pavement creates an adverse lateral drift, the degree of severity being dependent upon speed and the effective length of wheel contact with the side and top surfaces which increases in direct relation with the skew angle. Skewed placements create a real operational hazard, especially when considered with wet or icy travel conditions and should not be employed.

6. Rounded noses or ends of the individual units eliminates the possibility of sharp-corner contact and presents a more-pleasing appearance.

RECOMMENDATIONS

The tests further established an acceptable combination of the variable conditions, based upon physical sensations and their effectiveness as a warning of encroachment: (1) a spacing of approximately 12.5 feet center to center of individual vibrator units; (2) a 1-inch height of section; (3) a face slope of 30 deg.; (4) a length of base of 6 inches; (5) placement of the individual units normal to the centerline; and (6) rounded noses or ends of the individual units.

Statistical Methods Applied to Highway-Research Experimentation

DAVID B. DUNCAN
Virginia Agricultural Experiment Station
Virginia Polytechnic Institute

THIS paper demonstrates the application of three or four basic statistical principles to a hypothetical problem in testing pavement design. An experiment is designed in which a new type of pavement is compared with a standard one, and a new type of subbase is compared with a standard subbase. The experiment is planned for three locations. Simple hypothetical data consisting of maintenance costs in dollars per mile of highway per year are used for illustrative purposes and an analysis of the data is given in some detail.

One valuable property of this experimental design is that the results not only afford the required comparisons between the new and standard types of pavement and subbase, they also afford precise measures of the reliabilities of these comparisons. This is a most-important feature which receives much emphasis in the discussion and which is characteristic of most statistical methods.

Another useful feature of the experiment is that the reliability of each of the desired comparisons is maximized. This same feature incorporated in other experiments allows the experimenter to trim his plans down to the very minimum required to afford comparisons of a given reliability, or put in another way, it enables him to maximize the reliability of the comparisons for a given cost to be incurred.

Similar advantages afforded by statistical designs are also indicated. These methods and theory are largely concerned with the problems of designing and analyzing experiments and may be of considerable value in the field of highway research.

● A LARGE body of theory and methods for conducting scientific experiments has been developed within the last 20 or 30 years under the name of *Statistical Methods for Research Workers*. Many fields of scientific endeavor are putting this work to good use with considerable advantages. This is so in experimental work in agriculture, biology, medicine, and in many branches of engineering. Due to the recent nature of this work, however, these methods are not yet being fully exploited in many problems in which they would be of great value, and this is no doubt true of some projects in highway research.

In the following discussion we will demonstrate the application of three or four basic statistical principles to a hypothetical problem in pavement construction. This will serve best in a short paper of this nature to indicate how statistical methods in general may be of use in the field of highway research.

HYPOTHETICAL EXAMPLE

Suppose it is desired to compare a new type of pavement, which we will denote by p_1, with a standard type of pavement, p_0, and also a new type of subbase, s_1, with a standard type of subbase, s_0. We will assume that the comparisons are to be made on the basis of, say, *highway maintenance cost in dollars per mile of highway per year*. Figure 1 shows a statistical design (not to be confused with a construction design) known as a *2 × 2 Factorial Design* which might be chosen for such purposes. Hypothetical cost results are also included in the figure for the sake of illustration.

The figure indicates an arrangement for four experimental sections at each of three locations. The symbols under each section denote the construction for each section. For example, the symbols p_0s_1 in Section 2 at Location 1, denote the construction of standard pavement, p_0, on new type subbase, s_1.

The aims of the experiment may be expressed in three parts as follows: (1) to decide whether there is any appreciable change in maintenance costs due to the use of the new pavement instead of the standard pavement; (2) to decide whether there is any appreciable change in maintenance costs due to the use of the new subbase instead of the standard subbase, and (3) if there are appreciable changes in Items 1 or 2, to obtain an order of merit for the various combinations of pavement and subbase.

The average maintenance costs for each type of construction, for each location, and for the whole experiment are as shown in Table 1.

Now, considering the results for types of construction in more detail, we see that there are reductions in maintenance costs due to using the new type of pavement p_1 instead of the standard type p_0 and these may be summarized as shown in Table 2. Also, there

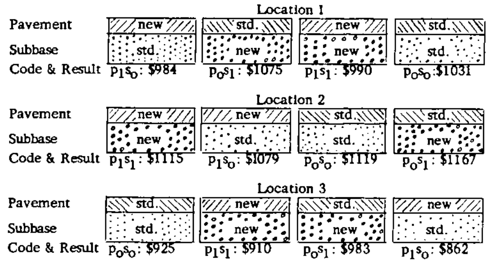

Figure 1. Hypothetical experiment: Statistical design and results (maintenance costs).

TABLE 1
AVERAGE MAINTENANCE COSTS
(In Dollars)

Type of Construction		Location	
p_0s_0: standard pavement, standard subbase	1,025	1	1,020
p_0s_1: standard pavement, new subbase	1,075	2	1,120
p_1s_0: new pavement, standard subbase	975	3	920
p_1s_1: new pavement, new subbase	1,005	Overall Mean	1,020

TABLE 2
REDUCTIONS IN COST DUE TO NEW PAVEMENT

a) When used with new subbase	$70
b) When used with standard subbase......	$50
Average reduction......................	$60

are increases in cost due to using the new subbase which may be summarized in a similar way as shown in Table 3.

The reliabilities of the comparisons in Tables 2 and 3 may be assessed by a statistical method termed an *analysis of variance*. For example, consider the difference of $20 between the cost reduction due to the new pavement when it is used with the new subbase and when it is used with the standard subbase. In statistical terminology this difference is called the *interaction* between types of pavement and types of subbase. (It will be observed in Table 3 that the difference between the cost increase due to the new subbase when it is used with the new pavement and when it is used with the standard pavement is also $20. This interaction is actually identical with the interaction in Table 2, since $(p_0s_1 - p_1s_1) - (p_0s_0 - p_1s_0)$, giving $70 - 50 = 20$, is algebraically identical with $(p_0s_1 - p_0s_0) - (p_1s_1 - p_1s_0)$, giving $50 - 30 = 20$.) From

TABLE 3
INCREASES IN COST DUE TO NEW SUBBASE

a) When used with new pavement	\$30
b) When used with standard pavement	\$50
Average increase	\$40

TABLE 4
ESTIMATES AND ERRORS
(Maintenance Costs in Dollars)

(1) Location	(2) Type of Construction	(3) Observed Result	(4) Estimate	(5) Error
1	p_0s_0	1031	1025	6
	p_0s_1	1075	1075	0
	p_1s_0	984	975	9
	p_1s_1	990	1005	−15
2	p_0s_0	1119	1125	−6
	p_0s_1	1167	1175	−8
	p_1s_0	1079	1075	4
	p_1s_1	1115	1105	10
3	p_0s_0	925	925	0
	p_0s_1	983	975	8
	p_1s_0	862	875	−13
	p_1s_1	910	905	5

an analysis of variance it is established that this interaction is *not significant.* That is, speaking roughly, it is probably nothing more than a random fluctuation in the results to which no importance should be attached.

Because of this, the reduction in cost due to the new pavement and the increase in cost due to the new subbase may be satisfactorily measured in the form of the averages, \$60 and \$40 respectively, shown at the bottoms of Tables 2 and 3 respectively. (If the interaction were significant the cost reductions \$70 and \$50 would be considered separately, and so would the increases \$30 and \$50.)

On testing the significances of the two averages \$60 and \$40 in the analys s of variance each one is found to be *significant* Broadly speaking, this means in the case of the cost reduction due to the new pavement, for example, that the new pavement can be expected to have lower maintenance costs in general than the standard pavement under the conditions represented by the experiment.

In this paper it will not be possible to give the complete details and methods of the analysis of variance which makes the careful assessment of the significances of the interaction and the main comparisons possible. The basic principles involved, however, are indicated in the following discussion.

ASSESSMENT OF ERROR AND TESTS OF SIGNIFICANCE

From the averages in Table 1 we can calculate an estimate corresponding to each result observed in the experiment. These estimates are shown in Column 4 of Table 4. The estimate corresponding to any result is obtained by taking $C + L - M$ where C represents the average for the type of construction involved, L represents the location average involved and M is the overall average. For example, for p_0s_1 at Location 3, we have $C + L - M = 1075 + 920 - 1020 = 975$. These estimates represent, in a sense, the results one might have expected if there were no experimental errors present. Consequently, the deviations, observed results minus estimates, give measures of errors in the experiment and these are shown in Column 5 of Table 4.

These errors are combined into one overall assessment of error called the *standard deviation*, s, using the formula $s = \sqrt{S/n}$, where S is the sum of the squares of the errors (called the *error sum of squares*) and n is the number of *degrees of freedom* among the errors. From the way the errors are calculated it is clear that they are not independent; for example, the fourth error at each location is determined by the preceding three, since the method of calculation forces all four to add to zero. Broadly speaking, the degrees of freedom n, represent the number of independent errors to which the given errors are equivalent. In designs of this type the number of degrees of freedom is $n = (r - 1)(c - 1)$, where r is the number of types of constructions and c is the number of locations; thus, in the example, $n = (4 - 1)(3 - 1) = 6$.

Proceeding with the example, we get $S = 6^2 + 0^2 + ... + 5^2 = 816$ and thence $s = \sqrt{816/6} = \sqrt{136} = 11.66$. This represents the best estimate of the standard deviation of any of the experimental results shown in Column 3 of Table 4. More important, we can now assess the *standard errors* of the various differences between means in which we are interested. The general formula for the standard error s_L of any linear combination

$$L = k_1x_1 + k_2x_2 + ... + k_{12}x_{12} \quad (1)$$

where $k_1, k_2, ..., k_{12}$ are any constants and

$x_1, x_2, ..., x_{12}$ are the twelve observed results, is given by

$$s_L = \sqrt{(k_1^2 + k_2^2 + ... + k_{12}^2)s^2} \quad (2)$$

where s is the standard deviation of a single result as before. For example, the interaction \$20 above can be written as a linear function of the form (1) in which six of the k's are $1/3$ and the other six are $-1/3$. Thence $\Sigma k^2 = 12(1/9) = 4/3$ and the standard error of this difference is $\sqrt{(4/3)s^2} = \sqrt{(4/3)\ 136} = 13.46$. Likewise, the \$60 reduction in cost due to the new pavement, or the \$40 increase in cost due to the new subbase, can be written as a linear function of the form (1) in which six of the k's are $1/6$ and the other six are $-1/6$. Thence the standard error of either difference is $\sqrt{12(1/36)s^2} = 6.73$.

To assess the significance of a difference, it is divided by its standard error to give a *"Student" t ratio with n* degrees of freedom. Tables[1] are available which give *α% significant values* of this ratio; an α% significant value $t(\alpha\%)$ being defined as that value of t which has an α% chance of being exceeded without any real difference being present at all. With $n = 6$ degrees of freedom, the tables give: $t(20\%) = 1.44$, $t(10\%) = 1.94$, $t(5\%) = 2.45$, $t(1\%) = 3.71$ and $t(0.1\%) = 5.96$. For the interaction above, $t = 20/13.46 = 1.49$. This barely exceeds $t(20\%) = 1.44$ which means that the probability of getting the observed result (\$20) by chance alone is almost 20%. This probability is so high that the interaction may be considered not significant. On the other hand, the t values for the reduction in cost due to the new pavement and the increase in cost due to the new subbase are $60/6.73 = 8.91$ and $40/6.73 = 5.94$; the first exceeds and the second almost exceeds $t(0.1\%) = 5.96$ and both may be considered highly significant.

CONCLUSIONS FROM THE EXPERIMENT

To summarize, the main conclusions are that, under the conditions represented by the experiment, (1) the new pavement has a lower maintenance cost than the standard pavement, and (2) the new subbase has a higher maintenance cost than the standard subbase. It has been further concluded that these two effects are independent, i.e., the reduction in maintenance costs due to the new pavement is the same when used with either type of subbase.

[1] For example, Table B in A. Duncan (1952).

Other things being equal, the order of merit for the various types of construction is (1) new pavement with standard subbase; (2) new pavement with new subbase; (3) standard pavement with standard subbase; and (4) standard pavement with new subbase.

STATISTICAL PRINCIPLES

Various general statistical principles which may be of use in many experiments are illustrated in the above example.

One of these is the *factorial* principle of testing each "level" of a given factor at each level of the other factor. This allows all of the results to be used in assessing the importance of changes of level in each factor, and also allows a similarly optimum assessment of interaction. It is not always possible to design experiments in this way, but in many cases the incorporation of this principle is of great value.

Another principle is the use of *replication*, that is, the repetition of the basic design at least twice. Only by replication is it possible to get measures of error of the type indicated and, thence, the consequent assessments of the significances of the results. In the above example, three replications are used, one at each of three locations. By increasing the number of replications one can expect to reduce the standard errors of the comparisons and interactions, thus increasing the precision of the experiment. *The number of replications in an experiment should be such that the benefit of any additional precision would be counterbalanced by the cost of the additional replication required to achieve it.*

A further feature in this and many other statistical designs is the use of *randomization*. The validity of the tests of significance in the above experiment is dependent on the requirement that the four types of construction at each location be assigned entirely at random to the four test sections involved and it has been assumed that this was done. If a nonrandom scheme were used in which the types of construction were consistently laid down

TABLE 5
RESULTS FOR EXPERIMENT WITH LARGER ERRORS
(Maintenance Costs in Dollars)

(1) Loc.	(2) Type of Constr.	(3) Obs. Result	(4) Est.	(5) Error	(6) *t* ratios
1	p_0s_0	1049	1025	24	For interaction 20/53.84 = .37
	p_0s_1	1075	1075	0	
	p_1s_0	1011	975	36	
	p_1s_1	945	1005	−60	
2	p_0s_0	1101	1125	−24	For pavement 60/26.92 = 2.23
	p_0s_1	1143	1175	−32	
	p_1s_0	1091	1075	16	
	p_1s_1	1145	1105	40	
3	p_0s_0	925	925	0	For subbase 40/26.92 = 1.49
	p_0s_1	1007	975	32	
	p_1s_0	823	875	−52	
	p_1s_1	925	905	20	

in the order p_0s_0, p_0s_1, p_1s_0, p_1s_1, for example, unknown biasses might enter into the result. If sections were being laid concurrently at each location, a change in the weather could occur which could give sections already laid some advantage over those that remain. The averages for p_0s_0 and p_0s_1 might thus get favorable biasses in this way. A change in workmanship during the course of construction could have a similar effect, and many other known or unknown factors could vitiate the results of the experiment in the same manner. When designs are randomized, due allowances for the effects of factors of this type are made in the tests of significance.

The tests of significance with the above data are only limited examples of a large section of useful statistical theory entitled *statistical inference*. To understand more of the importance of these tests it may be instructive to consider the following example:

Suppose that the errors (Column 5, Table 4) in the above experiment were say four times larger, without the averages for each type of construction and for each location (Table 1) being changed. This would happen if the observed results were as shown in Table 5, Column 3, in which the estimates, Column 4, are the same as those of Table 4 and the errors, Column 5, are four times those of Table 4.

Working from the new errors it is readily seen that each of the standard errors for the tests of significance would be four times larger. At the same time the differences (Tables 2 and 3) to be tested would be unaltered, hence the t ratios for the tests would be four times smaller as shown in Column 6 of Table 5.

The largest of these does not exceed $t(5\%) = 2.45$, hence none of the results would be significant. Thus, even though the averages are the same as before, the presence of higher experimental errors would make it too uncertain to conclude that any one type of construction is better than any other.

A common practice in many applications of these methods is to class differences as significant or not significant depending on whether or not they exceed $t(5\%)$ or $t(1\%)$.

The choice of the critical value in a rule like this should be made keeping in mind the relative seriousness of (1) the error of deciding that a new type of construction is an improvement when it isn't and (2) the error of deciding that a new type of construction is not an improvement when it is. If a conservative critical value, say $t(0.1\%)$ were used this would give good protection against errors of the first kind but poor protection against errors of the second kind. If a more liberal value of t, say $t(20\%)$, were used this would protect much less against errors of the first kind and much more against errors of the second kind.

CONCLUDING REMARKS

The foregoing hypothetical experiment and tests of significance are but limited examples of a large range of statistical methods and principles which are available. It is emphasized that much of the modern theory and methods of statistics are not concerned with routine tabulation and calculation problems as is often assumed. The modern methods and theory are largely concerned with the problems of designing and analysing experiments and as such may be of considerable value in the field of highway research.

For text books on this subject the reader is referred to Duncan (1952) or Davies (1949) for a general discussion and to Fisher (1951) and Cochran and Cox (1950) for a fuller treatment of experimental designs. For readers interested in statistical theory a good text is Anderson and Bancroft (1952) or Mood (1950).

REFERENCES

1. Anderson, R. L. and Bancroft, T. A., Statistical Theory in Research, McGraw-

Hill Book Company, Inc., New York, 1952.
2. COCHRAN, W. G. AND COX, G. M., Experimental Designs, John Wiley and Sons, Inc., New York, 1950.
3. DAVIES, O. L., Statistical Methods in Research and Production, two editions, Oliver and Boyd, Ltd., London, 1947–1949.
4. DUNCAN, A. J., Quality Control and Industrial Statistics, Richard D. Irwin, Inc., Chicago, Illinois, 1952.
5. FISHER, R. A., Design of Experiments, six editions, Oliver and Boyd, Ltd., London, 1935–51.
6. MOOD, A. M., Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, 1950.

DISCUSSION

PAUL IRICK, *Research Associate, Statistical Laboratory, Purdue University*—It should not go unnoticed that statistical methods have been applied to many of the investigations reported in the literature of highway research. A somewhat-casual survey of this literature reveals that use has been made of various significance tests and estimation procedures, sampling techniques, analysis of variance and covariance, multiple regression analysis, and other statistical methods. However, it appears that of the reported problems which seem to be statistical in nature, only a minority have actually made use of statistical designs and analyses. The preponderance of this usage has been in the area of traffic engineering rather than in areas of laboratory research such as materials testing. Many of the investigations in highway research are really factorial experiments and are therefore candidates for the use of statistical designs.

Duncan's relatively simple but well-chosen example has brought most of the basic aspects of statistical methodology to the reader's attention. He has ably shown that the principles involved are not difficult to comprehend when they are amply explained. It should be clear that the major concern of this paper is not to solve the type of problem presented by the illustrative example, but rather to promote a more-widespread use of statistical methods whenever it is clear that the latter can be appropriate and beneficial. We shall devote our discussion to three issues: (1) the benefits that may be expected from the use of statistical methods, (2) the major obstacles to the use of such methods, and (3) possible means for reducing the effect of such barriers.

BENEFITS FROM STATISTICAL METHODS

Statistical analyses require the experimenter to state explicitly just what hypotheses he is testing and what assumptions he has had to make concerning his measured variables. These details are in accord with the scientific method, wherever it is used, and may possibly remain obscure or uncertain when statistical methods are not used. Duncan did not give these details for his example, but he could have done so, and his assumptions are implied by the way in which his analysis was carried out. As is the case with the example, appropriate statistical procedures will be in line with commonsense directives. Statistical methods make it possible to estimate the amount of data needed to achieve desired information, and they provide for an optimum allocation of the experimental dollar.

Another benefit accrues from the use of statistical methods when the experimental data are summarized so that inferences can be drawn and conclusions reached. For example, if analysis of variance is used or if curve-fitting results are accompanied by a regression analysis, then the major results of the investigation may be displayed in a compact form, no matter how complex the experimental design may have been. And so we should expect a more-uniform method of presentation of results when statistical methods are used. As soon as the consumer of such reports becomes technically acquainted with these displays, he can readily realize what has happened in the data. A single line in an analysis-of-variance table may indicate that a whole series of plotted curves may be regarded as parallel or even coincident, whether the curves are actually shown or not. Again, from a regression-analysis table it may be learned whether or not a fitted curve should be interpreted as indicating any real relationship between the variables involved.

Once a statistical design has been chosen, the analysis of the data will be objective to the point that independent experimenters will go through the same procedures to reach inferences and conclusions. Although it is true that nonstatistical inferences may lead to the same conclusions and decisions as statistical inferences, one of the major distinctions between the two is that valid statistical inferences are accompanied by known risks of making incorrect conclusions. In the case of nonstatistical inference, there is also the probability for incorrect decisions, but usually little is known concerning the magnitude of this probability in specific cases. In his paper Duncan has discussed the bases for choosing the α levels of risk which may be appropriate to a particular inference.

And so, statistical methods eventually provide for conciseness and objectivity in the analysis of experimental data in any area of research. Standardized procedures, such as are specified by ASTM and other organizations, have somewhat the same aims, but perhaps do not encompass the experiment as a whole nor always take into account that any one experiment is generally but a component part of a still-broader investigation.

MAJOR OBSTACLES

It is our observation that statistical methods will seldom be used whenever the experimenter is not able to recognize at what point and to what extent his problem is statistical in nature. Duncan has pointed out that for nearly half a century the body of statistical methodology has increased to meet the needs of experimental research, especially in the fields of agriculture, biology, and psychology. During the last 15 years statistical quality control has become an important phase of production engineering. One thing that is more or less common to these listed fields is the presence of relatively large amounts of variation which cannot be assigned to the factors controlled by the experimental techniques. It is a pretty sure bet that the experimenter has a statistical problem whenever the "error" variation approaches the magnitude of the variations expected from experimental factors. Although it is true that unassignable variations are negligible in some types of engineering research, one has only to look, for example, at the flexural strengths of a few concrete beams of "identical composition" or at a few skid distances under "identical test conditions" to decide that the random variations of data in highway experiments can be quite appreciable.

Even if the experimenter is pretty sure that his problem has statistical aspects, he may assume, more or less rightfully, that his experimental procedure will coincide with some statistical design and consequently not bother to see if a more-efficient design is available. The nonstatistician engineer might also take the point of view that the nonengineer statistician could not possibly appreciate the engineering aspects of his problem, and be no more inclined to turn to statistical methods than to pharmaceutical methods, say.

Another obstacle arises whenever the experimenter decides that the difficulties in understanding statistical designs and analyses are too great when weighed against the time it would take to become familiar with these methods. He may choose to regard statistical procedures as being completely mysterious and place them in his interesting-if-true file. Not only may he be unable to make statistically derived results intelligible to his readers and backers, but he runs the risk of misusing and misinterpreting statistical designs and analysis himself. An experimenter who would use statistical methods must ordinarily be prepared to understand more-complicated techniques than those which are involved in the 2-by-2-factorial experiment of Duncan's paper. In general, the experimenter may be unable to fully appreciate the language of statistical inference and therefore arrive at conclusions not to be inferred from his analysis. On the other hand, the statistician may leave the conclusions in terms that are technically correct but which are more or less meaningless to the experimenter and his public. For example, the word "significant" in this paper refers to the inference, say, that there really is some difference, perhaps in the neighborhood of \$60, in mean maintenance costs between old and new pavement types at many similar locations. Later Duncan shows that the same observed difference would not be called significant under the conditions of inflated errors. Although the example makes it clear that the significance of differences is determined from the standard errors involved, the practical man might declare \$60 per mile per year to be significant or not, independently of any statistical analysis. This difficulty arises because the statistical meaning of significant is not in accordance with its connotation of worthwhile.

The most-serious barrier to a more-widespread use of statistical methods may be the inability of the experimenter and the statistician to coöperate closely from the time of the formulation of the hypotheses until the time that conclusions are reached. By implication, the paper points out that the principles of randomization and replication must be applied to the particular statistical design selected for an experiment. This obviously means that the choice of design and the procedures for taking the data must be planned in full view of the purpose of the experiment and the analysis which is to achieve this purpose. A common misunderstanding with respect to the use of statistical methods arises when it is supposed that there will necessarily be a valid statistical analysis available whatever may have been the experimental procedure. Even when the data are taken in accordance with a particular design, it can be that a different design would be more in accordance with the hypotheses the experimenter is testing and the assumptions he must make.

A vital step in the planning stage concerns the variable which is to be measured. Had the illustrative example of this paper been an actual one, one would have to consider at some length how many years' experience should go into a datum value. Perhaps time itself should be a factor in the design. One would have to decide whether it is correct to assume the nonexistence of interactions of locations with construction type as Duncan has done. If an effective statistical evaluation of experimental results is to be made, both the experimenter and the statistician must understand the purpose of the experiment, the experimental procedure, and the interpretation of the results.

MEANS FOR REDUCING OBSTACLES

The obstacles we have just listed do not really exist if the experimenter himself has had a rather thorough training in experimental statistics. There are several ways in which such training can take place. The research worker may study statistical methods during the course of his academic training or take advantage of the short courses in statistics which are offered by several universities from time to time. Staff seminars in any organization can become vehicles for the group study of statistical methods. Duncan's list of references provides a thoroughgoing curriculum in statistical methods for those who can become self-taught.

If the experimenter himself cannot become a statistician on the side, then the obstacles we have discussed can be minimized only as soon as he becomes operational-research minded. This is to say that the experimenter and the statistician become a team, complementing each other's understanding of the problem. As each becomes more confident of the other's ability to see the experiment in its true light, both become more open minded and willing to modify experimental procedures to any extent that does not jeopardize the purpose of the experiment. Both members realize that if the experiment is to benefit from statistical procedures, then the teamwork must be in evidence at each stage of the experiment. The experimenter can expect that the statistician will neither want to become involved in investigations which are not statistical in nature nor will he try to force an experiment into a statistical design indiscriminately just to ply his wares. The statistician can be expected to do his best to use existing statistical methods only where they are warranted. If appropriate methods are unknown to the statistician, he must either do some consulting himself or evolve new techniques of statistical methodology. Although either the statistician or the experimenter may blunder upon occasion, it can be expected that the team will perform effective experiments which lead to correct decisions with an economy of time and money.

Our final suggestion is relevant to the difficulty in making statistical analyses and conclusions less mystical to the readers of the report of the experiment. If the simpler techniques, and gradually more-complicated ones, were explained in the context of their application, just as Duncan has done with his illustrative example, it would not be long before many of the commonly used statistical procedures would become clear to all individuals working in a particular area. The journals in several fields of experimental research have reached this stage by virtue of somewhat similar processes.

It appears that as statistical methods are introduced in a given area, their use and understanding will necessarily pass through an evolution quite similar to the historical development of the statistical methods themselves.

R. B. Goode, *Highway Finance Research Associate, Virginia Council of Highway Investigation and Research*—L. H. C. Tippett, a British statistician, has written: "Engineers have not generally regarded themselves as needing statistical methods, but in recent years an increasing number have realized how useful the methods may be, and have applied them." This statement could be extended to include many nonengineers active in the highway field.

In this interesting paper, Duncan has reiterated some basic statistical principles which could be of immense value to many persons engaged in highway research. It is to be regretted that the limitation of time restricted the application of these principles to a single, though very pertinent, example. This in no way detracts from the paper, yet thereis reason to believe that the full import of this paper may not be grasped by those who could benefit most by using these and other statistical methods. Hence, I feel it is desirable to provide a few additional illustrations.

The use of a factorial design enables a competent observer to allocate to the factors under study, the variations attributable to each of them, *per se*, as well as indicating the extent of the intereaction between various combinations of factors.

Such a design might be used in an experiment to ascertain the durability and lowest annual cost for highway signs. The factors would be the different reflective materials and the various types of sign backings. Further, the combination of a factorial design with other principles could be adapted to a project directed toward determining the effects of various axle or wheel loads on pavements of different thicknesses.

Replication enhances the precision of any statistical result and that precision grows as the number of repetitions is increased. Single observations must be handled cautiously, but a great number of observations of the same phenomenon will normally be much-more representative. Perhaps the major advantage of repeated observations is that the effects of variations induced by uncontrolled factors can be minimized.

In determining the average compressive strength of concrete, a few persons might be satisfied with testing a single specimen. But, undoubtedly, a much better result could be obtained if several cores or samples were tested. An attempt to find the average life of a certain type of pavement probably would not be very accurate if only one section were studied. Utilizing several sections would be preferable and would provide more-valid results.

Consciously or unconsciously, bias of one sort or another enters most experimental procedures. Perhaps the simplest and best means of avoiding the adverse effects of such bias is through the randomized selection of test items. (Haphazard selection is not the same as random selection.) The use of random sampling is fairly well accepted in some highway-research fields, notably traffic surveys. In such instances, much is learned of a large group by observing only a relatively small portion of that group. This procedure justifies itself through savings of time and money without substantial sacrifice of accuracy. But the random choice or location of test items may also be applied in other areas.

Many testing and quality control procedures have been improved by the incorporation of randomization. The wider application of this principle in these fields seems desirable. Similarly, traffic compaction of flexible pavements might be observed by testing cores from randomly located pavement sections.

The judicious application of statistical methods can facilitate a scientific analysis and permit the assessment of error. However, mere determination of error is of little value unless the resultant variations are tested for their significance or lack of significance. These tests furnish an objective basis for decision making and they do so with substantial economies. As Duncan has stated, this is the great merit of statistical methods. Further application of these methods to highway problems is not only warranted, it is imperative.

D. B. Duncan, *Closure*—Both Irick and Goode have made pertinent and helpful contributions. The additional illustrations provided by Goode and the discussion by Irick of the benefits, obstacles, and possible ways of circumventing obstacles to the use of statistical methods are very relevant and are much appreciated.