NATIONAL BUREAU OF STANDARDS REPORT

NBS Project 3011-60-0002

26 March 1952

NBS Report 1540

CONTROL AND MEASUREMENT OF EXPERIMENTAL ERROR

by

W. J. Youden

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

FOREWORD

The widespread adoption of quality control charts in industry has aroused interest in statistical techniques useful in research and development.

The paper covers some elementary aspects of experimental design and was prepared for a joint meeting of the Baltimore and Washington Sections of the American Society for Quality Control.

J. H. Curtiss
Chief, National Applied
Mathematics Laboratories

A. V. Astin
Acting Director
National Bureau of Standards
26 March 1952

CONTROL AND MEASUREMENT OF

EXPERIMENTAL ERROR

W. J. Youden*

- - - - - -

ABSTRACT

- - - - - -

This paper discusses in general terms the problem of determining experimental error and the modern development of methods of so ordering the schedule of taking the measurements as to achieve the maximum possible precision for comparisons among the measurements.

Experimenters in the physical sciences endeavor to control the conditions under which measurements are made. It is generally considered that disagreement among repeated measurements arises from some failure to specify and maintain these conditions. This emphasis upon the specification and maintenance of the experimental conditions is inevitable whenever the experimenter seeks to determine absolute magnitudes. An investigator wishing to calibrate a thermometer by setting up the ice point and boiling point of water must observe a great many precautions that are not necessary if the thermometer can be compared with a thermometer having known corrections. The existence of national laboratories charged with the establishment of physical standards and the testing of reference standards shows that the making of absolute measurements requires the greatest care.

It is not often explicitly pointed out that the availability of secondary reference standards contributes greatly to the accuracy of scientific measurements wherever they are made. Comparative measurements avoid many difficulties because it is usually not necessary to make them under precisely specified and attained conditions. This follows because, in the neighborhood of these conventionally specified conditions, all items under comparison are affected in the same way and to the same degree by departures from the specified conditions. For example, the correction to a thermometer at 25° C. may be found by comparison with a known standard using a bath that is in the vicinity of 25°.

---

*National Bureau of Standards, Washington 25, D. C.

The difference between the readings for the two thermometers may be considered constant in the region near 25° C. When both readings are taken by the same technique, by the same observer, and closely enough together so that the environment is virtually constant, all biases arising from these and possibly other sources, are the same for both readings and drop out when the difference is taken. This difference, when applied to the known value for the standard, yields the desired value for the object under measurement and with sensibly the same accuracy with which the standard itself is known.

The considerations just mentioned are so well recognized in simple situations, such as the one just discussed, that it is surprising that experimenters have not applied these principles more extensively in lengthy series of measurements. Indeed, it has only recently been generally understood that these same factors have frequently led experimenters to form unduly optimistic estimates of the real errors in their measurements. Whenever several objects are under measurement an estimate of the error of measurement is obtained by performing two or more measurements on each object. If these repeated measurements are always made in immediate succession the agreement obtained is enhanced by reason of the identity of the circumstances prevailing over the interval of time required for taking the readings. On the other hand, the entire series of measurements for the array of objects under examination usually extends over a much longer period of time. Often there is no assurance that the environment is maintained with the same constancy which held for the repeated measurements on the same object. In consequence, the error of measurement is obtained under relatively constant conditions but applied to the comparison of objects which were measured under much less constant conditions. It has long been the practice of the analytical chemist to run his duplicate analyses in parallel and carry over the apparent precision so obtained to the comparison of results obtained at different times. When the same material is analyzed at different times and results are found to be more divergent than the expected precision would have predicted, the usual recourse is to try to run down the responsible environmental condition and take steps to control or allow for this source of error. Much can be done in this way, but it is quite unlikely that complete success can be achieved.

The objective of the experimenter, when he seeks to obtain as precise comparisons as possible among a group of objects, may often be realized without the painful searching out and elimination of the various factors which creep into a program extending over a considerable period of time. Often a simple rearrangement of the order of performing the work is all that is required to nullify the effect of changing conditions in so far as comparisons among the objects are concerned. Consider the simple situation confronting

CONTROL AND MEASUREMENT OF

EXPERIMENTAL ERROR

W. J. Youden*

- - - - -

ABSTRACT

- - - - -

This paper discusses in general terms the problem of determining experimental error and the modern development of methods of so ordering the schedule of taking the measurements as to achieve the maximum possible precision for comparisons among the measurements.

Experimenters in the physical sciences endeavor to control the conditions under which measurements are made. It is generally considered that disagreement among repeated measurements arises from some failure to specify and maintain these conditions. This emphasis upon the specification and maintenance of the experimental conditions is inevitable whenever the experimenter seeks to determine absolute magnitudes. An investigator wishing to calibrate a thermometer by setting up the ice point and boiling point of water must observe a great many precautions that are not necessary if the thermometer can be compared with a thermometer having known corrections. The existence of national laboratories charged with the establishment of physical standards and the testing of reference standards shows that the making of absolute measurements requires the greatest care.

It is not often explicitly pointed out that the availability of secondary reference standards contributes greatly to the accuracy of scientific measurements wherever they are made. Comparative measurements avoid many difficulties because it is usually not necessary to make them under precisely specified and attained conditions. This follows because, in the neighborhood of these conventionally specified conditions, all items under comparison are affected in the same way and to the same degree by departures from the specified conditions. For example, the correction to a thermometer at 25° C. may be found by comparison with a known standard using a bath that is in the vicinity of 25°.

---

*National Bureau of Standards, Washington 25, D. C.

The difference between the readings for the two thermometers may be considered constant in the region near 25° C. When both readings are taken by the same technique, by the same observer, and closely enough together so that the environment is virtually constant, all biases arising from these and possibly other sources, are the same for both readings and drop out when the difference is taken. This difference, when applied to the known value for the standard, yields the desired value for the object under measurement and with sensibly the same accuracy with which the standard itself is known.

The considerations just mentioned are so well recognized in simple situations, such as the one just discussed, that it is surprising that experimenters have not applied these principles more extensively in lengthy series of measurements. Indeed, it has only recently been generally understood that these same factors have frequently led experimenters to form unduly optimistic estimates of the real errors in their measurements. Whenever several objects are under measurement an estimate of the error of measurement is obtained by performing two or more measurements on each object. If these repeated measurements are always made in immediate succession the agreement obtained is enhanced by reason of the identity of the circumstances prevailing over the interval of time required for taking the readings. On the other hand, the entire series of measurements for the array of objects under examination usually extends over a much longer period of time. Often there is no assurance that the environment is maintained with the same constancy which held for the repeated measurements on the same object. In consequence, the error of measurement is obtained under relatively constant conditions but applied to the comparison of objects which were measured under much less constant conditions. It has long been the practice of the analytical chemist to run his duplicate analyses in parallel and carry over the apparent precision so obtained to the comparison of results obtained at different times. When the same material is analyzed at different times and results are found to be more divergent than the expected precision would have predicted, the usual recourse is to try to run down the responsible environmental condition and take steps to control or allow for this source of error. Much can be done in this way, but it is quite unlikely that complete success can be achieved.

The objective of the experimenter, when he seeks to obtain as precise comparisons as possible among a group of objects, may often be realized without the painful searching out and elimination of the various factors which creep into a program extending over a considerable period of time. Often a simple rearrangement of the order of performing the work is all that is required to nullify the effect of changing conditions in so far as comparisons among the objects are concerned. Consider the simple situation confronting

Consequently any recurrent early morning condition which affects the measurement in a given way will, over the course of four days, have operated on each one of the four items. Obviously this represents an improvement in the equalizing of the conditions and should improve the precision of the comparisons of the averages. Again it is necessary to devise some means of obtaining from these data a measure of the experimental error which is in fact, in keeping with the real precision of the comparisons. Such a measure of the error may be obtained by noticing that if the average for the two B measurements on the first two days is subtracted from the average of the two A measurements on the same days the difference is free from between day and within day effects. The third and fourth days provide a second estimate of the difference between A and B also automatically balanced for between and within day effects. Now two estimates of the difference between A and B are in hand and, as before, the discrepancy between these two estimates is the basis for estimating the experimental error.

The appeal that the Latin Square arrangement makes for the equalizing of the environmental conditions is somewhat tempered by the fact that the number of repetitions of each measurement must keep pace with the number of objects under comparison. Thus 7 objects would require 7 measurements on each, and 13 objects would require 13 measurements on each. It is rather remarkable that, in many cases, it is possible to perform only a part of the Latin Square and still be able to obtain for the various objects numerical estimates which have been adjusted to compensate for the effects associated with particular days and particular times of day. These curtailed Latin Squares are typical examples of a group of arrangements known as balanced incomplete blocks. The following arrangement shows 7 objects each measured 3 times.

Balanced Incomplete Blocks

Order within the day	Day 1	Day 2	Day 3
1st	A	B	D
2nd	B	C	E
3rd	C	D	F
4th	D	E	G
5th	E	F	A
6th	F	G	B
7th	G	A	C

The Latin Square would require 7 days but four of these have been omitted with a saving of 4/7 of the work. At first it would seem that the seven objects are treated fairly only in the sense that each one is measured every day. They are not treated fairly in respect to the time of day. For example, only objects A,B and D are measured at the first hour of the day. The strenght of the arrangement resides in the following state of affairs:

Time period within the day	
1st	A may be compared with B and D
5th	A may be compared with E and F
7th	A may be compared with C and G

That is, A is found in three time periods which also bring up for measurement all six of the objects with which A must be compared. This property holds for all the letters and makes possible a simple arithmetical procedure for correcting the simple averages for any persistent biases associated with the different times of the day. When the designs become as complex as in the present example the formula for the estimate of the experimental error is not at all obvious. The formula has been derived mathematically and presents no difficulty in computation.

One further example of arrangements which have as their purpose the improvement of the precision of the experimental comparisons will be mentioned. The goal is to select subsets from a group of objects under measurement in such a way that, as nearly as possible, the good precision that applies to comparisons between the objects within a small subset can be legitimately extended to comparison among the whole group. Keeping the subsets small makes it easier to maintain constant conditions for the measurements in the set. In general the available arrangements call for three or more repeat measurements on every object. Recently a class of arrangements have been found which accomplish the desired ends and require only two measurements for each object.

One of these arrangements, called Linked Blocks, is shown.

Linked Blocks

Time of measurement	Day 1	Day 2	Day 3	Day 4	Day 5
Morning	A	B	C	D	E
	F	G	H	I	J
Afternoon	I	H	F	J	G
	E	A	D	B	C

The name Linked Blocks comes from the provision that every block (day in this case) is linked to every other block by one or another of the objects in the block. Thus the objects A,F,I and E scheduled for Day 1 are found in turn in Days 2,3,4 and 5. Once more a restriction has been placed on the order of the objects within the block so that a complete set of the 10 objects is measured during the morning hours and the second set measured in the afternoon.

The nature of the adjustment to be made to take care of experimental conditions peculiar to a given day may be seen by noticing that object A may be compared (Days 1 and 2) with 6 of the objects (F,I,E,B,G,H) run on the same days as A. The other three objects (C,D,J) are measured on days 3,4 and 5. But on these three days F,I,E,B,G and H are also measured so that C,D and J may be compared with these six and through them finally with A.

In summary, it has been the purpose of these remarks to point out the well known fact that measurements made closely together in time tend to agree better than measurements taken at widely separated times. This is the foundation for the application of planned arrangements for taking scientific measurements. These arrangements often make it unnecessary to strive to maintain comparable conditions for the entire duration of an extensive program of measurements. The arrangements lend themselves to the equalization of biases introduced by the uses of different operators or different instruments since these may easily replace the rôles taken by days and time of day in the illustrations given in the paper.

W. J. Youden
October 24, 1951
National Bureau of Standards