This macro computes one-sided upper confidence limits for the number of defective units in a finite population (lot) using the hypergeometric probability distribution. A method is described which allows the results of the macro to be compared with similar results found in published tables. The macro is useful to industrial quality control personnel in estimating the maximum number of defective units in finite-sized lots from sampling inspection results where prior information about the process producing the lots is unknown.

INTRODUCTION

Industrial quality control practitioners often estimate the number of defective items in a lot from the number of defectives found in a sample of the lot. Typically, such estimates are assumed to follow the binomial distribution provided that:

- the lot size was "infinitely" large, or
- the product comes from a process in continuous production where series of lots are assumed to be formed at a constant fraction defective (p).

If the lot is small-to-moderate in size (200 or less), or of an isolated nature (e.g., product supplied from producers only once or during irregular production periods), the binomial distribution tends to yield inaccurate estimates of lot defectives and should be replaced with the hypergeometric distribution. (See Duncan, 1974; Tomsky, Nakano and Iwashita, 1979; Grant and Leavenworth, 1980).

The hypergeometric is applicable because inspections are conducted over a finite, rather than an infinite, set of units. Nondefective units are kept while defective units are returned for rework, or scrap (i.e., sampling without replacement). Also the probability of finding a defective unit (p) changes since the lot size reduces each time a unit is inspected. Thus p is not constant. In a binomial process, sampled units would be placed back into the lot and another sample from the lot is drawn (i.e., sampling with replacement). It should be apparent that the hypergeometric represents a more realistic situation than the binomial. The hypergeometric does approximate the binomial if the sample size is less than ten percent of the lot size.

According to Calvin (1984), the hypergeometric distribution also provides better assurance of lot quality in instances:

- where knowledge (or data) about the variation of the process which produced the lot is unknown or cannot be assumed; or,
- when supplier lot quality history is suspicious.

If lot quality is important, the hypergeometric serves as a more appropriate model to use in computing confidence limits on the number of defectives in finite-sized lots, as poor quality lots could significantly delay consumer assembly operations which use the lot components.

Although tables for binomial confidence limits are prevalent throughout the literature, tables for hypergeometric confidence limits are rare. This lack is largely due to the effort involved in computing hypergeometric probabilities.

Chung and D'Arcy (1950) presented charts for hypergeometric confidence limits for lot sizes of 500; 2,500; and 10,000. Odoh and Owen (1983) presented tables for one and two-sided confidence limits on the number of defective units for selected lot sizes of 400 or more. Tomsky et al. (1979) furnished tables for selected lot sizes up to 100 and described conditions for which hypergeometric confidence limits should satisfy. As with most tables, the tables for the hypergeometric tables, one is restricted to nominal lot sizes, sample sizes and risk levels. One would have to employ tedious calculations and/or interpolations in order to derive user-specific confidence limits. The SAS macro \texttt{(LOTDEF)} in this paper runs interactively under MVS/TSO and applies the conditions described by Tomsky et al. (1979) to obtain the one-sided upper confidence limits for the number of defectives ($D_+$) in a lot. The computer and macro free the user from computational burdens and allow greater flexibility in meeting specific user needs.

HOW IS THIS MACRO USEFUL?

Suppose a lot is received from a distributor or supplier who submits the product for the first time. Even though incoming sampling inspection results may indicate that the lot is acceptable, there is no guarantee that the remaining uninspected items in the lot are nondefective.

The macro thus serves as an effective means of estimating the lot defectives. Statistical control limits may be established about the proportion (fraction) of lot defectives. Such results may serve as a basis for initiating "certified material" where materials may be exempt from sampling inspection as more acceptable quality history is accumulated.
Factory or assembly rejections from the inventory of the accepted lots may be subtracted from the estimated lot defectives. The estimated lot defectives may be stored in a "bucket" of a quality history database as an adjustment of a lot quality rating (e.g., lot fraction defective).

The macro may also be useful in the development of more realistic sampling plans. Such plans based on the macro would be in closer agreement with quality goals than would traditional plans, especially when applied to lots undergoing incoming inspection. Should tables be necessary, the macro may be modified to generate the tabulated values of lot defectives.

LIMITATIONS

The same tabulated values obtained by Tomsky et al. (1979) were obtained using the macro. However, some of the macro-derived values were one less than the tabular values of the upper confidence limits obtained by Odeh and Owen (1983). Because the hypergeometric is a discrete distribution, exact probabilities cannot be calculated for the risk (x) levels given. The macro therefore determines the confidence limits which best approximate the closest risk level for the user-supplied parameters.

MACRO DESCRIPTION

The following algorithm, used to compute the maximum defectives in isolated lots, assumes 100% (1-α) percent confidence that the lot is hypergeometrically distributed. The procedure calls for the user to logon to TSO; allocate files; invoke SAS; and call the macro.

Macro \( \text{LOTDEF} \) executes the following steps:

Step 1. Users input Lot Size (N), Sample size (n), Number of defectives in the sample (d), and risk or error (α) one is willing to accept. Note the confidence level is 1-α.

Step 2. Set the upper 100*(1-α)% confidence limit of the number of defectives in the lot to \( D_\alpha = d+1 \). Note that the domain of d is in the interval, \( \max(0,n-N+n,n) \leq d < \min(D_\alpha,n) \).

Step 3. Compute the probability of finding d defectives in a sample of n units obtained from a lot size of \( N \) having \( D_\alpha \) defectives as:

\[
P(d;N,D_\alpha,n) = \sum \left[ \binom{N-D_\alpha}{d-N} \binom{D_\alpha}{N-d} \right] / \binom{N}{n}
\]

In SAS the PROBHT function was used to compute the hypergeometric probabilities.

Step 4. If \( P(d;N,D_\alpha,n) < \alpha \), then the maximum \( D_\alpha \) equals \( D_\alpha - 1 \). Otherwise increase \( D_\alpha \) by 1 and repeat step 3.

The macro prints the user supplied inputs, the estimated number of lot defectives \( (D_\alpha) \), and the additional information to the terminal screen.

The user may change the macro to work "backwards" by:

- Setting \( D_\alpha \) equal to the lot size \( (N) \) in step 2, and
- Decreasing \( D_\alpha \) by 1 until \( P(d;N,D_\alpha,n) \geq \alpha \) in step 4. The same results will also be obtained.

EXAMPLES

Three examples are presented to illustrate the use of the macro with lots of varying sizes. Example 1 describes the use of the macro with a lot containing 100 items. Examples 2 and 3 compared confidence limits obtained by other researchers for the estimated lot defectives with those obtained using the macro.

Example 1: The author randomly selected a sample of 13 items from a received lot containing 100 items. After inspecting the sample, zero defectives were found. The macro was used to calculate the upper confidence limit on the number of defectives expected in the lot.

\[
\text{THIS MACRO ESTIMATES ONE-SIDED UPPER CONFIDENCE LIMITS FOR THE NUMBER DEFECTIVES IN A LOT BASED ON THE HYPERGEOMETRIC DISTRIBUTION.}
\]

INPUT THE LOT SIZE \((L)\), DEFECTIVES FOUND IN THE SAMPLE \((D)\), SAMPLE SIZE \((n)\), AND RISK \((\alpha)\).

<table>
<thead>
<tr>
<th>LOT SIZE= 100</th>
<th>ESTIMATED LOT DEFECTIVES= 19</th>
<th>SAMPLE SIZE=n= 13</th>
<th>ESTIMATED IN SAMPLE= 0</th>
<th>LOT FRACTION DEFECTIVE= 0.1900</th>
<th>PROPORTION GOOD= 0.8100</th>
<th>CONFIDENCE LEVEL= 94.722</th>
</tr>
</thead>
</table>

This means that there is 94.72 percent confidence that:

- There are no more than 19 defectives in the lot.
- The lot fraction defective is no more than 19/100=0.19.
- There are at least 100-19=81 good items in the lot.
- The proportion of good items in the lot is no less than 81/100=0.81.
Tomasky et al. (1979) reported that a randomly selected sample of 10 items was taken from a lot having 30 items. The sample was tested and two defective items were found. The macro was used to calculate the number of defectives expected in the lot.

\[
\text{Example 2}
\]

This macro estimates one-sided upper confidence limits for the number of defectives in a lot based on the hypergeometric distribution.

Input the lot size (L), defectives found in the sample (d), sample size (n), and risk (alpha).

Lot size = 30
Defectives found in the sample = 2
Sample size (n) = 10
Risk (alpha) = 0.05

Estimated lot defectives = 15
Estimated proportion defective = 0.15
Confidence level = 95.26%

The macro confirmed the confidence limit found by Tomasky et al. (1979).

Example 3

The third example is from Odeh and Owen (1983). Twenty defective units were observed in a sample of 170 units from a lot of 1600 units with a risk of 0.0234. The macro was used to calculate the number of defectives expected in the lot.

The macro estimated the number of defective units in the lot to be at most 777 which assures a confidence level of 97.58%. This confirms Odeh and Owen's (1983) upper confidence limit.

CONCLUSION

The macro just described provides a systematic method for estimating the maximum number of defective items in finite-sized lots. The macro utilizes SAS' hypergeometric distribution function (PROBHYPER) which yields more accurate limits than the traditional binomial distribution estimates. The binomial limits tend to yield inaccurate estimates of the number of defectives in small or moderate-sized lots. These limits agree appreciably with existing tables and can serve as a gauge of lot quality which can result in the construction of more realistic sampling plans. Periodic 100 percent inspections of small-sized lots may be used to verify the accuracy of the estimates derived by the macro. As more quality history is accumulated, the estimates may be refined, via Bayesian techniques, to reflect more realistic and adjusted measures.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Lucia Ward-Alexander and Janice Staiger for their assistance in the preparation of this paper.

REFERENCES


The author may be contacted at:

M. T. Alexander
Procurement Quality Engineering
Westinghouse Electric Corporation
314 S. Hammond Ferry Road, M.S. 5551
Glen Burnie, MD 21061
(301) 785-7686

* SAS is the registered trademark of SAS Institute, Inc., Cary, NC USA
**MACRO LOTDEF**

**PROGRAM LISTING**

**LOTDEF: CONFIDENCE LIMITS FOR THE NUMBER OF DEFECTIVES IN A LOT**

**USING THE HYPERGEOMETRIC PROBABILITY DISTRIBUTION**

**AUTHOR: M.T. ALEXANDER, DATE: 03/20/65**

**COPYRIGHT 1965 WESTINGHOUSE ELECTRIC CORPORATION**

**ALL RIGHTS RESERVED**

******************************************************************************

TSO ALLOC F(IN) DA(*); TSO ALLOC F(OUT) DA(*);
CLEAR PAUSE;
DATA _NULL_;
INFILE IN UNBUFFERED EOF=LAST;FILE OUT;
   PUT 'THIS MACRO ESTIMATES ONE-SIDED UPPER CONFIDENCE LIMITS FOR ' /
   'THE NUMBER DEFECTIVES IN A LOT BASED ON THE HYPERGEOMETRIC ' /
   'DISTRIBUTION.' /
   'INPUT THE LOT SIZE (L), DEFECTIVES FOUND IN THE SAMPLE (D) ' /
   'SAMPLE SIZE (N) AND RISK (ALPHA). '
INPUT NN X N ALPHA;
   I=X;
   TWO: I=I+1;
   IF X ( MAX(0,N+I-NN) THEN P=0;
   ELSE IF X =) MIN(I,N) THEN P=1;
   ELSE P=PROBHYPR(NN,I,N,X);
   IF P < Alpha THEN DO;
      .'
      LAST, STOP;
   RUN,
   XEND LOTDEF
   DD=I-I,
   LFD=DD/NN,
   GOOD=I-LFD;
   PR=PROBHYPR(NN,DD,N,X);
   CON=100*( I-PR);
   PUT 45*'-';
   PUT 'LOT SIZE= ' NN,
   PUT 'ESTIMATED LOT DEFECTIVES= ' DD;
   PUT 'SAMPLE SIZE(N)= ' N,
   PUT 'DEFECTIVES IN SAMPLE(D)= ' X;
   PUT 'LOT FRACTION DEFECTIVE= ' LFD 6.4;
   PUT 'PROPORTION GOOD= ' GOOD 6.4;
   PUT 'RISK(ALPHA)= ' PR 6.4;
   PUT 'CONFIDENCE LEVEL= ' CON 6.2 '7',
   PUT 45*'-';
   GOTO LAST;
   END;
   ELSE GOTO TWO;
   LAST: STOP;
   RUN;
XEND LOTDEF