach combined with acceptance sampling is a viable tool for controlling MIS error rates. As experience with utilizing the present system is gained, more sophisticated refinements can be incorporated.

---

* McCarn, D.B., and Moriarty, D.G. "Computers in Medicare," <u>Hospitals</u>, 45:37, 1971.