RICHARD W. MENSING
DAH-YINN LEE
HERBERT T. DAVID
FEBRUARY 1973

FINAL REPORT
ISU – ERI – AMES – 72251

EVALUATION OF GAP-GRADED ASPHALT CONCRETE MIXTURES

PART II: STATISTICAL DESIGN AND ANALYSIS

Iowa Highway Research Board
Project HR-157

ERI Project 900-S

Prepared in cooperation with the
Iowa State Highway Commission
and the U. S. Department of Transportation
Federal Highway Administration

ENGINEERING RESEARCH INSTITUTE
IOWA STATE UNIVERSITY
AMES, IOWA 50010 USA

ENGINEERING
RESEARCH

ENGINEERING
RESEARCH

ENGINEERING
RESEARCH

ENGINEERING
RESEARCH

ENGINEERING
RESEARCH

FINAL REPORT

EVALUATION OF GAP-GRADED ASPHALT CONCRETE MIXTURES

PART II: STATISTICAL DESIGN AND ANALYSIS PART

R. W. Mensing
D. Y. Lee
H. T. David

February 1973

Prepared in cooperation with the
Iowa State Highway Commission
and the U. S. Department of Transportation
Federal Highway Administration

ISU – ERI – AMES – 72251
Project 900-S

ENGINEERING RESEARCH INSTITUTE
IOWA STATE UNIVERSITY AMES

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle

EVALUATION OF GAP-GRADED ASPHALT CONCRETE MIXTURES
Vol. 2: Statistical Design and Analysis

5. Report Date

February 1973

6. Performing Organization Code

7. Author(s)

Richard W. Mensing, Dah-yinn Lee, and Herbert T. David

8. Performing Organization Report No.

72251

9. Performing Organization Name and Address

Engineering Research Institute
Iowa State University
Ames, Iowa 50010

10. Work Unit No.

11. Contract or Grant No.

HR-157 (ERI 900S)

12. Sponsoring Agency Name and Address

Iowa State Highway Commission
Ames, Iowa 50010

13. Type of Report and Period Covered

Final Report

14. Sponsoring Agency Code

15. Supplementary Notes

The study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

This report presents the results of a comparative laboratory study between well- and gap-graded aggregates used in asphalt concrete paving mixtures. A total of 424 batches of asphalt concrete mixtures and 3,960 Marshall and Hveem specimens were examined.

The main thrust of the statistical analysis conducted in this experiment was in the calibration study and in Part I of the experiment. In the former study, the compaction procedure between the Iowa State University Lab and the Iowa Highway Commission Lab was calibrated. By an analysis of the errors associated with the measurements we were able to separate the "preparation" and "determination" errors for both laboratories as well as develop the calibration curve which describes the relationship between the compaction procedures at the two labs.

In Part I, the use of a fractional factorial design in a split plot experiment in measuring the effect of several factors on asphalt concrete strength and weight was exhibited. Also, the use of half normal plotting techniques for indicating significant factors and interactions and for estimating errors in experiments with only a limited number of observations was outlined.

17. Key Words

asphalt concrete, gap-graded aggregate

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	61	

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LABORATORY COMPACTION CALIBRATION

Introduction

So that results obtained at the Iowa State University (ISU) Laboratory can be compared with results obtained at the Iowa State Highway Commission (ISHC) Laboratory, a laboratory compaction calibration was undertaken as part of this project. To calibrate the results obtained at the two laboratories, eight asphalt concrete mixes were used in the study. Ideally, the mixes used represented the range of mixes to be encountered in the later parts of the project. Each laboratory received half of each of the eight mixes. One Marshall specimen was prepared in each of four molds (designated A, B, C, D) for each mix at both laboratories following the ISHC molding procedure. The same molds were used at both laboratories. Thus, in all 64 specimens were prepared. Six additional specimens were prepared at the Iowa State University Laboratory, using cold extraction; they were used in a comparison of hot and cold extraction. For each specimen prepared, the bulk specific gravity (Iowa Test Method No. 503A) was determined at both laboratories, resulting in 128 measurements used for the calibration analysis. Since some specimens from Mix 8 were destroyed during hot extraction at the Iowa State University Laboratory[1] it was decided not to use Mix 8 in the analysis. This reduced the number of measurements used in the analysis to 112.

A plot of the differences in the measurements between the two labs versus the sum of the observations (Fig. 1a) indicated that for material 2, although the average difference between the measurements obtained at

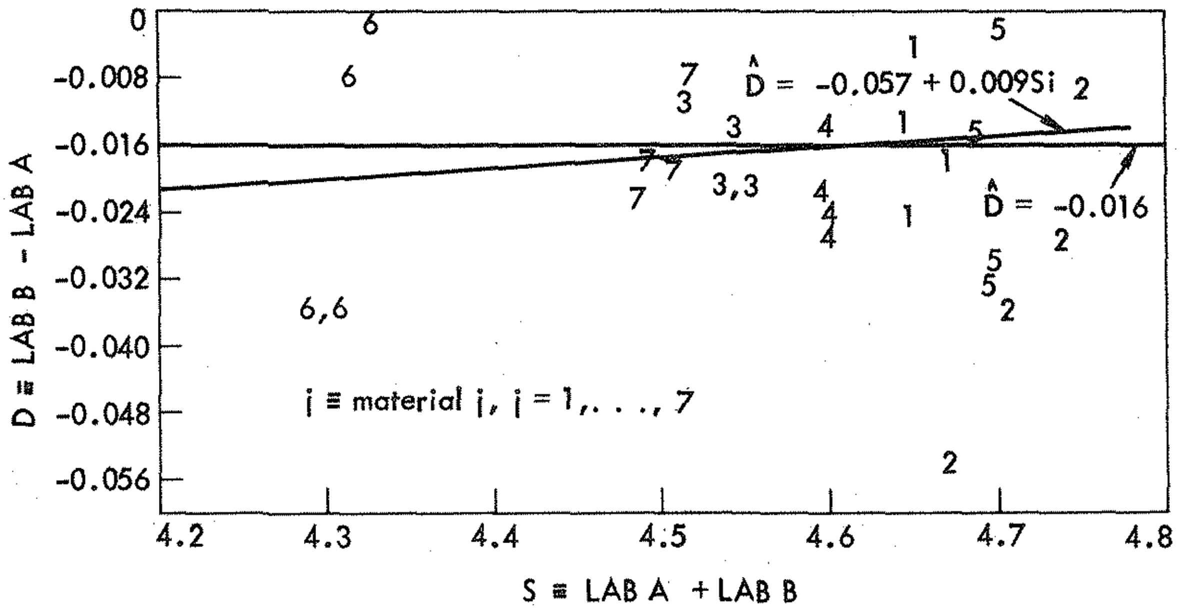

Fig. 1a. Plot of specific gravity measurements for specimens prepared and determined at the respective labs.

Lab A and Lab B was smaller, the differences appeared to have a larger variability than the other materials. From this initial observation and a test based on statistical "outlier" analysis, it was decided to exclude the data from Mix 2 when developing the calibration curve. Mix 2 was still used in the analysis of variance when estimating variance components. The outlier analysis used is summarized in the following paragraph.

To test if the average specific gravity difference $G_{AA} - G_{BB}$ for Mix 2 could be considered an outlier, the average differences were ordered:

Order:	1	2	3	4	5	6	7
Mix:	1	3	7	6	5	4	2
Means:	0.0150	0.0165	0.0168	0.0195	0.0203	0.0215	0.0313

The test statistic used was[2]

$$T_n = \frac{x_n - \bar{x}}{s} \qquad (1)$$

where x_n is the largest value, $\bar{x}$ is the mean and s is the standard deviation. Based on the mean differences observed in the calibration study, the value of T_n is 2.216. The probability of observing a value $T_n \geq 2.216$, if the difference for Mix 2 is not an outlier, is between 0.02 and 0.05. Since this probability is small it is reasonable to conclude that the difference for Mix 2 is an outlier and should be eliminated.

Data Analysis

The bulk specific gravity measurements were initially divided into the following four sets:

a) Lab A (ISHC) compacted material, readings taken at Lab A.

b) Lab A compacted material, readings taken at Lab B (ISU).

c) Lab B compacted material, readings taken at Lab A.

d) Lab B compacted material, readings taken at Lab B.

Five separate analyses of variance were run on the data. Each of the four sets listed above were run separately and then all the data were combined to get a total analysis of variance. The purpose of these runs was to get error variance estimates which, when combined, could be used to estimate the preparation and determination error variances at each of the laboratories.

For each of the four sets of data listed above the following model was used:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \qquad i = 1, \ldots, 7; \; j = 1, \ldots, 4 \qquad (2)$$

where

μ is the overall average bulk specific gravity

α_i is the deviation of the average specific gravity for material i from μ (i.e., the effect of the ith material);

β_j is the deviation of the average specific gravity for mold j from μ (i.e., the effect of the jth mold); and

ϵ_{ij} is the experimental error associated with the specimen made in the jth mold from material i.

The analysis of variance results are summarized in Table 1a through 1d.

Table 1a. Analysis of variance of specific gravity data for specimens prepared and tested at Lab A.

Source of variation	d.f.	Sum of squares	Mean squares
Materials	6	0.127445	0.021241
Molds	3	0.000332	0.000111
Error	18	0.000595	0.000033
Total	27		

Table 1b. Analysis of variance of specific gravity data for specimens prepared at Lab A and tested at Lab B.

Source of variation	d.f.	Sum of squares	Mean squares
Materials	6	0.129207	0.021534
Molds	3	0.000397	0.000132
Error	18	0.000567	0.000031
Total	27		

Table 1c. Analysis of variance of specific gravity data for specimens prepared at Lab B and tested at Lab A.

Source of variation	d.f.	Sum of squares	Mean squares
Materials	6	0.120957	0.020159
Molds	3	0.001068	0.000356
Error	18	0.002652	0.000147
Total	27		

Table 1d. Analysis of variance of specific gravity data for specimens prepared and tested at Lab B.

Source of variation	d.f.	Sum of squares	Mean squares
Materials	6	0.119776	0.019963
Molds	3	0.001349	0.000450
Error	18	0.002483	0.000138
Total	27		

The above four analyses resulted in the following error variance estimates. The experimental error associated with each set of data can be thought to consist of two components — a preparation or "compaction" error and a measurement or "determination" error. Assuming these two errors to be independent and additive with variances σ^2 and σ_D^2, respectively, estimates of these variances based on the analysis of variance given in Table 1 are:

a) $\sigma_A^2 + \sigma_{DA}^2 = 0.000033$

b) $\sigma_A^2 + \sigma_{DB}^2 = 0.000031$

c) $\sigma_B^2 + \sigma_{DA}^2 = 0.000147$

d) $\sigma_B^2 + \sigma_{DB}^2 = 0.000138$

Combining the data from all four sets into a composite sample, an analysis of variance was performed with four sources of variation separated: materials, molds, lab compaction and lab determination. The model used is:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \delta_k + (\alpha\delta)_{ik} + (\beta\delta)_{jk} + \epsilon_{ijk}$$
$$+ \gamma_l + (\alpha\gamma)_{il} + (\beta\gamma)_{jl} + (\alpha\beta\gamma)_{ijl} + (\alpha\delta\gamma)_{ikl} + (\beta\delta\gamma)_{jkl}$$
$$+ \eta_{ijkl} \qquad (3)$$

$i = 1, \ldots, 7$ $\qquad$ $j = 1, \ldots, 4$

$k = 1, 2$ $\qquad$ $l = 1, 2$

where

α_i is the effect of the ith mix,

β_j is the effect of the jth mold,

δ_k is the effect of the kth lab compaction,

γ_l is the effect of the lth lab determination,

$(\alpha\beta)_{ij}$, $(\alpha\delta)_{ik}$, ..., $(\beta\delta\gamma)_{jkl}$ are the interactions between the various sources,

η_{ijkl} is measurement error, and

ϵ_{ijk} is preparation error.

The analysis of variance based on the composite sample is summarized in Table 2. Error variance estimates based on the composite sample are:

$$\frac{1}{2}\left(\sigma_{DA}^2 + \sigma_{DB}^2\right) = 0.0000022$$
$$\sigma_A^2 + \sigma_B^2 + \frac{1}{2}\left(\sigma_{DA}^2 + \sigma_{DB}^2\right) = 0.000136. \qquad (4)$$

combining these estimates with the estimate obtained in the previous individual analyses results in the following individual estimates of the error variances.

Table 2. Analysis of variance of composite specific gravity data.

Source of variation	d.f.	Sum of squares	Mean squares
Materials (M)	6	0.496677	0.082779
Molds (m)	3	0.002426	0.000809
Lab prep. (P)	1	0.012878	0.012878
Mm	18	0.003774	0.000209
MP	6	0.000665	0.000111
mP	3	0.000706	0.000235
Error	18	0.002452	0.000136
Lab deter. (D)	1	0.000050	0.000050
MD	6	0.000017	0.000003
mD	3	0.000009	0.000003
PD	1	0.000172	0.000172
MmD	18	0.000039	0.000002
MPD	6	0.000025	0.000004
mPD	3	0.000004	0.000001
Error	18	0.000033	0.000002
Total	111	0.519927	

$$\hat{\sigma}_A^2 = 0.000012$$

$$\hat{\sigma}_B^2 = 0.000122$$

$$\hat{\sigma}_{DA}^2 = 0.000004$$

$$\hat{\sigma}_{DB}^2 = 0 \quad (5)$$

Variation due to specimen preparation and compaction is considerably larger than the determination (measurement) error. The measurement error variance for ISHC lab was of the same order, 0.002, as the rounding error introduced by recording the measurements to 0.5 or 0.0 and not interpolating between the measuring scale.

The compaction error for the Iowa State University Laboratory is considerably larger than that of the Iowa Highway Commission Laboratory. This large error could have resulted from: (a) frequent occurrences, during the compaction, of a malfunction of the newly constructed compactor (four or five specimens had to be finished with hand compaction) and (b) inexperience on the part of the ISU lab personnel in using the ISHC compaction procedure (such as the use of a pronged tunnel in introducing material into the mold, hot extraction and removal of hot specimens, etc.). Both of these problems have been eliminated since the formal commencing of HR-157[1].

To develop a calibration curve for laboratory compaction between the Iowa State Highway Commission Laboratory and the Iowa State University Laboratory, a regression analysis was conducted. The data used in the analysis were the specific gravity measurements obtained at Lab A and Lab B respectively on specimens prepared at the same laboratory. The independent variable included in the regression model is the sum of the specific gravity readings ($G_{BB} + G_{AA}$) at Lab A and Lab B. The dependent variable is the difference of the readings ($G_{BB} - G_{AA}$) from Lab B and Lab A. Using the sum of the measurements ($G_{BB} + G_{AA}$) as the independent variable, and the difference ($G_{BB} - G_{AA}$) as the dependent variable is just a transformation of axis when comparing Lab A versus Lab B data.

In regression analysis, the dependent variable is generally treated as the fixed known constant. In this study, the "true" specific gravity values for the mixes are unknown. To regress the measurements of one lab on the measurements of the second lab would be treating the measurements from the second lab as fixed. Although the transformation does not alleviate this problem, it is felt that the sum $(G_{BB} + G_{AA})$ or the average $(G_{BB} + G_{AA})/2$ is closer to the true value than the measurement obtained from either of the two labs. The regression model of the differences versus the sum would be the zero function if the only variation in the data is due to experimental error. On the other hand, if there is a possible constant bias, or, if the difference is a function of the type of materials used in the asphalt concrete, the functional relationship between the differences and the sum can be described by some polynomial. After looking at the data as plotted in Fig. 1a it was decided to use the simple linear regression model as the initial model. As indicated earlier only the data from Mixes 1, 3, 4, 5, 6, and 7 were used in this analysis. Using the simple linear model,

$$D_i = \beta_o + \beta_1 S_i + \epsilon_i \qquad i = 1, \ldots, 24$$

$$\epsilon_i \sim NID\ (o, \sigma^2) \tag{6}$$

where

D_i is the difference in specific gravity readings, $G_{BB} - G_{AA}$, and

S_i is the sum of the specific gravity readings, $G_{BB} + G_{AA}$.

The estimated regression equation is:

$$\hat{D}_i = -0.057 + 0.009S_i \ . \tag{7}$$

The analysis of variance associated with this model is summarized in Table 3.

Table 3. Analysis of variance for a simple linear equation.

Source of variation	d.f.	Sum of squares	Mean squares
Regression (β_1)	1	0.000033	0.000033
Residual	22	0.004091	0.000186
Total	23	0.004124	

A test of significance of β_1, i.e., a test of the hypothesis

$$H: \quad \beta_1 = 0$$
$$A: \quad \beta_1 \neq 0 \qquad (8)$$

indicated that the hypothesis $\beta_1 = 0$ could not be rejected. Based on this result, the model

$$D_i = \alpha_o + \epsilon_i \qquad i = 1, \ldots, 24$$
$$\epsilon_i \sim \text{NID}\ (o, \sigma^2) \qquad (9)$$

is appropriate. The estimated regression equation based on this model is the constant

$$\hat{D}_i = -\ 0.016\ . \qquad (10)$$

The two estimated regression equations are included in Fig. 1a. A 95% confidence interval for the true difference, using the variance estimates indicated earlier, is given by the limits (- 0.006, - 0.026). As indicated by this confidence interval, the observed difference is significantly different from zero. That is, the difference is larger than can be expected from random variations due to compaction and/or determination variability. The conclusion that can be drawn from these results is that there exists a constant difference in the specific gravity measurements taken at the two laboratories; the specific gravity determined on specimens compacted at the ISU laboratory is 0.016 lower than the readings recorded at the ISHC Laboratory on ISHC specimens. This difference was independent of the material tested, at least for the range of mixes used in this experiment.

It should be noted that this difference, 0.016, is the combined difference due to compaction and specific gravity determination at the two labs. A review of Fig. 1b, in which several comparisons of the data compacted and/or determined at the two labs are plotted, indicates that the majority of the difference is due to the difference in compaction between the labs. As was indicated earlier, problems with the new compactor and training personnel in using the ISHC compaction procedures could possibly explain the observed difference (especially the compactor adjustment). It is recommended that, should a comparison of results between the two labs be made, the two compactors should be first calibrated with respect to each other.

A comparison of specific gravity for cold and hot extracted specimens was made for the five material-and-mold combinations for which both types of specimens were prepared. The results are summarized in Table 4. Based on these few observations there is no reason to conclude that there is any difference in the specific gravity between cold and hot extracted specimens. The test for cold versus hot extraction is: Under the

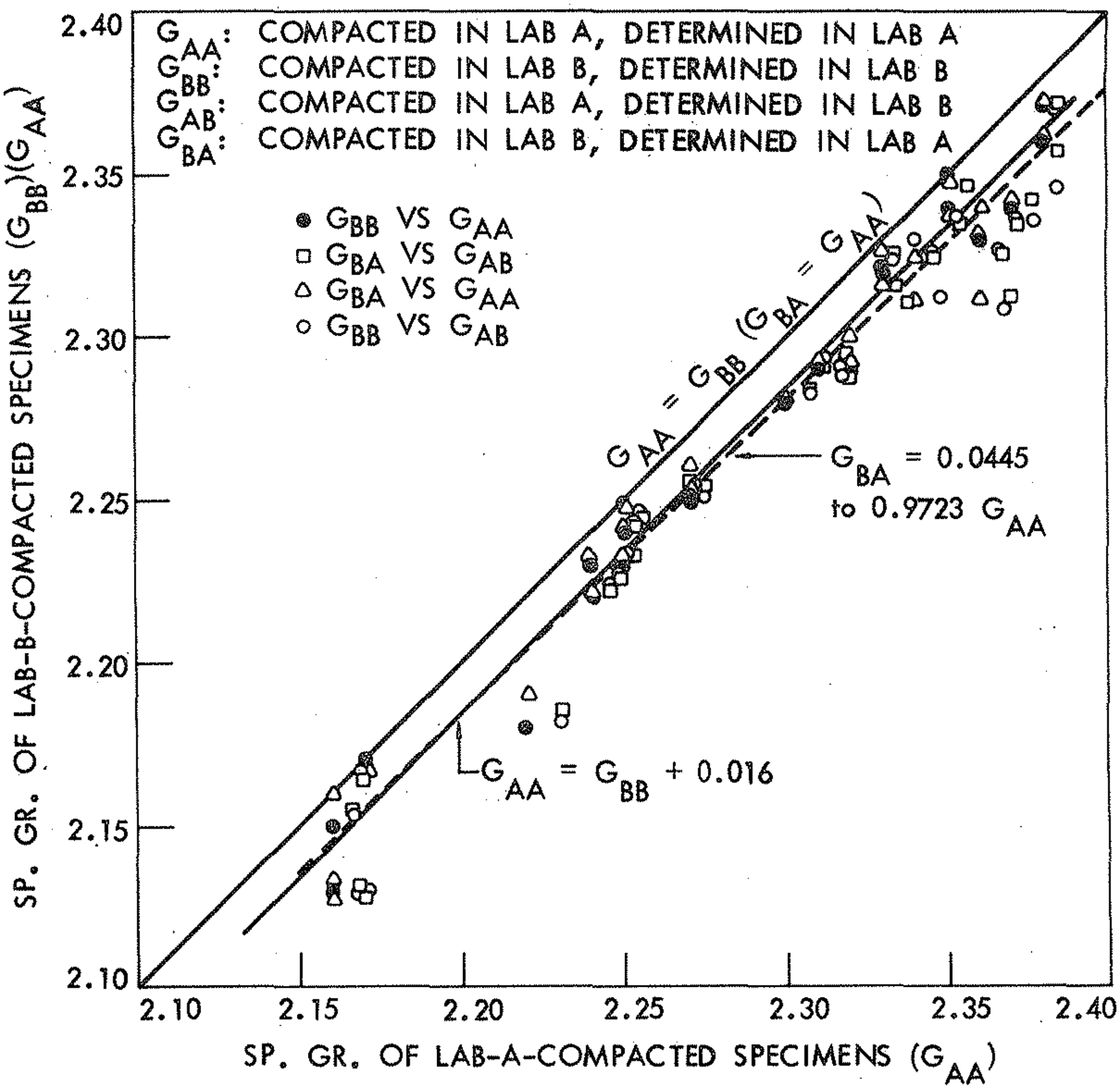

Fig. 1b. Comparison of data from labs.

Table 4. Comparison of cold and hot extraction.

Batch	Hot	Cold	Differences, D
B3D	2.242	2.236	0.006
B4A	2.288	2.305	- 0.017
B4B	2.295	2.288	0.007
B4C	2.292	2.294	- 0.002
B4D	2.284	2.399	- 0.015

$\overline{D} = -0.0042$; $S_D^2 = 0.00129$; $T = \overline{D}/\sqrt{S_D^2/n} = 0.83$

hypothesis, H: $\mu_c = \mu$, i.e., the true average specific gravity for cold extracted specimens is the same as that for hot extracted specimens, the test statistic

$$T = \overline{D}/\sqrt{s_D^2/n} \tag{11}$$

is a t-statistic with four degrees of freedom. Since $t = 0.83 < t_{0.90,4} = 1.533$, there is no reason to reject the hypothesis that the specific gravity is the same for both cold and hot extraction.

PART I: STUDY OF THE EFFECT OF SEVERAL FACTORS ON THE ASPHALT CONCRETE UNIT WEIGHT AND STABILITY

Objective

The primary purpose of Part I of the experimental program was to evaluate the effect of several variables (factors) on the strength of asphalt concrete mixtures. The factors considered to affect asphalt concrete strength and which were included in this part of the experiment were: aggregate type, aggregate gradation (maximum size and size distribution), asphalt grade, percent asphalt and amount of compaction. These factors and the levels selected for study in Part I are summarized in Table 5.

Table 5. Factors and levels included in Part I.

Factors[a]	Levels
Aggregate type (A)	Limestone; gravel
Aggregate gradation	
Maximum size (S)	3/8 in.; 3/4 in.
Size distribution (D)	BPR grading; gap 30 grading
Asphalt grade (G)	60 pen; 100 pen
Percent asphalt (P)	4%; 5%; 6%; 7%
Amount of compaction (C) (Marshall)	50 blows; 75 blows

[a]The symbols in parentheses are the letters used to identify each of the factors in the analysis of the data.

Additional related points of interest which were investigated in Part I are:

(1) Investigation of the removal of outlier observations prior to the statistical analysis.

(2) Derivation of the response curve of asphalt concrete strength as a function of the significant factors determined by the experiment.

(3) Investigation of optimum strength as a function of the factors included in the study.

(4) Discussion of several sources of experimental error as determined by the experimental design and analysis.

(5) Investigation of the effect, if any, of several different extraction procedures (hot, air cooled, water cooled) and curing times (1/2 hour, 2-4 weeks, 180 days, 360 days) on the unit weight and stability.

Experimental Design

Of the six factors included in Part I of the experimental program, five factors (A, S, D, G, and C) are at two levels and one factor (P) is at four levels. A complete analysis of these factors includes measuring the effect of each factor alone as well as measuring the joint effect of more than one factor. Such joint effects are called interactions e.g., a two-way interaction, is a measure of the change in the effect of one factor at different levels of a second factor. A full factorial experiment provides the capability of measuring all these effects and interactions since all levels of each factor are represented in combination with all levels of every other factor. In this experiment, the full factorial analysis would require the preparation of 64 batches. To reduce the number of batches to be prepared, a one-half fraction (i.e., 32 batches) of the factor level combinations were chosen. The factor level combinations selected were based on the assumption that certain higher-order interactions could be assumed to be negligible.

With percent of asphalt at four levels it is desirable to evaluate the linear, quadratic, and cubic effects of this factor. In designing the experiment and in determining the appropriate percent of asphalt levels to be included in the experiment two pseudofactors (labeled P1 and P2), each at two levels, were introduced. The four levels of percent of asphalt were associated with the four combinations of levels of P1 and P2 so that the quadratic effect of percent of asphalt is equivalent to the main effect of one pseudofactor. The linear and cubic effects of percent asphalt are equivalent to combinations of the main effects and interaction of P1 and P2.

The 1/2 fraction used in this experiment is based on a 1/2 fraction of a 2^6 factorial experiment involving the six factors A, S, D, G, P1, P2 and the confounding identity

$$I = ASGP1P2 \ . \qquad (12)$$

Both levels of C (compaction) were used for every batch of material prepared. The factor level combinations run in Part I are given in Table 6.

Table 6. Estimable main effects and interactions.

Main effects:	A, S, D, G, P [P linear (P_L); P quadratic (P_Q); P cubic (P_C)], C
Two-way interactions:	AXS, AXD, SXD, AXG, SXG, DXG, AXP_L, SXP_L, DXP_L, GXP_L, AXP_Q, SXP_Q, DXP_Q, GXP_Q, DXP_C, AXC, SXC, DXC, GXC, CXP_L, CXP_Q, CXP_C
Three-way interactions:	AXSXD, AXDXG, SXDXG, $AXDXP_L$, $SXDXP_L$, $DXGXP_L$, $AXDXP_Q$, $SXDXP_Q$, $DXGXP_Q$, AXSXC, AXDXC, SXDXC, AXGXC, SXGXC, DXGXC, $AXCXP_L$, $SXCXP_L$, $DXCXP_L$, $GXCXP_L$, $AXCXP_Q$, $SXCXP_Q$, $DXCXP_Q$, $GXCXP_Q$, $DXCXP_C$
Four-way interactions:	AXSXDXC, AXDXGXC, SXDXGXC, $AXDXCXP_L$, $SXDXCXP_L$, $DXGXCXP_L$, $AXDXCXP_Q$, $SXDXCXP_Q$, $DXGXCXP_Q$

With the 1/2 fraction factorial experiment it is possible to estimate 63 main effects and interactions under the assumptions that all the remaining interactions are negligible. The estimable main effects and interactions are listed in Table 6. Note that all main effects are estimable including the linear, quadratic, and cubic effects of percent of asphalt. All two-way interactions are estimated, except for the interactions AXP_C, SXP_C, and GXP_C. Also, all three-way and some four-way interactions involving aggregate size distribution (D) are estimated. Although this is different than the one-half fraction originally proposed, we feel that the experiment, as run, puts more emphasis on the difference between aggregate size distributions (gap graded versus conventional grading) which was of major interest in this project.

Since the linear and cubic effects of percent of asphalt are combinations of the effects P1 and P2 and the interaction P1P2, we observed the fact that with the one-half fraction used some of the interactions involving P_L and P_C, although not confounded (i.e. estimable), were correlated with (i.e., not orthogonal to) some of the interactions not involving percent of asphalt. The majority of the estimated main effects and interactions are orthogonal.

The experimental procedure used in carrying out Part I can be thought of as a two-phase mixing-compaction sequence as follows:

Mixing Phase

Thirty-two different batches were prepared. Each batch represented one of the 32-factor-level combinations of the five factors: aggregate, maximum size, aggregate size distribution, asphalt grade and percent of asphalt. The order of preparation was completely randomized.

Four additional batches were prepared. These were a single replication of 4 of the 32-factor-level combinations used. The factor level combinations replicated were:

(L, 3/4", Gap 30, 100, 6)

(G, 3/4", BPR, 60, 6)

(G, 3/8", Gap 30, 60, 5)

(G, 3/4", BPR, 100, 7)

The replicated batches were used: (1) to get an estimate of experimental error and (2) to investigate the effect of different extractions and different curing times.

Compaction Phase

For each of the 36 batches mixed, 14 specimens were prepared from each batch. Seven specimens were subjected to 50 compactive blows per side and the remaining seven were subjected to 75 blows per side. The assignment of specimen to compactive efforts was done at random.

An experiment using the type of experimental procedure described above is often referred to as a split plot experiment. The main plots (whole plots) in this experiment are the batches, the subplots (split plots) are the 14 specimens prepared from each batch. The different number of blows used in compacting a specimen is the split plot factor. Since there were seven specimens prepared using 50 blows and seven specimens prepared using 75 blows, this represented a replication of the basic two levels of the split plot factor 7 times.

Two additional phases which can be considered in the preparation of test specimens are: (1) extraction of a specimen from the mold and (2) curing time of a specimen prior to testing. The bulk of the specimen were extracted while they were hot and were cured for two to four weeks before testing. To study the effect of the type of extraction and curing time on stability, some of the specimens in the replicated batches were subjected to different types of extraction (air and water cooled) and/or different curing times (1 day, 180 days and 360 days).

Since the experimental procedure is two phase, it is necessary that the analysis of the data be cognizant of the errors in the data introduced by the experimental procedure. For the procedure described above two types of errors must be considered. These are split plot error and whole plot error.

Split Plot Error

For a given batch of material (i.e., combinations of levels of factors A, S, D, G, and P) 14 specimens were prepared. Seven were compacted using 50 blows and seven compacted using 75 blows. After compaction, the specimens were tested. Two measurements observed were unit weight and stability. Random variation in the measurements taken on specimens made from the same batch of material is due to several experimental conditions. Some of these are variation in the material from specimen to specimen, variation in compaction from one specimen to another and variation in making the tests and taking the observations. All these sources of variation are combined into a single error called the split plot error.

Whole Plot Error

In addition to the split plot errors above, specimens prepared from different batches have a batch to batch variation due to the variation in the mixing process and the difference in materials used in the different batches. This error is called the whole plot error.

These two errors affect the analysis of the data observed in a split plot experiment. Comparison of the observations between the two compactive efforts have only the split plot error associated with them.

On the other hand, comparisons among the other factors, i.e., A, S, D, etc., are based on comparisons between batches and so have both the split plot and batch errors (i.e., the whole plot error) associated with them.

Analysis of the Data

The data observed in Part I of this project is given in Table 7. Although several measurements, unit weight, stability, flow, air voids and VMA were taken on each specimen, the analyses were only made on unit weight and stability because their known sensitivity with respect to the variables involved.

The analysis of the data was based on a split plot experiment with replicated subplots involving a one-half fraction of the factor level combinations of six factors (five factors at two levels and one factor at four levels).

As indicated in the description of the experimental design, seven specimens were to be prepared at each factor-level combination. An examination of the data indicated that for some combinations data were available from only six specimens. Further examination of the data suggested that when seven measurements were recorded the first observations deviated considerably from the remaining six measurements. This was generally not true when only six observations were available.

The reasons for the deviant behavior of the first specimen in each batch were due mainly to:

(a) Specimen 1 was always the trial specimen; due to differences in gradation, asphalt content, etc., it may or may not have been within 2.50 $\pm$ 0.05 in. in height. Specimens 2 to 7, on the other hand, were prepared with the same adjusted weight of mixtures to give sample height within the limits.

Table 7. Physical properties of Marshall specimens — Series A.

Batch No.	Gradation	Type agg.	AC grade pen.	Compactive blows	Extraction & time of testing	Design AC by wt. agg., %	AC by wt. mix, %	Unit wt., pcf	Marshall stab., lb	Flow, in. (x 0.01)	Air voids, %	VMA, %
2	A-P	G	100	50	2-h[a]	6	5.60	152.5	1998	18	0.74	12.76
35					2-h	7	6.57	152.5	1394	35+	0.00	13.81
36					2-h	7	6.57	152.0	1557	25	0.07	13.88
					3-w[b]			151.8	1650	34	0.21	13.63
					3-a[c]			152.4	1632	24	0.00	13.32
35					1-w	7	6.57	151.5	1296	35+	0.00	13.81
					1-a	7	6.57	151.6	1339	35+	0.00	13.81
					4-a	7	6.57	151.8	1425	32	0.00	13.63
					4-w	7	6.57	152.3	1425	34	0.00	13.39
2	A-P	G	100	75	2-h	6	5.60	153.6	2023	20	0.00	12.08
					2-h	7	6.57	152.4	1449	35+	0.10	13.56
					3-w			152.4	1617	31	2.44	13.28
					3-a			152.4	1559	25	0.00	13.28
					4-w			152.9	1600	22	0.00	13.03
					4-a			152.9	1625	31	0.00	13.03
36					2-h	7	6.57	152.7	1510	26	0.00	13.63
					1-w	7	6.57	152.1	1669	20	0.00	13.63
					1-a	7	6.57	151.8	1625	23	0.10	13.62
3	A-P	G	60	50	2-h	4	4.05	149.4	2317	12	5.40	13.07
30					2-h	6	5.47	153.0	2300	21	2.25	12.33
33					2-h	6	5.74	154.2	2499	21	1.05	12.08
30					1-a	6	5.47	152.6	2294	16	2.25	12.33
					3-a			153.7	2250	23	2.03	11.76
					4-a			152.8	2025	17	2.39	12.08
33					1-w	6	5.74	153.1	2880	18	1.05	12.08
					3-w			153.6	2450	23	2.23	11.86
					4-w			153.4	2700	20	2.34	11.97
3	A-P	G	60	75	2-h	4	4.05	152.5	2920	13	3.45	11.28
30						6	5.47	154.3	2367	18	1.69	11.83
33						6	5.74	154.0	2352	22	1.01	12.04
30					1-w	6	5.47	152.7	2496	17	1.69	11.83
					3-w			153.6	2303	17	1.87	11.61
					4-w			153.3	2625	20	2.03	11.76
33					1-a	6	5.74	153.7	2420	28	1.01	12.04
					3-a			153.8	2100	23	2.11	11.76
					4-a			153.6	2400	24	2.19	11.83
15	A-P	L	100	50	2-h	4	4.05	148.5	3464	12	4.13	11.56
25	A-P	L	10C	50	2-h	6	5.74	150.4	2057	23	0.08	12.02
15	A-P	L	100	75	2-h	4	4.05	149.8	3957	12	3.25	10.74
25	A-P	L	100	75	2-h	6	5.74	150.5	2244	23	0.00	11.95
17	A-P	L	60	50	2-h	5	4.76	151.3	3170	17	1.76	10.55
28						7	6.43	149.9	2162	21	0.12	12.95
17	A-P	L	60	75	2-h	5	4.76	152.4	2738	19	1.07	9.92
28						7	6.43	150.8	2279	22	0.00	12.41
4	A-30	G	100	50	2-h	5	5.04	153.9	2948	15	2.04	11.32
24						7	6.57	152.7	1505	31	0.40	13.44
4	A-30	G	100	75	2-h	5	5.04	154.4	2697	16	1.76	11.06
24						7	6.57	153.1	1582	30	0.12	13.19
16	A-30	G	60	50	2-h	4	3.90	149.5	3766	11	5.35	12.82
21						6	5.60	153.8	2631	23	0.00	11.90
16	A-30	G	60	75	2-h	4	3.90	150.2	4124	10	4.92	12.42
21						6	5.60	153.6	2231	24	0.00	12.00
14	A-30	L	100	50	2-h	4	3.90	147.1	3937	10	6.49	12.25
29						6	5.60	150.1	2059	22	0.00	11.94
31						6	5.60	150.6	2130	20	0.00	11.69
29					1-a	6	5.60	150.5	2190	17	0.00	11.94
					3-a			150.6	1326	16	0.00	11.19
					4-a			150.3	2475	19	0.00	11.38
31					1-w	6	5.60	150.6	1987	16	0.00	11.69
					3-w			151.0	2171	19	0	10.97
					4-w			150.9	2325	17	0	11.01
14	A-30	L	100	75	2-h	4	3.90	148.1	4554	10	5.86	11.66
29						6	5.60	150.9	2107	23	0.00	11.65
31						6	5.60	151.1	2327	19	0.00	11.39
29					1-w	6	5.60	150.6	2470	18	0.00	11.65

Table 7. Physical properties of Marshall specimens — Series A, continued.

Batch No.	Gradation	Type agg.	AC grade pen.	Compactive blows	Extraction & time of testing	Design AC by wt. agg., %	AC by wt. mix, %	Unit wt., pcf	Marshall stab., lb	Flow, in. (x 0.01)	Air voids, %	VMA, %
					3-w			150.5	2373	20	0.00	11.30
					4-w			150.8	2225	17	0.00	11.08
31					1-a	6	5.60	151.1	2640	18	0.00	11.39
					3-a			151.4	2303	20	0.00	10.71
					4-a			151.4	2375	18	0.00	10.71
12	A-30	L	60	50	2-h	5	4.76	149.1	2971	17	2.52	11.86
18	A-30	L	60	50	2-h	7	6.57	150.2	1539	33	0.34	12.88
12	A-30	L	60	75	2-h	5	4.76	150.6	3006	18	1.54	10.99
18	A-30	L	60	75	2-h	7	6.57	150.9	1776	28	0.00	12.45
10	C-P	G	100	50	2-h	6	5.74	152.0	2354	21	0.83	12.77
13						4	4.05	143.2	2974	10	9.48	16.38
10	C-P	G	100	75	2-h	6	5.74	152.5	2496	22	0.62	12.59
13						4	4.05	147.1	4381	10	6.99	14.09
5	C-P	G	60	50	2-h	7	6.71	151.0	1712	27	0.55	14.27
19						5	4.76	150.9	4427	13	2.31	12.48
1	C-P	L	100	50	2-h	8	7.51	151.2	1595	24	0.69	15.00
23						5	4.90	146.7	4125	11	4.89	14.99
1	C-P	L	100	75	2-h	8	7.51	151.8	1851	24	0.00	14.40
23						5	4.90	149.9	5049	12	2.79	13.11
22	C-P	L	60	50	2-h	6	5.60	147.1	4146	12	3.27	15.33
26						4	4.05	145.4	4110	10	7.91	14.96
22	C-P	L	60	75	2-h	6	5.60	149.7	5059	13	1.59	13.86
26						4	4.05	145.8	5300	11	7.64	14.70
11	C-30	G	100	50	2-h	6	5.74	151.9	3215	16	0.26	12.75
20						4	3.90	145.6	3228	9	7.90	14.74
11	C-30	G	100	75	2-h	6	5.74	152.1	2778	16	0.26	12.75
20						4	3.90	149.4	4823	10	5.45	12.47
6	C-30	G	60	50	2-h	7	6.71	150.8	2204	24	1.06	14.32
32					2-h	5	5.47	150.3	4711	11	4.47	13.58
					3-a			151.1	5500	9	4.23	12.65
					3-w			147.3	3626	12	6.61	14.82
					4-a			150.1	4250	12	4.87	13.23
					4-w			149.9	4250	12	4.98	13.34
34					2-h	5	4.76	149.3	3326	11	4.59	13.47
34					1-w	5	4.76	149.0	3792	11	4.59	13.47
34					1-a	5	4.76	148.5	3360	11	4.59	13.47
6	C-30	G	60	75	2-h	7	6.71	151.1	2528	22	0.82	14.11
32					2-h	5	5.47	152.3	5301	13	3.47	12.32
					3-a			152.3	2131	9	3.44	11.93
					3-w			151.7	4999	8	3.82	12.29
					4-a			152.1	5900	14	3.60	12.07
					4-w			152.1	4900	9	3.60	12.07
34					2-h	5	4.76	149.3	3703	11	4.23	13.15
					1-w	5	4.76	149.8	4262	11	4.23	13.15
					1-a	5	4.76	151.2	5261	10	4.23	13.15
8	C-30	L	100	50	2-h	5	4.62	147.5	4979	11	5.60	13.87
27						7	6.71	150.0	2426	18	0.32	14.34
8	C-30	L	100	75	2-h	5	4.62	150.4	5374	10	3.73	12.16
27						7	6.71	150.8	2730	18	0.00	13.87
7	C-30	L	60	50	2-h	6	5.74	149.3	4139	13	1.06	13.59
9						4	4.05	144.6	5563	11	8.98	15.08
7	C-30	L	60	75	2-h	6	5.74	150.2	3678	16	0.76	13.34
9						4	4.05	148.4	6883	12	6.51	12.77

Extraction:
(a) Hot extraction.
(c) Air cooled extraction.
(b) water cooled extraction.

Time of stability test:
1: 1 day after compaction.
2: regular (3 days).
3: 180 days after compaction.
4: 360 days after compaction.

(b) For maximum efficiency, the daily mixing and compaction procedure was to mix batch 2, compact specimen 1, weigh out specimens 2 to 7 and maintain these at the compaction temperature; mix batch 2, compact specimen 1 of batch 2, weigh out specimens 2 to 7 of batch; mix batch 3, etc. ... After all 5 batches were mixed and specimen 1 of the 5 batches was compacted, then specimens 2 to 7 of batch 1, 2 to 7 of batch 2, etc., were compacted. Consequently, specimen 1 was compacted immediately after mixing and specimens 2 to 7 were compacted at about the same time but one to two hours later.

Although not all first specimens were deviant for all the batches, the convenience in the analysis gained by having an equal number of specimen per factor level combination led us to eliminate the first observation whenever 7 measurements were recorded. Thus, the total number of observations used in the analysis was 432 (384 observations from the 64-factor-level combinations plus 48 observations from the replicated batches).

A preliminary regression analysis was performed on the 384 observations obtained from hot extracted specimen which were cured for two to four weeks prior to testing. Using the symbols indicated in Table 5 to indicate the effects and interactions of the six factors, the model used in the analysis is:

$$Y = \mu + A + S + AS + D + AD + SD + ASD + G + AG + SG + DG + ADG$$
$$+ SDG + P_L + AP_L + SP_L + DP_L + ADP_L + SDP_L + GP_L + DGP_L + P_Q$$
$$+ AP_Q + SP_Q + DP_Q + ADP_Q + SDP_Q + GP_Q + DGP_Q + P_C + DP_C + \epsilon$$
$$+ C + AD + SD + ASD + DC + ADC + SDC + ASDC + GC + AGC + SGC$$
$$+ DGC + ADGC + SDGC + CP_L + ADP_L + SCP_L + DCP_L + ADCP_L + SDCP_L$$
$$+ GCP_L + DGCP_L + CP_Q + ACP_Q + SCP_Q + DCP_Q + ADCP_Q + SDCP_Q$$
$$+ GCP_Q + DGCP_Q + CP_Q + DCP_Q + \delta \qquad (13)$$

The terms P_L, P_Q and P_C represent the linear, quadratic, and cubic effects of the percent of asphalt in the mix. Also, ϵ and δ refer to the whole plot and split plot errors, respectively.

As discussed in the previous section, almost all of the terms in the model are orthogonal. The few nonorthogonal terms were checked and found to fall well within the overall plots, and hence were included in the analysis.

An analysis of the variation observed in the data, based on the model in Eq. 13 is given in Table 8. Each term in the model represents a potential source of variation. These are listed in the table. The sum of squares (or mean squares) reflects the significance of each factor and/or combination of factors in explaining the total variation in the data.

To assess the significance of each term in the model given in Eq. 13 it was necessary to estimate the appropriate error variances. These estimates were derived using half normal plotting techniques. Among the terms (main effects and interactions) in the model in Eq. 13, the sum of squares associated with any term including compaction (C) involves only the compaction (split plot) error. Thus, the effects and interactions which include C, if they are nonsignificant, can be used to estimate

Table 8. Preliminary analysis of variance for Series A.

Source of variation	d.f.	Unit weight, sum of squares	Stability, sum of squares
Aggregate (A)	1	349.034[a]	44259257.0[a]
Max size (S)	1	403.645[a]	125962616.6[a]
A×S	1	47.250[b]	12201212.5
Aggr. distribution (D)	1	1.438	14395244.3[c]
A×D	1	2.958	770147.9
S×D	1	14.531	939609.4
A×S×D	1	3.136	1845514.7
Asphalt grade (G)	1	0.970	16333762.5[c]
A×G	1	8.079	44483.6
S×G	1	7.964	5024036.3
D×G	1	8.313	1572736.0
A×D×G	1	18.859	1526743.1
S×D×G	1	8.138	11018.9
Asphalt percent (P)	3	983.682	261590685.1
Linear (P_L)	1	698.298[a]	254968493.9[a]
Quadratic (P_Q)	1	267.501[a]	6614737.5
Cubic (P_C)	1	17.883	7453.7
A×P_L	1	4.063	4880496.6
S×P_L	1	77.490[b]	10755852.3
D×P_L	1	1.593	8337009.0
A×D×P_L	1	7.069	729672.2
S×D×P_L	1	21.235[c]	2847120.4
G×P_L	1	10.369	905107.9

Table 8. Continued.

Source of variation	d.f.	Unit weight, sum of squares	Stability, sum of squares
D×G×P_L	1	23.850[c]	3082562.6
A×P_Q	1	47.250[b]	267073.3
S×P_Q	1	7.123	17980329.8[b]
D×P_Q	1	0.475	6492980.4
A×D×P_Q	1	1.563	1299094.3
S×D×P_Q	1	0.255	1900547.5
G×P_Q	1	2.266	101367.5
D×G×P_Q	1	0.658	3562214.1
D×P_C	1	4.135	27808.5
Compaction (C)	1	147.634[a]	11023942.6[a]
A×C	1	1.138	338022.0
S×C	1	22.282[a]	6657330.0[a]
A×S×C	1	2.888	103983.8
D×C	1	0.040	275900.6
A×D×C	1	5.631	22955.6
S×D×C	1	0.266	424735.5
A×S×D×C	1	0.532	817796.5
G×C	1	1.116	228198.8
A×G×C	1	4.356	23390.6
S×G×C	1	8.730[c]	152601.6
D×G×C	1	0.058	790.6
A×C×G×C	1	1.272	20871.3
S×D×G×C	1	1.138	42525.2

Table 8. Continued.

Source of variation	d.f.	Unit weight, sum of squares	Stability, sum of squares
$C \times P_L$	1	41.389[a]	6929411.0[a]
$C \times P_Q$	1	1.052	3712870.0[a]
$C \times P_C$	1	1.235	233311.1
$A \times C \times P_L$	1	4.748	343862.2
$S \times C \times P_L$	1	5.925	3423359.5[a]
$D \times C \times P_L$	1	0.466	143884.2
$A \times D \times C \times P_L$	1	7.849[c]	363649.7
$S \times D \times C \times P_L$	1	0.518	46838.8
$G \times C \times P_L$	1	3.431	110195.4
$D \times G \times C \times P_L$	1	2.236	248219.2
$A \times C \times P_C$	1	15.480[b]	623634.4
$S \times C \times P_Q$	1	0.001	183618.8
$D \times C \times P_Q$	1	3.546	1505880.8[b]
$A \times D \times C \times P_Q$	1	3.245	318839.1
$S \times D \times C \times P_Q$	1	2.958	1106499.4[c]
$G \times C \times P_Q$	1	0.001	33432.0
$D \times G \times C \times P_Q$	1	0.206	108709.7
$D \times C \times P_C$	1	0.627	455501.2
Residual	320	148.528	39316020.8
Total	383	2507.913	628987086.2

[a] A significant effect at 1% level of significance.

[b] A significant effect at 5% level of significance.

[c] A significant effect at 10% level of significance.

the compaction error. On the other hand, the effects and interactions involving the whole plot factors (A, S, D, G, and P) include both compaction and batch error. These effects can be used to estimate the whole plot error.

To estimate the two error variances, separate half normal plots were plotted for the two errors (whole plot and split plot error). The plots for both unit weight and stability are given in Figs. 2a through 2d.

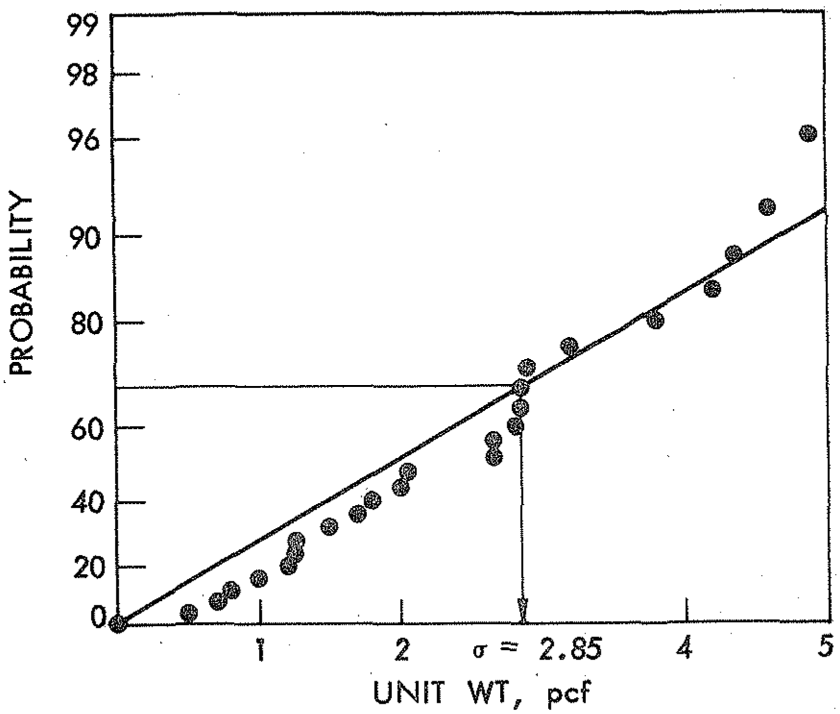

Fig. 2a. Half normal plot used to determine the whole plot error for unit weight.

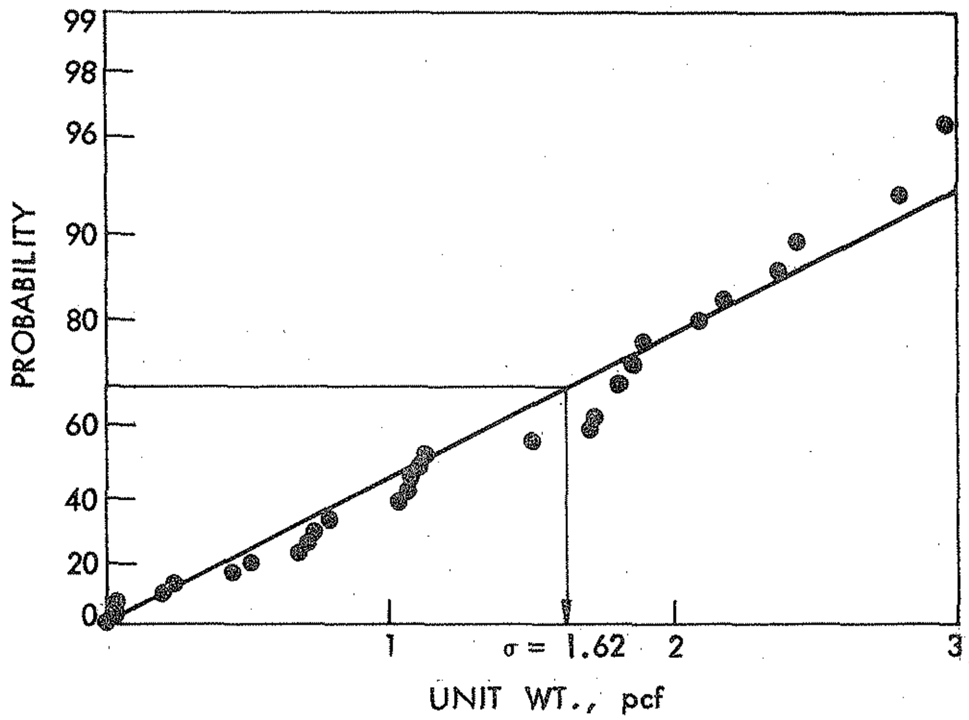

Fig. 2b. Half normal plot used to determine the split plot error for unit weight.

In addition to the error variance estimates, use of the half normal plots identifies the significant effects and interactions. A summary of this analysis is given in Tables 9 and 10 for unit weight and stability respectively.

One must be careful in drawing conclusions based on the analysis using the half normal plots. Using the results in Table 10, the conclusion would be that the only factors which have a significant effect on asphalt concrete stability are aggregate type, aggregate maximum size, percent of asphalt and compaction. This seems to contradict previous studies[3,4] which indicate that both aggregate distribution and asphalt grade are also significant factors. One explanation for this discrepancy

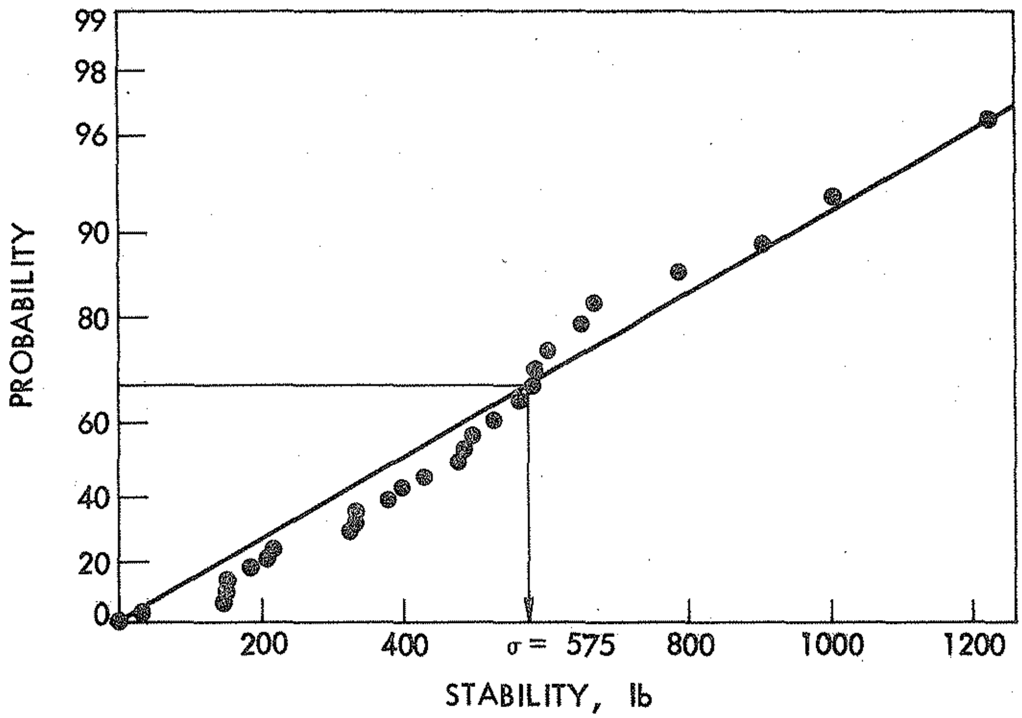

Fig. 2c. Half normal plot used to determine the split plot error for stability.

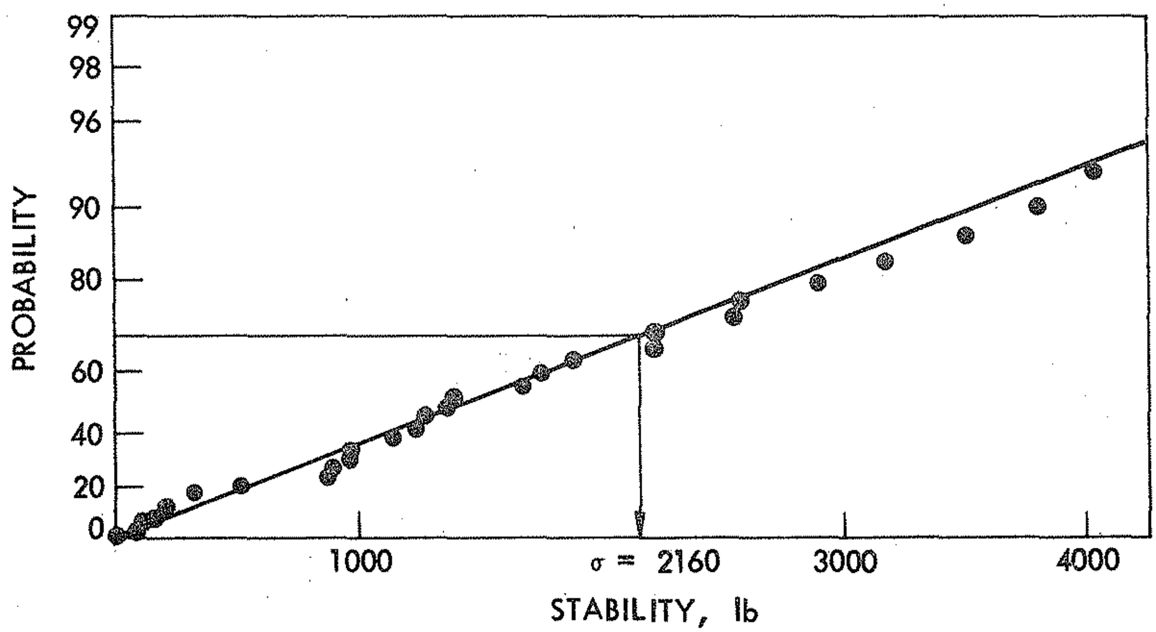

Fig. 2d. Half normal plot used to determine the whole plot error for stability.

Table 9. Analysis of variance for Series A based on half normal plot analysis; unit weight data.

Source of variation	d.f.		Sum of squares		Mean squares
Replication	1		3.321		3.321
Aggregate (A)	1		349.034		349.034
Max size (S)	1		403.645		403.645
A×S	1		47.250		47.250
Asphalt percent (P)	2		965.799		
Linear (P_L)		1		698.298	
Quadratic (P_Q)		1		267.501	
S×P_L	1		77.490		77.490
A×P_Q	1		47.250		47.250
Whole plot error	27		185.291		6.863
Effects		24		176.923	
R×T		3		8.368	
Compaction (C)	1		147.634		147.634
S×C	1		22.282		22.282
C×P_L	1		41.389		41.389
A×C×P_L	1		15.480		15.480
Split plot error	32		73.504		2.297
Effects		28		65.210	
R×C, R×T×C		4		8.294	
Total	71				

Table 10. Analysis of variance for Series A based on half normal plot analysis; stability data.

Source of variation	d.f.	Sum of squares	Mean squares
Replication	1	270991.5	279001.5
Aggregate (A)	1	44259257.0	44259257.0
Max size (S)	1	125962616.6	125962616.6
Asphalt content (P)			
Linear (P_L)	1	254968493.9	254968493.9
Whole plot error	31	131263735.6	4266572.1
Effects	28	124455936.7	
R×T	3	6807798.9	
Compaction (C)	1	11023942.6	11023942.6
S×C	1	6657330.0	6657330.0
C×P_L	1	6929411.0	6929411.0
S×C×P_L	1	3423359.5	3423359.5
C×P_Q	1	3712870.0	3712870.0
Split plot error	31	9809693.5	316441.4
Effects	27	8287847.9	
R×C, R×T×C	4	1521835.6	
Total	71		

is the conservative nature of the half normal plot analysis. Any statistical test of significance of a factor has associated with it the possibility of wrongly concluding the factor is significant when it is not. The half normal plot technique, when used to identify the significant factors and interactions has associated with it a "small" probability of making such a wrong decision. Rather than controlling error rates per contrast of the first kind, it can be thought of as controlling error rates per experiment. The net effect of this is to decrease error rates of the first kind and increase error rates of the second kind. The latter phenomenon may be involved in the case of the aggregate distribution and asphalt gradation factors.

Referring to Table 8, the level of significance associated with all the factors and interactions are based on the error variance estimates derived from the half normal plots. [Note: the whole plot variance estimates could be somewhat inflated since the estimates include both the distribution and gradation effects.] Note that both aggregate distribution and asphalt gradation are significant at a 10% significance level but not at a 5% significance level — thus the half normal plot analysis conclusion that these factors are not significant.

A second reason for the apparent contradictory results could be the fact that only two distributions and two asphalt gradations were used in the analysis. It could be, in particular, that the two distributions used in the experiment are not significantly different. This, of course, does not allow the inference that there is no difference among <u>all</u> distributions.

Also included in the tables are the estimates of the error variances obtained by an analysis of the four factor level combinations which were replicated. These are identified as R×T, R×C, and R×T×C in the analysis of variance tables, Table 9 and 10.

A regression analysis was performed on the unit weight and stability data observed in Part I. The data used in this analysis was that measured for specimen extracted while hot and for which the test took place two to four weeks after preparation. The independent variables included in the regression model were those variable (factors) found to be significant in the half normal plot analysis. The estimated regression equations for the expected unit weight (Wt) and stability (St) are:

$$\begin{aligned} Wt = {} & 93.493 - 0.109\delta_1 + 23.105S + 3.519S\delta_1 + 13.251P - 3.108PS \\ & - 0.821P^2 + 0.014P^2\delta_1 + 0.228C - 0.019SC - 0.018CP \\ & - 0.0004\delta_1 CP^2 \end{aligned} \tag{14}$$

$$\begin{aligned} St = {} & 3181.552 + 319.509\delta_1 + 407.326S - 194.225P + 67.576C \\ & - 81.710SC + 4.897CP + 4.017SCP - 1.560CP^2 \end{aligned} \tag{15}$$

where

δ_1 = 1 if the aggregate is limestone, or

- 1 if the aggregate is gravel; and where

S is maximum size in inches,

P is percent of asphalt by weight, and

C is the number of compactive blows used in the preparation of the specimen.

The response curves of unit weight and/or asphalt concrete stability can be derived from the estimated regression equations given in Eq. 14 and 15, respectively. Again one must use these equations with caution since they are based on the half normal plot analysis. The effect of aggregate distribution and asphalt gradation, along with several interactions, are not reflected in the above equations. Including these latter variables would likely alter the equations. For example, it is known that asphalt concrete stability is not a linear function of asphalt content. From this and consideration of Eq. 15, one is led to the conclusion that maximum stability occurs below four percent of asphalt. For all combinations of size and compaction the highest value of stability, using Eq. 15, occurs at four percent. Inclusion of aggregate distribution as a factor effecting stability led to an optimal (maximum) stability at a percent asphalt between four and seven percent for several aggregate gradations. This is discussed in Part II of this report.

In general, equations such as 14 and 15 can be used to predict the strength of asphalt concrete as a function of the several factors which effect stability. Thus, such equations could be used to determine the optimal percent of asphalt for a given aggregate, gradation, etc.

Using data for the four-factor-level combinations which were replicated, an analysis of the effect of three different types of extraction (hot, air cooled, water cooled) and four different curing times (1 day, 2-4 weeks, 180 days, 360 days) stability was performed. The appropriate analysis of variance is given in Table 11 for the stability measurements. As can be observed from the results in this table there is no reason to reject the hypothesis that different extractions and different curing lengths have no effect on stability measurements.

Table 11. Analysis of variance for testing the effect of extraction and curing times on stability.

Source of variation	d.f.	Sum of squares	Mean squares
Treatments (replicated asphalts)	3	105422813	
Curing time	3	693260	231086.7
Extraction	2	57825	28912.7
Experimental error	87	24368629	280099.2
Total	95	130542527	

PART II: DETAILED EVALUATION

Objective

In Part II of the experimental program, a detailed evaluation of Marshall and Hveem properties of asphalt concrete mixtures was undertaken. Of particular importance is the evaluation of the effect of different aggregate gradations on these properties. A comparison of the Marshall and Hveem properties for 33 different aggregate gradations was undertaken for two different limestones, 60- and 100-pen. asphalt grades and asphalt content between 3 and 7%. Also included in this part is an investigation of several procedures for determining the combination of aggregate gradation and asphalt content which maximized asphalt concrete strength. A discussion of these procedures is included in Vol. I of this report and hence is not included in this section. An additional series of batches were prepared using crushed gravel for several aggregate gradations. These are used for a comparison with natural gravel.

Experimental Design

The aggregate gradations included in Part II are listed in Table 12. Also included in the table are the other factors, type aggregate, asphalt grade and asphalt content, and the levels of these factors used.

As originally designed, all 33 aggregate gradations were combined with all 5 asphalt contents at each of the 4 combinations of type aggregate and asphalt grade. Thus, 660 batches were to be prepared. This design would have allowed a complete analysis of the 33 aggregate

Table 12. Factors and levels included in Part II.

1.	Aggregate type:	Limestone: L1, L2	
2.	Aggregate gradation[a]:	A-F, A-P, A-I, A-4, A-4L, A-8, A-8L, A-30,	(1-8)
		A-30L, A-100, A-100L, A-4H, A-4LH,	(9-13)
		A-8H, A-8LH, A-30H, A-30LH,	(14-17)
		B-P, B-B, B-8, B-8L, B-30, B-30L,	(18-23)
		B-100, B-100L,	(24-25)
		C-P, C-I, C-8, C-8L, C-30, C-30L,	(26-31)
		C-100, C-100L	(32-33)
3.	Asphalt grade:	60 pen., 100 pen.	
4.	Asphalt content:	3%; 4%; 5%; 6%; 7%	

[a] The paired symbols refer respectively to the maximum size (A: 3/4 in., B: 1/2 in., C: 3/8 in.), and to size distribution (F: Fuller's curve, P: Bureau of Public Roads curve, I: Iowa Highway Commission curve, 4: gap 4, 8: gap 8, 30: gap 30, 100: gap 100, L: Below-the-curve gap, and H: half gap).

gradations at all levels of the other factors. Also, rankings of the 33 gradations for all combinations of limestone and asphalt grade would have been available. The experiment as designed was to be run in four series (Series B-E), each series identified by the combination of limestone and asphalt grade to be used in the batches prepared in that series.

Due to time and material limitations only 335 batches were prepared. The combinations of factors used in preparing these batches are listed in Table 13. All 165 batches in Series B were nearly complete at the time of redesigning the experimental program. Thus, the decision was made to complete that series. Once that series was completed, there

Table 13. List of factor-level combinations used in making the batches in Series B, C, and D.

1. Series B, 165 batches: Limestone L1, 100 pen. combined with all 33 asphalt gradations and all 5 asphalt percents.

2. Series C, 85 batches: Limestone L1, 60 pen. combined with all 5 asphalt percents and the 17 asphalt gradations[a].

3. Series D, 85 batches: Limestone L2, 60 pen. combined with all 5 asphalt percents and the 17 asphalt gradations[b].

(a)			(b)		
A-F	B-P	C-8	A-P	B-8	C-P
A-4	B-B	C-30	A-I	B-30	C-I
A-8	B-8L	C-30L	A-4L	B-30L	C-8L
A-8L	B-100	C-100L	A-30		C-100
A-100	B-100L		A-30L		C-100L
A-100L			A-4H		
A-4LH			A-8H		
A-30H			A-8LH		
			A-30LH		

was only a limited amount of 100-pen. asphalt available for further use. Thus, all future testing was done with 60-pen. asphalt. Also, only about 165 more batches could be prepared. Since one of the purposes of Part II was to rank the asphalt gradations at each combination of limestone and asphalt grade, Series C and D were redesigned with one half of the gradations prepared with L1 and 60-pen. and the other half of the gradations prepared with L2 and 60-pen. The gradations included in each series were chosen at random. One gradation, C-100L, was included in both series. In all, there were 335 batches prepared.

The decision to include all gradations in either Series C or D was motivated by a desire to maximize the information regarding the ranking of the 33 gradations. Thus, in addition to the comparison among the gradations at L1 and 100 pen. (Series B) all gradations are included in the comparison at either L1 and 60 pen. (Series C) or L2 and

60 pen. (Series D). Thus, the largest number of gradations possible were included in comparing the rankings for a change from L1 to L2 or from 100 pen. to 60 pen. Some information regarding the joint effect of type aggregate and asphalt grade was sacrificed in this design. To gain information about the joint effect of aggregate type and asphalt grade would have required some gradations to have been used in both Series C and Series D. Thus, some gradations would have been only included in Series B and information on the change in the rank of these gradations with changes in aggregate or asphalt grade would be lost.

To gain information regarding the effect of type aggregate and asphalt grade and the interaction on the rankings of the asphalt gradations the following design (Table 14) would be preferred. Of course, this design assumes the existence of sufficient time and material, more than was available for this experiment. In this design, all gradations are included in a comparison at L1 and 100 pen. as well as in a comparison at one other combination of aggregate and grade. The six common gradations used in all four series would allow for a test of interaction.

Table 14. Design of four series for measuring effects and interaction.

Series B	(L1, 100 pen.)	– all 33 gradations (165 batches).
Series C	(L1, 60 pen.)	– 1/3 (11) of the gradations plus 6 gradations chosen at random (85 batches).
Series D	(L2, 60 pen.)	– 11 gradations not included in the 1st group of Series C plus the same 6 as in the 2nd group of Series C (85 batches).
Series E	(L2, 100 pen.)	– remaining 11 gradations not included in C or D plus the same 6 as in C and D (85 batches).

Series F, run separately from the above series, included either crushed gravel or natural gravel as the aggregate. Series F consisted of

(a) 45 batches (crushed gravel, 100-pen. asphalt) — gradations A-I, A-4, A-4L, A-8, A-8L, A-30, A-30L, A-100, A-100L;

(b) 5 batches (natural gravel, 100 pen.).

Analysis of the Data

The analysis of the rankings of the mixes with respect to gradations are reported in Vol. I of this report. Some additional statistical analyses were run and the results are outlined below. These results are summarized for the data from Series B only and are based on only strength (Marshall stability) data observed on the specimen prepared and tested using the normal Marshall test procedure:

(1) An analysis of variance, for testing the significance of aggregate gradation and percent of asphalt, based on the model

$$S_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i = 1, \ldots, 33$$

$$j = 1, \ldots, 5$$

$$\epsilon_{ij} \sim NID(0, \sigma^2) \qquad k = 1, \ldots, n_{ij} \qquad (16)$$

where α is effect of gradation,

β is effect of percent of asphalt, and

$(\alpha\beta)$ is interaction,

was performed. The results are reported in Table 15.

(2) Since the interaction of gradation and asphalt content was significant, the relationship of strength as a function of

Table 15. Analysis of variance for Marshall stability, Series B.

Source of variation	d.f.	Sum of squares	Mean squares	F
Aggregate gradation	32	36459892	1139371.6	22.3
Percent of asphalt	4	42708146	10677036.6	209.3
Interaction	127[a]	63344554	498776.0	9.8
Exp error	162	8293525	51194.6	
Total	325			

[a] There are only 127 d.f. instead of 128 since there were no measurements taken for gradation 32 and 7% asphalt.

percent of asphalt varies among the gradations. The effect of asphalt percent was measured by estimating the functional relationship of strength to percent asphalt for each gradation using the model:

$$S_{ijk} = \mu_i + \beta_{1i} PA_j + \beta_{2i} PA_j^2 + \epsilon_{ijk} \qquad i = 1, \ldots, 33$$

$$j = 1, \ldots, 5 \qquad (17)$$

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \qquad k = 1, \ldots, n_{ij}$$

where S_{ijk} is asphalt concrete strength

μ_i is average (over percent) strength using the ith asphalt gradation ($i = 1, \ldots, 33$)

PA_j is jth value of percent asphalt ($j = 1, 2, 3, 4, 5$); actual percent of asphalt used in the mix was used in the analysis, and

ϵ_{ijk} is experimental error.

The values of the estimated parameters are given in Table 16.

Table 16. Estimates of parameters in Eq. (17).

i	Gradation	$\hat{\mu}_i$	$\hat{\beta}_{1i}$	$\hat{\beta}_{2i}$	Stability, lb Adjusted means	Maximum
1	A-F	5716.74	- 970.83	69.51	2794.2	3440
2	A-P	- 4463.31	3451.89	- 376.66	2809.5	3100
3	A-I	1539.21	1218.72	- 159.73	3470.4	3740
4	A-4	5780.72	- 1064.92	90.56	2866.9	3060
5	A-4L	867.14	1250.64	- 163.80	2849.3	3300
6	A-8	1382.01	1172.32	- 158.17	3294.5	3830
7	A-8L	15165.13	- 4562.56	388.21	2817.2	4130
8	A-30	- 6892.88	5082.70	- 570.39	3425.9	4480
9	A-30L	- 2049.24	2273.03	- 257.08	2586.3	3070
10	A-100	5601.24	- 1283.60	99.31	1880.5	2290
11	A-100L	5183.91	- 873.49	69.45	2720.1	2900
12	A-4H	1115.74	1222.65	- 170.54	2850.2	3350
13	A-4LH	1725.62	881.75	- 124.25	2917.5	3160
14	A-8H	1260.88	1279.06	- 173.31	3181.8	3650
15	A-8LH	565.91	1340.37	- 174.68	2663.8	3150
16	A-30H	1526.90	1104.70	- 175.39	2528.4	4140
17	A-30LH	- 1604.31	2097.52	- 235.79	2682.3	3260
18	B-P	611.08	1520.25	- 209.69	2788.4	3800
19	B-B	- 2508.78	2451.40	- 235.91	3418.3	3850
20	B-8	- 1070.86	1803.47	- 206.33	2550.9	3390
21	B-8L	2292.60	1192.53	- 192.43	3411.7	3860
22	B-30	5833.64	- 418.17	- 24.32	3153.7	4640
23	B-30L	- 7136.88	4470.83	- 462.51	2953.9	3500

Table 16. Continued.

i	Gradation	$\hat{\mu}_i$	$\hat{\beta}_{1i}$	$\hat{\beta}_{2i}$	Stability, lb Adjusted means	Maximum
24	B-100	997.04	1411.67	- 192.57	3137.7	3460
25	B-100L	244.54	1258.99	- 141.02	2849.3	3280
26	C-P	69.24	1606.25	- 197.42	2984.2	3720
27	C-I	7208.43	- 1768.09	186.84	3308.7	3430
28	C-8	- 3188.46	3058.53	- 338.63	3172.7	3440
29	C-8L	4042.50	247.11	- 84.57	3274.2	4060
30	C-30	471.32	1061.33	- 92.28	3289.7	3310
31	C-30L	- 2150.56	2218.39	- 238.13	2686.6	2950
32	C-100	6841.32	- 490.92	- 48.06	2923.5	4450
33	C-100L	2840.22	210.35	- 47.29	2688.3	2900

Also included in Table 16 are the adjusted mean stability as well as the maximum stability for each gradation. The means are the average of all the strength observations for the given gradation after the observation is adjusted for the level of asphalt content. The effect of asphalt content is eliminated, thus leaving an average which measures the "average" strength for the given gradation. The maximum stability values were obtained from plots of stability vs asphalt content.

(3) Using the estimated adjusted means, all pairs of gradations were compared using Duncan's multiple range test[5]. The results, separating the three maximum sizes A, B and C, are summarized in Table 17.

Table 17. Results of Duncan's test[a].

Gradation: Max size 3/4 in.

A-100	A-30H	A-30L	A-8LH	A-30LH	A-100L	A-F	A-P	A-8L	A-4L	A-4H	A-4	A-4LH	A-84	A-8	A-30	A-I
			———	———	———	———	———	———	———	———	———	———		———	———	———
		———	———	———	———	———	———	———								
	———	———	———	———	———											

Gradation: Max size 1/2 in.

B-8	B-P	B-100L	B-30L	B-100	B-30	B-8L	B-B
			———	———	———	———	———
	———	———	———				

Gradation: Max size 3/8 in.

C-30L	C-100L	C-100	C-P	C-8	C-8L	C-30	C-I
———	———			———	———	———	———
			———	———			
		———	———				

[a] The gradations joined by a line underneath them cannot be considered to be different regarding the average strength.

Combining all sizes together and comparing all possible pairs of gradations using Duncan's test, the following groups of decreasing strength (stability) (all gradations within a group cannot be considered as different) were identified:

Table 18. Comparison of all possible pairs of gradations.

Group	Range- and mean-adjusted stability	Gradations within the group
1	(3154-3470) 3295	A-I A-8 A-30 A-8H B-B B-8L B-30 B-100 C-I C-8 C-8L C-30
2	(2867-3173) 3033	A-4 A-4LH A-8H B-30 B-30L B-100 C-P C-8 C-100
3	(2850-3154) 2952	A-4 A-4L A-4H A-4LH B-30 B-30L B-100 B-100L C-P C-100
4	(2664-2984) 2815	A-F A-P A-4 A-4L A-8L A-100L A-4H A-4LH A-8LH A-30LH B-P B-30L B-100L C-P C-30L C-100 C-100L
5	(2528-2924) 2744	A-F A-P A-4 A-4L A-8L A-30L A-100L A-4H A-8LH A-30H A-30LH B-P B-8 B-100L C-30L C-100L
6	(1881) 1881	A-100

From this analysis apparently there is no one gradation or group of gradations which uniquely results in significantly higher stability of asphalt concrete. It does appear, though, that gradation A-100 leads to a significantly lower stability level than all the other gradations included in this experiment. Also, it appears that the gradations in group 1 have a significantly higher stability value than the gradations in group 4. Otherwise, as indicated by the range of stability values in each group, there seems to be overlap between groups. This result could be the consequence of many factors, one of which is that there is indeed no difference in Marshall stability for many gradations. Or, the experimental error may be too large, or the test not powerful enough to "recognize" the true difference.

Although the "adjusted mean strength" of asphalt concrete was used in the comparison among the gradations, this is not the only usable measure. Another strong candidate would be the maximum strength (maximized over asphalt content) but this was used in the analysis described in Vol. 1 and thus was not used here. Using maximum strength, of course, is likely to result in different conclusions. For example, the adjusted mean stability for A-8 is 3294.5, higher than 2817.2 for A-8L; however, the peak or maximum stability for A-8 is 3830, lower than 4130 for A-8L. Similarly, the respective adjusted mean stabilities for C-P and C-I are 2984.2 and 3308.7; but the respective maximum stabilities are 3720 and 3430 (Table 17).

As mentioned earlier, the data from Series F was analyzed separately with emphasis put on the comparison between crushed and natural gravel. Given the gradation included in Series F, some additional comparisons can be made, e.g., a comparison of gradation A-I with some of the gap graded gradations and a comparison of normal gap gradations and "low" gap gradations. These and other comparisons are summarized below.

The results of the analysis for Series F are outlined for the specimen tested using the normal Marshall testing procedure. Initially the linear statistical model

$$S_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i = 1, \ldots, 10$$

$$\epsilon_{ijk} \sim NID(0, \sigma^2) \qquad j = 1, \ldots, 5 \qquad (18)$$

$$k = 1, 2$$

was fit to the data to test for differences due to gradations and percent of asphalt. The analysis of variance is given in Table 19.

Table 19. Analysis of variance for Marshall stability.

Source of variation	d.f.	Sum of squares	Mean squares	F
Gradation	9	10022592.3	1113621.4	25.3
Percent of asphalt	4	1449678.5	362419.6	8.24
Interaction	36	4552581.5	126460.6	2.88
Exp error	50	2198512.5	43970.3	

As was done for the data from Series B, since the interaction of gradation and asphalt content were significant, a separate regression model for stability vs percent asphalt was fit for each gradation. The regression equation is the same as the model in Eq. 17. The estimates of the model coefficients as well as the adjusted means are listed in Table 20.

The comparisons which were considered are:

1. Natural gravel vs gradated crushed gravel.
2. A-P vs gap-graded gradations.
3. Average of A-4 and A-8 vs A-100; a comparison of gaps in larger sizes vs gaps in smaller sizes.
4. A-4 vs A-4L.
5. A-8 vs A-8L.
6. A-30 vs A-30L.
7. A-100 vs A-100L.

The value of the comparisons and the sums of squares associated with each comparison are listed in Table 21.

Table 20. Regression estimate for Series F data.

i	Gradation	μ_i	β_{1i}	β_{2i}	Stability, lb Adjusted means	Stability, lb Max[a]
2	A-P	278.75	994.25	- 112.03	2288.0	2620
3	A-4	1560.39	396.50	- 54.90	2130.0	2160
5	A-4L	1085.23	522.93	- 73.28	1800.0	2140
6	A-8	- 360.06	1091.94	- 117.02	2007.5	2190
7	A-8L	1620.37	276.15	- 43.80	1875.0	2310
8	A-30	- 1016.97	1553.97	- 176.70	2095.0	2280
9	A-30L	1374.17	164.03	- 14.75	1795.0	2310
10	A-100	- 1613.08	1534.36	- 157.33	1860.0	2170
11	A-100L	2974.55	- 510.28	44.82	1630.0	1770
34[b]	Natural	2452.14	- 639.44	68.71	1080.0	1180

[a] From stability vs asphalt content plots.

[b] i = 34 refers to natural gravel.

Table 21. Summary of comparisons for Series F data.

Comparison	Estimate of comparison	Sum of squares	F
1	7760.5	66.92×10^5	15.21
2	3111.5	13.45×10^5	3.06
3	417.5	2.91×10^5	< 1
4	330.0	5.45×10^5	1.2
5	132.5	0.88×10^5	< 1
6	300.0	4.50×10^5	1.0
7	230.0	2.65×10^5	< 1

The column of F values are the values of the F ratio

$$F = \frac{\text{Sum of squares due to comparison}}{\text{Error mean square}} \tag{19}$$

which is useful for testing the significance of that comparison. Clearly, there is a significant difference in strength between asphalt concrete mixed with graded crushed gravel and ungraded natural gravel with the latter being significantly lower. The comparison between gradation A-P and the average of the gap gradings is significant at $\alpha = 0.1$ but not at $\alpha = 0.05$. The remaining comparisons are not significantly different. Again, as in the case of analysis in Series B, there is the question of experimental error and whether the "adjusted mean strength" (rather than the maximum strength) is a good parameter for the comparisons made.

SUMMARY

The main thrust of the statistical analysis conducted in this experiment was in the calibration study and in Part I of the experiment. In the former study, the compaction procedure between the Iowa State University Laboratory and the Iowa Highway Commission Laboratory was calibrated. By an analysis of the errors associated with the measurements we were able to separate the "preparation" and "determination" errors for both laboratories as well as develop the calibration curve which describes the relationship between the compaction procedures at the two labs.

In Part I, the use of a fractional factorial design in a split plot experiment in measuring the effect of several factors on asphalt concrete strength and weight was exhibited. Also, the use of half normal plotting techniques for indicating significant factors and interactions and for estimating errors in experiments with only a limited number of observations was outlined.

The statistical analysis outlined for Part II of this report only represents a small portion of the statistical analyses that could be done on the available data. The major thrust in Part II was on ranking the gradations and observing how the rankings varied as the experimental parameters (type aggregate, asphalt grade) were varied. For this reason only a limited amount of statistical analysis was undertaken.

There was a considerable amount of data accumulated during the course of this experiment. Many measurements, e.g., percent of air voids, flow, unit weight and others, were observed and are recorded on computer cards along with the necessary identification. Thus, this data could be used for a more extensive statistical analysis than project time and resources allowed.

ACKNOWLEDGMENTS

In addition to the many sources of assistance acknowledged in Vol. 1 of this report, we would like to especially acknowledge the assistance of the Numerical Analysis section of the Statistical Laboratory at Iowa State University for helping with much of the data punching and running of computer programs, and the Engineering Research Institute, for their support.

REFERENCES

Lee, D. Y., "Evaluation of Gap-Graded Asphalt Concrete Mixtures," Progress Report No. 1, Engineering Research Institute, Iowa State University (1971).

Grubbs, F. E., "Sample Criteria for Testing Outlying Observations," Annals of Mathematical Statistics, 21, 27-58 (1950).

Goetz, W. H., "Comparison of Triaxial and Marshall Test Results," Proc. AAPT, 20, 200 (1951).

Lees, G., "The Influence of Binder and Aggregate in Bituminous Mixtures," J. Inst. of Highway Engineers, 16: 1, 7 (1969).

Duncan, D. B., "Multiple Range and Multiple F-Tests," Biometrics, 11, 1-42 (1955).