NATIONAL BUREAU OF STANDARDS REPORT

2135

MAKING ONE MEASUREMENT DO THE WORK OF TWO

by

W. J. Youden

and

W. S. Connor

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE
Charles Sawyer, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

Electricity. Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Applied Electricity. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

Heat and Power. Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels. Cryogenic Engineering.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Measurements. Infrared Spectroscopy. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Organic Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

Mineral Products. Porcelain and Pottery. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

Building Technology. Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Machine Development.

Electronics. Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

Radio Propagation. Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

Ordnance Development.
Electromechanical Ordnance.
Electronic Ordnance.
These three divisions are engaged in a broad program of research and development in advanced ordnance. Activities include basic and applied research, engineering, pilot production, field testing, and evaluation of a wide variety of ordnance matériel. Special skills and facilities of other NBS divisions also contribute to this program. The activity is sponsored by the Department of Defense.

Missile Development. Missile research and development: engineering, dynamics, intelligence, instrumentation, evaluation. Combustion in jet engines. These activities are sponsored by the Department of Defense.

- Office of Basic Instrumentation
- Office of Weights and Measures.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT **NBS REPORT**

3737-60-0002 30 December 1952 2135

MAKING ONE MEASUREMENT DO THE WORK OF TWO

by

W. J. Youden
and
W. S. Connor

FOREWORD

Frequently measurements are made under conditions which are either hard to specify precisely or difficult to hold constant for any considerable period. Corrections for drifts or shifts arising from these uncontrolled conditions are often based on measurements made upon control or standard samples periodically introduced in the work schedule. These standard samples make possible the adjustment of the measurements on the test samples at the price of diverting effort that might otherwise be spent on test samples. The standard samples may be dispensed with by picking out certain ones of the test samples for measurement at a later time. This paper presents some schedules for the selection of test samples for remeasurement. When the schedule possesses a balanced symmetry the arithmetical operations for adjusting the observations become simple and easy. Furthermore all the measurements made contribute information on the test sample.

This paper was prepared for presentation before the meeting of the American Institute of Chemical Engineers in Cleveland, Ohio.

J. H. Curtiss
Chief, National Applied
Mathematics Laboratories

A. V. Astin
Director
National Bureau of Standards

MAKING ONE MEASUREMENT DO THE WORK OF TWO*

by

W. J. Youden and W. S. Connor**

Engineers expend much effort to bring about improvements in the reliability of their measurements. As a result remarkable advances in the sensitivity and performance of instruments have been achieved. More attention has been given to the careful specification and control of external factors which may influence the results obtained with these better instruments. The objective has been to reduce the uncontrolled residual variation in measurements to the point where it would be of minor importance; and even to achieve a state of affairs where such chance variations in measurements could be altogether ignored. It turns out that it is an unending struggle to achieve this objective because engineers put ever increasing demands upon their measurements. In a competitive world small effects may have large economic consequences. Forthermore, scientific discoveries sometimes depend upon the detection of differences of small magnitude.

It is worthwhile to pause and consider what sort of success can be achieved by the control of disturbing external factors. It is first necessary to ascertain what factors are operating in this manner. Complete enumeration of the factors is not easy and the search often ends when further improvement in this direction appears to require an undue amount of work. When factors have been identified it is sometimes evident that the necessary control is a costly and tedious undertaking. One of the most common factors is temperature. In consequence there exist a great variety of devices for controlling the temperature. Alternatively, if the temperature is not controlled it is recorded and an appropriate adjustment made to the measurement. After all has been done that can be done has success been achieved? Almost invariably the measurements are still influenced by factors not specified or by factors imperfectly controlled or allowed for. That this is so is demonstrated by the universal experience that two measurements will show better agreement when they are made in the same laboratory than when each of two laboratories reports one measurement.

---

* Presented before the meeting of the American Institute of Chemical Engineers, Cleveland, Ohio, 9 December 1952.

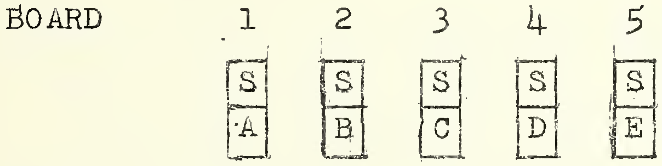
** National Bureau of Standards, Washington 25, D. C.

Of course there is no substitute for this tracking down of factors that disturb the measurements. It is highly necessary to get results that others can verify. But even today it comes as something of a shock to observe how much variation there is between laboratories when judged by the reproducibility of results within a laboratory. This has stimulated the undertaking of large scale and costly interlaboratory studies. Far too often the study only confirms what one already knew: that the laboratories disagree. The studies do not point the way to the elimination of the disagreement.

One more remark may be made regarding success, or rather the lack of success, in the reduction of variation to the point where it does not obscure the small differences that may be of great importance to the investigator. The rapid growth of interest in statistical techniques probably reflects a reluctant awareness that it will not be possible to eliminate variation in measurements. Statistical methods of interpretation appear to many as a way of living with this variation and making the best of it.

Most researches are carried out in a given laboratory. At this stage it is, as a rule, not necessary to be concerned about the between laboratory disagreements in absolute values. Usually the series of results obtained in a particular investigation are to be compared with one another. Relative precision is all that is required. Presumably a sister laboratory that is known to get high results would obtain a similar series all displaced on the high side and would draw the same conclusions from an intercomparison of its results. What may be overlooked is the existence of subdivisions within a laboratory that produce the same sort of disturbing influence that is found between laboratories. These subdivisions may be of many kinds such as different operators, machines, days, batches of reagent, etc. Even when it is recognized that such subdivisions contribute to the variation in the measurements nothing may be done about it. The reason is that the series of measurements is often too long to make it feasible to obtain them in one time period with one operator on one machine using one homogeneous lot of material. If the investigation involved only two or three or some other small number it might be convenient to hold many or even all of the above factors constant and secure for the intercomparison of this small set a very high precision. There is now available an extensive array of schemes that extend to large sets the high precision associated with small sets of measurements. Statisticians refer to such schemes as Experimental Designs. The remainder of the paper deals with some of the most recent of these schemes.

It is worth noting that, long before statisticians began to explore the subject of experimental design, scientists in certain circumstances made use of the basic idea involved. The situation that virtually forced an experimenter to use the idea back of the design of experiments usually arose when circumstances beyond the control of the experimenter provided him with an extremely limited quantity of homogeneous material. An additional supply would differ markedly from the first supply. Experimenters learned, how long ago no one knows, that the thing to do was run a control measurement on each of the several lots of material and express the test results in terms of the control. A typical example of this practice exists in exposure tests with paints. It is known that the performance of a paint is greatly influenced by the character of the surface to which it is applied. If the substratum is not uniform for all tests then differences will be ascribed to the paints that are, in reality, due to the substratum. A common test surface is wood, supplied by Nature, and subject to all of Nature's vagaries. Neither are there any obvious tests to satisfy the investigators that various pieces of wood are equivalent for the purposes of the exposure test. The experimenter falls back on a reasonable assumption: that is, that two *adjacent* pieces cut from the same board would be as much alike as possible. The device employed is simple. Take as many pieces of board as there are paints to compare and cut each board in two pieces keeping careful track of their identity. A control or standard paint is applied to one half piece from each board and the various test paints alloted to the remaining half, somewhat as shown.



Obviously the comparison of any test paint, say D, with *its* standard or control gives the best chance of a fair comparison. If the 10 half pieces had been mixed up and 5 painted with S and 5 with the other paints there would be no way to match up like halves. It is quite conceivable, that the variation shown among the 5 pieces painted with S in the scheme above, exceeds the actual differences among paints A through E. This does not matter so long as each paint is judged against its own standard. Clearly the success of this scheme depends upon it being possible to show that, if the two halves from one board are painted with S, the results agree much better than when the pieces come from different boards. Presumably this must have been shown or half the exposure facilities would not be expended on the standard paint.

The performance of each paint may be expressed as a per cent of the control from the same board and then these figures used to rate the paints. Alternatively, the difference between the results for each paint and its own standard could be taken and applied to the average of all five results obtained with the standard pieces. This might be more useful, as the average for the standard so obtained should give a fair idea of what might be the life of the paint on an 'average' board.

If five boards are used only one piece of board has received a given test paint and it might be well to use 10 or 15 boards instead of 5. Smaller differences between the paints will be detectable if the average of two or three test results can be used.

The employment of a control paint has overcome the diversity shown by different pieces of board but at a heavy price of using up half the test boards with the control paint. There is an alternate way of setting up a "control" that was used at least 35 years ago in another connection. The painting schedule is changed. First, all possible pairings of the letters S A B C D E are formed. There are 15 different pairs and these are assigned to 15 test boards as shown.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
S	S	S	S	S	A	A	A	A	B	B	B	C	C	D
A	B	C	D	E	B	C	D	E	C	D	E	D	E	E

There is an immediate consequence of this revised arrangement. The 15 test boards make available five pieces for each of the test paints and the control paint. The preceding scheme, using 15 boards, would have provided three pieces for each test paint and 15 control pieces. The three pieces for a given paint would have been compared with the three corresponding control pieces.

The problem is to set up a suitable "control" for the revised arrangement. Consider test paint A, tested on boards 1, 6, 7, 8, and 9, and paint B tested on boards 2, 6, 10, 11, and 12. Set up the following comparisons:

Board 1	$A - S = d_1$	Board 2	$B - S = d_2$
" 6	$A - B = d_6$	" 6	$B - A = d_6$
" 7	$A - C = d_7$	" 10	$B - C = d_{10}$
" 8	$A - D = d_8$	" 11	$B - D = d_{11}$
" 9	$A - E = d_9$	" 12	$B - E = d_{12}$
Total	$5A - (S+B+C+D+E) = \Sigma d_a$		$5B - (S+A+C+D+E) = \Sigma d_b$

Notice that A is compared with a "composite control" consisting of S, B, C, D, and E, while B is matched against a "composite control" consisting of S, A, C, D, and E. These "composite controls" are unfortunately not identical as they must be if they are to serve as a go between. It is, however, very easy to bring the two controls into agreement. If there had been a board, both ends of which had been painted with A, the difference between the ends should be zero. That is, A is equal to A. If to

$$5A - (S+B+C+D+E) = \Sigma d_a$$

there is added $$A - A = 0$$

The result is $$6A - \underbrace{(S+A+B+C+D+E)}_{\text{"composite control"}} = \Sigma d_a$$

the "composite control" is made to include all six paints.

A similar operation for paint B yields

$$6B - \underbrace{(S+A+B+C+D+E)}_{\text{"composite control"}} = \Sigma d_b$$

The same procedure is followed for paints S, C, D, and E, all of which give an expression with the same "composite control". The appropriate Σd, when divided by 6, gives for each paint the difference between the paint and the average of the "composite control". These average differences, which, incidentally, are based upon five test pieces instead of three, serve to rank the six paints. It was suggested above that, when S was the control, the average difference between a paint and its matched control could be added to or subtracted from the average of all the control pieces. Equally here, the obvious value to use for the "composite control" is the average of all 30 results, i.e., the five pieces available for all six paints.

If there is any hesitancy to adopt the above synthetic control as a reference value one may, if one desires, compare any paint A with the control S by taking $1/6(\Sigma d_a - \Sigma d_s)$. The "composite control", whatever its value, drops out of the picture. This average difference may, if desired, be applied to the average of the absolute values for S obtained from the five pieces painted with S. The various paints may thus still be expressed in terms of the performance of the standard. If the purpose is merely to rate the test paints among themselves then S may be omitted altogether. Ten boards would then suffice and provide four test pieces for each paint, instead of two when ten boards are used with a single control.

This example with the test boards illustrates a very general situation. Whenever there are unknown factors, or factors that are difficult to evaluate and control, recourse may be had to this device of picking some small area and assuming that these disturbing factors operate in the same way over the area picked. Test results obtained within this small area are presumed to be equally influenced by these unknown factors. The effects of these unknown factors "drop out" in the comparisons. The term 'area' is used in a generalized sense. The experimenter has the responsibility to define the limits of the 'area' within which comparisons may be made to particular advantage. The existence of such homogeneous areas is the indispensable condition for the profitable use of most experimental designs. A growing body of experimental evidence provides testimony that such homogeneous areas do exist in the majority of experimental programs.

Much work has been done in recent years to determine the most advantageous way of assigning the test items to the homogeneous area. All of the ways can be resolved into one or another manner of picking out the pairs to be formed. The use of a single control, assigned invariably to a part of each area is the most primitive and least efficient manner of making use of homogeneous areas. There are, however, some other considerations that are important to the experimenter. For example, ten paints, by the single control system, require only ten boards whereas the "composite control" system calls for 45 boards. But the ten boards with the single control make no provision for estimating the precision of the comparisons. If the set is repeated, using 20 boards in all, then the precision of the comparisons may be computed. This is still much less than 45. The 45 boards will of course give a much better experiment and also give a very good estimate of the precision, but many will object to such an increase in the size of the program.

The attractiveness of the "composite control" arrangement would be much enhanced if the requirement for a complete set of all possible pairs could be relaxed. Such is the case. For example, with ten paints, instead of 45 pairs 15, 25, or 30 may be selected. The 15 pairs cannot be any chance selection but must fulfill certain requirements of symmetry. The following 15 pairs link the letters together, either directly or indirectly.

AH
AI
AJ
BF
BG
BJ
CE
CG
CI
DE
DF
DH
EJ
FI
GH

Each letter, such as A, occurs three times and is paired with just three other letters, as H, I, and J. These three other letters, it turns out, are also paired with the remaining six letters, as JB, IC, HD, JE, IF, HG. To put it another way: let the 10 letters represent football teams. Pick a team. This team plays three other teams. These three other teams meet the six teams that did not play directly with the team first picked.

Short of each team playing all other nine teams, which would be a heavy schedule, the above scheme provides a satisfactory basis for rating the teams.

The same "composite control" can be set up for this selection of pairs but a little more algebraic maneuvering is required. The goal is to set up, by using differences obtained from the pairs, a comparison between a given letter and a "composite control" made up of a complete set of letters. This may be done for the letter B by adding up the following pairings:

$$\begin{array}{rr}
 & 3(B - F) \\
 & 3(B - G) \\
 & 3(B - J) \\
 & F - D \\
 & F - I \\
 & G - C \\
 & G - H \\
 & J - A \\
 & J - E \\
\text{and, of course,} & B - B \\
\hline
\end{array}$$

$$\text{Resulting in} \qquad 10B - \underbrace{(A+B+C+D+E+F+G+H+I+J)}_{\text{"composite control"}} = \Sigma d_b$$

The factor of 3 for the first three differences was determined by inspection. It gave the desired composition for the control. Notice that this really amounts to a system of weighting. In comparing a given letter with a "composite control" more weight should be given to the letters met directly than to the letters met indirectly through the good services of the intermediaries.

Now the total number of boards for the ten paints has been reduced to 15, and an estimate of precision is still available. Three test boards have been used for each paint. To get three boards for each paint with the single control system 30, or twice as many, boards would be required.

Picking a subset of 15 pairs from the complete set of 45 pairs leaves 30 pairs unused. These 30 pairs, too, as might be guessed, possess the necessary symmetry to make possible setting up a "composite control". Thirty boards would be needed and these would provide six pieces for each paint. Still another way of picking pairs which works for all numbers is to write half the letters at the head of a series of columns and the remaining letters at the left of a series of rows. The required pairs are given by the intersection of the rows and columns. Ten letters give the following arrangement.

	A	B	C	D	E
F	.	.	.	.	.
G	.	.	.	.	.
H	.	.	.	.	.
I	.	.	.	.	.
J	.	.	.	.	.

If there were nine letters the last row is omitted. The pairs are AF, AG, AH, AI, AJ, BF, BG, and so on. It is apparent that when A and B are compared F, G, H, I, J all serve as controls because A and B have each met these five letters. The pairs thus selected do make it possible to set up as before a "composite control" and use the test material previously expended on a single control for additional measurements on the items under test.

The discussion thus far has been carried on as if the "homogeneous area" was sufficient to accomodate two, and only two, experimental items. Very often three, four, five or more units constitute a natural area or block. A piece of wood might be sawn lengthwise and the resulting pieces then sawn in half. The four quarter pieces of the original piece of wood may be considered as very much alike. Single controls are often used in such cases. Indeed, the larger block cuts down the proportion of work expended on the control item. There is a very great gain also for the method of "composite controls", because, each squad of block sets up a considerable number of the required pairings. If the blocks contain four items then six pairs per block are immediately obtained. A very pretty example of how this may work out is afforded by the assignment of ten paints in sets of four to five pieces of wood cut into quarters.

1		2		3		4		5	
A	B	A	E	B	E	C	F	D	G
C	D	F	G	H	I	H	J	I	J

There are six pairs per square, 30 pairs in all. A little checking will satisfy the reader that these 30 pairs are exactly the 30 pairs left over when the 15 pairs previously discussed were picked from the complete set of 45 pairs which can be formed from the ten letters. It has already been stated that the residual 30 pairs are readily grouped so that the "composite control" can be set up for each of the letters. The five original pieces of wood have provided two test pieces for each letter, so that the number of test pieces (20) is even fewer than the 30 test pieces required to form just 15 pairs when the area accomodates only two paints.

The same 30 pairings can also be set up by using ten blocks each of which accomodates three items.

1	2	3	4	5	6	7	8	9	10
A	B	C	D	E	F	G	H	I	J
B	E	J	A	F	G	I	C	D	H
C	H	D	G	A	J	E	F	B	I

The intention at this time is not to provide a catalog of such arrangements. Rather, the desire is to show how a basic principle, long employed in scientific work, has undergone some elegant elaboration with an obvious gain in effectiveness. Less than twenty years ago Yates (5) wrote the initial theoretical paper in this field. Much of the important further development of the theory of these arrangements is due to Bose (1, 2). The earlier arrangements put little stress on holding to a minimum the number of test pieces, chiefly because it was thought that the major field of application would be in agriculture. There are two recent books on the design of experiments; one by Cochran and Cox (3), and one giving more of the mathematical theory by Kempthorne (4). A very elementary introduction is available in the author's book (6).

These arrangements are not an academic pastime. They have been found useful in the comparison of standard meter bars, Weston standard cells, temperature standards, and radio-activity standards. Such examples show that these arrangements are not limited to relatively crude measurements as may be the case in exposure tests. The utility of the device of a "composite control" is probably limited only by the ingenuity of the experimenters in recognizing areas of homogeneity in their operations and their taking good advantage of the increase in precision that such areas make possible.

LITERATURE CITED

1. Bose, R. C., and K. R. Nair. Partially balanced incomplete block designs. SANKHYA, 4, pp. 337-372, 1939.

2. Bose, R. C., and T. Shimamoto. Classification and analysis of partially balanced incomplete block designs with two associate classes. Jour. Amer. Stat. Assoc., 47, pp. 151-184, 1952.

3. Cochran, W. G. and G. M. Cox. Experimental Designs. John Wiley and Sons, Inc. New York, 1950.

4. Kempthorne, Oscar. The Design and Analysis of Experiments. John Wiley and Sons, Inc. New York, 1952.

5. Yates, F. A new method of arranging variety trials involving a large number of varieties. Jour. Agr. Sci., 26, pp. 424-455, 1936.

6. Youden, W. J. Statistical Methods for Chemists. John Wiley and Sons, Inc. New York, 1951.

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.


NBS