# CHAPTER 4

# Use of Supplementary Observations to Reduce Error

## 4.1 INTRODUCTION

The methods described in the previous chapter use a qualitative grouping of the units. For instance, even in Example 3.3, where a quantitative measurement, body weight, is used to group experimental animals into blocks, no use of the measurements is made once the blocks have been formed, so that in effect only a ranking of the animals in order of increasing weight is used. It is natural to consider whether effective use can be made of the actual value of weight, either in addition to, or instead of, the grouping into blocks.

We shall deal with this topic in the present chapter. It is convenient to reserve the term *concomitant observation* for a supplementary observation that may be used to increase precision. An essential condition has to be satisfied in order that after use of the concomitant observation, estimated treatment effects for the desired main observation shall still be obtained. This condition is that the concomitant observations should be quite unaffected by the treatments.

## 4.2 NATURE OF CONCOMITANT OBSERVATIONS

We shall, therefore, consider situations in which in addition to the main observations, for which we want to find the treatment effects, we have for each experimental unit one or more concomitant observations. The essential point in our assumptions about these observations is that the value for any unit must be unaffected by the particular assignment of treatments to units actually used. In practice this means that either

(*a*) the concomitant observations are taken before the assignment of treatments to units is made; or

(*b*) the concomitant observations are made after the assignment of treatments, but before the effect of treatments has had time to develop.





This case occurs, for example, in some agricultural field trials, and in some laboratory experiments with animals; or

(*c*) we can assume from our knowledge of the nature of the concomitant observations concerned, that they are unaffected by treatment differences. For example in comparing a number of textile spinning processes, a main observation might be the end breakage rate and a concomitant observation the relative humidity in the spinning shed during processing. Both are taken during processing, after the treatments have been assigned to units, but it is clear that the relative humidity for any particular period of processing is unaffected by the process applied during that period, and so this observation is a concomitant observation in our sense.

Examples of concomitant observations have already been mentioned. Some more are the yield of product on a plot in years previous to the experimental year (which is particularly useful in experiments on perennial crops), the purity of the raw material in a chemical process, the weight of a particular organ (for example, the heart) of an experimental animal used in a biological assay, the score attained by a subject in a preliminary test in a psychological experiment, and so on.

## 4.3 THE USE OF A CONCOMITANT OBSERVATION AS AN ALTERNATIVE TO BLOCKING

Consider a situation, such as that of Example 3.3, where one concomitant observation on each unit is available, and suppose for simplicity that the same number of units are to be devoted to each treatment. We have seen in Example 3.3 how such an observation can be used to effect a grouping into blocks; suppose now either that the concomitant observations are of types (*b*) or (*c*) and so are not available at the time the treatments are allotted, or that it is desired to use the observations quantitatively in the analysis of the results instead of in the formation of blocks.

Suppose first that no alternative system of blocking suggests itself, so that the treatments are assigned to experimental units completely randomly. That is to say if five units are to receive treatment $T_1$, these are selected completely randomly from all the units available, using the methods described in the next chapter. Five more units are selected at random from the remainder for the second treatment, and so on. This is in a sense the simplest and most flexible design that can be used. Its disadvantage, if there is no concomitant observation, is that no attempt is made to reduce the effect of uncontrolled variation.

Suppose, however, that a concomitant observation, denoted by $x$, is made on each unit and that the main observation is denoted by $y$. Thus the full set of observations consists of a series of pairs $(x, y)$, one pair for each experimental unit. Thus if the experiment concerned alternative chemical processes, $x$ might be a measure of the purity of the raw material and $y$ the yield of product.

Consider first what would happen if there were no treatment effects, i.e., if the observation obtained in any unit did not depend on the treatment applied to it. Imagine the value of $y$ plotted against the corresponding value of $x$. Qualitatively two things may happen. There may

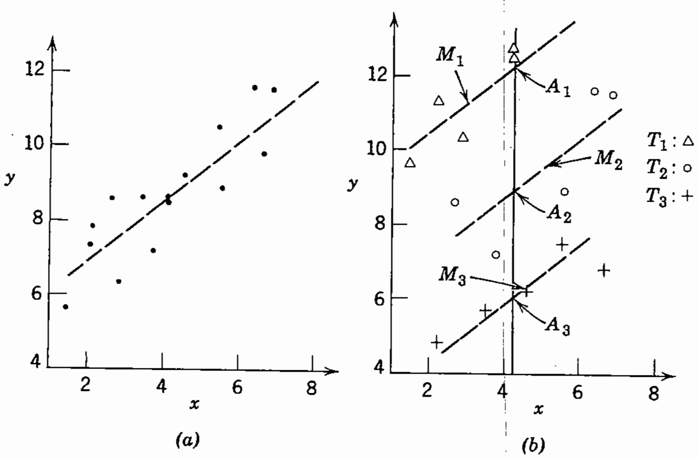


Fig. 4.1. The calculation of adjusted treatment means. (a) Plot of $y$ against $x$ in the absence of true treatment effects; (b) Corresponding plot after imposing treatment effects. The points $M$ are unadjusted means, the points $A$ are adjusted means, and the lines are fitted treatment lines.

be no appreciable relation between $y$ and $x$, the points forming a random scatter. In this case no useful information about $y$ can be obtained from the values of $x$. The interesting case is when the values of $x$ and $y$ cluster reasonably closely around a smooth curve. It is often convenient to assume that this curve is a straight line and we shall do this, but this assumption is not necessary; any simple smooth curve can be dealt with in a similar way.

Figure 4.1($a$) shows a typical case for fifteen experimental units. The line that has been drawn through the points is known statistically as the regression line of $y$ on $x$ and its physical interpretation is that it gives, corresponding to any particular value of $x$, an estimate of what the mean value of $y$ would be for a large number of experimental units similar to those used and all having the particular value of $x$. A detailed method



for the calculation of the line is described in textbooks on statistics; for a rough estimate it is sometimes enough to divide the points into four or five sets in order of increasing $x$, to find, perhaps by eye, the centroid of each set and then to fit a line by eye to these five centroids, checking that there is no evidence of systematic departure from linearity.

Now consider the situation when there are treatment effects. Plot a graph corresponding to Fig. 4.1($a$) but with the units receiving different treatments distinguished in some way. Figure 4.1($b$) shows what will

TABLE 4.1

FICTITIOUS OBSERVATIONS TO ILLUSTRATE THE CALCULATION OF ADJUSTED TREATMENT MEANS

| $x$ | Imaginary Value of $y$ before Applying Treatment | Treatment | Observed Value of $y$ | $x$ | Imaginary Value of $y$ before Applying Treatment | Treatment | Observed Value of $y$ |
|---|---|---|---|---|---|---|---|
| 1.5 | 5.6 | $T_1$ | 9.6 | 4.1 | 8.6 | $T_1$ | 12.6 |
| 2.2 | 7.8 | $T_3$ | 4.8 | 4.6 | 9.2 | $T_3$ | 6.2 |
| 2.2 | 7.3 | $T_1$ | 11.3 | 5.5 | 10.5 | $T_3$ | 7.5 |
| 2.7 | 8.6 | $T_2$ | 8.6 | 5.6 | 8.9 | $T_2$ | 8.9 |
| 2.9 | 6.3 | $T_1$ | 10.3 | 6.4 | 11.6 | $T_2$ | 11.6 |
| 3.5 | 8.6 | $T_3$ | 5.6 | 6.6 | 9.8 | $T_3$ | 6.8 |
| 3.8 | 7.2 | $T_2$ | 7.2 | 6.8 | 11.5 | $T_2$ | 11.5 |
| 4.1 | 8.5 | $T_1$ | 12.5 | | | | |

| | Unadjusted Treatment Means | Adjusted Treatment Means |
|---|---|---|
| $T_1$ | 11.26 | 12.20 |
| $T_2$ | 9.56 | 8.87 |
| $T_3$ | 6.18 | 5.94 |

result; the data on which this figure is based have been constructed from Fig. 4.1($a$) by adding $+4$ to the $y$ values of the five units selected at random for the treatment $T_1$, leaving five more unchanged for $T_2$ and adding $-3$ to the remaining units considered to have received $T_3$; the data themselves are given in full in Table 4.1.

What can we say in general about the form of the graph corresponding to Fig. 4.1($b$)? First we expect to find that so far as the values of $x$ are concerned, there are no systematic differences between treatments; for example we would have been surprised to find that the five units with lowest values of $x$ all received $T_1$. This is because of our initial assumption that $x$ is a concomitant variable and because the treatments were assigned to units randomly. If inspection of the results suggests that there are nevertheless systematic variations in $x$ between treatments, there are three possible explanations;

(i) the effect is a chance one. Whether or not this is likely to be so can be assessed by a statistical significance test;

(ii) our belief that the allotment of treatments to units is random may be false. Now if the methods of objective randomization described in the next chapter have been used, this possibility can be disregarded, but it may happen that for some reason of convenience, or by an oversight, the assignment of treatments has been left systematic or determined subjectively by some procedure other than strict randomization. In such a case the assignment of treatments could be correlated with $x$.

(iii) in the case of concomitant measurements of type (*b*) and (*c*), our assumption that $x$ is unaffected by treatments may be false, i.e., the variations in $x$ may represent a genuine treatment effect.

In certain applications (ii) or (iii), or both, can be ruled out. If the significance test indicates that it is very unlikely that the $x$ differences are chance ones and (ii) is the only possible explanation, we may proceed with the methods to be described below, although the possibility should not be overlooked that correlation with $x$ is not the only peculiarity in the allocation of treatments to units. However, if (iii) is a possibility, the methods must not be used in order to estimate the simple treatment effects of $y$; this case will be discussed in Example 4.6.

To sum up the discussion of this first point, we normally expect there to be no systematic differences between treatments in the values of $x$, but in certain cases it is all right to go ahead cautiously even if there are such differences.

The second general point follows from the basic assumption of Chapter 2 that the treatment effects are represented by the addition of constants. Therefore if, in the absence of treatment effects, we have a reasonably linear set of values such as Fig. 4.1(*a*), we shall find that the points for each treatment tend to cluster around a line and that the lines for the different treatments are parallel. This has happened in Fig. 4.1(*b*) and it will be clear why this is from the way the values for this diagram have been constructed in Table 4.1.

Therefore if the graph corresponding to Fig. 4.1(*b*) shows definite evidence of nonparallelism, the fundamental assumption about the treatment constants must be false. An apparent nonparallelism may arise from random fluctuations, i.e., from chance properties of the particular arrangement of treatments actually used, and the statistical significance of the nonparallelism can be tested by standard statistical methods. If real



nonparallelism is established beyond reasonable doubt, it would usually be required to estimate the treatment effects separately for a number of values of $x$. Alternatively if the nonconstancy can be removed by a simple transformation of $y$, for example, to $\log y$ or to $y/x$, this would be done. These points will not be gone into in detail here, since they concern the analysis rather than the design of the experiment; the general effect on the interpretation of the experiment has already been discussed in §§ 2.2, 2.3 and will be mentioned again below. The main implication for experimental design is that if the treatment effects are suspected to vary systematically from unit to unit, it will be wise to record the values of suitable supplementary observations, in order that the variations in treatment effect may be detected and explained.

If the true treatment effects are constant but the initial values, in the absence of treatment effects, tend to cluster not around a line but around a curve, the final graph corresponding to Fig. 4.1(*b*) will consist of a series of parallel curves, one for each treatment. The nonlinearity makes no difference in principle to the argument, but complicates the procedure. It is therefore desirable to make the relation effectively linear if this can be done easily. The method is to use a transformed concomitant variable, such as $\log x$, $1/x$, $\sqrt{x}$, etc., selected after an initial graphical analysis in terms of the original variable, $x$. We shall not go into details.

To sum up, we usually expect to find that there are no systematic differences between treatments in the values of $x$, and that the points for different treatments lie along parallel lines (or curves). We can now see how to use the diagram to obtain improved estimates of the treatment effects. Fit parallel lines to the sets of points and call these *treatment lines*. Take any convenient value of $x$, say the overall mean value of $x$, and find the corresponding value of $y$ on each treatment line. Call these values the *adjusted treatment means*. Then the differences among these quantities are estimates of the true treatment effects. This process can sometimes be done adequately by purely graphical methods, but if an objective answer is required, or if the standard error of the adjusted values is wanted, the whole procedure should be done arithmetically by the statistical technique called analysis of covariance (Goulden, 1952, p. 153). Again the details of this need not concern us; it is simply a question of fitting the parallel lines (or curves) arithmetically by statistically efficient methods,* rather than graphically.

The reason why the adjusted treatment means give a better estimate of the treatment effects than the uncorrected treatment means can be seen by thinking about the results for $T_1$ and $T_2$ in the numerical example.

* Some complications will arise if the scatter of $y$ about the fitted line or curve varies appreciably with $x$.

By chance $T_1$ has been tested on units with on the whole lower values of $x$ than those for $T_2$, and this means that the uncorrected mean $M_1$ of observations on $T_1$ will tend to be low. The adjusted value of $A_1$, obtained by sliding along the treatment line, is in effect an estimate of what the treatment mean for $T_1$ would have been had the units for $T_1$ had average values of $x$. The comparison of the adjusted treatment means for $T_1$ and $T_2$ therefore corrects the error arising from the differing values of $x$. Note that in fact the differences among the adjusted means are nearer the "true" values 4, 3 than are the corresponding differences for the unadjusted means.

The precision of comparisons based on adjusted treatment means depends on the standard deviation of $y$ about the regression line, i.e., roughly speaking, on the standard deviation that $y$ would have over a set of units all with the same value of $x$, treatment effects being absent. The formula for the standard error of the difference between two adjusted treatment means is complicated by the fact that the adjustments themselves have error, due to random error in estimating the slope of the regression line. This results in the standard error not being the same for all pairs of treatments. However for quick comparison the following approximate formula may be used, provided that there are no systematic differences between treatments in the value of $x$ and that there are the same number of observations on each treatment;

$$\begin{pmatrix}\text{standard error of}\\ \text{difference between two}\\ \text{adjusted means}\end{pmatrix} = \begin{pmatrix}\text{standard deviation}\\ \text{about regression}\\ \text{line}\end{pmatrix}$$
$$\times \sqrt{\left\{\frac{2}{\text{no. of obs. per treatment}}\right\}} \times \sqrt{\left\{1 + \frac{1}{\text{no. of treatments} \times \left(\text{no. of obs. per treatment} - 1\right)}\right\}}. \quad (1)$$

This should be compared with the formula (1) of § 1.2, namely that the standard error in an experiment not involving correction for a concomitant variable is equal to

$$\text{standard deviation} \times \sqrt{\left(\frac{2}{\text{no. of obs. per treatment}}\right)}. \quad (2)$$

The factor in the second square root in formula (1) is the contribution arising from the error in the slope of the fitted treatment lines.

There are several consequences of (1) and (2). First, if there is really no relation between $y$ and $x$, the two standard deviations in (1) and (2) are equal, so that the precision is lower after adjustment. This is because for nearly every particular arrangement of treatments there will be some apparent dependence of $y$ on $x$, so that a nonzero adjustment will be applied. This is an additional random term, inflating the error. However, if the number of units is large the quantity under the second square root in (1) will not be much greater than unity and the additional error is unlikely to be appreciable. The second, and more important, consequence is that, provided that this second square root can be ignored, the ratio of the standard error with and without adjustment for $x$ is



$$R_s = \frac{\text{standard deviation of } y \text{ about regression line}}{\text{overall standard deviation of } y}, \quad (3)$$

both true standard deviations being calculated in the absence of treatment effects. The ratio $R_s$ is a measure of the degree of relation between $y$ and $x$. Readers familiar with the definition and meaning of the population correlation coefficient $r$ may note that $R_s = \sqrt{(1 - r^2)}$. To give a rough picture of the strength of relation between $y$ and $x$ that will lead to a given value of the ratio $R_s$, Fig. 4.2 shows a number of scatter diagrams of $y$ and $x$ together with the corresponding values of the ratio. For example if, in the absence of treatment effects, $y$ and $x$ are related as in Fig. 4.2($c$), the use of $x$ as a concomitant variable would halve the standard error and thus be equivalent to a fourfold increase in the number of units.

## 4.4 ALTERNATIVE PROCEDURES

The procedure just given uses only the observations from the experiment and does not depend on assuming a completely specified relation between $y$ and $x$; the relation is in fact estimated from the data and is not regarded as known a priori. All that we assume is that the relation is approximately linear. The disadvantage of the method is that it can be rather tedious to apply.

A simpler method, and one that in special cases is often used implicitly, is the construction of a suitable *index of response*, i.e., a combination of $y$ and $x$ which is treated as a new observation for analysis as a single quantity. This method is frequently used when the concomitant observation is of the same nature as the main observation, differing from it in being taken before the treatments are applied. Thus $x$ may be the initial score in a spelling test, or the initial weight of an animal, $y$ being the score or weight after treatment. An index of response widely used in such cases is $y - x$, the improvement or increase during test, a comparison of treatments being made in terms of this, the separate values of $y$ and $x$ being ignored.

It can be seen that if the slope of the treatment lines, fitted by the method of § 4.3, happens to be exactly unity, the index of response and the method of adjustments give identical estimates of the treatment effects. In other words the decision to use a simple analysis of the differences, $y - x$, amounts to assuming a special form for the residual relation between $y$ and $x$, whereas the method of adjustments in effect finds in an objective way from the data, the most suitable linear combination of $y$ and $x$ for analysis.

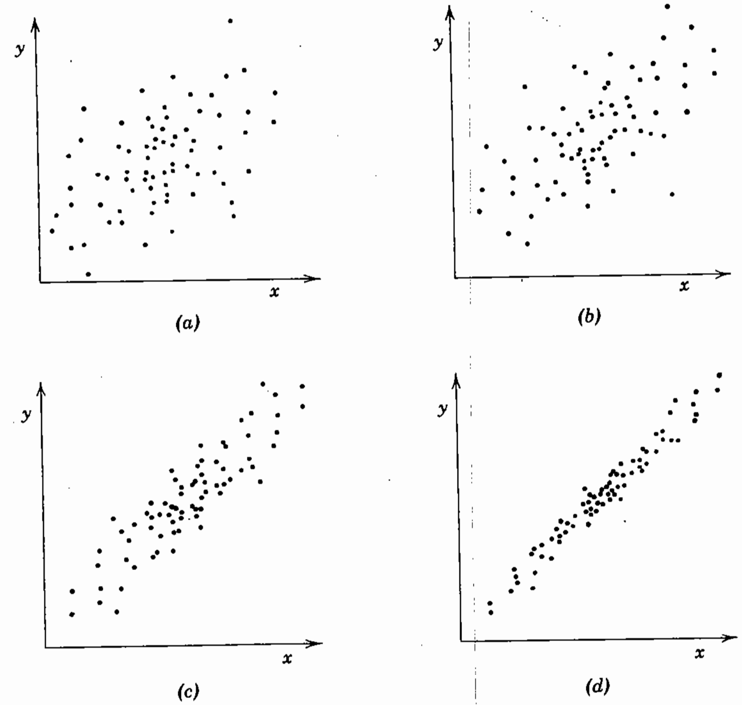


Fig. 4.2. Scatter diagrams to illustrate various ratios $R_s$ and correlations $r$.
(a) $R_s = 0.9$, $r = 0.44$ (b) $R_s = 0.8$, $r = 0.60$
(c) $R_s = 0.5$, $r = 0.87$ (d) $R_s = 0.2$, $r = 0.98$

Quite generally, whenever theory or prior knowledge suggest a particular relation between $y$ and $x$, an alternative to the method of adjustments is to assume that the relation is approximately true and to make a simple analysis of $y - kx$, where $k$ is the expected slope of the relation between $y$ and $x$ (see Example 4.4 in the next section). In the previous paragraph $k$ is taken to be unity. The objection is of course that if the value of $k$ chosen is seriously different from the true slope, an appreciable loss of precision will result. The uncritical use of indices of response was



criticized on these grounds by Fisher (1951, p. 161), who introduced the method of adjustment of § 4.3.

Gourlay (1953) and Cox (1957) have investigated how close $k$ need be to the true slope in order to avoid a serious loss of information. Table 4.2, taken from the second of these papers, sums up the conclusions.

TABLE 4.2

LOSS OF PRECISION FROM USING WRONG INDEX OF RESPONSE

| Correlation between $y$ and $x$ | Range within which Ratio of True to Assumed Slope must Lie to Avoid a Loss of Precision of | | |
|---|---|---|---|
| | 10% | 20% | 50% |
| 0.4 | (0.28, 0.72) | (−0.02, 2.02) | (−0.62, 2.62) |
| 0.6 | (0.68, 1.32) | ( 0.40, 1.60) | ( 0.06, 1.94) |
| 0.8 | (0.76, 1.24) | ( 0.67, 1.33) | ( 0.47, 1.53) |
| 0.9 | (0.82, 1.18) | ( 0.74, 1.26) | ( 0.59, 1.41) |
| 0.95 | (0.90, 1.10) | ( 0.85, 1.15) | ( 0.77, 1.23) |

A loss of precision of say 50 per cent means here that the standard error with the assumed slope is $\sqrt{1.5}$ times what it would have been with the correct slope.

For example, suppose that the correlation between main and concomitant observations is 0.8; see equation (3) and Fig. 4.2 for an explanation of what this means. Then the loss of precision arising from using the wrong index of response is less than 20 per cent provided that the assumed slope $k$ is not less than 0.7 times or more than 1.3 times the true slope; a precise definition of loss of precision is given in the footnote to the table. A reasonable general conclusion from the table is that the index of response does not have to be too close to the best one to give quite good results.

It is not possible to give a general rule about when to use an index of response rather than the method of adjustments. For the decision rests on, for example, the importance of saving time in analysis, the strength of belief in the assumed relation, and the importance of objectivity in the answer. However, if a number of similar experiments have to be analyzed, a sensible thing may be to analyze one or two experiments by the lengthier method before deciding what to do with the whole series. In any case a quick graphical analysis of all or part of the data will often show whether the assumed value of $k$ is a reasonable one and whether the treatment effects appear to be independent of $x$, i.e., whether the treatment lines are parallel.

When the concomitant variable is available before the treatments are assigned to the experimental units the method used in Example 3.3 is a further alternative; that is, the units can be grouped in randomized blocks on the basis of the values of $x$, putting in one block those units with the lowest values of $x$, etc. If there is a nearly perfect relation between $y$ and $x$, with negligible random scatter, the use of adjustments based on $x$ will give nearly zero random errors, whereas the randomized block design will give appreciable error arising from the dispersion of $x$ within blocks. However this is an extreme case and in many practical situations the randomized block method is adequate unless there are one or two very extreme values of $x$, or the number of units per treatment is small, or the correlation between the two variables is high, say 0.8 or more. The randomized block approach is of course simpler in analysis since it does not involve the calculation of adjustments. Some quantitative comparisons of blocking and adjustment have been given by Cox (1957).

The most profitable uses of the adjustment method are likely to be when there is no known index of response and when

(*a*) the value of $x$ is not available until after the treatments have been assigned to the units; or

(*b*) the relation between $y$ and $x$ is of intrinsic interest; or

(*c*) it is important to examine whether the treatment effects vary with the value of $x$; or

(*d*) it is desired to block the experimental units on the basis of some other property, so that the information contained in $x$ has to be used in another way. We consider this in the next section.

## 4.5 THE USE OF A CONCOMITANT OBSERVATION IN ADDITION TO BLOCKING

In § 4.3 we described a method for adjusting the treatment means to take into account a concomitant variable $x$, it being assumed that there is no grouping of the experimental units into blocks. Suppose now that the units are arranged in a randomized block, Latin square, or other similar design and that it is required to make an adjustment for a concomitant variable. Then the assumptions and general method discussed in § 4.3 above apply, with the single change that instead of plotting the observations of $y$ directly against those of $x$, we plot a partial residual value of $y$ against a corresponding partial residual value of $x$. These partial residuals differ slightly from the residuals defined in §§ 3.3, 3.4 in that, in order to get a more meaningful graph, the treatment means are



not to be subtracted. For example in a randomized block experiment, the partial residual corresponding to a particular observation $y$ is

$$\text{observation} - \begin{pmatrix}\text{mean observation}\\ \text{on the}\\ \text{corresponding block}\end{pmatrix}, \tag{4}$$

whereas for a Latin square the partial residual is

$$\text{observation} - \begin{pmatrix}\text{mean observation}\\ \text{on the}\\ \text{corresponding}\\ \text{row}\end{pmatrix} - \begin{pmatrix}\text{mean observation}\\ \text{on the}\\ \text{corresponding}\\ \text{column}\end{pmatrix} + \begin{pmatrix}\text{overall}\\ \text{mean}\end{pmatrix}. \tag{5}$$

The general idea is that before plotting, variation in $y$ and $x$ accounted for by the grouping of the experimental units should be removed.

The procedure is best explained in detail by an example.

*Example* 4.1. Pearce (1953, p. 113) has illustrated the statistical technique for calculating adjustments in a randomized block experiment. His data are reproduced in Table 4.3. The main observation, $y$, is the yield in pounds of apples over a four-year experimental period and the concomitant observation, $x$, is the yield in bushels in the preceding four-year period, during which no differing treatments were applied to the trees. Thus $x$ is a proper concomitant observation of type (*a*). The six treatments under comparison are denoted by $T_1, \dots, T_6$ and the observations in the table have been set out ordered by treatments, rather than randomly.

The first step in the analysis is to work out the mean values of $y$ and of $x$ by treatments and by blocks (the first part of Table 4.3(*b*)). The adjusted treatment means can be obtained directly by analysis of covariance without the calculation of partial residuals, but to do the adjustment semigraphically and to obtain a general understanding of what is being done, we proceed as follows.

The partial residuals denoted by $Y$ and $X$ are set out in Table 4.3(*c*), and have been calculated from equation (4). Thus for the observation $y$ on $T_2$ in block III, the partial residual is $243 - 268.7 = -25.7$, since the mean of all observations, $y$, in block III is 268.7. Figure 4.3 shows the corresponding partial residuals $Y$ and $X$ plotted against one another for each treatment; this figure is analogous to Fig. 4.1(*b*). Inspection of the graph shows that within any one treatment there is a strong effectively linear relation between the values of $Y$ and $X$ and that the lines for each treatment are substantially parallel. Thus there is no evidence that the true treatment effects depend on $x$; a line for one treatment rising steeply from a relatively low value of $Y$ for negative $X$ to a relatively high value of $Y$ for positive $X$ would have suggested that the corresponding treatment was relatively good for "good" trees with a high initial yield and relatively bad for "poor" trees with a low initial yield. In the absence of such effects, we may fit parallel treatment lines. The slope as estimated graphically is about 30 units, that is an increase of 30 in $Y$ for unit increase in $X$. The slope calculated from the analysis of covariance is 28.4 and this value has been used in what follows.

Treatment lines are drawn with this slope for the points corresponding to each

TABLE 4.3

THE CALCULATIONS OF ADJUSTMENTS IN A RANDOMIZED BLOCK EXPERIMENT

(a) *The Data*

| | Block I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|---|
| | $y$ | $x$ | $y$ | $x$ | $y$ | $x$ | $y$ | $x$ |
| $T_1$ | 287 | 8.2 | 290 | 9.4 | 254 | 7.7 | 307 | 8.5 |
| $T_2$ | 271 | 8.2 | 209 | 6.0 | 243 | 9.1 | 348 | 10.1 |
| $T_3$ | 234 | 6.8 | 210 | 7.0 | 286 | 9.7 | 371 | 9.9 |
| $T_4$ | 189 | 5.7 | 205 | 5.5 | 312 | 10.2 | 375 | 10.3 |
| $T_5$ | 210 | 6.1 | 276 | 7.0 | 279 | 8.7 | 344 | 8.1 |
| $T_6$ | 222 | 7.6 | 301 | 10.1 | 238 | 9.0 | 357 | 10.5 |

(b) *Some Mean Values*

(i) *by blocks*

I 235.5 7.10 II 248.5 7.50 III 268.7 9.07 IV 350.3 9.57

(ii) *by treatments*

| | Mean $y$ | Mean $x$ | Mean $x$ − Overall Mean | Adjustment* | Adjusted Mean |
|---|---|---|---|---|---|
| $T_1$ | 284.5 | 8.45 | 0.14 | −3.98 | 280.5 |
| $T_2$ | 267.8 | 8.35 | 0.04 | −1.14 | 266.7 |
| $T_3$ | 275.2 | 8.35 | 0.04 | −1.14 | 274.1 |
| $T_4$ | 270.2 | 7.92 | −0.39 | 11.08 | 281.3 |
| $T_5$ | 277.2 | 7.48 | −0.83 | 23.57 | 300.8 |
| $T_6$ | 279.5 | 9.30 | 0.99 | −28.12 | 251.4 |
| Overall | 275.75 | 8.31 | | | |

(c) *The Partial Residuals*

| | Block I | | II | | III | | IV | |
|---|---|---|---|---|---|---|---|---|
| | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ | $Y$ | $X$ |
| $T_1$ | 51.5 | 1.1 | 41.5 | 1.9 | −14.7 | −1.4 | −43.3 | −1.1 |
| $T_2$ | 35.5 | 1.1 | −39.5 | −1.5 | −25.7 | 0.0 | −2.3 | 0.5 |
| $T_3$ | −1.5 | −0.3 | −38.5 | −0.5 | 17.3 | 0.6 | 20.7 | 0.3 |
| $T_4$ | −46.5 | −1.4 | −43.5 | −2.0 | 43.3 | 1.1 | 24.7 | 0.7 |
| $T_5$ | −25.5 | −1.0 | 27.5 | −0.5 | 10.3 | −0.4 | −6.3 | −1.5 |
| $T_6$ | −13.5 | 0.5 | 52.5 | 2.6 | −30.7 | −0.1 | 6.7 | 0.9 |

(d) *Some Estimates of Precision*

The estimated standard error of the difference between two uncorrected treatment means is 28. The estimated standard error of the difference between two adjusted treatment means depends slightly on which pair of treatments is being compared; an average value is 12.

* The adjustment is equal to minus the slope (28.4) times the preceding column.



treatment. The position of each line is chosen so that it passes through the centroid of the appropriate points. Thus, for $T_6$ the mean values of $x$ and $y$ are 9.30 and 279.5. Therefore, in terms of residuals the mean point for $T_6$ has $X$ equal to $9.30 - 8.31 = 0.99$ and $Y$ equal to $279.5 - 275.75 = 3.75$, and this is the point $M_6$ in Fig. 4.3. To avoid complicating the figure unduly, the treatment lines are shown only for $T_5$ and $T_6$. The difference between the $Y$ values for the centroids $M_5$ and $M_6$ is just the difference between the unadjusted treatment means, Table 4.3(*b*), column 1.

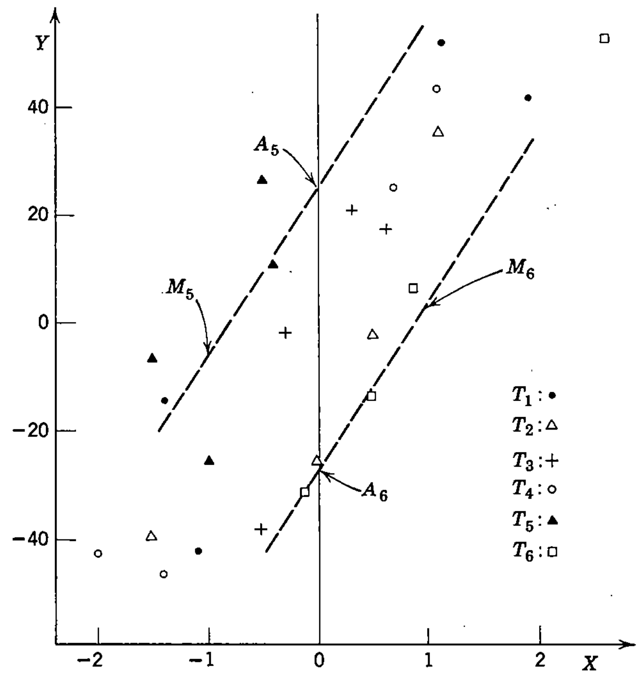


Fig. 4.3. Adjusted means in a randomized block experiment.

To find the adjusted estimates we take a standard value of $X$, say zero, in Fig. 4.3 and read off, or calculate, the corresponding values of $Y$ on the treatment lines. These correspond to the points $A_5$ and $A_6$ in the figure; when the overall mean value of $y$, 275.75, is added to the $Y$ values, we obtain the adjusted treatment means in the last column of Table 4.3(*b*). Thus the $Y$ value for $A_6$ is about −28, agreeing with the value in the last column but one of the table. An arithmetical method for obtaining the adjustments is indicated in the table.

The common-sense justification of this procedure is exactly the same as that for the simple example in § 4.3. The only new point is the elimination of block differences prior to the plotting of the scatter diagram. The reasonableness of this step can be seen by considering that it should be possible to superimpose on the data arbitrary block effects for both $x$ and $y$ without affecting the conclusions.

The main effect of the adjustments has been on the treatments $T_5$ and $T_6$. The treatment $T_6$, for example, was applied to trees having on the whole high values of initial yield, so that the adjustment to $y$ has lowered the yield. Some estimates of precision obtained from the full statistical analysis are given in Table 4.3(*d*) and show that the use of $x$ has apparently appreciably increased the precision of the experiment.

The description of this example in terms of graphical methods is, of course, not to be taken as an implied criticism of the usefulness or appropriateness of the arithmetical technique of analysis of covariance.

Quite generally, in an experiment in which the concomitant observation $x$ is available before the allocation of treatments to units, it would be possible to group the units into blocks on the basis of $x$ and in addition to apply adjustments to remove the effect of variations in $x$ not accounted for in the blocking. It would, however, very rarely be worth doing this solely for the purpose of increasing precision. The main value of the method of adjustments is when $x$ represents some property of the experimental unit not directly connected with that used in grouping the units into blocks, or into the rows and columns of a Latin square.

## 4.6 SOME GENERAL POINTS

The following examples serve both to illustrate applications of the method and to bring out points of general interest connected with concomitant variables.

*Example* 4.2. In experiments on spinning textile yarn, it is sometimes difficult to ensure that each batch of yarn spun has the same mean weight per unit length. Yet one of the observations of most interest, the end-breakage rate in spinning, is quite critically dependent on the weight per unit length. Therefore it is natural to take the weight per unit length as a concomitant variable in an analysis of the type just described. Randomized block or Latin square designs are frequently useful to control systematic differences between machines, times, etc. This is an example of the use of a concomitant variable of type (*c*) (§ 4.2), in that the estimate of mean weight per unit length does not become available until after the completion of spinning. However, differences in mean weight per unit length are accidental in the sense that they could have been eliminated by minor adjustment of the machinery had sufficient initial information been available. Hence there is no sense in which treatments "cause" differences in yarn mean weight per unit length, so that use of this quantity as a basis for adjustment is in order.

This example illustrates the use of adjustments to correct for failure to control completely the experimental conditions.

*Example* 4.3. Pearce (1953, p. 34) has given an excellent account of the use of concomitant variables for adjustment in experiments on fruit trees and other perennial crops. He points out that when an experimental unit consists of one tree or bush, positional variations, which can conveniently be controlled by blocking, are likely to be relatively less important than in experiments on, say,



wheat, in which each unit contains at least several thousand plants and in which individual differences are therefore likely to balance out. Consequently in the experiments Pearce is describing, it is worthwhile not only to control positional effects by blocking, but also to obtain concomitant variables for eliminating, as far as possible, the effect of variations arising from the peculiarities of individual trees. He lists suitable concomitant observations for various crops.

It would be possible, although inconvenient, to use these observations as a basis for blocking, following the method of Example 3.3. This would, however, mean that blocks would not be formed from adjacent units and this would make spatial variations more difficult to control, as well as probably making the experimental work more difficult to organize.

The general conclusion to be drawn from this example is that the use of concomitant observations is especially worth consideration whenever a major portion of the uncontrolled variation is associated with the particular objects, animals, plants, people, etc. forming the experimental units, rather than with the "external" conditions under which the experiment is carried out.* On the other hand variation associated with "external" conditions (observer, time, spatial, etc. differences) is often most conveniently controlled by the randomized block or Latin square devices. Experimental psychology and education are two fields where these remarks are particularly relevant, since in them a major source of uncontrolled variation lies in differences between subjects.

*Example* 4.4. Finney (1952, p. 45) has discussed an experiment of Chen et al. (1942) relating to the assay of digitalis-like principles in ouabain and other cardiac substances. The method was to infuse slowly a suitable dilution of drug into an anaesthetized cat and to record the dose at which death occurred. Twelve drugs were under comparison, ouabain being taken as a standard. There were three observers each testing four cats per day and the experiment was repeated on twelve days. A $12 \times 12$ Latin square was used with each column representing a day's work and each row a combination of observer and time of day. The details of this are not relevant to the discussion of adjustments, but the reader not too familiar with the use of Latin squares should think over this use of the design, consulting the above references if necessary.

The concomitant observation was the heart weight of the cat, determined, of course, at the end of each test, therefore being a concomitant variable of type (*c*). There were sound reasons for regarding the heart weight as unaffected by drug differences. Finney discusses the use of a second concomitant observation, the body weight, but we shall disregard this here. Notice how this use of the Latin square and a concomitant variable fits in with the general remarks at the end of Example 4.3.

There were physiological reasons for expecting the fatal dose of a particular drug to be proportional to the surface area of the cat's heart, which in turn is roughly proportional to the two-thirds power of heart weight. This power-law dependency is converted into a linear relation by working with logarithms, namely

$$\log(\text{fatal dose}) = \tfrac{2}{3}\log(\text{heart weight}) + \text{constant}. \tag{6}$$

* Another method of controlling such variations is to use each object as a unit several times (Chapter 13). However, this may, as in the present example, be impossible from the nature of the experiment, and in other cases may lead to troublesome "interference" effects between different units.

Even if this relation does not hold exactly, it seems more reasonable to expect an approximately linear relation between logarithms than between original observations. Also, as we have seen in Example 2.1, it is natural to consider drug differences as affecting doses multiplicatively (e.g., one drug always requiring a 10% greater dose than a second drug would have on the same cat) and thus affecting log dose additively. Hence there are two reasons for converting the observations of dose and heart weight into logarithms. When this was done the fitted slope for the relation of log dose on log heart weight came to 0.676, in close agreement with the theoretical value of $\frac{2}{3}$; hence the use of an index of response (§ 4.4) log (dose) $-$ $\frac{2}{3}$ log (heart weight) would have given excellent results.

This example illustrates the point that we should consider whether there is any prior reasoning to suggest the general form of the relation between the main observation and the concomitant observation, and exemplifies also that the relation between the two may be of some intrinsic interest.

Another general question, of analysis rather than of design, raised by the previous example concerns the desirability of transforming the observations mathematically, for example by taking logarithms, before analysis. This brings up some difficult issues. The concomitant variable $x$ does not enter into the definition of what we are trying to estimate, namely the comparison of doses, and it is sometimes very helpful to transform $x$ in order to get a linear relation between the variables. When we apply a mathematical transformation to the main observation, however, we are, if we work with means of transformed quantities, estimating differences between treatments on the transformed scale, not on the original scale.

In the example, the transformation to log dose is suggested both because ratios of doses provide the natural measure of relative potency and also because the theoretical relation between main and concomitant observations is simplified thereby. Things are not always so easy. If there is a definite reason for regarding treatment differences in terms of an observational scale $z$ as particularly meaningful and for expecting treatment differences to be constant on this scale, then it seems wrong to estimate treatment effects on some other scale such as log $z$ just for reasons of statistical or arithmetical convenience. In many applications, however, there may be no particular reason to expect that the scale on which the main observation is recorded is the one most helpful for analysis and interpretation of the results.

Common transformations usually affect the results materially only if the total fractional variation of results around the average is very appreciable, representing say a twofold variation or more.

*Example* 4.5. In all our examples so far the concomitant variable $x$ has been a quantitative measurement. We can, however, sometimes usefully employ a "dummy" variable $x$ to represent a qualitative division of the units into two classes. Thus, in an experiment with animals, a randomized block design might be used, taking all the animals in any one block from a single litter. In general, however, it would not be possible in doing this to ensure that all the animals in one block are of the same sex. The problem may therefore arise of adjusting the treatment means for the effect of any systematic difference between sexes, since in general each treatment will not occur equally frequently on males and females. To achieve this adjustment, introduce a concomitant variable taking the value 0 for males and 1 for females. Then the adjusted treatment means, calculated by the procedure of § 4.3, give the estimated treatment effects corrected for variation in the sex-ratio between treatment groups.



It might also be interesting in this application to examine whether the treatment effects for males are different from those for females, and this too can be done by an extension of the above procedure. Other groupings of the experimental units into two* sets not controlled in the blocking can be handled in the same way, provided of course, that the condition for a concomitant variable is satisfied, i.e., that the concomitant variable is unaffected by the treatments.

Throughout the discussion of adjustments it has been assumed that $x$ is a concomitant variable, i.e., that the value of $x$ for any unit is unaffected by the treatment applied to that unit. A further important aspect of the method is the examination of whether the treatment effects are constant. This is done, as explained in § 4.3, by looking for nonparallelism in the treatment lines or curves. We now consider a different type of application in which $x$ is not a concomitant variable.

*Example* 4.6. The following example is suggested by Gourlay (1953) in his discussion of the analysis of covariance applied to psychological research. To compare a number of methods of teaching composition, the methods are assigned randomly to a number of experimental groups, with several groups for each treatment. After an appropriate time, scores are obtained for each group to measure (*a*) ability in composition and (*b*) knowledge of the mechanical aspects of English. Call these two scores $y$ and $x$.

Then $x$ is certainly not a concomitant variable, since it is quite likely to be influenced by methods of teaching composition. If, however, we go ahead with the method of adjustments, we are answering the question: what would the mean values of $y$ have been had there been no differences in $x$? In other words we are asking whether any differences in ability in composition can be accounted for in terms of the effect of teaching on a knowledge of the mechanical aspects of English.

Figure 4.4 shows four cases that could arise with two treatments in a completely randomized experiment. In Fig. 4.4(*a*) differences between treatments in $y$ are clearly not accounted for by differences in $x$. We may, in terms of the particular example, conclude that the treatment $T_2$ has improved ability in composition and that the improvement is not wholly accounted for by improvements in the knowledge of mechanical aspects of English. In Fig. 4.4(*b*) the difference between treatments in $y$ is less than would be expected on the basis of the increase in $x$. In Fig. 4.4(*c*) differences in $x$ account for the differences in $y$. In Fig. 4.4(*d*) the interpretation is in doubt because, although the fitting of treatment *lines* might suggest unaccounted differences in $y$ the data are also reasonably consistent with

* Groupings into, say, three or four sets can be handled reasonably easily by the methods for several concomitant variables, to be described in § 4.6.

a single smooth curve. Similar difficulties are likely to arise whenever there are large differences in $x$, so that substantial extrapolation is involved. In practice, statistical analysis would usually be desirable to examine the precision of the conclusions suggested by the diagrams.

It should be noted that we talk about differences in $y$ being "accounted for" by differences in $x$. That this is the right thing is best seen from the consideration that $x$ and $y$ are variables of the same nature and that, from a statistical point of view, any arguments that purported to show that differences in $x$ caused

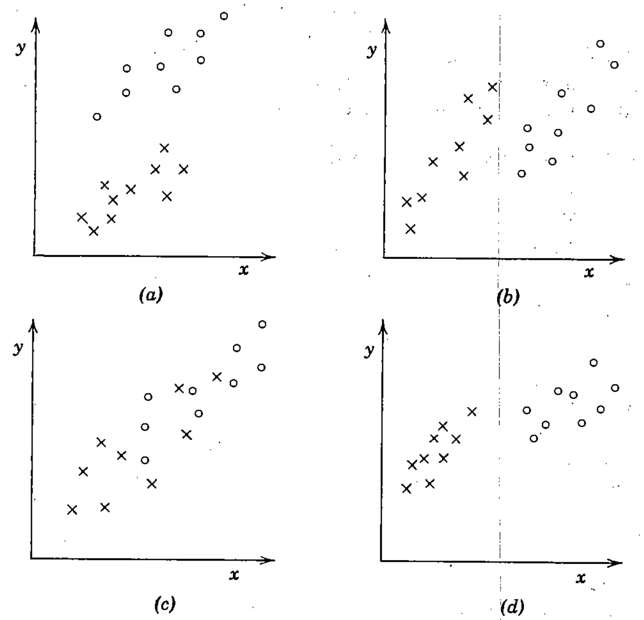


Fig. 4.4. Some possible types of relation between $y$ and $x$ when there are two treatments $T_1$, ×; $T_2$, ○.

differences in $y$ would equally prove that differences in $y$ caused differences in $x$. If we were to conclude that differences in $x$ cause differences in $y$, this could only be because of hypotheses about the nature of the variables $x$ and $y$ extraneous to the experimental results themselves. Such hypotheses may be perfectly in order, but the scientist should always be aware when they are being appealed to.

This application, when $x$ is not a concomitant variable, is clearly quite different from the use of a concomitant variable to increase the precision of treatment comparisons. The types of variable that are legitimate for increasing precision have been set out in § 4.2 and it is important to make sure in using the method that the conditions given there are satisfied. In particular it is in principle quite wrong to use as a concomitant variable



a quantity $x$ which there is no strong prior reason to regard as unaffected by treatments, but which has happened, in the particular experiment, to show nonsignificant treatment effects.

## 4.7 SEVERAL CONCOMITANT VARIABLES

It may happen that instead of one concomitant observation on each experimental unit there are several, that we think that they all give useful information about the main observation $y$ and that we therefore wish to use them all to adjust the estimated treatment effects for $y$. For example, in an animal experiment we may have for each animal the initial weight and also other measurements that are considered important. In a psychological experiment we may have for each subject initial scores in several tests and other measurements on the subject relevant to the performance of the experimental task. In an industrial experiment we may have several different measurements on each batch of raw materials. We shall assume that all these observations are genuine concomitant observations.

A simple way of dealing with this problem is to combine the concomitant observations into one. Thus, if we had percentage scores in several tests, we could take the total score in all tests as a single concomitant variable. Similarly, if the concomitant variables $x_1, x_2, \ldots, x_k$ are not commensurable, we could construct a combination of them by some intuitively reasonable process. Thus, if $s_1, s_2, \ldots, s_k$ are the standard deviations of $x_1, x_2, \ldots, x_k$ measuring the variation between units and if $w_1, w_2, \ldots, w_k$ are rough measures* of the probable importance of $x_1, x_2, \ldots, x_k$ we could take a new preliminary variable $w_1x_1/s_1 + w_2x_2/s_2 + \ldots + w_kx_k/s_k$. If we are fortunate enough to have previous data linking $y$ to $x_1, \ldots, x_k$, the best regression formula for predicting $y$ from $x_1, \ldots, x_k$ gives a good single variable.

These methods are useful but are, except for the last one, open to obvious drawbacks. We may, by using an inappropriate single variable, sacrifice a lot of information; if we make a really bad choice we may do worse than using just one of the original variables. Also we lose objectivity in that there is no guarantee that another worker, faced with the same experimental data, even though asking the same questions, would not have reached somewhat different conclusions. This is not always important.

If we decide to use all the concomitant observations, without prior combination into a single quantity, the appropriate numerical technique,

* If $x_1, x_2, \ldots, x_k$ represent roughly independent properties of the experimental units, we should take $w_1$ proportional to the correlation coefficient we expect to find between $y$ and $x_1$, etc.

multiple analysis of covariance, is a direct extension of the corresponding technique with one concomitant variable. The general interpretation of what is being done is the same too. Thus, with two concomitant variables $x_1$, $x_2$, we are effectively plotting a three-dimensional diagram of $y$ against $x_1$ and $x_2$, separately for each treatment, and then fitting parallel treatment planes to each set of points. The values of $y$ at some standard values of $x_1$ and $x_2$ give the adjusted treatment means. The same idea applies when there are three or more preliminary variables, although more than three dimensions are required to visualize the procedure geometrically.

We shall not go into details of the numerical method here, but it should be noted that the arithmetic, while straightforward, gets rapidly more laborious as the number of concomitant variables is increased. For this reason the method is not recommended as a routine tool, although there is no doubt that it is on occasion valuable and is worth knowing about. It would be possible to find the adjustments semigraphically, but the procedure is rather involved and not of much interest. The numerical calculations when we use a single concomitant variable, but fit treatment curves instead of lines, are very similar to those with several variables. This method of increasing precision by adjustment for several concomitant variables suffers from the general disadvantage that the gain is obtained by indirect calculation, i.e., by obtaining quantities that are comparatively remote from the original observations.

Sometimes, it may be required to use two or more concomitant observations in order to group the units into blocks, just as the single concomitant observation, body weight, was used in Example 3.3. If we have, from previous work, a regression formula for the main observation in terms of the concomitant observations there is no difficulty. If we have no such prior quantitative information, a sensible procedure with two variables $x_1$ and $x_2$ is the following. Plot a scatter diagram of $x_2$ against $x_1$ choosing a scale such that the dispersions in the two directions are approximately equal. Thus, with sixteen experimental units we should have a diagram with sixteen points, one for each unit. To group into four blocks of four units each, we take that division which makes each block of four points as compact as possible in the scatter diagram; if $x_1$ is thought to be a better variable than $x_2$, we pay more attention to dispersion parallel to the $x_1$ axis than to that parallel to the $x_2$ axis. Various ways of introducing a third variable, $x_3$, will occur to the reader.

## SUMMARY

A concomitant observation is one whose value for any experimental unit is independent of the arrangement of the treatments under comparison.



Suppose that there is available on each unit a concomitant observation in addition to the main observation in terms of which it is required to compare the treatments.

The concomitant observations can be used to increase the precision of the treatment comparisons, provided that the concomitant and main observations on a unit would have been closely correlated in the absence of treatment effects. One method, which is essentially the statistical technique called analysis of covariance, is to obtain from the data adjustments that should be applied to the treatment means, in order to estimate what main observations would have been obtained had it been possible to make the concomitant variable the same for all experimental units. A second, simpler but usually less efficient, method is to analyze the data in terms of a suitable index of response formed by combining the main and the concomitant observations on a unit into a single quantity.

The skilful choice of concomitant observations can lead to an appreciable increase in precision, particularly when the main uncontrolled variation arises from the peculiarities of individual experimental units (animals, subjects, etc.).

A further important use of the concomitant observations is in the detection and explanation of variations in treatment effect from unit to unit.

## REFERENCES

Chen, K. K., C. I. Bliss, and E. B. Robbins. (1942). The digitalis-like principles of *Calotropis* compared with other cardiac substances. *J. Pharmac. and Exptl. Therapeutics*, **74**, 223.

Cox, D. R. (1957). The use of a concomitant variable in selecting an experimental design. *Biometrika*, **44**, 150.

Finney, D. J. (1952). *Statistical method in biological assay*. London: Griffin.

Fisher, R. A. (1951). *The design of experiments*. 6th ed. Edinburgh: Oliver and Boyd.

Goulden, C. H. (1952). *Methods of statistical analysis*. 2nd ed. New York: Wiley.

Gourlay, N. (1953). Covariance analysis and its applications in psychological research. *Brit. J. Statist. Psychol.*, **6**, 25.

Pearce, S. C. (1953). *Field experimentation with fruit trees and other perennial crops*. East Malling, England: Commonwealth Bureau of Horticulture and Plantation Crops.