Appendix A. QUALITY CONTROL PRINCIPLES

General

Controlling the quality of an operation in the Bureau involves (1) measuring the quality of the process or product being controlled, (2) a timely reporting of information on the quality of the production process to its supervisor, and (3) correcting a faulty process, or rejecting poor quality products and replacing them by good quality products, or both. Information, communication, and action are inseparable elements in the operation of a quality control program.

In general, there are three types of statistical quality control programs:

1. **Process control** measures the quality of a process and provides for adjustments in the process when necessary.

2. **Acceptance sampling** determines which lots of items meet quality standards and are acceptable and which lots are unacceptable and should be reworked or replaced by good items.

3. **Hybrid procedures** utilize both process control and acceptance sampling.

Any of these three types of statistical quality control programs may either measure a specified multivalue characteristic of the product (inspection by variables) or classify the product as defective or nondefective (inspection by attributes). Inspection by the measurement of variables deals with fluctuation in the quality of the characteristic being controlled. This type of measurement is likely to be found where performance of equipment rather than performance of people is being controlled. Inspection in the measurement of the attributes is of the zero-one variable type—the item is defective or nondefective. Operations within the Census Bureau usually are controlled by use of the attributes type of measurement.

**Process Control**

The underlying principle of process control is that every work process is subject to a certain amount of variation due to unassignable causes that is regarded as chance variation. When fluctuations in the measurements are within the range of chance variation, the process is said to be in a "state of statistical control." A frequency distribution of measurements from such a process will have one mode, will tail off on both sides of the mode, and will be symmetrical or moderately skewed. A distribution of this type can be characterized by its mean and its variance, or some other measure of dispersion such as the range, which can be used to establish limits on a control chart.

As long as the process remains undisturbed by extraneous causes (assignable causes), the process will exhibit a state of statistical control. On the other hand, when important assignable causes operate to disturb the process, measurements will reflect this by unduly large deviations from the average of observations. The control chart will reflect this by showing points outside control limits; it is then necessary to look for causes of disturbance in the process and to correct them.

The theory underlying the use of the control chart works splendidly when the process primarily involves machinery of some type. In such an instance, when there is an out-of-control situation, some adjustment in the mechanism is usually required to bring the process back to a state of statistical control. However, many Census Bureau operations involve human effort more than machine effort. For example, when the performance of a human census coder drifted out of statistical control, a mechanical adjustment could not be made. It was necessary to inform the coder that his work was considered out of control and that he had to improve his performance. Presumably, when a clerk's output was out of control and he was told of this, and perhaps additional instruction or training was given him, he began to produce acceptable work again.

An individual may be regarded as having a "natural level" of error within limits at which he is capable of doing the job to which he is assigned. Then, the state of being out of control arises whenever his error rate exceeds his limits. Of course, in regard to a clerical job, consideration is given to the varying degrees of difficulty in the work assignments within the same job, supervision, and other factors. Effort is exerted to bring into control the work of clerks who go outside the control limits specified for the job. The action taken regarding a clerk who does not demonstrate the capability of performing within specifications is similar to that taken on a machine which cannot meet product tolerances. The clerk, like the machine, has to be replaced by one who can perform satisfactorily.

**Acceptance Sampling**

In an acceptance sampling plan for quality control, two curves provide information on its characteristics:

1. The operating characteristic (OC) curve. In general, the operating characteristics curve for a single sampling plan is determined by three numbers: the size of the work lot, the sample size, and the acceptance number. Points on the OC curve represent coordinates of the probability of acceptance and the true fraction defective. If the sample size is large relative to the lot size, the lot size will have an effect on the shape of the OC curve, and the formula for sampling from a finite population, the hypergeometric distribution, is used for computing the probability of acceptance of a lot for true values of the fraction defective. On the other hand, for most sampling problems the lot size is large relative to the sample size and the selection of successive items in the sample changes the probability negligibly and has practically no effect on the OC curve. Under such conditions, the assumption of sampling with replacement from an infinite population is made and probabilities are usually computed by means of the binomial or Poisson distributions.

When random samples are drawn, chance alone will inevitably operate to give rise to two types of wrong decisions. Some of the time, the sample information will indicate that acceptable lots or good quality producers are to be rejected. Also, some of the time, the sample information will indicate that substandard lots or producers are to be accepted. The ability to discriminate between good and bad lots or producers is reflected in
the shape of the OC curve. As the sample size is increased, the ability to discriminate between good and bad lots or producers is increased, but at an increased cost.

2. Points on the average outgoing quality (AOQ) curve represent the coordinates of the outgoing percent defective (ordinate) and the incoming percent defective (abscissa). For very good quality (as defined by specifications), the incoming and outgoing quality levels are virtually the same. This is because there are very few lots in which so many defectives are found in the sample that the lots are subjected to 100-percent inspection and correction of errors. For very bad quality (as defined by specifications), the outgoing quality levels show considerable improvement over incoming quality levels. This is achieved, however, by paying a price in the form of 100-percent inspection and correction. The AOQ curve has a maximum point called the average outgoing quality limit. This point does not represent the maximum error rate to be expected in any lot. It represents, just as all points on the AOQ curve do, the average outgoing quality expected for a given error rate over all lots having that error rate. The average outgoing quality limit, then, represents the average for those lots having an incoming error rate corresponding to the maximum point on the AOQ curve.

Hybrid Procedures

In the Bureau, inspection of clerical processes takes advantage of both process control and acceptance sampling. Consequently, the procedures are of a hybrid type in which there are both process control limits and lot rejection limits. When a process involving clerical work has gone out of statistical control, screening of the items within the work lot takes place. This process of screening and correcting work lots, actually within the area of acceptance sampling, operates regardless of whether the clerical process is in a state of control or not. When work produced by a clerk is outside process control limits, action is taken to locate the assignable cause and bring the work produced within statistical control. The work under these circumstances is not usually subjected to rescreening and correction. However, when the work quality is beyond the specification limits for the job for rejection of a lot, not only is the search for assignable cause made but in addition the substandard work is subjected to screening and correction. The quality control of the coding of population and housing census responses operated under such a hybrid plan.

Defects and errors arising in census operations affect the cost or duration of the work or the quality of the final statistics or both. Unfortunately, the effect of such errors is not always clear. Consequently, the design of plans for the control of quality in operations depends on judgments as to the levels of quality that should be specified and paid for. Once these levels have been decided upon, statistical theory replaces judgment and provides an answer as to whether the control level and expenditures are consistent. If the control level and cost are inconsistent, one or the other must change, and again judgment plays a part. If they are consistent, theory again plays a role in guiding the choice of the "best plan" among the "admissible" ones. "Best," as used here, means the lowest average outgoing percent defective for a fixed cost or lowest cost for a fixed outgoing percent defective.

Quality control plans are selected and compared on the basis of their costs and their power to discriminate between acceptable and unacceptable work lots or between the producers of these lots. If the preponderance of the lots and of producers are acceptable, the principal cost is determined by sample verification. If a substantial proportion of the lots or of the producers are not acceptable, the major part of the cost is determined by the action taken to improve outgoing quality.

Whenever possible, quality control plans in the Census Bureau have emphasized process control, which implies the prevention of defects, rather than acceptance sampling which implies the correction of defects. However, few of the plans employed during the 1960 census operations were all one or the other; most of them were hybrids which provided for correction of very poor work when it occurred as well as control of the process to prevent it.
Appendix B. FORMULAS

The formulas are shown below:

Average fraction inspected curve (AFI):
\[ \text{AFI} = \frac{f}{(1-f)(1-P)^i + f} \]

Probability of acceptance under sampling (Pa):
\[ \text{Pa} = \frac{(1-P)^i}{(1-f)(1-P)^i + f} \]

Average outgoing quality (AOQ):
\[ \text{AOQ} = (1 - \text{AFI})P \]

Regression Analysis

Symbols used in the formulas for calculating regression coefficients, correlation coefficients, and analysis of variance for regression (see chapter III) are defined as follows:

\( X_3 \), dependent variable (density)
\( X_1 \), independent variable (voltage)
\( X_2 \), independent variable (foot candles)
\( n \), number of sample readings
\( r_{31} \), simple correlation coefficient of density and voltage
\( r_{32} \), simple correlation coefficient of density and foot-candles
\( SSE \), sum of squares of errors or residuals
\( b_{31} \), regression coefficient of density on voltage
\( b_{32} \), regression coefficient of density on foot-candles
\( a_{31} \), intercept of regression line of density and voltage
\( a_{32} \), intercept of regression line of density and foot-candles
\( S_{b_{31}} \), standard error of regression coefficient \( (b_{31}) \)
\( S_{b_{32}} \), standard error of regression coefficient \( (b_{32}) \)
\( S_{a_{31}} \), standard error of constant
\( S_{a_{32}} \), standard error of constant
\( S^2_3 \), estimated variance of dependent variable \( (X_3) \)
\( X_{31} \), deviation of each reading on dependent variable from its mean \( (X_3) \)
\( X_{11} \), deviation of each reading on voltage from its mean \( (X_1) \)
\( X_{21} \), deviation of each reading on foot-candles from its mean \( (X_2) \)

Continuous Production Plan

Symbols used in the formulas for calculating characteristics of a continuous production plan (see chapter II) are defined as follows:

\( P \), true fraction defective
\( f \), sampling fraction
\( i \), number of units inspected in interval to be defect-free before sampling
The formulas were the following:

1. **Formulas for calculating sample estimates of parameters:**

   **Correlation coefficient:**
   
   \[ r_{31} = \frac{\Sigma x_{31}x_{11}}{\sqrt{\Sigma x_{31}^2}(\Sigma x_{11}^2)} \]

   \[ r_{32} = \frac{\Sigma x_{31}x_{31}}{\sqrt{\Sigma x_{31}^2}(\Sigma x_{31}^2)} \]

   **Regression coefficient:**
   
   \[ b_{31} = \frac{\Sigma x_{31}x_{11}}{\Sigma x_{11}^2} \]
   
   \[ b_{32} = \frac{\Sigma x_{31}x_{31}}{\Sigma x_{31}^2} \]

   **Intercept of regression line:**
   
   \[ a_{31} = \bar{x}_{31} - b_{31} \bar{x}_{11} \]
   
   \[ a_{32} = \bar{x}_{31} - b_{32} \bar{x}_{31} \]

   **Estimating equation:**
   
   \[ \tilde{X}_s = a_{31} + b_{31} X_{11} \]
   
   \[ \tilde{X}_s = a_{32} + b_{32} X_{31} \]

2. **Formulas for sample estimates of standard errors:**

   **Unexplained variation:**
   
   \[ 1 - r_{31}^2 \]
   
   \[ 1 - r_{32}^2 \]

   **Sum of squares of errors:**
   
   \[ SSE = \left(1 - r_{31}^2\right) \Sigma x_{31}^2 \]
   
   \[ SSE = \left(1 - r_{32}^2\right) \Sigma x_{32}^2 \]

   **Estimated variance:**
   
   \[ S_m^2 = \frac{SSE}{n-2} \]

   **Standard error of regression coefficient:**
   
   \[ S_{b_{31}} = \frac{S_m}{\sqrt{\Sigma x_{11}^2}} \]
   
   \[ S_{b_{32}} = \frac{S_m}{\sqrt{\Sigma x_{32}^2}} \]

   **Standard error of intercept:**
   
   \[ S_{a_{31}} = \sqrt{\frac{1}{n} + \frac{\bar{x}_{11}^2}{\Sigma x_{11}^2}}S_m \]
   
   \[ S_{a_{32}} = \sqrt{\frac{1}{n} + \frac{\bar{x}_{32}^2}{\Sigma x_{32}^2}}S_m \]

3. **Formulas for analysis of variance for regression:**

   **Mean squares:**
   
   **Regression:**
   
   \[ r_{31}^2 \Sigma x_{31}^2 \]

   **Error:**
   
   \[ (1-r_{31}^2) \Sigma x_{31}^2 \]

   **F Computed:**
   
   \[ F = \frac{(n-2)r_{31}^2}{(1-r_{31}^2)} \]

   **Degrees of freedom:**
   
   **Regression.................1**
   
   **Error.................n-2**

   **Coder Control**

   Symbols used in formulas in this part of the appendix are defined as follows:

   - \( x \), random variable in attributes measurement which is assigned a value of one if the item inspected in the sample is defective ("in error"), or assigned a zero otherwise.
   - \( p \), true error rate (fraction defective)
   - \( n \), probability that a dependent verifier will detect an item in error
Formulas used for calculating operating characteristic curves for lot acceptance, coder acceptance, and coder qualification (see chapter IV) are shown below.

Assumptions are as follows: (a) The product to be inspected comprises a series of successive work lots produced by a continuing process; (b) under normal conditions the lots are expected to be of the same quality; (c) the samples \( n_1 \) and \( n_2 \) are inspected independently of each other; and (d) the sample sizes are small relative to the lot size so that the binomial distribution can be used to compute probabilities.

1. **Lot acceptance**

The formula for computing the probability of accepting a work lot is as follows:

\[
L_1 = \sum_{x=0}^{n_1} \binom{n_1}{x} p^x (1-p)^{n_1-x} ; \quad \text{probability of accepting a work lot under independent verification}
\]

\[
L_2 = \sum_{x=0}^{a_3} \binom{n_2}{n_2} (\pi p)^{\pi p} (1-\pi p)^{n_2-x} ; \quad \text{probability of accepting a work lot under dependent verification}
\]

(Formula i) \( L = (L_1) \cdot (L_2) \); joint probability of accepting a work lot

2. **Coder acceptance**

**Probability coder will be accepted:**

\[
L_1 = \sum_{x=0}^{a_1} \binom{n_1}{x} p^x (1-p)^{n_1-x} ; \quad \text{probability of accepting a coder on the basis of the results of a sample where independent verification is used}
\]

\[
L_2 = \sum_{x=0}^{a_2} \binom{n_2}{n_2} (\pi p)^{\pi p} (1-\pi p)^{n_2-x} ; \quad \text{probability of accepting a coder on the basis of the results of a sample where dependent verification is used}
\]

(Formula ii) \( L = (L_1) \cdot (L_2) \); joint probability of accepting a coder on the basis of results from both the independent and dependent verification samples

In most sampling inspection plans, it is generally assumed that the process of inspecting sample items is carried out with negligible error. The assumption made is that verifiers examining items invariably classify them correctly as defective or nondefective. As the text material indicates, experience and research have shown
that inefficiency in inspection can be a very important factor in nullifying the theoretical characteristics of sampling inspection plans. A discussion of inspection inefficiency and sampling plans may be found in the following reference: Lavin, Marvin, "Inspection Efficiency and Sampling Inspection Plans," Journal of the American Statistical Association, Volume 41, No. 236, December 1946, pp. 432-438. Lavin presents formulas for determining the effect of verifier inefficiency on operating characteristic curves. Formulas \( L_{a} \) and \( L_{b} \) above are special cases of formulas presented in the article under the assumption that the probability of a verifier classifying a nondefective item as defective is small enough to be considered zero. The demonstration of this follows:

1. Definitions of symbols are as follows:
   - \( P_{a} \), probability of classifying a nondefective item as defective ("in error")
   - \( P_{b} \), probability of classifying a defective item as nondefective
   - \( p \), true fraction defective
   - \( p' \), effective fraction defective

2. The case is considered where \( P_{a} \) and \( P_{b} \) are constants and \( p' \) is a linear function of \( p \)

\[
p' = (1 - P_{a}) p + P_{b} (1 - p)
= P_{a} + (1 - P_{a} - P_{b}) p
\]

3. The probability of accepting a work lot under a dependent verification system assuming perfect efficiency in inspection would be given by the formula below.

\[
L_{a} = \sum_{x=0}^{n_{a}} \binom{n_{a}}{x} \left( P_{a} + (1 - P_{a} - P_{b}) p \right)^{x} \left( 1 - P_{b} \right)^{n_{a} - x}
\]

Relaxation of this assumption and inserting \( p' \) for \( p \) yields the following:

\[
L_{a} = \sum_{x=0}^{n_{a}} \binom{n_{a}}{x} \left( P_{a} + (1 - P_{a} - P_{b}) p' \right)^{x} \left( 1 - P_{b} \right)^{n_{a} - x}
\]

where \( (q = 1 - p) \) and \( L_{a} \) is the effective operating characteristic curve.

4. The probability of correctly classifying an item is given by \( \pi = 1 - P_{a} - P_{b} \). If the assumption is made that the probability of classifying a nondefective item as defective is small enough to be ignored, then \( \pi \) is equal to the probability of not classifying a defective item as nondefective (\( \pi = 1 - P_{a} \)). This assumption is plausible in certain types of coding processes where the incoming percent defective is of the order of 3 to 5 percent, and the dependent review clerk is likely to agree with the code assigned by the production coder.

5. By inserting \( \pi = 1 - P_{a} \) and assuming \( P_{a} \) equal to zero in the formula for the effective operating characteristic curve, the formula for \( L_{a} \) is changed as follows:

\[
L_{a} = \sum_{x=0}^{n_{a}} \binom{n_{a}}{x} \left( P_{a} + (1 - P_{a} - P_{b}) p' \right)^{x} \left( 1 - P_{b} \right)^{n_{a} - x}
= \sum_{x=0}^{n_{a}} \binom{n_{a}}{x} \left( \pi p' \right)^{x} \left( 1 - \pi \right)^{n_{a} - x}
= \sum_{x=0}^{n_{a}} \binom{n_{a}}{x} \left( \pi (1 - \pi)^{n_{a} - x}
\]

6. Similarly, the formula for \( L_{b} \), probability of accepting a coder on the basis of the results of a sample where dependent verification is used, is derived.

3. **Coder qualification**

Let \( M_{t} = (1 - 1)^{i} + 1 \left( (r - 1) s + k \right) \)

Let \( N_{t} = (1 - 1)^{i} + 1 \left( (r - 1 - 1) s + k \right) \)

Formula (iii)

\[
Q_{rs} = L^{*} + \sum_{i=1}^{r} \binom{L}{i} \left( 1 - L \right)^{i} - \frac{1}{2} \binom{N_{t}}{i} \left( L \right)^{i-1} \left( 1 - L \right)^{s} \]

probability of "s" successive acceptable decisions on a coder with a maximum of "d" decisions

The general formula above yields the following for computing the probability of coder qualification:

a. **General coding**

Probability of 4 successive accepted samples in a maximum of 12.

\[
d = 12; \quad s = 4; \quad r = \frac{4}{2} = 3; \quad k = 0
\]

\[
Q_{(4)(4)} = \left( \frac{8}{1} \right) L^{4} \left( 1 - L \right)^{4} \left( 1 - L \right)^{4} L^{4} \left( 1 - L \right)^{2}
\]

b. **Industry and Occupation Coding**

Probability of 5 successive accepted samples in a maximum of 25.

\[
Q_{(5)(5)} = \frac{L}{6} + \left( \frac{20}{1} \right) \left( 1 - L \right)^{5} \left( 1 - L \right)^{5} \left( 1 - L \right)^{5} 
\]

4. **Point system** — Probability that a coder will not be removed from production after training.

\[
U_{b,c} = \frac{D}{Z} \left[ \left( \frac{D}{Z} \right)^{2} \right] \left( \frac{D}{Z} \right) \left( \frac{C}{Z} \right) L_{b} \left( \frac{D}{Z} \right) \left( 1 - L_{b} \right) \left( \frac{C}{Z} \right)
\]

where \( F = D - \frac{Z}{2} \) and \( Z = (Q_{b,c}) \),

decisions, and \( C \) the initial credit. The parentheses denote the largest integral value contained in \( D + C - 1 \). This notation for the lower limit value of \( Z \) takes care of the case when \( D + C \) is even or odd. If \( D+C = 2k \), then \( Z = -C+2 \) is the lower limit. Admissible values for \( Z \) are even or odd for \( D = A_{D} - R_{D} \) are even or odd for \( D = A_{D} + R_{D} \) even or odd.

The quantity \( A_{D} \) is the cumulative number of decisions to accept, and \( R_{D} \) is the cumulative number of deci-
sions to reject based upon results of the independent verification samples.

The formula presented above can be simplified so that binomial tables such as those cited in reference (7) can be used for computations. This can be accomplished by making substitutions as follows:

\[ E = \frac{D - g}{2} \]
\[ V = E - C = \frac{D - g}{2} - C \]

The result is Formula (iv), below, used in computing the probability of a coder not being removed on or before D decisions.

(Formula (iv))

\[ U_{0, c} = f_1 - \left( \frac{1-L}{L} \right)^C f_2 \]

For simplification in writing the formulas for \( f_1 \) and \( f_2 \), the following substitutions are made:

\[ G = \frac{D + C - 1}{2} \]
\[ H = \frac{D - C - 1}{2} \]

then, \( f_1 \) and \( f_2 \) are as follows:

\[ f_1 = \sum_{E=0}^{G} \left( \frac{E}{G} \right) L_1^{D-E} (1-L_1)^E \]
\[ f_2 = \sum_{V=0}^{H} \left( \frac{V}{H} \right) L_1^{D-V} (1-L_1)^V \]

To illustrate the application of Formula (iv), the probability of a coder not being removed on or before ten decisions is computed below under the following conditions: \( L_1 = .60; C = 3; D = 10 \)

\[ G = \frac{D + C - 1}{2} = \frac{12}{2} = 6; \text{ upper limit for } E \]
\[ H = \frac{D - C - 1}{2} = \frac{6}{2} = 3; \text{ upper limit for } V \]
\[ f_1 = \sum_{E=0}^{6} \left( \frac{10}{E} \right) (1.60)^{10-E} (1.40)^E = 0.8452 \]
\[ f_2 = \sum_{V=0}^{3} \left( \frac{10}{V} \right) (1.60)^{10-V} (1.40)^V = 0.3923 \]

The values for \( f_1 \) and \( f_2 \) can be determined from those in Tables of the Cumulative Binomial Distribution, Cambridge, Mass., Harvard University Press, 1955, p. 308.

\[ U_{10,3} = f_1 - \left( \frac{1-L_1}{L_1} \right)^3 f_2 \]
\[ = (0.9452) - \left( \frac{0.40}{0.60} \right)^3 (0.3923) \]
\[ = 0.832 \]

This value for \( U_{10,3} \) can be found in the following table which shows the probability of survival \( U_{D,C} \) when points are permitted to be accumulated over D decisions and an initial state of C credits is granted to the coder.

<table>
<thead>
<tr>
<th>Prob. of acceptance (L)</th>
<th>Prob. of rejection (1-L)</th>
<th>Initial credit C = 1</th>
<th>Initial credit C = 3</th>
<th>Initial credit C = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(one decision)</td>
<td>(one decision)</td>
<td>Number of decisions (D)</td>
<td>Number of decisions (D)</td>
<td>Number of decisions (D)</td>
</tr>
<tr>
<td>1.00</td>
<td>.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.90</td>
<td>.10</td>
<td>.89</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>.80</td>
<td>.20</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td>.70</td>
<td>.30</td>
<td>.69</td>
<td>.57</td>
<td>.57</td>
</tr>
<tr>
<td>.60</td>
<td>.40</td>
<td>.59</td>
<td>.37</td>
<td>.34</td>
</tr>
<tr>
<td>.50</td>
<td>.50</td>
<td>.41</td>
<td>.18</td>
<td>.11</td>
</tr>
<tr>
<td>.40</td>
<td>.60</td>
<td>.22</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td>.30</td>
<td>.70</td>
<td>.12</td>
<td>.04</td>
<td>.01</td>
</tr>
<tr>
<td>.20</td>
<td>.80</td>
<td>.04</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>.10</td>
<td>.90</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>.00</td>
<td>1.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>
Matcher Control

A truncated sampling inspection plan is discussed in the text in connection with control of the matcher’s performance in detecting disagreements between punch coders and production coders. A decision to accept the work of a matcher is based on missing “a” or fewer planted errors in a sample of “n” errors. However, whenever in the course of inspecting a matcher’s work it is discovered that he has missed “a+1” planted errors he receives a “rejection.” The formulas in this part of the technical appendix deal with two aspects of the sampling inspection plan: (1) Determining the probability of accepting the matcher in one decision, and (2) determining the expected sample size (average sample number) per decision.

Symbols used in formulas in this part of the appendix are defined as follows:

x, random variable in attributes measurement which is assigned a value of one if the matcher fails to detect a planted error and zero otherwise

P, true error rate of the matcher

a, acceptance number for accepting a matcher in one decision

b, rejection number for a matcher; b = a + 1

n, sample size for reaching a decision to accept the work of a matcher

L, probability of accepting the work of a matcher in one decision

B, binomial distribution; for example

\[
B(n, p, b) = \sum_{x=b}^{n} \binom{n}{x} p^x (1-p)^{n-x}
\]

1. **Probability of matcher acceptance in one decision**

The probability of accepting the work of a matcher in one sample of “n” planted errors is computed by the following formula:

Formula (vi)

\[
L = B(n, P) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}
\]

The failure to detect a planted error is represented by the random variable \(x = 1\). The acceptance number is denoted by “a”, and the rejection number by \((b = a + 1)\).

2. **Average sample size per decision**

The plan calls for accepting the work of a matcher on the basis of a sample of “n” planted errors with “a” or fewer failures to detect coder and/or puncher errors. Rejection of a matcher occurs whenever “b” failures occur in the process of inspecting planted errors introduced in the matcher’s work. Thus, the plan involves curtailed sampling with truncation occurring in the process of rejection. The acceptance region is fixed at the coordinates \((n, a)\). The rejection points are denoted by coordinates \((j, b)\); the letter \(j\) is used here to denote a trial and to distinguish it from \(x\), used as the random variable for acceptance. The formula for determining the probability associated with the coordinates \((j, b)\) is the negative binomial.

\[
\text{Probability (j,b) = } \binom{j-1}{a} p^a (1-p)^{j-a}
\]

\[
\text{where j takes on integral values in the range b to n (inclusive). This is equivalent to the probability of “a” failures on the (j-1)st trial and a failure on the jth trial.}
\]

In developing the formula for computing average sample numbers for curtailed binomial sampling plans, use is made of the expected value of “j”, written as follows:

\[
E(j) = \sum_{j=a}^{n} j \binom{j-1}{b-1} p^b (1-p)^{j-b}
\]

It can be shown that the expected value of “j” is related to the binomial distribution in the following way:

\[
E(j) = b \left( \frac{1-P}{P} \right) B(n,b+1) + b B(n,P,b)
\]

The interested reader will find the mathematical proof in references (5) and (8). By use of this relation, the formula for the average sample number may be written as follows:

\[
\text{ASN} = n L + E(j) = n B(n, P) + b \left( \frac{1-P}{P} \right) B(n, P, b+1) + b B(n, P, b)
\]

\[
\text{ASN} = n \left( 1 - B(n, P, b) \right) + b \left( \frac{1-P}{P} \right) B(n, P, b+1) + b B(n, P, b)
\]

Formula (vii)

\[
\text{ASN} = n + b \left( \frac{1-P}{P} \right) B(n, P, b+1) - b n B(n, P, b)
\]
Appendix C. BIBLIOGRAPHY

Reference Books


Tables


Articles and Papers


Manuals


Vol. IV, Stage II Population and Housing Data Processing.


Part III, General Coding Instructions. 48, [5] pp., and attachments A-K (code lists, about 800 pp.)


Part VI, Industry and Occupation Coding. 9, 1 pp.


for the data you depend on...

STATISTICAL ABSTRACT of the UNITED STATES, 1965

Keeping current and conversant with America's dynamic economic, political, and social tempo is crucial, whether you're in business, education, research or in any one of a host of other jobs where you need figures that fit the facts.

The 1965 edition of the Abstract, published by the Bureau of the Census, is as authoritative, comprehensive, and timely, as the assembled research from 200 government, private, and international agencies can make it.

Organized by subject into 33 major sections, each with introductory text, the 1,299 tables and 50 charts in the 1965 edition are significant records of American activity and achievement. The detailed alphabetical index helps find the facts at hand and the extensive Guide to Sources tells where to find more.

$3.75 • Clothbound • Illustrated • 1059 pages

DON'T LET OBSOLESCENCE DATE THE DATA YOU DEPEND ON
Use the order form below and bring the facts into focus.

To: Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Enclosed find $____ (check, money order or Supr. of Documents coupons). Please send ____ copies of the Statistical Abstract of the United States, 1965, at $3.75 per copy.

PLEASE FILL IN MAILING LABEL BELOW

Name__________
Street address__________
City, State, ZIP code__________

POSTAGE AND FEES PAID
U.S. GOVERNMENT PRINTING OFFICE

OFFICIAL BUSINESS
RETURN AFTER 5 DAYS

FOR USE OF S.D.