Application of Statistical Methods to Laboratory Concrete Freeze-Thaw Test Data

P. E. IRICK, *Research Associate, Statistical Laboratory,* and
J. B. BLACKBURN, *Research Engineer*
Joint Highway Research Project, Purdue University

THIS paper is written for researchers in the field of construction materials in an effort to demonstrate and explain a fairly complicated statistical analysis. The authors discuss experimental designs, the mathematical model, the assumptions which must be satisfied before the analysis of variance may be used, the mechanics of the analysis of variance, and a method for determining where the differences lie among a set of means. An example is used to help those who are unfamiliar with statistical terms and to illustrate the mechanics of the analysis of variance.

The authors believe a better understanding of the statistical aspects of experimental design and a more widespread use of the analysis of variance will lead to: (1) a more complete analysis of data than is now possible by interpreting graphs or tables; (2) better agreement among researchers as to the meaning of data because their methods of analysis will be more similar and less subject to personal bias; and (3) greater validity in test data once the importance is recognized of the roles played by replication and randomization in the design of experiments.

1. INTRODUCTION

● AT the present time there is a great need for more uniform methods of analyzing research data in the field of highway construction materials. This need can hardly be overemphasized in view of the ever-increasing research activity by private, federal and state agencies. Furthermore, the ever-increasing complexity of experimental designs and the growing tendency toward cooperative research involving several laboratories makes it more important than ever to use a method of analysis of the cumulative data that will yield comparable results for each analyst. Investigations such as studies of the durability of portland cement concrete or soil-cement mixtures; or the permeability of different concretes; or the compressive strengths of specimens from different concrete mix designs; or the stability of bituminous-aggregate mixtures; or any other investigation where numerical data are recorded, are particularly suited to the method of analysis that will be described in this paper. One value of the analysis lies in the fact that much personal bias is eliminated in interpreting the results. Further advantages are that the inferences one may make about his data are based on mathematical probabilities. For example, confidence limits may be computed for means, and differences between means may be determined to be real or non-existent, on the basis of mathematical probabilities.

The method of analysis that is referred to is known as the *analysis of variance.* It is widely used in many scientific areas including agriculture and the biological sciences. Only in recent years have the principles of experimental design and analysis begun to be used in engineering research.

Very few investigators appear to have used the analysis of variance in the field of construction materials testing. In 1953 Haskell (*1*) reported the use of the analysis of variance for analyzing concrete strength test data and Wright (*2*) referred to it in 1954 in a similar example. Duncan (*3*) showed how the analysis of variance could be used to advantage in a study of highway maintenance costs. It should

be mentioned in passing that researchers in the field of traffic engineering have used statistical techniques, including the analysis of variance, for some time.

It is the purpose of this paper to show how the analysis of variance may be used as an aid in interpreting materials test data. Our discussion will be facilitated by the use of an example for which we have chosen the data from a study of the freeze-thaw durability of laboratory concrete specimens. The purpose of the experiment was to investigate the freeze-thaw durability of 3 x 4 x 16 inch concrete specimens made from certain cement-aggregate combinations when duplicate sets of specimens were subjected to thawing in water at different temperatures. The cement-aggregate combinations involved four aggregates and two cements which were chosen for their range of field performance record. In highway pavements and exposed structural concrete two of the four aggregates have good service records and the other two have bad service records. One of the cements (alkali equivalent of 0.40) has a good service record and the other (alkali equivalent 0.75) has a doubtful service record. The four aggregates are gravels, the best of which is predominantly of igneous origin and the poorest of which contains approximately 50 percent chert. A simpler example could have been chosen but the premise is that many experiments will be at least as involved as this one.

In order to realize the benefits which arise from the use of statistical methods, as much attention must be paid to the experimental design as to the analysis of the experimental data. Conclusions which are drawn from a statistical analysis are always relative to the mathematical model which represents the design, and cannot be properly interpreted until the model is clearly understood.

This presentation is intended to be self contained, and so we have given an extensive explanation of the use of statistical methods in each of the phases of design, analysis, and interpretation. The only part of the development that has been intentionally slighted is concerned with algebraic mathematical manipulations.

The mechanics of the analysis of variance and other statistical procedures are rather completely discussed in Sections 3, 4.1, 5.2, 6.2 and 6.3. Sections 2, 4.2, 5.1, and 6.1 are mainly concerned with the fundamental concepts which underlie the statistical procedures. Section 7 is a discussion of the statistical inferences for the experiment that was used as an example.

The entire presentation has been directed at just one experimental design, that of the experiment described above, with the attitude that it is of importance for the reader to fully understand the fundamental statistical concepts which are connected with a single but fairly complicated analysis of variance. It would appear, however, that other experiments in the field of materials testing will have designs quite similar to that of the illustration, and that many of the principles discussed here are general and are readily adapted to other experiments. In the long run it will be advantageous to have a statistician assist in the design of experiments and to have computers to do the analyses. The fact remains, however, that the research person must finally interpret the results of the analysis of his data and the more familiar he is with statistical concepts the more valid will be his conclusions.

2. DESIGN OF THE EXPERIMENT

The design of an experiment may be regarded as a blueprint for the way in which data are to be collected on one or more dependent variables. The broad objective of most experiments in materials testing is to determine the effects of selected controlled factors upon some measurable characteristic, Y, of the material being tested.

In the example to be used throughout this paper, the dependent variable is a conventional measure of the durability of a concrete specimen,

$$Y = 100 \times \frac{E_f}{E_0},$$

where E_0 and E_f are the respective dynamic moduli of elasticity of the specimen at the outset and after f cycles of freezing and thawing.

2.1 *Nature of the Controlled Factors*

The experimentally controlled factors fall

logically into two classes. The first of these consists of factors which determine the nature of the material to be tested. These factors will be called formulation factors since their complete specification leads to a formula for preparing the material from which the test specimens are to be made. The second class consists of factors which specify how the materials are to be tested, and will be called test condition factors. In each of these two categories there may be factors which are held at one level throughout the experiment, and other factors which are varied over two or more levels.

An experiment is said to be a complete *factorial experiment* if each level of any factor is tested in every combination of levels of the remaining factors. This is the only type of experiment to be discussed in this paper, although much use is currently made of experimental designs which call for only a fraction of the complete factorial experiment (*4* and *5*). If the experimenter has enough beforehand information, or is willing to make the necessary assumptions concerning his variables, he may take advantage of the economies of these *fractional replications*.

The controlled factors for the illustrative example are listed and classified below.

Formulation Factors

One Level	*Several Levels*
Water-cement ratio	Sources of aggregates (4 levels)
Cement factor	Sources of ASTM Type 1 Cements (2 levels)
Condition of aggregate at time of mixing	
Ratio of fine to coarse aggregate	
Percent of entrained air	

Test Condition Factors

Type of concrete mixer	Thawing temperatures (3 levels)
Source of fine aggregate	
Number of freeze-thaw cycles per day	
Freezer temperature	

It was not feasible to attain a single value in each mix for the one level formulation factors, and so minor variations were permitted. Although these variations were small, they contributed to the uncontrolled variations to be discussed in the next section.

Still another distinction must be made among the experimental factors before an appropriate statistical analysis can be made. Some of the factors may be called *fixed* factors, while the remaining factors are called *random* factors. A factor is said to be fixed if the experimenter purposely selects its levels and wishes to generalize the results of the experiment to only these selected levels. On the other hand, if the experimental results are to be generalized to a larger set of levels for the factor than actually appear in the design, then the factor is said to be a random factor, provided that each level in the larger set had an equal chance to become one of the levels used in the design. For example, the four aggregates used in this experiment would be called fixed if they were the only aggregates about which conclusions were to be drawn. But if these four aggregates had been randomly selected from one hundred aggregates, say, then aggregates would be a random factor, and it would be valid to generalize the results of the experiment to the larger set of aggregates provided that an appropriate analysis of variance were performed. All the factors listed above will be called fixed factors.

2.2 *Nature of the Uncontrolled Variables*

The measured or dependent variable Y is expected to take on somewhat different values for any two mixes, or specimens, even though the mixes or specimens represent the same combination of controlled factor levels. This variation is called *error* variation and is generally brought about by many uncontrolled variables.

The uncontrolled variables may represent departures in the controlled factors from their purported levels. For example, not all the mixes in the experiment had identical percents of entrained air. Other uncontrolled variables may reflect variations which could have been controlled but were not. For example, laboratory personnel may change during the course of a long time experiment, and this could lead to variations in methods of handling and measurement of the test specimens. Another

specific source of error variation in measurements on concrete specimens arises because of different distributions of aggregate types, particle size and placement within the specimen.

It would be nearly impossible to list all uncontrolled variables which operate during the course of any particular experiment, but if a complete statistical analysis is to be made, the effects of these variables must be quantitatively estimated. Unless simplifying assumptions are made, it is necessary to make repeated measurements on Y for each combination of levels of the controlled factors. These *replications* will then provide a measure of the effect of the uncontrolled variables which are at work. If the replications can be regarded as random samplings from all replications which could have been made using the same experimental design, then the observed results can be generalized to a larger set of experiments, all performed with the same design and under the influence of the same uncontrolled variables.

If, for example, only one mix were made from each formulation, there would be no way to evaluate the effect of variables which have different effects from mix to mix, no matter how many specimens or tests were made from a single mix.

In the illustrative example there are eight combinations of formulation factor levels, four aggregates times two cements. For each of these eight, two mixes were prepared, and so each factor level combination was replicated twice. Without prior information on the extent of mix-to-mix variability, there is no way to tell whether two mixes represent a sufficient amount of replication.

Figure 1 shows a block diagram of the experimental design. Nine test beams were molded from each of the sixteen mixes, then three of these were assigned to each of the three temperature levels. Thus any aggregate-cement-temperature combination is represented by replication in two stages, two mixes and three specimens from each mix. Replication of beams within mixes provides a comparison of mix to mix variability with beam to beam variability.

The design of the experiment is not complete with the specification of controlled factor levels and the number of replications which are to be made. It must still be decided in what manner the replications are to be obtained. If the factor level combinations and the replications are to receive the effects of uncontrolled variation in a random fashion, then provision must be made for this in the design.

Randomization procedures are necessary to validate the significance tests and probability statements which accompany an analysis of variance. Randomization essentially insures that any net effect of the uncontrolled variables has the same chance to be associated with one combination of experimental factor levels as with any other, at least within the same replication. It is intuitively clear that one can never randomize with respect to all the uncon-

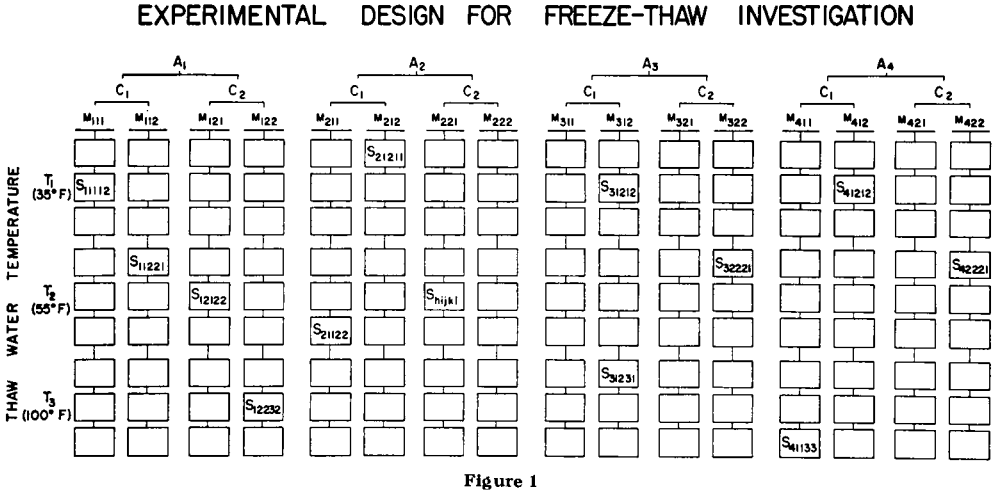


Figure 1

trolled variables, and in practice, randomization is usually performed with respect to time or location. In field experiments, factor level combinations might be assigned randomly to stretches of highway to be used in the experiment. In materials testing experiments it would seem that the major opportunities for randomization lie in manufacturing the mixes in a random order, and in assigning test specimens randomly to the levels of the test condition factors. Randomization can also be accomplished with respect to the placement of the specimens in the freeze-thaw apparatus.

The order of manufacture of the sixteen mixes can be randomized completely by, say, drawing the numbers from one to sixteen from a hat, then making the mixes in the order drawn. Another procedure would consist in making one mix of each of the eight formulation factor combinations in a random order, then repeating this procedure in a second block of the eight mixes. Whether these procedures are equivalent depends upon whether the mean value of Y in one block is expected to be the same as the mean of Y in the second block. Whether a completely randomized design, a complete blocks design, or an incomplete blocks design is used will depend largely upon how many mixes are prepared in a well defined time interval such as a day. Only one mix was prepared each day in the illustrative experiment. Complete randomization of the order of preparation of the mixes gives uncontrolled variables associated with day to day changes in the experimental conditions an equal chance to exert their effect on any one of the mixes to be prepared. Randomization of the temperature treatments to the nine specimens in each mix gives each specimen an equal chance to be subjected to any one of the three thawing temperatures. The order in which test beams are measured could also be randomized. To whatever extent the experimenter follows randomization procedures, he takes out insurance against making unfounded probability statements concerning similar experiments to which he may generalize (*6*).

At some point in the design stage of an experiment, a mathematical model should be written down. A mathematical model is essentially an equation which expresses the value of each observation in terms of symbols which represent the effects of the controlled factors and uncontrolled variables upon the observed value. It will be simpler to explain what is meant by the model after the data have been presented for the example. It should be emphasized that the experimental design, the mathematical model, and the statistical analysis are all interdependent. Although analyses of variance may have been performed without an explicit statement of the model, it is nevertheless true that every analysis of variance is determined by a particular model which is itself dictated by the design of the experiment.

3. PRESENTATION OF THE DATA

It will only be necessary to present a portion of the data that were taken for this experiment since the primary purpose is to illustrate the use of the analysis of variance. However, enough of the raw data will be shown to make it clear how the data were handled.

In the statistical analysis of durability test data it is necessary to decide at which intervals the data should be analyzed. Usually a study of the accumulated data will give some indication although not always. In the illustrative experiment the data at 40, 80, 120 and 160 cycles were chosen for analysis. A separate analysis was made at each of these cycle times although the bulk of the discussion of this paper will relate to the analysis of the data at 40 cycles. The results of these analyses are correlated since the same specimens appear in each analysis.

Table 1 shows the forty cycle data for each specimen from the two mixes containing Aggregate One and Cement One. The values recorded are those for the dependent variable

TABLE 1

TYPICAL TABLE SHOWING PERCENTAGE OF ORIGINAL DYNAMIC E RETAINED AFTER 40 CYCLES OF FREEZING AND THAWING FOR SPECIMENS CONTAINING AGGREGATE 1 WITH CEMENT 1

Mix No.	Specimen No.	Temperature		
		T_1	T_2	T_3
M_1	1	96.5	95.6	84.1
	2	94.9	95.5	82.8
	3	97.2	90.8	84.5
	$M \times T$ totals	288.6	281.9	251.4
	$M \times T$ means	96.2	94.0	83.8
M_2	1	93.9	94.6	78.4
	2	97.5	93.8	81.5
	3	96.5	94.5	77.8
	$M \times T$ totals	287.9	282.9	237.7
	$M \times T$ means	96.0	94.3	79.2

TABLE 2
TOTALS OF PERCENTAGE OF ORIGINAL DYNAMIC *E* AFTER 40 CYCLES OF FREEZING AND THAWING FOR THE INTERACTION EFFECTS *ACMT*, *ACT*, AND *ACM* AND FOR MAIN EFFECT *A*. ALSO SHOWN ARE *ACM* AND *ACT* MEANS

Aggregate No.	Cement No.	Mix No.	Temperature T_1	Temperature T_2	Temperature T_3	*ACM* Totals	*ACM* Means	*A* Totals
A_1	C_1	M_1	288.6*	281.9	251.4	821.9†	91.3	
		M_2	387.9	282.9	237.7	808.5	89.8	
		ACT totals	576.5‡	564.8	489.1	1630.4	90.6	
		ACT means	96.1	94.1	81.5			
	C_2	M_1	294.1	291.7	280.5	866.3	96.3	
		M_2	288.6	288.3	277.3	854.2	94.9	
		ACT totals	582.7	580.0	557.8	1720.5	95.6	
		ACT means	97.1	96.1	93.0			3350.9
A_2	C_1	M_1	235.5	262.4	238.4	736.3	81.8	
		M_2	262.8	257.6	258.5	779.2	86.6	
		ACT totals	498.3	520.0	497.2	1515.5	84.2	
		ACT means	83.1	86.7	82.9			
	C_2	M_1	256.2	268.6	268.5	793.3	88.1	
		M_2	252.7	260.3	277.3	790.3	87.8	
		ACT totals	508.9	528.9	545.8	1583.6	88.0	
		ACT means	84.8	88.2	91.0			3099.1
A_3	C_1	M_1	275.0	252.1	204.9	732.0	81.3	
		M_2	259.6	253.4	198.5	711.5	79.1	
		ACT totals	534.6	505.5	403.4	1443.5	80.2	
		ACT means	89.1	84.2	67.2			
	C_2	M_1	273.1	281.6	225.4	780.1	86.7	
		M_2	278.7	277.9	245.2	801.8	89.1	
		ACT totals	551.8	559.5	470.6	1581.9	87.9	
		ACT means	92.0	93.2	78.4			3025.4
A_4	C_1	M_1	264.6	264.7	205.3	734.6	81.6	
		M_2	257.6	263.4	219.3	740.3	82.2	
		ACT totals	522.2	528.1	424.6	1474.9	81.9	
		ACT means	87.0	88.0	70.8			
	C_2	M_1	293.8	293.7	270.1	857.6	95.3	
		M_2	265.7	271.4	208.1	745.2	82.8	
		ACT totals	559.5	565.1	478.2	1602.8	89.0	
		ACT means	93.2	94.2	79.6			3077.7
Grand total								12553.1

* Each *ACMT* entry is the total for three specimens.
† Each *ACM* entry is the total for nine specimens.
‡ Each *ACT* entry is the total for six specimens.

TABLE 3
TOTALS FOR $A \times C$ INTERACTION EFFECTS AND FOR MAIN EFFECTS *A* AND *C* AT 40 CYCLES

Agg. No.	Cement No. C_1	Cement No. C_2	*A* Totals
A_1	1630.4*	1720.5	3350.9
A_2	1515.5	1583.6	3099.1
A_3	1443.5	1581.9	3025.4
A_4	1474.9	1602.8	3077.7
C totals	6064.3	6488.8	Grand total; 12553.1

* Each $A \times C$ entry is the total for eighteen specimens.

for this experiment, namely the percentage of 0 cycle dynamic moduli of elasticity of the specimens. These data will be used in Sec. 5.2 along with those in Tables 2 through 5 to show the calculations for the analysis of variance.

Table 2 is an extension of the totals from Table 1 to all aggregates and cements. The totals which arise when only two controlled factors are considered at a time are shown in Tables 3, 4, and 5.

Table 6 shows all of the means for all the data taken. It is from this table that means are

obtained for making comparisons at 40, 80, 120 and 160 cycles. A method for making these comparisons will be explained in Section 6.3.

4. THE MATHEMATICAL MODEL

4.1 *Description of Terms in the Model*

The analysis of variance has been developed for a linear model, which is to say that Y is to be expressed as a linear combination of its sources of variation. The model must include a term to represent each source of variation which the experimental design gives to Y. If the linear model is not appropriate for Y, either a different dependent variable, or some transformed value of Y, e.g. log Y, should be used. The mathematical model which has been chosen for the illustrative experiment may be written as follows:

$$Y_{hijkl} = \mu + A_h + C_i + (AC)_{hi} + M_{hij} + T_k + (AT)_{hk} + (CT)_{ik} + (ACT)_{hik} + (MT)_{hijk} + S_{hijkl}$$

Values for the terms in the model are not actually observable from the experimental data, but are *parameters* which can only be evaluated when the experiment has been performed indefinitely many times with the same design and under the influence of the same uncontrolled variables. In what follows this larger set of experiments will be called the *generalized experiment.* It is precisely to this hypothetical set of experiments that the experimenter should be able to generalize his results and conclusions.

The above model is for any particular cycle of freezing and thawing. Five subscripts are necessary to locate any single specimen in Figure 1, or its Y value in Table 1. These subscripts are coordinates taking on the values

$h = 1, 2, \cdots, a = 4$ (aggregates)

$i = 1, \cdots, c = 2$ (cements)

$j = 1, \cdots, m = 2$ (mixes in an hi combination)

$k = 1, \cdots, t = 3$ (temperatures)

$l = 1, \cdots, s = 3$ (specimens in an $hijk$ combination)

For example, Y_{11232} is found to be 81.5 after 40 cycles since this is the value of Y in Table 1 for the first aggregate, first cement, second mix, third temperature, and the second such specimen.

TABLE 4
TOTALS FOR $A \times T$ INTERACTION EFFECTS AND FOR MAIN EFFECTS A AND T AT 40 CYCLES

Agg. No.	Temperature			A Totals
	T_1	T_2	T_3	
A_1	1159.2*	1144.8	1046.9	3350.9
A_2	1007.2	1048.9	1043.0	3099.1
A_3	1086.4	1065.0	874.0	3025.4
A_4	1081.7	1093.2	902.8	3077.7
T totals...	4334.5	4351.9	3866.7	Grand total; 12553.1

* Each $A \times T$ entry is the total for twelve specimens.

TABLE 5
TOTALS FOR $C \times T$ INTERACTION EFFECTS AND FOR MAIN EFFECTS C AND T AT 40 CYCLES

Cement No.	Temperature			C Totals
	T_1	T_2	T_3	
C_1	2131.6*	2118.4	1814.3	6064.3
C_2	2202.9	2233.5	2052.4	6488.8
T totals...	4334.5	4351.9	3866.7	Grand total; 12553.1

* Each $C \times T$ entry is the total for twenty-four specimens.

Without going into the mathematical analysis of the model let it suffice to say that if a model parameter is a mean in the generalized experiment, then the best estimate of this parameter is the corresponding mean obtained from the observed data. Particular values for each of the terms in the model can be estimated at 40 cycles from the data given in Tables 1, 2, and 6. We shall now explain briefly the meaning to be attached to each of the eleven terms in the model, giving a numerical estimate for the individual terms when the observation is Y_{11232} at 40 cycles.

μ is the overall parameter mean effect of all factors and variables upon Y, and is estimated by the grand mean of all the data. From Table 6, the mean of all 144 specimens at 40 cycles was 87.2, and this is the 40 cycle estimate of μ. Each possible experiment in the generalized experiment will produce a grand mean such as the 87.2 of this experiment, and the parameter μ is the average of all such grand means in the generalized experiment. We shall use the symbols $\mu \doteq 87.2 = \hat{\mu}$ where $\doteq$ is used in place of "is estimated by", and where a circumflex

TABLE 6
COMPLETE TABLE OF MEANS FOR PERCENTAGE OF 0 CYCLE DYNAMIC E

Thaw Temp.	Cycles of Freeze-Thaw	A_1		A_2		A_3		A_4		Across Cements				Across Aggregates		Across Aggregate and Cements
		C_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	A_1	A_2	A_3	A_4	C_1	C_2	
T_1 (35°F)	40	96.1[a]	97.1	83.1	84.8	89.1	92.1	87.0	93.3	96.6[b]	83.9	90.5	90.1	88.8[c]	91.8	90.3[d]
	80	96.5	96.6	80.6	85.0	84.6	88.1	84.9	90.9	96.5	82.7	86.3	87.8	86.6	90.1	88.4
	120	96.7	94.5	76.2	83.6	82.2	84.7	78.6	89.6	95.7	79.9	83.6	84.6	83.7	87.8	85.8
	160	94.3	96.7	75.1	82.6	82.3	85.8	78.3	92.5	95.4	79.2	84.2	85.4	83.0	88.9	86.0
T_2 (55°F)	40	94.1[a]	96.7	86.7	88.2	84.3	93.3	88.0	94.2	95.4[b]	87.4	88.7	91.1	88.3[c]	93.1	90.7[d]
	80	92.7	96.6	84.5	90.0	85.1	91.2	81.4	93.1	94.6	87.2	88.1	87.2	85.9	92.7	89.3
	120	94.5	93.8	83.7	89.4	77.8	87.2	78.0	91.2	94.1	86.3	82.9	84.6	83.3	90.4	86.9
	160	93.5	95.5	77.9	89.5	76.3	86.5	73.3	91.9	94.3	83.7	81.8	82.6	79.8	90.7	85.5
T_3 (100°F)	40	81.5[a]	93.0	82.9	91.0	67.2	78.4	70.8	79.7	87.2[b]	86.9	72.8	75.2	75.6[c]	85.5	80.6[d]
	80	76.2	89.4	79.7	91.7	55.2	71.3	51.4	69.4	82.8	85.6	63.2	60.4	65.6	80.4	73.0
	120	60.9	80.3	75.8	88.8	42.3	61.3	40.8	59.2	70.6	82.3	52.6	49.2	55.2	72.6	64.0
	160	59.7	78.9	68.1	85.8	41.0	62.3	37.6	60.8	69.3	76.9	53.8	48.1	52.2	72.1	62.4
Across temperatures	40	90.6[e]	95.6	84.2	88.0	80.2	87.9	81.9	89.1	93.1[f]	86.1	84.0	85.5	84.2[g]	90.1	87.2[h]
	80	88.5	94.2	81.6	88.9	75.0	83.5	72.6	84.5	91.3	85.2	79.2	78.5	79.4	87.7	83.6
	120	84.0	89.5	78.6	87.3	67.4	77.7	65.8	80.0	87.3	82.7	73.0	73.2	77.2	84.6	81.0
	160	82.5	90.4	73.7	86.0	66.5	78.2	63.1	81.7	86.8	79.9	73.9	71.9	71.8	84.1	78.1

[a] Mean of six specimens.
[b] Mean of twelve specimens.
[c] Mean of twenty-four specimens.
[d] Mean of forty-eight specimens.
[e] Mean of eighteen specimens.
[f] Mean of thirty-six specimens.
[g] Mean of seventy-two specimens.
[h] Mean of one-hundred-forty-four specimens.

accent denotes an estimate of the parameter having the accent.

A_h, C_i, and T_k are called *main effects*. A_h is the difference between μ and the parameter mean of all specimens made with aggregate h. If A_h is zero, aggregate h will be said to have no effect on Y. If A_h is positive or negative then specimens made with aggregate h have either more or less durability, respectively, than the mean of all the aggregates in the design. Corresponding statements can be made in describing the nature of C_i and T_k. To estimate the main effects from Table 6 we must subtract the table mean from the various factor level means. All 40 cycle estimates of main effects are listed below.

$$A_1 \doteq 93.1 - 87.2 = 5.9 = \hat{A}_1$$
$$A_2 \doteq 86.1 - 87.2 = -1.1 = \hat{A}_2$$
$$A_3 \doteq 84.0 - 87.2 = -3.2 = \hat{A}_3$$
$$A_4 \doteq 85.5 - 87.2 = -1.7 = \hat{A}_4$$

$$C_1 \doteq 84.2 - 87.2 = -3.0 = \hat{C}_1$$
$$C_2 \doteq 90.1 - 87.2 = 2.9 = \hat{C}_2$$

$$T_1 \doteq 90.3 - 87.2 = 3.1 = \hat{T}_1$$
$$T_2 \doteq 90.7 - 87.2 = 3.5 = \hat{T}_2$$
$$T_3 \doteq 80.6 - 87.2 = -6.6 = \hat{T}_3$$

Two letter combinations of controlled factors represent two factor *interaction effects* of the corresponding factors. Although a different explanation of interaction will be given later, at this point an interaction effect will be defined to be the amount by which the parameter mean of specimens, for the indicated combination of factor levels, fails to differ from μ by exactly the sum of the corresponding main effects. For example, $(AC)_{hi}$ is zero only if the parameter mean of specimens from aggregate h and cement i differs from μ by $A_h + C_i$. If such is not the case, then this combination of levels is said to contribute to the interaction of aggregates with cement. Similarly there may be interactions of aggregates with temperatures and of cements with temperatures. Using Table 6 and the previous estimates of main effects we have estimated half of the AC and AT interaction effects and all of the CT interaction effects in the following list:

$$(AC)_{11} \doteq (90.6 - 87.2) - \hat{A}_1 - \hat{C}_1$$
$$= (3.4) - (5.9) - (-3.0) = .5 = (\widehat{AC})_{11}$$
$$(AC)_{12} \doteq (95.6 - 87.2) - \hat{A}_1 - \hat{C}_2$$
$$= (8.4) - (5.9) - (2.9) = -.4 = (\widehat{AC})_{12}$$
$$(AC)_{41} \doteq (81.9 - 87.2) - \hat{A}_4 - \hat{C}_1$$
$$= (-5.3) - (-1.7) - (-3.0) = -.6$$
$$= (\widehat{AC})_{41}$$
$$(AC)_{42} \doteq (89.1 - 87.2) - \hat{A}_4 - \hat{C}_2$$
$$= (1.9) - (-1.7) - (2.9) = .7 = (\widehat{AC})_{42}$$

$$(AT)_{11} \doteq (96.6 - 87.2) - \hat{A}_1 - \hat{T}_1$$
$$= (9.4) - (5.9) - (3.1) = .4 = (\widehat{AT})_{11}$$
$$(AT)_{12} \doteq (95.4 - 87.2) - \hat{A}_1 - \hat{T}_2$$
$$= (8.2) - (5.9) - (3.5) = -1.2 = (\widehat{AT})_{12}$$
$$(AT)_{13} \doteq (87.2 - 87.2) - \hat{A}_1 - \hat{T}_3$$
$$= (0) - (5.9) - (-6.6) = .7 = (\widehat{AT})_{13}$$
$$(AT)_{41} \doteq (90.1 - 87.2) - \hat{A}_4 - \hat{T}_1$$
$$= (2.9) - (-1.7) - (3.1) = 1.5 = (\widehat{AT})_{41}$$
$$(AT)_{42} \doteq (91.1 - 87.2) - \hat{A}_4 - \hat{T}_2$$
$$= (3.9) - (-1.7) - (3.5) = 2.1 = (\widehat{AT})_{42}$$
$$(AT)_{43} \doteq (75.2 - 87.2) - \hat{A}_4 - \hat{T}_3$$
$$= (-12.0) - (-1.7) - (-6.6)$$
$$= -3.7 = (\widehat{AT})_{43}$$

$$(CT)_{11} \doteq (88.8 - 87.2) - \hat{C}_1 - \hat{T}_1$$
$$= (1.6) - (-3.0) - (3.1) = 1.5 = (\widehat{CT})_{11}$$
$$(CT)_{12} \doteq (88.3 - 87.2) - \hat{C}_1 - \hat{T}_2$$
$$= (1.1) - (-3.0) - (3.5) = .6 = (\widehat{CT})_{12}$$
$$(CT)_{13} \doteq (75.6 - 87.2) - \hat{C}_1 - \hat{T}_3$$
$$= (-11.6) - (-3.0) - (-6.6) = -2.0$$
$$= (\widehat{CT})_{13}$$
$$(CT)_{21} \doteq (91.8 - 87.2) - \hat{C}_2 - \hat{T}_1$$
$$= (4.6) - (2.9) - (3.1) = -1.4 = (\widehat{CT})_{21}$$

$$(CT)_{22} \doteq (93.1 - 87.2) - \hat{C}_2 - \hat{T}_2$$
$$= (5.9) - (2.9) - (3.5) = -.5 = (\widehat{CT})_{22}$$
$$(CT)_{23} \doteq (85.5 - 87.2) - \hat{C}_2 - \hat{T}_3$$
$$= (-1.7) - (2.9) - (-6.6)$$
$$= 2.0 = (\widehat{CT})_{23}$$

$(ACT)_{hik}$ represents the *three factor interaction* effect of aggregates with cements with temperatures. This parameter measures the amount by which the difference between the parameter mean of all specimens in an *hik* combination and μ fails to equal the sum of the three main effects and the three two-factor interaction effects which are involved. From Table 6 and the previous estimates of main effects and two factor interaction effects, we may now estimate three factor interaction effects at 40 cycles. Only a few of these estimates are given below.

$$(ACT)_{111} \doteq (96.1 - 87.2) - \hat{A}_1 - \hat{C}_1 - \hat{T}_1$$
$$- (\widehat{AC})_{11} - (\widehat{AT})_{11} - (\widehat{CT})_{11}$$
$$= 8.9 - (5.9) - (-3.0) - (3.1) - (.5)$$
$$- (.4) - (1.5) = .5 = (\widehat{ACT})_{111}$$
$$(ACT)_{112} \doteq (94.1 - 87.2) - (5.9) - (-3.0)$$
$$- (3.5) - (.5) - (-1.2) - (.6)$$
$$= .6 = (A\hat{C}T)_{112}$$
$$(ACT)_{113} \doteq (81.5 - 87.2) - (5.9) - (-3.0)$$
$$- (-6.6) - (.5) - (.7) - (-2.0)$$
$$= -1.2 = (A\hat{C}T)_{113}$$
$$(ACT)_{421} \doteq (93.3 - 87.2) - (-1.7) - (2.9)$$
$$- (3.1) - (.7) - (1.5) - (-1.4)$$
$$= 1.0 = (A\hat{C}T)_{421}$$
$$(ACT)_{422} \doteq (94.2 - 87.2) - (-1.7) - (2.9)$$
$$- (3.5) - (.7) - (2.1) - (-.5)$$
$$= 0.0 = (A\hat{C}T)_{422}$$
$$(ACT)_{423} \doteq (79.7 - 87.2) - (-1.7) - (2.9)$$
$$- (-6.6) - (.7) - (-3.7) - (2.0)$$
$$= -1.1 = (A\hat{C}T)_{423}$$

The three remaining terms in the model account for the effects of the uncontrolled variables upon Y. M_{hij} measures the difference between the parameter mean of all specimens in an *hi* combination which could have been made from the jth mix, and the parameter mean of all mixes which could have been made from aggregate h and cement i in the generalized experiment. This term, therefore, evaluates the extent of mix-to-mix variation within any particular combination of formulation factor levels. M_{hij} is estimated by subtracting the mean of all Y values in the *hi* combination from the mean of all Y values for mix j in the *hi* combination. Table 2 provides the necessary estimates, and a few of the M_{hij} estimates are listed below.

$$M_{111} \doteq 91.3 - 90.6 = .7 = \hat{M}_{111}$$
$$M_{112} \doteq 89.8 - 90.6 = -.8 = \hat{M}_{112}$$
$$M_{121} \doteq 96.3 - 95.6 = .7 = \hat{M}_{121}$$
$$M_{122} \doteq 94.9 - 95.6 = -.7 = \hat{M}_{122}$$

Since each mix occurs in combination with three temperature levels, the model provides a term $(MT)_{hijk}$ to evaluate the interaction of mixes with temperatures within the *hi* combination of aggregates and cements. It would perhaps be appropriate to assume that this interaction is zero but we have not done so in the present model. The subsequent analysis of variance will point out whether or not such an assumption is reasonable. The estimate of $(MT)_{hijk}$ is obtained from the same considerations that were used in deriving estimates of the two factor interaction effects among the controlled factors. For a fixed aggregate-cement combination, $(MT)_{hijk}$ is the amount by which a mix-temperature mean fails to differ from the aggregate-cement mean by exactly the sum of the differential effects of mix and temperature. The added complication here is that (MT) interaction effects are determined for each *hi* combination separately, just as though each *hi* combination represented the entire experiment. Thus, for example, $(MT)_{11jk}$ estimates are obtained from the part of Table 2 which pertains to the first aggregate and the first cement. A few of these estimates are given below.

$$(MT)_{1111} \doteq (96.2 - 90.6) - (91.3 - 90.6)$$
$$- (96.1 - 90.6) = -.6 = (\widehat{MT})_{1111}$$

$$(MT)_{1112} \doteq (94.0 - 90.6) - (91.3 - 90.6) - (94.1 - 90.6) = -.8 = (\widehat{MT})_{1112}$$

$$(MT)_{1113} \doteq (83.8 - 90.6) - (91.3 - 90.6) - (81.5 - 90.6) = 1.6 = (\widehat{MT})_{1113}$$

$$(MT)_{1121} \doteq (96.0 - 90.6) - (89.8 - 90.6) - (96.1 - 90.6) = .7 = (\widehat{MT})_{1121}$$

$$(MT)_{1122} \doteq (94.3 - 90.6) - (89.8 - 90.6) - (94.1 - 90.6) = 1.0 = (\widehat{MT})_{1122}$$

$$(MT)_{1123} \doteq (79.2 - 90.6) - (89.8 - 90.6) - (81.5 - 90.6) = -1.5 = (\widehat{MT})_{1123}$$

Finally, the parameter S_{hijkl} measures the difference between Y_{hijkl} and the mean of all possible specimens that could have been made from that part of the jth mix, in the *hi* combination, which was subjected to temperature k. There were only three such specimens in the design of the single experiment, but it is clear that even a single mix could be molded into nine beams in an infinite number of ways, since any interchange of aggregate particles would result in somewhat different specimens. The numerical estimate of S_{hijkl} is simply the difference between Y_{hijkl} and the mean of the three companion specimens in the *hijk* combination. The following examples are derived from Table 1.

$$S_{11111} \doteq (96.5 - 96.2) = .3 = \hat{S}_{11111}$$

$$S_{11112} \doteq (94.9 - 96.2) = -1.3 = \hat{S}_{11112}$$

$$S_{11113} \doteq (97.2 - 96.2) = 1.0 = \hat{S}_{11113}$$

$$S_{11231} \doteq (78.4 - 79.2) = -.8 = \hat{S}_{11231}$$

$$S_{11232} \doteq (81.5 - 79.2) = 2.3 = \hat{S}_{11232}$$

$$S_{11233} \doteq (77.8 - 79.2) = -1.4 = \hat{S}_{11233}$$

Now that each term in the model has been characterized verbally and numerically, we can show how Y_{11232} gets its value from the model. By substituting the respective estimates of parameters in the model, the equation becomes as follows.

$$Y_{11232} = \hat{\mu} + \hat{A}_1 + \hat{C}_1 + (\widehat{AC})_{11} + \hat{M}_{112} + \hat{T}_3 + (\widehat{AT})_{13} + (\widehat{CT})_{13} + (\widehat{ACT})_{113} + (\widehat{MT})_{1123} + S_{11232},$$

$$Y_{11232} = (87.4) + (5.9) + (-3.0) + (.5) + (-.8) + (-.6) + (.7) + (-2.0) + (1.2) + (-1.5) + (2.3) = 81.7$$

The observed value of Y_{11232} in Table 1 is 81.5 instead of the 81.7 which was just obtained from the model. This discrepancy is due to the fact that we have rounded off the means in Tables 2 and 6. It can be shown that the model becomes an identity for each Y when the described estimation procedure has been used.

The analysis of variance will consist in averaging, on a mean square basis over all 144 observations, each of the effects which are represented in the model for the observed value of Y.

4.2 *Assumptions Concerning the Terms in the Model*

An analysis of variance cannot be properly performed unless the statistical analyst knows what assumptions have been made on the terms in the mathematical model. Conversely, if the data have arisen from an experimental design which is in accord with an acceptable model, and if more or less conventional assumptions can be made concerning the terms in the model, then the data can be analyzed statistically. It is for these reasons that the statistical analysis of experimental data cannot be divorced from the design of the experiment itself.

From the definition of the main effects it follows that

$$\sum_{h=1}^{a} A_h = \sum_{i=1}^{c} C_i = \sum_{k=1}^{t} T_k = 0$$

since these effects were simply deviations from μ, and the sum of deviations about μ is zero for each controlled factor. Had one of these factors been a random factor, it would not be supposed that the sum of its effects in the experimentally used levels would be zero, but that the corresponding sum over all possible levels, called the expected sum, would be zero.

The respective variances of the level parameters for the main factors are defined by

$$V(A) = \sum_{h=1}^{a} A_h^2/(a-1);$$

$$V(C) = \sum_{i=1}^{c} C_i^2/(c-1);$$

$$V(T) = \sum_{k=1}^{t} T_k^2/(k-1).$$

The divisor in these variances has been taken to be one less than the number of levels for convenience of notation in formulas which are to follow. The only way in which any of these variances can be zero is for all the terms under the summation sign to be zero. For example, $V(A)$ will be zero if, and only if, $A_1 = A_2 = A_3 = A_4 = 0$. Otherwise $V(A)$ will be some positive value since each parameter is squared under the summation.

From the previous definitions of two and three factors interaction effects of the controlled factors it follows that

$$\sum_{h,i} (AC)_{hi} = \sum_{h} (AC)_{hi} = \sum_{i} (AC)_{hi} = 0$$

$$\sum_{h,k} (AT)_{hk} = \sum_{h} (AT)_{hk} = \sum_{k} (AT)_{hk} = 0$$

$$\sum_{i,k} (CT)_{ik} = \sum_{i} (CT)_{ik} = \sum_{k} (CT)_{ik} = 0$$

$$\sum_{h,i,k} (ACT)_{hik} = \sum_{h,i} (ACT)_{hik} = \sum_{h,k} (ACT)_{hik}$$

$$= \sum_{i,k} (ACT)_{hik} = \sum_{h} (ACT)_{hik}$$

$$= \sum_{i} (ACT)_{hik} = \sum_{k} (ACT)_{hik} = 0$$

The variances of these interaction effects are defined to be

$$V(AC) = \sum_{h,i} (AC)_{hi}^2/(a-1)(c-1);$$

$$V(AT) = \sum_{h,k} (AT)_{hk}^2/(a-1)(t-1)$$

$$V(CT) = \sum_{i,k} (CT)_{ik}^2/(c-1)(t-1);$$

$$V(ACT) = \sum_{h,i,k} (ACT)_{hik}^2/(a-1)\ (c-1)\ (t-1)$$

Again, none of these variances can be zero unless each term within the summation is zero.

For each combination of aggregates and cements, only m mixes ($m = 2$ in the example) appear in the experimental design. Since generalizations are to be made to an indefinite number of mixes in each combination, it is not supposed that $\sum_{j=1}^{j=m} M_{hij} = 0$, but that $E(M)_{hij} = 0$, where E stands for the average or expected sum of the M_{hij} over all possible mixes in an hi combination. The variance of all possible M_{hij} will be denoted by $\sigma_M^2 = E(M_{hij}^2)$. It is *assumed* that σ_M^2 is the same for any hi combination. It is further assumed that the complete set of M_{hij} follow a normal distribution and that the value of any particular M_{hij} is independent of any other M_{hij}. Thus, for each hi combination, we have assumed a normal distribution of independent M_{hij} values, and that each of these normal distributions has mean 0 and variance σ_M^2.

For the random variables $(MT)_{hijk}$ it is assumed that, for a particular hi combination, $\sum_{k=1}^{k=t} (MT)_{hijk} = 0$ and that the expected value of $(MT)_{hijk}$ over all possible mixes for any hi combination is zero. Furthermore, for a particular hi combination, all possible values of $(MT)_{hijk}$ are assumed to be normally and independently distributed with mean zero and variance of σ_{MT}^2. Symbolically, we have said that

$$E(MT)_{hijk} = 0 \qquad \sigma_{MT}^2 = E(MT)_{hijk}^2$$

where the expected values refer to all possible mixes in *any* hi combination.

For any $hijk$ combination, it is assumed that all possible specimens give values to the random variable S_{hijkl} which are independent and have a normal distribution with mean zero and variance σ_S^2.

That is, $E(S_{hijkl}) = 0$, $E(S_{hijkl}^2) = \sigma_S^2$ where the expectation refers to all possible specimens in any $hijk$ combination.

It should be noted that we have used the symbol V for variances of a finite number of parameters and σ^2 for variances of an indefinite number of values of random variables. In the illustrative model, it happens that these two symbols also distinguish between controlled factor variances and uncontrolled variation. This is so only because all the controlled factors were fixed factors in the design.

The estimates for the terms in the model will

sum to zero whenever the corresponding model terms either sum to zero or have expected zero sums. That this is true can be verified for the estimates given in the last section. For example, $\sum_i C_i = 0$, and $\sum_i \hat{C}_i = (-3.0) + (2.9) = 0$ to within round off error. Again, $E(M_{hij}) = 0$, and $\sum_j \hat{M}_{11j} = (.7) + (-.8) = 0$.

The foregoing list of conditions and assumptions on the quantities which are to be estimated in the experiment is rather imposing, and one may well wonder whether *any* of them have ever been precisely met in a particular experiment. Randomization procedures have the express purpose of eliminating biases and correlations in the observed values and therefore make the assumptions of independence and homogeneity of variance more tenable. The normality assumptions as well as the homogeneity of variance assumptions are often enhanced by transformations on the measured variable.

For a given experimental design the precision and accuracy of the estimates of the model terms can be increased only through proper randomization and sufficient replication. The degree of precision and accuracy can be stated in terms of mathematical probabilities only when assumptions can be made concerning the probability distributions of the random variables which affect the observations.

5. ANALYSIS OF VARIANCE

5.1 *Estimates of Variances of the Model Terms*

An analysis of variance is essentially a procedure for determining the magnitude of the variances of the terms in the mathematical model. In section 4.1 each term in the model was described and estimated numerically from the data. The variance of each term was defined and briefly discussed in section 4.2. In the summary given in Table 7 we have brought together the notation for the estimates of the model terms and their variance. The first term in the model, μ, has no variance since it has only one value in the generalized experiment. In addition, the quantities, Q, have been introduced as variances of data estimates of the terms in the model. These new variances have formulas which are quite similar to those for the variances of the corresponding model terms. In the generalized experiment there is only one variance for each of the model terms, but the value of the corresponding Q will change from experiment to experiment since new estimates will be obtained for the model terms from each experiment. It might be supposed that if the observed estimates were substituted into any particular formula for the variance of a model term, the result would be the desired estimate of the variance of the model term, but such is not the case in general. The effects have been listed in Table 7 in the same order by which they were discussed in section 4.1; main effects first, then two factor interactions, then three factor interactions, then effects of the uncontrolled variables, or error effects.

Many of the estimates indicated in the second column of Table 7 have been given in section 4.1. We shall now illustrate the use of these estimates in calculating a few of the Q's:

$$A \text{ effect: } Q(\hat{A}) = \sum_h \hat{A}_h^2/(a-1)$$

$$= [(5.9)^2 + (-1.1)^2 + (-3.2)^2 + (-1.7)^2]/3 = 16.4$$

$$CT \text{ effect: } Q(\hat{C}T) = \sum_{i,k} (\widehat{CT})_{ik}^2/(c-1)(t-1)$$

$$= [1.5)^2 + (.6)^2 + (-2.0)^2 + (-1.4)^2 + (-.5)^2 + (2.0)^2]/1 \times 2 = \frac{12.8}{2} = 6.4$$

$$M \text{ effect: } Q(\hat{M}) = \frac{1}{ac} \sum_{h,i,j} \hat{M}_{hij}^2/(m-1)$$

$$= [(.7)^2 + (-.8)^2 + \cdots\cdots + (6.2)^2 + (-6.3)^2]/4 \times 2 \times 1 = \frac{97.6}{8} = 12.2$$

The last column in Table 7 shows the expectations of the Q values, or what the Q values in any line would "average out" to be over the generalized experiment. These expectations can be derived mathematically when, and only when, decisions have been reached which specify the experimental design and its mathematical model, the assumptions on the terms in the model, and the randomization procedures which are to be used. The results in the last column are all-important since they form the basis for the statistical inferences which are to be made from the analysis of variance.

TABLE 7

Effect	Notation for the Model Terms and their Estimates	Variances of the Model Terms and Corresponding Variances, Q of the Estimates of the Model Terms	Expectations of the Corresponding Variances, Q
A (aggregates)	$A_h \doteq \widehat{A}_h$ $(h = 1, \cdots, a)$	$V(A) = \Sigma_h A_h^2/(a-1)$ $Q(\widehat{A}) = \Sigma_h \widehat{A}_h^2/(a-1)$	$\frac{\sigma_S^2}{cmts} + \frac{\sigma_M^2}{cm} + V(A)$
C (cements)	$C_i \doteq \widehat{C}_i$ $(i = 1, \cdots, c)$	$V(C) = \Sigma_i C_i^2/(c-1)$ $Q(\widehat{C}) = \Sigma_i \widehat{C}_i^2/(c-1)$	$\frac{\sigma_S^2}{amts} + \frac{\sigma_M^2}{am} + V(C)$
T (temperatures)	$T_k \doteq \widehat{T}_k$ $(k = 1, \cdots, t)$	$V(T) = \Sigma_k T_k^2/(t-1)$ $Q(\widehat{T}) = \Sigma_k \widehat{T}_k^2/(t-1)$	$\frac{\sigma_S^2}{acms} + \frac{\sigma_{MT}^2}{acm} + V(T)$
AC (agg. with cem.)	$(AC)_{hi} \doteq (\widehat{AC})_{hi}$	$V(AC) = \Sigma_{hi}(AC)_{hi}^2/(a-1)(c-1)$ $Q(\widehat{AC}) = \Sigma_{hi}(\widehat{AC})_{hi}^2/(a-1)(c-1)$	$\frac{\sigma_S^2}{mts} + \frac{\sigma_M^2}{m} + V(AC)$
AT (agg. with tem.)	$(AT)_{hk} \doteq (\widehat{AT})_{hk}$	$V(AT) = \Sigma_{hk}(AT)_{hk}^2/(a-1)(t-1)$ $Q(\widehat{AT}) = \Sigma_{hk}(\widehat{AT})_{hk}^2/(a-1)(t-1)$	$\frac{\sigma_S^2}{cms} + \frac{\sigma_{MT}^2}{cm} + V(AT)$
CT (cem. with tem.)	$(CT)_{ik} \doteq (\widehat{CT})_{ik}$	$V(CT) = \Sigma_{ik}(CT)_{ik}^2/(c-1)(t-1)$ $Q(\widehat{CT}) = \Sigma_{ik}(\widehat{CT})_{ik}^2/(c-1)(t-1)$	$\frac{\sigma_S^2}{ams} + \frac{\sigma_{MT}^2}{am} + V(CT)$
ACT (agg. with cem. with tem.)	$(ACT)_{hik} \doteq (\widehat{ACT})_{hik}$	$V(ACT) = \Sigma_{hik}(ACT)_{hik}^2/(a-1)(c-1)(t-1)$ $Q(\widehat{ACT}) = \Sigma_{hik}(\widehat{ACT})_{hik}^2/(a-1)(c-1)(t-1)$	$\frac{\sigma_S^2}{ms} + \frac{\sigma_{MT}^2}{m} + V(ACT)$
M (mixes in agg.-cem.)	$M_{hij} \doteq \widehat{M}_{hij}$ $(j = 1, \cdots, m)$	$\sigma_M^2 = EM_{hij}^2$ $Q(\widehat{M}) = \Sigma_{hij}\widehat{M}_{hij}^2/ac(m-1)$	$\frac{\sigma_S^2}{ts} + \sigma_M^2$
MT (mixes with tem. in agg.-cem.)	$(MT)_{hijk} \doteq (\widehat{MT})_{hijk}$	$\sigma_{MT}^2 = E(MT)_{hijk}^2$ $Q(\widehat{MT}) = \Sigma_{hijk}(\widehat{MT})_{hijk}^2/ac(m-1)(t-1)$	$\frac{\sigma_S^2}{s} + \sigma_{MT}^2$
S (specimens in agg.-cem.-mix-tem.)	$S_{hijkl} \doteq \widehat{S}_{hijkl}$ $(l = 1, \cdots, s)$	$\sigma_S^2 = ES_{hijkl}^2$ $Q(\widehat{S}) = \Sigma_{hijkl}\widehat{S}_{hijkl}^2/acmt(s-1)$	σ_S^2

Any changes in the design, model, or assumptions may lead to different results in this column, and consequently to different conclusions from the experimental data. This being the case, it is all the more obvious that the experimenter and statistician (perhaps the same person) must cooperate closely from the start. For the experimenter cannot understand the implications of the statistical analysis unless he knows the nature of the mathematical model and the assumptions which are placed on the model. Neither can the statistician derive appropriate results for the last column of Table 7 unless he clearly understands the experimental design and is able to work from an acceptable model.

A comparison of the third and fourth columns of Table 7 shows that, in general, the

expectations of the Q values are not equal to the variances of the model terms to which they correspond. Effects M, MT, and S, which represent uncontrolled variables, yield expectations which involve only the error variances σ^2_M, σ^2_{MT}, and σ^2_S. Effects which represent the controlled factors yield expectations which involve one or more of the error variances as well as the variances of the controlled factor effects.

For example, the expectation of $Q(\hat{A})$ is not $V(A)$ but is $\frac{\sigma^2_S}{cmts} + \frac{\sigma^2_M}{\text{cm}} + V(A)$. Furthermore the expectation of $Q(\hat{M})$ is $\frac{\sigma^2_S}{\text{ts}} + \sigma^2_M$, and not σ^2_M alone. In other words, $Q(\hat{A})$ does not estimate $V(A)$ directly, but also is expected to reflect mix to mix variation, while $Q(\hat{M})$ would estimate σ^2_M directly only if the specimens did not vary within mixes, in which event σ^2_S would be zero.

Now to decide whether the aggregates have any real (parameter) differential effects on durability, we must decide whether or not $V(A)$ is zero, since $V(A)$ is zero only if the four aggregates give the same mean for Y in the generalized experiment. To settle this question on a non-statistical basis, the experimenter might compare the variation in the four aggregate means against the variation found between mixes within the same aggregate. This is precisely the procedure suggested by the analysis of variance. By comparing the expectations of $Q(\hat{A})$ and $Q(\hat{M})$ it is clear that unless $Q(\hat{A})$ exceeds $(1/cm)\ Q(\hat{M})$ there is no reason to suspect that any real aggregate effects exist. The ratio of the expectations of $Q(\hat{A})$ and $(1/cm)\ Q(\hat{M})$ will be unity if $V(A)$ is zero, and greater than one if there really are different effects among the four aggregates. In section 6.1 we shall discuss *how much* greater than one this ratio must be before the conclusion is reached that there are real differential effects of the aggregates upon the durability of the concrete.

Thus, the analysis of variance makes it possible to systematically compare the effects of the controlled factors with the effects of the uncontrolled variables upon the experimental units. Moreover, the comparisons are made on the basis of expectations over the generalized experiment, rather than by using rules which would depend only upon the data from a single experiment.

Table 7 is really an analysis of variance table, and could be completed numerically by determining the remaining Q values. However it is more conventional to carry out the calculations using the formulas to be given in the next section. The conventional formulas have two advantages. In the first place, totals will be used instead of the means which have been used up to this point, and thus round off errors are minimized. Secondly, the conventional formulas yield weighted Q quantities. For example, $\hat{A}_h$ involves $cmts = 2 \times 2 \times 3 \times 3 = 36$ specimens for aggregate h. The weighted value of $Q(\hat{A})$ then becomes $cmts\ Q(\hat{A}) = 36\ (16.4) = 590.4$. This weighted Q value is called the *mean square* for aggregates.

The formula for the mean square, MS, for aggregates may be written in the form: $MS(A) = cmts\ Q(\hat{A}) = cmts \sum_{h=1}^{a} \hat{A}^2_h/(a-1)$. The numerator of this formula is called the aggregates *sum of squares*, and the denominator is called the degrees of freedom for aggregates. Numerically the formula gives $MS(A) = (36 \times 49.2)/3 = 1771.2/3 = 590.4$, where 1771.2 and 3 are the respective sum of squares and degrees of freedom for aggregates.

Similarly, each mix parameter estimate, $\hat{M}_{hij}$, represents $ts = 3 \times 3 = 9$ specimens, and so the mean square for mixes may be written

$$MS(M) = ts\ Q(\hat{M})$$

$$= \left(ts \sum_{h,i,j} \hat{M}^2_{hij}\right) \Big/ ac(m-1) = (9 \times 97.3)/8$$

$$= 875.7/8 = 109.5$$

In this case the sum of squares for mixes is 875.7 and has 8 degrees of freedom.

For each effect line of Table 7, the denominator of the Q value is the degrees of freedom for the effect, and the weighted numerator of the Q is the sum of squares for the effect.

These values for mean squares are not as accurate as those shown in Table 8, Section 5.2, because of round-off error resulting from using means.

The expectations of the mean squares are called *expected mean squares*, and will be the values shown in the last column of Table 7 when weighted. The weights are such that all divisors disappear in the expected mean squares. Thus the expected mean square for the A effect becomes $\sigma^2_S + ts\,\sigma^2_M + cmts\ V(A)$, while the expected mean square for the

M effect becomes $\sigma_S^2 + ts\,\sigma_M^2$. The ratio of these expected mean squares will be unity only if $V(A)$ is zero, and will exceed one otherwise.

Although the analysis of variance in Table 7 has been presented in the order in which the model terms were discussed in Section 4, it is more appropriate to order the lines in accordance with the sequence of terms in the mathematical model. In the model, the first terms (after μ) were for the formulation factor effects, A, C, AC, and the error effect, M, for mixes replicated in formulation factor combinations. Next came the test condition factor and its interactions with the formulation factor effects, T, AT, CT ACT. Finally, the model included error terms for the interaction of mixes with temperatures, MT, and for the specimen to specimen variation, S. Subsequent analysis of variance tables will be arranged in the order of the model terms rather than in the order of Table 7. Moreover, the tables which follow will be in terms of the weighted Q values. The advantage of the weighted Q's lies in the fact that every mean square in the table has been put on a per observation basis, and hence the mean squares become directly comparable.

5.2 *Conventional Computational Procedure*

The principal computation for the analysis of variance is that for finding sums of squares. The sum of squares must be determined for each effect term in the mathematical model. Each sum of squares, SS, is then divided by its degrees of freedom to obtain the mean square, MS, for the effect. Ratios are then formed using the mean square for the effect being tested as the numerator, and the appropriate error mean square as the denominator. These ratios are then compared with standard tables to determine whether or not there are significant differences among the levels of the effect being tested. These ratios and the subsequent significance tests will be discussed in Section 6.2. If differences are found to be significant, then it will be determined which of the levels are different from the others by a procedure to be described in Section 6.3.

The computation for the sum of squares for each term of the model proceeds as follows below. The data shown in the computations were taken from Tables 1 through 5. Restating the model:

$$Y_{hijkl} = \mu + A_h + C_i + (AC)_{hi} + (M)_{hij} + T_k + (AT)_{hk} + (CT)_{ik} + (ACT)_{hik} + (MT)_{hijk} + S_{hijkl}$$

where:

$$h = 1, \cdots, a = 4 \qquad k = 1, \cdots, t = 3$$

$$i = 1, \cdots, c = 2 \qquad l = 1, \cdots, s = 3$$

$$j = 1, \cdots, m = 2$$

SS for Aggregates, A (From Table 3 or 4)

$$\frac{\sum_h \left(\sum_{ijkl} Y_{hijkl}\right)^2}{cmts} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$= \frac{(3350.9)^2 + (3099.1)^2 + (3025.4)^2 + (3077.7)^2}{36} - \frac{(12{,}553.1)^2}{144} = 1{,}754.3$$

SS for Cements, C (From Table 3 or 5)

$$\frac{\sum_i \left(\sum_{hjkl} Y_{hijkl}\right)^2}{amts} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$= \frac{(6064.3)^2 + (6488.8)^2}{72} - \frac{(12{,}553.1)^2}{144} = 1{,}251.4$$

SS for the interaction AC (From Table 3)

$$\frac{\sum_{hi} \left(\sum_{jkl} Y_{hijkl}\right)^2}{mts} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts} - SS \text{ for } (A + C)$$

$$= \frac{(1630.4)^2 + \cdots\cdots + (1602.8)^2}{18} - \frac{12{,}553.1)^2}{144} - (1754.3 + 1251.4) = 89.4$$

SS for Mixes in Aggregate-Cement Combinations, M in AC (From Table 2)

$$\sum_{hi}\left[\frac{\sum_{j}\left(\sum_{kl} Y_{hijkl}\right)^2}{ts} - \frac{\left(\sum_{jkl} Y_{hijkl}\right)^2}{mts}\right]$$

$$= \frac{(821.9)^2 + (808.5)^2}{9} - \frac{(1630.4)^2}{18}$$

$$+ \frac{(866.3)^2 + (854.2)^2}{9} - \frac{(1720.5)^2}{18} + \cdots$$

$$\cdots + \frac{(857.6)^2 + (745.2)^2}{9} - \frac{(1602.8)^2}{18} = 874.1$$

In practice one would accumulate the positive terms, then all the negative terms, before making the indicated subtraction.

SS for Temperatures, *T* (From Table 4 or 5)

$$\frac{\sum_{k}\left(\sum_{hijl} Y_{hijkl}\right)^2}{acms} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$= \frac{(4334.5)^2 + (4351.9)^2 + (3866.7)^2}{48}$$

$$- \frac{(12{,}553.1)^2}{144} = 3{,}156.7$$

SS for the Interaction $A \times T$ (From Table 4)

$$\frac{\sum_{hk}\left(\sum_{ijl} Y_{hijkl}\right)^2}{cms} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$- SS(A + T) = \frac{(1159.2)^2 + (1007.2)^2 + \cdots + (874.0)^2 + (902.8)^2}{12}$$

$$- \frac{(12{,}553.1)^2}{144} - (1{,}754.3 + 3{,}156.7) = 1{,}729.5$$

SS for the Interaction $C \times T$ (From Table 5)

$$\frac{\sum_{ik}\left(\sum_{hjl} Y_{hijkl}\right)^2}{ams} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$- SS(C + T) = \frac{(2131.6)^2 + (2202.9)^2 + \cdots + (1814.3)^2 + (2052.4)^2}{24}$$

$$- \frac{(12{,}553.1)^2}{144} - (1{,}251.4 + 3{,}156.6) = 311.6$$

SS for the Interaction $A \times C \times T$ (From Table 2)

$$\frac{\sum_{hik}\left(\sum_{jl} Y_{hijkl}\right)^2}{ms} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$- SS(A + C + T + AC + AT + CT)$$

$$= \frac{(576.5)^2 + (564.8)^2 + \cdots + (565.1)^2 + (478.2)^2}{6} - \frac{(12{,}553.1)^2}{144}$$

$$- (1{,}754.3 + 1{,}251.4 + 3{,}156.7 + 89.4 + 1{,}729.5 + 311.6) = 89.6$$

SS for the Interaction $M \times T$ (From Table 2)

$$\sum_{hi}\left\{\left[\frac{\sum_{jk}\left(\sum_{l} Y_{hijkl}\right)^2}{s} - \frac{\left(\sum_{jkl} Y_{hijkl}\right)^2}{mts}\right] - \left[\frac{\sum_{k}\left(\sum_{jl} Y_{hijkl}\right)^2}{ms} - \frac{\left(\sum_{jkl} Y_{hijkl}\right)^2}{mts}\right]\right\}$$

$$- SS(M)$$

which collapses to:

$$\frac{\sum_{hijk}\left(\sum_{l} Y_{hijkl}\right)^2}{s} - \frac{\sum_{hik}\left(\sum_{jl} Y_{hijkl}\right)^2}{ms}$$

$$- SS(M) = \frac{(288.6)^2 + (287.9)^2 + \cdots + (270.1)^2 + (208.1)^2}{3}$$

$$- \frac{(576.5)^2 + (564.8)^2 + \cdots + (565.1)^2 + (478.2)^2}{6} - 874.1 = 405.7$$

SS for Specimens (From Table 1)

$$\sum_{hijkl} Y^2_{hijkl} - \frac{\sum_{hijk}\left(\sum_{l} Y_{hijkl}\right)^2}{s}$$

$$= (96.5)^2 + (94.9)^2 + \cdots + (81.5)^2 + \cdots$$

$$- \frac{(288.6)^2 + (281.9)^2 + \cdots}{3} = 2035.6$$

If all ten of the foregoing formulas are added together, the result is called the total sum of squares.

Total Sum of Squares (From Table 1)

$$\sum_{hijkl} Y_{hijkl}^2 - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts}$$

$$= (96.5)^2 + (94.9)^2 + \cdots + (81.5)^2 + (77.8)^2 + \cdots - \frac{(12{,}553.1)^2}{144} = 11{,}697.9$$

In the same way that each observation is broken up into additive terms in the mathematical model, the total sum of squares is thus broken up into sums of squares, one sum for each of the model terms with the exception of μ. The total sum of squares can be used as a partial check on the computation of the remaining sums of squares.

It is worthy of note that, in all of the computational formulas, each time a squared total enters into a computation it is divided by the number of observations that are in the total. This can be more clearly seen, perhaps, from the numerical illustrations of the formulas.

In Section 5.1, the sum of squares for aggregate effects was said to be the weighted sum of squared estimates of the aggregate parameter effects. That is, $SS(A) = cmts \sum_h \hat{A}_h^2$, where $cmts$ is the weight, or number of specimens entering into an aggregates mean. In Section 4.1, A_h was defined to be the difference between the mean for the hth aggregate and the grand mean. Putting these two definitions together in summation notation we have

$$SS(A) = cmts \sum_h \left(\frac{\sum_{ijkl} Y_{hijkl}}{cmts} - \frac{\sum_{hijkl} Y_{hijkl}}{acmts} \right)^2$$

Squaring the binomial and expanding,

$$SS(A) = cmts \left[\sum_h \left(\frac{\sum_{ijkl} Y_{hijkl}}{cmts} \right)^2 - 2 \left(\frac{\sum_{hijkl} Y_{hijkl}}{acmts} \right) \left(\frac{\sum_h \sum_{ijkl} Y_{hijkl}}{cmts} \right) + a \left(\frac{\sum_{hijkl} Y_{hijkl}}{acmts} \right)^2 \right]$$

Now the last two terms can be directly combined, and upon removing the square brackets we have the desired result,

$$SS(A) = \frac{\sum_h \left(\sum_{ijkl} Y_{hijkl} \right)^2}{cmts} - \frac{\left(\sum_{hijkl} Y_{hijkl} \right)^2}{acmts}.$$

Next we shall look at the derivation of the sum of squares for mixes-in-aggregates. In Section 4.1, $\hat{M}_{hij}$ was defined to be the difference between the mean of the jth mix in an hi aggregate-cement combination and the mean of the hi combination. That is,

$$\hat{M}_{hij} = \frac{\sum_{kl} Y_{hijkl}}{ts} - \frac{\sum_{ikl} Y_{hijkl}}{mts}.$$

Mixes are said to *nested* in aggregate-cement combinations because the mix parameters are the differential effects of mix means relative to the aggregate-cement means within which the mix occurs. From Section 5.1, the sum of squares for mixes is

$$SS(M) = ts \sum_{hij} \hat{M}_{hij}^2$$

$$= ts \sum_{hij} \left(\frac{\sum_{kl} Y_{hijkl}}{ts} - \frac{\sum_{ikl} Y_{hijkl}}{mts} \right)^2$$

Squaring the binominal and expanding the summation on the j subscript only,

$$SS(M) = ts \sum_{hi} \left[\sum_j \left(\frac{\sum_{kl} Y_{hijkl}}{ts} \right)^2 - 2 \left(\frac{\sum_{jkl} Y_{hijkl}}{mts} \right) \left(\frac{\sum_{kl} Y_{hijkl}}{ts} \right) + m \left(\frac{\sum_{jkl} Y_{hijkl}}{mts} \right)^2 \right]$$

Combining the last two terms and bringing ts inside the square brackets we have the formula as it was given above,

$$SS(M) = \sum_{hi} \left[\frac{\sum_i \left(\sum_{ki} Y_{hijkl} \right)^2}{ts} - \frac{\left(\sum_{jkl} Y_{hijkl} \right)^2}{mts} \right].$$

The only other derivation we shall give is for

the sum of squares for the AC interaction. In Section 4.1, $(\widehat{AC})_{hi}$ was defined to be the amount by which an hi aggregate-cement mean failed to differ from the grand mean by exactly $\hat{A}_h$ and $\hat{C}_i$. In summation form this means that

$$(\widehat{AC})_{hi} = \left(\frac{\sum_{jkl} Y_{hijkl}}{mts} - \frac{\sum_{hijkl} Y_{hijkl}}{acmts}\right) - (\hat{A}_h + \hat{C}_i).$$

From section 5.1 and Table 7, $SS\ (AC) = mts \sum_{hi} (\widehat{AC})_{hi}^2$. Substituting, we have

$$SS(AC) = mts \sum_{hi} \left[\left(\frac{\sum_{jkl} Y_{hijkl}}{mts} - \frac{\sum_{hijkl} Y_{hijkl}}{acmts}\right) - \hat{A}_h - \hat{C}_i\right]^2.$$

After squaring and expanding, this quantity reduces to

$$SS(AC) = \left[\sum_{hi}\left(\frac{\sum_{jkl} Y_{hijkl}}{mts}\right)^2 - \left(\frac{\sum_{hijkl} Y_{hijkl}}{acmts}\right)^2\right] - amts \sum_h \hat{A}_h^2 - amts \sum_i \hat{C}_i^2$$

$$= \sum_{hi} \frac{\left(\sum_{jkl} Y_{hijkl}\right)^2}{mts} - \frac{\left(\sum_{hijkl} Y_{hijkl}\right)^2}{acmts} - SS(A) - SS(C).$$

as was given in the computational formula.

We have given these derivations to indicate that the computational formulas are algebraically equivalent to the sums of squares which were defined in terms of the Q values of Table 7. The computational formulas are more precise in that totals are used, thus avoiding round off errors which accrue when means are used as in Section 5.1. The sum of squares for aggregates was found to be 1771.2 in Section 5.1, whereas the more precise formula of this section gave $SS(A) = 1754.3$. The derivations we have given should also serve to point out that the computational procedures depend entirely upon the model and that Sections 4.1, 4.2, and 5.1 do give the necessary background for understanding the computational formulas.

Table 8 is a conventional analysis of variance table, showing the results of the computations of this section in the sum of squares and mean square columns. The degrees of freedom for any effect is the denominator of the respective Q value given in Table 7. The expected mean squares in Table 8 are the expectations of the Q values which were shown in Table 7 after the latter have been weighted or multiplied by the number of specimens entering into each of the respective parameter estimates. For example, in Table 7,

$$EQ(\hat{A}) = \frac{\sigma_S^2}{cmts} + \frac{\sigma_M^2}{cm} + V(A).$$

Each $\hat{A}_h$ is based on $cmts$ specimens, and so the expected mean square for aggregates becomes $cmts\ E\,Q(\hat{A}) = \sigma_S^2 + ts\,\sigma_M^2 + cmts\ V(A)$ as shown in Table 8.

If one or more specimens are out of the test at any particular cycle chosen for analysis, a method must be selected for the analysis of the remaining data. The simplest procedure, although perhaps not the best, is to assign an arbitrary value to each missing specimen, and then perform the analysis as described above. The degrees of freedom for the specimens sum of squares must be reduced by one for each missing specimen.

The rule used in analyzing the illustrative data has been to assign to a missing specimen the mean value of the remaining specimens in any such $hijk$ cell. Other rules would perhaps give less upward bias to these cell means. A different dependent variable, such as a durability factor, might make the above modifications unnecessary. Any modification of the data will affect the analysis, however slightly, and the results must be interpreted with the modifications in mind.

6. STATISTICAL INFERENCE

6.1 *General Nature of Significance Tests and Confidence Interval Estimates*

If there is any basis at all for generalizing the results of a single experiment to a larger set of experiments, then to make any quantitative *inference* from the experimental data is to draw a conclusion about the value of one or more parameters which exist in the complete

TABLE 8
ANALYSIS OF VARIANCE TABLE FOR 40-CYCLE DATA
Model: $Y_{hijkl} = \mu + A_h + C_i + (AC)_{hi} + M_{hij} + T_k + (AT)_{hk} + (CT)_{ik} + (ACT)_{hik} + (MT)_{hijk} + S_{hijkl}$

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	"F" Ratio	F .05	Significance	Expected Mean Square
A	$(a-1) = 3$	1,754.3	584.8	5.4	4.07	*	$\sigma^2_S + ts\,\sigma^2_M$ + cmts $V(A)$
C	$(c-1) = 1$	1,251.4	1,251.4	11.4	5.32	*	$\sigma^2_S + ts\,\sigma^2_M$ + amts $V(C)$
AC	$(a-1)(c-1) = 3$	89.4	29.8	.3	4.07	NS	$\sigma^2_S + ts\,\sigma^2_M$ + mts $V(AC)$
M in AC	$ac(m-1) = 8$	874.1	109.3	5.1	2.05	*	$\sigma^2_S + ts\,\sigma^2_M$
T	$(t-1) = 2$	3,156.7	1,578.3	62.2	3.63	*	$\sigma^2_S + s\,\sigma^2_{MT}$ + acms $V(T)$
AT	$(a-1)(t-1) = 6$	1,729.5	288.3	11.4	2.74	*	$\sigma^2_S + s\,\sigma^2_{MT}$ + cms $V(AT)$
CT	$(c-1)(t-1) = 6$	311.6	155.8	6.1	3.63	*	$\sigma^2_S + s\,\sigma^2_{MT}$ + ams $V(CT)$
ACT	$(a-1)(c-1)(t-1) = 6$	89.6	14.9	.6	2.74	NS	$\sigma^2_S + s\,\sigma^2_{MT}$ + ms $V(ACT)$
$(M\text{in}AC)T$	$ac(m-1)(t-1) = 16$	405.7	25.4	1.2	1.75	NS	$\sigma^2_S + s\,\sigma^2_{MT}$
S	$acmt(s-1) = 96$	2,035.6	21.2				σ^2_S
Total....	$acmts - 1 = 143$	11,697.9					

set of experiments. By a statistical *inference* we shall mean a probability statement which specifies a value, or range of values, for one or more parameters. The statistical inferences which follow an analysis of variance will result either from tests of hypotheses on the parameters or from interval estimates of these parameters.

As a simple example, we note from Table 6 that the mean for Y at 40 cycles was 93.1% for all specimens made from the first aggregate. As in section 4.1, this mean leads to an estimate of the parameter A_1, by $\hat{A}_1 = 93.1 - 87.2 = 5.9\%$. In order to generalize this observed differential effect of the first aggregate, we must ask what value A_1 really has in the generalized experiment.

To answer this question without using statistical inference, the experimenter would logically note the extent of random variation from mix to mix, and then perhaps conclude either that "it appears that the first aggregate really does have some differential effect on durability", or that "there does not seem to be any reason to conclude that the first aggregate is any different from the average, when error variation is taken into account". Thus, the experimenter might use his own judgment to reach one or the other of these two conclusions.

If statistical inference is used, the inferential process becomes more systematic and less subjective. First, the experimenter hypothesizes that A_1 is really zero. Then he turns to statistical methodology for a rule which will be based on probability theory, and which will lead him to an acceptance or a rejection of his hypothesis. This rule is called the *test* of the hypothesis, and should have several desirable properties. It should be objective in that it can be stated before any data at all are taken. It should dispose of the hypothesis only after quantitatively comparing the parameter estimate with an appropriate measure of the reliability of the estimate. The rule should lead to an acceptance of the hypothesis if the data estimate is "close enough" to the hypothetical value of the parameter, and to a rejection of the hypothesis if the estimate is "too far" from the value given to the parameter by hypothesis. The rules of statistical tests will define "close enough" and "too far" in terms of the risks that erroneous inferences will be drawn. It is this last property that represents the major distinction between statistical and non-statistical inference. Finally, the conclusions from statistical inference will have the same interpretation by all who understand the rule.

Whether statistical inference is used or not, there is the risk that the hypothesis will be rejected even though it is true. For example, there is some likelihood that A_1 will be con-

cluded to be different from zero when in fact A_1 is zero. We shall call this risk α, where α is between zero and one and is called the *significance level* of the test. We shall suppose that the α risk has been specified before the data are analyzed, and that this risk has been taken to be a conventional value such as .05 or .01.

We shall not actually demonstrate the test for the hypothesis that A_1 is zero, but if the appropriate rule were used for this test with $\alpha = .05$, then the inference would be either that A_1 is zero or that A_1 is different from zero. In the latter case, the hypothesis would have been rejected by the rule, and the estimate A_1 would be said to differ significantly from zero at the 5% level. For brevity, such a conclusion may be stated by saying that there is a *significant* effect of the first aggregate on durability, but this statement is ambiguous unless the value of α and the hypothetical value of A_1 are clearly implied.

In particular, if 5.9% is declared to be significant at the 5% level, the meaning to be conveyed by this statement is (1) a rule has been used to accept or reject the hypothesis that $A_1 = 0$, (2) the rule led to a rejection, and (3) there was a 5% risk that a rejection would occur even though A_1 were truly zero.

A common misconception concerning the outcome of a significance test is that a significant effect is a large effect. It can be that a statistically significant effect is of no practical importance. On the other hand, an observed effect may be large enough to be of practical importance and still not be declared to be significant by the significance test. This latter situation usually arises when there is a relatively large amount of error variation and/or a relatively small amount of data information on the effect.

If an effect is not declared significant, then the hypothesis of zero effect cannot be rejected, and may therefore be said to be acceptable or tenable. Such an outcome does not *prove* there to be a zero parameter effect, but presumably results in no different behavior on the part of the experimenter than if a zero effect has been proved to be the case. Another pitfall in the interpretation of the outcome of a significance test is to suppose that $1 - \alpha$ is the probability of making a correct inference about the parameter being tested. In the example of this section, 5% is the risk that the effect of the first aggregate will be declared significant when the truth is that the parameter A_1 is zero, and $1 - \alpha = .95$ is the degree of assurance that A_1 will not be declared significant when in fact A_1 is zero. Now if A_1 is *not zero*, there is also the risk that the significance test will lead to an acceptance of the hypothesis that A_1 is zero. Although the risk of this second type of erroneous conclusion can be calculated (7), we shall not do so in this paper. It turns out that this risk of falsely accepting the tested hypothesis, i.e. failure to claim significance even though a non zero effect exists, depends upon the true (non-zero) value of A_1, upon the degrees of freedom in the error term, and upon the significance level used in the test. For a fixed α, the risk of the second kind decreases as the true A_1 value is farther and farther from the hypothetical value, and for a fixed non-zero A_1, this risk decreases as a larger α is used. Suppose, for example, that A_1 is 1% instead of zero. It could be that there is a .30 risk of accepting the hypothesis that $A_1 = 0$ if $\alpha = .01$, whereas this risk might be only .10 if α is taken to be .05. For a specified non zero A_1 and α, the risk of the second kind can also be reduced by increasing the degrees of freedom for the measure of unreliability in the data estimate of A_1. In the tests of hypotheses which are to follow in the next section, we shall use a 5% level of significance.

It may be that the experimenter wishes to estimate the value of a parameter, rather than to test some hypothesis on the parameter. For example, he may ask for the magnitude of A_1 instead of for the conclusion that the first aggregate either has or does not have any effect on durability.

A non-statistical estimate of A_1 might be, say, 5.9% ± 2%, where the ±2% represents a more or less intuitive judgment upon the reliability of $\hat{A}_1 = 5.9\%$. The corresponding statistical inference might turn out to be the claim that, with 95% confidence, A_1 is somewhere in the interval 5.9% ± 4%. In the statistical inference, the ±4% has been determined by deciding that it is permissible to run a 5% risk that the interval to be given will not, in fact, contain the parameter being estimated. It could be that there is a .75 risk that 5.9 ± 2% does not contain A_1. Or it might happen that the estimate 5.9 ± 2% is associated with 99.99% confidence, whereas a 95% confidence interval would be, say, 5.9 ± .5%. To determine confidence interval estimates for param-

eters, the experimenter must select the appropriate procedure from the body of statistical methodology and also specify the risk he is willing to assume in making an erroneous estimate of the parameter.

And so these two techniques of statistical inference, testing hypotheses and interval estimation, lead to conclusions from the experimental data to the generalized experiment, even though only one experiment has been performed. The inferences are accompanied by stated risks which relate to the complete set. For if a significance level of .05 is used in testing a hypothesis *which is really true* in the generalized experiment, then it can be expected that in 5 times out of each 100 experiments the tested hypothesis will be erroneously rejected. Or if a parameter were to be estimated with 95% confidence from each experiment in the generalized experiment then it would be expected that in 95 experiments out of each 100, the interval estimates would in truth contain the parameter being estimated.

If the experimental procedure does not conform with the experimental design and its mathematical model, or if the model is inappropriate for the measured variable, or if the assumptions on the terms in the model are invalid, then the actual risks for drawing erroneous conclusions may be considerably different than those which are implied by the statistical inferences. If, for example, randomization is neglected, then it is practically impossible to know to what extent the risks are altered from their purported levels. If some of the assumptions such as normality of the distribution of experimental errors, or homogeneity of error variation, are untenable, then it may be that, say, a 90% confidence interval estimate is really only an 80% confidence interval. Although no blanket statement is possible, much of the research in mathematical statistics has shown that the statistical analyst does not need to feel uncomfortable in stating his risks of error unless the assumptions on the model are rather grossly invalid.

Before deciding that he will have nothing to do with gambling as he draws conclusions from his experimental data, the experimenter must come to realize that any generalization represents a gamble, and that the attitude of the scientist must surely be to know of the odds with which he is betting. Through the use of statistical inference, the experimenter can give himself as long odds as he feels is necessary, but if no risks are permissible, then the experimenter must either draw no inferences at all, or perform the same experiment ad infinitum.

6.2 *Significance Tests on Means*

The first inferences that are to be drawn from the analysis of variance concern the variances of the controlled factor parameters. These seven variances, $V(A)$, $V(C)$, $V(AC)$, $V(T)$, $V(AT)$, $V(CT)$, and $V(ACT)$ appear in the last columns of Tables 7 and 8.

If, for example, we hypothesize that the aggregate effect parameters, A_1, A_2, A_3, and A_4 are all zero, then $V(A) = \sum_h A_h^2/a - 1$ is zero under this hypothesis. Conversely, if $V(A) = 0$, then the four parameters do not vary and must each be zero since, as was stated in section 4.2, $\sum_h A_h = 0$. And so the hypothesis that $A_1 = A_2 = A_3 = A_4 = 0$ is equivalent to the hypothesis that $V(A) = 0$. If this hypothesis is rejected by a significance test, the conclusion will be that at least one aggregate parameter mean is different from the others. On the other hand, an acceptance of this hypothesis will amount to the inference that there are no parameter differences among the four aggregates. We shall now discuss the appropriate statistical rules for testing the seven hypotheses of zero controlled factor parameter effects.

It must first be decided what α risk is allowable for the event that any of the tested hypotheses will be rejected even though it is true. We shall let $\alpha = .05$ for each test, which is to say that we are willing to claim significant effects on the average of once in twenty tests even though the tested effects are truly zero.

For each hypothesis, the statistical rule, or test, consists in forming the ratio of two mean squares from Table 8. The numerator of this ratio is the mean square found in the line whose effect corresponds to the hypothesis being tested, and the denominator is the mean square whose expected value is that of the numerator if the hypothesis is true. For example, the expected mean square for aggregates is $\sigma_S^2 + ts\,\sigma_M^2 + cmts\,V(A)$ and that for mixes is $\sigma_S^2 + ts\,\sigma_M^2$. If $V(A)$ is zero then both mean squares have expectations $\sigma_S^2 + ts\,\sigma_M^2$, and so the required test ratio is that of the

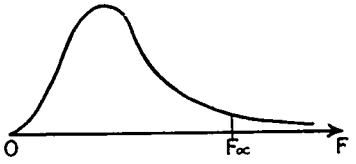
aggregates mean square to the mixes mean square. The mix mean square is said to be the error mean square for testing aggregate effects, or it is conventional to say that aggregate effects are tested against mix effects. The ratio of these two mean squares is called an "F" ratio because it follows the Fisher variance ratio probability law if the tested hypothesis is true.

Thus, at 40 cycles, we have the test ratio $F = (584.8/109.3) = 5.4$ for testing the hypothesis that $V(A) = 0$.

Following the same procedure, it can be seen from the last column of Table 8 that cement effects, C, and aggregate with cement interaction effects AC, are also tested against the mix mean square, M in AC. Thus, all formulation factor effects are tested against the mix-to-mix variation. In order to test the temperature effects, we note that the expected mean square for temperature is $\sigma_S^2 + s\,\sigma_{MT}^2 + acms\ V(T)$. Under the hypothesis that there are no differential effects due to temperature, $V(T) = 0$, and this expectation becomes $\sigma_S^2 + s\,\sigma_{MT}^2$ which is the expected mean square for the MT effects. Hence the F ratio for testing temperature effects is $F = \frac{MS(T)}{MS(MT)} = \frac{1578.3}{25.4} = 62.2$. The interactions of temperature with aggregates and with cements are also tested against the $M \times T$ mean square. The resulting F ratios for all of these tests are shown in the F ratio column of Table 8.

The next step is to decide which F ratios should lead to a rejection of any tested hypothesis. It is intuitive that if the F ratios are in the neighborhood of unity then the hypothesis is tenable and that a rejection should occur if the F ratios are "too large". The dividing line for this decision is in terms of the predetermined α risk. Figure A is that for a typical F probability distribution. Values of F may range from zero to infinity, and the total area under the curve is taken to be unity. The horizontal scale corresponds to all values that F might have *if the tested hypothesis* is true, and any segment of this scale cuts out an area under the curve which corresponds to the probability that F will fall in this segment.

Now we are committed to a rejection of the tested hypothesis with probability $\alpha = .05$ even though the hypothesis is true. This means that we must designate some segment of the F



Probability Distribution of the F Ratio

Figure A

scale, which cuts out .05 of the total area, as a region for rejection of the hypothesis. There are many such segments that could be chosen, but as has been said, we shall prefer to reject the hypothesis of zero effects whenever F is excessively large. This means that the F values which lead to rejection should be those under the right hand tail of the F curve. The critical point, F_α, is the value of F which will be exceeded by chance only $100\alpha\%$ of the time when the hypothesis is true. If the observed F falls in the rejection region, then the inference will be that the hypothesis is not true, and that there really are some non-zero parameter effects.

The F_α values for $\alpha = .05$ and $\alpha = .01$ are given in Tables 9 and 10 respectively. The shape of the F curve, and consequently the value for F_α depends upon the degrees of freedom associated with the numerator and denominator mean squares in the observed F ratio. For the hypotheses that $V(A) = 0$ the F ratio was 5.4. The numerator mean square had 3 degrees of freedom for aggregates $(a - 1)$ and the denominator mean square had 8 degrees of freedom for mixes $ac(m - 1)$. With $\alpha = .05$, Table 9 shows that under these conditions, $F_{.05} = 4.07$. Since the observed F exceeds $F_{.05}$, we infer that the hypothesis is not true, or we may simply say that the aggregate effects are significant at the 5% level. Table 8 shows the value of $F_{.05}$ for each significance test. At 40 cycles, for example, *all controlled factor* effects were significant at the 5% level except for the AC and ACT interactions. The same significances prevailed at each of the other cycles for which analyses were made, although it must be remembered that effects may be of different magnitudes and still be significant at the same level. In the significance column an asterisk appears for each mean square which was significantly larger than its appropriate error mean square,

TABLE 9
UPPER 5% POINTS ($F_{.05}$) OF THE F DISTRIBUTION

Degree of Freedom for Denominator Mean Square	Degrees of Freedom for Numerator Mean Square																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.59	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.36	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.81	1.75	1.76	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

* Adapted from tables computed by Merrington and Thompson and appearing in *Biometrika*, vol. 33, 1943 with the permission of the editor.

while the letters *NS* stand for non-significant effects.

For each of the four cycles analyzed, the aggregates means are shown in Fig. 2, cement means in Fig. 3, and temperature means in Fig. 4. Whenever a set of more than two means produces a significant F ratio in the analysis of variance, it does not necessarily follow that each mean in the set is significantly different from every other mean in the set. This point will be discussed in a later section.

And so by impartial and predetermined rules we have drawn statistical inferences about the controlled factor effects as they would occur in the generalized experiment. There is a known risk, .05, that a true hypothesis will be erroneously rejected, and a calculable risk that a false hypothesis will be erroneously accepted. This latter risk has been minimized by choosing an acceptance region for the F ratio wherein an observed F is not so likely to fall if the tested hypothesis is really false. From the sketch of the F distribution it can be seen that the risk of the second type of error will be increased if a smaller α level is used (e.g., $\alpha = .01$) for then the acceptance region will be larger.

Now that it has been decided which effects are significant, some interpretations are in order. Interpretations of significant main effects, i.e. aggregates, cements, and temperatures, must be withheld until the interaction effects are discussed. For example, the interaction of aggregates with cements was not found to be significant. Although a mathematical interpretation of two factor interactions was given in section 4, it is simpler to say that if two factors do not interact, then the difference in Y between any two levels of one factor is constant across all levels of the second factor. Since we have inferred the absence of aggregate-cement interaction, then we conclude that differences in durability caused by the two cements are the same for any aggregate, or that the difference in durability caused by two aggregates is the same for both cements. Graphically this means that parallel broken lines will result from a plot of

TABLE 10
UPPER 1% POINTS ($F_{.01}$) OF THE F DISTRIBUTION

Degrees of Freedom for Denominator Mean Square	Degrees of Freedom for Numerator Mean Square																		
	1	2	3	5	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,336
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	27.6	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	6.51	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.10	4.19	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.22	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.57	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

* Adapted from tables computed by Merrington and Thompson and appearing in *Biometrika*, Vol. 33, 1943 with the permission of the editor.

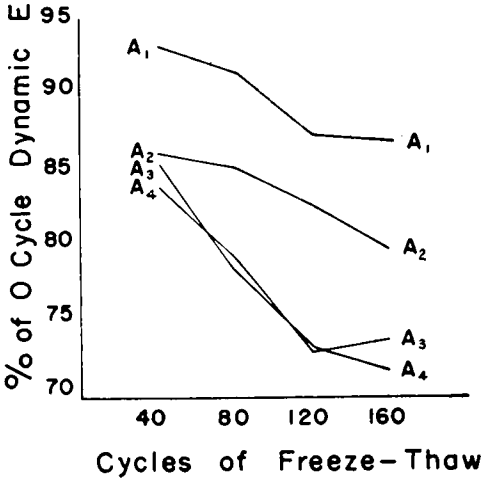


Figure 2

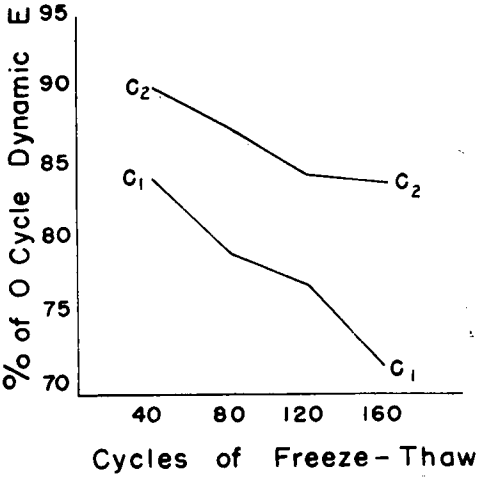


Figure 3

the AC means averaged across temperatures. This phenomenon is illustrated in Figure 5. The lines are not precisely parallel, but departures from parallelism are inferred to arise from error variation rather than from any real variation. It must be remembered that the AC effects have been averaged across temperatures, and so the question must be raised as to

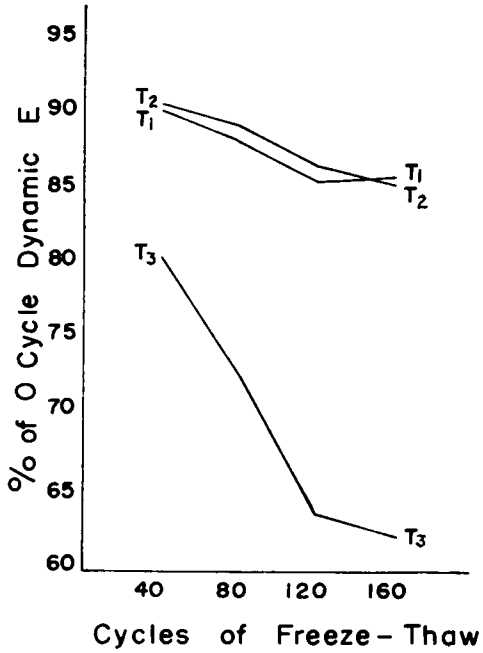


Figure 4

whether the absence of AC interaction prevails at each temperature level. Our inference is that it does, for the ACT interaction was not significant. If the ACT effects had been significant, the interpretation would be that the extent of AC interaction is not the same for each level of temperature.

The interaction of aggregates with temperature was significant as was the interaction of cements with temperatures. In the absence of ACT interaction we infer that the AT interaction is of the same magnitude for both cements, and that the CT interaction was about the same for all four aggregates. The AT and CT interactions result in the non-parallel graphs of Figures 6 and 7. Of course the scale used in constructing the graphs can be made to reduce or emphasize the non-parallelism, and it is also rather clear from the graphs as well as from the F ratios that the AT interactions are more pronounced than are the CT interactions. For these reasons it would seem to be all the more necessary to have predetermined rules for reaching decisions on the interaction effects.

The significance of main effects cannot be logically interpreted until hypotheses have been tested on the interactions of the main effects factors. Since the aggregates have been judged to interact with the temperatures, the fact that the aggregate main effects are significant does not carry much meaning since the aggregate effects have been averaged across temperatures. It may be that the four aggregate effects are significantly different at one temperature but not at some other temperature. It may also turn out, if there is an AT interaction, that the aggregates differ significantly at each temperature, but that the aggregate effects across temperatures are not significantly different. This situation would arise if the graphs of Figure 6 crossed each other in a suitable pattern. And so the experimenter may have only incidental interest in the significance achieved by the main effects if he has found significant interaction effects, for then he must study the effects of one factor at *each* level of the other factor. It is perhaps for this reason that the complete factorial design has great utility. If the procedure of "holding every factor at one level save for one factor which is allowed to vary" is followed, then the experimenter is not able to study interactions at all. He cannot place any important interpretation on the results for the one factor which was varied unless, by some fortuitous circumstance, his varied factor does not interact with any of the one level factors.

Thus far in this section we have only been concerned with significance tests on the controlled factor effects. We may also test hypotheses on the error variances, σ_M^2 and σ_{MT}^2. From the last column of Table 8, it is seen that, under the hypothesis that σ_M^2 is zero, the mix mean square has the same expectation as the specimens mean square. At 40 cycles, the appropriate F ratio for testing the hypothesis of no mix effects is $F = 109.3/21.2 = 5.1$. From Table 8 the degrees of freedom are 8 for the numerator and 96 for the denominator. From Table 9, with 8 degrees of freedom for the numerator and 96 degrees of freedom for the denominator, a value of 2.05 or larger, for F is cause for rejection of the hypothesis that $\sigma_M^2 = 0$. The conclusion is that the uncontrolled variables *do* bring in variation between mixes above and beyond the variability to be expected within a single mix. The implication is that the mixes *must* be replicated if an appropriate error term is to be available for testing the formulation factor effects. If the mix mean square had not been significant, a single mix would suffice for each aggregate cement combination, and the last line in the

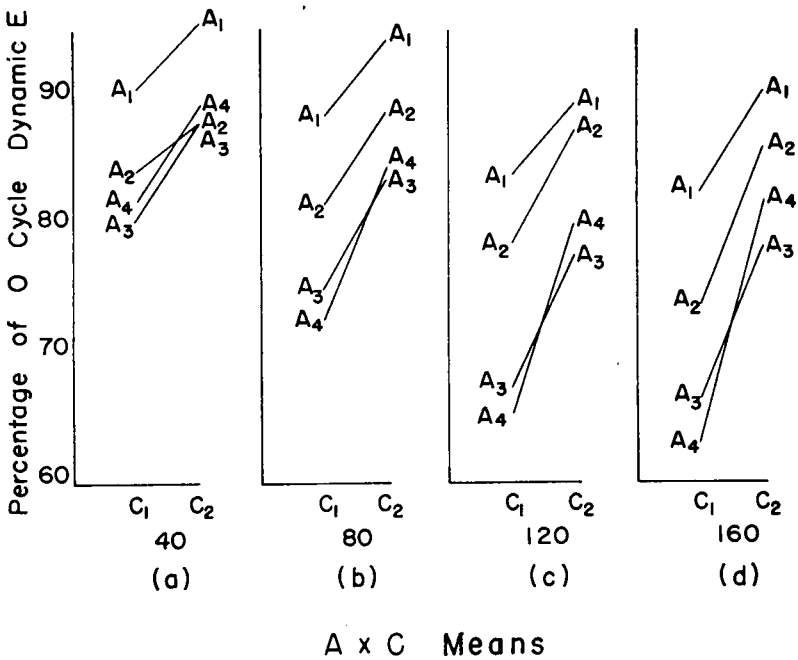


Figure 5. Graphical illustration of non-significance of aggregate × cement interaction effects.

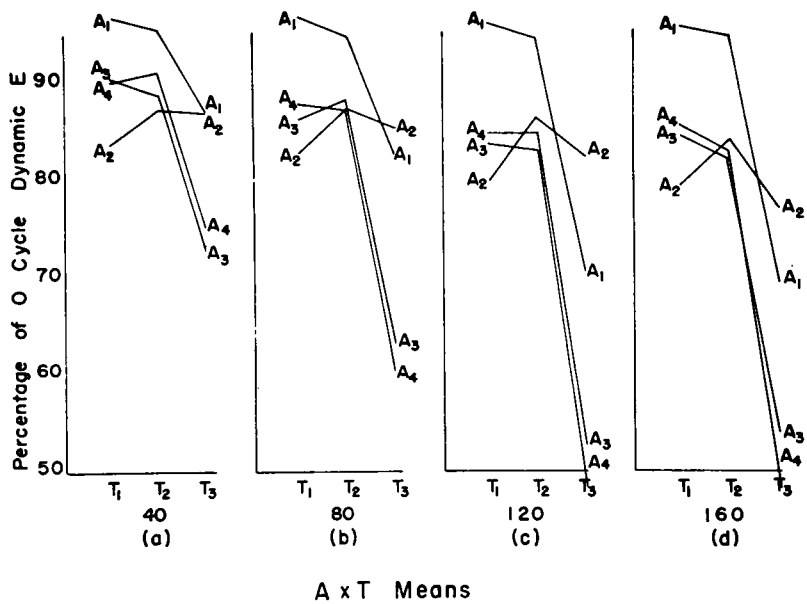


Figure 6. Graphical illustration of significant aggregate × temperature interaction effects.

analysis of variance table would provide the error term for testing the formulation factors.

To test the hypothesis that σ_{MT}^2 is zero, we must form the F ratio of the MT mean square to the specimens mean square. This ratio is not significant at 40 cycles, but is significant at 80, 120 and 160 cycles. From the expected mean squares column of Table 8 it follows that

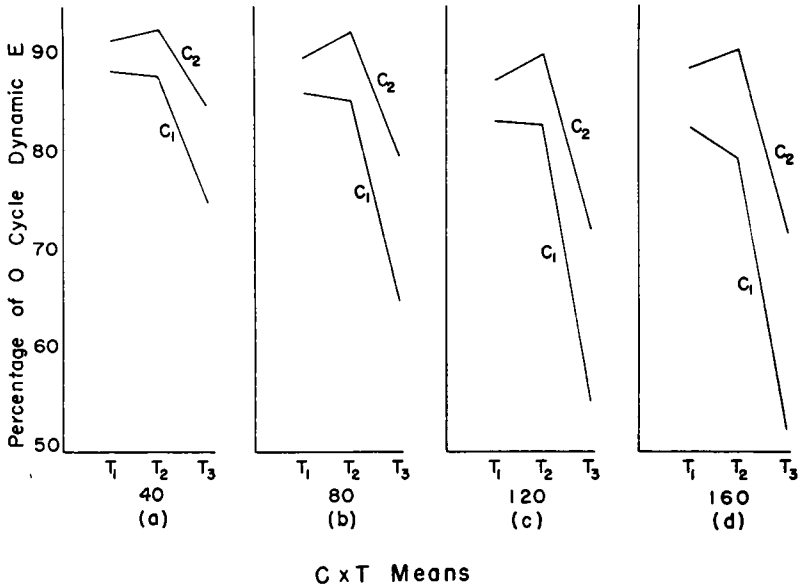


Fig. 7. Graphical illustration of significant cement × temperature interaction effects.

if σ_{MT}^2 is zero then all effects involving temperatures may be tested against the specimens mean square. One might follow this procedure at 40 cycles, but a more cautious approach would be to use the MT mean square as the error line for any of these tests at all cycles.

If no replication of mixes had been incorporated into the design, the expected mean squares would *still* be those given in Table 8. In such an event it would not be possible to test for the controlled factor effects without making some further assumptions. For example, the experimenter might know before running the experiment that he could assume $V(AC)$ to be zero, as turned out to be the inference from Table 8. Then with only one mix of each aggregate cement combination it is still possible to test for the aggregate and for the cement effects. For under the hypothesis that $V(A) = 0$ and with the assumption that $V(AC) = 0$, then the expectations of the aggregate mean square and the AC interaction mean square are both $\sigma_S^2 + st\,\sigma_M^2$, and even though σ_M^2 is unestimable, the F ratio is still appropriate for the test. It is exactly these kinds of assumptions that are necessary if fractional replications are to be used in order to reduce the scope of the experiment.

One of the major assumptions underlying the statistical tests is that the random effects have homogeneous variances.

Any set of independent observed mean squares may be tested for homogeneity by a test such as Bartlett's test (*8*). The mix to mix mean square is really the average of eight mix mean squares, one for each aggregate cement combination. The MT mean square also arises from eight separate mean squares, while the specimen mean square is the average of the 48 mean squares for specimens in the various $ACMT$ combinations. Before performing the analysis of variance, tests of homogeneity should be performed upon each of these sources of error variation. If there are significant departures from homogeneity an investigation into the cause is called for. If nothing is done to correct for existing heterogeneity of variance, the significance levels of the F tests will be somewhat different than their purported levels, depending upon the degree of heterogeneity. For example, an F ratio which has apparent significance at the 5% level may actually be significant at only the 10% level. The magnitude of this disturbance has been the subject of mathematical investigation (*9*).

Heterogeneous variances sometimes arise because of transcribing errors, computational errors, or changes in the measuring device. If the magnitude of an error mean square is a function of the mean value of the measured variable, heterogeneity of variance can be ex-

pected whenever the measured variable changes markedly in value. Measures of durability of concrete appear to have this characteristic, and the data of the illustrative experiment do not have a desirable degree of homogeneity. It would seem that those aggregates which have intermediate durability values are likely to lead to larger error mean squares than those which have durabilities which are either relatively high or low. It may be that a different scale of measurement for durability would lead to more homogeneity. An alternative measure is to analyze the experimental data in segments for which the error variation is homogeneous.

The *sensitivity* of the significance tests described in this section can be described in terms of the relative ease or difficulty with which differential effects which are non zero in the generalized experiment will be detected as such by the test. Since F ratios are used as the test criteria, it is apparent that for a fixed parameter variance, say $V(A)$, the observed F ratio will be larger if the denominator mean square, for mixes in this case, is relatively small. From an inspection of the F values in Tables 9 and 10 it can be seen that any observed F ratio will achieve significance for large denominator degrees of freedom whereas it may not for smaller denominator degrees of freedom. But these degrees of freedom are essentially proportional to the number of replications that were in the experimental design. And so, as is intuitively clear, the significance tests will be made more sensitive to the existence of parameter differences in the controlled factors either as the extent of uncontrolled variation is reduced, or as the number of replications is increased.

After making the analysis of variance and significance tests for a particular experiment, the estimates of error variation serve as excellent guides for the design of future experiments which will be subject to the same uncontrolled variables. The analysis may suggest that a change in the mathematical model is in order, or that more (or fewer) replications are necessary for the experiment to have a required degree of sensitivity to non-zero parameter effects. Each statistical analysis, therefore, should serve to provide conclusions relative to the experiment that has been performed, but perhaps just as important, should serve to give very specific information on how to design subsequent experiments. Although this principle is generally observed whether or not a statistical analysis is made, the systematic statistical approach can eliminate much of the guess work as the experimenter searches for better and more economical experimental designs.

Just as any design engineer turns to a new model as soon as his last one has been put into production, the experimenter who uses statistical principles of design and analysis should strive to find a better mathematical model from experiment to experiment as he studies the physical phenomena in his field. Even a small change in the mathematical model, or perhaps in the assumptions underlying the model, will ordinarily result in a different analysis of variance and in modified significance tests. For these reasons, we have made no assertion that the model used in the illustrative experiment is an optimum one. As was stated in the introduction, our main motivation has been to discuss and explain just one model, in some great detail.

6.3 *Contrasts of Means*

If none of the F ratios from the analysis of variance had been significant, the conclusion would have to be either that all of the effects parameters are zero, or that they are too small to assert themselves over and above the variation arising from the uncontrolled variables. But whenever any effect is shown to be significant by an F test, then the parameters which represent such an effect are concluded to be different from one another in some way. In this case the experimenter will ordinarily try to determine which parameters are different and by how much. We shall do this by establishing 95% confidence intervals for any comparison of the parameters, using a method which was presented by D. B. Duncan (*10*).[1]

For any set of k parameter means, $\mu_1, \mu_2, \cdots \mu_k$, let $\bar{Y}_1, \bar{Y}_2, \cdots, \bar{Y}_k$ be the corresponding observed means. Let n be the number of data observations that enter into the determination of each mean, $\bar{Y}$. Now a *contrast* of parameter means is any linear combination of

[1] Statistical methods for contrasting means after an analysis of variance are in an evolutionary state. The procedures discussed here have been adapted from a paper, "Simultaneous Confidence Intervals Derived from Multiple Range and Multiple F Tests," by D. B. Duncan and R. G. Bonner. This paper was presented to the American Statistical Association at Montreal, Canada in September, 1954, and is somewhat different from Reference 10.

these means $D = \sum_{i=1}^{k} \lambda_i \mu_i$, where the coefficients, λ_i, are any numbers such that $\sum_{i=1}^{k} \lambda_i = 0$. Since $\bar{Y}_i$ is the observed estimate of the parameter, μ_i, the estimate of the parameter contrast is the observed contrast, $\hat{D} = \sum_{i=1}^{k} \lambda_i \bar{Y}_i$.

For example, we inferred from Table 8 that the differential effects of aggregates, A_1, A_2, A_3 and A_4 were not all zero since the F ratio for aggregates was significant. If the differential effects are not all zero, then the parameter means for the aggregates must be somewhat different from one another. The observed aggregates means are found, in Table 6, to be 93.1, 86.1, 84.0 and 85.5 at 40 cycles, each based on $n = 36$ specimens. Suppose we wish to compare the first mean with the average of the second and fourth means. This contrast may be written, $D = \mu_1 - (\mu_2 + \mu_4)/2$, so that $k = 4$, $\lambda_1 = 1$, $\lambda_2 = -\frac{1}{2}$, $\lambda_3 = 0$, $\lambda_4 = -\frac{1}{2}$, and $\sum_{i=1}^{4} \lambda_i = 0$. The estimate of this contrast is $\hat{D} = 93.1 - (86.1 + 85.5)/2 = 7.3$. Although the true contrast was estimated to be 7.3 in this experiment, $\hat{D}$ will vary from one experiment to another because of uncontrolled variation. The variance of all $\hat{D}$ values in the generalized experiment will be called $\sigma_{\hat{D}}^2$ and $\sigma_{\hat{D}}$ is said to be the *standard error* of the estimated contrast. The mathematical formula for $\sigma_{\hat{D}}^2$ is derived in much the same way as were the expected mean squares of Table 8, and turns out to be $\sigma_{\hat{D}}^2 = (\sigma^2/n) \sum_{i=1}^{k} \lambda_i^2$, where σ^2 is estimated from one or more of the error mean squares in the analysis of variance table. The estimate of σ^2 will be called $\hat{\sigma}^2$, and the nature of $\hat{\sigma}^2$ depends upon which observed means enter into the contrast, $\hat{D}$. Substituting $\hat{\sigma}^2$ for σ^2, the estimated standard error of the contrast becomes

$$\hat{\sigma}_{\hat{D}} = \sqrt{\frac{\hat{\sigma}^2}{n} \sum_i \lambda_i^2}$$

and this standard error will have degrees of freedom *df*, depending upon which mean squares enter into $\hat{\sigma}^2$. In general, $\hat{\sigma}^2$ will be the mean square in the analysis of variance which was used to test for the significance of the differences among the means which enter into the contrast, $\hat{D}$, but there are exceptions to this rule.

For the above example, $\hat{\sigma}^2$ is the mix mean square, from Table 8, which was used to test for the significance of the aggregates effects. This mean square was 109.3 at 40 cycles, based on 8 degrees of freedom, and so

$$\hat{\sigma}_{\hat{D}}^2 = 109.3/36\ [(1)^2 + (-\tfrac{1}{2})^2 + (0)^2 + (-\tfrac{1}{2})^2] = 4.55,$$

and $\hat{\sigma}_{\hat{D}} = \sqrt{4.55} = 2.14$. So far, we have determined that the contrast in question has the estimated value of 7.3 with an estimated standard error of 2.14.

We now wish to give a confidence interval estimate for the true contrast, D. The statistical procedure is to claim with $100(1 - \alpha)\%$ confidence that D lies between the two numbers

$$\hat{D} \pm \frac{r_{F(\alpha)}}{\sqrt{2}} \hat{\sigma}_{\hat{D}} = \hat{D} \pm \frac{r_{F(\alpha)}}{\sqrt{2}} \sqrt{\frac{\hat{\sigma}^2}{n} \sum_i \lambda_i^2},$$

where $r_{F(\alpha)}$ is a probability factor to be found in Table 11 for $\alpha = .05$, or in Table 12 for $\alpha = .01$. $r_{F(\alpha)}$ is found in these tables at the intersection of k, the number of means from which the contrast is made, and *df*, the degrees of freedom for $\hat{\sigma}^2$. In our example $k = 4$ and $df = 8$, and if we let $\alpha = .05$, Table 11 gives $r_{F(.05)} = 3.800$. Hence the required confidence interval has the limits $7.3 \pm \frac{3.800}{\sqrt{2}}(2.14) = 7.3 \pm 5.7$, and we may claim with 95% confidence that the true contrast has a value somewhere between 1.6 and 13.0. Since the confidence interval does not contain zero, we may say that the observed contrast is significantly different from zero at the 5% level, or simply say that the contrast is significant.

It should be observed from the formula for the confidence interval that there are three ways for obtaining shorter confidence intervals. One is to design the experiment so that $\hat{\sigma}^2$, the mix mean square, will be smaller, perhaps by controlling additional factors. A second way is to obtain more than $n = 36$ observations on each aggregate, or to design a larger experiment. A third way is to use a smaller value for $r_{F(\alpha)}$, which amounts to requiring less confidence in the interval estimate since $r_{F(\alpha)}$ will be smaller if a larger α risk is assumed. Now the $r_{F(\alpha)}$ values in Tables 11 and 12 are large enough "safety" factors to insure $100(1 - \alpha)\%$ confidence for *any* of the many possible contrasts that can be made from the k means through different choices of the

TABLE 11*
SIGNIFICANT RANGES FOR A 5% LEVEL MULTIPLE COMPARISONS TEST†, $r_{F(.05)}$

df	k																
	2	3	4	5	6	7	8	9	10	11	13	15	17	19	21	51	101
3	4.500	4.711	4.889	5.032	5.146	5.24	5.33	5.40	5.47	5.53	5.64	5.73	5.80	5.87	5.93	6.43	6.72
4	3.926	4.198	4.412	4.590	4.742	4.87	4.99	5.10	5.19	5.28	5.43	5.56	5.68	5.78	5.87	6.67	7.19
5	3.636	3.921	4.153	4.348	4.517	4.67	4.80	4.92	5.03	5.13	5.31	5.47	5.61	5.74	5.85	6.85	7.56
6	3.461	3.751	3.991	4.195	4.374	4.53	4.68	4.81	4.93	5.04	5.24	5.41	5.57	5.71	5.84	6.99	7.85
7	3.345	3.637	3.880	4.090	4.274	4.44	4.59	4.73	4.85	4.97	5.18	5.37	5.54	5.69	5.83	7.11	8.09
8	3.261	3.554	3.800	4.013	4.201	4.37	4.52	4.67	4.80	4.92	5.14	5.34	5.51	5.68	5.82	7.21	8.30
9	3.199	3.492	3.739	3.955	4.145	4.32	4.48	4.62	4.76	4.88	5.11	5.31	5.50	5.67	5.82	7.30	8.48
10	3.151	3.444	3.691	3.908	4.101	4.28	4.44	4.58	4.72	4.85	5.08	5.30	5.48	5.66	5.82	7.37	8.63
11	3.113	3.405	3.653	3.871	4.065	4.24	4.40	4.55	4.69	4.82	5.06	5.28	5.48	5.65	5.82	7.43	8.77
12	3.082	3.373	3.621	3.840	4.036	4.21	4.38	4.53	4.67	4.80	5.05	5.27	5.47	5.65	5.82	7.49	8.89
13	3.055	3.346	3.595	3.814	4.011	4.19	4.36	4.51	4.65	4.79	5.03	5.26	5.46	5.65	5.82	7.54	9.00
14	3.033	3.324	3.572	3.792	3.989	4.17	4.34	4.49	4.64	4.77	5.02	5.25	5.45	5.64	5.82	7.58	9.10
15	3.014	3.304	3.553	3.773	3.971	4.15	4.32	4.48	4.62	4.76	5.01	5.24	5.45	5.64	5.82	7.62	9.20
16	2.998	3.288	3.536	3.757	3.955	4.14	4.30	4.46	4.61	4.75	5.00	5.23	5.44	5.64	5.82	7.66	9.28
17	2.984	3.273	3.522	3.742	3.941	4.12	4.29	4.45	4.60	5.74	4.99	5.23	5.44	5.64	5.82	7.70	9.36
18	2.971	3.260	3.508	3.729	3.929	4.11	4.28	4.44	4.59	4.73	4.99	5.22	5.44	5.64	5.82	7.73	9.43
19	2.960	3.248	3.497	3.718	3.918	4.10	4.27	4.43	4.58	4.72	4.98	5.22	5.44	5.64	5.82	7.76	9.50
20	2.950	3.238	3.487	3.707	3.908	4.09	4.26	4.42	4.57	4.71	4.97	5.21	5.43	5.64	5.83	7.78	9.56
22	2.933	3.220	3.469	3.690	3.890	4.08	4.25	4.41	4.56	4.70	4.96	5.21	5.43	5.63	5.83	7.83	9.68
24	2.919	3.206	3.454	3.675	3.876	4.06	4.23	4.39	4.55	4.69	4.96	5.20	5.42	5.63	5.83	7.87	9.78
26	2.908	3.193	3.441	3.663	3.864	4.05	4.22	4.38	4.54	4.68	4.95	5.19	5.42	5.63	5.83	7.91	9.87
28	2.896	3.183	3.431	3.652	3.854	4.04	4.21	4.38	4.53	4.67	4.94	5.19	5.42	5.63	5.83	7.94	9.95
30	2.888	3.174	3.422	3.643	3.845	4.03	4.20	4.37	4.52	4.67	4.94	5.19	5.42	5.63	5.83	7.97	10.02
40	2.858	3.143	3.390	3.611	3.814	4.00	4.18	4.34	4.50	4.64	4.92	5.17	5.41	5.63	5.84	8.08	10.30
60	2.828	3.112	3.358	3.580	3.783	3.97	4.15	4.31	4.47	4.62	4.90	5.16	5.40	5.63	5.84	8.21	10.64
100	2.804	3.086	3.333	3.555	3.758	3.95	4.12	4.29	4.45	4.60	4.89	5.15	5.40	5.63	5.85	8.33	10.97
∞	2.772	3.051	3.297	3.518	3.722	3.91	4.09	4.26	4.42	4.57	4.86	5.13	5.39	5.63	5.86	8.54	11.66

For $df = 1$ or $df = 2$ use $r_F = 17.969$ or 6.085 respectively for all values of k.
*See Table 12 for acknowledgment.
†Using special protection levels based on degrees of freedom.

coefficients λ_i. It may be that the experimenter only wishes to compare the observed means in pairs, or two at a time. In such a case, the true contrast is of the form $D = \mu_i - \mu_j$, so that $\lambda_i = 1$, $\lambda_j = -1$, and the other $k - 2$ λ's are all zero. The estimate of D is then $\hat{D} = \bar{Y}_i - \bar{Y}_j$, or simply the difference between the two observed means.

Since there are relatively few possible contrasts of this sort, the $r_{F(\alpha)}$ values of Tables 11 and 12 give more protection against erroneous confidence intervals than is necessary. The $r(\alpha)$ values in Tables 13, for $\alpha = .05$, and 14, for $\alpha = .01$, have been developed for confidence intervals for contrasts of just two means at a time. In this case $\sum_i \lambda_i^2 = (1)^2 + (-1)^2 = 2$, and the formula for the desired confidence intervals becomes

$$\hat{D} \pm \frac{r(\alpha)}{\sqrt{2}} \hat{\sigma}_{\hat{D}} = (\bar{Y}_i - \bar{Y}_j) \pm \frac{r(\alpha)}{\sqrt{2}} \sqrt{\frac{\hat{\sigma}^2}{n}(2)}$$

$$= (\bar{Y}_i - \bar{Y}_j) \pm r(\alpha) \sqrt{\frac{\hat{\sigma}^2}{n}}.$$

In what follows, all contrasts will be for just two means at a time.

For the four aggregates means, there are six possible contrasts taking two means at a time. As in the above example, $\hat{\sigma}^2$ is the mix mean square of 109.3, at 40 cycles, based on 8 degrees of freedom. For $\alpha = .05$, Table 13 gives $r(.05) = 3.47$ for $k = 4$ and $df = 8$. Hence the 95% confidence interval for the difference between any two of the aggregates parameter means is given by

$$\hat{D} \pm 3.47 \sqrt{\frac{109.3}{36}} = \hat{D} \pm 6.0,$$

and the only significant contrasts are those involving two means whose difference exceeds 6.0. Only the differences between the first aggregate and each of the other three are significant, and the confidence intervals for the significant contrasts are roughly between 1% and 15% of original dynamic E. The confidence intervals for contrasts among the second, third, and fourth aggregates overlap

TABLE 12*
SIGNIFICANT RANGES FOR A 1% LEVEL MULTIPLE COMPARISONS TEST†, $r_{F(.01)}$

df	k																
	2	3	4	5	6	7	8	9	10	11	13	15	17	19	21	51	101
3	8.260	8.701	9.044	9.325	9.564	9.77	9.96	10.13	10.27	10.41	10.65	10.86	11.04	11.21	11.36	12.70	13.69
4	6.511	6.979	7.359	7.679	7.957	8.20	8.42	8.62	8.81	8.98	9.29	9.56	9.80	10.02	10.22	12.07	13.55
5	5.702	6.157	6.535	6.858	7.144	7.40	7.63	7.84	8.04	8.22	8.55	8.85	9.11	9.35	9.58	11.71	13.49
6	5.242	5.682	6.050	6.370	6.655	6.91	7.15	7.36	7.56	7.75	8.09	8.40	8.67	8.93	9.16	11.47	13.47
7	4.948	5.373	5.733	6.048	6.329	6.58	6.82	7.04	7.24	7.43	7.77	8.09	8.37	8.63	8.88	11.30	13.45
8	4.745	5.158	5.509	5.819	6.097	6.35	6.58	6.80	7.00	7.19	7.54	7.86	8.15	8.42	8.66	1.18	13.45
9	4.596	4.998	5.344	5.649	5.923	6.17	6.41	6.62	6.82	7.02	7.36	7.68	7.98	8.25	8.50	11.08	13.45
10	4.482	4.876	5.216	5.517	5.788	6.04	6.27	6.48	6.68	6.87	7.23	7.55	7.84	8.11	8.37	11.00	13.45
11	4.393	4.780	5.114	5.411	5.680	5.93	6.16	6.37	6.57	6.76	7.11	7.43	7.73	8.00	8.26	10.93	13.46
12	4.320	4.702	5.032	5.326	5.592	5.84	6.06	6.28	6.48	6.67	7.02	7.34	7.64	7.91	8.17	10.88	13.46
13	4.260	4.637	4.963	5.254	5.519	5.76	5.99	6.20	6.40	6.59	6.94	7.26	7.56	7.84	8.09	10.83	13.47
14	4.210	4.582	4.905	5.194	5.457	5.70	5.92	6.14	6.33	6.52	6.87	7.19	7.49	7.77	8.03	10.79	13.48
15	4.168	4.536	4.856	5.143	5.404	5.64	5.87	6.08	6.28	6.46	6.82	7.14	7.43	7.71	7.97	10.75	13.48
16	4.131	4.496	4.814	5.098	5.358	5.60	5.82	6.03	6.23	6.42	6.76	7.08	7.38	7.66	7.92	10.72	13.49
17	4.098	4.461	4.777	5.060	5.317	5.56	5.78	5.99	6.18	6.37	6.72	7.04	7.34	7.62	7.88	0.70	13.49
18	4.070	4.431	4.744	5.025	5.282	5.52	5.74	5.95	6.14	6.33	6.68	7.00	7.30	7.58	7.84	10.67	13.50
19	4.046	4.404	4.716	4.995	5.250	5.49	5.71	5.92	6.11	6.30	6.64	6.96	7.26	7.54	7.80	10.65	13.50
20	4.023	4.379	4.689	4.968	5.222	4.46	5.68	5.88	6.08	6.27	6.61	6.93	7.23	7.51	7.77	10.63	13.51
22	3.987	4.338	4.646	4.922	5.174	5.41	5.63	5.83	6.03	6.21	6.56	6.88	7.18	7.46	7.72	10.59	13.52
24	3.956	4.304	4.609	4.883	5.134	5.37	5.58	5.79	5.98	6.17	6.51	6.83	7.13	7.41	7.67	10.56	13.53
26	3.930	4.276	4.579	4.851	5.100	5.33	5.55	5.75	5.94	6.13	6.47	6.79	7.09	7.37	7.63	10.54	13.54
28	3.907	4.252	4.553	4.824	5.072	5.30	5.52	5.72	5.91	6.10	6.44	6.76	7.06	7.34	7.60	10.52	13.55
30	3.889	4.231	4.531	4.800	5.048	5.28	5.49	5.69	5.89	6.07	6.41	6.73	7.03	7.31	7.57	10.50	13.55
40	3.824	4.160	4.454	4.719	4.962	5.19	5.40	5.60	5.79	5.97	6.31	6.63	6.92	7.20	7.47	10.43	13.58
60	3.762	4.091	4.379	4.640	4.879	5.10	5.31	5.51	5.70	5.88	6.21	6.53	6.82	7.10	7.36	10.36	13.62
100	3.712	4.036	4.320	4.577	4.813	5.03	5.24	5.44	5.62	5.80	6.13	6.44	6.74	7.01	7.28	10.30	13.66
∞	3.643	3.958	4.235	4.486	4.716	4.93	5.13	5.32	5.51	5.68	6.01	6.32	6.61	6.88	7.14	10.20	13.75

For $df = 1$ or $df = 2$ use $r_F = 90.024$ or 14.036 respectively for all values of k.
* Permission granted by the Editor, *Virginia Journal of Science* to reproduce Tables 11 and 12 from paper by D. B. Duncan, "A Significance Test for Differences Between Rank Treatments in an Analysis of Variance." 2, 171–189, 1951.
† Using special protection levels based on degrees of freedom.

zero and cannot be called significant at the 5% level.

Since the method of confidence intervals for contrasts is relatively simple, and since it picks out the significant differences in observed means and estimates the magnitude of the corresponding parameter differences, it might be wondered why one should bother with an analysis of variance at all. Contrasts can be made without performing the entire analysis of variance, but the analysis of variance systematically produces the mean squares which are necessary for the estimates, $\hat{\sigma}^2$. Furthermore, the F tests from the analysis of variance point out which effects need to be investigated for significant contrasts, and which effects are not significant to begin with and therefore warrant no further investigation.

There were just two cement means, averaged across temperatures, and the difference between them has already been declared significant by the F test shown in Table 8. The magnitude of this difference can now be estimated to be, with 95% confidence,

$$(90.1 - 84.2) \pm 3.26 \sqrt{\frac{109.3}{72}} = 5.9 \pm 4.0,$$

or somewhere between 1.9 and 9.9. The value 3.26 is found in Table 13 with $k = 2$ and $df = 8$, the degrees of freedom for the mix mean square, $\hat{\sigma}^2 = 109.3$. Each cement mean was based on $n = 72$ specimens.

In making contrasts of the overall temperature means, at 40 cycles, $k = 3$, and the appropriate mean square for $\hat{\sigma}^2$ is the MT mean square of Table 8, and so $\hat{\sigma}^2 = 25.4$ based on 16 degrees of freedom. Each temperature mean was based on $n = 48$ specimens, and $r(.05) = 3.15$ from Table 13. 95% confidence intervals for differences between two temperatures means then become

$$\hat{D} \pm 3.15 \sqrt{\frac{25.4}{48}} = \hat{D} \pm 2.3.$$

In Table 6, the three means in question are found to be 90.3, 90.7 and 80.6, and the significant contrasts are those which exceed 2.3,

TABLE 13*
SIGNIFICANT RANGES FOR A 5% LEVEL NEW† MULTIPLE RANGE TEST, $r_{.05}$

df	k															
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.58	3.58	3.58	3.58	3.58	3.58	3.58	3.58	3.58
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.44	3.46	3.46	3.46	3.46	3.47	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.45	3.46	3.46	3.47	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.45	3.46	3.46	3.47	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.44	3.45	3.46	3.47	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.44	3.45	3.46	3.47	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.43	3.45	3.46	3.47	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.43	3.44	3.46	3.47	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.43	3.44	3.46	3.46	3.47	3.47	3.47
22	2.93	3.08	3.17	3.24	3.29	3.32	3.35	3.37	3.39	3.42	3.44	3.45	3.46	3.47	3.47	3.47
24	2.92	3.07	3.15	3.22	3.28	3.31	3.34	3.37	3.38	3.41	3.44	3.45	3.46	3.47	3.47	3.47
26	2.91	3.06	3.14	3.21	3.27	3.30	3.34	3.36	3.38	3.41	3.43	3.45	3.46	3.47	3.47	3.47
28	2.90	3.04	3.13	3.20	3.26	3.30	3.33	3.35	3.37	3.40	3.43	3.45	3.46	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.40	3.43	3.44	3.46	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.39	3.42	3.44	3.46	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.37	3.40	3.43	3.45	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.36	3.40	3.42	3.45	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.34	3.38	3.41	3.44	3.47	3.61	3.67

* See Table 14 for acknowledgment.
† Using special protection levels based on degrees of freedom.

or between either of the first two temperatures and the third. The confidence intervals for the true differences extend from 7.4% of original dynamic E to 12.4%.

It was observed in Section 6.2 that a study of differences between overall means is not very informative if the factor in question is known to interact with other factors in the experiment. Since the F ratios gave significant interactions of temperatures with both cements and aggregates, the most interesting contrasts should be among the aggregates means at each temperature, or vice versa, and between the cement means at each temperature, or vice versa.

There are twelve aggregate-temperature combination means at each cycle, and we shall let $k = 12$ for contrasts which involve these means. This is being somewhat conservative for we shall not want to make all 66 possible contrasts of these means taken two at a time, and so $r(\alpha)$ for $k = 12$ is larger than is actually needed for the 27 possible contrasts of aggregate means for each temperature and temperature means for each aggregate. The appropriate value for $\hat{\sigma}^2$ in these contrasts depends upon whether the contrasts are between aggregate means for a given temperature or between temperature means for a given aggregate. A discussion of this phenomenon may be found in reference (11).

For contrasts of temperature means at each aggregate level, $\hat{\sigma}^2$ is the MT mean square of Table 8, $\hat{\sigma}^2 = 25.4$, on 16 degrees of freedom. Each of these means is based on $n = 12$ specimens, and so 95% confidence intervals for these contrasts are given by

$$\hat{D} \pm 3.44 \sqrt{\frac{25.4}{12}} = \hat{D} \pm 5.0,$$

where $r(.05) = 3.44$ from Table 13 with $k = 12$, $df = 16$.

In order to determine whether the aggregate means are significantly different at a single temperature, an approximation must be made for $\hat{\sigma}^2$. A mathematical analysis similar to that found in reference (11) will show that the vari-

TABLE 14*

SIGNIFICANT RANGES FOR A 1% LEVEL NEW† MULTIPLE RANGE TEST, $r_{.01}$

df	k															
	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.0	9.1	9.2	9.3	9.3	9.3	9.3
4	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.3	7.4	7.4	7.5	7.5	7.5	7.5
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.6	6.6	6.7	6.7	6.8	6.8	6.8
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.1	6.2	6.2	6.3	6.3	6.3	6.3
7	4.95	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	5.8	5.9	5.9	6.0	6.0	6.0	6.0
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.6	5.7	5.7	5.8	5.8	5.8	5.8
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.5	5.5	5.6	5.7	5.7	5.7	5.7
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.36	5.42	5.48	5.54	5.55	5.55	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.24	5.28	5.34	5.38	5.39	5.39	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.13	5.17	5.22	5.24	5.26	5.26	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.04	5.08	5.13	5.14	5.15	5.15	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	4.96	5.00	5.04	5.06	5.07	5.07	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	4.90	4.94	4.97	4.99	5.00	5.00	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.84	4.88	4.91	4.93	4.94	4.94	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.72	4.75	4.80	4.83	4.86	4.88	4.89	4.89	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.76	4.79	4.82	4.84	4.85	4.85	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.72	4.76	4.79	4.81	4.82	4.32	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.69	4.73	4.76	4.78	4.79	4.79	4.79
22	3.99	4.17	4.28	4.36	4.42	4.48	4.53	4.57	4.60	4.65	4.68	4.71	4.74	4.75	4.75	4.75
24	3.96	4.14	4.24	4.33	4.39	4.44	4.49	4.53	4.57	4.62	4.64	4.67	4.70	4.72	4.74	4.74
26	3.93	4.11	4.21	4.30	4.36	4.41	4.46	4.50	4.53	4.58	4.62	4.65	4.67	4.69	4.73	4.73
28	3.91	4.08	4.18	4.28	4.34	4.39	4.43	4.47	4.51	4.56	4.60	4.62	4.65	4.67	4.72	4.72
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.54	4.58	4.61	4.63	4.65	4.71	4.71
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.46	4.51	4.54	4.57	4.59	4.69	4.69
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.39	4.44	4.47	4.50	4.53	4.66	4.66
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.35	4.38	4.42	4.45	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.26	4.31	4.34	4.38	4.41	4.60	4.68

* Permission granted by the Editor, *Biometrics*, to reproduce Table 13 and 14 from paper by D. B. Duncan, "Multiple Range and Multiple F Tests," March, 1955.
† Using special protection levels based on degrees of freedom.

ance of such contrasts must be estimated by using

$$\hat{\sigma}^2 = [MS(M) + t\ MS(MT) - MS(S)]/t$$

where $MS(M)$, $MS(MT)$, and $MS(S)$ are the three error mean squares of Table 8, $MS(M) = 109.3$, $MS(MT) = 25.4$, $MS(S) = 21.2$. Numerically, $\hat{\sigma}^2 = [(109.3) + 3(25.4) - (21.2)]/3 = 54.8$. Now when estimated variances involve *combinations* of mean squares from the analysis of variance table, the proper degrees of freedom must also be approximated. Instead of going into such approximations here, we shall use a conservative value for df, namely, the smallest of the degrees of freedom for any of the mean squares involved. In the case at hand, this was the mix mean square, based on 8 degrees of freedom. From Table 13, $r(.05) = 3.56$ for $k = 12$ and $df = 8$. So the (approximate) 95% confidence limits for contrasts of aggregates means at a given temperature are

$$\hat{D} \pm 3.56 \sqrt{\frac{54.8}{12}} = \hat{D} \pm 7.6.$$

Further discussion of the confidence intervals for differences among aggregate-temperature means will be given in Section 7.

There are $k = 6$ cement-temperature means, averaged across aggregates. Confidence intervals for contrasts of these means are determined in exactly the same way as they were for the contrasts of the aggregate-temperature means, with the exceptions that $k = 6$ and $n = 24$ specimens for each such mean.

To compare two temperature means in a given cement, the 95% confidence limits are given by

$$\hat{D} \pm 3.34 \sqrt{\frac{25.4}{24}} = \hat{D} \pm 3.4.$$

Approximate 95% confidence limits for the contrast of the two cement means at a given temperature become

$$\hat{D} \pm 3.55 \sqrt{\frac{54.8}{24}} = \hat{D} \pm 5.3.$$

In the next section these confidence limits will be used in a specific discussion of the cement-temperature interaction.

The analysis of variance did not show any significant ACT interaction. This result implies that, for example, the cement-temperature interaction effects exist in essentially the same degree for each of the aggregates, and that the aggregate-temperature interaction effects are practically the same for both cements. Thus there is no justification for examining the aggregate-cement-temperature means for contrasts of the type which have been made in the preceding paragraphs.

Although many other contrasts can be made among the various means in Table 6, we have restricted our discussion to those contrasts which can account for the significant F ratios found in the analysis of variance. It is possible to perform the analysis of variance in such a way that many of the interesting contrasts can be studied by F tests in the analysis of variance table. We have preferred to present the methods of this section because they can have quite widespread application in situations where the experimenter wishes to estimate the true contrast of any set of parameter means.

We should like to point out that contrasts which are not significant at some predetermined α level will be significant if a smaller confidence coefficient is used. The experimenter might choose to use whatever reasonable confidence coefficient becomes necessary in order to claim "significance" for those contrasts which did not quite achieve the required value.

Such a procedure is equivalent to making the data show whatever is desired and gives rise to the statement that "anything can be inferred through the use of statistical methods with a given set of data". Statistical inferences are based on rules which should be decided upon in the design stage of the experiment, and not be changed at the end of the analysis. On the other hand, a statistical analysis of experimental data will ordinarily suggest better mathematical models for the phenomena being studied, and should provide the experimenter with new hypotheses to be tested. In addition, the analysis can be expected to reveal the amount of data which will be necessary to give the next experiment a desirable degree of sensitivity to the effects of the controlled factors upon the experimental material.

TABLE 15
RESULTS OF ANALYSIS OF VARIANCE AT 40, 80, 120 AND 160 CYCLES

Source of Variation	Degrees of Freedom *df*	Cycles of Freezing and Thawing								$F_{.95}$	Significance			
		40		80		120		160			40	80	120	160
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*					
A	3	584.8	5.4	1287.8	9.6	1774.7	14.2	1587.5	11.5	4.07	*	*	*	*
C	1	1251.4	11.4	2517.6	18.8	3393.9	27.2	5893.2	42.5	5.32	*	*	*	*
AC	3	29.8	.3	61.2	.5	120.5	1.0	190.1	1.4	4.07	NS	NS	NS	NS
M in *AC*	8	109.3	5.1	134.0	3.7	124.8	2.9	138.6	2.6	2.05	*	*	*	*
T	2	1578.3	62.2	4008.2	52.2	8149.5	83.4	8910.0	65.0	3.63	*	*	*	*
AT	6	288.3	11.4	659.1	8.6	924.8	9.5	821.7	6.0	2.74	*	*	*	*
CT	2	155.8	6.1	408.8	5.3	623.2	6.4	601.8	4.4	3.63	*	*	*	*
ACT	6	14.9	.6	6.7	.1	57.5	.6	35.8	.3	2.74	NS	NS	NS	NS
(*M* in *AC*)*T*	16	25.4	1.2	76.8	2.1	97.9	2.3	137.1	2.6	1.75	NS	*	*	*
Spec	96, 92, 86, 84	21.2		36.4		43.0		52.4						

7. PRACTICAL INFERENCES FROM CONTRASTS OF MEANS

The inferences relating to the experiment must now be evaluated. In order to utilize all the information at hand in a proper manner the experimenter must know something of statistical concepts. Of prime importance to the experimenter is the column in the analysis of variance table showing the results of significance tests, a table displaying means (such as Table 6), and a brief discussion which points out those means that were determined to be significantly different from others in a set. To present such a discussion we must first show the results of the analysis of variance at 40, 80, 120 and 160 cycles and then show the results of contrasts of means for those effects which proved to be significant. The procedures have been discussed in previous sections for the data at 40 cycles, thus it is only necessary to show the results in this section.

Table 15 shows the results of the analyses of variance for the data at 40, 80, 120 and 160 cycles of freezing and thawing. The significance columns, using an α risk of .05, show that essentially the same effects are significant at each cycle time. If it can be shown that the same contrasts of means are significant at each cycle-time, it may then be concluded that essentially all the significant differences to be obtained from this experiment were detected in the 40 cycle data.

7.1 *Contrasts Among Means for Main Effects*

The main effects A, C and T are significant at each cycle-time as shown in Table 15. This fact implies that at least two of the levels of the main effects, A, C and T, are really different at each cycle-time. This situation is shown graphically in Figures 2, 3 and 4 where per cent of 0 cycle dynamic E is plotted as the ordinate and cycle-times are plotted as the abscissa. For example, it is undoubtedly true that aggregate A_1 is superior to aggregates A_2, A_3 and A_4 at 40 cycles. Also it is quite likely that at 160 cycles, A_1 is superior to A_2 which is in turn superior to A_3 and A_4. Similar discussions could be made concerning the cement (Fig. 3) and temperature (Fig. 4) means at each cycle-time. However, since the interaction effects of aggregate with temperature, AT, and cement with temperature, CT, are significant it is more important to study the results at each level of temperature as though three independent experiments had been conducted simultaneously.

7.2 *Contrasts of AT and CT Means, Temperature Fixed*

The results of contrasts of AT and CT means are shown in Table 16 where the contrasts have been made across aggregates, or across cements, at a fixed thawing temperature. Thawing the 3 x 4 x 16-in. beams at 35 deg. F. (T_1) resulted in less durability for beams containing aggregate A_2 than for those containing aggregate A_1. The durability of beams containing aggregates A_3 or A_4 is neither better than that for beams containing aggregate A_2, or worse than that for beams containing aggregate A_1.

At the second thawing temperature, 55 deg. F., the contrasts between aggregates A_1 and A_2 barely achieved significance at forty

TABLE 16*

RESULTS OF CONTRASTS OF AT AND CT MEANS TWO AT A TIME (TEMPERATURE FIXED)

Thaw-water Temperature, Deg. F.	No. of Cycles	Std. Error of AT Contr.	Diff. Req'd. to be Sign. ($r_{.05}$ = 3.56)	Per Cent of Zero Cycle Dynamic E for Specimens Made with Coarse Aggregate				Std. Error of CT Contr.	Diff. Req'd to be Sign. ($r_{.05}$ = 3.55)	Per Cent of 0 Cycle Dynamic E for Specimens Made with Cement	
				A_1	A_2	A_3	A_4			C_1	C_2
T_1 (35° F.)	40	2.14	7.6	96.6	83.9	90.5	90.1	1.51	5.4	88.8	91.8
	80	2.96	10.5	96.5	82.7	86.3	87.8	2.09	7.4	86.6	90.1
	120	3.23	11.5	95.7	79.9	83.6	84.6	2.28	8.1	83.7	87.8
	160	3.72	13.2	95.4	79.2	84.2	85.4	2.63	9.3	83.0	88.9
T_2 (55° F.)	40	2.14	7.6	95.4	87.4	88.7	91.1	1.51	5.4	88.3	93.1
	80	2.96	10.5	94.6	87.2	88.1	87.2	2.09	7.4	85.9	92.7
	120	3.23	11.5	94.1	86.3	82.9	84.6	2.28	8.1	83.3	90.4
	160	3.72	13.2	94.3	83.7	81.8	82.6	2.63	9.3	79.8	90.7
T_3 (100° F.)	40	2.14	7.6	87.2	86.9	72.8	75.2	1.51	5.4	75.6	85.5
	80	2.96	10.5	82.8	85.6	63.2	60.4	2.09	7.4	65.6	80.4
	120	3.23	11.5	70.6	82.3	52.6	49.2	2.28	8.1	55.2	72.6
	160	3.72	13.2	69.3	76.9	53.8	48.1	2.63	9.3	52.2	72.1

* The AT and CT means are taken from Table 6 and rearranged in order to show contrasting means.

Note: Significantly different means are indicated thus X↔X. Groups of means in which no differences occur are indicated thus X X X X.

cycles. With an α risk of .05 the difference had to be at least 7.6 per cent to be declared significant and it was only 0.4 per cent greater than that, or 8.0 per cent. This particular contrast was not significant at the later cycle times.

The greatest differences in durability occurred when the specimens were thawed at 100 deg. F. Specimens which contained aggregates A_1 and A_2 were more durable than those containing aggregates A_3 and A_4. Furthermore, this was the only freeze-thaw cycle in which differences between the cements were consistently declared to be real. Field observations of the durability of concrete made with these aggregates and cements correspond closely with the results obtained from this last freeze-thaw cycle. An obvious conclusion is that this is the most efficient of the three cycles for evaluating the relative quality of gravels similar to those tested for use in portland cement concrete.

7.3 *Contrasts of AT and CT Means Across Temperatures*

Table 17 shows the results of contrasts between temperature means for each aggregate and for each cement. These contrasts show that the results obtained when the thaw-water temperature was 100 deg. F. (T_3) were significantly different from those obtained with either 35 (T_1) or 55 (T_2) deg. F. thaw-water with the exception of aggregate A_2. The durability of specimens made with aggregate A_2 was essentially the same for each of the three freeze-thaw cycles. These results are shown graphically in Fig. 6.

Contrasting the temperature means for each aggregate and for each cement tells us nothing about which of the thaw-water temperatures best reflected the durability of concrete in service. However, such contrasts did point out the fact that aggregate A_2 was in some way different from aggregates A_1, A_3 and A_4. Thus it is important to contrast interaction

TABLE 17
RESULTS OF CONTRASTS OF *AT* AND *CT* MEANS TWO AT A TIME (ACROSS TEMPERATURES)

No. of Cycles	Thaw Temp.	Std. Error of *AT* Contr.	Diff. Req'd. to be Sign. ($r_{.05}$ = 3.44)	Per Cent of Zero Cycle Dynamic *E* for Specimens Made with Coarse Aggregate				Std. Error of *CT* Contr.	Diff. Req'd. to be Sign. ($r_{.05}$ = 3.34)	Per Cent of On Dynamic *E* for Specimens made with Cement	
				A_1	A_2	A_3	A_4			C_1	C_2
40	T_1	1.46	5.1	96.6	83.9	90.5	90.1	1.03	3.4	88.8	91.8
	T_2			95.4	87.4	88.7	91.1			88.3	93.1
	T_3			87.2	86.9	72.8	75.2			75.6	85.5
80	T_1	2.53	8.7	96.5	82.7	86.3	87.8	1.79	6.0	86.6	90.1
	T_2			94.6	87.2	88.1	87.2			85.9	92.7
	T_3			82.8	85.6	63.2	6.4			65.6	80.4
120	T_1	2.86	9.9	95.7	79.9	83.6	84.6	2.02	6.8	83.7	87.8
	T_2			94.1	86.3	82.9	84.6			83.3	90.4
	T_3			7 .6	82.3	52.6	49.2			55.2	72.6
160	T_1	3.38	11.6	95.4	79.2	84.2	85.4	2.39	8.0	83.0	88.9
	T_2			94.3	83.7	81.8	82.6			79.8	90.7
	T_3			69.3	76.9	53.8	48.1			52.2	72.1

Note: Significantly different means are indicated thus; x x.

means across each factor in the interaction. When interactions involving three or more factors are significant, contrasts may still be made among the means although the interpretation becomes more difficult.

8. FUTURE INVESTIGATIONS

The *F* tests from the analysis of variance and the contrasts of means show significance with a remarkable consistency at each cycle time. An important implication to the experimenter is that the number of cycles needed for the detection of true differences is perhaps no more than forty. Subsequent experiments will indicate whether or not this circumstance is generally true of concrete durability data.

The statistical treatment of the data from this experiment is now finished. It is left to the experimenter to explain *why* there were differential effects among the levels of the factors and their interactions. For this experiment lithological classification of aggregate particles; heavy-media separation of aggregate particles into specific gravity groups; quick-chemical analysis for alkali-reactivity of the finely-crushed gravels; determination of the alkali content of the cements; and mortar bar expansion tests of aggregate-cement combinations have been determined for the purpose of explaining the differences that occurred among the results. This information has been supplemented by field observations.

Each of these supplemental experiments is subject to statistical analysis and is based on a particular mathematical model. The discussion of the results of these experiments is beyond the scope of this paper and thus is omitted. They were mentioned here to show how the authors attempted to explain why certain differences occurred in the freezing and thawing tests.

If future freezing and thawing experiments were to be run on similar concretes, a great deal of useful information has been obtained from this experiment. For instance it appears that with as few as forty cycles all the important differences will be present in the data. Further, it is clear that thawing the specimens at 100 F. better reflects the differences in field durability of concrete made with the aggregates and cements used in this study and the other two thawing temperatures may be discarded. Thus, we have defined a new experiment which will yield essentially the same information with only a fraction of the effort. The new experiment would involve duplicate

mixes of three beams each for each aggregate-cement combination. These beams would be subjected to forty cycles of freezing and thawing at the rate of two cycles per day. The thaw water would be maintained at 100 deg. F. and the freezer set for −18 deg. F., as in the previous experiment. The mathematical model for such an experiment would be

$$Y_{hijl} = \mu + A_h + C_i + (AC)_{hi} + M_{hij} + S_{hijl}$$

if a completely randomized design is again used. Since the thaw temperature has been fixed at one level there is no longer a term in the model for it or for interactions with it. Since there are two mixes of each aggregate-cement combination, there will always be the opportunity to test for significant aggregate-cement interaction although the results from this experiment indicate that it does not exist. This type of experiment would have great utility for anyone using gravels and cements from the same sources for a long period of time. It could be repeated from time to time to check on the uniformity of the durability of concrete being produced.

REFERENCES

1. W. E. Haskell, "Example of Use of Analysis of Variance in Investigation of Some Properties of Portland Cement Concrete," *California Highways and Public Works*, May–June, 1953.
2. P. J. F. Wright, "Statistical Methods in Concrete Research," *Magazine of Concrete Research* (London), March, 1954.
3. D. B. Duncan, "Statistical Methods Applied to Highway-Research Experiments," *Proc.*, Highway Research, Board, Vol. 33, 1954.
4. O. L. Davies, and W. A. Hay, "The Construction and Use of Fractional Factorial Designs in Industrial Research," *Biometrics*, Vol. 6, 1950.
5. O. Kempthorne, "A Simple Approach to Confounding and Fractional Replication in Factorial Experiments," *Biometrika*, Vol. 34, 1947.
6. O. Kempthorne, *The Design and Analysis of Experiments*. John Wiley and Sons, Inc., New York, 1952, Ch. 7.
7. W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis*. McGraw-Hill, New York, 1951, Ch. 14.
8. R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*. McGraw-Hill, New York, 1952, pps. 141–144.
9. G. E. P. Box, "Effects of Inequality of Variance and of Correlation Between Errors in the Two-Way Classification," *The Annals of Mathematical Statistics*, Sept., 1954, Table 2, pg. 493.
10. D. B. Duncan, "Multiple Range and Multiple F Tests," *Biometrics*, March, 1955.
11. R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research*. McGraw-Hill, New York, 1952, Article 23.2.