Statistical Experiment Design of Laboratory Experiments

W. R. HUDSON, University of Texas

•IN BOTH the university and in the real world one often sees engineers and students, and professors for that matter, working with test data, usually plotting them in an effort to find out what they think the data are telling them. In doing this, we often massage the data in various ways. We may find it desirable to throw out "bad" data points or to bias the curve so that it looks good to us. In doing this we ought to keep in mind 2 things:

1. We can do anything we desire to the data as plotted on that graph and we can do anything we desire with the shape of the curve, and our judgment may make the line better or worse.
2. We are not usually interested in those data or that line because we will never use those data again. They are samples from the real world. What we are really interested in, usually, is what we can expect to happen under similar conditions either in the laboratories or in the outside world. In this case, what we do with the data or with the line makes quite a bit of difference because what the "truth" is or what nature intends to do will not be affected one bit by what we say about the data or the line.

I think this is best illustrated by the following example. I walked into a research laboratory one day and observed a friend of mine carefully plotting 3 points on a graph like that shown in Figure 1. The graph represents the results from a complex test (Test A) compared to the results of a very simple test (Test B) for many pairs of tests. Observe the perfect correlation line that he already had plotted and the large number of points that seem to fit the line very closely. I observed my friend plotting from his data sheets the 3 points shown in Figure 2 by crosses through dots. As I entered the room he seemed to be very disturbed. I said, "Friend, what is the matter?" He said, "Oh heck, I thought I had 3 new points for my test correlation, but as you can see I have only one." With that he erased the 2 external points. There is no question that the set of points does represent a group of soils for which the 2 tests did correlate well. Figure 3, however, shows the set of all correlations run by my friend. He has thrown away all the points not shown on Figures 1 and 2. Unfortunately, he proposes to use the correlation in all circumstances, that is, to substitute Test B for Test A as his specifications test. I think that you can easily see the error in this approach. He threw away the unwanted points, but nature will not throw them away.

There is another area of major concern in our laboratory work at a different level. Many of us do not treat our data as shown in Figure 1. We do, however, divide very complex problems into parts and look at them with a one-factor-at-a-time approach, ignoring the effects of the other variables and the resulting interaction. I would like to illustrate this with a little myth. Now remember that a myth while not literally true represents a truth.

> Once upon a time there were 3 boys. Their names were Ronnie Hudson, Johnny Beaton, and Isaac Newton. Now, Ronnie lived down in Texas where lemons and grapefruit are grown. As he rested under the trees, he noted with his high intelligence that grapefruit falling on his head from the tree hurt much more than lemons. Being very astute and very scientifically minded, he ran several experiments involving objects and the resulting force with which they

Paper sponsored by Special Committee on the Conduct of Research and presented at the 49th Annual Meeting.

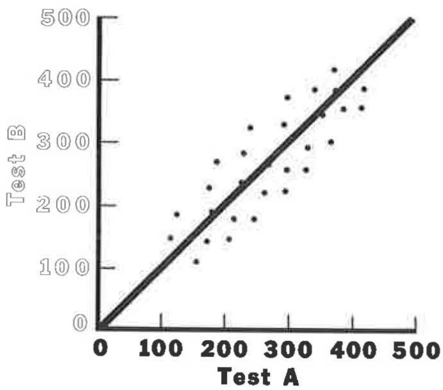

Figure 1.

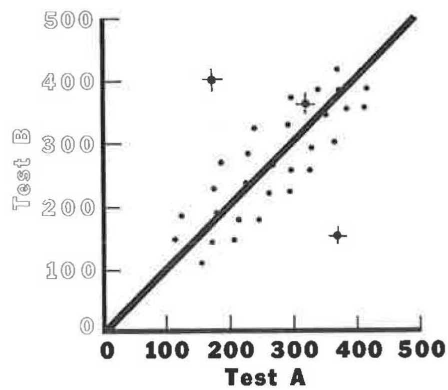

Figure 2.

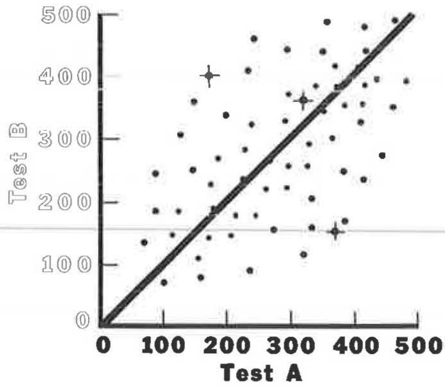

Figure 3.

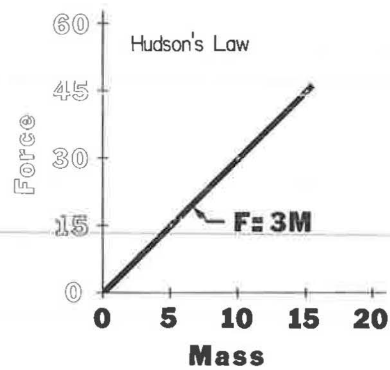

Figure 4.

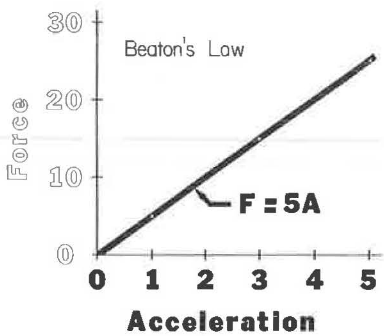

Figure 5.

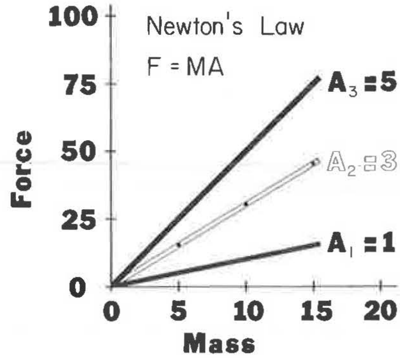

Figure 6.

struck his head. That year he went to the Fruit Research Board and reported that empirically grapefruit were more dangerous than lemons. After several more years of study, while obtaining a very sore noggin, he also refined his data as shown in Figure 4. From this he deducted Hudson's Law: Force is equal to a constant of 3 times the mass of the falling object. He dutifully went back to the national convention and reported these findings.

Not long thereafter young Johnny Beaton, living out in California, lay around under the date trees and grape arbors. He noticed with brilliance equal to Hudson's that when dates falling from tall trees struck his head they hurt much more than grapes falling from the low arbors. He also ran a scientific experiment and reported that the height of the tree was very important. After studying the data (Fig. 5), he noted carefully that the force of a falling object is really equal to a constant, 5 times the acceleration of the object. He dutifully went to the national convention and reported Beaton's Law, which seemed to discredit, by innuendo although not directly, Hudson's Law.

Shortly thereafter young Newton, who lived somewhere in the northwest among the apple trees, often sat to think under small apple trees and large apple trees, under trees growing big apples and little apples. He noted very astutely that the pain due to an apple hitting his head depended on the size of the apple as well as the height of the tree; however, the problem was too complex for easy solution. He promptly called in 3 of his friends (Paul Irick, Virgil Anderson, and Jack Youden) who considered the problem with him in some detail. Finally, an appropriate experiment was designed and apples of various size were dropped from various heights onto poor Isaac. Figure 6 shows very succinctly the data that resulted. Lo and behold, the mass really did not govern the results directly nor did the acceleration. Newton's Law was therefore born: Force is proportional to mass times acceleration. When this result was reported at the national convention, Beaton and Hudson were discredited, and it was reported that they subsequently have spent their declining years playing with dirt, cement, and asphalt.

To me this myth illustrates why our literature is full of contradictory conclusions. Like the blind men and the elephant, each of us deals with one part of the problem, such as the mass, while ignoring an equally important part, such as the acceleration.

A serious example is shown in Figure 7. If I looked only at an asphalt content of 7 percent, I would report that fine gradations are better than coarse gradations. However, at an asphalt content equal to 4 percent, the exact reverse is true. At a recent conference, a colleague of mine reported that the tensile strength of asphaltic mixes did not depend on gradation. Figure 8 shows the experiment design that led to his conclusions. In the meantime, we ran a rather large experiment involving 7 variables, one of which was gradation. The results of these tests show that, in truth, gradation has no direct effect. Figure 7 shows, however, an important interaction effect, and the tensile properties do vary with gradation because the effect of asphalt content changes as gradations get finer because more asphalt is absorbed by the mix.

Table 1 gives the factor space over which an asphalt stabilization experiment is to be conducted. Seven factors are included at 2 levels such. Figure 9 shows another presentation of the same factor space. In the figure, a better relationship can be seen among the variables because all possible combinations of the variables are shown and there is one block for each combination in the factorial. Furthermore, this form can be used to collect data. The answers for a specimen applicable to a particular set of factors can be written down in that block. If the rows and columns are summed, you have a quick look at the data, and it is possible to make quick estimates of factor effect.

If it is impossible to conduct as many tests as you desire, such as in the case of the factorials shown in Figure 9—128 specimens plus some replicates—we can run a special experiment taking some carefully selected samples. This is called a fractional factorial. Figure 10 shows a one-quarter fractional factorial, that is, one-quarter of the total number of blocks. It is not always desirable to take partial factorials; however, under certain circumstances they can be especially useful in cutting down on the size of an experiment while obtaining satisfactory results.

If a quarter fraction is too small, then a half fraction, as shown in Figure 11, can be tried; this is the first quarter plus the addition of a second quarter. You can see in Figure 11 the symmetry in the selection of samples to be taken. There are equal

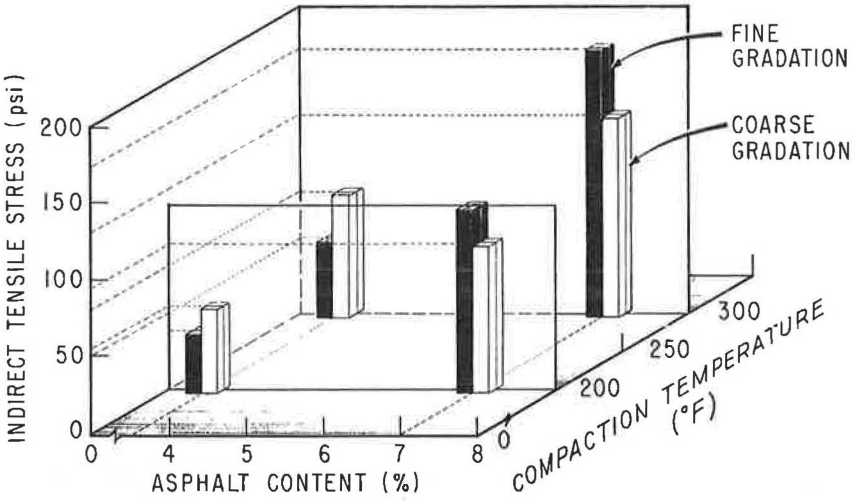

Figure 7.

numbers of all factors involved in the samples to be chosen so that there is balance within the experiment. Certainly there is not as much information as in a full factorial, because there are many fewer degrees of freedom in the experiment.

Table 2 gives the factors in a cement treatment experiment involving 5 factors at 3 levels and 4 factors at 2 levels each, or 5,832 specimens, an almost impossible number to deal with. In this case it is almost essential to deal with a one-quarter or

Aggregate Type / Asphalt Type / Gradation	Basalt			Limestone			Granite			Gravel		
	40-50	60-70	85-100	40-50	60-70	85-100	40-50	60-70	85-100	40-50	60-70	85-100
Coarse									6.0			
Medium		5.3-8.7				4.7	6.0	6.0	5.2 & 6.0			4.6
Fine									6.0			
X							7.9					

Figure 8.

even a one-eighth fractional factorial. As a matter of fact, we ran an even more complicated experiment by breaking the experiment up into 3 separate parts, but that is too complex to discuss here.

Let us turn our attention now to the results of some of these experiments. Figure 12 shows what statisticians call a main effect. This is the effect that a particular variable, in this case molding water content, has on the results of a test, in this case, indirect tensile strength. As we can see, increasing molding water content from 3 to 7 percent increases tensile strength from 60 to 200 psi, We might write an equation for the effect in the form

$$T = C_0 + C^1W$$

This is the type of effect that engineers most often model.

Figure 13, however, shows a 2-factor interaction. Unfortunately, if we look at the main effects and think that we know what the effect of molding water content is on tensile strength, we will be badly mistaken, because molding water content also interacts with gradation to affect the results. That is, the effect of water content is greater on fine gradations than it is on coarse gradations. An interaction is the term we use to describe the cross-product terms in a mathmatical model when the effect of one factor depends on the level of another factor in the experiment.

As if that were not complicated enough, there are sometimes 3-factor interactions present, which really causes problems. The 3-factor interaction involving water content, cement content, and type of curing is shown in Figure 14. As you can see, the

Curing Temp., °F	Compaction Temp., °F	Mixing Temp., °F	Asphalt Content, %	Aggregate: Limestone				Gravel			
			Gradation →	Fine		Coarse		Fine		Coarse	
			Viscosity (Space) →	AC 5	AC 20	AC 5	AC 20	AC 5	AC 20	AC 5	AC 20
75	200	250	5.5								
			8.5								
		350	5.5								
			8.5								
	300	250	5.5								
			8.5								
		350	5.5								
			8.5								
110	200	250	5.5								
			8.5								
		350	5.5								
			8.5								
	300	250	5.5								
			8.5								
		350	5.5								
			8.5								

Figure 9.

TABLE 1

FACTORS INCLUDED IN ASPHALT STABILIZATION EXPERIMENT

Factor	Level	
	Low	High
Aggregate type	Crushed limestone	Rounded gravel
Aggregate gradation	Fine	Coarse
Asphalt viscosity	AC-5	AC-20
Asphalt content, percent	5.5	8.5
Mixing temperature, deg F	250	350
Compaction temperature, deg F	200	300
Curing temperature, deg F	40	110

TABLE 2

FACTORS INCLUDED IN A CEMENT TREATMENT EXPERIMENT

Factor	Level		
	Low	Medium	High
Molding water content, percent	3	5	7
Curing time, days	7	14	21
Aggregate gradation	Fine	Medium	Coarse
Type of curing	Air dried	—	Sealed
Aggregate type	Gravel	—	Limestone
Curing temperature, deg F	40	75	110
Compactive effort	Low	—	High
Type of compaction	Impact	—	Gyratory shear
Cement content, percent	4	6	8

effect of cement content is much stronger at high water contents than at low water contents, but it is also stronger in general for sealed than for air-dried specimens as shown by the greater height of the black columns. Thus we see that we cannot show the effect of water content alone, because it causes other effects in combination with other factors; in this case, cement content and type of curing.

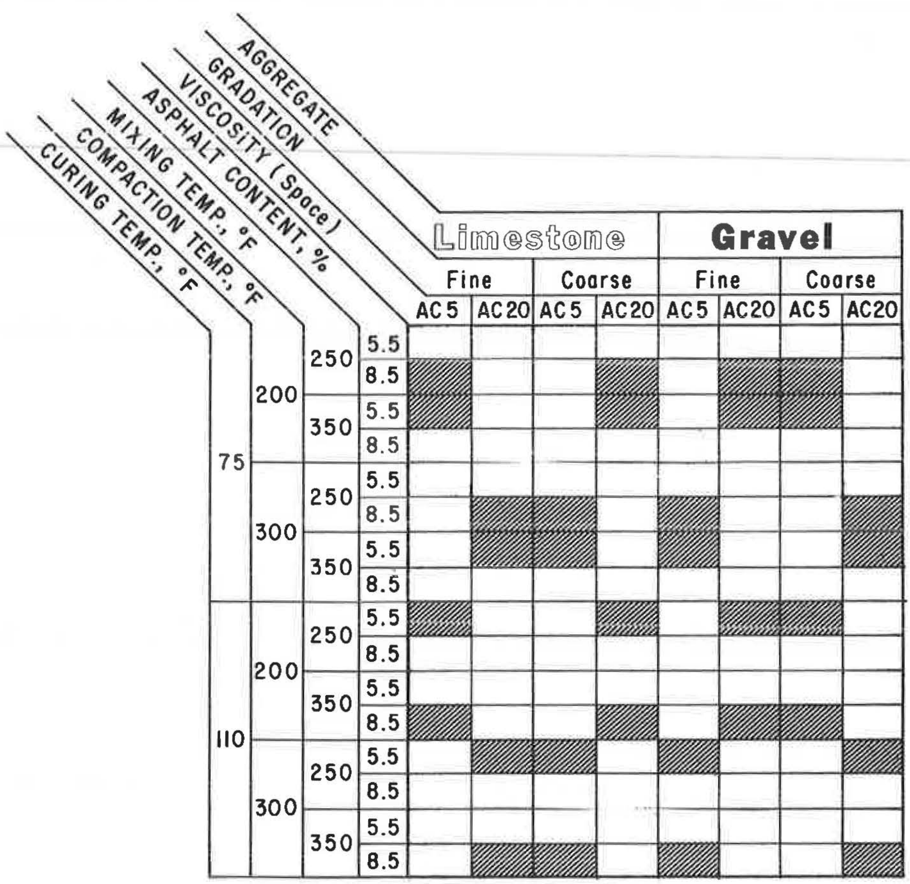

Figure 10.

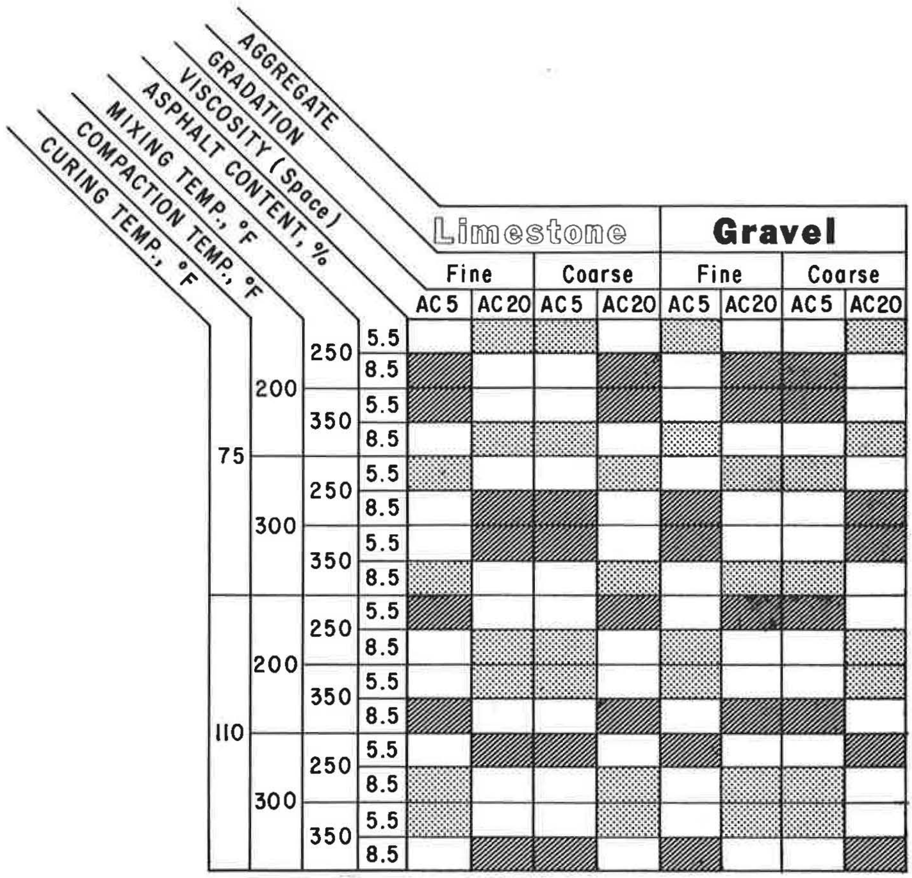

Figure 11.

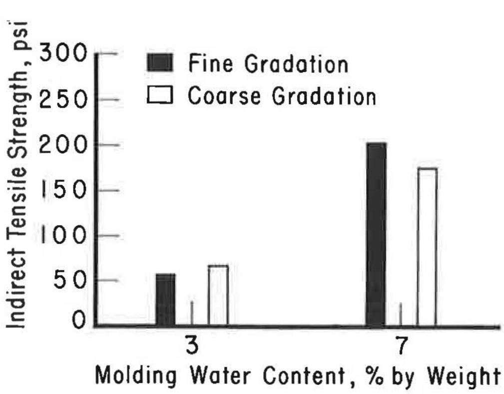

Figure 12.

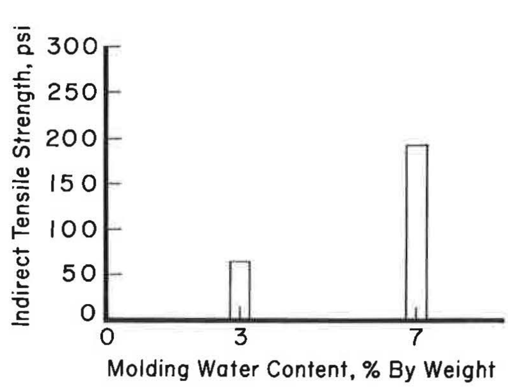

Figure 13.

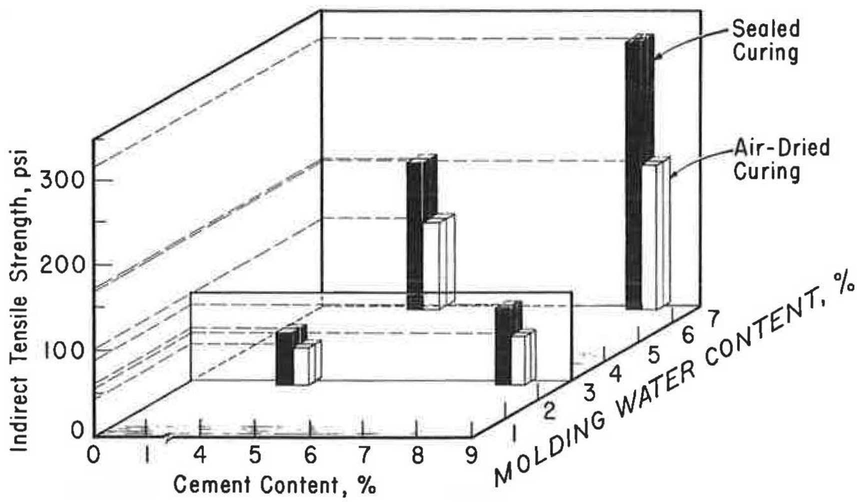

Figure 14.

I would summarize this discussion as follows:

1. The design of experiments forces an explicit look at the problem, the factors involved, and the levels of interest. This look improves our understanding of the problem. I have seen this so often with graduate students. They struggle for months to define and design their experiments, until they get a little exasperated with me because I do not do it for them. After they get the experiments designed, however, they find that the completion of the tests and the final analyses become much easier.

2. On complex problems, as most engineering research problems these days are, statistics is essential for getting the correct information from the test, getting the most information for your money in an experiment, and understanding quantitatively the findings of an experiment, that is, how good the answers are.

3. On very complex problems, the assistance of a statistician can be helpful; in fact, it is sometimes almost essential. However, for many simple problems, and after some consultations with a statistician, most engineers can, with the use of a good statistics book and some hard work, because reasonably good at designing and analyzing simple experiments.

4. Experimental design can be very valuable in conducting laboratory research experiments.