109

W. E. DEMING

DESIGN OF A REPLICATED SAMPLE TO MEASURE THE CHANGE IN VALUE OF AN INVENTORY

Reprinted from the

Bulletin of the International Statistical Institute

VOL. XXXVIII: PART III

TOKYO 1961

DESIGN OF A REPLICATED SAMPLE TO MEASURE THE CHANGE IN VALUE OF AN INVENTORY

by

W. Edwards Deming
Consultant in Statistical Surveys, Washington

Advantages of a replicated design

As is well known, 3 standard errors measured above and below the result of the complete coverage of a specified frame will (with rare and recognizable exceptions) contain practically all the results that would arise from repetitions of a prescribed sampling procedure applied to this complete coverage. In practice we are able to estimate the standard error of a sampling procedure from a single sample, provided the sample is properly designed and is carried out in reasonable conformance to the specifications.

The procedure by which to estimate a standard error may be complex unless the design facilitates this calculation. Fortunately, it is possible by replication of a sample-design to reduce to simple arithmetic the calculation of the standard error, and to simplify greatly the study of a statistical control or audit for evaluation of the nonsampling errors. Replication of a sample-design is simple: one merely draws and processes the sample, not as one entire sample, but as 2 or more valid, independent, interpenetrating subsamples. The variance between the results of the individual subsamples provides a rapid and valid measure of the variance of repeated estimates obtained from the entire sample. Formulas appear later.

Control of nonsampling errors

It is important to know the range of sampling variation in a result. It is equally important to know the possible effect of the nonsampling errors, such as faulty use of instruments, faulty interviewing, measuring the wrong pieces, interviewing in the wrong area, failing to cover an area completely, reading wrong prices, copying figures incorrectly, nonresponse. Nonsampling errors are detectable and measurable by a careful recoverage of a subsample of the main sample, with comparison and explanation of any differences found. This careful recoverage is properly called a statistical control or audit.

If a test is destructive, then we can of course not re-test units that were in the main sample. We can, however, still carry out a careful investigation of a small interpenetrating sample of units that were not drawn into the original sample, and discover the types of blunders and blemishes that could afflict the main sample.

Incidentally, it is as necessary to learn about the nonsampling errors in a complete coverage as it is in a sample, and there is only one way to get this information—viz., to recover very carefully a sample of the complete coverage.

Nonsampling errors that arise from blemishes in procedure are not to be confused with built-in deficiencies in the method of test or in the questionnaire. These deficiencies arise from incomplete knowledge of the subject-matter, or from the necessity to use a second-class test of some kind, either to save expense, or for political or administrative reasons.

Application to a problem in accounting[1)]

I present here a simple example wherein the purpose of the study was to compare the dollar-value of an inventory at the end of the year with the dollar-value that the same inventory would have had at the beginning of the year. The difference is known in accounting circles as the LIFO adjustment. The company may decrease taxes, and acquire meaningful information for management, by having a reliable estimate of this figure. There are many angles to the problem, both substantive and statistical, but I shall omit the complexities.

The frame for the portion of the study that I shall report here was a list of about 20,000 parts in one plant of a large manufacturing concern. This list, according to the proper official of the company, if studied 100%, would constitute an entire solution to the problem of measuring the LIFO adjustment. The list was therefore a satisfactory frame.

The frame is by definition the list of items, levels, or conditions that we should wish to study for a complete solution to a problem. Every unit of material in a frame must bear a serial number, as random numbers will make the selections for the sample.

Certain items were of great value and were placed in a stratum for 100% treatment (*vide* Table 1). Other items of lesser value were divided into 3 other strata, to take different sampling ratios, based on Neyman allocation in an attempt to achieve the best precision possible for the manpower expended. The stratification and size of sampling plan were as follows:

a. 100% sample in Class 1

b Neyman allocation in the remaining 3 classes, by which

$$n_i = N_i \sigma_i / \bar{\sigma}_w \tag{1}$$

n_i is the size of sample in Class i; N_i the number of items in the frame in this class; σ_i the standard deviation of the LIFO adjustment in this class.

$$\bar{\sigma}_w = \frac{1}{N}(N_2\sigma_2 + N_3\sigma_3 + N_4\sigma_4) \tag{2}$$

There is no term for Class 1, because the sample in Class 1 is 100%. Class 1 contributes 0 variance. We define also at this time, for use later,

1) Portions of this paper from here on were presented at the meeting of the American Society for Quality Control in New York on 26 Feb. 1960. Details of the sample-design, in the notation of the formulas used here, appear in my book *Sample Design in Business Research* (Wiley, 1960), Chapter 15.

(3) $$\sigma_w^2=\frac{1}{N}(N_2\sigma_2^2+N_2\sigma_2^2+N_4\sigma_4^2)$$

c. 10 systematic subsamples

d. Standard error to aim at, \$50,000. This amount was 1/3d the 3-sigma limit of \$150,000, which the management decided was about the maximum allowable sampling tolerance that would serve their needs.

The formula for the variance of the LIFO adjustment is

(4) $$\text{Var}(X-Y)=\frac{1}{n}N^2(\bar{\sigma}_w)^2-N\sigma_w^2$$

$$=\frac{1}{n}(\sum_2^4 N_i\sigma_i)^2-\sum_2^4 N_i\sigma_i^2$$

$X-Y$ is the LIFO adjustment that originates in Classes 2, 3, 4. X is the estimate of the total value of the inventory today in these classes, and Y is the estimate for a year ago.

To find n, we put $\text{Var}(X-Y)=50{,}000^2$, and look in Table 1 for the numerical values of the sums. Then

(5) $$n=\frac{(\sum N_i\sigma_i)^2}{\text{Var}(X-Y)+\sum N_i\sigma_i^2}$$

$$=\frac{2.45\times10^{12}}{(25+9.8)10^8}=\frac{6\times10^4}{34.8}=1720$$

The sample-sizes in Classes 2, 3, and 4 will be proportional to $N_i\sigma_i$ in Table 1. That is,

(6) $$n_2=1.92n/2.45,\qquad n_3=.37n/2.45,\qquad n_4=.16n/2.45$$

wherein $n=1720$.

Table 1. *Basic date and basic calculations for the sample-design*

Class	N_i	σ_i	$N_i\sigma_i$	$N_i\sigma_i^2$	n_i	Z
1 \$10,000 and over	395	2 750	—	—	395	—
2 \$1,000–\$9,999	3 837	500	1.92×10^6	9.60×10^8	1350	30
3 \$100–\$999	7 467	50	.37	.19	260	280
4 0–\$99	8 022	20	.16	.03	110	700
Sum	1 9721	—	$N\bar{\sigma}_w=2.45$	$N\sigma_w^2=9.82$	2 115	—

The zoning interval for 10 subsamples will be $N_i/.1n_i$, rounded to some convenient figure. The next step is to make sure that every lot in every class has a serial number 1, 2, 3, etc.; then to read out from a table of random numbers 10 unduplicated random starts in each class between 1 and the zoning interval for that class (Table 2). We may then build up the sampling tables by adding the proper zoning interval to the random starts. The sampling tables so produced will draw 10 systematic subsamples in each class.

To process the sample, determine the number of items in each lot in the sample, the standard cost per item both now and a year ago, and by multiplication (extension) the dollar-value of each lot now (x) and a year ago (y). Sums of the estimated dollar-values by subsample in each class will furnish the figures needed in the equations for the estimates.

Table 2. *The random starts in each class*

Class	Zoning interval	1	2	3	4	5	6	7	8	9	10
2	30	15	30	25	16	04	06	22	24	03	23
3	280	119	108	040	098	168	106	058	070	076	224
4	700	124	486	661	056	229	181	575	689	473	630

The estimate from Subsample i for the total inventory is calculable by the formula

$$X^{(i)}=A+30x_2^{(i)}+280x_3^{(i)}+700x_4^{(i)} \tag{7}$$

where A is the present value of Class 1, $x_2^{(i)}$ is the present value of the inventory in Subsample i in Class 2, with similar definitions in Classes 3 and 4. The estimate X from the total sample will be the average of the individual 10 estimates. Or, we may write directly

$$X=A+3\,x_2+28\,x_3+70\,x_4 \tag{8}$$

where x_2, x_3, and x_4 are the present values of the inventories in Class 2, 3, and 4 in all 10 subsamples combined. There will be similar equations for the estimates of the inventory a year ago, written with Y in place of X, and with B in place of A.

The results are in Table 3. The averages at the bottom give

$X=\$21{,}062{,}000$, the estimate of the inventory now

$Y=\$19{,}977{,}000$, the estimate of the inventory a year ago

$X-Y=\$1{,}085{,}000$, the estimate of the LIFO adjustment

$f=X-Y=1.0544$, the estimate of the relative LIFO adjustment, or the LIFO index

For the standard error of $X-Y$ we compute from Table 3 the 10 values of $X^{(i)}-Y^{(i)}$ in the accompanying array. The maximum is 1288, the minimum is 850;

1.	1215	6.	1142
2.	1129	7.	1288
3.	862	8.	1210
4.	1005	9.	850
5.	1100	10.	1058

hence an estimate of the standard error of $X-Y$ is

$$\hat{\sigma}_{X-Y}=\frac{1}{10}(1288-850)10^3=\$43,800 \tag{9}$$

For the standard error of f we note from Table 3 that the maximum estimate of f is 1.0639 and that the minimum is 1.0418, whence

$$\hat{\sigma}_f=\frac{1}{10}(1.0639-1.0418)=.0022 \tag{10}$$

The standard errors calculated above do not take advantage of the finite multiplier in Class 2. As Class 2 dominates that variance, and as n_i/N_i in Class 2 is 1/3, a new calculation with retention of the finite correction would reduce the standard error by about 11%. Our corrected estimate of the standard error of $X-Y$ is thus about \$40,000, which is safely below the standard error aimed at, and in any event is good enough.

Table 3. *Results of the inventory now and year ago*
$A=\$7,714,000$ $B=\$7,226,000$

Sub-sample	$30\,x_2$	$280\,x_3$	$700\,x_4$	$X^{(i)}$	$30\,y_2$	$280\,y_3$	$700\,y_4$	$Y^{(i)}$	$f^{(i)}$
$i=1$	10068	3118	231	21131	9364	3020	306	19916	1.0610
2	9515	4319	440	21088	9141	3191	402	19959	1.0566
3	10044	2823	168	20746	9650	2796	212	19884	1.0434
4	10859	2614	228	21417	10365	2597	224	20412	1.0492
5	9769	2938	319	20741	9206	2894	315	19641	1.0560
6	9763	2959	549	20985	9405	2667	545	19843	1.0576
7	10317	3200	197	21428	9617	3103	194	20140	1.0639
8	9703	2808	326	20551	9045	2756	314	19341	1.0626
9	10150	3048	275	21186	9833	3014	264	20336	1·0418
10	9803	3566	270	21353	9384	3415	270	20295	1.0521
Average	9999	3049	300	21062	9501	2945	305	19977	1.0544

Add 000 to each figure

Results of a replicated design of an audit

The statistical control or audit is a complete and independent re-test and computation of a sample of the items drawn from the main sample. Study of the statistical control consists of a comparison item by item of the results obtained in the main sample with the results obtained in the statistical control, with attempt to find the explanation and to remove the causes of differences. A convenient design is to apportion the statistical control to the subsamples, so that one may calculate the ratio of the results obtained from the audit with the results calculated with the values reported in the main sample for the same items that fell into the audit.

One way to facilitate study of an audit is to show by subsample, and by meaningful subtotals (as for a plant, or for a division or other area), the ratio of the estimate calculated from the results of the audit to the estimate calculated with the values reported in the main sample for the same items. I present here a set of 10 ratios calculated for one of the characteristics studied in an audit of another survey. These ratios are not typical: no illustration is. I present them because it so happened here that all 10 ratios were greater than 1, which fact pointed definitely

Ratio of audit to main sample, by subsample

1.	1.1045	6.	1.0144
2.	1.0066	7.	1.0974
3.	1.0019	8.	1.0167
4.	1.0122	9.	1.0250
5.	1.0274	10.	1.0150

to the existence of persistent errors in the main sample that caused it to produce an underestimate. In fact, the lower limit of underestimate is 1.0019. An underestimate of 1% meant loss of $100,000 in taxes to the company that owned the inventory. Investigation and study of the errors in the main sample, and consideration of possible losses that might arise from wrong figures, led to a decision of management to reprice the main sample with improved definitions and with special training and care.

RÉSUMÉ

L'auteur resouvient de le point que l'écart type d'un result d'un echantionage est très important. De plus, il avance le point qu'il est maintenant possible est simple, par *replication* (itération; interpenetrating samples de Mahalanobis) de faire le plan d'un echantionage d'obtenir simultanément l'écart type d'aucun result avec le result le même. Replication est aussi utile pour le control. Le control statistique est le soigneu examen de nouveau d'un petit echantionage tiré au hasard d'echantionage principal. Les results, l'un d'echantionage principal, l'autre du control, fait l'évaluation des erreurs operationales, avec son écart type, par l'itération que est le nature de plan replicaté. L'auteur illustre les principes de la replication pour l'echantionage principal et pour le control, tous les deux, par les exemples tirés de l'évaluation d'un inventaire.