

New Experimental Designs for Paired Observations

W. J. Youden and William S. Connor

There are many experimental situations in which observations naturally occur in pairs. From v quantities to be compared, $v(v-1)/2$ distinct pairs can be formed, so that for even moderately large v , the accompanying number of observations may be prohibitively large. Or, even if the work can be done, the precision of the estimates of the effects of the quantities and the experimental error may be greater than is needed. In either event, a subset of the pairs should be used. If the arithmetic is to be manageable, this subset must be chosen with care. Such a subset is described and its analysis is given in detail.

1. Introduction

In many experimental situations only two quantities can be observed at a time under comparable conditions, so that observations on a set of quantities naturally divide into pairs. If each quantity is compared with every other quantity of the set, the statistical analysis of the observations is simple. However, if the number of quantities is large, the amount of work required to make observations on all of the pairs may be prohibitive. Even if the work can be done, the results of the experiment may be more precise than is needed. In either event, it is desirable to observe only part of the pairs. The selection of this subset of pairs must be made with care if the statistical analysis is to remain relatively simple.

An example will help to clarify the problem. Suppose that 36 quantities are to be studied, two at a time. From them there are 630 distinct pairs that can be formed. In this paper consideration will be given to particularly attractive subsets called two-group arrangements, which require from 324 down to 35 of these pairs.

To illustrate the statistical analysis of such an arrangement, a typical subset for the case of eight quantities is analyzed, using data on thermometer calibrations.

2. The Two-Group Arrangement

It is common scientific practice to compare new objects under study with one or more standards. An important function of the National Bureau of Standards is the calibration of thermometers, meter bars, and other devices for industrial and other uses. These calibrations are made by comparing the new objects with established standards.

Perhaps the most usual situation is that in which there is a single standard, which may be designated by S. If there are six new objects to be calibrated, then a common practice is to pair each new object with S. Thus, denoting the new objects by numbers, the pairs are as follows:

S1	S3	S5
S2	S4	S6

This traditional experimental procedure suggests the two-group arrangement, which consists of dividing the v objects under study into two groups of m and n objects, respectively, ($v=m+n$), and of pairing every object from one group with every object from the other group.¹ No other pairs are formed.

In the situation just considered the standard is the only object in one group so that $m=1$ and $n=6$. This arrangement provides information about the standard, the new objects, and the pairs, but gives no information about the experimental error. To obtain such information it is natural to run each pair again, so that there are 24 observations altogether.

Among these observations the standard occurs 12 times and the new objects twice each. This lays heavy emphasis on comparisons between the new objects and the standard but less emphasis on comparisons among the new objects. Thus if σ is the true standard deviation of an observation, then the standard deviation of a comparison of the first kind is σ and of the second kind is 1.4σ .

Another two-group arrangement results from assigning the standard and new objects 1 and 2 to one group and new objects 3, 4, 5, and 6 to the other. In this case $m=3$, $n=4$ and the pairs are the following.

S3	S4	S5	S6
13	14	15	16
23	24	25	26

There are 24 observations and as much information about the experimental error as in the preceding arrangement. The standard is put on the same footing as the new objects because it is observed 4 times and the new objects either 3 or 4 times. This is reflected among the comparisons, for the standard deviation of the comparison between any two objects in the first group is σ , in the second group is 1.15σ , and between an object in the first group and one in the second is σ . Thus there is no loss in the precision of comparisons that involve the standard, and there is a substantial gain in the precision of other comparisons.

¹ When $m=n$, the arrangement has been called Group Divisible. See, R. C. Bose and T. Shimamoto, Classification and analysis of partially balanced incomplete block designs with two associate classes, *J. Am. Statistical Assoc.* **47** 151 (1952).

The problem posed in the introduction can now be resolved in many ways. If the 36 quantities are divided into two groups of 18 each, 324 pairs will be formed. At the other extreme is the division into 1 and 35, which results in only 35 pairs.

3. Application to Thermometer Calibration

The authors asked the Thermometry Section of the National Bureau of Standards to intercompare eight thermometers, using the two-group arrangement. The usual practice of the section is to read the thermometers in sequence in a bath with slowly rising temperature and then to read them in reverse order. This device effectively compensates for changes in the bath temperature, provided that the temperature changes at a constant rate. *The effectiveness of the two-group arrangement, however, does not depend on a constant rate of change in temperature.*

The thermometers were partly immersed in a bath of distilled water, and were read through a telescope mounted a short distance away. The temperature of the bath was at approximately 40° C at the start of the readings, but rose gradually throughout the experiment. There were short pauses of irregular length between pairs of readings.

The eight thermometers were divided into 2 groups of 4 each, containing thermometers 1, 2, 3, and 4 and 5, 6, 7, and 8, respectively. The readings are given in table 1 in the order in which they were obtained.²

The computations can be simplified by subtracting some convenient number from each observation. Accordingly, all subsequent calculations are based on the observations in table 1 after subtracting 40 from each of them.

² The thermometers were randomized within the pairs and the pairs within the runs.

The mathematical model underlying the statistical analysis is based on the following considerations. Let M be a reference temperature in the range of temperatures of the bath during the experiment. At the time of measurement of the j th pair of thermometers, the temperature of the bath will be $M+p_j$, where p_j is defined by this condition.

Next, suppose that the i th thermometer belongs to the j th pair, and let x_{ij} denote the observed temperature for this thermometer when the j th pair is read. Then the difference between the observed temperature x_{ij} and the true bath temperature $M+p_j$ will consist of two parts: a systematic error t_i , peculiar to the i th thermometer, and a random reading error e_{ij} , i. e.,

$$x_{ij} - (M + p_j) = t_i + e_{ij}$$

or

$$x_{ij} = M + t_i + p_j + e_{ij}.$$

By imposing the restrictions $\sum_{i=1}^p t_i = \sum_{j=1}^{mn} p_j = 0$, M is uniquely defined.

The constants M , t_i , and p_j and the error e_{ij} are unknown but can be estimated from the data. It is assumed that the errors associated with different readings are independent and come from the same population of errors. This population is assumed to have mean zero and standard deviation σ , which may or may not be known.

The following calculations will show how to estimate the constants and the standard deviation.³ Estimates of the t 's are of especial interest, since they may be used to calibrate a new thermometer in terms of a standard. Estimates are denoted by

³ Derivations of formulas are given in the appendix.

TABLE 1. Temperature readings in order of time

Run											
1			2			3			4		
Pair	Ther-mometer	Reading									
		°C			°C			°C			°C
1	{ 1 7	40.00 39.99	5	{ 3 8	40.18 40.18	9	{ 2 6	40.23 40.22	13	{ 6 3	40.26 40.28
2	{ 5 3	40.08 40.13	6	{ 7 2	40.07 40.19	10	{ 8 4	40.24 40.15	14	{ 7 4	40.15 40.20
3	{ 8 2	40.15 40.17	7	{ 1 6	40.10 40.18	11	{ 7 3	40.12 40.20	15	{ 5 2	40.27 40.30
4	{ 6 4	40.13 40.05	8	{ 5 4	40.17 40.13	12	{ 5 1	40.23 40.16	16	{ 1 8	40.21 40.31

carets. For example, \hat{t}_i is the estimate of t_i . To analyze the coded data it is convenient to compute an auxiliary quantity, D , for each thermometer. Thus D_i , the D for the i th thermometer, is computed as follows. For each pair that contains the i th thermometer the difference between the reading for the i th thermometer and the reading for the other thermometer of the pair is computed. The sum of these differences is D_i . For example,

$$D_1 = [0 - (-.01)] + (.10 - .18) + (.16 - .23) + (.21 - .31) = -.24.$$

Let the group that contains m thermometers be called group 1, and the group that contains n thermometers be called group 2. Let the sum of the D 's for the thermometers in group 1 be denoted by S_1 , and in group 2 by S_2 . Then the D 's may be used to estimate the correction for the i th thermometer by the following formulas:⁴

$$\hat{t}_i = (vD_i - S_1)/vn$$

if i is in group 1, and

⁴ It sometimes happens that the temperatures or other quantities are not observed directly, but instead the differences between the quantities in the pairs are recorded. Although in this case M and the p 's cannot be estimated, the t 's still are estimable by these formulas.

$$\hat{t}_i = (vD_i - S_2)/vm$$

if i is in group 2. For example, for the first thermometer

$$\hat{t}_1 = (8D_1 - S_1)/32 = (-1.92 + .07)/32 = -.05781.$$

If σ is unknown from past experience, it may be calculated from the data. This calculation is made quite simply by working with the differences between the readings within a pair. Let the difference without regard to sign for the j th pair be designated by d_j . Then σ is estimated from the formula⁵

$$2(mn - m - n + 1)\hat{\sigma}^2 = \sum_{j=1}^{mn} d_j^2 - \sum_{i=1}^v \hat{t}_i D_i.$$

The computations may be systematized by use of table 2, in which the estimates of the t 's and $\sum_{i=1}^v \hat{t}_i D_i$ are found.

⁵ When just the differences are observed, it is convenient to do the analysis in terms of the standard deviation of the differences, which may conveniently be denoted by σ_d . This formula and others below apply in this case, too, provided σ is replaced by $\sigma_d/2$.

Table 2. Calculation of the thermometer effects

		Group 2 (n) thermometer				Calculations					
		5	6	7	8	Σ	D	$8D$	$32\hat{t}$	\hat{t}	$D\hat{t}$
Group 1 (m) thermometer	1	.23	.16	.10	.21	.47	-.24	-1.92	-1.85	-.05781	.01388
	2	.27	.30	.23	.17	.89	.18	1.44	1.51	.04719	.00849
	3	.08	.13	.28	.18	.79	.15	1.20	1.27	.03969	.00595
	4	.17	.13	.05	.15	.53	-.16	-1.28	-1.21	-.03781	.00605
Calculations	Σ	.75	.72	.59	.71	5.43					
	D	.03	.13	-.26	.17		-.07				
	$8D$.24	1.04	-2.08	1.36			-.56			
	$32\hat{t}$.17	.97	-2.15	1.29				-.28		
	\hat{t}	.00531	.03031	-.06719	.04031					-.00874	
	$D\hat{t}$.00016	.00394	.01747	.00685						.03437
										.02842	

The coded readings are entered in the upper left-hand part of the table, where every cell corresponds to some pair. For example, the first pair is put into the cell in row 1 and column 7, with the reading for thermometer 1 recorded in the upper right-hand corner and for thermometer 7 in the lower left-hand corner. By so recording the readings, each row and column is divided into subrows and subcolumns.

The remaining rows and columns are for calculations, which it is believed are self-evident. In general, row $8D$ is replaced by vD and $32\hat{t}$ by $(vn)\hat{t}$. Likewise, column $8D$ is replaced by vD and $32\hat{t}$ by $(vn)\hat{t}$. Several checks are available: (1) the sum of the entries in row Σ must equal the sum of the entries in column Σ , and (2) the sums of the other corresponding rows and columns, except the last, must be of different sign but of the same absolute value. In the table these quantities appear along the diagonal.

The standard deviation is estimated from the formula given above. The differences d may easily be calculated from table 1, and $\sum_{i=1}^v t_i D_i$ from table 2. The differences and calculations on them are given in table 3.

TABLE 3. Calculation of the standard deviation

Pair (j)	d_j	d_j^2	Pair (j)	d_j	d_j^2
1.....	0.01	0.0001	9.....	0.01	0.0001
2.....	.05	.0025	10.....	.09	.0081
3.....	.02	.0004	11.....	.08	.0064
4.....	.08	.0064	12.....	.07	.0049
5.....	.00	.0000	13.....	.02	.0004
6.....	.12	.0144	14.....	.05	.0025
7.....	.08	.0064	15.....	.03	.0009
8.....	.04	.0016	16.....	.10	.0100

$\sum_{j=1}^{16} d_j^2 = .0651,$	$\sum_{i=1}^8 \hat{t}_i D_i = .03437 + .02842 = .0628,$
$18\hat{\sigma}^2 = .0651 - .0628 = .0023, \hat{\sigma} = .0114.$	

Two thermometers can be compared by finding the difference between their estimated effects. To judge the significance of such a difference, it is desirable to know the standard deviation of the difference. If i and i' both are in group 1, then the square of the standard deviation of the difference is

$$\sigma_{\hat{t}_i - \hat{t}_{i'}}^2 = 4\sigma^2/n;$$

if both are in group 2, then

$$\sigma_{\hat{t}_i - \hat{t}_{i'}}^2 = 4\sigma^2/m;$$

and if i is in group 1 but i' is in group 2, then

$$\sigma_{\hat{t}_i - \hat{t}_{i'}}^2 = 2(v-1)\sigma^2/mn;$$

If σ^2 is not known, then its estimate is used.

As an example, consider thermometers 1 and 2. The appropriate formula is the first one above, so that

$$\sigma_{\hat{t}_1 - \hat{t}_2}^2 = .0114.$$

Just as it has been possible to intercompare the thermometers even though in some cases a particular pair of thermometers were never at the same temperature, so also it is possible to determine the relative temperatures of the bath when each of the mn pairs of thermometers were read even though the temperatures were read with different thermometers with unknown corrections. It may sometimes be important to ascertain the character of the drift or changes taking place in the experimental system. In the example given, matters were arranged so that there was an approximately linear drift upward in the bath temperature. Table 4 reflects this condition, the values being computed as is indicated below.

TABLE 4. Average temperatures of the pairs referred to $40^\circ C$

Pair	Uncorrected average	Corrected average	Pair	Uncorrected average	Corrected average
	$^\circ C$	$^\circ C$		$^\circ C$	$^\circ C$
1.....	-0.005	0.058	9.....	0.225	0.186
2.....	.105	.082	10.....	.195	.194
3.....	.160	.116	11.....	.160	.174
4.....	.090	.094	12.....	.195	.221
5.....	.180	.140	13.....	.270	.235
6.....	.130	.140	14.....	.175	.228
7.....	.140	.154	15.....	.285	.259
8.....	.150	.166	16.....	.260	.269

The averages after correction for thermometers exhibit the upward trend much more clearly than do the crude, uncorrected averages.

The uncorrected averages for the j th pair is simply the arithmetic average of the two readings in the pair. The corrected average is the uncorrected average adjusted for the systematic errors of the thermometers that occur in the j th pair. In symbols it is $\hat{M} + \hat{p}_j$.

The estimate of M is

$$2mn\hat{M} = \sum_{i=1}^v \sum_{j=1}^{mn} x_{ij} + (m-n) \sum_{i=1}^m \hat{t}_i,$$

which, in the case at hand, reduces to

$$\hat{M} = \left(\sum_{i=1}^8 \sum_{j=1}^{16} x_{ij} \right) / 32,$$

the grand mean of the readings. Thus $\hat{M} = 5.43/32 = .16969$. These formulas should be used with the understanding that $x_{ij} = 0$ if the i th thermometer does not occur in the j th pair.

Thus far all values have been given in coded form and the adjusted thermometer readings in terms of systematic deviations from the reference temperature M . It may be of interest to estimate readings for all thermometers at temperature M . These decoded estimated readings, calculated by the formula $\hat{M} + 40 + \hat{t}_i$, are as follows:

Thermometer	Temperature	Thermometer	Temperature
	° C.		° C.
1	40.11	5	40.18
2	40.22	6	40.20
3	40.21	7	40.10
4	40.13	8	40.21

The estimate of p_j is obtained by a simple adjustment of the observations in the j th pair. If i and i' are the thermometers in the j th pair, then

$$2\hat{p}_j = x_{ij} + x_{i'j} - 2\hat{M} - \hat{t}_i - \hat{t}_{i'}$$

For example, for $j=2$, $i=5$, $i'=3$, and

$$2\hat{p}_2 = .08 + .13 - 2(.16969) - .00531 - .03969,$$

so that $\hat{p}_2 = -.08719$.

It now is possible to exhibit the decomposition of x_{52} into its parts. Thus

$$x_{52} = \hat{M} + \hat{t}_5 + \hat{p}_2 + \hat{e}_{52}$$

$$.08 = .16969 + .00531 + (-.08719) + (-.00781).$$

It is interesting to note that the estimated error in this particular reading is of about the same magnitude as $\hat{\sigma}$.

The fundamental importance of the arrangement is that it makes it possible to intercompare the thermometers and to limit the error arising from fluctuations in the bath temperature to those temperature changes that take place in the very short interval required to read two thermometers. Temperature changes from one pair to another do not contribute to the error of measurement. This technique is applicable in all cases where either the apparatus or the environment may drift or undergo unpredictable changes.

4. Appendix

4.1. Derivation of Estimates

Let the group that contains m objects be denoted by G_1 , and the group that contains n objects be denoted by G_2 . Then the reduced normal equations

for estimating the treatment (thermometer) effects are

$$n\hat{t}_i - \sum_{u=m+1}^v \hat{t}_u = D_i \quad (1)$$

for i in G_1 and

$$m\hat{t}_i - \sum_{u=1}^m \hat{t}_u = D_i \quad (2)$$

for i in G_2 .

Summing over the treatments in G_1 , eq (1) becomes

$$n \sum_{u=1}^m \hat{t}_u - m \sum_{u=m+1}^v \hat{t}_u = \sum_{u=1}^m D_u. \quad (3)$$

Then imposing the restriction

$$\sum_{u=1}^v t_u = 0, \quad (4)$$

it is clear that

$$(n+m) \sum_{u=1}^m \hat{t}_u = \sum_{u=1}^m D_u. \quad (5)$$

Similarly, using eq (2), summing over the treatments in G_2 , and applying eq (4), obtain

$$(n+m) \sum_{u=m+1}^v \hat{t}_u = \sum_{u=m+1}^v D_u. \quad (6)$$

From eq (1) and (6) it follows for i in G_1 that

$$n(n+m)\hat{t}_i = (n+m)D_i + \sum_{u=m+1}^v D_u$$

or since $\sum_{u=1}^v D_u = 0$,

$$vn\hat{t}_i = vD_i - \sum_{u=1}^m D_u. \quad (7)$$

Similarly, for i in G_2 ,

$$vm\hat{t}_i = vD_i - \sum_{u=m+1}^v D_u. \quad (8)$$

4.2 Derivation of Variance

For random variables x and y let $V(x)$ and $Cov(x,y)$ denote, respectively, the variance of x and the covariance of x and y . Then for i and i' in G_1 ,

$$V(D_i) = 2m\sigma^2, \quad Cov(D_i, D_{i'}) = 0. \quad (9)$$

From eq (7) and (9),

$$n(\hat{t}_i - \hat{t}_{i'}) = (D_i - D_{i'}),$$

$$V(\hat{t}_i - \hat{t}_{i'}) = 4\sigma^2/n. \quad (10)$$

Similarly, for i and i' both in G_2 ,

$$V(\hat{t}_i - \hat{t}_{i'}) = 4\sigma^2/m. \quad (11)$$

For i in G_1 and i' in G_2 it is convenient to use the formula

$$V(\hat{t}_{i'} - \hat{t}_i) = (C_{ii} + C_{i'v} - 2C_{ii'})\sigma^2, \quad (12)$$

where C_{rs} is the element in the r th row and s th column of the inverse of the coefficient matrix of the reduced normal equations. From eq (7) and (8),

$$C_{ii} = 2(v-1)/vn, \quad C_{i'v} = 2(v-1)/vm, \quad C_{ii'} = 0.$$

Hence eq (12) becomes

$$V(\hat{t}_i - \hat{t}_{i'}) = 2(v-1)\sigma^2/mn. \quad (13)$$

4.3 Derivation of Estimate of σ

The differences d_j form a basis for the space, which consists of the error space and the space of the \hat{t} 's. Therefore, the sum of squares due to the d 's can be partitioned into two orthogonal parts, one due to error and one due to treatments. Since the sum of squares due to treatments is $\left(\sum_{i=1}^v \hat{t}_i D_i\right)^2/2$, twice the sum of squares due to error is

$$2(mn - m - n + 1)\hat{\sigma}^2 = \sum_{j=1}^{mn} d_j^2 - \sum_{i=1}^v \hat{t}_i D_i.$$

WASHINGTON, September 25, 1954.