The Use of Multiple Regression and Correlation in Test Road Data

P. E. Irick, *Statistical Laboratory, Purdue University*,
R. C. Leathers, *Bureau of Public Roads*,
R. A. Lill, *American Trucking Associations*, and
W. L. Mertz, *Bureau of Public Roads*

There are certain questions that may be asked in nearly any experimental investigation. In this paper seven such inquiries are listed and discussed as they might pertain to an experimental road test:

1. What governs the selection of dependent and independent variables?
2. How may mathematical models be conceived for expressing the experimental conclusions?
3. What levels of the independent variables need to be investigated?
4. What techniques may be used for estimating constants in the mathematical models?
5. How precisely are the estimated constants determined by the experiment?
6. How valid is the mathematical model?
7. How far, in space and time, may the results of the experiment be generalized?

From a most general point of view all experimental findings can be characterized by mathematical models, and much of this paper is concerned with the development of this concept. None of the extemporaneous models used in this paper, however, should be considered as representing any more than possible points of departure for further efforts to interrelate the many experimental results from a road test.

The attitude has been taken that the experimental objectives, the design of the experiment, and the analysis of the experimental data are so closely connected that none of these three aspects of the experiment can be adequately described without detailed reference to the other two.

With this in mind, the paper outlines a possible analytical approach, assuming that the over-all experimental objectives are to produce certain relationships between observed behavior variables and those variables which are either set by the experimental design or which arise through the forces of nature during the course of the experiment.

An effort has been made to give a somewhat detailed discussion of the concepts and roles of randomization and replication.

Data from the WASHO road test have been used to demonstrate analytical techniques rather than to produce conclusions from the illustrative data, for little effort has been made to refine those analyses which have been used in the examples.

● NEARLY any dependent variable in experimental research is a function of more than one independent variable. In many cases it is appropriate or even necessary for the experiment to produce an overall functional relationship among the many variables. It is the purpose of this paper to demonstrate multivariate methods of analysis with respect to data that might arise in the course of an experimental test road. For numerical illustration, certain data have been chosen from the WASHO road test with the aim of clarifying techniques of analysis, rather than for the purpose of reaching conclusions which necessarily follow from the WASHO data. Much of the discussion of the analytical methods may be found in textbooks on experimental design and analysis (*1*, *2*), but in this paper the procedures are pointed at the setting of a road test.

In order to start on familiar ground, we shall briefly discuss a hypothetical example which is not directly related to highway research. It is intended that this example will produce certain situations which may parallel some of those found in road test research. The example will also serve to raise several basic questions to be answered in the course of the paper.

A hypothetical example: Let us suppose that the law of gravitation,

$$F = 6.66 \times 10^{-8}\, m_1 m_2/d^2$$

were unknown, and that three different experimenters were asked to determine the functional relationship of F with m_1, m_2, and d.

The first experimenter might decide to fix m_2 and d, then take observations on F for several values of m_1. The rest of his experiment would consist in observing the relation of F to m_2 while m_1 and d were held fixed, then the relation between F and d for fixed values of m_1 and m_2. The results of this experiment might be shown graphically as in Figure 1, where we may suppose that the experimenter deduced empirical values for the constants a, b, and c. It may be noted that the curve in Figure 1c has the wrong equation. This error is understandable since the functions c/d (incorrect) and c/d^2 (correct) have graphs which are quite similar in appearance.

We shall assume that the second experimenter was more cautious, choosing to fix each pair of independent variables in several combinations while studying the relation of F to the remaining independent variable. The experimental results might have turned out as shown in Figure 2, where again we suppose that the experimenter was able to deduce values for the indicated constants. These plots show that there are interacting effects of the independent variables upon F. For example, Figure 2*a* shows that the relation of F to m_1 is not the same for all values of m_2 when d is held constant. Thus, the second experiment not only yielded more information, but served to point out that the conclusions from the first experiment were quite limited in view of the interacting effects of m_1, m_2, and d upon F. The second experimental results may have great utility, especially if interpolation and extrapolation are permitted in Figure 2, but the fact remains that neither of the first two experiments produced the law of gravitation.

Suppose that the third experimenter took a more general view of the problem and interpreted his assignment to mean that he must

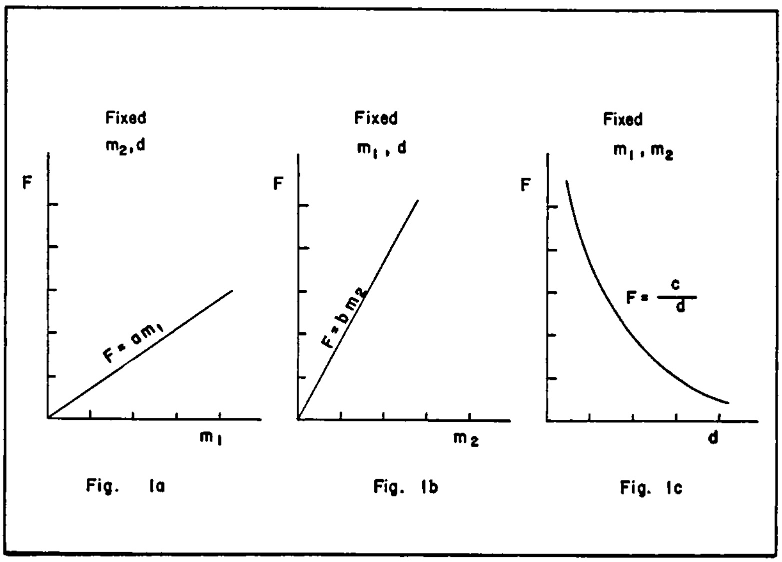


Figure 1. Special cases of the law of gravitation, $F = 6.66 \times 10^{-8}\, m_1\, m_2/d^2$.

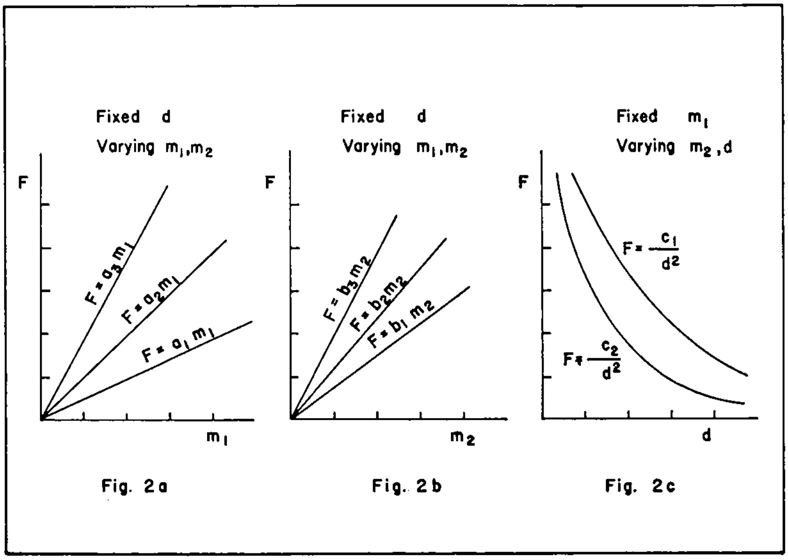


Figure 2. Interacting effects of the independent variables in the law of gravitation, $F = 6.66 \times 10^{-8} m_1 m_2/d^2$.

find a single equation involving the four variables F, m_1, m_2, and d. He might express his objective by writing the functional expression

$$F = \phi(m_1, m_2, d; \quad a_1, a_2 \cdots)$$

where $a_1, a_2, \cdots$ are undetermined constants. The implication of this symbolism is that the variable F depends upon m_1, m_2, and d in some unknown mathematical form called ϕ, and that in whatever form ϕ has, certain constants, $a_1, a_2, \cdots$ will appear. Since the experimenter did not know the form of ϕ, nor how many constants were in ϕ, nor the values of the constants, we shall assume that he was willing to "guess" the form of ϕ, taking into account any relevant theoretical or empirical information gained from his past experience.

Let $\hat{\phi}$ represent his guessed function, or mathematical model, for the correct function ϕ. The graph of $\hat{\phi}$ cannot be drawn, since four coordinate axes are required, but the experimenter could conceive that $\hat{\phi}$ represents a surface whose ordinates are determined by various triplets of values for m_1, m_2, and d. We might suppose that his first model was the simple linear function

$$\hat{F} = \hat{\phi}(m_1, m_2, d; \quad a_1 a_2, \cdots) = a_1 + a_2 m_1 + a_3 m_2 + a_4 d.$$

He may have quickly discarded this model upon observing that contrary to his experience, F does not necessarily become zero when either m_1 or m_2 is zero. Then, perhaps through long thought and inspiration, he may have chosen another linear model, but this time with all variables appearing in logarithmic form. Such a model would have to be written as

$$\log F = a_1 + a_2 \log m_1 + a_3 \log m_2 + a_4 \log d.$$

Upon deciding that this model was quite reasonable and in accord with his past experience, he would naturally have turned to the problem of estimating the constants a_1, a_2, a_3 and a_4 by means of an empirical investigation. Since there are four unknowns, his experiment must be designed to yield at least four sets of numerical values (F, m_1, m_2, d) to be substituted back into the model. We assume that his experiment led to four such sets of data, and that upon substitution into the model, four independent simultaneous equations were available for solution. Upon solving these equations, the experimenter may have determined that

$$a_1 = 8.13 \times 10^{-8}, \qquad a_2 = 0.98,$$

$$a_3 = 1.22, \qquad \text{and} \qquad a_4 = -2.04$$

Finally, substituting these empirical estimates back into his model, and taking antilogarithms, his conclusions would be that

$$F = 6.51 \times 10^{-8} \, m_1^{.98} m_2^{1.22}/d^{2.04}$$

It is clear that the third experimenter found the law of gravitation to within experimental error, mainly because of a fortuitous choice for his mathematical model. We can see that his experimental design and analysis were of such a nature that estimates for the constants were fairly accurate. However, he had no way of knowing whether his conclusion was valid, nor any estimate of the precision associated with his empirical constants. Nevertheless, his result enabled him to draw all the curves shown in Figures 1 and 2, and many more, since he could regard any of these graphs as but special cases of his own result.

In the ficticious example just related, the multivariate approach of the third experimenter was more productive than the methods of the first two experimenters. However, if his result is to have true scientific worth and acceptability, and is to have practical application, he should be prepared to answer the seven questions which follow.

1. What governed the selection of the particular independent variables m_1, m_2, and d?
2. What thinking went into the construction of the mathematical model which was used?
3. What levels of the independent variables needed to be investigated experimentally so that the unknown constants could be satisfactorily determined?
4. What mathematical techniques were used, and which are best for estimating the undetermined constants?
5. How precise are the estimates of the constants, and with what precision can F be estimated from the empirical equation?
6. How valid is the model?
7. How far, in space and time, may the results of this experiment be generalized?

In this paper we shall use the term *validity* in describing the degree to which any assumed mathematical model, $\hat{\phi}$, gives an adequate mathematical representation of the true function, ϕ, over some (perhaps) restricted range of the independent variables, and over some (perhaps) restricted range in space and/or time. For example, the third experimenter's first model was invalid, but his final model was valid for all possible values of m_1, m_2, and d, over all the universe, and for all time, we may suppose. The assumed model does not have to have the correct form in order to have a high degree of validity. For example, anyone who makes a linear interpolation in a table of logarithms is momentarily assuming a linear model to represent a logarithmic model.

We shall use the word *precision* to describe the extent to which empirical estimates may vary from experiment to experiment and/or within the same experiment. For example, the third experimenter's estimate for a_2 was 0.98, but if the experiment were repeated he might obtain the estimate $a_2 = 1.04$. Such fluctuation in the estimates reflects imprecision or unreliability in the experimental results. A high degree of precision might be attained even though the model were invalid, or it could be that no satisfactory degree of precision can be reached even though a valid model is employed.

Since the law of gravitation is known, and is held to be universal, the answers to many of these questions are apparent as we look at the results of the third experimenter. The fact remains, however, that *he* may have been able to give only very few satisfactory answers, and most of those in a completely subjective manner—with no real objective backing from his data and analysis.

In the remainder of this paper we shall attempt to answer these seven questions, although not in the order listed above, as they might apply to an analysis of road test data. Nearly all of the questions are profound and difficult, and so we shall expound and illustrate at some great length and with considerable detail.

THE MULTIPLE REGRESSION PROBLEM

We may suppose that an objective in a road test might be to determine a functional relationship between some dependent variable and two or more independent variables, just as in the introductory example. The general aim of a multiple regression analysis is to produce such a relationship in a mathematical form, and in such a way that the last five questions in the preceding introduction may be answered.

In order to achieve this aim, the objective must clearly describe those variables which are to appear in the function. Data must be taken on these variables through a design which will make the desired analysis possible. The statement of objectives, the experimental design, and the data analysis should be regarded as inseparable phases of the experiment, from its conception on through to the final report.

To illustrate, suppose that an objective for a road test is to determine a functional relationship between the maximum amplitude of deflection in a flexible pavement and certain pavement design variables, load characteristics, and environmental conditions.

In order for this objective to be reached it must be translated into specific descriptions of the variables and the manner in which the data are to be collected.

Nature of the Variables

Let the dependent variable, deflection, be designated by the letter, b. For our numerical illustration, b will be the Benkelman beam deflections taken at creep speed on sections of the WASHO Road Test in June, 1954. The independent variables will be selected from certain of the design variables as well as from measurements obtained in the WASHO subsurface testing program (*3*).

A complete listing of the variables selected for the illustrative regression analysis is given below.

x_1 = Design thickness of the asphaltic concrete surface, in inches.
x_2 = Design total thickness of the gravel base and subbase, in inches.
m_1 = Design load on a rear dual trailer wheel, in kips.
m_2 = Distance from rear trailer axle to next axle ahead, in feet.
m_3 = Number of loaded axle applications during the WASHO Road Test, in thousands.
m_4 = Lateral placement of rear dual wheel under which deflection is measured, distance from centerline in feet.
y_1 = Temperature of the asphaltic concrete surface at time of deflection measurement, degrees Fahrenheit.
y_2 = Gravel density in top 6 inches, in pounds per cubic foot.
y_3 = Basement soil density in top 6 inches, in pounds per cubic foot.
y_4 = Moisture content at top of basement soil, in percent.
$\bar{b}$ = Mean deflection for approximately 25 readings taken longitudinally with constant transverse placement, in inches.

In Table 1*a* is listed a portion of the actual data to be used in the regression analysis.

From the WASHO design there are 80 possible simultaneous sets of the selected variables. Twenty-two of these occur in the thinner sections which were not studied in the trench program, and so the illustrative data consists of 58 sets of values.

The objective of the analysis is now to express deflection as a multivariate function of the variables x_1, x_2, m_1, m_2, m_3, m_4, y_1, y_2, y_3, and y_4. Thus we have come to the first question listed in the introductory section, "How were the independent variables selected?" The answer must be that they have been named in a particular objective for the experiment, and were selected on the basis of past experience as well as from current hypotheses. The implication is that there must exist a documented statement of objectives, one of which reads, in symbolic form,

Find: $b =$

$$\phi(x_1, x_2, m_1, m_2, m_3, m_4, y_1, y_2, y_3, y_4).$$

Without such a list, both the design of the experiment and its analysis can easily remain ambiguous and somewhat unrelated throughout the experiment.

It will be assumed in the analysis that the independent variables actually do take on their stated values for each recorded deflection value. For example, if a mean deflection of 0.042″ is recorded opposite $x_1 = 2''$ and $y_4 = 22.4$ percent, it is assumed that these particular values for x_1 and y_4 are fixed and correct. Each of the values for y_1, y_2, y_3, and y_4 in Table 1*a* are mean values for representing the environmental conditions in the test sections at the time of the trench program. The mathematical premise in a regression problem is that only the dependent variable is subject to random variation, and that the values of the independent variables are fixed in the experimental design.

Before proceeding with the illustrative example, it is necessary to discuss those variations in the dependent variable which may arise *in addition* to that which results from its functional relationship with the independent variables, but before such variation can be meaningfully interpreted, we must come to grips with question 7 in the introduction.

The Connection Between Randomization and Generalization

The last question asked in the introductory section was, "How far in space and time may

TABLE 1a
ACTUAL DATA—WASHO

Set	b	x_1	x_2	m_1	m_2	m_3	m_4	y_1	y_2	y_3	y_4
1	.061	2	8	8.0	4.0	238	3.5	91	128.0	88.3	25.6
2	.029	2	12	8.0	4.0	238	3.5	84	129.0	95.9	20.2
3	.032	2	12	8.0	4.0	238	9.5	84	127.9	89.7	21.2
4	.033	2	16	8.0	4.0	238	3.5	82	128.7	92.1	21.8
56	.038	4	14	11.2	21.5	119	9.5	81	132.0	93.2	24.3
57	.029	4	18	11.2	21.5	119	3.5	78	133.1	92.9	22.0
58	.037	4	18	11.2	21.5	119	9.5	78	130.6	94.0	24.0

b = Mean of observed deflection - inches
x_1 = Thickness of asphaltic concrete - inches (design)
x_2 = Gravel thickness - inches (design)
m_1 = Wheel load - kip
m_2 = Spacing of loaded axles - feet
m_3 = Axle applications - thousands
m_4 = Lateral placement - feet from centerline
y_1 = Temperature of asphaltic concrete - degrees F.
y_2 = Gravel density - pcf (0 - 6 inch level)
y_3 = Basement soil density - pcf (0 - 6 inch level)
y_4 = Basement soil moisture content - percent (0 inch level)

the results of an experiment be safely generalized?" Although this question sounds straightforward enough, it is actually somewhat ambiguous until the meaning of the word *generalize* is established. Before attempting to formalize the concept of a generalization, let us examine two or three instances which might illustrate rather common applications of this subtle concept. We might, for example, state that the hourly traffic count on maintenance section J of Route 6 is *generally* higher than that on section M of Route 23. This generalization obviously refers to an average over time units, rather than to every particular hour that might be observed for there may be some hours during which the volume of traffic is greater on Route 23 than on Route 6. Furthermore, in making such a generalization, we likely have some rather definite overall period of time in mind, say from 1940 to 1955.

As a second type of generalization, we might say that, *in general*, heavier pavement structures are required over clay soil than over sandy soil. Again we must be speaking of an average condition, for there are certainly some locations where the statement would not hold true. In this case the generalization refers to an average condition over some collection of geographical regions.

Perhaps the whole aim of experimental research is to become able to make generalizations from observed data. If we were to observe that the temperature of water at its boiling point can be expressed either by 212 degrees F or by 100 degrees C, and at its freezing point either by 32 degrees F or 0 degrees C, we might plot these two points on a graph, draw a line through them, determine the equation $F = 9\,C/5 - 32$, then state that this must be the relation between the two observed temperature scales. Now this is a very sweeping generalization, for the implication is that this relation holds for all temperatures over all space units and time periods, and that it not only holds true on the average, but for any individual pair of measurements. Generalizations abound in everyday life where we find some who will say that men are generally more intelligent than women, etc.

Now if we look carefully at each of the preceding instances in which generalizations were being made, we find that in each case we were discussing an observable characteristic of some elemental unit: e.g., traffic count in one hour on Section J of Route 6, required pavement thickness of a highway section, Fahrenheit temperature of water, intelligence of a man, etc. Secondly, our generalization consisted of stating something about the average value of these characteristics, either over other elemental units, or over the same elemental unit as it is moved about in space or time. In each case the generalization was presumably meant to have some boundary or scope, in space, time, or with respect to type of elemental unit.

In order to make a generalization we must first have in mind an observable characteristic in an elemental unit, the latter being specified as to type, position in space, and position in time. The generalization itself consists in making a statement about the average value of the observable characteristic over some collection of elemental units. And the generalization remains vague unless its scope can be stated as the generalization is made, for only then is it clear just what elemental units are in the collection.

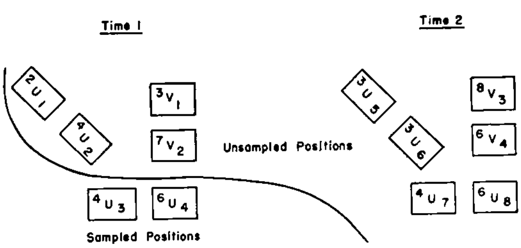
In experimental research, an elemental unit may be called an experimental unit, such as a test section in a road test. The observed characteristic may be some dependent variable, such as mean deflection in the section. Any generalization must then be from the observed deflection to an average deflection in some larger collection of experimental units, and the scope of the generalization is simply the listing of those units which appear in this larger collection. The scope may possibly include different types of units than were observed. It is quite likely to include units found in other positions in space, and is almost certain to include other positions in time than those which were observed.

In the sketch below we show 12 experimental units, 8 of type U and 4 of type V, laid out in space and in time. We may think of the U units as being "identical" road test sections occurring in different positions at the site of the road test. The V units may be nearby sections of highway built to the same specifications as the U units, and the two times might be day versus night, spring versus summer, 1955 versus 1965, etc.

Next we select some dependent variable, b, which unbeknownst to the experimenter, will take on the values shown in the upper left-hand corner of each unit. Now suppose that an experiment is designed to measure b, but that the experiment design allows only 1 of the 12 units in the sketch to be observed. We are immediately confronted with two questions. How shall we determine which of the 12 to observe? To which of the other 11 units may we generalize; that is, what will be our scope of generalization? It is fairly certain that no matter which unit is selected for the experiment, there will be varying opinions among the observers as to the proper scope of generalization. These opinions may vary from "not one step beyond the selected unit," to various combinations of the 12 units, to all 12 units, and perhaps on beyond those units shown in the sketch.

If it is not permissible to generalize at all from the observation on the selected unit, the experiment becomes a case study and has little or no significance in the scientific world.

Before going further, we must distinguish between what we shall call *scientific* generalizations and *faith* generalizations. As we try to decide which of the 12 experimental units to observe in the experiment, it may be that we automatically exclude some on the grounds of undue expense, inconvenience, etc. Such units will be said to fall in the unsampled positions since there is no chance for any of them to become the observed experimental unit. The remaining units will be said to fall in the *sampled positions* if each of them has some known positive chance of becoming the selected unit. If all units in the sampled positions have the same chance of being selected, the choice will be said to be *random*. In the sketch we have arbitrarily put only U_3 and U_4 among the sampled positions. This means that the experiment must consist in observing either U_3 or U_4 since none of the other 10 units has any chance of being selected.



Now, if the generalization from the single observed unit includes one or more units in the unsampled positions, we shall say that the generalization was made on faith. If the scope of generalization includes only units in the sampled positions, the generalization will be said to be scientific.

To point out why a generalization over the sampled points is any more scientific than a generalization which extends into unsampled positions we must refer back to the numerical values shown in the upper left-hand corner of each unit in the sketch. These numbers represent the value of some dependent variable, and as has been stated, a generalization is a statement concerning the average of *some* collection of these numbers.

The *general* value of the numbers in the sampled positions, U_3 and U_4 is $(4 + 6)/2 = 5$. If one of U_3 or U_4 is chosen at random, over and over, each will occur with nearly the same frequency, and the average numerical value in repeated experimentation will be 5, which was the general value. In other words, values for the dependent variable in the sampled positions will "average out" to be the general value in the scope of scientific generalization. Any value in the sampled region is said to be an unbiased estimate for the mean of that region. On the other hand, if we wished to generalize to all the six units shown at time 1 in the sketch, we should be estimating the mean $(2 + 4 + 4 + 6 + 3 + 7)/6 = 4.5$. Thus, repeated selections from the sampled positions will average to give the general values in the sampled positions, but *not necessarily* the average value in some larger scope of generalization. We may have the faith that such will be the case, and justifiably. For example, a faith generalization from U_3 and U_4 to include V_1 and V_2 happens to be correct, since the mean over this scope is $(4 + 6 + 3 + 7)/4 = 5$, just as in the sampled positions.

Two roles played by randomization are to increase the scope of generalization as well as to provide unbiased estimates within this scope.

Perhaps it is a common paradox for an experimenter to fail to randomize wherever he might, then go home at night to play bridge and insist rather strongly that the cards be shuffled before each hand is dealt. The shuffling process increases the sampled positions from one arrangement to something like 8×10^{67} possible arrangements of the card deck. He may also be aware that his hand will be unbiased even if he has no face cards, and be willing to state that he got a "fair deal."

In a concrete laboratory, the only store of aggregate may be a small pile of gravel of somewhat dubious origin, and so the experimenter knows that generalizations from beams made from this aggregate to outside aggregate sources are made strictly on faith. Nevertheless, he may mix his little pile well before scooping up a shovelfull. Again, he randomizes in order to get unbiased results for as large a set of sampled positions as possible.

Faith generalizations are quite prevalent in the world of science as well as in the everyday lives of all men. Perhaps no sicentific law has been shown to hold true in all places and for all times, but if nearly all scientists have complete faith that such would be the case, it is customary to say that the law has been proven. Faith generalizations to different types of experimental units than were observed are also common. Experimental results in biological experimentation may have been observed on animals, and yet generalizations made to the human race. Such a situation is also found in engineering research whenever results on laboratory-sized specimens are generalized to life-sized units, say from concrete cylinders to concrete slabs or columns. It may even be true that test road sections represent a different type of experimental unit than corresponding sections of highway.

Assuming that it will be desirable, if not necessary, to make some generalizations beyond the sampled positions, the experimenter should logically extend the scope of the sampled positions as far as he possibly can without causing himself undue expense or hardship. It is usually the case that if reliable results are obtained within a large scope of scientific generalization, then more confidence will be placed in those generalizations which must be made on faith.

Errors in the Dependent Variable

In the previous paragraph it has been pointed out that one role of randomization is to permit generalizations from observed experimental units to the mean of the experimental units which lie within the scope for scientific generalization. A random selection of only one experimental unit from the sampled positions will give an unbiased result as was previously explained. However, we

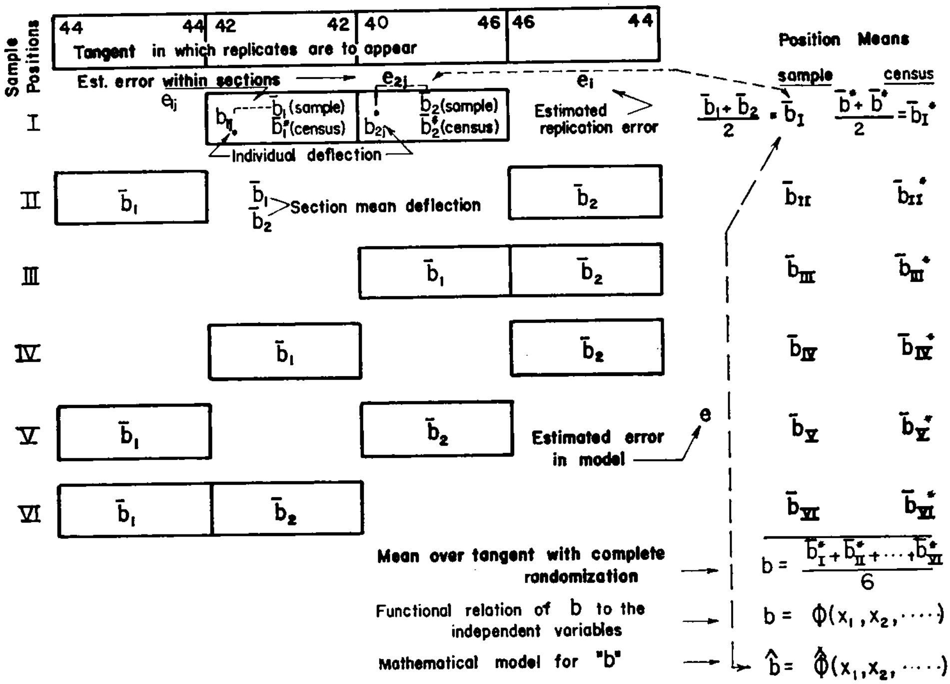


Figure 3a.

shall be unable to appraise the precision and validity of conclusions unless the analysis can give some numerical estimate of the variation among the experimental units in the sampled positions.

For example, the top diagram in Figure 3*a* might represent a test road tangent having four sections. Suppose a test section of given specifications is to be constructed somewhere in one of the four possible positions. If its position is selected at random, then the whole tangent lies in the scope of scientific generalization. But in order to estimate the variation in any dependent variable among positions in the tangent, a second, or *replicate*, section must be built to the same specifications as for the first section. It may be required that the two replicates be constructed independently of each other, at least insofar as the experimental budget will permit, for this procedure will again increase the scope of scientific generalization. The need for replication will again be illustrated in the paragraphs which follow.

Assuming that two replicate sections are to be placed somewhere in the tangent shown in Figure 3*a*, it is apparent that there are six possible sample positions for the two replicates, positions I, II, $\cdots$ VI, as labeled in the figure. If it is decreed that only position VI may be used, then this single position represents the scope of scientific generalization. If it is felt that the two replicates must be adjacent, then there are three possible positions, I, III, and VI, and the scope is three times larger than before. If one of these three is selected at random, the randomization is said to be *restricted.*

Complete randomization gives any one of the positions I, II, III, IV, V, or VI an equal chance to become the observed sections. In order to show the distinction between restricted and complete randomization, "true mean deflection" figures have been placed in the upper left-hand corner of the rectangles of the tangent layout. These numbers, 44, 42, 40, and 46, represent hypothetical observations that might occur for the specified section were it to be built in each of the respective positions along the tangent. Since the mean of these numbers is 43, any scientific generalization to the whole tangent layout should be an unbiased estimate of the number 43 by our definition of scientific generalization.

The restricted randomization described above will give a mean of $(42 + 40)/2 = 41$ for position I, $(40 + 46)/2 = 43$ for position III, and 43 for position VI. Since each of these three positions has the same chance of being selected, this randomization procedure will average, in repeated sampling, to give a mean of $(41 + 43 + 43)/3 = 42.3$, a biased result for the intended generalization of 43. Complete randomization gives each of the position means equal probability, and it can be seen that the mean deflection in all six positions is indeed the mean for the tangent.

If the "true mean deflections" were the numbers 44, 42, 46, and 44, as shown in the upper *right*-hand corner of the tangent rectangles, then the restricted randomization gives unbiased results since the respective means in positions I, III, and VI are 44, 45, and 43, and these average out to be 44, just as for the whole tangent. We have given this numerical illustration mainly to point out that restricted randomization assigns different probabilities of selection to the possible tangent sections, and that bias *may* arise in any generalization to the whole tangent unless the results are weighted in accordance with these probabilities. It is again a matter of *faith* to suppose that experimental results from a purposive selection, or that unweighted results from restricted randomizations, will give the unbiased estimates that are obtained from complete randomization.

After much discussion, we have come to the point of answering question 7 as it was raised in the introduction to this paper. Our answer is that the experimental results may scientifically be generalized to those positions or experimental units in space and time, which were *actively* sampled through proper randomization procedures, and that the soundness of faith generalizations must rest, in some appreciable degree, on the reliability and validity of the scientific generalizations.

As a final word on the subject of randomization, we point out that position I in Figure 3*a* might rather easily be the constructed location of the two replicate sections whether by decree, restricted randomization, or complete randomization. The experimental results would, of course, be the same in any of these three events. And so the skeptic (who presumably shuffles the cards each time he plays bridge or poker) may be unable to appreciate that randomization has changed the scope of generalization not one whit, nor that his result is any more or any less biased with or without randomization. In the last analysis, it may be that the most valuable benefits of active randomization accrue from the experimenter's serene knowledge that the laws of probability were available and were invoked, and that the searching question, "Did my subjective choice influence the selection of the experimental units?" can be answered unequivocally and affirmatively.

In order to proceed, it is necessary to explain the notation in Figure 3*a*. In the first replicate of position I, we let b_{1j} represent an individual deflection measurement; $\bar{b}_1$ represents the mean of a *sample* of such measurements, and $\bar{b}_1^*$ is the mean deflection which would occur if a census of all possible individual measurements were taken in the section. The difference, $b_{1j} - \bar{b}_1$, is a sampling error within the section and has been called e_{ij}. Corresponding notation appears in the second replicate of position I. The mean of the two replicate sample means appears in the right margin of the figure as $\bar{b}_{\mathrm{I}}$, while the mean of the two replicate census means is $\bar{b}_{\mathrm{I}}^*$.

Had any other position than I been selected through complete randomization, similar deflection means would have occurred in the right margin as $\bar{b}_{\mathrm{II}}$, $\bar{b}_{\mathrm{II}}^*$, $\bar{b}_{\mathrm{III}}$, $\bar{b}_{\mathrm{III}}^*$, etc. The mean of all six possible $\bar{b}^*$'s is called b at the bottom of the right margin, and it is precisely this value to which we are entitled to generalize by having used complete randomization. It is to this variable that we hope to relate the independent variables in a function, ϕ, which applies to the whole tangent. But just as was the case with the experimenter in the introductory example, we shall undoubtedly have to estimate the function ϕ by a mathematical model, $\hat{\phi}$. Whatever value of deflection the model gives is then called $\hat{b}$.

In addition to the sampling error, e_{ij}, within sections, there are two other errors that appear in Figure 3*a*. The first of these is called the replication error, e_i, and is the difference between the mean deflection in a single replicate and the mean of the two replicate means. That is, $e_i = \bar{b}_2 - \bar{b}_{\mathrm{I}}$, as is shown in the figure. Replication error is the

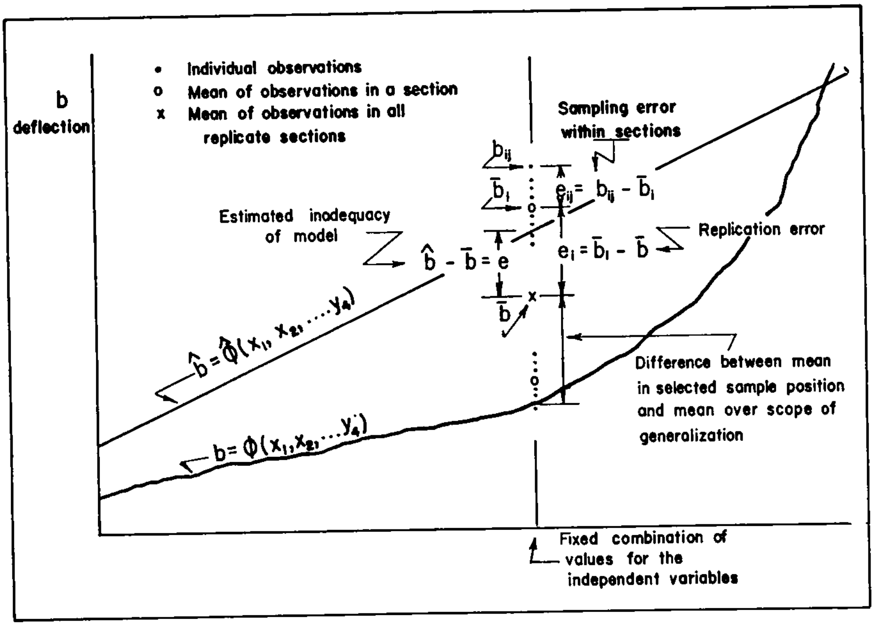


Figure 3b. The errors in a regression analysis.

only available estimate of variation among sampled positions, such as was manufactured in the numbers 44, 42, 40, 46 shown in the tangent layout at the top. *Without an estimate of replication error, there will be no way to test the validity of the model* $\hat{\phi}$, unless further (perhaps unwarranted) assumptions are made.

The error in the model, for one particular set of values of the independent variables, e. g., for two replicate test sections, is the difference between the value given for mean deflection by the model, $\hat{b}$, and that which was actually observed, $\bar{b}_I$. This error, $e = \hat{b} - \bar{b}_I$, should logically be compared with the replication error, e_i, in appraising the validity of the model.

In order to emphasize the notation and the meaning of these errors, in Figure 3*b* we have essentially reproduced Figure 3*a* in a more conventional manner.

The horizontal scale in Figure 3*b* represents an independent variable such as pavement temperature, and all the discussion of Figure 3*a* relates to variation in the dependent variable at a single point on the horizontal scale. The wavy curve through Figure 3*b* represents what might be called the "true relation" between the dependent and independent variables, and it is the equation of this curve, $b = \phi\ (x_1, x_2, \cdots, y_4)$, that we wish to approximate by a mathematical model, $\hat{\phi}$. The straight line through Figure 3*b* is the graph for the assumed model, $\hat{b} = \hat{\phi}\ (x_1, x_2, \cdots, y_4)$. Although we shall not single them out, all the various symbols and errors shown in Figure 3*a* can be found in Figure 3*b*. Before proceeding, the reader perhaps should verify that such is the case.

The three types of error, model errors, replication errors, and sampling errors, will be analyzed numerically in a later section. Since our illustration is from WASHO data, since there was no replication in the WASHO test road, and since we must have replication errors to determine the validity of the model, it has been necessary to manufacture mean deflections for dummy replicates to complement the data shown in Table 1*a*. These "replicate" values are partially shown in Table 1*b*.

The dummy values are not completely fictitious in that, for each of the 58 sets of data in Table 1*a*, they reflect the observed differences in mean deflection between adjacent traffic lanes in June, 1954, when all sections were tested under 9-kip wheel loads.

TABLE 1b
DUMMY REPLICATION VALUES

Set	b_1	b_2	$\bar{b}$
1	.066	.056	.061
2	.028	.030	.029
3	.035	.029	.032
4	.036	.030	.033
56	.037	.039	.038
57	.029	.029	.029
58	.037	.037	.037

b_1 = Mean deflection first replicate - inches
b_2 = Mean deflection second replicate - inches
$\bar{b}$ = Mean deflection first and second replicate - inches

THE MULTIPLE REGRESSION EQUATION

Perhaps from time immemorial, man has used models in order to clarify his own thinking, in order to explain his thoughts to others, and in order to study those physical phenomena which tend to make his life complex. Models, therefore, take on a great variety of forms. Some are true miniatures of their life-size prototypes, some are well chosen verbal illustrations, some are symbolic doodles, and some are more mathematical in their nature.

It is common to observe that technically trained men start drawing graphs upon the slightest provocation, thereby constructing mathematical models—for *every* graph has a mathematical representation and vice versa.

The purpose of a multiple regression analysis is to produce a mathematical equation which may be analyzed to determine the degree to which it accounts for values taken on by a dependent variable. The second question in the introduction was "How does one construct a mathematical model?" Although we make no pretense of being able to answer this question adequately, we shall list three considerations that are usually taken into account.

Construction of the Mathematical Model

The problem is to remove ambiguity from the letter ϕ in the expression, for example,
$b = \phi\ (x_1, x_2, m_1, m_2, m_3, m_4, y_1, y_2, y_3, y_4)$
We must ask whether ϕ simply adds the variables together, or whether it squares some, cubes others, puts others into exponents, etc. With 10 independent variables, it might be felt that it is impossible, or too complicated to construct such a model. However, it is not required that the model fit the data precisely, nor that it be the "true law" for deflection, if such exists, but only that it satisfactorily represent the values of the dependent variable over the selected range of the independent variables.

The first step should be to bring together previously determined mathematical relationships and theories which relate b to any of the independent variables. Most of the mathematical results in physical research have been deduced from assumptions on the various derivatives of the desired functions. It is common practice to use fundamental laws of physics to set up differential equations whose solutions become the desired mathematical models. In other problems the laws of probability are used to produce models which account for mean behavior in animate and inanimate objects. In the present problem we should perhaps study all that is known mathematically about elasticity and rupture, stress and strain, entropy and capillarity, and other basic phenomena—then begin to fit these relations together as well as to derive still others.

In the second place we shall probably require that the model reduce, in its special cases, to empirical results in which strong faith has generally been placed. Constructed or derived models should logically either be compatible with observed results from past experiments, or should more adequately account for the interplay of the variables than had previously been done.

In addition to pure mathematical development and reliance upon past experimental results, models may evolve from the intuition of the experimenter. Such models are rightfully called hypotheses, and may be tested by new experimentation. The third experimenter in the introductory example started as simply as he could, using a linear model. His success arose from his willingness, figuratively, to paste different scales on the dials of his instruments as he converted the original units to logarithms. Very few variables are stated in fundamental units, and in fact, much of the time there are no fundamental units except through convention. For example, a degree of temperature is simply a step, and how broad this step is must surely have to do with the velocity of molecules. But velocity is measured in still other arbitrary units, and so on, ad infinitum. Our point is that a mathematical model should permit much freedom with the scales for the observations.

It is well established that many mathematical functions may be satisfactorily represented by a polynomial expansion of the independent variables, and so a linear model consists of only the first few terms in such an expansion. For example, in a two-dimensional graph, no matter how curvilinear, linear approximations are said to be valid for certain restricted ranges of the independent variable.

Even if he has little mathematical talent, and no prior evidence concerning the nature of ϕ, the experimenter may wish to avoid certain absurdities in his model construction. In the present problem, we might require that any model used must give zero for deflection whenever m_1 (wheel load) is zero, and that if the other variables are fixed, deflection should increase as m_1 increases. Such a list of restrictions, or conditions to be satisfied by the model, constitutes the third of our three considerations in model construction.

In this paper we shall consider the two models shown below as possibilities for expressing the multivariate relationship between deflection and the 10 independent variables which have been selected.

$$\text{Model I:} \qquad \hat{b} = a_0\, m_1{}^{(X)(M)(Y)}$$
$$\text{Model II:} \qquad \hat{b} = m_1{}^c\, a_0{}^{(X)(M)(Y)}$$

where in both cases, the exponent $(X)(M)(Y)$ is an abbreviation for the product

$$(a_1 + a_2x_1 + a_3x_2 + a_4x_1x_2)\cdot (a_5m_2 + a_6m_3 + a_7m_4 + a_8m_2m_3 + a_9m_2m_4 + a_{10}m_3m_4)\cdot(a_{11}y_1 + a_{12}y_2 + a_{13}y_3 + a_{14}y_4 + a_{15}y_1y_2 + a_{16}y_1y_3 + a_{17}y_1y_4 + a_{18}y_2y_3 + a_{19}y_2y_4 + a_{20}y_3y_4).$$

Neither of these models has yet been carefully studied with respect to the three considerations of model construction, (1) known theory, (2) past experimental findings, and (3) logical mathematical behavior. It is almost certain that these relationships are not entirely appropriate for representing the function ϕ, but we have not searched the literature for those theories and empirical results which would indicate irregularities in these models.

In both cases, $\hat{b}$ is zero when m_1 is zero, and may increase as m_1 increases if all the other variables are fixed. In both models this "stress-strain" relation can turn out linear or curvilinear, depending upon whether or not the power of m_1 is one. In each model, if m_1 is fixed along with all other independent variables save one, the resulting equation connecting $\hat{b}$ with this last variable is for an exponential curve. This is equivalent to the assumption that the rate of change of $\hat{b}$ with respect to this last variable is proportional to the value of $\hat{b}$.

It is rather clear that this assumption cannot hold true for all of the variables x_1, x_2, m_2, m_3, m_4, y_1, y_2, y_3, and y_4, but in many cases the exponential curve may give a satisfactory representation.

If the three parentheses in the exponent of either Model I or Model II are multiplied out, there will be hundreds of resulting terms. These terms will account for two, three, and four variable linear interactions of the independent variables. In the illustrative analysis no more than 19 of these terms will be included since this is the present capacity of the computer which was used to achieve the analysis. The authors have arbitrarily selected certain terms for the exponent in each model. With these selections, the models become as shown below.

Model I: (Reduced)

$$\hat{b} = a_0 m_1 \exp \left\{ \begin{array}{l} (a_1 + a_2 m_2 + a_3 m_3 + a_4 m_4 \\ + a_5 x_1 + a_6 x_2 + a_7 y_1 + a_8 y_2 \\ + a_9 y_3 + a_{10} y_4 + a_{11} x_1 x_2 \\ + a_{12} m_3 m_4 + a_{13} m_4 x_2 \\ + a_{14} m_4 y_4 + a_{15} x_2 y_1 + a_{16} x_2 y_2 \\ + a_{17} x_2 y_3 + a_{18} x_2 y_4 \\ + a_{19} y_1 y_2) \end{array} \right\}$$

Model II: (Reduced)

$$\hat{b} = m_1^{a_{19}} a_0 \exp \left\{ \begin{array}{l} (a_1 + a_2 m_2 + a_3 m_4 + a_4 x_1 \\ + a_5 x_2 + a_6 y_1 + a_7 y_2 \\ + a_8 y_3 + a_9 y_4 + a_{10} m_4 x_2 \\ + a_{11} m_4 y_4 + a_{12} x_2 y_1 \\ + a_{13} x_2 y_2 + a_{14} x_2 y_3 \\ + a_{15} x_2 y_4 + a_{16} y_1 y_2 \\ + a_{17} y_1 m_2 + a_{18} x_1 m_2) \end{array} \right\}$$

In order to show the role of product terms such as $x_1 x_2$ in the models, consider the two equations:

$$b = 40 - 2x_1 - 3x_2 \quad \text{and} \quad b = 40 - 2x_1 - 3x_2 - x_1 x_2 .$$

Now suppose that $x_1 = 4$. Then these two equations become:

$$b = 40 - 8 - 3x_2 \quad \text{and} \quad b = 40 - 8 - 3x_2 - 4x_2 ,$$

or

$$b = 32 - 3x_2 \quad \text{and} \quad b = 32 - 7x_2 .$$

These last two equations are supposed to show the association between b and x_2, and the coefficients of x_2 indicate the rate of change of b with x_2. In the first equation this rate is the coefficient -3, and would have been so no matter what value had been selected for x_1. On the other hand, the rate of change of b with x_2 in the second equation depended entirely upon the value assigned to x_1, turning out to be -7 for $x_1 = 4$. We say that the second equation exhibits an interacting effect of x_2 with x_1. So the inclusion of the product terms in the exponents of our models was for the express purpose of accounting for possible interactions among certain of the independent variables. It is somewhat common to hear the statement made in many areas of research that "there are so many variables which contribute to fluctuations in my observed results, and the influence of any one of them depends entirely upon how the other variables are operating." Changed into the language of this paper, the previous quotation would become "my dependent variable is a multivariate function of a rather large number of independent variables many of which interact with one another."

Although we have certainly not exhausted the subject, we have tried to illustrate how multivariate models might be constructed, and have discussed at least some of the basic considerations in such constructions. Upon due reflection, it becomes rather clear that all experimenters construct multivariate mathematical models for their results. For if only a graph is shown which relates two variables, the experimenter will ordinarily list a number of other variables which the reader is to hold fixed in his mind as he studies the graphical conclusion. Geometrically, the experimenter has admitted that he really studied the trace of a multivariate surface in a single coordinate plane, just as did the first experimenter in the law of gravitation example which was used in the introduction to this paper.

Estimation of the Constants in the Model

In this section we deal with questions 3 and 4 of the introduction—both of which were concerned with the estimation of any constants that appear in the mathematical model being used.

In each reduced version of models I and II there were 20 undetermined constants. Our technique for estimating these constants from the data will require two steps, (1) linearization of the model, and (2) the application of the principle of least squares to the linearized model.

In order to demonstrate what we mean by linearization, consider the equation for a parabola in two dimensions, $y = a_0 + a_1x + a_2x^2$. If we substitute $X_1 = x$, and $X_2 = x^2$, the equation of the parabola in two dimensions becomes $y = a_0 + a_1X_1 + a_2X_2$, the equation of a plane in three dimensions. Any equation of the form $X_0 = A_0 + A_1X_1 + A_2X_2 + \cdots + A_kX_k$ is said to be a linear equation. If all the variables on the right side are held fixed, save one, the equation becomes that of a straight line in two dimensions. This straight line would be called a *trace* on the original $k + 1$ dimensional plane.

Many curvilinear models can be linearized through substitutions such as was just done; through transformations on the variables, such as taking logarithms; by using polynomial expressions to represent more complicated functions; or by some combination of these procedures.

Either Model I or Model II of our illustrative example may be linearized by taking logarithms of both sides of the model. The result of so doing is shown below.

Model I: (Linearized)

$$\log \hat{b} = \log a_0 + (X)(M)(Y) \log m_1$$

Model II: (Linearized)

$$\log \hat{b} = (\log a_0)\ (X)(M)(Y) + c \log m_1$$

When only the 19 selected terms from the product $(X)(M)(Y)$ are used, we have, showing only a portion of the terms,

Model I: (Reduced and Linearized)

$$\log \hat{b} = (\log a_0) + (\log m_1)\ (a_1 + a_2m_2 + \cdots + a_{19}y_1y_2)$$

Model II: (Reduced and Linearized)

$$\log \hat{b} = (\log a_0)\ (a_1 + a_2m_2 + \cdots + a_{18}x_1x_2m_2) + a_{19} \log m_1$$

To show these equations in their linear form we make the substitutions

For Model I: Let

$$A_0 = \log a_0, \quad A_1 = a_1, \quad A_2 = a_2, \cdots, A_{19} = a_{19},$$

and let

$$\hat{B} = \log \hat{b}, \quad X_1 = \log m_1, \quad X_2 = m_2 \log m_1, \cdots, X_{19} = y_1y_2 \log m_1.$$

For Model II: Let

$$A_0 = a_1 \log a_0, \quad A_1 = a_2 \log a_0, \quad A_2 = a_3 \log a_0, \cdots, \quad A_{18} = a_{18} \log a_0, \quad A_{19} = a_{19}$$

and let

$$\hat{B} = \log \hat{b}, \quad X_1 = m_2, \quad X_2 = m_4, \quad \cdots, X_{18} = x_1x_2m_2, \quad X_{19} = \log m_1.$$

With these substitutions, either model may be written in the desired linear from as

$$\hat{B} = A_0 + A_1X_1 + A_2X_2 + \cdots + A_{19}X_{19}.$$

Using the definitions for $X_1, X_2, \cdots, X_{19}$ which were just given for Model I, the values shown in Table 1*a* have been transformed into those shown in Table 2. The decimal points have been moved so that all the new variables are coded as integers in Table 2.

Since both models call for the logarithms of deflection, all original readings are to be used in logarithmic units, where for convenience, we have used $\log (100\ b) = B$, thus dropping the characteristic of each logarithm.

For each of the 58 sets of readings in Table 1*a*, there is now an observed "replicate mean logarithm," $\bar{B}$, and any difference between $\bar{B}$ and the model's estimate, $\hat{B}$ is the error which has been labelled *e* in both Figures 3*a* and 3*b*.

The principle of least squares says that those estimates are best for the coefficients $A_0, A_1, A_2, \cdots, A_{19}$ which make the sum of the squared model errors as small as possible. That is, application of the least square criterion will minimize the sum of $(\hat{B} - \bar{B})^2$.

Since there are 20 unknowns, the least squares procedure must produce, as must any other method, 20 independent equations to be solved simultaneously. Not many years ago such a solution would have been virtually impossible, but present day high speed electronic computers solve the whole problem in a matter of minutes from data to solutions. And so it becomes quite feasible to construct a great variety of multivariate mathematical models since any of them can be quickly analyzed.

We shall not go into the details of the least squares solutions here, but the coefficients shown in the fourth and seventh columns of Table 3 represent the solutions of 20 simultaneous equations for Model I and Model II

TABLE 2
VARIABLES IN THE LINEAR REGRESSION EQUATION

Model I

B	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}
7853	0903	0361	2149	3161	1806	0722	8217	1558	7973	2312	1445	0752	0253	0809	0657	0924	0637	1848	1052
4624	0903	0361	2149	3161	1806	1804	7585	1649	8660	1624	2167	0752	0379	0638	0910	1398	1040	2170	0978
5051	0903	0361	2149	8578	1806	1084	7585	1549	8100	1914	2167	2042	1029	1818	0910	1386	0972	2298	0970
5186	0903	0361	2149	3161	1806	1445	7405	1622	8317	1996	2890	0752	0506	0689	1185	1860	1331	3160	0963
4624	1049	2255	1248	3671	4196	1888	8182	3962	9745	2308	7553	0437	0661	0808	1473	2573	1754	4154	1089
5682	1049	2255	1248	9966	4196	1888	8182	3670	9861	2518	7553	1186	1794	2392	1473	2466	1775	4531	1069

$B = 10^4 \log 100b$
$X_1 = 10^3 \log m_1$
$X_2 = 10^2 m_2 \log m_1$
$X_3 = 10 \; m_3 \log m_1$
$X_4 = 10^3 m_4 \log m_1$
$X_5 = 10^3 x_1 \log m_1$
$X_6 = 10^2 x_2 \log m_1$
$X_7 = 10^2 y_1 \log m_1$
$X_8 = (10^2 y_2 \log m_1) - 10^4$
$X_9 = 10^2 y_3 \log m_1$
$X_{10} = 10^2 y_4 \log m_1$
$X_{11} = 10^2 x_1 x_2 \log m_1$
$X_{12} = m_3 m_4 \log m_1$
$X_{13} = 10 \; m_4 x_2 \log m_1$
$X_{14} = 10 \; m_4 y_4 \log m_1$
$X_{15} = y_1 x_2 \log m_1$
$X_{16} = x_2 y_2 \log m_1$
$X_{17} = x_2 y_3 \log m_1$
$X_{18} = 10 \; x_2 y_4 \log m_1$
$X_{19} = 10^{-2} y_2 y_1 \log m_1$

TABLE 3
COEFFICIENTS FOR REGRESSION MODELS

Model I				Model II		
Variables		Coefficients		Variables		Coefficients
Symbol	Coded values for Set 58, Table 1a	6 variables	19 variables	Symbol	Coded values for Set 58, Table 1a	19 variables
		$A_0 = -2411.7$	$A_0 = 1218.8$			$A_0 = -40157.$
$\log m_1$	$X_1 = 1049$	$A_1 = 14.339$	$A_1 = -73.341$	m_2	$X_1 = 215$	$A_1 = -1.3727$
$m_2 \log m_1$	$X_2 = 2255$	$A_2 = -.30855$	$A_2 = 18.719$	m_4	$X_2 = 95$	$A_2 = 24.258$
$m_3 \log m_1$	$X_3 = 1248$		$A_3 = 27.326$	x_1	$X_3 = 4$	$A_3 = -2296.0$
$m_4 \log m_1$	$X_4 = 9966$	$A_4 = .03623$	$A_4 = .35187$	x_2	$X_4 = 18$	$A_4 = 2079.4$
$x_1 \log m_1$	$X_5 = 4196$	$A_5 = -1.1315$	$A_5 = -1.0662$	y_1	$X_5 = 78$	$A_5 = 439.45$
$x_2 \log m_1$	$X_6 = 1888$	$A_6 = -3.7396$	$A_6 = 6.6584$	y_2	$X_6 = 1306$	$A_6 = 5.7436$
$y_1 \log m_1$	$X_7 = 8182$		$A_7 = .88441$	y_3	$X_7 = 940$	$A_7 = -8.7913$
$y_2 \log m_1$	$X_8 = 3670$		$A_8 = -.03518$	y_4	$X_8 = 240$	$A_8 = 21.591$
$y_3 \log m_1$	$X_9 = 9861$		$A_9 = -1.0049$	$x_1 x_2$	$X_9 = 72$	$A_9 = 130.01$
$y_4 \log m_1$	$X_{10} = 2518$		$A_{10} = 3.1894$	$x_2 m_4$	$X_{10} = 171$	$A_{10} = -.84108$
$x_1 x_2 \log m_1$	$X_{11} = 7553$	$A_{11} = .59987$	$A_{11} = .44801$	$m_4 y_4$	$X_{11} = 2280$	$A_{11} = -.75521$
$m_3 m_4 \log m_1$	$X_{12} = 1186$		$A_{12} = -.39097$	$x_2 y_1$	$X_{12} = 1404$	$A_{12} = -23.146$
$x_2 m_4 \log m_1$	$X_{13} = 1794$		$A_{13} = -.08540$	$x_2 y_2$	$X_{13} = 2351$	$A_{13} = -5.4094$
$m_4 y_4 \log m_1$	$X_{14} = 2392$		$A_{14} = -1.0277$	$x_2 y_3$	$X_{14} = 1692$	$A_{14} = 4.5818$
$x_2 y_1 \log m$	$X_{15} = 1473$		$A_{15} = -7.7762$	$x_2 y_4$	$X_{15} = 4320$	$A_{15} = -.93146$
$x_2 y_2 \log m_1$	$X_{16} = 2466$		$A_{16} = -3.3395$	$y_1 y_2$	$X_{16} = 1019$	$A_{16} = -.57110$
$x_2 y_3 \log m_1$	$X_{17} = 1775$		$A_{17} = 4.1040$	$x_1 m_2$	$X_{17} = 860$	$A_{17} = 5.2518$
$x_2 y_4 \log m_1$	$X_{18} = 4531$		$A_{18} = -1.1519$	$x_1 x_2 m_2$	$X_{18} = 1548$	$A_{18} = -3.5669$
$y_1 y_2 \log m_1$	$X_{19} = 1069$		$A_{19} = 4.5571$	$\log m_1$	$X_{19} = 1049$	$A_{19} = 11.265$
$10^4 \log 100b$	$B = 5682$			$10^4 \log 100b$	$B = 5682$	
$\hat{B} = A_0 + A_1 X_1 + \ldots =$		5018	4987	$\hat{B} = A_0 + A_1 X_1 + \ldots =$		5031
$\log 100\hat{b} = 10^{-4} \hat{B} =$		.5018	.4987	$\log 100\hat{b} = 10^{-4} \hat{B} =$		.5031
$\hat{b} =$		.032″	.032″	$\hat{b} =$		.032″

respectively. The computing routine makes it possible to further reduce the model through selecting any or all subsets of the original variables. The coefficients for one such selection of six variables from Model I are given in the third column of Table 3. It will be shown later that none of these three selected models will meet standards which will be set for validity. Although the 6-variable version of Model I is the poorest fitting equation for the observed deflection data, it will be used for most of the graphical presentation in a later section.

Our answer to the fourth original question, "What mathematical techniques have been used to estimate the constants in the model?" has been "to linearize the constructed model, then apply least squares methods to the linearized model." Such a procedure is not necessarily an optimum way to estimate the constants in a curvilinear multivariate regression model. Even for models which are linear at the outset, it may be that other criteria than least squares should be used in arriving at the necessary simultaneous equations. We shall not discuss the relative merits of alternative methods of estimation, but these need to be studied as carefully as the model construction itself. It is obviously important to supplement good models with adequate means for estimation of the constants in the model.

The objectives of an experiment must specify those independent variables that are to appear in the mathematical model for the analysis. A simpler way of putting this is that the objectives must make clear just what graphs are to be drawn in the final report. The experimental design must then assign values to the independent variables so that it is possible to achieve the analysis which will lead to these graphs. If a multiple regression analysis is to be performed so that the effects and interactions of the independent variables on the dependent variable can be truly studied, then it is required that there be a linear independence among the columns of values for those sets of the independent variables which enter into the analysis. This requirement means that it must not be possible to derive all the values of one independent variable from the values of one or more of the other independent variables by using a linear equation for the derivation. For example, the variable m_2 in the illustrative example has been designated as axle spacing in order to distinguish between single and tandem axle loads. It is not necessary that the values we have arbitrarily assigned to m_2 are in their proper units, since the multiple regression analysis will presumably give coefficients (as in Table 3) which properly modify the chosen values of roughly 4 feet for tandem axles and 21 feet for single axles. Now the values which identify the number of previous applications are shown as m_3 in Table 1*a*, and if the whole table were available, it would be seen that the only combinations of m_2 and m_3 in the data table, or in the WASHO experimental design, are those below on the left.

m_2	m_3	m_2	m_3
4.0	238	4.1	238
4.2	238	4.1	238
21.1	119	21.3	119
21.5	119	21.3	119

Perhaps for all practical purposes, the pairs of m_2, m_3 values should be considered as being those on the right above. In the latter case there is a complete linear dependence between m_2 and m_3, for it can be verified that all the m_2 values can be determined by substituting m_3 in the formula $4.1 - 0.1445\, m_3$. It follows that m_2 and m_3 are completely dependent since knowledge of one, for any set of data values, is equivalent to knowledge of the other. Whatever effect m_2 alone has on the dependent variable is confused with the effect of m_3, and so there is no point in using both variables in the mathematical model. Actually as is shown in the table on the left above, m_2 and m_3 were "not quite" dependent although there is a high correlation between their values.

It is not incorrect to design the experiment so that there are linear dependencies of the type just illustrated, unless the objectives of the experiment are to isolate the separate effects of such independent variables. It is partly for this reason that experimental objectives, in terms of particular dependent and independent variables, must be stated before there is any hope of knowing if an analysis can be achieved to meet the objectives by means of an appropriate experimental design. The analysis being illustrated in this paper can be accomplished as long as there are no linear dependencies among the independent variables which appear in the linearized regression model. In the illustrative problem, this means that in Table 2, none of the columns of figures $X_1, X_2, \cdots, X_{19}$ must be derivable from any linear combination of the other columns. It may be noted, however, that certain non-linear dependencies exist in Table 2. For example, it can be seen in Table 3, columns 5 and 6, that $X_9 = X_3X_4$ for Model II.

For a multipurpose experiment, such as a road test, it is almost certain that some of the independent variables will take on linearly dependent values, that is, not be truly independent of one another. When these dependencies exist among those independent variables which are specified in the experimental objective, the objective cannot be reached. Suppose, for example, that an experimental objective were stated in the form

$$\text{Find: } b = \phi\ (x_1, x_2, m/M, Y)$$

where b is some performance variable, x_1 and x_2 are variables yielding different pavement thicknesses, m is a variable for a single load application, M is a total mass to which a test section is subjected, and Y is descriptive of the environmental conditions for the section during the time that M occurs. The significance of the dash in the function ϕ is that the letters on the left of the dash are varied at the test site, while the letters on the right of the dash are common to all test sections. So this objective may be verbalized by asking "What is the relation of performance to design thickness and method of transporting loads if all test sections have to carry the same total load under the same environmental conditions?" Now if the experimental design assigns values to x_1, x_2, and m as shown in the table below, the separate effects of these variables on b cannot be appraised, for they are linearly dependent through the relation $m = 2\ (x_1 + x_2)$ as can be verified arithmetically.

x_1	x_2	m
2	4	12
4	2	12
3	6	18
5	4	18

In this case, it becomes impossible to distinguish between the effect of m and the effect of $x_1 + x_2$ on the dependent variable. We have not meant to imply that this last example necessarily represents an objective of a road test, nor that such dependencies would be permitted in the experimental design, but we have attempted once more to show how close is the connection between successful analyses and clearly stated experimental objectives.

To summarize this section, we have answered questions 3 and 4 of the introduction by saying that we shall estimate the constants in the linearized mathematical model by least squares methods. Furthermore, the experimental levels of the chosen independent variables must be independent in the linearized model, and so the experiment must be designed to meet this requirement.

Analysis of the Errors

In order to find the individual errors that the model gives in estimating mean deflection, values for the independent variables must be substituted in the linearized regression equation $\hat{B} = A_0 + A_1X_1 + A_2X_2 + \cdots$.

In order to illustrate this substitution, the X values for set 58 in Table 1*a* are shown in Table 3. The X values in the second column have already been given in the last line of Table 2, while those in the sixth column have been coded from Table 1*a* in accordance with the terms in Model II.

For any of these models, the estimate of B is given by multiplying successive coefficients by their respective X values, then adding, including the value of A_0. The resulting estimates, $\hat{B}$, are shown in the third from the last line in Table 3. For example, the 6-variable Model I estimates B to be 5018, the 19-term Model I estimate is 4987, and the 19-variable Model II gives the estimate 5031. The observed value of B was 5682, and so the respective errors in these models are $5682 - 5018 = 664$; $5682 - 4987 = 695$; and $5682 - 5031 = 651$. It is the sum of squares of these errors that has been minimized by the least squares solutions. To produce estimates of deflection in inches, antilogarithms must be taken of the $\hat{B}$ values in their decimal form. Estimates for inches of deflection are shown to be 0.032″ in the last line of Table 3 for all three illustrative models, although this is quite coincidental. Thus these models have estimated 0.032″ mean deflection for the outer wheel path of the 22.4-kip axle load on the WASHO section having 22″ total thickness and 4″ asphaltic concrete surface. The observed mean deflection was $b = 0.037''$, and so the model errors are $0.037'' - 0.032'' =$

TABLE 4
ANALYSIS OF ERRORS

Type of Error	Degrees of Freedom for Error	Mean Square Error in Log Units	Criterion Ratio
Error in Model $e = \hat{b} - \bar{b}$	Model I A 51	.00684	.00684/.00148 = 4.6
	Model I B 38	.00461	.00461/.00148 = 3.1
	Model II B 38	.00399	.00399/.00148 = 2.7
Replication Error Between Sections $e_i = \bar{b}_i - \bar{b}$	58	.00148	$\frac{.00148}{.00011} = 13.5$
Sampling Error Within Sections $e_{ij} = b_{ij} - \bar{b}_i$	2784	.00011	

A = Six variables
B = Nineteen variables

0.005″ for set 58 of the original data. There are 58 such errors for each of the three models which have been fitted to the data. To express these errors as a single value, it is conventional to find an average squared error. Each model error is squared, then these squares are added, and the sum is divided by a number called the degrees of freedom for error. The resulting mean squares for the model errors are given in the third column of Table 4.

Degrees of freedom for model error are always the number of sets of observations less the number of constants estimated in the model. If only two sets of data were observed, and two constants appeared in the mathematical model, there would be no degrees of freedom for model errors. In fact, perfect fits can be expected whenever there are as many estimated constants as there are sets of data. As a rule of thumb, it is desirable that as many as 5 or 10 degrees of freedom be available for measuring errors in the model.

If only the mean square of the model errors were available, we would know what these errors averaged to be, but could not tell whether they arose because of improper choice of model or were simply reflections of unreliability in the data.

By definition, the discrepancy between an individual replicate mean, and the mean of two replicates represents differences which arise from variables other than those which were selected in the model. If all available replication errors are squared, then averaged through division by an appropriate number of degrees of freedom, the result is called the mean square error for replicate sections. This value is shown as 0.00148 in Table 4.

For any single set of data values, the degrees of freedom for replication error are one less than the number of replicates in the set. In the present illustration, each set of data contributes one degree of freedom for replication error since there were two "replicates." Summing for the 58 sets of data, there are 58 degrees of freedom for replication error as is shown in Table 4.

Sampling errors within replicate sections are averaged by first squaring each observed error, $e_{ij} = b_{ij} - \bar{b}_i$, (c. f., Figures 3*a* and 3*b*) then summing for all individual observations. The degrees of freedom for sampling error are the total number of observations less the number of replicate sections. In our illustration this is approximately $(25 \times 2 \times 58) - (2 \times 58) = 2784$ as shown in Table 4. The mean square for sampling error has been put on the same unit basis as the previously described mean squares, and is shown as 0.00011 in Table 4.

All three of these mean squares are directly comparable in Table 4. The three types of error are assumed to be additive, and it follows mathematically that, if the model were perfectly valid, then the ratio of the model mean square error to the replication mean square error can be expected to be in the neighborhood of one. We shall adopt the rule that if this ratio exceeds two, there is just cause to discard, or look for means of improving, the current model, at least for the number of degrees of freedom in the example. As is shown in the criterion ratio column of Table 4, the 19-term Model I yielded errors of such magnitude that the criterion ratio was 3.1. When all 19 variables of Model II were used, the criterion ratio was down to 2.7. So it appears that other variables need to be accounted for in the model, or that the reduced models include somewhat the wrong selection from the hundreds of terms that were available in the respective exponents, or that the mathematical form of the models needs revision. It is not necessarily our present purpose to determine an adequate model for representing these deflections, but it seems rather certain that such can be done without too much difficulty. It must be remembered that our "replicate" values are not really such, and that true replicates might produce either larger or smaller replication errors. In the process of computation, multiple correlation coefficients are determined as a matter of course. For Model I the multiple correlation coefficient was 0.93, while for Model II this value was 0.94. Although these values would seem phenomenally high in many fields of investigation, they are not high enough to satisfactorily account for deflections in the present illustration, as the criterion ratios show.[1] Although these models show promise, they need further investigation along the lines of the three considerations mentioned in the preceding section on model construction.

The ratio of the replicate mean square to the within replicates mean square is an indication of the need for replication. If this ratio exceeds two (say), then it is pretty certain that there are differences in the dependent variable for other causes than those which bring about errors within replicates, and that if the scope of generalization is to include these, replication is necessary.

The ratio of replicate mean square to within replicate mean square is shown as 13.5 in the last column of Table 4. Since this is much greater than our standard of two, it would follow that within the replicate mean square does not account for all uncontrolled variation that occurs in the test data. However, it must be remembered that the replicate values shown in Table 1*b* are fictitious.

If one wished to have some notion of the variation in compressive strength over a sizeable construction made of concrete, he might form 25 cylinders from one mix, test them, find the standard deviation to be 50 psi, then say he knew the variability in compressive strength over the whole construction. But if the mean value of compressive strength of cylinders from another mix is 200 psi higher than that from the first sampled mix, then his first standard deviation does not reflect the mix to mix variation at all. If the mix mean differences were of the order of 20 psi however, then it is quite possible that his standard deviation does apply to the whole construction. In the road test example, the ratio of replicate-mean-square to within-replicate-mean-square is used for the same purpose, to see if, on the whole "job," there is any more "mix to mix" variation than "within mix" variation, and if so, to make it possible to relate the experimental results to the whole "job." Of course if it is a foregone conclusion that the replication error *will* be no more than that error which occurs within replicates, there is no need for replication. For some variables, however, there can be no within replicate variation at all, for the value of the variable may relate to the whole section. Amount of maintenance required is a prime example of this sort of variable. For such a variable, it would seem mandatory to provide replication, if the precision of the results is to be scientifically estimated.

Although the 19-variable model mean squares in Table 4 are about three times the replication mean square, the average error made by the models is about 1.6 to 1.8 times the average replication error, the latter being approximately 0.003 inches of deflection. With Model II, therefore, estimated deflections will average to be off from the observed values by about 0.005 inches.

There are varying magnitudes over the range of the independent variables for both the model errors, e, and the replication errors, e_i. Such heterogeneity of error can become the basis for valuable conclusions in itself. For although the squared model errors may average (in Table 4) to be more than twice that of the squared replication errors, this situation does not necessarily prevail over the whole range of data. It can easily happen that the model is much more valid in some ranges than in others. But in order to make such a study of the errors, either there will

[1] A multiple correlation coefficient of 0.94 implies that $100(.94)^2 = 88$ percent of the variation in the 58 mean deflections has been accounted for by the 19 terms of the Model II regression equation.

TABLE 5
ESTIMATED VERSUS OBSERVED DEFLECTIONS

	x_1	2 inch									
	x_2	4		8		12		16		20	
	m_4	3.5	9.5	3.5	9.5	3.5	9.5	3.5	9.5	3.5	9.5
$m_1 = 8.0$ $m_2 = 4.0$	$\bar{b}$	*	*	.061	*	.029	.032	.033	.029	.029	.026
	$\log 100\bar{b}$	*	*	.79	*	.46	.51	.52	.46	.43	.42
	$\hat{B}_{I}$	.76	.78	.67	.69	.57	.59	.48	.50	.39	.41
	$\hat{B}_{II}$	*	*	.78	*	.47	.55	.44	.46	.42	.43
	$\hat{b}_{I}$	.057	.060	.046	.049	.038	.039	.030	.032	.025	.026
	$\hat{b}_{II}$	*	*	.060	*	.030	.035	.028	.029	.026	.027
	e_{I}	*	*	.015	*	−.009	−.007	.003	−.003	.002	0
	e_{II}	*	*	.001	*	−.001	−.003	.005	.000	.001	−.001
$m_1 = 10.0$ $m_2 = 4.2$	$\bar{b}$	*	*	*	*	.036	.047	.039	.034	.031	.030
	$\log 100\bar{b}$	*	*	*	*	.56	.67	.59	.53	.49	.48
	$\hat{B}_{I}$	.86	.89	.76	.78	.66	.68	.56	.58	.46	.48
	$\hat{B}_{II}$	*	*	*	*	.58	.66	.53	.57	.51	.50
	$\hat{b}_{I}$	.073	.077	.059	.061	.046	.048	.036	.038	.028	.030
	$\hat{b}_{II}$	*	*	*	*	.038	.045	.034	.037	.033	.032
	e_{I}	*	*	*	*	−.010	−.001	.003	−.004	.002	0
	e_{II}	*	*	*	*	−.002	.002	.005	−.003	−.002	−.002
$m_1 = 9.0$ $m_2 = 21.1$	$\bar{b}$	*	*	.060	*	.037	.038	.030	.029	.026	.026
	$\log 100\bar{b}$	*	*	.78	*	.57	.58	.48	.46	.42	.42
	$\hat{B}_{I}$	.76	.78	.67	.69	.57	.59	.47	.49	.38	.40
	$\hat{B}_{II}$	*	*	.71	*	.57	.61	.48	.48	.37	.40
	$\hat{b}_{I}$	.058	.061	.046	.049	.037	.039	.030	.031	.024	.025
	$\hat{b}_{II}$	*	*	.052	*	.037	.041	.030	.030	.023	.025
	e_{I}	*	*	.014	*	0	−.001	0	−.002	.002	.001
	e_{II}	*	*	.008	*	.000	−.003	.000	−.001	.003	.001
$m_1 = 11.2$ $m_2 = 21.5$	$\bar{b}$	*	*	.073	*	.039	.053	.033	.035	.030	.034
	$\log 100\bar{b}$	*	*	.86	*	.59	.72	.52	.54	.48	.53
	$\hat{B}_{I}$	.86	.88	.76	.78	.65	.67	.54	.57	.44	.46
	$\hat{B}_{II}$	*	*	.81	*	.68	.72	.59	.57	.48	.48
	$\hat{b}_{I}$	.073	.077	.057	.060	.045	.047	.035	.037	.028	.029
	$\hat{b}_{II}$	*	*	.065	*	.047	.053	.039	.037	.030	.030
	e_{I}	*	*	.016	*	−.006	.006	−.002	−.002	.002	.005
	e_{II}	*	*	.008	*	−.008	.000	−.006	−.002	.000	.004

	x_1	4 inch									
	x_2	2		6		10		14		18	
	m_4	3.5	9.5	3.5	9.5	3.5	9.5	3.5	9.5	3.5	9.5
$m_1 = 8.0$ $m_2 = 4.0$	$\bar{b}$	*	*	.042	.041	.036	.044	.031	.031	.030	.031
	$\log 100\bar{b}$	*	*	.62	.61	.55	.64	.49	.49	.48	.49
	$\hat{B}_{I}$	.62	.64	.57	.59	.52	.54	.48	.50	.43	.45
	$\hat{B}_{II}$	*	*	.55	.63	.55	.63	.51	.52	.48	.48
	$\hat{b}_{I}$	.042	.044	.037	.039	.033	.035	.030	.031	.029	.028
	$\hat{b}_{II}$	*	*	.036	.042	.035	.042	.032	.033	.030	.030
	e_{I}	*	*	.005	.002	.003	.009	.001	0	.003	.003
	e_{II}	*	*	.006	−.001	.001	.002	−.001	−.002	.000	.001
$m_1 = 10.0$ $m_2 = 4.2$	$\bar{b}$	*	*	.047	.061	.037	.045	.041	.045	.035	.038
	$\log 100\bar{b}$	*	*	.67	.69	.57	.65	.61	.65	.54	.58
	$\hat{B}_{I}$	.71	.73	.66	.68	.61	.63	.55	.57	.50	.52
	$\hat{B}_{II}$	*	*	.69	.76	.65	.68	.62	.62	.54	.55
	$\hat{b}_{I}$	.052	.054	.046	.048	.040	.042	.036	.038	.032	.033
	$\hat{b}_{II}$	*	*	.049	.058	.045	.048	.042	.042	.034	.035
	e_{I}	*	*	.001	.013	−.003	.003	.005	.007	.003	.005
	e_{II}	*	*	−.002	.003	−.008	−.003	−.001	.003	.001	.003
$m_1 = 9.0$ $m_2 = 21.1$	$\bar{b}$	*	*	.032	.038	.029	.028	.027	.025	*	.023
	$\log 100\bar{b}$	*	*	.51	.58	.46	.45	.43	.40	*	.36
	$\hat{B}_{I}$	.62	.64	.57	.59	.52	.54	.47	.49	.41	.44
	$\hat{B}_{II}$	*	*	.50	.57	.51	.49	.42	.44	*	.39
	$\hat{b}_{I}$	.042	.044	.037	.039	.033	.034	.029	.031	.026	.027
	$\hat{b}_{II}$	*	*	.032	.037	.033	.031	.026	.028	*	.025
	e_{I}	*	*	−.005	−.001	−.004	−.006	−.002	−.006	*	−.004
	e_{II}	*	*	.000	.001	−.004	−.003	.001	−.003	*	−.002
$m_1 = 11.2$ $m_2 = 21.5$	$\bar{b}$	*	*	.042	.052	.039	.044	.032	.038	.029	.037
	$\log 100\bar{b}$	*	*	.62	.72	.59	.64	.51	.58	.46	.57
	$\hat{B}_{I}$	.70	.73	.65	.67	.59	.61	.53	.56	.48	.50
	$\hat{B}_{II}$	*	*	.63	.65	.61	.64	.51	.54	.46	.50
	$\hat{b}_{I}$	.051	.053	.044	.047	.039	.041	.034	.036	.030	.032
	$\hat{b}_{II}$	*	*	.043	.045	.041	.044	.033	.035	.029	.032
	e_{I}	*	*	−.002	.005	0	.003	−.003	.002	−.001	.005
	e_{II}	*	*	−.001	.007	−.002	.000	−.001	.003	.000	.005

$\bar{b}$ = Observed Mean Deflection—inches; $\hat{B}$ = Estimated $\log 100\,\bar{b}$; $\hat{b}$ = Estimated $\bar{b}$

have to be full replication, or else the errors in unreplicated sections will have to be assumed to be of the same magnitude as those which were actually observed.

For the sake of comparison we have reduced Model I to just six variables in order to see to what extent the model errors would increase from the original 19-variable case. The selected variables are shown in Table 3, as are the coefficients for the 6-variable model. It can be noted that each of the selected terms has different coefficients, in the 6-variable equation than in the 19-variable equation. Perhaps the simplest analogy to this situation is the case of a basketball team of five players. If one were removed, the remaining four would probably change their playing in accordance with whichever substitute the coach sent in. Or if there were no more players on the bench, the play of the remaining four would have to change radically if the team were to continue to score. It is possible to make a great variety of selections for the independent variables in the model, but we know that the criterion ratio cannot be expected to decrease as selections are made from the 19 variables used in Models I and II. For the 6-variable version of Model I, the model mean square is 4.6 times the replication mean square, and so the model makes errors about twice the size of the replicate errors, or about 0.006 inches on the average.

In Table 5 we have shown all 80 combinations of the WASHO pavement design, wheelpath, and load characteristics. For each cell the observed mean deflections at the conclusion of the road test are listed as $\bar{b}$. The estimated logarithms of these values from the 6-variable version of Model I are labelled $\hat{B}_{\mathrm{I}}$, and those from the 19-variable Model II as $\hat{B}_{\mathrm{II}}$. All these estimates have been converted back to inches of deflection, $\hat{b}_{\mathrm{I}}$ and $\hat{b}_{\mathrm{II}}$. Although the error analysis of Table 4 is in logarithmic units, we have shown model errors in Table 5 in the original units. It can be seen that the errors are not homogeneous over all 80 cells, and also that most Model II errors are smaller than are the 6-variable Model I errors, as has already been indicated by the analysis in Table 4. A more complete study of these errors will undoubtedly point the way to a more valid model.

In this section we have attempted to present answers for questions 5 and 6 of the introduction by illustrating what would appear to be logical and sound bases for determining the validity of the mathematical model and the precision of its estimates. We have not discussed the question of precision of the regression coefficients themselves, as they are shown in Table 3 for example. This is quite an important question since if the coefficient of X_5 turns out to be -1.1315 as in column three of Table 3, we should have a clear notion as to how many of these digits are statistically significant. We shall let it suffice to say here that, in the course of the same computing routine which produces the estimated coefficients, an inverse matrix for the coefficients of the least squares simultaneous equations is calculated. Then from the elements of this inverse matrix, and from the replication mean square of Table 4, it becomes a relatively simple matter to estimate the degree of precision which the regression coefficients have.

From the point of view of the scientific method, the objectives in any experiment consist of testing hypotheses on the characteristics of mathematical models—through prescribed experimental design and analysis. For many of the objectives in road test research, it would appear that a multiple regression analysis can satisfactorily estimate the constants in whatever model is hypothesized, and can objectively determine the validity of the model as well as the precision of its constants and the reliability of its estimates.

Orthogonal Analyses

It is highly desirable that the levels of independent variables be arranged by design so that there is no correlation between any two effects of these variables. Experiment designs having this property may be called orthogonal designs since the estimate of any particular regression coefficient is truly independent of the values taken on by the remaining coefficients. In orthogonal analyses, the reliability of each regression coefficient may be estimated separately from that of the others, and the computational effort is greatly reduced.

In view of these advantages, every effort should be made to design an experiment so that it leads to orthogonal analyses, and much of the literature (*1*, *2*) is devoted to the construction of orthogonal experiment designs. In general, such experiments are factorial in nature, with any level of one experimental factor occurring with every combination of levels of the remaining factors.

Orthogonal analyses are special cases of the general regression analysis we have used in this paper. Although we have stated that no two independent variables may be perfectly correlated, many of the variables of our illustrative analysis are intercorrelated to various degrees, and thus the analysis is non-orthogonal.

Even though a road test experiment may be set up in an orthogonal design, there are perhaps three reasons why at least some of the data may have to be analyzed through general regression procedures.

In the first place, it may be desirable to include terms in the regression model that reflect effects of uncontrolled variables, such as moisture and temperature. Secondly, many of the original experimental units may be lost through the failure of test sections. As a consequence, the balance of any original orthogonal design may be lost. Finally it seems that it may be worthwhile to investigate mathematical models in which the original variables are transformed in such a way that the effects of the transformed variables are correlated even though the design levels of these variables are independent.

It would seem therefore, that an optimum compromise for a road test experiment would be to design the experiment for orthogonal analyses, but to make provision for those non-orthogonal analyses that may be required for one or another of the reasons just mentioned.

Graphical Presentation of the Mathematical Model

The criterion ratios in Table 4 indicate that neither Model I nor Model II has a desirable level of validity, although it is quite possible that lower criterion ratios might have resulted from a different choice of interaction terms. The great majority of errors of estimate shown in Table 5 are under 0.005 in., but some of these errors are quite large and systematically so.

If each of the levels of one independent variable were to occur in combination with all levels of the remaining variables, then the road test would be said to be a complete factorial experiment. In this event all possible interactions of the independent variables could be studied systematically. It would seem that, for one dependent variable or another, nearly all interactions of two, three, and perhaps four or five variables will become significant. The regression analysis described in this paper supposes that these interactions can be studied through error analysis and subsequent model adjustment. For each selected model, any systematic arrangement of the errors will suggest that important interacting effects of the independent variables have not yet been taken into account. If the purpose of this paper were to analyze the WASHO data, we should have to do a much more comprehensive analysis than is being presented.

In this section we wish to point out how a multivariate model can yield a great variety of graphical results. For this demonstration we shall choose the simplest, and poorest fitting, of the three models that have been studied in this paper, the 6-variable version of Model I.

If the coefficients in the next to last column of Table 3 are substituted into the linearized regression model, then antilogarithms taken, Model I becomes

$$\hat{b} = 0.5739 m_1 \exp\left\{\begin{array}{l}(1.43389 - 0.0031 m_2 \\ \quad - 0.00362 m_4 \\ \quad - 0.11315 x_1 \\ \quad - 0.03740 x_2 \\ \quad + 0.00600 x_1 x_2)\end{array}\right\}$$

There are five independent variables on the right hand side of this result. If values are assigned to any four of these, the resulting equation may be plotted in two dimensions. Such a graph would then be a 2-variable trace of the original multidimensional surface. If the model were valid by the criterion of Table 4, and if the errors in Table 5 were homogeneous, or at least in rather constant ratio to the replication errors over the entire table, then any such trace should be regarded as truly indicative of the relationship between the two plotted variables when other variables are held fixed.

In Figure 4 we show traces of this surface when deflection is plotted against wheel load, all other variables in *this* model being held constant. Dotted curves are used for those parts of the traces that are outside the range of the experimental levels of the independent variables.

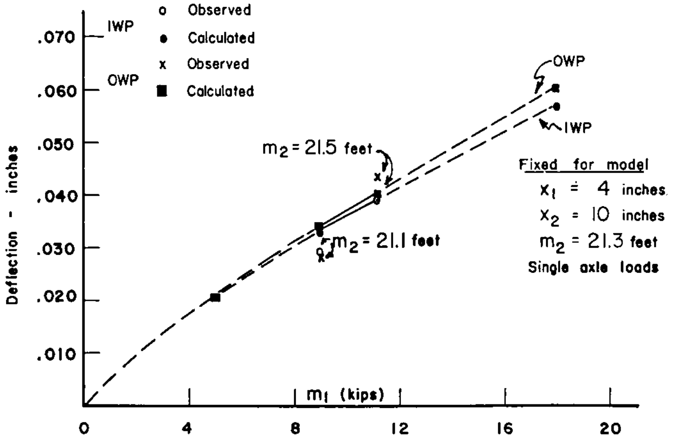


Figure 4a

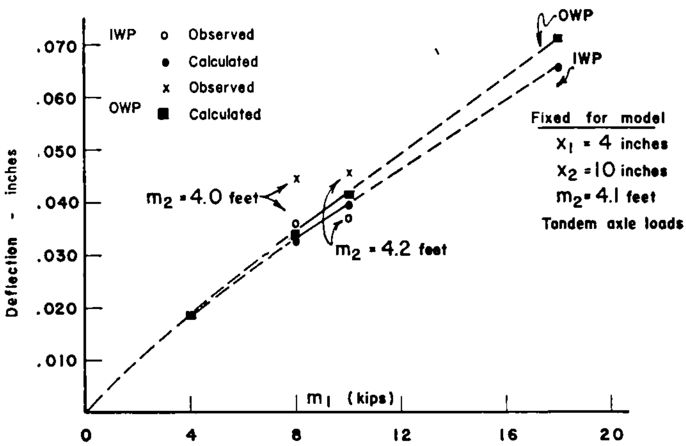


Figure 4b

Figure 4. Deflection versus wheel load. (Model I—6 variables)

First letting $x_1 = 4$ in. (asphaltic concrete thickness) and $x_2 = 10$ in. (gravel base thickness), the model becomes

$$\hat{b} = 0.5739 m_1 \exp$$

$$(0.8473 - 0.0031 m_2 - 0.0036 m_4)$$

In Figure 4*a*, m_2 is taken to be 21.3 ft. (axle spacing), and for the two values of transverse placement, m_4, the model reduces to the equations

$$m_4 = 3.5 \text{ ft.}: \hat{b} = 0.5739 m_1^{.7940}$$

$$m_4 = 9.5 \text{ ft.}: \hat{b} = 0.5739 m_1^{.8156}$$

When $m_2 = 4.1$ ft. as in Figure 4*b*, the corresponding equations are

$$m_4 = 3.5 \text{ ft.}: \hat{b} = 0.5739 m_1^{.8472}$$

$$m_4 = 9.5 \text{ ft.}: \hat{b} = 0.5739 m_1^{.8690}$$

The observed deflections are also shown on the figure, and it is rather apparent that the observations have characteristics which differ from those of the model. If the mathematical form of the model is appropriate, then it is likely that certain important interaction terms have been erroneously omitted in the reduced model being used. For example, if the interaction of y_4 (moisture content) with any of x_1, x_2, m_2, m_4, were introduced into the model, then each single observation in Figure 4 would be represented by a different curve.

In Figure 5 we have let $m_1 = 11.2$ kips, $m_2 = 21.5$ ft., and $x_1 = 4$ in.. The (invalid) model being used reduces to the equation

$$\hat{b} = 0.5739\ (11.2)\ \exp\left\{\begin{array}{l}(1.3672 + 0.0036m_4 \\ \quad - 0.0374x_2 \\ \quad - 0.1132x_1 \\ \quad + 0.0060x_1x_2)\end{array}\right\}$$

In order to plot $\hat{b}$ against gravel thickness, x_2, this equation has been reduced to the four

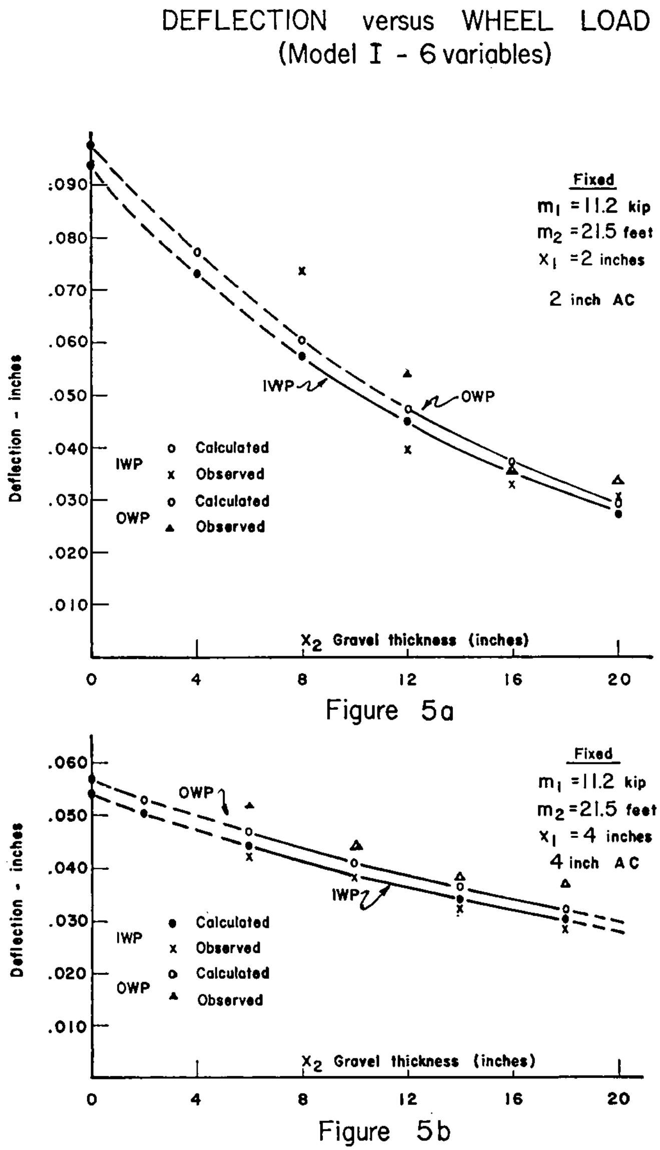


Figure 5. Deflection versus gravel thickness. (Model I—6 variables).

cases shown in Figure 5. The reduced equations for the plotted curves are shown below.

$x_1 = 2$ in., $m_4 = 3.5$ ft.:

$$\hat{b} = 0.5739\ (11.2)^{(1.1534 - 0.0254x_2)}$$

$x_1 = 2$ in., $m_4 = 9.5$ ft.:

$$\hat{b} = 0.5739\ (11.2)^{(1.1750 - 0.0254x_2)}$$

$x_1 = 4$ in., $m_4 = 3.5$ ft.:

$$\hat{b} = 0.5739\ (11.2)^{(0.9270 - 0.0134x_2)}$$

$x_2 = 4$ in., $m_4 = 9.5$ ft.:

$$\hat{b} = 0.5739\ (11.2)^{(0.8486 - 0.0134x_2)}$$

Upon comparing the observed points with the calculated points in Figure 5, it is apparent that this model does not yet account for all interactions among the selected variables since some of the errors are quite systematic. It would be quite natural to draw a different set of curves through these data than are given by the four equations above. If this is done, the implication is that a more valid model is being employed. Let us call this better fitting model Model III.

If Model III fits the observations better than the 6-variable Model I curves, then it must be that Model III differs from Model I either in mathematical form or in selected variables. But it would seem that unless the nature of Model III can be determined and analyzed as Model I has been, there will be little hope for generalizing from the observed data to other sections of pavement in time and space. On the other hand, if valid mathematical models which interrelate the independent variables can be found, then the experimental results are all contained in the model which may become the basis for both scientific generalizations and those which are made on faith.

It seems reasonable to suppose that nearly all of the objectives of a test road can be stated in terms that specify graphs which are to appear in the final report. The design of the experiment may then ensure that datum points are spotted in such a way that both the validity and precision of the reported graphs can be estimated from the data.

THE MULTIPLE CORRELATION PROBLEM

Nature of the Variables

In the mathematical structure of a multiple regression problem, only the dependent variable is assumed to be subject to sampling variation, while the independent variables are assumed to take on mean values which are fixed in the analysis. In the sketch below this situation is illustrated for two independent variables, x and y, and a dependent variable, b. The squares above each x and y combination represent experimental units which are sampled at each combination.

b_3		b_3		b_3		b_3
b_1 b_2		b_1 b_2		b_1 b_2		b_1 b_2
$x = 2,\ y = 3$		$x = 2,\ y = 4$		$x = 3,\ y = 2$		$x = 3,\ y = 3$

As has been described in the previous section, a regression problem consists in determining b as a function of x and y.

On the other hand, suppose that x and y are not assigned particular levels, but that they occur jointly with the variable b in each of a large collection of experimental units as indicated in the sketch below.

|x_{13}, y_{13}, b_{13}|x_{14}, y_{14}, b_{14}|

|x_{10}, y_{10}, b_{10}|x_{11}, y_{11}, b_{11}|x_{12}, y_{12}, b_{12}|

|x_6, y_6, b_6|x_7, y_7, b_7|x_8, y_8, b_8|x_9, y_9, b_9|

|x_1, y_1, b_1|x_2, y_2, b_2|x_3, y_3, b_3|x_4, y_4, b_4|x_5, y_5, b_5|

Now if a sample of experimental units is chosen at random from the "pile" shown above, we may say that the determination of the joint association among b, x, and y is a multiple correlation problem. The values taken on by all three of the variables are subject to the "luck of the draw." The nature of the second sketch is meant to imply that the main interest in a correlation problem is likely to be in the *degree* to which the three variables shown are jointly correlated. We may ask whether changes in x and y are accompanied by changes in b, and if so, in what way, or whether changes in b and x are associated with changes in y, etc. If all three variables are jointly associated, or correlated, then there are regression equations for relating x and y to b, x and b to y, or y and b to x, although one of these regression equations may be of more practical interest than another. However, we shall assume that the main goal in the multiple correlation problem is to arrive at an index which measures the degree of association

TABLE 6
ACTUAL DATA—WASHO—JUNE, 1954

SECTION 14 - 4 - ALL

	SET	log 100b	x_1	x_2	m_4	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
SINGLE AXLE	1	.36	4.00	10.25	9.5	60	133.4	96.4	22.5	134.4	93.0	22.0	26.3	24.7	24.6	25.3
	2	.34	4.00	9.50	9.5	62	137.3	97.4	23.8	136.5	90.4	23.0	26.6	24.8	25.0	29.1
	14	.48	4.25	8.75	9.5	62	131.8	87.6	26.2	137.6	95.7	26.3	24.7	23.9	27.2	25.4
TANDEM AXLE	1	.60	3.75	8.75	9.5	68	129.5	89.7	27.6	132.1	82.7	26.8	26.7	23.1	24.5	22.3
	2	.58	3.50	8.25	9.5	68	133.1	84.0	26.0	128.4	93.0	24.7	24.1	24.6	21.9	22.5
	14	.56	3.75	8.75	9.5	68	136.9	95.0	22.3	132.6	87.8	22.8	25.6	23.0	19.0	23.6

SECTION 22 - 2 - ALL

	SET	log 100b	x_1	x_2	m_4	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
SINGLE AXLE	1	.46	2.00	19.00	9.5	90	129.4	90.6	24.5	131.9	87.5	26.3	27.2	25.0	27.3	28.1
	2	.51	2.00	18.25	9.5	90	130.7	85.6	24.2	152.4	87.2	27.7	26.7	27.5	26.2	30.2
	14	.46	2.00	18.50	9.5	90	139.7	91.6	21.9	133.2	90.0	23.8	24.1	30.1	25.5	28.2
TANDEM AXLE	1	.46	2.25	19.00	9.5	81	134.6	91.8	25.1	130.9	86.0	25.4	24.1	25.9	25.8	25.1
	2	.38	2.25	19.00	9.5	81	129.2	91.2	24.0	129.6	85.1	26.1	26.3	19.2	27.1	27.5
	14	.36	2.50	18.00	9.5	81	143.3	88.4	21.7	134.2	89.5	26.0	23.1	21.3	27.2	26.1

among the variables. As an illustration of a multiple correlation problem, we have chosen sets of simultaneous data which arose at each of 14 spots within each of four test sections at the conclusion of the WASHO road test. Each set of observations listed in Table 6 corresponds to a particular spot on the pavement. The variables shown in Table 6 are defined as follows.

b = Deflection under 9-kip wheel load, in inches.

x_1 = Actual asphaltic concrete thickness, in inches.

x_2 = Actual gravel thickness, in inches.

m_4 = Distance of spot from center line, in feet.

y_1 = Temperature of surface, in degrees Fahrenheit.

y_2 = Gravel density in top 6 inches, in pounds per cubic foot.

y_3 = Soil density in top 6 inches, in pounds per cubic foot.

y_4 = Soil moisture content at top, in percent.

y_5 = Gravel density in 6 to 12-inch layer, in pounds per cubic foot.

y_6 = Soil density in 6 to 12-inch layer, in pounds per cubic foot.

y_7 = Soil moisture content at 4 inches below gravel, in percent.

y_8 = Soil moisture content at 9 inches below gravel, in percent.

y_9 = Soil moisture content at 15 inches below gravel, in percent.

y_{10} = Soil moisture content at 21 inches below gravel, in percent.

y_{11} = Soil moisture content at 27 inches below gravel, in percent.

For a given section, there are many possible observation spots, and each of these corresponds to a rectangle in the second sketch shown above. If we were to redraw the second sketch to conform with the illustrative example, each rectangle would appear as shown below.

$$\boxed{b, x_1, x_2, m_4, y_1, y_2, y_3, y_4, \cdots, y_{11}}$$

The 14 sets of data for each section would then represent a selection from a large "pile" of such rectangles. For illustrative purposes, we shall assume that the selection may be regarded as random. In this particular problem we are most interested in knowing how much of the variability in b can be accounted for by the variation in the other listed variables. The conventional index for measuring this association is the square of the multiple correlation coefficient, R^2. If we let s_0^2 be the variance of the b values in the sample of sets, then R^2 is the fraction of s_0^2 which can be accounted for by a linear relationship of b with the other variables. In these paragraphs, multiple correlation procedures are being used to study the within-replicate variations in deflection. In other words we are now discussing a means for analyzing the errors labelled e_{ij} in Figures 3*a* and 3*b*, and which have been averaged in the last line of Table 4. It might turn out, for example, that within-section variability in deflections is almost entirely explainable in terms of thickness, temperature, density, and moisture fluctuations in the pavement components. It is also possible that the same mathematical models may be used to represent within-replicate variation as are used in the regression problem described in the preceding sections. This question is partly a matter of generalizing from one type of experimental unit to another, for the experimental units are now pavement "spots" rather than pavement sections, as were the units for the regression model.

Multiple Correlation Analysis

The computational procedures in multiple correlation are identical with those used for a multiple regression analysis, although the interest now centers about correlation coefficients rather than regression equations. All possible simple correlation coefficients for the variables taken two at a time are shown in Table 7 for two of the selected WASHO sections. Each coefficient is based on 14 pavement spots, and deflections are in logarithmic form. For example, the simple r between deflection, log 100 b, and soil moisture, y_4, at the top of the basement soil, is shown in Table 7 to be 0.815 for section 14-4-T and -0.051 in section 22-2-T. The scatter diagrams for the observations from which these correlations were calculated are shown in Figure 6. These two examples were selected to show that the relation of deflection to environmental variables might be considerably different from one pavement design to another. In the present case, and with no other information to go on, it would appear that soil moisture and pavement thickness might have an interacting effect upon deflection. Furthermore, it could be that this interaction itself interacts with wheel load, for example. If we studied only the relation of deflection to moisture for fixed

TABLE 7
SIMPLE INTERCORRELATIONS OF THE VARIABLES FOR MULTIPLE CORRELATION ANALYSIS

SECTION 14-4-T

	$\log_{100b}$	x_1	x_2	m_4	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
log 100b		-.105	-.312	.369	.494	.031	.815	.531	-.219	.571	.562	.788	.536	.765
x_1	.196		.016	.072	-.286	.236	-.077	-.293	-.292	-.138	.094	-.325	-.249	-.133
x_2	-.137	-.901		-.630	-.659	.092	-.248	-.474	-.550	-.267	-.118	-.543	-.421	-.410
m_4	.354	.766	-.708		.531	.273	.448	.425	-.058	.421	.500	.627	.163	.341
y_2	.030	.427	-.468	.171		.197	.295	.538	.117	.102	.191	.629	.211	.594
y_3	.321	-.323	.231	-.115	-.375		-.130	.423	-.359	-.011	.221	-.044	-.276	.110
y_4	-.051	-.229	.399	-.242	-.237	-.166		.339	-.322	.689	.736	.756	.573	.462
y_5	-.019	.232	-.433	.525	.347	-.161	-.557		.096	.516	.382	.598	.420	.301
y_6	-.411	-.400	.169	-.562	-.173	-.056	.116	-.274		-.251	-.483	-.046	.215	-.022
y_7	-.160	-.130	.298	.007	.057	-.462	.389	-.170	.053		.767	.690	.346	.267
y_8	-.075	-.111	.277	.083	-.177	.057	-.177	.123	-.404	.091		.653	.067	.220
y_9	.430	.063	-.050	.265	.247	-.273	.366	.346	-.390	.101	-.086		.456	.619
y_{10}	.094	.292	-.390	.373	.015	.114	-.472	.124	.236	.134	.203	-.424		.437
y_{11}	.138	.187	-.220	-.005	-.061	.158	-.459	-.312	.293	.010	-.034	-.708	.698	

SECTION 22-2-T

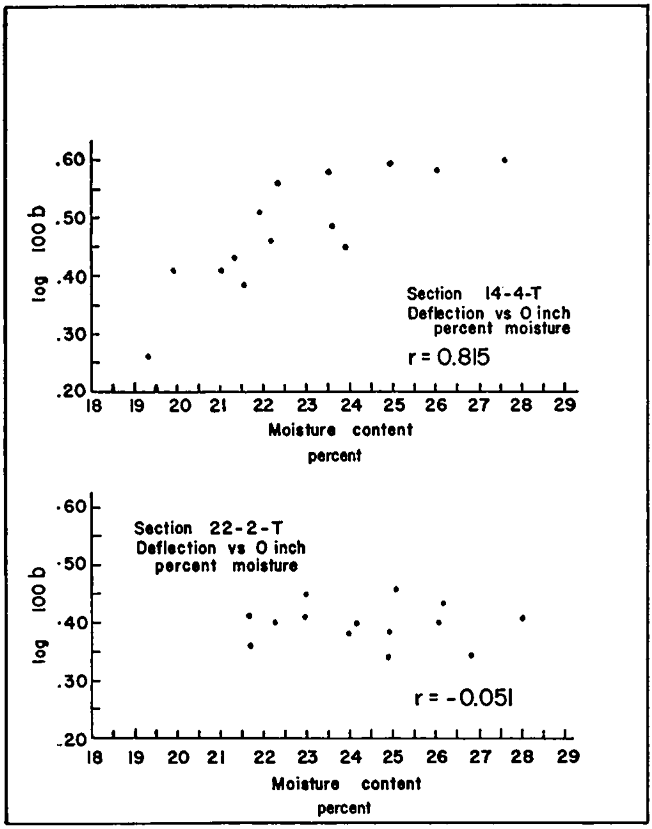


Figure 6.

pavement thickness and fixed wheel load, we would be doing precisely what the first experimenter did in the introductory example of this paper, namely to suppose that there were no interactions of the type just described. To put it another way, we should be assuming that an extra percent of moisture would have the same effect on deflection irrespective of the values for wheel load, pavement thickness, gravel density, etc. It is for this reason that an optimum analysis is one which will produce an overall model for representing the interacting effects of all selected variables, and on a test section basis, since the latter are presumably the basic experimental units of a test road. The linear regression equation for deflection versus the remaining variables in this correlation analysis is of the same form as those used in Models I and II of the previous sections. In fact we have used log 100 b as the dependent variable so that the last statement *would* be true. If the correlation analysis is to account for the errors in the last line of Table 4, then we should require that the same units be used for the dependent variable. The present regression model may be written as

$$\log 100\, b = A_0 + A_1x_1 + A_2x_2 + \cdots + A_{13}y_{11}.$$

Comparing this equation with the logarithm form of either Model I or II it may be seen that all three are of the same form, since now m_1 is constant due to the fact that all values of b are for a 9-kip wheel load. The equation just written differs, however, from Models I and II in that no interaction terms of the type x_1x_2, etc., have been included.

Just as in the multiple regression analysis, simultaneous linear equations must be solved to produce regression coefficients for the equation written above. We shall not show these coefficients in this paper since, as has been

TABLE 8
MULTIPLE CORRELATION ANALYSIS

Sections	Selected Variables x			Moisture Variables Only xx		
	S_o^2	R^2	S_e^2	S_o^2	R^2	S_e^2
14-4-S N = 14	27.30	.446	28.09	27.30	.562	22.21
14-4-T N = 14	96.75	.773	40.79	96.75	.862	24.79
22-2-S N = 14	11.49	.759	5.14	11.49	.787	4.55
22-2-T N = 14	12.95	.310	14.52	12.95	.672	6.90

x $x_1, x_2, m_4, y_2, y_3, y_4$

xx $y_4, y_7, y_8, y_9, y_{10}, y_{11}$

stated, we are more concerned with the value of the multiple correlation coefficient, and the amount of within replicate variability that is *not* accounted for by the equation above.

In Table 8, the results of the correlation analysis are shown, for two different selections of variables, and for each of the four chosen sections. The values of s_0^2 in Table 8 measure the variability among 14 deflections in a section, including differences in wheel paths. Such was not the case for the within-replicate mean square error for Table 4, and so the values of s_0^2 are expected to be higher in Table 8 than was the error mean square in the last line of Table 4. To put the values for s_0^2 in Table 8 into the units of Table 4, the former must be divided by 25×10^4. If this is done, it will be found that the average s_0^2 for the four sections shown in Table 8 is 0.00015, whereas the within replicate mean square of Table 4 was 0.00011. For each of two selections of variables, the remaining columns in Table 8 show the multiple R^2 and error variance, s_e^2, which estimates variability in deflections after accounting for the selected variables through their multiple linear association with b. If s_e^2 is the same as, or little less than, s_0^2, the conclusion is that the selected variables did not effectively reduce, or account for, the original variability of deflections within the section. Upon comparing the values of s_e^2 with the corresponding values of s_0^2 in Table 8, it can be noted that in these four sections, the moisture variables alone generally reduced the error variance by 50 percent or more. This has been rather consistently found to be the case in more complete studies than are reported here. In some of the thinner sections, over 90 percent of the within section variability in deflection has been attributable to moisture fluctuations through a linear relationship. For the selection x_1, x_2, m_4, y_2, y_3, y_4, there was a significant reduction in s_0^2 for two of the sections shown, but not for the other two. Such inconsistencies may possibly indicate that the linear regression equation should include more of the original variables and their interacting effects. However, the variances s_0^2 have only 13 degrees of freedom from the 14 deflection values which they represent. Six of these were used in fitting the regression equation, and so the error variances, s_e^2, are based on only 7 degrees of freedom. Inclusion of any more terms in the regression equation would make it very difficult to determine whether there is any real linear connection between b and the other variables, for as the degrees of freedom for s_e^2 are reduced, its value becomes quite unstable and unsatisfactory for comparison with the original variance s_0^2.

We have said that R^2 is that fraction of the original variance, s_0^2, which can be allocated to the linear regression relationship being used. It is not true, however, that the mean square of the unallocated variation is $(1 - R^2)s_0^2$ unless s_0^2 is based on a large number of observations. If there are k constants in the linear regression equation, then the connection among s_0^2, R^2, and s_e^2 is that

$$s_e^2 = \frac{(n-1)(1-R^2)}{(n-1-k)} s_0^2$$

where n is the number of sets of observations in the analysis.

In the separate section analyses of Table 8, $k = 6$ and $n = 14$, so $s_e^2 = {}^{13}\!/_7(1 - R^2)s_0^2$. It can be seen that with a small number of observations it is possible for s_e^2 to exceed s_0^2.

In the event that there is a significant reduction in s_0^2 through a sufficiently high R^2, it is pertinent to learn if the regression coefficients may be considered to be the same in replicate sections, and whether the same associations hold on up through the overall regression analysis. This is like asking whether the same relationships which are observed in the small, on "spot" type experimental units, will hold between replicate life-size experimental units, or test sections, and whether both of these investigations are in accord with the mathematical models for the component parts of the experiment, such as tangent layouts.

In conclusion we should like to say that the methods for analysis presented in this paper are meant to be systematic and logical, although they may not be optimum. The main attempt has been to outline analytical procedures which, upon development and improvement, may lead to a more uniform and objective interpretation of those experimental data for which the methods are applicable. We do not mean to imply that these methods are always appropriate, nor that they preclude other numerical analyses of the data. It would appear, however, that the procedures which have been described can lead to compact summarizations of the data. Regression models can serve as effective and efficient vehicles for generalizations which are to be made, as well as for bringing together experimental evidence and the theory of pavement behavior.

ACKNOWLEDGMENT

The senior author of this paper is heavily indebted to William N. Carey, Jr. for the many informative and stimulative discussions and debates which led to the conception and fulfillment of this paper. Much of the illustrative material has resulted from these conversations.

The other authors, Rex Leathers, Richard Lill, and Lee Mertz have made it possible to supplement the textual material with calculations, figures, and tables, and have contributed to an increased appreciation of those problems existing in the field of highway research.

Recognition is given to the numerical analysts of the Purdue Statistical Laboratory for their concern and interest in the development of high speed computer routines for meeting the requirements of those analyses which are set forth in the text.

Appreciation is also expressed to Robert Damon of the AASHO field office, for his difficult hours spent in the translation and transcription of the manuscript.

REFERENCES

1. Davies, O. L., *The Design and Analysis of Industrial Experiments*. London, Oliver and Boyd, 1954.
2. Bennett, C. A., and Franklin, N. L., *Statistical Analysis in Chemistry and the Chemical Industry*. New York, John Wiley & Sons, 1954.
3. "The WASHO Road Test," Part 2, Special Report 22, Highway Research Board, Washington, D. C., 1955.