Instrumental Drift *

W. J. Youden
National Bureau of Standards, Washington 25, D.C.

THE developments in instrumentation and control devices in recent years are manifest in most laboratories. These advances have brought better measurements and have eased the labor of obtaining and recording them. Furthermore, improved instrumentation often has made it feasible to take more measurements. There is another consequence, one that many experimenters will consider an advantage, to be credited to better instruments. Better measurements, and more of them, have made it possible to interpret most data without recourse to statistical techniques.

Experimenters habitually try to select instruments and to control measurement procedures in order to get reproducible measurements that are good enough for their immediate purposes. These purposes generally fall into two classes: either the experimenter wants to keep the uncertainty in the result below some specified value, or else he wants to be able to distinguish between objects if these differ by some minimum amount in the measured property. If the worker succeeds in these respects, the interpretation of the data is simplified, because the uncertainties in the measurements can be, and usually are, ignored.

Apparently it is easier for many people to obtain elaborate and expensive control devices than it is to delve into the subject of the statistical design of experiments. Or they may be unaware that statistical design can bring the same kind of improvement in the data that comes from providing a uniform environment and will do this with little or no expense. The ideal measurement procedure should give results that the experimenter can accept without worrying about their reliability. The experimenter is then free for the task of studying the relationships that are involved in his scientific problem. In most cases the measurements are subject to random and other unknown sources of error that may either obscure relationships or even give the appearance of relationships when in fact there are none.

A good place to introduce statistics is in the preliminary trials an experimenter makes to assure himself that his apparatus and instruments are in a satisfactory operating condition. Consider the question of whether or not the instrument is subject to drift. Drift is usually explored by making a series of repeated measurements on the same object. Another question then plagues the worker. How can these repeated measurements be made independent of one another? How can the operator "forget" previous readings so that subsequent readings will not be influenced by earlier ones? These matters will be considered later.

Suppose the experimenter has made a series of measurements on the same object and has plotted the values as ordinates against the serial numbers of the measurements. A line drawn parallel to the x-axis with y equal to the average of all the readings will provide a visual test to detect trends in the sequence of readings. The experimenter would like to have the measurements indiscriminately scattered about the line and confined between two bracketing parallel lines as close as possible to the average line. If there is a pronounced trend, the visual test will reveal it. On the other hand, the experimenter may not be sure. Here, then, is the opportunity to use an objective statistical criterion to bolster his own judgment.

Table 1 lists measurements $y_1, y_2, \ldots, y_n$ in the order in which they were obtained. Two quantities, S^2 and D^2, may be computed from the observations in Table 1. The ratio of D^2 to S^2 should fall within predictable limits about the integer 2 if the results are free from trends. The quantity S^2 is the sum of the squares of the deviations of the plotted points from the horizontal line through the average. The formula

$$S^2 = \Sigma(y_i - \bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n},$$

where $\bar{y}$ is the average value, is a convenient way to obtain this sum of squares. Incidentally, the estimate of the standard deviation for these measurements is $\sqrt{S^2/(n-1)}$.

The quantity D^2 is the sum of the squares of the differences between successive measurements: $D^2 = \Sigma d_1^2$. It will be noted that the interval between two *successive* measurements gives only slight opportunity for the trend to operate. The d's are nearly what they would be if there were no trend at all. In contrast, the deviations between the individual y's and $\bar{y}$ are susceptible to the trend, and S^2 will be larger than it would otherwise be. The value of the ratio D^2/S^2 will then fall below 2.0.

It remains to set up some criterions for the allowable ratio of D^2/S^2. In any set of observations, the

Table 1. Successive differences between measurements.

Order of measurement	Measurement	Successive difference
1	y_1	
		$d_1 = y_1 - y_2$
2	y_2	
		$d_2 = y_2 - y_3$
3	y_3	
•		
•		
•		
$n-1$	y_{n-1}	
		$d_{n-1} = y_{n-1} - y_n$
n	y_n	

errors of measurement may, by chance, fall into suspicious configurations even when there is no trend. This is more likely to happen if the series is a short one so the limits for the ratio D^2/S^2 will depend on the number n of measurements.

Bennett (*1, 2*) has adapted some tables, published by Hart (*3*), that list limits of D^2/S^2, each of which will be exceeded on the average in 1 out of 20 sequences (or 1 in 100) for sequences that are not afflicted by any trend whatsoever. Thus, if a particular sequence does transgress these boundaries, it is usual to consider this as evidence of a trend rather than as a very improbable occurrence. Table 2 shows some specimen values of the limits taken from Bennett's table.

Sufficiently low values of the ratio D^2/S^2 are evidence of a trend. Overly large values of D^2/S^2 also indicate that the data depart from the expected random scatter. One way that the ratio may be inflated is by changing the zero setting or making other adjustments between successive readings. In general these adjustments will lead to a succession of large differences between successive readings and therefore will inflate D^2.

The following 20 determinations of the percentage of nickel were made on 20 successive segments of a rod of alloy by a spectrochemical procedure: 42.4, 40.8, 41.0, 41.8, 40.3, 40.8, 40.8, 39.6, 41.5, 41.5, 40.2, 40.4, 41.0, 42.2, 39.4, 41.0, 41.4, 40.6, 42.4, 40.8. It was important to know whether there was a trend along the rod. The computation for D^2/S^2 gave 31.32/12.99, or 2.41. The quotient is well within the listed limits for the ratio with n equal to 20, and there is no convincing evidence for a trend. The scatter of the data about the average line is shown in Fig. 1.

One obvious way to avoid the effect of remembering previous readings, referred to earlier, is to change the object being measured. At first thought, this would appear to make it impossible to detect any trend or drift in the measuring equipment. Certainly each reading will now depend on which object is measured and, if there is a drift, where the measurement is in the series. Such entanglement of effects can, however, be readily resolved if the objects are measured in an appropriate sequence. The devising of these sequences is one of the activities in the field of statistical design.

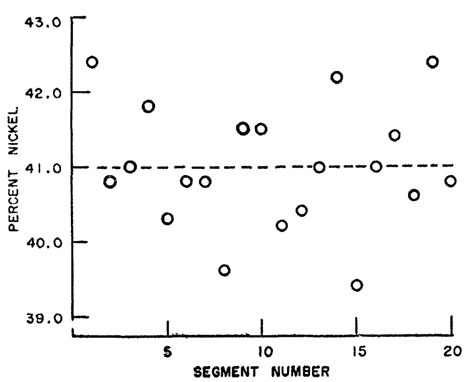

Fig. 1. Percentage of nickel in successive segments of an alloy rod.

Five objects, *A, B, C, D,* and *E,* may be available, and each could be measured four times in a sequence of 20 measurements. The problem is to set up a schedule that will still make it possible to detect the trend. Obviously nothing will be gained if four measurements are made on *A*, then four measurements on *B*, and so on. The memory difficulty is still present, and the values obtained for each object are inseparably combined with the drift, if any, in the instrument.

An alternative arrangement that begins to get into the problem is one that divides the sequence into four parts, each part containing all five objects. Thus,

BAEDC | *BEDAC* | *EDCBA* | *CAEDB*

The order of the objects within each part should be random. Now the average for the five objects in a particular block should be the same as the average in any other block except insofar as a trend happens to be present. In a coarse way, these averages, when plotted opposite 3, 8, 13, and 18, begin to reveal any instrument trend. The actual trend in any block that would be revealed by five ordinates is replaced by the average of these ordinates and centered in the middle of the block.

A modification of the afore-mentioned procedure will delineate the presumably rather smooth curve that corresponds to the true trend line during the measurements. The curve can be approximated by drawing short horizontal lines in a stepwise fashion along the curve. Each short horizontal line replaces the slant and slightly curved line in its vicinity. This horizontal line is located at a height equal to the average ordinate of the curve in the narrow band covered by the curved short line. If there were some way to determine the position of these short horizontal lines, the curve, or trend line, would stand revealed. It is better to have as many short lines as possible and to have them as short as possible. Ten short lines, each covering two measurements, afford a better approximation to the trend curve than four lines, each covering five measurements.

A difficulty then arises in the fact that the pair of objects used in any part will not be the same as the pair used in some other part of the curve. This would appear to make the averages for each pair useless for comparison, because the objects are different. If the pairs are formed in an appropriate manner, there is a simple procedure for comparing the parts, despite the fact that different objects occur in the different parts.

Five objects can be used to form 10 different pairs, each object appearing in four of the pairs.

Part	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
Object	*AB*	*DE*	*BC*	*EA*	*CD*	*EB*	*AC*	*BD*	*CE*	*DA*

These pairs break the trend curve into 10 parts. The order of the pairs is immaterial. The purpose is to determine the average values of the ordinates for each of the 10 parts, just as if all the measurements had been made on one object.

First use is made of the fact that the objects in any part, say *A* and *B* in part *a*, appear in six other parts. Thus, by using object *A*, the differences between part

Table 2. Limits for the ratio D^2/S^2.

No. in series n	1 in 20		1 in 100	
	Lower	Upper	Lower	Upper
5	0.82	3.18	0.54	3.46
10	1.06	2.94	0.75	3.25
15	1.21	2.79	0.92	3.08
20	1.30	2.70	1.04	2.96

Table 3. Determination of average value of ordinate *a*.

Using object	Difference between ordinates
A	$3(a-d)=x_1$
A	$3(a-g)=x_2$
A	$3(a-j)=x_3$
B	$3(a-c)=x_4$
B	$3(a-f)=x_5$
B	$3(a-h)=x_6$
C	$(c+g)-(e+i)=x_7$
D	$(h+j)-(b+e)=x_8$
E	$(d+f)-(b+i)=x_9$
Sum	$18a-2(b+c+d+\ldots+j)=\Sigma x_i$
Equivalently	$20a-2(a+b+c+\ldots+j)=\Sigma x_i$
And	$a-$ (average ordinate over all parts) $=\Sigma x_i/20$

Table 4. Comparison of actual and calculated instrument drift.

Reading number	Instrument drift	Object and its value		Observed reading	Part	Calculated deviation from mean drift	Calculated drift*
1	0	*A*	75	75	*a*	− 2.9	2.1
2	4	*B*	85	89			
3	7	*D*	55	62	*b*	3.7	8.7
4	10	*E*	45	55			
5	13	*B*	85	98	*c*	8.5	13.5
6	15	*C*	65	80			
7	17	*E*	45	62	*d*	12.0	17.0
8	17	*A*	75	92			
9	16	*C*	65	81	*e*	10.7	15.7
10	15	*D*	55	70			
11	13	*E*	45	58	*f*	6.0	11.0
12	10	*B*	85	95			
13	7	*A*	75	82	*g*	0.5	5.5
14	4	*C*	65	69			
15	0	*B*	85	85	*h*	− 6.2	− 1.2
16	− 3	*D*	55	52			
17	− 7	*C*	65	58	*i*	− 14.1	− 9.1
18	− 10	*E*	45	35			
19	− 13	*D*	55	42	*j*	− 18.2	− 13.2
20	− 15	*A*	75	60			

*** It is possible to determine this calculated drift only when the mean value of drift is known or determinable. In this example, the value 5.0 is used for the average of all ordinates from the curve of Fig. 2.**

a and parts *d, g,* and *j* can be estimated; by using object *B*, the differences between part *a* and parts *c, f,* and *h* can be estimated. This leaves parts *b, e,* and *i* to be considered. Notice that, by using object *C*, parts *c* and *g* can be compared with parts *e* and *i*; by using object *D*, parts *h* and *j* can be compared with parts *b* and *e*; and finally *E* gives parts *b* and *i* in terms of parts *d* and *f*. The lower-case letters are used to represent the average ordinates of the parts. These differences are shown in Table 3.

When a result located in a part, say *d*, is subtracted from a result in another part, say *a*, using the *same* object (*A*), the value of *A*, whatever it may be, drops out. The first six differences tabulated are multiplied by 3 to bring the sum to the form shown in Table 3. All letters other than *a* have the coefficient −2. The ordinate for part *a*, multiplied by 18, has twice the sum of the ordinates for all other parts subtracted from it, and Σx_i is the result. The difference is unchanged if twice ordinate *a*, or 2*a*, is added and subtracted. Division by 20 then gives the ordinate for *a* when added to the *average* ordinate over all parts.

A constructed example illustrates how well the scheme works. Suppose an instrument drifts as shown in Fig. 2. The curve shows the drift expressed in units of the terminal figure recorded. The instrument starts out and drifts so that after a time the readings are too high by about 17 units in the last place; then the trend reverses and drops until at the end readings are too low by about 15 units.

Imagine that five objects, *A, B, C, D,* and *E,* are available and that these, when measured, should give the values 75, 85, 65, 55, and 45, respectively. By reading from the drift curve and by assigning the values for the objects, one obtains a sequence of 20 readings, as shown in the fourth column of Table 4.

The only information that is assumed available for the statistical analysis is the column of observed readings together with the identities of the objects. It is assumed that the objects themselves do not change in value during the observations. The calculation of the average drift corresponding to part *a*, using the data of Table 4 and the equations of Table 3, is shown in Table 5.

Table 5. Calculation of average drift corresponding to part *a*.

Using object	Difference between ordinates
A	$3(75-92)=-51$
A	$3(75-82)=-21$
A	$3(75-60)=\ 45$
B	$3(89-98)=-27$
B	$3(89-95)=-18$
B	$3(89-85)=\ 12$
C	$(80+69)-(81+58)=\ 10$
D	$(52+42)-(62+70)=-38$
E	$(62+58)-(55+35)=\ 30$
Sum	$18a-2(b+c+d+\ldots+j)=-58$
Equivalently	$20a-2(a+b+c+\ldots+j)=-58$
And	$a-$ (average ordinate over all parts) $=-58/20$
Therefore calculated deviation from mean drift $=-2.9$	

Table 6. Determination of average value of ordinate b.

Using object	Differences between ordinates	
	In symbols	Using data of Table 4
D	$3(b-e)=x_1$	$3(62-70)=-24$
D	$3(b-h)=x_2$	$3(62-52)=\ 30$
D	$3(b-j)=x_3$	$3(62-42)=\ 60$
E	$3(b-d)=x_4$	$3(55-62)=-21$
E	$3(b-f)=x_5$	$3(55-58)=-\ 9$
E	$3(b-i)=x_6$	$3(55-35)=\ 60$
A	$(d+j)-(a+g)=x_7$	$(92+60)-(75+82)=-\ 5$
B	$(f+h)-(a+c)=x_8$	$(95+85)-(89+98)=-\ 7$
C	$(e+i)-(c+g)=x_9$	$(81+58)-(80+69)=-10$
Sum	$18b-2(a+c+d+\ldots+j)=\Sigma x_i$	$=\ 74$
Equivalently	$20b-2(a+b+c+\ldots+j)=\Sigma x_i$	$=\ 74$
And	$b-$ (average ordinate over all parts) $=\Sigma x_i/20$	$=\ 74/20$
		Calculated deviation from mean drift $=\ 3.7$

To calculate the ordinate for b, we must set up another series of differences similar to the series used for the calculation of a (Table 3). These new differences are given in Table 6.

In setting up the series, note that the objects appearing in part b are objects D and E. Therefore the first 3 differences (x_1, x_2. x_3) are obtained by taking the value of object D in part b and subtracting from it the respective values of object D in the other three parts in which it appears. The next three differences (x_4, x_5, x_6) are obtained using the values of E in similar fashion. The difference x_7 is obtained by taking the sum of the values of A in the two parts where A appears with D and E and subtracting the sum of the two values of A that appear with B and C.

Similar sets of differences must be set up for all 10 parts in order to calculate the ordinates. In each instance the sum of all nine equations will be of the form shown in the sets given for a and b and, therefore, will provide a check that the proper differences have been set up. As a further check, when all 10 values of deviation from mean drift have been calculated, their sum should equal 0.

The numerical procedure outlined in the preceding paragraphs, leads to the estimates shown in the last column of Table 4. These, unavoidably, apply to both observations in the pair to which they are attached. Inspection reveals that the calculated drift is in excellent agreement with the averages of the two drifts recorded for each pair in the second column. Furthermore, by the pattern of deviations from mean drift (Fig. 2), the drift of the instrument stands revealed through the overlay of the different objects measured.

The drift curve was plotted on the assumption that the 20 observations were taken at equal intervals of time. This restriction may be relaxed, provided that the two observations forming any pair are taken in close succession and provided that the times are recorded. The x-axis becomes a time scale and the average ordinate for each part is located at the average time for the two observations.

One of the merits of using different objects is the fact that the observer cannot anticipate the next reading and this assists in the attainment of objectivity in the readings. This objectivity is particularly desirable in the matter of estimating the precision of the readings. Precision is usually estimated from immediately successive readings on the same object, and it is difficult to avoid forming an optimistic appraisal of the precision. The present scheme also makes possible an estimate of the precision. The numerical details are available (4–7).

So far all the emphasis has been placed on the performance of the instrument. The instrument will be used to measure objects, and it is reasonable to inquire whether the 20 observed readings in Table 4 can also be used to estimate the values of the five objects.

The pairs were formed in all possible ways from the five objects. Consequently, any given object has been matched with the four others in some four of the 10 parts. And, most important, in any part made up of two readings it can be assumed that the instrument drift error is approximately the same for each reading. In taking the difference between the readings for two objects in a part, the instrument drift, whatever in may be at that time, virtually drops out. The difference obtained is just about what it would be if there were no drift at all.

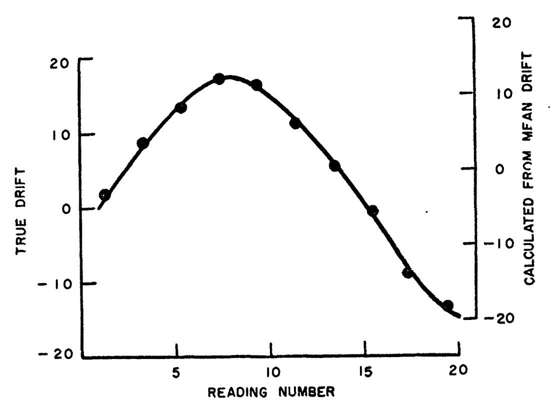

Fig. 2. Instrument drift in units of the terminal figure recorded.

Table 7. Comparison of correct and calculated values for the objects measured.

Object	Correct value	Calculated value	Calculated value less 5.0
A	75.0	79.4	74.4
B	85.0	90.4	85.4
C	65.0	70.6	65.6
D	55.0	59.0	54.0
E	45.0	50.6	45.6

The arithmetic for evaluating the objects is less involved than that used for the drift. To calculate an average for *A*, form the following differences:

Using part *a*, $A-B=-14$
Using part *g*, $A-C=13$
Using part *j*, $A-D=18$
Using part *d*, $A-E=30$
Sum, $4A-(B+C+D+E)=47;$
Equivalently, $5A-(A+B+C+D+E)=47;$
And A − average of all $=9.4;$
Average of all 20 readings $=70.0$; $A=70.0+9.4=79.4$.

Table 7 shows the calculated averages for the objects alongside the correct values. There is evidently a marked discrepancy between the correct and calculated values. The fourth column shows the calculated values all diminished by 5.0, and now the two sets show good agreement. The correction, 5.0, cannot be evaluated in any actual case. It is, in fact, the average value of the drift introduced by the instrument. There is no way, short of the good fortune in having one of the objects a known standard, to separate out the average drift from the average of all the objects.

In much experimental work the *difference* between test items is all that is important to establish. Where absolute values are required, a standard object is indispensable. If the absolute value of one object is known, all other objects can then be determined.

Many choices are available in the construction of the sequence used. The parts or blocks may be of any size. For example, seven objects can be arranged in seven triads, or 10 objects in 10 triads.

ABD | *BCE* | *CDF* | *DEG* | *EFA* | *FGB* | *GAC*
ABE | *HIJ* | *BHC* | *GEI* | *IDB* | *EFH* | *CJD* | *JGF* | *DAG* | *FCA*

The first of these sequences is an example of a class of designs called balanced incomplete blocks. The second sequence is a partially balanced incomplete block design. Various discussions of these designs are available (*4, 5, 7*).

There is a final important comment to make. Comparisons of objects can be made even with a drifting instrument. Even when the instrument has been operating satisfactorily, the experimenter perforce usually has had to *assume* that this state was maintained while making the critical measurements. Statistical design makes it possible to show that the instrument did stay in adjustment and, if not, to introduce appropriate adjustments.

References and Notes

* Based on a talk given at the Gordon Research Conference on Instrumentation in 1954.
1. C. A. Bennett, *Ind. Eng. Chem.* **43**, 2063 (1951).
2. ——— and N. L. Franklin, *Statistical Analysis in Chemistry and the Chemical Industry* (Wiley, New York, 1954), p. 677.
3. B. I. Hart, *Ann. Math. Statistics* **13**, 445 (1942).
4. R. C. Bose and T. Shimamoto, *J. Am. Statistical Assoc.* **47**, 151 (1952).
5. W. G. Cochran and G. M. Cox, *Experimental Designs* (Wiley, New York, 1950).
6. W. S. Connor and W. J. Youden, *J. Research Natl. Bur. Standards* **53**, R. P. 2532 (1954).
7. O. Kempthorne, *The Design and Analysis of Experiments* (Wiley, New York, 1952).

Phase Microscopy 1950–1954

Oscar W. Richards

American Optical Co., Research Department, Southbridge, Massachusetts

THIS paper is an analytic summary of some 200 publications that have appeared for the most part since the publication of *Phase Microscopy* (*1*). Phase microscopy is now so generally used that it often does not appear in the titles or abstracts of papers. This makes a complete listing of papers nearly impossible, and omissions are the result of failure to find the publications. Phase microscopy is useful for the study of colorless transparent or nearly colorless transparent materials containing detail composed of small differences in optical path (refractive index × thickness).

The Phase Microscope

The principles on which the phase microscope is based are illustrated in Figs. 1, 2, and 3. On the left in Fig. 1 is shown light wave *A′* passing through a transparent object *C* and slowed down with respect to light wave *A*, which did not pass through the transparent object. Accordingly, light wave *A′* is out of phase with light wave *A*. However, both the human eye and photographic plates are insensitive to phase differences, and as a result the image can scarcely be seen or photographed. Light wave *A″* passing through an absorbing medium *E* is reduced in amplitude (dis-