What to do about the covariables that complicate almost every statistical comparison is the message of the last of three related papers. Perhaps its discussion of the permissible number of covariables is its most pertinent reminder for the nonstatistical researcher.

The Use of Analysis of Covariance and Balancing in Analytical Surveys*

BERNARD G. GREENBERG, PH.D., F.A.P.H.A.

Professor and Head, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, N. C.

SPECIFIC differences in mortality, morbidity, and other characteristics may be observed to occur in surveys of population groups. Before concluding that the occurrence of a difference in one of those variables can be ascribed to the factor(s) used in classifying the populations, it is important to investigate the composition of the populations with respect to other attributes related to that particular variable. For example, does a population of nonsmokers have less lung cancer than a population of smokers because of smoking habits, or is the former group less prone to develop lung cancer because of age, race, sex, socioeconomic, and occupational differences? Similarly, is the nutritional status of one racial group better than that of another because of real differences in food habits, or is it a reflection of the pattern of cash income? Is the mortality experience of community A lower than that of community B because of differing health conditions and facilities in the two areas, or does it merely indicate a difference related to the age, race, and sex distributions?

The comparison of two groups which are essentially alike in every detail except the one being tested is an ideal which is usually unattainable. This condition is probably inevitable since every individual in a population has multiple possible measurements each of which may contribute, in varying degrees, to produce the observable effect. Uncontrolled and associated variables are always present when population groups are studied, and the research worker in public health must be prepared to deal with variations in the above mentioned attributes as well as with differences in educational level, height, weight, birthweight, order of birth, period of gestation, amount of nutrient consumed, number of previous inoculations, or an almost infinite variety of other possibilities.

This problem is not uniquely limited to analytical surveys. Even in planned experiments † where a group of valid controls may have to be constructed, the presence of many variables other than the one under immediate consideration becomes a complicating factor. The associated variables are referred to in

* Presented before the Statistics Section of the American Public Health Association at the Eightieth Annual Meeting in Cleveland, Ohio, October 23, 1952.

† A planned experiment is distinguished from an analytical survey here by defining the former to include only those cases in which subjects can be assigned at random to treatment groups and the response is observed from a sequence of events progressing forward in time.

such cases as concomitant ones or covariables. Technics developed in experimental design to determine whether observed differences between experimental and control groups are due to differing treatments, the covariables, or both, will be found to be useful in the present problem.

In planned experiments, one technic used is to assign subjects at random to treatment groups in the hope that bias will be avoided from these associated sources of variation, known or unknown. But in such instances, it is impractical (and sometimes impossible) to expect randomization procedures to cancel the full impact of all covariables.* In surveys, furthermore, there may not be any opportunity for randomization of subjects to treatment or population groups since the latter already exist. The only freedom of design present is concerned with the sampling process.

When the number of observations at hand is large, however, one can sometimes solve this problem by dividing the data into several homogeneous subgroups and dealing with each of them individually. For example, when age, race, and sex are concomitant variables needing consideration, comparisons between treatment or population groups might be made separately for white males, ages 5–9, 10–14, 15–19, etc.

There are several objections to this type of analysis.

1. When the covariable is a continuous one, such as age, subdivision involves arbitrary decisions for grouping.
2. In most public health problems, numerous observations are not always possible because of the nature of the event. There is a limited frequency of occurrence of certain deaths, diseases, and conditions.
3. Subdivision into small homogeneous groups involves samples larger than may be needed to measure an over-all effect. This is inefficient and costly.

How, then, are these other variables efficiently dealt with so as to eliminate or adjust for their effects upon the measurement under study? A recommended procedure for treating them is called analysis of covariance. It is perhaps expedient to illustrate its use by an example. The illustration will also serve to bring out the assumptions involved in the method.

EXAMPLE

The data for the present example are from a section of a larger study in nutrition, part of which has been reported elsewhere.[1] At one point in that investigation, the question was raised whether the sample of 9- through 11-year-old children in a private urban school were taller than children in the same grades in a rural school. The children reported here from the rural school include only those whose parents were classified as tenant farmers. The data in Table 1 were available to answer this particular question.

The data for the present example involve a main variable (height) and a covariable (age) which are both quantitative, continuous measurements. The technic of covariance may also be extended to cover cases where some of the variables are qualitative and discontinuous.

Although the sample of children for the private school appears to be taller by (144.5–141.7) = 2.8 cm., this difference is not significant since the t-test yields a value of 1.06. On the other hand, examination of the ages of the two groups being compared shows that the rural school children in this sample are about one-half year older on the average. Taking this fact into consideration with the observed height difference might provide a new light on the difference in stature.

ANALYSIS OF COVARIANCE

Analysis of covariance is a recom-

* Randomization will ordinarily remove the bias but not the accompanying variability produced by the associated variables.

TABLE 1

Height and Age of Private and Rural School Children in a Study in North Carolina in 1948

	Private School		*Rural School*	
Students	*Age (x) (months)*	*Height (y) (cm.)*	*Age (x) (months)*	*Height (y) (cm.)*
1	109	137.6	121	139.0
2	113	147.8	121	140.9
3	115	136.8	128	134.9
4	116	140.7	129	149.5
5	119	132.7	131	148.7
6	120	145.4	132	131.0
7	121	135.0	133	142.3
8	124	133.0	134	139.9
9	126	148.5	138	142.9
10	129	148.3	138	147.7
11	130	147.5	138	147.7
12	133	148.8	140	134.6
13	134	133.2	140	135.8
14	135	148.7	140	148.5
15	137	152.0	..	..
16	139	150.6	..	..
17	141	165.3	..	..
18	142	149.9	..	..
Total	2,283	2,601.8	1,863	1,983.4
Mean	126.8	144.5	133.1	141.7
Sums of squares	291,331	377,329.00	248,469	281,478.10
Sums of squares of deviations	1,770.50	1,253.26	556.93	486.99
Sum of products	330,900.20		263,996.20	
Sum of products of deviations	905.23		62.33	

mended method of taking this age difference into account in order to compare stature. The procedure involves a combination of analysis of variance and standard regression technics.

The fact that the children differ in ages has introduced two limitations on the use of the straightforward t-test above.

The first has already been pointed out, viz., that the rural children are older by one-half year. The second and more important one is that the estimate of the standard error of the difference in height used in the preliminary test of significance was inflated because of the varying ages of children in those grades.

To consider the latter limitation first, a more valid estimate of this error can be obtained by measuring deviations from a fitted regression line of height on age instead of the average height in each school. In that way, the rural school child who is, say 140 months old, can be compared in height with an expected height for children of that age (Figure 1) rather than the observed average for the entire group of 9- through 11-year-old children, which was 141.7 cm.

The calculations for this procedure are outlined in Table 2 and will be discussed briefly here. It will be assumed that the steps for converting the data from Table 1 to the form indicated in the left hand side of Table 2, if not readily apparent, can be comprehended by consulting a good text on statistical methods (e.g., see Snedecor [2]).

The preliminary t-test employed the value of 1,740.25 in deriving an estimate of error, this term coming from the "Within schools" line for Sy^2 in Table 2. As pointed out, this term is based upon deviations from the mean in each school. A more valid estimate of error is ob-

TABLE 2

Analysis of Covariance of Stature of Children in Two Schools

Source of Variation	Degrees of Freedom	Sums of Squares and Products			Errors of Estimate		
		Sy^2	Sxy	Sx^2	Adjusted Sum of Squares *	Degrees of Freedom	Mean Square
Total	31	1,805.25	826.42	2,633.87	1,545.95	30	
Between schools	1	65.00	—141.14	306.44			
Within schools	30	1,740.25	967.56	2,327.43	1,338.02	29	46.14
For test of significance of adjusted means					207.93	1	207.93

$$F = \frac{207.93}{46.14} = 4.51$$

$$* \; Sy^2 - \frac{(Sxy)^2}{Sx^2}$$

tained if deviations are measured from the regression lines of height on age in each school. When this is done, the adjusted sum of squares is obtained, the result being equal to 1,338.02. This calculation was designed, therefore, to derive a proper estimate of error to be used in the denominator of the F-test.

The fact that the rural children were slightly older is taken into account when the numerator for the F-test is calculated. It is obtained from the difference between the total adjusted sum of squares and the previously calculated "Within schools" value, i.e.:

$$1,549.45 - 1,338.02 = 207.93$$

The test of significance is then significant at the 5 per cent level since the F value is 4.51.

The estimated difference in stature between the two schools, if age is adjusted to the same age for both groups, is represented by the vertical difference between the two parallel lines shown in Figure 1. This difference can be calculated as follows:

$$\begin{aligned} \text{Private school adjusted mean height} &= 144.5 - 0.42\;(126.8 - 129.6) = 145.7\\ \text{Rural school adjusted mean height} &= 141.7 - 0.42\;(133.1 - 129.6) = 140.2\\ \text{Difference} &= 5.5 \end{aligned}$$

Where $0.42 = \frac{967.56}{2,327.43}$ = common slope of the regression lines of height on age; 129.6 is the mean age of the 32 students from both schools and remaining values are mean heights and ages as reported in Table 1.

The adjustments calculated by the foregoing are represented graphically in Figure 1. Thus, the unadjusted mean height of private school children at point A is transferred to point B. Similarly, the unadjusted mean height of rural children at point C is adjusted to point D. Points B and D are at the average age of all 32 children and there is 5.5 cm. separating these two points.

SPECIAL ASSUMPTIONS UNDERLYING ANALYSIS OF COVARIANCE

In addition to the usual assumptions underlying the analysis of variance,[3] additional ones are introduced here because of the application of regression procedures. These are as follows:

1. ***Linearity of regression of height on age***—The appropriateness of assuming

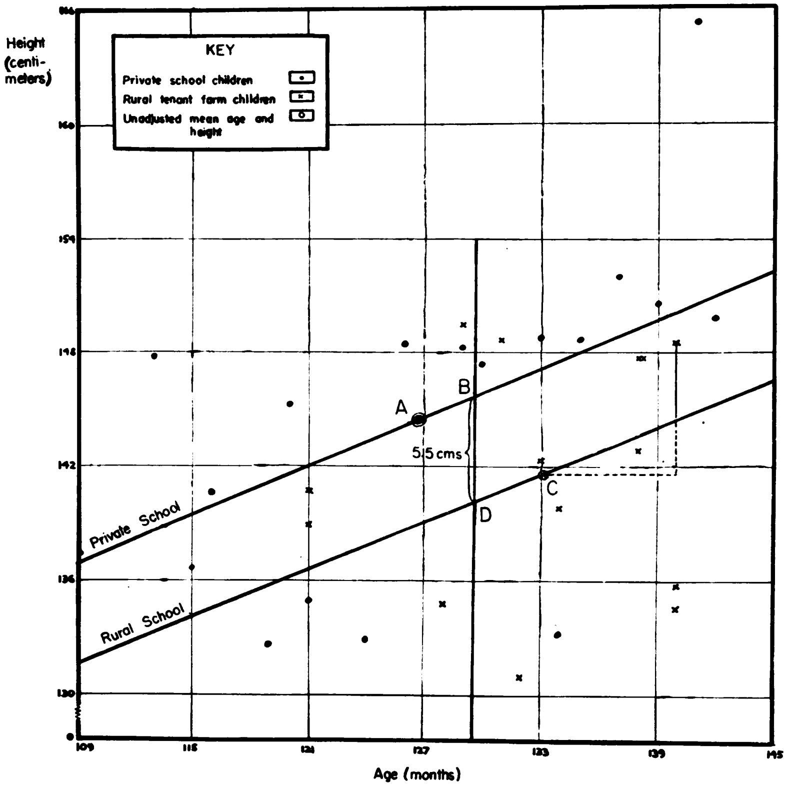

FIGURE 1—Height and Age of Private and Rural School Children in a Study in North Carolina in 1948.

that a straight line relationship exists between height and age should be investigated. For the relatively small interval of three years involved here, departure from linearity does not appear to be a serious shortcoming. For large ranges in age, however, the linearity of the regression should be reviewed and evaluated.

The technic of covariance is by no means restricted to applications where the regression is linear. Even though the relationship might assume a nonlinear form of mathematical curve, adjustments can still be made. This may involve, however, additional calculations similar to the use of more than one covariable. (See below for discussion of several covariables.)

2. *Similarity of regression slopes for each school*—It can be seen in Figure 1, that the deviations in each school are measured from two parallel lines. That is to say, the lines were determined so as to have the same slope. This assumption is necessary if there is to be a constant differential in stature between the two groups over the entire age range. If the two slopes are not identical, then the children in one school might be taller at age 9 but smaller at age 11.

The parallelism of the lines should be tested in every instance. The details of this test will not be discussed here. The test, nevertheless, was performed in this example and the departure from parallelism was not significant.

3. *Homogeneity of variances*—As in other applications of elementary regression technics, it is assumed that the

deviations about the line at age 9 are of the same order of magnitude as those at ages 10 and 11. From the growth-curve literature (Burgess[4]), it is known that this is not completely true. The degree to which it is not, however, is slight in this example and not serious enough to invalidate the inferences. For information on handling situations involving heterogeneity of variances, see Cochran[5] and Bartlett.[6]

4. *Overlapping of ages in both groups*—If the mean ages in the two groups are far apart, adjustment of the mean heights to points B and D at the point of over-all mean height may represent extrapolation. The uncertainties of extrapolation are greater in covariance analysis than in the usual case of regression since it is performed twice. In this example, there was considerable overlapping of the age distributions and no trouble was encountered on that score.

SEVERAL COVARIABLES

The use of more than one concomitant variable is frequently necessary to achieve desired results, particularly to obtain a valid estimate of error. The analysis of covariance permits the use of any number of covariables, provided it is less than the number of degrees of freedom in error.

The question of how many variables to include is identical to that which arises in most multiple regression problems. The decision depends upon a comparison of the expected gain in efficiency versus the cost of gathering that information and performing the additional calculations. Generally speaking, if the partial correlation coefficient of $r_{yx_p \, . \, x_1 x_2 \ldots x_{p-1}}$ is significantly different from zero, the inclusion of variable X_p (in addition to X_1, X_2, ... X_{p-1}) will reduce the magnitude of the denominator in the F-test.

BALANCING

In most planned experiments, the investigator has the opportunity of arranging the subjects so that the average ages, say, in the two groups are the same. This method of design has intuitive as well as popular appeal and is well known in some branches of research as *matching of groups* or *balancing.**

If balancing has value in planned experimentation, it appears at first glance that matching of groups might also be helpful in analytical surveys. The investigator can choose samples in such a way that the average values of the covariables are identical, or nearly so, by restricting the random element in the sampling procedure (e.g., stratified sampling in several stages). For example, a segment from each of two populations might be chosen so that both samples contain the same percentage of males.

If this is a true probability sample for a finite population, use of the correct estimation technic would weight the estimate of the main variable under study according to the true sex ratio in each population and nothing will have been gained.

If interest is centered upon an infinite population, does matching of groups produce a worth-while result? Generally in public health problems, I think not. The justification for this can be demon-

* The procedure of balancing is a tricky one if a random element in the sample is to be retained. In the case where subjects are available before the experiment is started, and they are to be assigned to either an experimental or control treatment group, a recommended procedure for matching the two groups with respect to a covariable is as follows:

Arrange the subjects according to the magnitude of the covariable to be balanced. There are many ways in which the individuals can now be grouped from this linear arrangement so that the term t_{xx} in expression (A) above is small. In many cases, one of the most advantageous groupings is to proceed in units of four subjects. The first of each unit of four persons is assigned at random to either the experimental or control treatment. Having made this initial random allocation, the following two individuals are assigned to the opposite treatment. Finally, the fourth individual is assigned to the same treatment as the first person. Repeat the identical procedure, including randomization, for each unit of four. The sequence appears as a sandwich of four which in symbolic notation can be written as ABBA.

strated by examining the claimed advantages of balancing in planned experiments.

When balancing is used, the reasoning goes that there is no need for adjustment of the covariables by covariance since the two distributions, or at least the first moments of the two distributions, are identical. Unfortunately, this logic is not quite true. Even if one assumes that the relationship is exactly linear so that equivalence of the means is all that is needed, matching of the treatment groups has done nothing about the error term, the denominator of the F-test. In this respect, covariance analysis is of much more value for it enables a valid estimate of the error term to be made.

It so happens, nevertheless, that when covariance analysis is used, there is still a slight additional gain to be achieved in the sensitivity of an experiment by balancing. This gain stems from the effect of balancing upon the numerator of the F-test.

If an equal number of subjects have been allocated to each treatment group at random, it can be shown that the average or expected value of the numerator of the F-ratio in the analysis of covariance is equal to *

$$\text{(A)} \qquad \sigma^2 + n \left\{ 1 - \frac{t_{xx}}{(p\text{-}1)\ S_{xx}} \right\} \Delta_t^{\,2} \ ,$$

where $\sigma^2 =$ true experimental error
$n =$ number of subjects on each treatment
$p =$ number of treatments
$t_{xx} =$ symbol for the sum of squares represented by 306.44 in the present example
$S_{xx} =$ symbol for the sum of squares represented by 2,633.87 in the present example
$(p\text{-}1)\ \Delta_t^{\,2} =$ sum of squares of the true treatment effects

The sensitivity or discriminatory power of a given experiment can be enhanced by maximizing the numerator of the F-ratio, or making $\dfrac{t_{xx}}{(p–1)\ S_{xx}}$ as small as possible. The effect of balancing is to make t_{xx} essentially equal to zero.

The term $\dfrac{t_{xx}}{(p–1)\ S_{xx}}$ has therefore been suggested by Lucas [7] as a measure of the loss in sensitivity due to failure to balance. It has an average or expected value equal to $\dfrac{1}{np–1}$. Also, 95 per cent upper limits for the loss in sensitivity due to failure to balance can be calculated. These values have been calculated in Table 3 for the special case where $p = 2$. It can be seen that the sensitivity loss is trivial (0.05) when there are about 20 subjects in all and only two treatments. With less than that many subjects, the loss becomes more substantial.

TABLE 3

Loss in Sensitivity of an Experiment with Two Treatments Due to Failure to Balance

No. of Subjects on Each Treatment	*Average Loss in Sensitivity*	*Five Per cent of Time Loss in Sensitivity Exceeds*
2	0.33	0.90
3	0.20	0.66
4	0.14	0.50
..	..	..
..	..	..
..	..	..
10	0.05	0.20
20	0.03	0.09
30	0.02	0.06
40	0.01	0.06

Thus, in cases where covariables exist, the preferred procedure is always covariance analysis. If the study consists of less than 10 subjects on each of two treatments, a slight gain in efficiency may be obtained by the technic of balancing.

* It has been assumed that the true mean of the covariable is identical in each of the treatment groups.

SUMMARY

In drawing inferences from surveys of population groups, it has been demonstrated how the effect of covariables upon the measurement under study can be taken into account by the use of covariance analysis. The assumptions underlying the use of the method were pointed out.

The advantage of balancing or matching of groups with respect to a covariable has been compared with that of covariance. The latter is the preferred procedure, since it enables a valid estimate of error to be made as well as removing any bias. In addition to covariance, however, balancing has further merit when surveys involve less than 10 subjects on each of two treatment groups. If many covariables exist, covariance analysis may be inconvenient to use for all of them, in which case balancing of the remainder will usually remove a considerable portion of the bias.

REFERENCES

1. Bryan, A. Hughes, and Greenberg, B. G. Methods for Studying the Influence of Socio-economic Factors on the Growth of School Children—Body Measurements. *J. Elisha Mitchell Scientific Soc.* 65:311–314 (Dec.), 1949.
2. Snedecor, George. ***Statistical Methods.*** Ames, Ia.: Iowa State College Press, 1946.
3. Eisenhart, Churchill. The Assumptions Underlying the Analysis of Variance. ***Biometrics*** 3:1–21 (Mar.), 1947.
4. Burgess, May Ayres. The Construction of Two Height Charts. *J. Am. Statist. A.* 32:290–310, 1937.
5. Cochran, W. G. Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied. ***Biometrics*** 3:22–38 (Mar.), 1947.
6. Bartlett, M. S. The Use of Transformations. ***Ibid.*** 3:39–52 (Mar.), 1947.
7. Lucas, Henry L., Jr. ***Design and Analysis of Feeding Experiments With Milking Dairy Cattle.*** Raleigh, N. C.: Institute of Statistics Mimeograph Series No. 18.

Medical Health Officer Examination

The U. S. Civil Service Commission announces an examination for medical officers to fill positions in various specialized fields, with salaries ranging from $5,940 to $10,800. The positions are principally in the Bureau of Indian Affairs located on reservations west of the Mississippi and in Alaska with a few positions in the Fish and Wild Life Service. Further information from U. S. Civil Service Commission, Washington 25, D. C.