BNWL-1385
UC-15

5-
5-70

THE ROLE OF STATISTICS IN SAFEGUARDS

K. B. Stewart and J. L. Jaech

May 1970

AEC RESEARCH &
DEVELOPMENT REPORT

BNWL-1385

BATTELLE NORTHWEST
BATTELLE MEMORIAL INSTITUTE
PACIFIC NORTHWEST LABORATORIES
BATTELLE BOULEVARD, P. O. BOX 999, RICHLAND, WASHINGTON 99352

PACIFIC NORTHWEST LABORATORY
RICHLAND, WASHINGTON
operated by
BATTELLE MEMORIAL INSTITUTE
for the
UNITED STATES ATOMIC ENERGY COMMISSION UNDER CONTRACT AT(45-1)-1830

AEC-RL RICHLAND, WASH.

BNWL-1385

UC-15, Safeguards and Nuclear Materials Management

THE ROLE OF STATISTICS IN SAFEGUARDS

By

K. B. Stewart and J. L. Jaech

Applied Mathematics Department
Systems and Electronics Division

May 1970

BATTELLE MEMORIAL INSTITUTE
PACIFIC NORTHWEST LABORATORIES
RICHLAND, WASHINGTON 99352

Printed in the United States of America
Available from
Clearinghouse for Federal Scientific and Technical Information
National Bureau of Standards, U.S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

THE ROLE OF STATISTICS IN SAFEGUARDS

K. B. Stewart and J. L. Jaech

ABSTRACT

This document, BNWL-1385, discusses certain specific uses of statistics in connection with safeguards, both current and projected. The principal theme is that the application of statistics plays an important role in safeguarding nuclear materials, and that statistics should become an increasingly important tool in the safeguards field.

TABLE OF CONTENTS

THE ROLE OF STATISTICS IN SAFEGUARDS

K. B. Stewart and J. L. Jaech

INTRODUCTION

This report is a study of ways and means in which statistics may be used more fully in maintaining accurate inventory control of nuclear materials. In a nuclear materials control system (which is a vital part of an entire safeguards system) data play a dominant role. An extensive portion of the available information about the system is contained in the data, and the decisions (of a safeguards nature) that must be made are heavily dependent upon the data generated. Because of this important role, it is essential that great care be used in collecting, storing, and analyzing the data. Any tool that can be used to assist in these tasks should be used. One important tool is statistics. The proper application of statistical principle is essential to ensure that conclusions drawn from the data are valid, to indicate the kinds of data that are needed, to specify the quality of the data required, and to extract maximum information from the available data.

This study gives some specific uses of statistics in connection with safeguards, both current and projected. In developing this theme, it is hoped that an increased awareness will be incurred of the importance of statistics to safeguards, and that concerned individuals will be stimulated to evaluate more fully the practicability of applying these modern techniques to safeguards problems.

SUMMARY

Because statistical procedures are far from routine, it is increasingly important that highly trained professional people be used in any endeavor to apply statistics as a tool in safeguards.

A wide range of measurements and more accurate measurements will accomplish much in obtaining a valid image of what a materials unaccounted for (MUF) incident may represent.

MUF significance must be weighed along with such other indications as the internal consistency of isotopic data and it is the integration of all such aspects that will provide the most valid answer as to whether a materials diversion has taken place.

Game theory or the application of strategy and counter-strategy is a worthwhile approach to identifying the cause of an apparent MUF.

CONCLUSIONS

(1) The statistical procedures required to obtain maximum, valid information from the data are, in many cases, far from routine. This fact, together with the extreme importance of safeguards, calls for added emphasis upon the necessity to utilize professionals trained in this discipline to the greatest extent possible.

(2) The MUF is a primary indicator of safeguards performance. There is strong motivation to reduce the variance of MUF. This can be accomplished in several ways with the most obvious way being to obtain better measurements, or perhaps more measurements. However, reduction in variance can be achieved at no additional cost of collecting data simply by using more sophisticated methods of data analysis to exhaust the specific information from the data. Under certain realistic conditions, the gains in efficiency achieved in this manner can be very large.

(3) Significant attention is being given the MUF as an indicator of missing material or diversion. However, there may be other indicators such as the internal consistency

of isotopic data. The integration of all of these aspects should lead to better decisions on whether diversion has taken place, how much has taken place, and where to look for the source of diversion.

(4) Consideration is being given other applications of statistics to the problems of safeguards. One good possibility for a significant contribution is the establishment of clear, quantitative goals for safeguards. Another broad area is the application of game theory. Since the control of nuclear materials in a true safeguards environment is essentially a game of strategy and counter-strategy, the theory of games provides a statistical subject-matter field pertinent to evaluating overall strategies.

DISCUSSION

SPECIFIC APPLICATIONS

Minimum Variance MUF

The MUF is a concept vital to the control of nuclear materials. The estimation of MUF appears to be very simple and straightforward, and yet, for certain conditions, more sophisticated estimation procedures have been developed which provide estimates having properties which represent significant improvements over those of the simple estimation method. The estimation procedure will, under these conditions, provide estimates with much greater precision and, in addition, will remove the correlations between successive MUF's that exist using the simple estimation procedure.

It is assumed initially that throughput is relatively constant, and that measurements in the system can be classified as input, output, and inventory measurements. Let

x_i = amount of input minus amount of output for month i (a time period other than one month may be used)

y_i = ending inventory physical measurement at the end of month i [or beginning of month (i + 1)].

μ_i = true amount of material in inventory at end of month i.

Further assume that

$$E(y_i) = \mu_i \tag{1}$$

$$\text{var } y_i = \sigma_y^2 \text{ for all } i \tag{2}$$

$$\text{var } x_i = \sigma_x^2 \text{ for all } i \tag{3}$$

The traditional simple estimate of the MUF at the end of month n is

$$M_n' = (y_{n-1} + x_n) - y_n \tag{4}$$

An estimation procedure is suggested which will provide a different estimate for μ_n and lead to a different estimate for MUF. Define a set of z values as follows:

$$z_n = y_n$$

$$z_{n-1} = y_{n-1} + x_n$$

$$z_{n-2} = y_{n-2} + x_{n-1} + x_n$$

$$\vdots$$

$$z_1 = y_1 + x_2 + x_3 + \cdots\cdots x_n$$

$$z_0 = y_0 + x_1 + x_2 + \cdots\cdots x_n \tag{5}$$

z_n is simply the ending physical inventory for month n and hence is an estimate of μ_n. With no losses, z_{n-1} is also an estimate of the inventory at end of month n. The same is true of all the z's; they all have expected value μ_n. This suggests that some averaging procedure applied to all the z's might provide a better estimate of μ_n than simply using the physical inventory at the end of month n.

Intuitively, one would believe that the "most recent" z's would be better estimates of the current inventory than would the earlier z's. This would suggest the use of some weighting procedure. Consider the estimate of μ_n consisting of the linear combination

$$I_n = \sum_{0}^{n} \alpha_i \; z_i \qquad (6)$$

where

$$\sum_{0}^{n} \alpha_i = 1$$

I_n is an unbiased estimate of μ_n under the assumptions of this model. In determining the set of α's to use to provide the "best" estimate of μ_n, a reasonable criterion is to select that set which results in a minimum variance for the estimator. The derivation of this estimator is somewhat complicated by the fact that the z's are not mutually independent. The covariance between z_i and z_j is

$$\text{cov}\ (z_i, z_j) = (n - j)\ \sigma_x^2,\ i < j \qquad (7)$$

Also, the variance of z_i is

$$\text{var}\ z_i = \sigma_y^2 + (n - j)\ \sigma_x^2 \qquad (8)$$

Denote by V the variance-covariance matrix for the z's. Then, one can use the generalized least squares estimate[1] for μ_n, denoted by $\hat{\mu}_n$, and appeal to the Gauss-Markoff theorem to assert that this is the minimum variance unbiased estimator. This becomes

$$\hat{\mu}_n = (T' \ V^{-1} \ T)^{-1} \ T' \ V^{-1} \ Z \tag{9}$$

where T is a column vector of ones with (n+1) rows, and Z is the column vector of z's. The variance of $\hat{\mu}_n$ is

$$\text{var } \hat{\mu}_n = (T' \ V^{-1} \ T)^{-1} \tag{10}$$

As a very simple example, suppose

$$\sigma_y^2 = 5 \text{ units}^2$$

$$\sigma_x^2 = 1 \text{ unit}^2$$

In this example, the variance of a physical inventory is five times as large as the variance of the total input less the total output. Consider the estimate of $\hat{\mu}_2$. The z's become

$$z_2 = y_2$$

$$z_1 = y_1 + x_2$$

$$z_0 = y_0 + x_1 + x_2 \tag{11}$$

The variance-covariance matrix is

$$V = \begin{pmatrix} 7 & 1 & 0 \\ 1 & 6 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

$$V^{-1} = \frac{1}{205}\begin{pmatrix} 30 & -5 & 0 \\ -5 & 35 & 0 \\ 0 & 0 & 41 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$(T' \; V^{-1} \; T)^{-1} = \frac{205}{96} = \text{var } \hat{\mu}_2$$

$$\hat{\mu}_2 = \frac{205}{96} T' \; V^{-1} \; Z$$

$$= (0.26042 \;\; 0.31250 \;\; 0.42708) Z$$

This indicates that the "best" linear combination to use in this example is

$$\hat{\mu}_2 = 0.26042 z_0 + 0.31250 z_1 + 0.42708 z_2$$

which has variance 205/96 or 2.135 units^2. Note that the usual estimate of μ_2, namely y_2, has a variance of 5 units^2 which says the gain in precision in this example is very significant.

It should be pointed out that the particular linear combinations of z's which should be used and the improvement in precision depend only on the ratio of σ_x^2 to σ_y^2, and not on their values. This permits the construction of tables of the α coefficients as a function of $c = \sigma_x^2/\sigma_y^2$ and n. Table 1 gives these coefficients for c = 0.10(0.10) 1.80 and n = 1(1) 10.

Given this new estimating procedure for μ_n, the next problem is to see how the result might be used in estimating the MUF. The reasonable choice is to use the best estimate of the ending inventory for month (n-1) along with the (inputs - outputs) and ending physical inventories for month n. Denote this estimate of MUF by M_n''.

TABLE 1

THE C VALUE IS .10

1	1.00000									
2	.47619	.52381								
3	.29326	.32258	.38416							
4	.19759	.21735	.25884	.32622						
5	.13854	.15239	.18149	.22873	.29885					
6	.09904	.10894	.12974	.16351	.21364	.28513				
7	.07150	.07865	.09367	.11805	.15424	.20585	.27804			
8	.05189	.05708	.06797	.08566	.11192	.14938	.20177	.27433		
9	.03775	.04153	.04946	.06233	.08144	.10869	.14681	.19961	.27237	
10	.02751	.03026	.03604	.04542	.05934	.07920	.10698	.14545	.19847	.27134

THE C VALUE IS .20

1	1.00000									
2	.45455	.54545								
3	.26042	.31250	.42708							
4	.16005	.19206	.26248	.38540						
5	.10095	.12114	.16556	.24309	.36925					
6	.06433	.07720	.10550	.15491	.23530	.36275				
7	.04117	.04940	.06751	.09913	.15057	.23212	.36010			
8	.02639	.03166	.04327	.06354	.09651	.14879	.23082	.35902		
9	.01693	.02031	.02776	.04076	.06191	.09544	.14805	.23028	.35857	
10	.01086	.01303	.01781	.02615	.03972	.06123	.09499	.14775	.23006	.35839

THE C VALUE IS .30

1	1.00000									
2	.43478	.56522								
3	.23310	.30303	.46387							
4	.13215	.17180	.26298	.43306						
5	.07625	.09913	.15175	.24988	.42299					
6	.04426	.05753	.08807	.14503	.24550	.41961				
7	.02574	.03346	.05122	.08434	.14276	.24402	.41847			
8	.01498	.01947	.02980	.04908	.08308	.14200	.24351	.41809		
9	.00872	.01133	.01735	.02857	.04835	.08265	.14174	.24335	.41796	
10	.00507	.00660	.01010	.01663	.02815	.04811	.08250	.14165	.24329	.41791

TABLE 1 (contd)

THE C VALUE IS .40

1	1.00000									
2	.41667	.58333								
3	.21008	.29412	.49580							
4	.11082	.15514	.26152	.47252						
5	.05918	.08285	.13966	.25234	.46596					
6	.03172	.04440	.07485	.13524	.24972	.46408				
7	.01701	.02382	.04015	.07255	.13396	.24896	.46354			
8	.00913	.01278	.02155	.03893	.07189	.13360	.24874	.46339		
9	.00490	.00686	.01156	.02089	.03858	.07169	.13349	.24868	.46334	
10	.00263	.00368	.00621	.01121	.02070	.03848	.07164	.13346	.24866	.46333

THE C VALUE IS .50

1	1.00000									
2	.40000	.60000								
3	.19048	.28571	.52381							
4	.09412	.14118	.25882	.50588						
5	.04692	.07038	.12903	.25220	.50147					
6	.02344	.03516	.06447	.12601	.25055	.50037				
7	.01172	.01758	.03223	.06299	.12525	.25014	.50009			
8	.00586	.00879	.01611	.03149	.06262	.12506	.25003	.50002		
9	.00293	.00439	.00806	.01575	.03131	.06253	.12502	.25001	.50001	
10	.00146	.00220	.00403	.00787	.01566	.03127	.06251	.12500	.25000	.50000

THE C VALUE IS .60

1	1.00000									
2	.38462	.61538								
3	.17361	.27778	.54861							
4	.08080	.12928	.25533	.53458						
5	.03785	.06057	.11962	.25044	.53152					
6	.01776	.02841	.05612	.11749	.24936	.53085				
7	.00833	.01333	.02634	.05514	.11703	.24913	.53070			
8	.00391	.00626	.01236	.02588	.05492	.11692	.24907	.53067		
9	.00184	.00294	.00580	.01215	.02578	.05488	.11690	.24906	.53066	
10	.00086	.00138	.00272	.00570	.01210	.02576	.05487	.11689	.24906	.53066

TABLE 1 (contd)

THE C VALUE IS .70

1	1.00000									
2	.37037	.62963								
3	.15898	.27027	.57075							
4	.07001	.11902	.25135	.55962						
5	.03098	.05267	.11123	.24766	.55745					
6	.01373	.02333	.04927	.10971	.24694	.55702				
7	.00608	.01034	.02183	.04861	.10941	.24680	.55694			
8	.00269	.00458	.00967	.02154	.04848	.10935	.24677	.55692		
9	.00119	.00203	.00429	.00954	.02148	.04845	.10934	.24676	.55692	
10	.00053	.00090	.00190	.00423	.00952	.02147	.04845	.10934	.24676	.55692

THE C VALUE IS .80

1	1.00000									
2	.35714	.64286								
3	.14620	.26316	.59064							
4	.06115	.11008	.24706	.58170						
5	.02568	.04622	.10373	.24424	.58013					
6	.01079	.01942	.04358	.10262	.24374	.57986				
7	.00453	.00816	.01831	.04312	.10242	.24365	.57981			
8	.00190	.00343	.00770	.01812	.04304	.10238	.24364	.57980		
9	.00080	.00144	.00323	.00761	.01808	.04302	.10238	.24363	.57980	
10	.00034	.00061	.00136	.00320	.00760	.01808	.04302	.10238	.24363	.57980

THE C VALUE IS .90

1	1.00000									
2	.34483	.65517								
3	.13495	.25641	.60864							
4	.05380	.10221	.24262	.60138						
5	.02151	.04086	.09699	.24042	.60022					
6	.00860	.01634	.03879	.09616	.24007	.60004				
7	.00344	.00654	.01552	.03846	.09603	.24001	.60001			
8	.00138	.00261	.00621	.01539	.03841	.09600	.24000	.60000		
9	.00055	.00105	.00248	.00615	.01536	.03840	.09600	.24000	.60000	
10	.00022	.00042	.00099	.00246	.00615	.01536	.03840	.09600	.24000	.60000

TABLE 1 (contd)

THE C VALUE IS 1.00

1	1.00000									
2	.33333	.66667								
3	.12500	.25000	.62500							
4	.04762	.09524	.23810	.61905						
5	.01818	.03636	.09091	.23636	.61818					
6	.00694	.01389	.03472	.09028	.23611	.61806				
7	.00265	.00531	.01326	.03448	.09019	.23607	.61804			
8	.00101	.00203	.00507	.01317	.03445	.09017	.23607	.61803		
9	.00039	.00077	.00193	.00503	.01316	.03444	.09017	.23607	.61803	
10	.00015	.00030	.00074	.00192	.00503	.01316	.03444	.09017	.23607	.61803

THE C VALUE IS 1.10

1	1.00000									
2	.32258	.67742								
3	.11614	.24390	.63995							
4	.04239	.08902	.23356	.63503						
5	.01550	.03255	.08540	.23218	.63437					
6	.00567	.01190	.03123	.08491	.23200	.63429				
7	.00207	.00435	.01142	.03105	.08485	.23197	.63427			
8	.00076	.00159	.00418	.01136	.03103	.08484	.23197	.63427		
9	.00028	.00058	.00153	.00415	.01135	.03103	.08484	.23197	.63427	
10	.00010	.00021	.00056	.00152	.00415	.01135	.03103	.08484	.23197	.63427

THE C VALUE IS 1.20

1	1.00000									
2	.31250	.68750								
3	.10823	.23810	.65368							
4	.03792	.08343	.22907	.64958						
5	.01331	.02928	.08039	.22796	.64907					
6	.00467	.01028	.02821	.08001	.22782	.64901				
7	.00164	.00361	.00990	.02808	.07996	.22780	.64900			
8	.00058	.00127	.00348	.00986	.02807	.07996	.22780	.64900		
9	.00020	.00044	.00122	.00346	.00985	.02807	.07996	.22780	.64900	
10	.00007	.00016	.00043	.00121	.00346	.00985	.02807	.07996	.22780	.64900

TABLE 1 (contd)

THE C VALUE IS 1.30

1	1.00000									
2	.30303	.69697								
3	.10111	.23256	.66633							
4	.03409	.07840	.22463	.66288						
5	.01150	.02646	.07582	.22373	.66249					
6	.00388	.00893	.02559	.07552	.22363	.66245				
7	.00131	.00302	.00864	.02549	.07549	.22361	.66244			
8	.00044	.00102	.00292	.00861	.02548	.07548	.22361	.66244		
9	.00015	.00034	.00098	.00290	.00860	.02548	.07548	.22361	.66244	
10	.00005	.00012	.00033	.00098	.00290	.00860	.02548	.07548	.22361	.66244

THE C VALUE IS 1.40

1	1.00000									
2	.29412	.70588								
3	.09470	.22727	.67803							
4	.03077	.07384	.22028	.67512						
5	.01000	.02401	.07163	.21954	.67481					
6	.00325	.00781	.02330	.07140	.21946	.67478				
7	.00106	.00254	.00758	.02322	.07138	.21946	.67477			
8	.00034	.00083	.00246	.00755	.02321	.07137	.21945	.67477		
9	.00011	.00027	.00080	.00246	.00755	.02321	.07137	.21945	.67477	
10	.00004	.00009	.00026	.00080	.00246	.00755	.02321	.07137	.21945	.67477

THE C VALUE IS 1.50

1	1.00000									
2	.28571	.71429								
3	.08889	.22222	.68889							
4	.02787	.06969	.21603	.68641						
5	.00875	.02187	.06780	.21542	.68617					
6	.00275	.00686	.02128	.06761	.21536	.68614				
7	.00086	.00215	.00668	.02122	.06759	.21535	.68614			
8	.00027	.00068	.00210	.00666	.02121	.06759	.21535	.68614		
9	.00008	.00021	.00066	.00209	.00666	.02121	.06759	.21535	.68614	
10	.00003	.00007	.00021	.00066	.00209	.00666	.02121	.06759	.21535	.68614

TABLE 1 (contd)

THE C VALUE IS 1.60

1	1.00000									
2	.27778	.72222								
3	.08361	.21739	.69900							
4	.02534	.06590	.21188	.69688						
5	.00769	.01999	.06427	.21138	.69668					
6	.00233	.00606	.01949	.06412	.21133	.69666				
7	.00071	.00184	.00591	.01945	.06410	.21132	.69666			
8	.00021	.00056	.00179	.00590	.01945	.06410	.21132	.69666		
9	.00007	.00017	.00054	.00179	.00590	.01944	.06410	.21132	.69666	
10	.00002	.00005	.00017	.00054	.00179	.00590	.01944	.06410	.21132	.69666

THE C VALUE IS 1.70

1	1.00000									
2	.27027	.72973								
3	.07880	.21277	.70843							
4	.02312	.06242	.20785	.70661						
5	.00679	.01832	.06101	.20742	.70645					
6	.00199	.00538	.01791	.06089	.20739	.70644				
7	.00058	.00158	.00526	.01788	.06088	.20738	.70644			
8	.00017	.00046	.00154	.00525	.01787	.06088	.20738	.70644		
9	.00005	.00014	.00045	.00154	.00525	.01787	.06088	.20738	.70644	
10	.00001	.00004	.00013	.00045	.00154	.00525	.01787	.06088	.20738	.70644

THE C VALUE IS 1.80

1	1.00000									
2	.26316	.73684								
3	.07440	.20833	.71726							
4	.02115	.05923	.20393	.71569						
5	.00602	.01685	.05800	.20357	.71556					
6	.00171	.00479	.01650	.05791	.20354	.71555				
7	.00049	.00136	.00469	.01647	.05790	.20354	.71555			
8	.00014	.00039	.00134	.00469	.01647	.05790	.20354	.71555		
9	.00004	.00011	.00038	.00133	.00468	.01647	.05790	.20354	.71555	
10	.00001	.00003	.00011	.00038	.00133	.00468	.01647	.05790	.20354	.71555

$$M_n'' = (\hat{\mu}_{n-1} + x_n) - y_n \quad (12)$$

This is the same as the traditional estimate of MUF except y_{n-1} is replaced by $\hat{\mu}_{n-1}$.

The variance of M_n'' is

$$\text{var } M_n'' = \sigma_y^2 + \sigma_x^2 + \sigma_3^2 \quad (13)$$

where $$\sigma_3^2 = (T' \, V^{-1} \, T)^{-1} \quad (14)$$

This compares with the variance of M_n', the traditional estimate.

$$\text{var } M_n' = 2\sigma_y^2 + \sigma_x^2 \quad (15)$$

A measure of the improvement in precision may be expressed as the ratio

$$R^2 = \frac{\text{var } M_n'}{\text{var } M_n''} \quad (16)$$

or, alternatively

$$R = \frac{\text{standard deviation } M_n'}{\text{standard deviation } M_n''} = \sqrt{\frac{\text{var } M_n'}{\text{var } M_n''}} \quad (17)$$

R is a function of $c = \sigma_x^2/\sigma_y^2$ and n, the number of months. For simplification, R is shown as a function of c only in Table 2 with infinite n. For all practical purposes, n becomes "infinite" very quickly, and so Table 2 is quite descriptive for any n. To illustrate, in the example used previously,

$$n = 3$$

$$\sigma_y^2 = 5$$

$$\sigma_x^2 = 1$$

$$\sigma_3^2 = 205/96$$

$$R = \sqrt{\frac{11}{8.1354}} = 1.163$$

TABLE 2. R Versus c for Infinite n

c	R
0.10	1.238
0.20	1.188
0.30	1.157
0.40	1.135
0.50	1.118
0.60	1.105
0.70	1.094
0.80	1.085
0.90	1.077
1.00	1.071
1.10	1.065
1.20	1.060
1.30	1.055
1.40	1.052
1.50	1.048
1.60	1.045
1.70	1.042
1.80	1.040

This compares with a value of 1.188 at Infinite n.

Note from Table 2 that the gain in precision can be large as c gets small, i.e., as the physical inventory uncertainty tends to dominate. This gain in precision can be equivalent to the information obtained by additional sampling and analytical effort.

It should be emphasized that although the foregoing discussion assumes that σ_y^2 and σ_x^2 are constant over time, i.e., that the variance of the physical inventory and the variance of the (inputs - outputs) are constant from month to month, these assumptions are made for simplicity in presentation and are not crucial to the argument. Similar results have been found for the situation in which σ_x^2 is not constant. Work is proceeding for the case in which σ_y^2 is not constant.

Some further results on MUF, and specifically the "minimum-variance" MUF, are given in the section of this report entitled "MUF as Criterion of Diversion." Before this discussion, however, another way is discussed briefly in which the variance of the MUF may be decreased.

Costwise Minimization of the Variance of MUF

In the previous section, a method was shown whereby different estimating procedures applied to a given set of data will reduce the variances of the estimates of the inventory and MUF. Any improvements that can be made in the quality of the data themselves will provide additional benefits. This section addresses itself to data improvements as opposed to "analysis of the data" improvements.

The two obvious ways in which data improvements may be accomplished are by using improved methods of making measurements and by making more measurements. Both steps cost money, so it is advantageous to know how many measurements to take at what places with what sampling and measurement devices. The particular combination of efforts which minimizes total cost, or alternatively, which maximizes information for a fixed cost, is the reasonable combination to employ.

In a simplified structure, assume that MUF is a linear combination of several (m) random variables, where the i^{th} one is denoted by $\bar{u}_i$ to show that it is the average of n_i measurements.

$$\text{MUF} = \sum_{i=1}^{m} \bar{u}_i \tag{18}$$

Assuming the $\bar{u}_i$ values are uncorrelated,

$$\text{var MUF} = \sum_{i=1}^{m} \text{var } u_i/n_i \tag{19}$$

Suppose that a total of C dollars are to be spent to obtain the measurements, and that it costs c_i dollars to obtain a single u_i measurement. Then it can be shown that n_i should be chosen directly proportional to the standard deviation of the measurement and inversely proportional to the square root of the cost.

$$n_i = k\sigma_i/\sqrt{c_i} \tag{20}$$

where k is chosen in such a manner that

$$\sum_{i=1}^{m} n_i c_i = C \tag{21}$$

To illustrate, suppose there are five measurement points with the following values for σ_i, c_i and C.

$\sigma_1 = 2$ $\quad c_1 = 4$
$\sigma_2 = 3$ $\quad c_2 = 4$
$\sigma_3 = 5$ $\quad c_3 = 9$
$\sigma_4 = 10$ $\quad c_4 = 1$
$\sigma_5 = 20$ $\quad c_5 = 1$

$C = 330$

Then,

$$n_1 = 2k/2 = k$$
$$n_2 = 3k/2$$
$$n_3 = 5k/3$$
$$n_4 = 10k/1 = 10k$$
$$n_5 = 20k/1 = 20k$$

From Equation 21:

$$k\left[4 + 6 + 15 + 10 + 20\right] = 330$$

$$k = 6$$

Therefore,

$n_1 = 6$ measurements on u_1
$n_2 = 9$ measurements on u_2
$n_3 = 10$ measurements on u_3
$n_4 = 60$ measurements on u_4
$n_5 = 120$ measurements on u_5

The variance of MUF in this case is, (by Equation 19)

$$\text{var MUF} = \frac{4}{6} + \frac{9}{9} + \frac{25}{10} + \frac{100}{60} + \frac{400}{120} = 9.17$$

Suppose we had chosen

$$n_1 = 20$$
$$n_2 = 20$$
$$n_3 = 5$$
$$n_4 = 50$$
$$n_5 = 75$$

which gives the same total cost. Then, the variance of MUF would be

$$\text{var MUF} = \frac{4}{20} + \frac{9}{20} + \frac{25}{5} + \frac{100}{50} + \frac{400}{75} = 12.98$$

This example is intended to illustrate the approach and is admittedly an oversimplification. However, the simplification is only one of computational ease and does not imply that such a cost minimization approach is not feasible in practice. It may be that, in a given situation, simple closed-form solutions would not be found. However, even in very complex situations, computer simulation may be used effectively to provide adequate answers.

This discussion on cost-minimization is very closely related to a statistical technique referred to as estimating components of variance. This refers to estimating the contribution to an overall variance that comes from different sources. An identification of the major contributing sources provides an indication of where the effort should be placed in reducing the overall variance.

Components of variance estimation plays a basic role in statistics. This is stated very well in a book by Anderson and Bancroft.[(2)] "However, this (components of variance) is actually the basis of all statistical concepts, because it deals with that particular aspect of data which requires statistical treatment - variability."

MUF as Criterion of Diversion

The preceding two sections were concerned with techniques for reducing the variance of MUF. The importance of this in safeguards is emphasized in this section.

To begin the discussion, some remarks are made on the statistical testing of hypotheses, since this is central to the discussion on the use of MUF as a criterion of diversion. It may be that in some situations MUF is not used formally as a statistic in hypothesis testing, yet inherent in the very motivation behind calculating MUF is the intent to use it as a measure of diversion whether or not there is loss of material by some other route. Thus, there is benefit in formalizing the process.

The general theory will be explained in terms appropriate to this discussion. A null hypothesis is set up which, on the basis of the data, will either be rejected in favor of an alternative hypothesis, or else it will not be rejected. Specifically, let the null hypothesis be

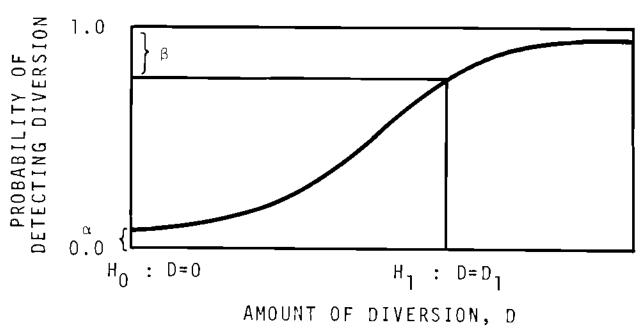
$$H_0 : D = 0 \tag{22}$$

that is, there is no diversion (or loss). Let the alternative hypothesis be that there is D_1 units of diversion over the period of time in question.

$$H_1 : D = D_1, \; D_1 > 0 \tag{23}$$

Data will be collected, D will be estimated from the data, and on the basis of this estimate, the null hypothesis will be tested. Denote the estimate by $\hat{D}$, which will be the MUF estimated by some procedure. Clearly, if $\hat{D}$ is "large," one would be inclined to reject the null hypothesis and conclude that a loss of materials has occurred. But how large is "large?" This depends upon the risks of making incorrect decisions which one is willing to take, and upon the variance of $\hat{D}$.

The figure sketched below is helpful in understanding the situation.



This is called a power curve. Its placement and shape depend on:

α = probability of asserting that there is diversion when indeed there is none (called a Type I error). α is called the significance level of the test.

$\beta(D_1)$ = probability of asserting there is no diversion when in fact D_1 units have been diverted (called a Type II error; β is function of D_1).

D_c = "critical value" for $\hat{D}$, i.e., the assertion is made that there is diversion when $\hat{D} > D_c$.

$\sigma^2_{\hat{D}}$ = variance of $\hat{D}$.

α is normally chosen quite small. β is, of course, a function of D, and is chosen small for that value of D which one would like to detect "most" of the time. D_c is a function of α, and $\sigma^2_{\hat{D}}$. This latter quantity is of crucial importance to the whole discussion

To illustrate, suppose

α = 0.05 (5% chance of asserting that there is diversion when in fact there is none.)

$\sigma^2_{\hat{D}} = 4 \text{ units}^2$ ($\sigma_{\hat{D}} = 2$ units)†

To find D_c, the critical value for $\hat{D}$, assume that $\hat{D}$ is normally distributed with average value (mean) zero (under the null hypothesis of no diversion) and standard deviation 2. Then the pertinent statistic to use is the ratio

$$\hat{D}/\sigma_{\hat{D}} = \hat{D}/2$$

Under the null hypothesis, this is normally distributed with mean zero and standard deviation one. (If $\sigma_{\hat{D}}$ is not known, but is estimated, Student's "t" distribution is used with the appropriate number of degrees of freedom.) Then, from a table of the normal distribution,

$$\text{Prob}\left[(\hat{D}/2) > 1.64\right] = 0.05 = \alpha$$

† *It is assumed in this example that this is either known or estimated with some degree of precision.*

from which Prob ($\hat{D} > 3.28$) = 0.05. Thus, D_c, the critical value, is 3.28 in this example. The "power curve" is then easily constructed. For example, find the point on the curve at $D_1 = 5$ units. The above process is reversed:

Prob (diversion is asserted)

$$= \text{Prob } (\hat{D} > 3.28)$$

$$= \text{Prob } \left(\frac{\hat{D}-5}{2} > \frac{3.28-5}{2}\right)$$

This last step is made to achieve a variable which has mean zero and standard deviation one, and hence will permit use of standard normal tables. Denoting this transformed variable by D*, the above expression becomes

$$\text{Prob } (D^* > -0.86)$$

which is 0.81 from the normal table. Thus,

$\beta(5) = 1 - 0.81 = 0.19$, or there is a

19% chance that a diversion of 5 units will not be detected. This gives one point on the power curve.

This example and discussion raises a few questions. What should be chosen as a value for α? (What percentage of the time should be spent trying to track down some material when in fact the evaluator is just observing a random fluctuation?) What about an observed large <u>negative</u> value of $\hat{D}$? What β values are reasonable for given D's? These questions involve extensive thought, and there are no easy answers. However, the point is that when formalizing the hypothesis testing approach, one is forced to face up to questions such as those, to recognize the risks involved, and to predetermine what action will be taken when, in fact, a significantly large MUF does occur. Furthermore, the power curve is extremely important in measuring the impact of a reduced variance in MUF. It tells precisely how important it is to reduce this variance, say, by a factor of two.

To illustrate this last point, suppose that in the previous example $\sigma_{\hat{D}}$ were 4 instead of 2. Then, the critical value D_c is changed.

$$\text{Prob}\left[(\hat{D}/4) > 1.64\right] = 0.05$$

implies D_c is now 6.56 rather than 3.28. Arguing intuitively, in the presence of a greater random error, more "evidence" of diversion is needed before it can be concluded that diversion has truly occurred. This will obviously reduce the power of the test. Specifically, again at $D_1 = 5$,

$$\text{Prob}\left[\hat{D} > 6.56\right]$$

$$= \text{Prob}\left[\frac{\hat{D}-5}{4} > \frac{6.56-5}{4}\right]$$

$$= \text{Prob}\left[D^* > 0.39\right] = 0.35$$

Thus, by doubling the standard deviation in this example, the power $(1-\beta)$ is reduced from 0.81 to 0.35, or, equivalently, the β error changes from 0.19 to 0.65. There is a 65% chance that a diversion of 5 units will not be detected. The power curve demonstrates the practical significance of decreasing the uncertainty in MUF by a given amount.

This is illustrated for the example in section (1) of this report where the variance of the usual estimate of MUF was compared with the variance of the minimum-variance estimator (see Equations 13 and 15). In the example treated there, for infinite n, and $c = \sigma_x^2/\sigma_y^2 = 0.2$, the standard deviation of the usual estimator M_n' was 18.8% larger than that of M_n'', the minimum variance estimator. The power curve data for this example are reproduced in Table 3. The power curve is a function of the sizes of σ_y^2 and σ_x^2 in addition to their ratio, and so the first column of Table 3 is given as a multiple of σ_y, the physical inventory standard deviation.

TABLE 3. Example of Power Curve

$$c = \sigma_x^2/\sigma_y^2 = 0.2$$

$$n = \infty$$

$$\alpha = 0.05$$

	Power for	
D_1/σ_y	M_n'	M_n''
0.6	0.11	0.12
1.2	0.20	0.24
1.8	0.33	0.41
2.4	0.48	0.60
3.0	0.64	0.77
3.6	0.77	0.89
4.2	0.88	0.95
4.8	0.94	0.98
5.3	0.97	1.00
5.9	0.99	1.00

Clearly, for "small" or "large" diversions, the reduction in standard deviation of 18.8% has little practical importance. It is in the "borderline" area where, in fact, it is most important that the data be as precise as possible, that the benefits become of practical importance.

Thus far, the MUF has only been considered as an indicator of when there might be diversion during a given month (or unit time period). There is significant additional information embedded in the data which can be extracted when one considers sequences of MUF's in addition to each individual one.

The two techniques to be discussed both depend on statistical independence between the successive MUF's. Such independence does not exist with the traditional estimate, M_n', since the beginning inventory for one month is the same as the ending inventory for the preceding month. This points out another advantage of the minimum-variance MUF. Somewhat surprisingly, perhaps, successive MUF's calculated by this procedure are statistically independent of one another. This is an important property.

This property has been proved rigorously, but will be demonstrated here with a simple example. Suppose n = 2 so that, from Equation 12

$$M_2'' = (\hat{\mu}_1 + x_2) - y_2$$

$\hat{\mu}_1$ comes from Equation 9 and specifically from Table 1. Say $c = \sigma_x^2/\sigma_y^2 = 0.5$. Then

$$\hat{\mu}_1 = 0.4\ z_0 + 0.6\ z_1$$

where $$z_0 = y_0 + x_1$$

$$z_1 = y_1$$

Therefore

$$\hat{\mu}_1 = 0.4\ y_0 + 0.4\ x_1 + 0.6\ z_1$$

the previous MUF, M_1'' is

$$M_1'' = \hat{\mu}_0 + x_1 - y_1$$

where $\hat{\mu}_0 = y_0$

So $M_1'' = y_0 + x_1 - y_1$

$$M_2'' = 0.4\ y_0 + 0.4\ x_1 + 0.6\ y_1$$

At first glance, these would appear to be strongly correlated, since both are functions of y_0, x_1, and y_1. However, in calculating the covariance between M_1'' and M_2'', remembering that y_0, x_1, and y_1 are themselves uncorrelated, this becomes

$$\text{Cov}\ (M_1'', M_2'') = 0.4\ \sigma^2_{y_0} + 0.4\ \sigma^2_{x_1} - 0.6\ \sigma^2_{y_1}$$

$$= 0.4\ \sigma^2_y + 0.4\ \sigma^2_x - 0.6\ \sigma^2_y$$

from (1) and (3)

$$= -0.2\ \sigma^2_y + 0.4\ \sigma^2_x$$

But, in this example, $c = \sigma^2_x/\sigma^2_y = 0.5$, or $\sigma^2_x = 0.5\ \sigma^2_y$, and the covariance becomes zero. This is a demonstration of the fact that the particular coefficients which define the linear combination of x's and y's to result in a MUF estimate are so chosen that statistical independence results.

This is a very important fact from a practical standpoint. It permits the use of some standard statistical procedures which would not be applicable in the presence of correlated MUF's. One such procedure is the mean square successive difference (MSSD) method of looking for trends in the data.

Consider the quantity

$$Q^2 = \sum_{i=1}^{n-1} \frac{\left(M_{i+1}'' - M_i''\right)^2}{2(n-1)} \qquad (24)$$

In the absence of trends in the data, Q^2 has an expected value equal to the variance of M_i''. (The first few terms in the sum will have unequal variances, but these will quickly approach a limiting value. One can perhaps remove the first few from the analysis.) Thus, Q may be compared with the usual variance estimator.

$$\text{var} = \sum_{i=1}^{n} \frac{\left(M_i'' - \bar{M}''\right)^2}{(n-1)} \tag{25}$$

If the ratio of Q^2 to Var is significantly different from one, significant nonrandomness exists in the data. Tables to test for the statistical significance of this ratio are contained in Reference 3.

In the absence of nonrandomness, Q^2 and Var are both what may be called "external" estimates of the variance of MUF, i.e., they are based on the amount of variation actually observed among the observed MUF's. When nonrandomness obtains, Q^2 would be the appropriate measure of this variance. It is very enlightening to compare these values with those obtained "internally," i.e., by propagating the errors of the quantities making up a given MUF.

If all the variance sources are accounted for and are estimated with sufficient precision, then the "external" and "internal" estimates of variance MUF should agree. Failure to agree points out that the situation has not been modeled properly, or that the input data are in error. It is very important to know this because otherwise the cost-minimization approach in the section devoted to costwise minimization of variance can lead to wrong answers as to where the effort should be expended. Even if formal cost minimization procedures are not followed, the evaluator is certainly guided by his assumed components of error structure in determining this effort allocation. It is important that he have the error structure properly appraised.

As a final note in this section, the evaluator should at the very least plot sequences of MUF's to see how they vary over time. An especially effective plot in detecting small, persistent biases and/or shifts in bias is the so-called Cusum plot which accumulates the sum of the MUF's and plots this sum as a function of time. Clearly, when in control, this plot should hover around zero.

In the next section, a different topic is considered with the treatment of biases in analyzing and interpreting data reviewed.

The Treatment of Biases in Uncertainty Statements

Suppose that a measurement, x, has expected value

$$E(x) = \theta + \delta \tag{26}$$

where θ is the "true" value of the measurement, and δ is the bias in the measurement. The precision of the measurement is obtained from

$$\sigma^2 = E[x - E(x)]^2 \tag{27}$$

The expected squared difference between the measurement and the true value is

$$E(x - \theta)^2 = E[x - E(x) + E(x)-\theta]^2$$

$$= E[x - E(x)]^2 + 2E\left\{[x - e(x)][E(x)-\theta]\right\} + E[E(x)-\theta]^2$$

The middle term is zero, and the last term is δ^2, from Equation (26). Thus,

$$E(x - \theta)^2 = \sigma^2 + \delta^2$$

$$= \text{"precision"}^2 + \text{"bias"}^2 \tag{28}$$

In this sense, bias^2 may be considered as a variance component. This concept is essential to an understanding of the nature of the uncertainty of a MUF. If desired, the bias^2 component may be thought of as a random component which persists over the time period in question. For example, it may represent a calibration error, and the instrument in question may be recalibrated at the beginning of a new time period, thus inducing (possibly) a new value for δ.

Precisions, expressed in relative terms, improve (i.e., random variances decrease) with increased numbers of measurements, while biases do not. For this reason, moderate biases for individual determinations can prove to be dominating in a campaign. When making cost-minimization studies, these biases certainly must be kept in mind.

Cumulative Error Model in Calibration Work

This subject will not be developed fully in this report, but it is important enough that mention must be made of it. In the previous section, it was pointed out that moderate biases will often dominate over a campaign, and calibration error was given as an example of such a bias. It is our concern that the calibration error assigned in a given situation may be an underestimate of the true error, and, in fact, it may well be a serious underestimate.

This concern arises because of the way in which calibrations are generally performed. In tank calibrations, for example, increments of liquid are added and successive readings are taken. These readings are not statistically independent of one another but rather, the random errors are cumulative in nature. The calibration data are often fit by least squares techniques, even though the assumption of an independent and equal error structure for each value of the dependent variable does not hold true. The net result is that the precision assigned the calibration may be grossly overstated. Since calibration precisions affect safeguards numbers as biases, the effect of the invalid analysis of the calibration data can be very misleading, and can lead to the assignment of uncertainty to a MUF that is grossly incorrect.

For a more complete discussion of this topic, see References 4, 5, and 8. The latter reference is specifically directed at tank calibrations.

Future and Extended Uses of Statistics in Safeguards

There are many exciting future possibilities for the applications of statistics and closely related techniques to the problems of safeguarding nuclear materials. Areas in which some attention has been given future applications are mentioned briefly in this closing section.

One of the aspects of future safeguards is the development of quantitative criteria for making objective judgements on the performance of safeguards activities. Are decisions to be made on the basis of absolute effectiveness alone? To what extent should cost-effectiveness play a role? What is the proper way to include the variables (amount diverted and time to detect) when evaluating them jointly? It is in answering questions such as these that statistics, in using the many subject-matter forms that the subject now takes, can be extremely helpful in setting forth in quantitative and unambiguous form the objectives of safeguards efforts. This will create an interacting relationship between the decision maker and the systems developers and will be helpful in directing the future thrust of activities into a commonly-agreed-upon quantitative channel.

Another exciting possibility is the application of the theory of games to safeguards problems. In the pure safeguards situation, the primary emphasis is in developing the optimum strategy for defeating an adversary. Adversary strategies, probabilities of attempts at diversion, and counter strategies should be formally studied. In view of the complexity of the situation, simple closed-form answers are probably not possible. Here again, though, with the availability of modern high speed computers, simulation studies which will provide adequate answers are entirely feasible.

Finally, consider the application of decision theory to safeguards. The general objective of decision theory, which is closely related to the theory of games, is to make the best decision on the basis of the available information. It accomplishes this by mapping an outcome space of possible results into a decision space. In the safeguards context, the kinds of statistical indicators of diversion would be mapped into the space of diversionary tactics, the intent being to infer the origin of the diversion. Such an approach, whether formalized or not, is really basic to good safeguards since it does little good to have evidence of diversion if one is at a loss to know what investigative and corrective action to take.

BIBLIOGRAPHY

1. *A. C. Aitken. "On Least-Squares and Linear Combinations of Observations," Proceedings Roy. Soc. of Edinburgh, vol. 55, pp. 42-8. 1935.*

2. *R. L. Anderson and T. A. Bancroft. Statistical Theory in Research, McGraw-Hill, New York, 1952. p. 374.*

3. *C. A. Bennett and N. L. Franklin. Statistical Analysis in Chemistry and the Chemical Industry, John Wiley and Sons, New York, N.Y., 1954. pp. 677-684. Table of significance for von Neumann ratio, p. 679.*

4. J. L. Jaech. "A Note on the Equivalence of Fitting a Straight Line Through Cumulative Data," J. American Statistical Association, vol. 59, pp. 863-866. 1964.

5. John Mandel. "Fitting a Straight Line to Certain Types of Cumulative Data," J. American Statistical Association, vol. 52, pp. 552-566. 1957.

6. K. B. Stewart. "Some Statistical Aspects of B-PID's and Ending Inventories," AEC and Contractor SS Materials Management Meeting, May 25-28, 1959, TID-7581. Available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia.

7. K. B. Stewart. B-PID and Inventory Estimates with Minimum Variance, HW-56536. Available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia, 1958.

8. C. G. Hough. Statistical Analysis--Accuracy of Volume Measurements in a Large Process Vessel, HW-62177. Available from Clearinghouse for Federal Scientific and Technical Information, Springfield, Virginia, 1959.

DISTRIBUTION

No. of Copies

OFFSITE

1 AEC Chicago Patent Group

G. H. Lee

25 AEC Division of International Affairs

W. A. Strauser

5 AEC Division of Nuclear Materials Safeguards

R. G. Page

186 AEC Division of Technical Information Extension

1 AEC New Brunswick Laboratory

C. J. Rodden

115 AEC Office of Safeguards and Materials Management

G. M. Inman (75)
S. C. T. McDowell (30)
J. M. Williams (10)

1 Allied Chemical Corporation
P.O. Box 35
Florham Park, New Jersey 07932

G. A. Huff

20 Argonne National Laboratory

M. A. Kanter

2 Arms Control and Disarmament Agency

L. Stieff

10 Brookhaven National Laboratory

H. Kouts

1 Idaho Nuclear Corporation

F. H. Tingey

1 Kerr McGee Corporation
Field Box 315
Crescent, Oklahoma 37028

R. E. Hubert

1 Lawrence Radiation Laboratory, Livermore

R. E. Batzel

1 Los Alamos Scientific Laboratory

Jane H. Hall

No. of Copies	
5	National Bureau of Standards G. Suzuki
1	Nuclear Audit Testing Co. Suite 312, Barr Bldg. 910 17th NW Washington, D. C. 20006 R. P. Wischow
2	Nuclear Fuel Services, Inc. Wheaton Plaza Office Building Wheaton, Maryland 20902 R. J. Jones
2	Stanford Research Institute E. M. Kinderman

ONSITE-HANFORD

1	AEC Chicago Patent Group R. K. Sharp
3	AEC Richland Operations Office H. R. Freitag (2) C. L. Robinson
1	Atlantic Richfield Company L. L. Zahn
3	Atlantic Richfield Hanford Company J. W. Jordan ARHCO Files (2)
3	Battelle Memorial Institute
1	Jersey Nuclear Company R. Nilson
2	RDT Assistant Director for Pacific Northwest Programs
74	Battelle-Northwest C. A. Bennett R. O. Budd A. W. DeMerschman K. Drumheller (3) J. L. Jaech R. A. Schneider (20) K. B. Stewart (30) L. F. Wirfs (10) Technical Publications (2) Technical Information Files (5)