

# Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings

J. Lieblein and M. Zelen

Fatigue is an important factor in determining the service life of ball bearings. Bearing manufacturers are therefore constantly engaged in fatigue-testing operations in order to obtain information relating fatigue life to load and other factors. Several of the larger manufacturers have recently pooled their test data in a cooperative effort to set up uniform and standardized ball-bearing application formulas, which would benefit the many users of anti-friction bearings. These data were compiled by the American Standards Association, which subsequently requested that the National Bureau of Standards perform the necessary analyses. This paper summarizes the principal results of the analyses undertaken by the Bureau, and describes the statistical procedures used in the investigation.

## 1. Introduction

### 1.1. Statement of Problem

The experience of ball-bearing manufacturers over many years has led to the acceptance of an equation of the form [15, p. 15, eq (53)]<sup>1</sup>

$$L = (C/P)^p, \quad (1)$$

relating fatigue life  $L$  to load  $P$  when other factors are kept constant. In the above equation,  $C$  is termed the "basic (dynamic) capacity," and is defined [15, p. 48] as the constant bearing load (in pounds) that 90 percent of a group of similar bearings can endure for one million revolutions under the given running conditions.

The quantity  $C$  in eq (1) depends upon the characteristics of the bearing type, as indicated in [15, p. 32, eq (120)]. When the expression cited is substituted in eq (1), the *fatigue-life formula for ball bearings* takes the form

$$L = \left[ \frac{f_c Z^{a_1} D_w^{a_2} (i \cos \alpha)^{a_3}}{P} \right]^p. \quad (2)$$

The symbols are defined as follows:

$Z$  = number of balls.

$D_w$  = ball diameter in inches.

$i$  = number of rows.

$\alpha$  = contact angle.

$P$  = bearing load in pounds.

$L$  = number of million revolutions that a specified percentage of bearings will fail to survive on account of fatigue causes. If the percentage is 10, then  $L = L_{10}$ , and is termed the *rating life*; if the percentage is 50, then  $L = L_{50}$ , the *median life*.

$p, a_1, a_2, a_3, f_c$  are taken as unknown parameters whose values have to be estimated from given data.

Since  $i = 1$  and  $\alpha = 0^\circ$  for *deep-groove* ball bearings, with which this paper is exclusively concerned, the life equation that will henceforth be considered takes the form

$$L = \left[ \frac{f_c Z^{a_1} D_w^{a_2}}{P} \right]^p. \quad (2a)$$

<sup>1</sup> Figures in brackets indicate the literature references at the end of this paper.

The main goal of this investigation was to determine the "best values" of the unknown parameters in the life equation from the experimental data. One of the major problems was to determine the value of the exponent  $p$ , as there was disagreement within the ball-bearing industry whether an appropriate value for  $p$  was 3, 4, or some other value.

### 1.2. Description of Data

The data available for analysis consisted of sets of records summarizing endurance tests for deep-groove ball bearings. These tests were carried out over a period of years by four major ball-bearing companies. In the interest of trade anonymity, these companies will henceforth be designated by A, B, C, and D. Each endurance test consisted of a number of bearings of the same type (the number varying from test to test), which were tested simultaneously under the same load and running conditions. Table 1 summarizes the number of test groups of data for each company. The data from company B were sufficiently extensive to permit a further breakdown into three bearing types, here denoted by B-1, B-2, and B-3.

TABLE 1. Summary of ball-bearing data

Company	Number of test groups	Total number of bearings in test groups
A.....	50	1,250
B.....	148	3,280
Type B-1.....	37	.....
Type B-2.....	94	.....
Type B-3.....	17	.....
C.....	12	291
D.....	3	109
Total (all companies).....	213	4,948

The worksheets, summarizing the tests, recorded the number of millions of revolutions reached by each bearing in the test group before fatigue failure. Information was also given for those tests terminated before all bearings in the test group failed. In addition to the test results, the worksheets included information on the characteristics of the bearing type (e. g., values for  $Z$ ,  $D_a$ ,  $i$ ,  $\alpha$ ) and load  $P$ , as well as other items of descriptive and identifying information. A specimen worksheet is reproduced in appendix A.

All necessary quantities for evaluating the unknown parameters in the life equation (2) were given directly on the worksheets except the fatigue life  $L$ .<sup>2</sup> This quantity can be estimated from the observed fatigue lives of individual bearings within a test group. As already noted, two concepts of fatigue life are used for  $L$ , namely, the *rating life*  $L_{10}$ , and the *median life*  $L_{50}$ . Separate analyses have been carried out with regard to each throughout.

Appendix A summarizes the data taken from the original worksheets that were used in the statistical analysis. Also given are the computed values for  $L_{10}$ ,  $L_{50}$ , and the "Weibull slope"  $e$  (which relates to the dispersion of fatigue lives). The methods for obtaining these quantities from the bearing data are given in detail in appendix B.

### 1.3. Assumptions for the Statistical Analyses

All conclusions reached in this report, and all statistical analyses employed, are based upon the following principal assumptions:

(a) The life formula (2) is the proper functional form for describing fatigue life in ball bearings.

(b) Differences in the measured life of bearings classed as identical, tested at the same load, reflect only the inherent variability of fatigue life, and are free from systematic errors that may arise from different test conditions, materials, manufacturing methods, etc.

<sup>2</sup> Certain estimates for  $L_{10}$  and  $L_{50}$  had been entered on the worksheets for many of the tests. However, these were not regarded as part of the data submitted for analysis.

(c) All the bearings in a test group can be regarded as a random sample from a homogeneous population of ball bearings.

(d) The probability distribution of the number of revolutions to fatigue failure is of the same form for each test group, although its parameters may differ from group to group.

(e) This fatigue-life distribution is of the type known as the "Weibull distribution."

The purely statistical assumptions, (c) to (e), served as the basis for the determination of  $L_{10}$ ,  $L_{50}$ , and  $e$  for each test group. Assumption (e), however, is not involved in the methods used to evaluate the parameters in the life formula (2) from given values of  $L_{10}$  or  $L_{50}$ . A different assumed form for the distribution of fatigue life might give somewhat different values for  $L_{10}$  and  $L_{50}$ , but the same methods could then be used to evaluate the unknown parameters in the life formula (2).

Other assumptions of a more technical nature were necessary in the course of the analyses. These are discussed in appendixes B and C.

As in all cases where inferences are made from given data, the conclusions reached here pertain only to the population from which the given data can be regarded as constituting a random sample.

## 2. Outline of Statistical Analyses

The statistical analyses were divided into two phases. The first phase considered the problem of finding estimates of  $L_{10}$ ,  $L_{50}$ , and the Weibull slope  $e$  from the given test data; the second phase used these estimates of  $L_{10}$  and  $L_{50}$  to evaluate the unknown values of the parameters in the life formula.

### 2.1. Estimation of $L_{10}$ and $L_{50}$

The quantity  $L$  depends upon the existence of an underlying probability distribution of bearing lives. Selection of a distribution or population is equivalent to specifying the probability that a bearing selected at random from such a population will survive any given number of revolutions,  $L$ , or, conversely, that if  $c$  is a specified probability, then  $L$  is the life period that will be survived with this probability, e. g.,

$$\text{Probability \{life} \geq L\} = c = \begin{cases} .90 \text{ for } L = L_{10}. \\ .50 \text{ for } L = L_{50}. \end{cases}$$

Accordingly, any  $L$ , such as  $L_{10}$  or  $L_{50}$ , must be obtained by estimating a characteristic of the assumed distribution. For reasons described in appendix B, the distribution characterizing ball-bearing fatigue life was taken to be the Weibull distribution. In brief, this distribution can be derived by assuming a "weakest-link" concept of fatigue strength. In addition, the suitability of the Weibull distribution for fatigue life has been verified in many cases by empirical plotting of data.

One method of estimating  $L_{10}$  or  $L_{50}$  makes use of special probability-plotting paper so designed that a theoretical Weibull distribution plots as a straight line, and treats the problem as one of straight-line fitting by conventional least squares procedures. However, the procedure usually followed does not take into full account the number of bearings that remain intact when tests are incomplete, nor the interdependence of successive points. Because of these and other limitations, it was decided to use an alternative approach in the estimation of  $L_{10}$  and  $L_{50}$  for each test group (see appendix B).

To this end, a method was developed that takes into account explicitly the number of bearings remaining intact at the termination of a test, and that also possesses several other advantages. This method makes use of certain specially determined linear functions of the observed failure times (in logarithms),  $x_i$ , arranged in order of size. These functions have the general form

$$T = \sum_{j=1}^k c_j x_j \quad (3)$$

As the method makes intimate use of the ordered arrangement of the data, it is termed an "order statistics" method.

The coefficients  $c_j$  in eq (3) allow great flexibility. They have been determined in such a manner that the method will have certain desirable objective characteristics, e. g., freedom from systematic error and a minimum standard error.

## 2.2. Evaluation of the Parameters in the Life Formula

Once the estimates for  $L_{10}$  and  $L_{90}$  are obtained, it is possible to evaluate the exponent  $p$  in the life formula. However, in order to make the most efficient use of the given data, it is necessary also to estimate the other parameters,  $f_c$ ,  $a_1$ , and  $a_2$ .

The methods for estimating the values of  $L_{10}$  and  $L_{90}$  for each test group actually yield results for  $\ln L_{10}$  and  $\ln L_{90}$ . Thus, taking logarithms<sup>3</sup> of the life equation (2a) gives

$$\ln L = (p \ln f_c) + (pa_1) \ln Z + (pa_2) \ln D_a - p \ln P. \quad (4)$$

This equation can be written more simply as

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3, \quad (5)$$

where

$$\left. \begin{aligned} Y &= \ln L \text{ (for either } L_{10} \text{ or } L_{90}), \\ b_0 &= p \ln f_c, \quad b_1 = pa_1, \quad b_2 = pa_2, \quad b_3 = -p, \\ x_1 &= \ln Z, \quad x_2 = \ln D_a, \quad x_3 = \ln P. \end{aligned} \right\} \quad (6)$$

The quantities  $x_1$ ,  $x_2$ , and  $x_3$  depend on the characteristics of the bearing type and test conditions, and can be regarded as known exactly. On the other hand, the variable  $Y$ , which depends on the outcome of the bearing tests, is subject to considerable dispersion. Thus, estimates can be found for the parameters  $b_0$ ,  $b_1$ ,  $b_2$ , and  $b_3$ , using standard least squares methods based on minimizing the sums of squared deviations in the  $y$  direction. These methods are discussed in detail in appendix C.

After the parameters  $b_0$ ,  $b_1$ ,  $b_2$ , and  $b_3$  are estimated, values for  $a_0$ ,  $a_1$ ,  $a_2$ , and  $p$  can be found from the relations

$$\begin{aligned} a_0 &= \ln f_c = -\frac{b_0}{b_3} & a_1 &= -\frac{b_1}{b_3} \\ a_2 &= -\frac{b_2}{b_3} & p &= -b_3 \end{aligned}$$

It is clear that the values for  $a_0$ ,  $a_1$ , and  $a_2$  depend on the value of  $p$ .

The estimates for  $p$  and the  $a$ 's are subject to some uncertainties because they are based on test results, which themselves are subject to considerable variability. Hence with every value of  $p$  and of the  $a$ 's calculated from the life data, there is given also an interval of uncertainty to indicate its precision. These intervals are "95-percent confidence limits."<sup>4</sup>

A large interval of uncertainty associated with an estimate indicates poor precision; a small interval of uncertainty is evidence of high precision. These intervals of uncertainty not only reflect the inherent variability of the test data, but are also affected by (a) how well the life equation (2a) is the proper functional form for bearing life, and (b) the suitability of the data (including the number of test groups) for estimating the parameters in the life formula.

Further technical details concerning the evaluation of the parameters in the life formula are given in appendix C.

<sup>3</sup> Natural logarithms to the base  $e$  are used throughout.

<sup>4</sup> Briefly, confidence intervals describe the compatibility of the observations with an unknown parameter estimated from them; 95-percent confidence limits are limits such that on the average, in repeated applications of the same procedure, 95 percent of intervals so calculated will contain the unknown true value of the parameter. The confidence limits associated with  $p$  are symmetric. However, the confidence limits associated with the  $a$ 's are asymmetric because of the dependence of the  $a$ 's on  $p$ .

### 3. Summary of Analyses

#### 3.1. Evaluation of Parameter $p$

The statistical analysis based on all deep-groove ball-bearing data from companies A, B, and C<sup>6</sup> yielded the final values for  $p$  shown in table 2. The separate values for each of the three companies are given in table 3. The intervals of uncertainty specified by  $\pm$  quantities in those tables refer to intervals within which, with reasonable assurance, the true value of the parameter is located. The fact that all of the intervals of uncertainty exhibit considerable overlap shows that the data are consistent with the supposition that all three companies have a common value of  $p$  for deep-groove bearings. The fact that all the intervals include 3 indicates that all the estimates of  $p$  are consistent with the practice of taking  $p=3$ . Moreover, the value of  $p$  for  $L_{10}$  was not significantly different from that for  $L_{50}$ .

TABLE 2. Final over-all values of  $p$  for deep-groove bearings

	$L_{10}$	$L_{50}$
$p$ .....	$2.87 \pm 0.35$	$2.80 \pm 0.31$

TABLE 3. Individual estimates of  $p$  for deep-groove bearings by company

Company	Number of test groups	$L_{10}$	$L_{50}$
A.....	50	$3.00 \pm 0.64$	$3.05 \pm 0.60$
B.....	148	$2.75 \pm .48$	$2.62 \pm 0.40$
C.....	12	$3.12 \pm .88$	$2.88 \pm 1.02$

The values given for  $p$  are based on analyses of all deep-groove ball-bearing data, irrespective of bearing type. Hence, the parameter estimates represent "omnibus" values. In order to investigate the dependence of the exponent  $p$  on bearing type, the data from company B, which was made up of three bearing types, were analyzed separately. The results for the exponent  $p$  are shown in table 4. These results are all compatible with the value  $p=3$ .

TABLE 4. Value of  $p$  by bearing type for company B

Type	Number of test groups	Value of $p$	
		$L_{10}$	$L_{50}$
B-1.....	37	$3.36 \pm 0.68$	$3.23 \pm 0.47$
B-2.....	94	$2.65 \pm 0.91$	$2.13 \pm 0.79$
B-3.....	17	$1.89 \pm 1.28$	$2.82 \pm 1.10$
Total.....	148		

<sup>6</sup>The data furnished by company D were too few to be included in the analysis.

### 3.2. Evaluation of Parameters $f_c$ , $a_1$ , and $a_2$

The computations that give estimates for the exponent  $p$  also yield estimates for the quantities  $\ln f_c$ ,  $a_1$ , and  $a_2$ . From the relations (6) it is clear that the values for these parameters depend on the value for  $p$ . Thus, associated with every value of  $p$  will be corresponding values for  $\ln f_c$ ,  $a_1$ , and  $a_2$ . Table 5 summarizes these parameter estimates associated with the final values of  $p$ . The estimates for  $a_0 = \ln f_c$ , rather than  $f_c$ , are given here, because this is the parameter that arises naturally in the life formula (cf. eq 4).

The analyses conducted separately for each company resulted in other values than those in the previous paragraph for  $a_0$ ,  $a_1$ ,  $a_2$ . These results are summarized in table 6. They show excellent agreement with the results in table 5, even though the values for  $p$  are somewhat different.

TABLE 5. Final values of  $a_0$ ,  $a_1$ ,  $a_2$  for  $L_{10}$  and  $L_{50}$

Company	$p$	$a_0$	Interval of uncertainty	$a_1$	Interval of uncertainty	$a_2$	Interval of uncertainty
$L_{10}$							
A.....	2.87	9.02	(7.31, 10.79)	0.380	(-0.454, 1.201)	1.72	(1.51, 1.92)
B.....	2.87	8.55	(7.98, 9.14)	.670	(0.418, 0.920)	1.81	(1.70, 1.92)
C.....	2.87	9.56	(8.85, 12.42)	-.174	(-1.750, 1.352)	1.37	(0.09, 2.67)
$L_{50}$							
A.....	2.80	10.36	(8.81, 11.98)	0.015	(-0.741, 0.751)	1.89	(1.50, 1.88)
B.....	2.80	9.05	(8.54, 9.60)	.695	(.470, .920)	1.91	(1.81, 2.01)
C.....	2.80	9.05	(6.61, 11.58)	.475	(-.921, 1.847)	1.76	(0.60, 2.93)

TABLE 6. Values of  $a_0$ ,  $a_1$ ,  $a_2$  for  $L_{10}$  and  $L_{50}$ , based on independent analyses for each company

Company	$p$	$a_0$	Interval of uncertainty	$a_1$	Interval of uncertainty	$a_2$	Interval of uncertainty
$L_{10}$							
A.....	3.00	8.97	(7.18, 10.90)	0.390	(-0.507, 1.249)	1.73	(1.50, 1.94)
B.....	2.75	8.59	(7.99, 9.24)	.666	(.398, 0.928)	1.80	(1.67, 1.92)
C.....	3.12	9.21	(7.29, 11.84)	-.041	(-1.326, .992)	1.36	(0.49, 2.30)
$L_{50}$							
A.....	3.05	10.13	(8.48, 12.00)	0.072	(-0.768, 0.855)	1.71	(1.50, 1.91)
B.....	2.62	9.15	(8.61, 9.76)	.690	(.456, .922)	1.90	(1.79, 2.00)
C.....	2.88	8.93	(6.58, 12.39)	.510	(-1.055, 1.810)	1.75	(0.66, 3.05)

Similarly, the values for  $a_0$ ,  $a_1$ , and  $a_2$ , arising from separate analyses made on the three types of bearings from company B, resulted in still other estimates for these parameters. Table 7 summarizes these estimates. These estimates are less precise than the corresponding omnibus values given for company B in table 6. This is a consequence of the fact that within a bearing type, the quantities  $Z$  and  $D_0$  hardly vary at all. This condition makes the data unsuitable for estimating the associated unknown parameters,  $a_0$ ,  $a_1$ , and  $a_2$ .

TABLE 7. Values of  $a_0$ ,  $a_1$ ,  $a_2$  for  $L_{10}$  and  $L_{30}$  by bearing type (company B)

Type	$a_0$	Interval of uncertainty	$a_1$	Interval of uncertainty	$a_2$	Interval of uncertainty
$L_{10}$						
B-1.....	7.25	( 5.21, 9.39)	1.07	( 0.35, 1.76)	1.68	(1.27, 2.08)
B-2.....	7.34	( 5.54, 9.33)	1.21	( .43, 2.00)	1.69	(1.98, 1.93)
B-3.....	2.50	(-7.96, 16.04)	3.70	(-2.25, 9.08)	1.27	(0.90, 1.65)
$L_{30}$						
B-1.....	7.39	( 5.92, 8.92)	1.23	( 0.73, 1.73)	1.79	(1.50, 2.08)
B-2.....	9.00	( 7.10, 11.67)	0.87	(-.05, 1.68)	1.77	(1.46, 2.03)
B-3.....	1.03	(-4.34, 6.53)	4.60	( 1.97, 7.19)	1.48	(1.21, 1.70)

### 3.3. Redetermination of the Estimates for $f_c$

The uncertainty intervals associated with estimates for the parameter  $a_0 = \ln f_c$  are quite large. This is primarily because the uncertainty associated with the estimate of  $a_0$  also depends on how well the other parameters,  $a_1$ ,  $a_2$ , and  $p$ , are estimated. Another way to evaluate  $a_0$ , which may result in smaller intervals of uncertainty, is to assume a priori values for  $a_1$ ,  $a_2$ , and  $p$ , and then determine the estimate for  $a_0$ . This procedure was followed by using the widely accepted values for the parameters given in [15], namely,  $a_1 = 2/3$ ,  $a_2 = 1.8$ ,  $p = 3$ .

However, if on such a calculation the values assumed for the parameters  $a_1$ ,  $a_2$ , and  $p$  are not compatible with the given data, then values of  $a_0$  (or  $f_c$ ) so calculated will not be correct determinations for these data. Accordingly, an analysis was made to determine whether the parameter values in [15] were compatible with the given data.

This analysis showed that these parameter values are compatible with the data, with respect to all individual companies for rating life  $L_{10}$ , but not for median life  $L_{30}$ . (Company A was the only company for which the parameter values are suitable for median life.) A further analysis, by bearing type for company B, showed that the above parameter values are not suitable for the rating life  $L_{10}$  with respect to B-3-type bearings.

In the light of this last analysis, redetermined values of  $a_0$ , taking  $a_1 = 2/3$ ,  $a_2 = 1.8$ , and  $p = 3$ , are only strictly valid with respect to company A, company B (B-1, B-2), and company C for rating life  $L_{10}$ . These values are summarized in table 8. For convenience, these new estimates are given for  $f_c = \ln^{-1} a_0 = \exp a_0$ .

TABLE 8. Values for  $f_c$  assuming  $a_1 = 2/3$ ,  $a_2 = 1.8$ ,  $p = 3.0$  for  $L_{10}$

Company	Number of test groups	$f_c$	Interval of uncertainty
A.....	50	4,538	(4,273, 4,817)
B (over-all value).....	148	4,925	(4,750, 5,105)
B-1.....	37	4,709	(4,403, 5,034)
B-2.....	94	5,033	(4,885, 5,187)
B-3*.....	17	.....	.....
C.....	12	3,294	(3,029, 3,583)
D.....	3	4,839	(3,478, 6,187)

\*Assumed values of parameters  $a_1$ ,  $a_2$ , and  $p$  not compatible with test results for bearings of this series.

The authors express their gratitude for the assistance rendered by the various staff members of the Bureau, without which the prosecution of this study and realization of its goal would not have been possible. Particular thanks go to Churchill Eisenhart, Chief of the Statistical Engineering Laboratory, for his many valuable suggestions and constructive criticisms which have been incorporated into this report. Thanks are also due to Joseph M. Cameron for many useful suggestions and for his liaison assistance with the work of the Computation Laboratory. Most of the large-scale calculations were performed on the Bureau's electronic computer (SEAC), and for these the authors are deeply indebted to the following members of the Computation Laboratory: I. Stegun, for her general supervision; A. Futterman, for her painstaking efforts in organizing and seeing the SEAC computations through to a successful conclusion; also to R. Capuano and R. Zucker, as well as others in the hand-computing department, for the large amount of hand computations. Our appreciation is also expressed to the following personnel of the Statistical Engineering Laboratory: M. Carson and M. L. Epling for their excellent computational work, and L. S. Deming, M. E. McKinley, C. Yick, and L. Hamilton for their efforts in connection with the various tables and charts.

The authors also express their thanks to the members of the American Standards Association Committee B-3 (Ball and Roller Bearings), Subcommittee 7, for their invaluable cooperation during the course of this study.

#### 4. Appendix A. Summary of Original Data

This appendix summarizes in tabular form the worksheets submitted by the American Standards Association Subcommittee to the National Bureau of Standards for statistical analysis. Separate tables are presented for deep-groove data from companies A, B, C, and D. These four tables (A-1 to A-4) are followed by table A-5, which gives a synopsis of the number of test groups and the number of bearings for each company.

Tables A-1 to A-4 give the size of test group, the values for quantities  $P$ ,  $Z$ ,  $D_n$ , and the estimates<sup>6</sup> for  $L_{10}$ ,  $L_{50}$ , and the "Weibull slope"  $e$ . All of these variables are directly observed or specified quantities except for the estimates  $L_{10}$ ,  $L_{50}$ , and  $e$ . These last three quantities are based on statistical calculations that made use of the results of individual endurance tests. These calculations are explained in appendix B.

The original data, as submitted, contained a few cases where companies tested bearings manufactured by other companies. Such test groups are not included in the summary tables, as these results confound differences in testing with differences in manufacturing. Therefore these test results were not used in any of the analyses. Thus, table A-3, for company C, omits 4 tests performed on other manufacturers' bearings; table A-4, for company D, omits 3 tests.

The five tables described above are followed by a specimen worksheet<sup>7</sup> with identifying information removed. A sample of Weibull-function coordinate paper is also included. This coordinate paper had been used for graphing the results of all the individual endurance tests and these graphs had accompanied the worksheets submitted to the Statistical Engineering Laboratory.

<sup>6</sup> The estimates for  $L_{10}$  and  $L_{50}$  are given in millions of revolutions for all companies except company D. The life estimates for D are shown in hours, the same units in which the original endurance data were given.

<sup>7</sup> Bearings marked "Omitted" were completely eliminated from consideration, as company representatives explained that these were non-fatigue failures and should not be regarded as part of the test group. As a result, the test group in the case of the specimen sheet shown was taken to consist of 23 bearings rather than the original number of 25. This type of situation appeared rather infrequently, however.



TABLE A-1. Summary ball-bearing data for company A, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$ 

Record No.	Year of test	Number in test group	Load	Z Number of balls	$D_a$ Ball diam.	$L_{10}$	$L_{50}$	Weibull slope $e$
			1b		in.			
1-1	1936	24	4240	8	11/16	19.2	84.5	1.27
1-2	1937	20	4240	8	11/16	26.2	74.2	1.81
1-3	1937	14	4240	8	11/16	11.1	68.1	1.04
1-4	1937	19	4240	8	11/16	11.8	66.8	1.09
1-5	1937	18	4240	8	11/16	13.5	79.4	1.06
1-6	1938	21	2530	9	1/2	5.80	25.7	1.27
1-7	1938	28	4240	8	11/16	18.3	44.7	2.10
1-8	1938	27	4240	8	11/16	5.62	73.2	.73
1-9	1940	20	4240	8	11/16	15.8	82.7	1.14
1-10	1940	22	4240	8	11/16	8.70	41.6	1.20
1-11	1940	19	4240	8	11/16	11.6	160	.72
1-12	1940	15	1940	9	7/16	20.6	71.4	1.52
1-13	1940	15	1940	9	7/16	14.5	88.2	1.04
1-14	1940	15	2536	9	1/2	12.1	33.1	1.87
1-15	1940	14	2536	9	1/2	15.1	46.4	1.67
1-16	1940	15	2536	9	1/2	14.0	43.6	1.66
1-17	1940	14	2536	9	1/2	19.3	51.8	1.91
1-18	1940	26	4240	8	11/16	46.2	110	2.17
1-19	1940	14	4240	8	11/16	30.0	88.2	1.74
1-20	1942	20	4240	8	11/16	21.1	57.4	1.89
1-21	1942	20	4240	8	11/16	17.3	45.7	1.94
1-22	1942	37	4240	8	11/16	37.5	118	1.64
1-23	1942	36	4240	8	11/16	20.3	77.1	1.41
1-24	1942	32	4240	8	11/16	4.03	42.5	.80
1-25	1941	28	2544	8	17/32	8.38	84.7	.81
1-26	1943	23	3975	8	19/32	1.79	13.5	.93
1-27	1942	30	4400	10	5/8	11.7	45.1	1.39
1-28	1942	31	6920	8	7/8	4.15	15.8	1.41
1-29	1943	30	990	9	5/16	7.23	41.0	1.09
1-30	1943	30	1509	7	7/16	22.9	110	1.20
1-31	1943	30	932	7	11/32	9.54	31.6	1.57
1-32	1944	26	3180	8	19/32	6.28	23.0	1.45
1-33	1944	29	3180	8	19/32	4.81	21.2	1.27
1-34	1944	33	8640	10	7/8	4.17	12.8	1.68
1-35	1944	26	14080	8	1-1/4	5.42	31.6	1.07
1-36	1951	28	1940	9	7/16	7.47	49.5	1.00
1-37	1951	34	2330	9	7/16	4.80	21.3	1.26
1-38	1951	27	1550	9	7/16	14.8	78.4	1.13
1-39	1951	29	1165	9	7/16	84.9	460	1.11
1-40	1951	27	2910	9	7/16	3.40	16.5	1.19
1-41	1951	27	3880	9	7/16	1.24	3.23	1.97
1-42	1951	26	776	9	7/16	241	951	1.37
1-43	1951	30	19750	8	1-3/4	3.01	12.6	1.31
1-44		30	2112	8	11/16	89.1	486	1.11
1-45		30	4224	8	11/16	15.2	104	.98
1-46		30	8448	8	11/16	2.04	10.2	1.17
1-47		30	2112	8	5/8	51.0	376	.94
1-48		30	4224	8	5/8	5.26	58.8	.78
1-49		30	8448	8	5/8	.883	4.94	1.09
1-50	1944	30	4224	8	11/16	14.8	57.4	1.39

TABLE A-2. Summary ball-bearing data for company B, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$ 

Record No.	Year of test	Number in test group	Load lb	Z Number of balls	$D_n$ Ball diam. in.	$L_{10}$	$L_{50}$	$e$ Weibull slope
2-1	1940	19	570	10	3/16	6.68	13.4	2.72
2-2	1944	20	570	9	5/16	29.8	70.0	2.22
2-3	1946	23	580	9	1/4	16.3	55.1	1.55
2-4	1946	23	580	9	1/4	28.5	69.2	2.13
2-5	1947	23	580	9	1/4	16.4	49.3	1.71
2-6	1943	10	665	9	1/4	10.3	40.1	1.50
2-7	1944	10	665	9	1/4	25.7	46.4	3.19
2-8	1942	19	580	10	1/4	9.55	39.6	1.32
2-9	1946	33	620	10	1/4	17.9	62.1	1.51
2-10	1947	15	620	10	1/4	19.9	73.2	1.45
2-11	1947	31	620	10	1/4	12.9	50.4	1.39
2-12	1944	19	625	10	1/4	19.3	46.2	2.18
2-13	1941	17	720	10	1/4	11.1	23.3	2.54
2-14	1946	60	980	11	9/32	15.7	43.5	1.85
2-15	1947	32	980	11	9/32	11.2	38.1	1.54
2-16	1950	49	600	11	5/16	417	809	2.85
2-17	1949	60	600	11	5/16	216	709	1.58
2-18	1943	20	900	11	5/16	35.6	100	1.82
2-19	1946	67	1220	11	5/16	12.0	42.2	1.50
2-20	1947	34	1220	11	5/16	8.53	46.6	1.11
2-21	1940	20	1370	11	5/16	6.77	18.9	1.85
2-22	1950	60	1425	11	5/16	13.5	46.5	1.53
2-23	1950	60	2243	11	5/16	2.32	8.06	1.51
2-24	1942	20	720	12	5/16	36.7	141	1.40
2-25	1946	55	1300	12	5/16	19.0	57.2	1.71
2-26	1947	30	1300	12	5/16	19.5	60.6	1.67
2-27	1944	20	1650	14	11/32	17.0	74.4	1.37
2-28	1946	59	1760	14	11/32	20.9	53.7	2.00
2-29	1947	34	1760	14	11/32	9.56	40.7	1.30
2-30	1940	20	2010	13	13/32	5.49	33.3	1.05
2-31	1940	9	2010	13	13/32	1.39	44.0	.54
2-32	1944	19	2140	13	13/32	9.80	82.7	.88
2-33	1943	11	2630	15	13/32	5.19	54.9	.80
2-34	1942	12	5900	14	19/32	6.36	17.5	1.86
2-35	1947	19	5900	14	19/32	3.68	22.1	1.05
2-36	1947	20	8070	14	23/32	8.34	23.6	1.81
2-37	1942	12	8075	14	23/32	6.78	36.4	1.12
2-38	1938	10	565	9	.210	9.27	18.4	2.75
2-39	1940	23	720	8	9/32	18.2	56.9	1.66
2-40	1940	24	720	8	9/32	22.8	56.2	2.09
2-41	1940	25	720	8	9/32	3.99	15.6	1.38
2-42	1940	21	720	8	9/32	9.07	29.4	1.60
2-43	1940	25	720	8	9/32	7.14	28.5	1.36
2-44	1940	25	720	8	9/32	12.5	26.4	2.51
2-45	1941	25	720	8	9/32	18.8	48.7	1.98
2-46	1941	23	720	8	9/32	21.5	53.2	2.08
2-47	1947	33	860	8	5/16	17.1	59.0	1.52
2-48	1948	8	860	8	5/16	15.2	87.6	1.08
2-49	1943	20	900	9	5/16	30.1	92.3	1.68
2-50	1944	18	900	9	5/16	15.0	47.6	1.63

TABLE A-2. Summary ball-bearing data for company B, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$ —Continued

Record No.	Year of test	Number in test group	Load	2 Number of balls	$D_m$ Ball diam.	$L_{10}$	$L_{50}$	$e$ Weibull slope
			1b		in.			
2-51	1945	27	940	9	5/16	17.5	52.8	1.71
2-52	1947	34	940	9	5/16	14.4	65.6	1.24
2-53	1938	10	1180	9	5/16	8.76	22.1	2.04
2-54	1945	30	1580	9	3/8	12.1	43.3	1.47
2-55	1947	33	1580	9	3/8	17.2	64.6	1.42
2-56	1948	8	1580	9	3/8	10.7	34.6	1.61
2-57	1945	31	2160	9	7/16	10.9	37.6	1.52
2-58	1947	30	2160	9	7/16	12.7	53.7	1.30
2-59	1938	9	2200	9	7/16	3.73	43.5	.77
2-60	1947	30	2480	9	7/16	16.6	78.3	1.21
2-61	1950	40	1340	9	15/32	180	275	4.44
2-62	1937	19	1660	10	7/16	85.2	234	1.86
2-63	1941	19	1700	9	15/32	57.1	230	1.35
2-64	1939	24	2480	9	15/32	15.7	55.8	1.48
2-65	1939	25	2480	9	15/32	27.1	97.8	1.47
2-66	1939	23	2480	9	15/32	21.7	122	1.09
2-67	1939	28	2480	9	15/32	13.2	42.3	1.62
2-68	1939	28	2480	9	15/32	35.8	145	1.35
2-69	1939	20	2480	9	15/32	12.7	34.7	1.87
2-70	1944	20	2480	9	15/32	10.1	27.8	1.87
2-71	1945	20	2480	9	15/32	8.83	34.3	1.39
2-72	1938	10	2480	9	15/32	16.5	60.3	1.45
2-73	1942	11	2480	9	15/32	17.9	65.8	1.45
2-74	1943	10	2480	9	15/32	15.7	63.1	1.35
2-75	1943	20	2480	9	15/32	10.8	42.1	1.38
2-76	1944	18	2480	9	15/32	14.2	39.9	1.83
2-77	1944	18	2480	9	15/32	19.0	67.8	1.48
2-78	1944	18	2480	9	15/32	16.3	57.7	1.49
2-79	1944	20	2480	9	15/32	2.93	18.0	1.04
2-80	1944	20	2480	9	15/32	5.69	25.4	1.26
2-81	1944	28	2480	9	15/32	9.54	39.9	1.32
2-82	1944	22	2480	9	15/32	12.6	55.7	1.27
2-83	1944	23	2480	9	15/32	5.10	37.5	.94
2-84	1944	18	2480	9	15/32	16.0	53.7	1.56
2-85	1944	20	2480	9	15/32	1.98	22.1	.78
2-86	1945	20	2480	9	15/32	5.65	28.8	1.16
2-87	1945	20	2480	9	15/32	12.8	43.6	1.58
2-88	1945	20	2480	9	15/32	9.84	32.3	1.59
2-89	1945	20	2480	9	15/32	12.1	43.0	1.48
2-90	1945	20	2480	9	15/32	5.48	40.8	.94
2-91	1945	20	2480	9	15/32	6.64	25.3	1.41
2-92	1945	32	2480	9	15/32	13.9	41.9	1.70
2-93	1946	35	2480	9	15/32	9.02	45.4	1.17
2-94	1946	34	2480	9	15/32	11.0	49.2	1.26
2-95	1947	31	2480	9	15/32	14.5	73.6	1.16
2-96	1944	9	2480	9	15/32	5.91	37.2	1.02
2-97	1944	10	2480	9	15/32	18.1	40.5	2.33
2-98	1945	10	2480	9	15/32	17.1	53.3	1.65
2-99	1945	10	2480	9	15/32	32.6	61.8	2.95
2-100	1945	10	2480	9	15/32	24.1	66.2	1.87

TABLE A-2. Summary ball-bearing data for company B, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$ —Continued

Record No.	Year of test	Number in test group	Load	Z Number of balls	$D_z$ Ball diam.	$L_{10}$	$L_{50}$	Weibull slope
			lb		in.			
2-101	1945	20	2480	9	15/32	36.1	71.6	2.75
2-102	1946	20	2480	9	15/32	63.3	104	3.82
2-103	1946	12	2480	9	15/32	14.4	59.0	1.33
2-104	1946	11	2480	9	15/32	15.1	92.9	1.04
2-105	1945	10	2480	9	15/32	18.8	39.4	2.55
2-106	1950	12	2480	9	15/32	5.63	34.7	1.04
2-107	1950	12	2480	9	15/32	7.23	34.5	1.21
2-108	1951	30	2480	9	15/32	16.7	71.8	1.29
2-109	1951	63	2480	9	15/32	26.5	90.3	1.54
2-110	1950	23	2480	9	15/32	8.35	49.1	1.06
2-111	1943	19	3250	9	15/32	3.79	9.30	2.10
2-112	1937	10	3470	10	7/16	9.05	36.6	1.35
2-113	1944	20	4000	9	15/32	2.98	7.35	2.08
2-114	1943	19	2300	10	15/32	22.5	73.4	1.59
2-115	1938	10	2730	10	15/32	3.82	31.7	.89
2-116	1946	22	2660	10	17/32	6.55	20.8	1.63
2-117	1944	20	2250	11	15/32	17.5	64.3	1.45
2-118	1943	16	2300	11	15/32	61.7	152	2.10
2-119	1945	48	2840	11	15/32	18.6	42.7	2.27
2-120	1947	28	2840	11	15/32	21.6	66.3	1.58
2-121	1947	8	2840	11	15/32	11.9	39.1	1.59
2-122	1948	8	2840	11	15/32	13.9	50.6	1.46
2-123	1943	19	3200	11	15/32	7.80	33.1	1.30
2-124	1944	28	4000	11	15/32	3.55	13.9	1.38
2-125	1943	19	4000	11	15/32	9.40	23.4	2.06
2-126	1947	23	6350	11	11/16	4.76	22.7	1.21
2-127	1944	20	12000	11	1-1/16	3.23	9.86	1.69
2-128	1944	20	12000	11	1-1/16	2.62	9.52	1.46
2-129	1944	9	12700	8	1-1/2	7.89	39.7	1.17
2-130	1949	18	16500	11	1-1/16	4.93	20.4	1.33
2-131	1950	20	16500	11	1-1/16	6.26	16.2	1.98
2-132	1938	8	565	7	5/16	37.3	103	1.85
2-133	1944	20	900	7	5/16	14.0	38.6	1.86
2-134	1938	10	1650	8	13/32	30.3	87.6	1.77
2-135	1944	20	2250	8	15/32	25.7	71.2	1.85
2-136	1943	20	2300	8	15/32	10.5	60.4	1.07
2-137	1944	19	3200	8	15/32	10.3	24.1	2.21
2-138	1944	19	4000	8	15/32	4.56	12.9	1.81
2-139	1937	10	1710	8	17/32	25.1	274	.79
2-140	1938	9	2360	8	17/32	48.8	264	1.12
2-141	1937	10	2680	8	17/32	7.53	60.7	.90
2-142	1937	11	3850	8	17/32	14.9	62.6	1.32
2-143	1947	21	7760	8	29/32	4.57	43.4	.84
2-144	1943	12	9550	8	1-1/16	3.90	40.7	.80
2-145	1947	21	9750	8	1-1/16	15.5	79.4	1.16
2-146	1948	16	11400	8	1-3/16	10.2	43.9	1.29
2-147	1948	20	11400	8	1-3/16	4.71	16.9	1.48
2-148	1947	18	11420	8	1-3/16	10.1	34.2	1.55

TABLE A-3. Summary ball-bearing data for company C, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$

Record No.	Year of test	Number in test group	Load	Z Number of balls	$D_a$ Ball diam.	$L_{10}$	$L_{50}$	$e$ Weibull slope
3-1	1942	94	1580	7	in. 9/16	16.9	64.8	1.40
3-2	1949	29	790	7	9/16	211	729	1.52
3-3	1949	35	1185	7	9/16	74.4	287	1.40
3-4	1940	29	1600	9	1/2	9.62	40.1	1.32
3-5	1945	10	1600	8	15/32	11.9	66.3	1.10
3-6	1943	9	2275	7	17/32	13.8	58.0	1.31
3-7	1946	13	2540	8	15/32	2.38	11.3	1.21
3-8	1946	12	2540	8	15/32	2.38	11.5	1.19
3-9	1949	12	1580	7	9/16	8.75	62.2	0.96
3-10	1949	12	1580	7	9/16	25.7	113	1.27
3-11	1947	24	1600	9	1/2	11.5	113	0.92
3-12	1949	12	610	8	5/16	26.8	65.6	2.10

TABLE A-4. Summary ball-bearing data for company D, with computed values for  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$

Record No.	Year of test	Number in test group	Load	Z Number of balls	$D_a$ Ball diam.	$L_{10}^*$	$L_{50}^*$	$e$ Weibull slope
4-1	1946	19	1750	9	in. 7/16	159	963	1.05
4-2	1951	34	1750	9	7/16	71.7	526	0.94
4-3	1951	56	1750	9	7/16	113	582	1.15

\* Life estimates are in hours.

TABLE A-5. Summary of test groups of ball-bearing data

Company	Number of test groups	Total number of bearings in test group
A.....	50	1,259
B.....	148	3,289
Type B-1.....	37	-----
Type B-2.....	94	-----
Type B-3.....	17	-----
C.....	12	291
D*.....	3	109
Total (all companies).....	213	4,948

\*These data were not used in the main analyses.

SPECIMEN WORKSHEET

Reference No. \_\_\_\_\_  
 Bearing Mfg. by \_\_\_\_\_  
 Bearing Tested by \_\_\_\_\_  
 Date of Test 8-26-46  
 Bearing No. \_\_\_\_\_  
 Load 500 B.L.  
 Speed 2000 r.p.m.  
 Lubrication: Type Jet Oil  
 Frequency \_\_\_\_\_  
 Ball No. and Dia. 9 - 1/4"  
 Contact Angle 0°  
 Groove Radius: Inner Ring 51.6%  
 Outer Ring 53.0%  
 Number of Rows 1  
 Bore 20 mm.  
 O.D. 42 mm.  
 Lot Size 25 Taken on 23

Bearing temperature measured on outer ring at point of maximum load \_\_\_\_\_  
 Material: Type \_\_\_\_\_  
 Source \_\_\_\_\_  
 Rockwell Hardness of:  
 Inner Ring 63.5  
 Outer Ring 64.  
 Balls \_\_\_\_\_

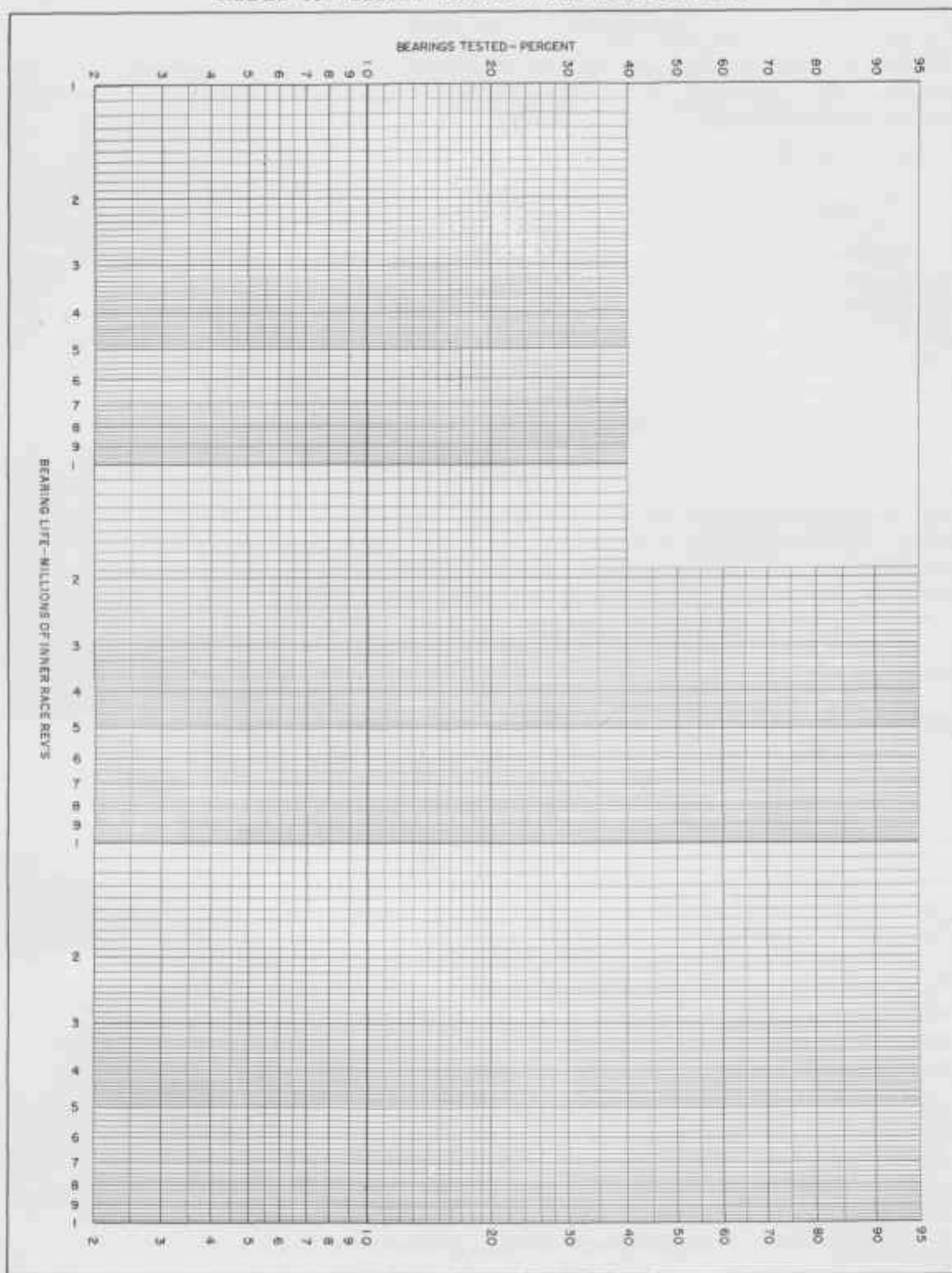
Ball Failure 13 52%  
 Inner Ring Failure 5 20%  
 Outer Ring Failure 1 4%

Test life in 10<sup>6</sup> revolutions:  
 Median 68.  
 Mean 71.  
 B-10 29.  
 Slope of Curve 2.23  
 Test No. 3183  
 Lot 71

Table Ordered According to  
Endurance Life

Brg. No.	Endurance Mill. Revs.	Type of Failure	Remarks
16	17.88	Ball	
10	28.92	Ball	
5	33.00	Ball	
19	41.52	I.R.	
9	42.12	Ball	
11	45.60	Ball	
15	48.48	Ball	
12	51.84	Ball	
20	51.96	Ball	
18	54.12	I.R.	
13	55.56	I.R.	
1	67.80	Ball	
<del>2</del>	<del>67.80</del>	<del>L. Bore</del>	<del>Omitted</del>
<del>3</del>	<del>67.80</del>	<del>L. Bore</del>	<del>Omitted</del>
4	60.64	Ball	
6	68.64	L. Bore	
25	68.88 →	Disc.	
22	84.12	Ball	
17	93.12	Ball	
7	98.64	I.R.	
23	105.12	I.R.	
24	105.84 →	Disc.	
21	127.92	Ball	
8	128.04	O.R.	
14	173.40 →	Disc.	

SAMPLE OF WEIBULL FUNCTION COORDINATE PAPER



## 5. Appendix B. Evaluation of $L_{10}$ , $L_{50}$ , and Weibull Slope $e$ , by Using Order Statistics for Censored Data

This is a technical appendix that gives the mathematical basis for estimating, for each test group, the values of  $L_{10}$  and  $L_{50}$  for use in the regression analysis discussed in appendix C, and also the Weibull slope  $e$ .

### 5.1. Weibull Distribution

#### a. Characteristics

As noted in the text, the basic assumption for estimating  $L_{10}$ ,  $L_{50}$ , and  $e$  for each test group was that the probability distribution of fatigue lives of individual bearings could be represented by a "Weibull distribution."<sup>6</sup> This means that the observed fatigue lives (number of revolutions) of all the bearings in a test group of, say,  $n$  bearings constitute a random sample of  $n$  independent observations from a distribution whose cumulative (from above) distribution function (hereafter denoted by cdf) is<sup>7</sup>

$$\begin{aligned} S(L) &= \text{Prob}\{\text{life} \geq L\} \\ &= \exp[-(L/a)^e], \quad 0 \leq L < \infty, \end{aligned} \quad (\text{B1})$$

where  $a$  and  $e$  are the two parameters to be fitted. They are related to  $L_{10}$  and  $L_{50}$  by eq (B2a) below. The function  $S(L)$  is also termed the "survivorship" function. This distribution is one of three limiting types to which the distribution of the smallest member of a sample, under general conditions, tends as the sample size is increased indefinitely. (Another type is discussed in the following section.) This matter was first studied chiefly by Fisher and Tippett [5], and for this reason the type (B1) is sometimes referred to as Fisher-Tippett type III for smallest values.

There are both theoretical and practical reasons for choosing the Weibull distribution (B1) as the underlying probability distribution for fatigue life.

*Theoretical.* Here it is assumed that fatigue is an "extreme-value" phenomenon, related in some manner to the strength at the weakest point in the material under stress. The theoretical reasoning that proceeds from this assumption is mentioned by a number of authors, and is given explicitly, for example, by Freudenthal and Gumbel in [6, p. 316 to 318]. It leads precisely to the form (B1) (see eq (2.9) in [6]). It is recognized that this statistical assumption has not received universal acceptance. This paper is, however, not concerned with the relative merits of various statistical theories of fatigue, but merely with consequences of a reasonable choice from among them.

*Practical.* Application of the Weibull distribution received extensive attention by W. Weibull in [19], where he showed that a distribution of the general type (B1) represented certain fatigue-life data quite satisfactorily. In addition, inspection of the special "Weibull" plots accompanying the worksheets suggests that many can be fitted satisfactorily by a straight line representing a Weibull distribution, as explained below.

The manner in which these graphs are constructed is described by Weibull in [19]. A sample of Weibull-function coordinate paper used for this purpose is included in appendix A. The essence of the method is that eq (B1) may be converted, by taking logarithms twice, into

$$e(\ln L) - (e \ln a) = \ln\{\ln(1/S)\}, \quad (\text{B2})$$

where "ln" denotes the natural logarithm (base denoted by  $e$ ) and  $S=S(L)$ . From eq (B1) and the definitions of  $L_{10}$  and  $L_{50}$ , when  $L=L_{10}$ ,  $S(L)=.90$ ; and when  $L=L_{50}$ ,  $S(L)=.50$ . These values substituted in eq (B2) give

<sup>6</sup> So named for W. Weibull (cf. [18], p. 16 ff.), who is considered to be one of the first to study it extensively.

<sup>7</sup> The use of a continuous instead of discrete probability distribution will introduce no appreciable error.



$$\left. \begin{aligned} e(\ln L_{10}) - e(\ln a) &= \ln[\ln(1/.90)] = -2.25037 \\ e(\ln L_{50}) - e(\ln a) &= \ln[\ln(1/.50)] = -0.36651, \end{aligned} \right\} \quad (\text{B2a})$$

the values on the right-hand side being obtained from [17, table 2]. These are the relationships between the parameters  $a$ ,  $e$ ,  $L_{10}$ , and  $L_{50}$ . The right-hand numerical values will later be denoted by  $y_{.30}$ ,  $y_{.50}$ , respectively. Equation (B2) may be written

$$ex - a' = y,$$

where

$$x = \ln L, \quad a' = e \ln a, \quad y = \ln[\ln(1/S)]. \quad (\text{B3})$$

The variables  $x$ ,  $y$  correspond to the two scales shown on the Weibull-function coordinate paper in appendix A. The variable  $x$ , with unrestricted values, corresponds to the horizontal scale "Bearing life," having a logarithmic scale. The variable  $y$  is represented through the percentage surviving,  $S$ , or rather through the (vertical) scale for "bearings tested—percent" = percent failed<sup>10</sup> =  $1 - S = P$ , which can vary only between 0 and 1. This scale also has non-uniform graduations, given by the iterated logarithm in (B3).

The Weibull distribution is thus seen to be equivalent to a straight-line relationship, with "Weibull slope"  $e$ , between the logarithm of fatigue life and an associated quantity  $y$  depending only on its relative rank when the fatigue lives are arranged in ascending order. Thus, goodness of fit of the straight line (B3) is equivalent to goodness of fit of a Weibull distribution to the fatigue lives  $L$  of an individual test group. In fact, one common method of statistical analysis of fatigue-life data (Freudenthal and Gumbel [6]) depends upon the use of the classical method of least squares for fitting this straight line. This method is, however, subject to certain limitations described below. Instead, an alternative method, presented in the following sections, is preferred that fits the distribution of  $x = \ln L$  directly by use of order statistics.

#### b. Limitations of Fitting by Least Squares

In the classical method of least squares for fitting the straight-line relationship (B3) to a test group of ball-bearing data, pairs of values  $(x_i, y_i)$ ,  $i=1, \dots, n$ , are required. The values of  $x = \ln L$  are obtained from the given data. However, the variable  $y$ , measured through the percentage failing,  $P = 1 - S$ , presents difficulties. The problem of how to plot  $P$  is known as the problem of "plotting position."

It seems clear that the values,  $P_i$ , of the plotted variable,  $P$ , must somehow be related to the rank order of the bearings as they fail. A natural choice is the percentage failing:  $P = f/n$ , where  $f$  is the rank order of failure in a test group of  $n$ . This is not advisable for reasons discussed at length by Gumbel in [9, p. 14], where he advocates the plotting position  $f/(n+1)$ .<sup>11</sup> Other workers take different positions, and the question of plotting position must be regarded as still unsettled.

A second difficulty with the use of least squares is that as usually used it fails to take adequate account of the number of items remaining intact ("runouts") in the incompleting tests. As a final point, it is to be noted that the successive plotted points are not independent, as they represent the observed lives in increasing order. A correct use of least squares procedures would have to take into account all the intercorrelations, which is not done in the usual application of the "method of least squares." The method of order statistics described in section 3 has the advantage of avoiding the above limitations of the least squares method.

<sup>10</sup> The symbol  $P$  as used here should not be confused with the same symbol for load used in the life formula. In any event, the meaning will be clear from the context.

<sup>11</sup> This plotting position was also used by Weibull in [16] (cf. eq (72) and the vertical scale in figures 3 and 4 therein).

## 5.2. The Extreme-Value Distribution

### a. Relation to Weibull Distribution

The preceding section indicates that logarithms of lives, rather than lives themselves, are the natural units in which to carry out the analysis. This idea has also been adopted by those who do not use the Weibull distribution, either because they are unaware of its existence or because they do not feel it fits their data.

If the Weibull distribution is adopted for fatigue life,  $L$ , then the variate,  $x = \ln L$ , has the nonnormal cumulative distribution function<sup>12</sup>

$$\begin{aligned} G(x) &= \text{Prob}\{\ln(\text{life}) \geq x\} = \text{Prob}\{\text{life} \geq e^x\} \\ &= S(e^x) = \exp(-a^{-e^x}) = \exp[-e^{(x-\ln a)/\beta}], \quad -\infty < x < \infty. \end{aligned}$$

This may be written

$$G(x) = \Phi(y) = \exp(-e^y), \quad (\text{B5})$$

where

$$y = (x - u)/\beta, \quad -\infty < x < \infty, \quad (\text{B6})$$

and

$$u = \ln a, \quad \beta = 1/e \quad (\text{B7})$$

are the two parameters. The distribution,  $\Phi(y)$ , considered as a distribution of the "reduced variable,"  $y$ , has standardized parameters  $u=0$ ,  $\beta=1$ , and is called the "reduced distribution."

The form (B5) is another of the three asymptotic distributions of extreme values, sometimes designated as Fisher-Tippett type I for smallest values. This distribution has been studied extensively, chiefly by E. J. Gumbel (e. g., [7, 8, 9]). In this paper, the term "extreme-value distribution" will be given to the distribution of smallest values (unless otherwise specified), although this name is frequently given to the largest-values case.

From the above discussion, it is apparent that methods pertinent to the type I extreme-value distribution (B5) are appropriate. For this purpose there is available a mathematical approach recently developed by one of the authors of this report, and described in detail in [13].

### b. Characteristics

A description of the extreme-value distribution (B5), together with an interpretation of its parameters in terms of life estimates (or rather their logarithms), is essential to an understanding of the application of the method of order statistics in this paper. It will be seen that the problem of estimating life is equivalent to that of estimating the parameters  $u$  and  $\beta$ .

The parameters of the extreme-value distribution (B5) are depicted in figure 1 (page 291). The quantity  $u$  is the position of the mode or highest point of the (frequency) distribution. The quantity  $\beta$  is a scale parameter, analogous to the standard deviation,  $\sigma$ , in the case of the normal distribution. In fact,  $\beta$  is  $\sqrt{6}/\pi$  (about  $\frac{2}{3}$ ) times the standard deviation of the extreme-value distribution.

Although the two parameters,  $u$ ,  $\beta$ , completely specify the distribution, it is very useful to introduce related quantities of the form

$$t = u + \beta y, \quad (\text{B8})$$

which are linear combinations of parameters  $u$  and  $\beta$  and may thus also be regarded as parameters when known values are later assigned to  $y$ . Introduction of  $t$  makes it possible to estimate  $u$  and  $\beta$  simultaneously. Thus if  $t$  can be obtained as  $a + by$  with  $a$  and  $b$  known and  $y$  arbitrary, then the values  $u=a$ ,  $\beta=b$  can be read off at once.

The parameter  $t$  has another highly important meaning. In figure 1 the area  $F$  under the distribution to the right of the ordinate erected at  $t$  represents the probability that a value

<sup>12</sup> Cf. Freundthal and Gumbel [6], eq (2.5), (2.6), (2.8), (2.9).

DENSITY FUNCTION,  $g(x) = G'(x)$

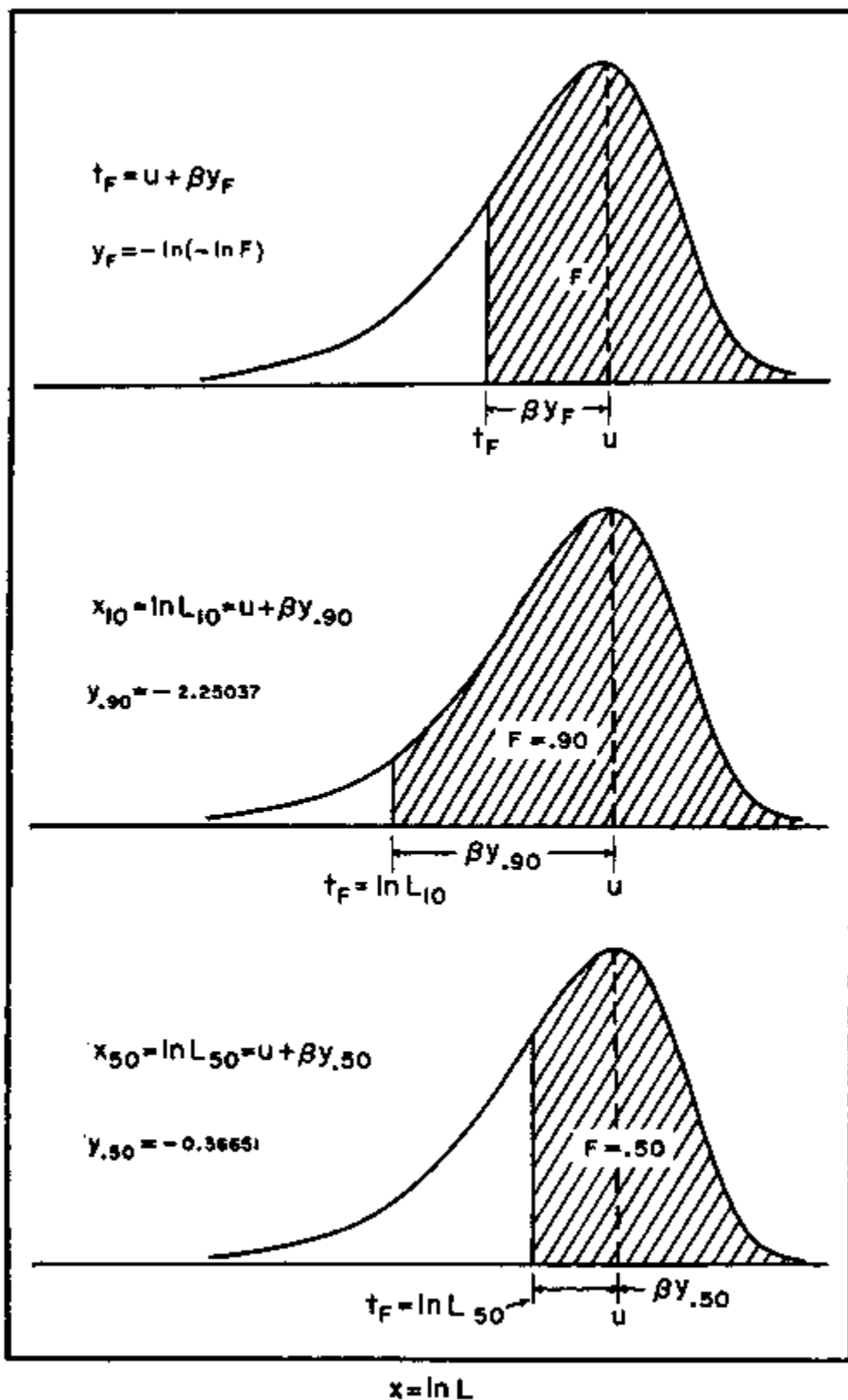


FIGURE 1. General form of extreme-value distribution (for smallest values) showing relationship of parameters  $t_F$ ,  $x_{10} = \ln L_{10}$ , and  $x_{50} = \ln L_{50}$ , to  $u$  and  $\beta$ .

larger than  $t$  will occur. Thus  $t$  is a function of  $F$  and may be written  $t_F$ , as shown; it is designated the "upper  $100F$ -percentage point" of the distribution. For example, if  $F=.90$ , then  $t=t_{.90}$  represents a value of  $x=\ln L$ , which will be exceeded by 90 percent of the population. This is associated with rating life  $L_{.90}$  (life exceeded by 90 percent of bearings) by the relation

$$t_{.90}=x_{10}=\ln L_{10}, \quad (\text{B9})$$

where  $x$  represents life in logarithmic units. Similarly, for median life,

$$t_{.50}=x_{50}=\ln L_{50}. \quad (\text{B10})$$

Since the  $t$ 's are regarded as parameters of the distribution, so also are  $x_{10}$  and  $x_{50}$ , and therefore  $L_{10}$  and  $L_{50}$ . These are not, of course, all independent.

In general, we have the percentage point  $t_F$ , which, expressed in terms of the original parameters  $u$  and  $\beta$ , may be written in the form (B8):

$$t_F = u + \beta y_F, \quad (\text{B8a})$$

where  $y$  is a quantity depending only on the probability  $F$ , determined as follows. We have from (B8a)

$$y_F = (t_F - u) / \beta, \quad (\text{B11})$$

i. e.,  $y_F$  is the value of  $(x-u)/\beta$  when  $x$  takes the value  $t_F$ . But by definition of the probability  $F$ , in view of (B5), (B6), and (B11),

$$F = \text{Prob}\{x \geq t_F\} = G(t_F) = \Phi(y_F) = \exp(-e^{y_F}). \quad (\text{B12})$$

Thus, solving for  $y_F$ , we obtain

$$y_F = \ln(-\ln F). \quad (\text{B13})$$

This is the reduced variable corresponding to the probability  $F$ , and may be obtained by a simple change in sign from table 2 of [17], which tabulates the function

$$-\ln(-\ln \Phi_y),$$

where  $\Phi_y$ , a probability, takes on values from 0 to 1. Thus,

$$\left. \begin{array}{l} \text{for } F=.90, y_F = -2.25037; \\ \text{for } F=.50, y_F = -0.36651. \end{array} \right\} \quad (\text{B14})$$

The above discussion shows that both  $x_{10}$  and  $x_{50}$  (rating and median lives in logarithmic units) may be determined once the general percentage point (B8a) is estimated by giving the two particular values (B14) to  $y_F$ .

#### c. Conversion From Largest to Smallest Values

The methods and numerical results developed in [13] were for problems, such as maximum gust-loads on airplanes, that required the distribution of *largest* sample values. In order to adapt this material to the distribution of *smallest* values (B5) required here, the relationships of symmetry involved in the reversal of direction must be examined with care. To avoid confusion, it is necessary to use subscripts  $L$  and  $S$  to distinguish between quantities related to the largest-value distribution from those related to the smallest-values case. No generality is lost by use of reduced variates. Thus, in (B5),  $x$  will be replaced by the reduced variate  $y$ , and, for simplicity, the symbol  $G(y)$  will be used instead of  $\Phi(y)$ :

$$G(y) = \Phi(y) = \exp(-e^y). \quad (\text{B15})$$

From this, the ("cumulative from above") distribution of smallest values is

$$\text{Prob}\{Y_S \geq y\} = G_S(y) = \exp(-e^y), \quad -\infty < y < \infty, \quad (\text{B16})$$

where  $Y_S$  denotes the reduced smallest value. The corresponding distribution of largest values is (see Gumbel [9, eq (I), p. 21])

$$\text{Prob}\{Y_L \geq y\} = H_L(y) = 1 - \text{Prob}\{Y_L < y\} = 1 - \exp(-e^{-y}) = 1 - G_S(-y), \quad -\infty < y < \infty, \quad (\text{B17})$$

from (B16).

The corresponding relation for the density functions is obtained by differentiation, with  $g_S(y) = G'_S(y)$ , and  $h_L(y) = H'_L(y)$ :

$$g_S(y) = h_L(-y). \quad (\text{B18})$$

Hence the two distributions are merely mirror images of each other. The moments of the distributions are related as follows:

$$\nu_{kS} = E_S(y^k) = \int_{-\infty}^{\infty} y^k g_S(y) dy = \int_{-\infty}^{\infty} (-y')^k h_L(y') dy' = (-1)^k \nu_{kL}. \quad (\text{B19})$$

Thus, the means differ in sign and the variances are identical:

$$\nu_{1S} = -\gamma = -\nu_{1L}, \quad (\text{B20})$$

$$\sigma_S^2 = \frac{\pi^2}{6} = \sigma_L^2. \quad (\text{B21})$$

These values are given, for example, in [9, p. 23, eq (3.27)].

Finally, we need the relationships between moments of the order statistics for the two distributions. As the smallest-value distribution is a reversal of the largest-value distribution, it is natural to reverse the arrangement of the order statistics as well. This gives simpler results. Thus we are interested in the  $i$ th order statistic in the series

$$(S): \quad y'_1 \geq y'_2 \geq \dots \geq y'_i \geq \dots \geq y'_n, \quad (\text{B22})$$

where the parent distribution is that of *smallest* values. Primes will be used as a reminder that the order is *descending*, not ascending. Thus in tables B-2 and B-3 the absence of primes indicates that the order statistics are in increasing order.

(B22) is the analogue of the series

$$(L): \quad y_1 \leq y_2 \leq \dots \leq y_i \leq \dots \leq y_n \quad (\text{B23})$$

of order statistics for the *largest*-value parent distribution. Whenever a distinction is necessary the subscripts  $S$  or  $L$  will be used with the  $y$ 's.

From (B18) it may seem intuitively (and may be justified rigorously) that the distributions and moments of the order statistics follow the same symmetry relationships as the parent distributions, namely,

$$\left. \begin{aligned} E_S(y_i^m) &= (-1)^m E_L(y_i^m) \\ E_S(y'_i) &= E_L(y_i) \\ \sigma_{S,i}^2 &= \sigma_S^2(y_i) = \sigma_L^2(y_i) = \sigma_{L,i}^2 \\ \sigma'_{S,i} &= \sigma_S(y'_i, y'_i) = \sigma_L(y_i, y_i) = \sigma_{L,i} \end{aligned} \right\} \quad (\text{B24})$$

In other words, the even moments remain the same; the odd moments change only in sign.

The above development shows that the numerical results for moments of order statistics previously obtained in [13] for the largest-value case can be used here for smallest values without any substantive change.

### 5.3. Method of Order Statistics for Censored Samples

#### a. For Small Samples

Consider an independent random sample of  $n$  items from the distribution of smallest values, of which only the  $k$  smallest values can be observed. In view of the preceding discussion, it is desirable in the theoretical development to deal with the order statistics in descending order:

$$(x'_1 \geq x'_2 \geq \dots \geq x'_{n-k}) \geq x'_{n-k+1} \geq \dots \geq x'_n, \quad (\text{B25})$$

where the parentheses denote the  $(n-k)$  (largest) unobservable values, and the remaining  $k$  values are known. This arrangement materially simplifies the exposition. Primes will again be used to denote descending order to distinguish from ascending order, which will occur in the later parts of this section.

From the  $k$  known values it is desired to determine an estimator

$$T'_{n,k} = w'_1 x'_{n-k+1} + w'_2 x'_{n-k+2} + \dots + w'_k x'_n, \quad k \leq n, \quad (\text{B26})$$

(i. e., the weights  $w'_j$ ) of the general parameter,

$$t_r = u + \beta y_r, \quad (\text{B27})$$

of the extreme-value population (B5), such that  $T'$  in (B26) is (1) unbiased and (2) of minimum variance. Mathematically, this means that

$$E(T') = t_r, \quad (\text{B28})$$

where  $E$  denotes mathematical expectation, and

$$\text{Var}(T') = \text{a minimum}, \quad (\text{B29})$$

subject to the above condition.

From (B6),

$$x = u + \beta y, \quad (\text{B30})$$

where  $y$  is the reduced variable and  $x$  the observed variable. From this the following relations for the order statistics  $x_r$  and  $y_r$  are apparent:

$$x'_j = u + \beta y'_j, \quad j = n-k+1, n-k+2, \dots, n, \quad (\text{B31})$$

$$x'_{n-k+1} \geq x'_{n-k+2} \geq \dots \geq x'_n, \quad (\text{B32})$$

$$y'_{n-k+1} \geq y'_{n-k+2} \geq \dots \geq y'_n, \quad (\text{B33})$$

$$E_S(x'_j) = u + \beta E_S(y'_j). \quad (\text{B34})$$

The values  $E_S(y'_j)$  may be obtained with the aid of the table in [14]. This table gives the values of  $E_L(y'_r)$  where the order statistics,  $y'_r$ , are in descending order (as indicated by the prime). The means needed in (B34) are obtained from (B24) and the evident symmetry relations

$$E_L(y_r^m) = (-1)^m E_L(y'_{n-r+1}^m)$$

as

$$E_S(y'_j) = -E_L(y'_{n-j+1}). \quad (\text{B35})$$

Reference [14] gives the values of  $E_L(y'_r)$  for  $r=1(1)\min(n,26)$ ,  $n=1(1)10(5)60(10)100$ .

From (B28) and (B31),

$$E(T') = \sum_{j=1}^k w'_j [u + \beta E(y'_{n-k+j})] = t_r = u + \beta y_r. \quad (\text{B36})$$

This is required to be an identity for all values of the parameters  $u, \beta$ . Equating their coefficients gives the two conditions on the weights,  $w'_j$ :

$$\sum_{j=1}^k w'_j = 1, \quad \sum_{j=1}^k [E(y'_{n-k+j})] w'_j = y_r, \quad (\text{B37})$$

where the numerical values  $E(y'_{n-k+j})$  may be obtained from [14] as already indicated.

For the variance condition (B29), we have, in view of (B26),

$$\sigma_3^2(T') = \sum_{j=1}^k w'_j{}^2 \sigma^2(x'_{n-k+j}) + \sum_{j=1}^k \sum_{\substack{l=1 \\ l \neq j}}^k w'_l w'_j \sigma(x'_{n-k+l}, x'_{n-k+j}). \quad (\text{B38})$$

$$\begin{aligned} \text{Var}(T') &= [\sum w'_j{}^2 \sigma_{n-k+j}^2 + \sum \sum' w'_l w'_j \sigma'_{n-k+l, n-k+j}] \beta^2 = V_k^{(n)'} \beta^2 \\ &= \text{minimum subject to (B28)}. \end{aligned} \quad (\text{B39})$$

Use of Lagrange multipliers in the same manner as in [13, pp. 50-52] gives, after differentiation, the conditions on the weights:

$$\sigma_{n-k+j}^2 w'_j + \sum_{\substack{l=1 \\ l \neq j}}^k \sigma'_{n-k+l, n-k+j} w'_l + \lambda + \mu E y'_{n-k+j} = 0, \quad j=1, 2, \dots, k. \quad (\text{B40})$$

For each fixed value of  $k \leq n$  there are  $k$  linear equations which, with the two in (B37), form a simultaneous system of  $(k+2)$  equations in the  $(k+2)$  unknowns,  $w'_1, w'_2, \dots, w'_k, \lambda, \mu$ . The values of  $\lambda$  and  $\mu$  are useful as a check, because, if (B40) is multiplied by  $w_j$  and summed for  $j=1, 2, \dots, k$ , the result is, in view of (B37) and (B39),

$$V_k^{(n)'} + \lambda + \mu Y_r = 0. \quad (\text{B41})$$

The minimum value,  $V_k^{(n)'}$ , will be denoted by  $Q'_{k,n}$ .

In general, there will be a set of  $(k+2)$  linear equations to solve for each  $k=2, \dots, n$ .

(1) *Case  $k=n$ .* For  $k=n$ , the matrix of coefficients and right-hand "constant terms" of (B40) and (B37) is the  $(n+2)$  by  $(n+3)$  matrix

$$\Lambda_n^0 = \begin{bmatrix} \sigma'_{11} & \sigma'_{12} & \dots & \sigma'_{1n} & 1 & E y'_1 & 0 \\ \sigma'_{21} & \sigma'_{22} & \dots & \sigma'_{2n} & 1 & E y'_2 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma'_{n1} & \sigma'_{n2} & \dots & \sigma'_{nn} & 1 & E y'_n & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & 1 \\ E y'_1 & E y'_2 & \dots & E y'_n & 0 & 0 & y_r \end{bmatrix}. \quad (\text{B42})$$

The ordinary  $(n+2)$  by  $(n+2)$  matrix of coefficients, without the constant terms, will be denoted by  $\Lambda_n$ . If  $\Gamma_n$  denotes the vector column of constant terms, then

$$\Lambda_n^0 = [\Lambda_n \Gamma_n], \quad (\text{B43})$$

and the linear system of  $(n+2)$  equations may be denoted

$$\Lambda_n W'_n = \Gamma_n, \quad (\text{B44})$$

where  $W'_n$  denotes the column vector of the  $(n+2)$  unknowns  $w'_1, w'_2, \dots, w'_n, \lambda, \mu$ .

The coefficients of the unknowns in (B44) involve the means  $E(y'_j)$ , already discussed, and covariances  $\sigma'_{ij}$ . These values are given in table B-1 for  $n=2$  to 6. The  $\sigma'_{ij}$  were computed by the method developed in [12]. Table B-1 also indicates how the moments for the largest values case can be obtained simply from those shown.

The  $(n+2)$  solutions of (B44) are all expressible linearly in terms of the components of  $\Gamma_n$ . Thus the solutions all take the form

$$\left. \begin{aligned} w'_j &= a'_j + b'_j y_F, & j=1, 2, \dots, n \\ \lambda &= c'_1 + d'_1 y_F \\ \mu &= c'_2 + d'_2 y_F \end{aligned} \right\} \quad (\text{B45})$$

Substituting these  $w'_j$  in (B39) gives an expression of the form

$$Q'_{n,n} = V_{n,n}^{(n)} = A'_n + 2B'_n y_F + C'_n y_F^2. \quad (\text{B46})$$

The quantities  $a'_j$ ,  $b'_j$  for the weights  $w'_j$ , for  $n=2$  to 6, are shown in table B-2. The coefficients  $A'_n$ ,  $B'_n$ ,  $C'_n$  of  $Q'_{n,n}$ , and the values of  $Q'_{n,n}$  evaluated at  $F=.90$ ,  $.50$ , corresponding to  $L_{10}$ ,  $L_{50}$ , respectively, are given in table B-3.

Calculations were limited to  $n=6$  in this paper, in view of the diminishing returns in "efficiency" (see below) for increasingly larger amounts of computing. Methods suitable for larger values of  $n$  are discussed later.

Table B-3 shows that as sample size increases from  $n=2$  to 6 (in the case  $k=n$ ), the variance diminishes for the percentage-point parameters  $t_F$  for  $F=.90$  and  $.50$ , i. e.,  $t_F = x_{10} = \ln L_{10}$  and  $t_F = x_{50} = \ln L_{50}$ . This is a common characteristic of the behavior of estimators for increasing sample size. Another method whereby estimators may be compared is through their efficiency.

Efficiency is a measure intended to provide a convenient standard of comparison for estimators. This is done for two estimators to be compared by dividing the variance of each into a theoretical "smallest" variance,  $Q_{LB}$ , known as the "Cramér-Rao lower bound." Further details in the case of complete samples where  $k=n$ , as here, may be found in [13, p. 14 and 15]; values of  $Q_{LB}$  are also indicated in this reference in table III (a).

Table B-4 shows the efficiency values so obtained, for the case  $k=n$ ,  $n=2$  to 6, as regards the order-statistics estimators for the parameters  $x_{10} = \ln L_{10}$ ,  $x_{50} = \ln L_{50}$ .

These values show that for  $x_{10}$ , the efficiency, starting with under 70 percent for  $n=2$ , increases rapidly until 89 percent, out of a possible maximum 100 percent, is reached for  $n=6$ . A 90-percent-efficient estimator is generally considered to be quite good. As regards  $x_{50}$ , the efficiency is well above 95 percent for all the values of  $n$ , and for  $n=6$  exceeds 99 percent. In view of results of this nature, and because of the increasingly heavy computations necessary, calculations were not carried beyond  $n=6$  in [13].

The above applies to estimation of the parameters  $x_{10}$  and  $x_{50}$ , which it will be recalled are the logarithms of the actual life estimates  $L_{10}$ ,  $L_{50}$ . It is believed that efficiency of the method of order statistics in obtaining estimates of actual life  $L_{10}$ ,  $L_{50}$  is probably reasonable, in view of its high efficiency in estimating the logarithms,  $x_{10}$ ,  $x_{50}$ .

(2) *Case  $k < n$ .* For the case  $k < n$ , the procedure is very similar. One starts with a  $(k+2)$  by  $(k+3)$  order matrix  $\Lambda_k^0$  derived from  $\Lambda_n^0$  in (B42) by striking out the first  $(n-k)$  rows and columns. One proceeds in this manner for  $k=n-1$ ,  $n-2$ , etc., until when  $k=2$  the matrix becomes

$$\Lambda_2^0 = \begin{bmatrix} \sigma'_{n-1, n-1} & \sigma'_{n-1, n} & 1 & E y'_{n-1} & 0 \\ \sigma'_{n, n-1} & \sigma'_{n, n} & 1 & E y'_n & 0 \\ 1 & 1 & 0 & 0 & 1 \\ E y'_{n-1} & E y'_n & 0 & 0 & y_F \end{bmatrix} \quad (\text{B47})$$

representing a set of 4 equations in 4 unknowns.



The resulting weights  $w'_i$  and variances  $Q'_{n,k}$  were obtained in similar manner to those for  $k=n$  in (B45) and (B46). These, it will be recalled, are primed quantities, associated with *descending* order of the order statistics. Because the observations,  $x_i$ , for successive failures naturally occur in ascending order, it is more useful for actual application, in contrast to theoretical development, to tabulate the weights and covariances for the order statistics in *ascending* order. This has been done in table B-2, giving the weights  $w_i = a_i + b_i y_p$ , and in table B-3, giving the variances  $Q_{n,k} = A + 2B y_p + C y_p^2$  for the estimators  $T_{n,k}$  formed with the above weights. These variances are also evaluated for the parameters  $x_{10} = \ln L_{10}$ ,  $x_{50} = \ln L_{50}$ . The relationships of these unprimed quantities to the primed ones of the previous theoretical development is merely a reversal of the order throughout, as indicated by subscripts: i. e., every  $a'_i$  is changed to the corresponding  $a_{k-i+1}$  and similarly for  $b'_i$  and  $w'_i$ . The variances  $Q$  may be shown to remain unchanged.

#### b. Extension to Larger Samples

Samples of more than six items are broken up into independent samples of 6 with a remainder subgroup, if necessary, of from 2 to 5 items. Because the endurance data were arranged in increasing order of life, independent random subgroups could not be obtained by simply taking groups of 6 in the (numerical) order in which they appeared on the worksheets. It was therefore first necessary to randomize the endurance lives on each data worksheet. This was accomplished by use of random numbers that were generated in the electronic computer (the SEAC) as needed.

Such artificial randomization is not desirable when it can be avoided, because the results of the calculations are then not unique, but may depend to a limited degree on the particular set of random numbers used.<sup>13</sup> It is therefore recommended that when the bearings in a test group are to be simultaneously run on a battery of fatigue-testing machines, the individual bearings should be recorded in advance in some more or less natural order *independent* of the order in which failure takes place in the course of the test. Natural order might be order of manufacture, order of testing, etc.

In the present investigation, each subgroup was treated as a random sample by the methods already developed for size 6 or less. That is, a "subestimator" was calculated for each subgroup and the results averaged to produce an over-all sample estimator.

An estimator, both for the individual subgroup and for the over-all sample, was obtained for each of the four population quantities:

$$u, \quad \beta, \quad t_{30} = x_{10} = \ln L_{10} = u + y_{.30}\beta, \quad t_{50} = x_{50} = \ln L_{50} = u + y_{.50}\beta. \quad (\text{B48})$$

For subgroups, these four parameter estimates are given by (caret denotes "estimate of")

$$\hat{u} = \sum_{i=1}^k a_i x_i, \quad \hat{\beta} = \sum_{i=1}^k b_i x_i, \quad \hat{x}_{10} = T_{n,k}(10) = \hat{u} + y_{.10}\hat{\beta}, \quad \hat{x}_{50} = T_{n,k}(50) = \hat{u} + y_{.50}\hat{\beta}, \quad (\text{B49})$$

where  $y_{.30} = -2.25037$ ,  $y_{.50} = -0.36651$ , and  $x_1 \leq x_2 \leq \dots \leq x_k$ ,  $2 \leq k \leq n \leq 6$ , are the logarithms of the actual observed lives in a subgroup arranged in ascending order, and the  $a_i$  and  $b_i$  are read directly from table B-2. For the over-all sample estimator, the subestimators  $T_{n,k}$  are merely averaged.

For later use (appendix C) the variance of the over-all estimator,  $T$ , and its relation to sample size will be considered here. Consider first the case of a *complete* sample, where no intact bearings are present because the test is run to completion. Let  $n$  be the sample size; then there are two cases, according as (1)  $n \leq 6$ , or (2)  $n > 6$ .

<sup>13</sup> This effect can be reduced somewhat by making a duplicate run and averaging the results, as was done in this study.

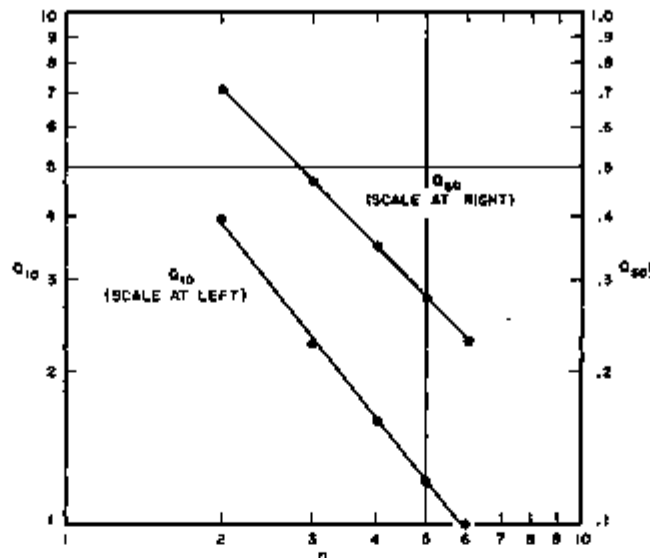


FIGURE 2. Relationship of variances  $Q_{10}$  and  $Q_{50}$  to sample size  $n$  for  $n=2$  to 6 (logarithmic scale in each direction).

$Q_{10}$  is variance of estimator of  $x_{10} = \ln L_{10}$   
 $Q_{50}$  is variance of estimator of  $x_{50} = \ln L_{50}$   
 All  $Q$ 's are in units of  $\beta^2$

(1)  $n \leq 6$ . Table B-3 gives the numerical variances,  $Q_{10}$  and  $Q_{50}$ , for  $n=2$  to 6. These values are plotted in figure 2 on double-logarithmic paper. The values for  $Q_{50}$  (right-hand scale) are seen to lie on a straight line of slope negative unity. This shows that at least in this case, variance is inversely proportional to sample size. For the other case,  $Q_{10}$ , a straight line also gives a reasonably good fit, and its slope appears to differ only a little from  $-1$ . Hence the underlined statement is approximately applicable here too.

(2)  $n > 6$ . If a sample of size  $n > 6$  is broken into equal subgroups (of size 6, for example) there will generally be a remainder of size less than 6. The preceding development, when suitably modified, shows that for large  $n$  the influence of this remainder is small compared to the remaining bulk of the sample and thus the rule in question holds approximately in this case. Agreement with the rule is less close for a few cases of moderate  $n$ , but for simplicity the inverse relationship will be taken as a reasonable rule of thumb in all cases for the over-all purposes of analysis.

Two complete runs were made on the SEAC for each of the 213 test groups of data, and the two results were averaged for each group, giving values of the averages

$$\bar{x}, \bar{\beta}, \bar{i}_{.10} = \overline{\ln L_{10}}, \bar{i}_{.50} = \overline{\ln L_{50}}. \quad (\text{B50})$$

From these, the values of  $L_{10}$  and  $L_{50}$  were obtained from a table of exponentials and the Weibull slope  $\epsilon = 1/\bar{\beta}$  obtained as a consequence of formula (B5). An example showing the steps in calculation of  $L_{10}$ ,  $L_{50}$ , and  $\epsilon$  is discussed below.

Because this investigation represents probably the most extensive mass fitting of the Weibull distribution made to date, a tabulation of the 213 values of the parameter  $\epsilon$  will be of considerable value to future applications of this distribution. This is shown in table B-6. The corresponding histogram is given in figure 3. Particular items of interest are

mean of  $\epsilon$  (all 4 companies) = 1.51,

median  $\epsilon$  (all 4 companies) = 1.43.

Note that 50 percent of the values are in the interval  $\epsilon = 1.17$  to 1.74.

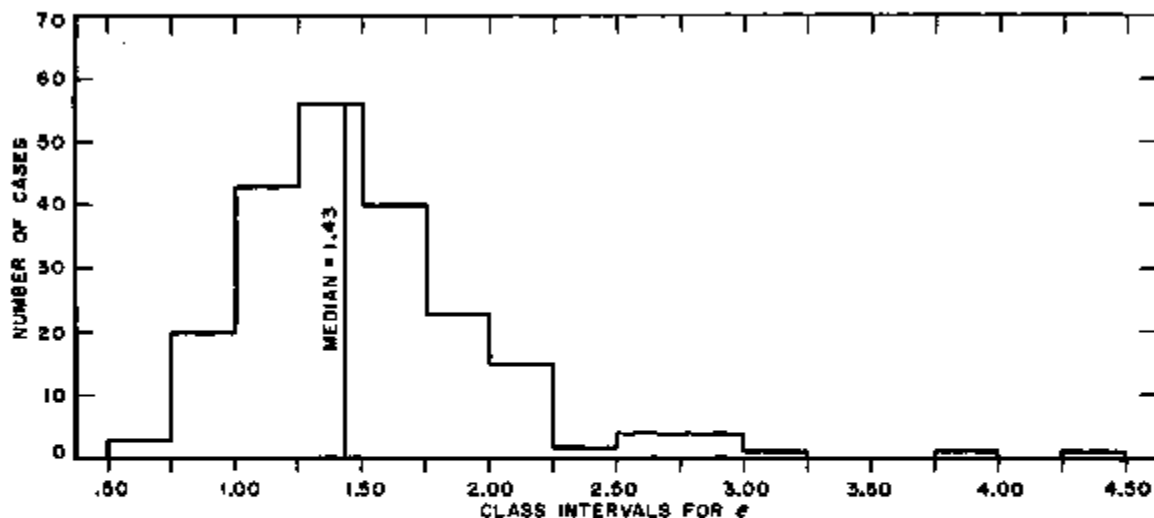


FIGURE 3. Histogram for the BIS estimates of parameter  $e$  for companies A, B, C, and D combined. Data in table B-a

### c. Worked Example

The example that will be given to illustrate the foregoing procedures will be the one that was worked out as a "test problem" for the SEAC before using the full set of data. The test group of bearings selected for this purpose was No. 1-1 in table A-1, for company A. The test group consisted of 24 ball bearings, of which 4 remained intact when the test was discontinued. The details of the computation for obtaining values of  $L_{10}$ ,  $L_{50}$ , and Weibull slope  $e$  from the test group of data are contained in table B-5 and described in the steps below.

The endurance lives in observed (increasing) order are listed in column (1). The arrows indicate the four "run-outs," or "intacts," whose testing was discontinued at the number of million revolutions indicated. All that is known about these four bearings is that their fatigue lives exceeded the values shown.

*Step 1. Randomization.* The order of endurance lives in column (1) was randomized by use of a set of random numbers generated in the SEAC as part of the computation work. The result is shown in column (2) of table B-5.

*Step 2. Subgroups.* The lives in randomized order were divided, as shown by the lines of separation, into subgroups of size  $n=6$ , the maximum size for which the order-statistics weights had been computed.<sup>14</sup> Each subgroup was then prepared for the application of the order-statistics method by rearranging in increasing order (column (3)). Natural logarithms were then taken as in column (4).

*Step 3. Weights.* Each subgroup was regarded as consisting of  $k$  actual observations out of a censored sample of  $n$ . It happened here that  $n$  was 6 for every subgroup;  $k$  took the values 6, 5, 6, 3. These values are shown in the subscripts of  $T_{n,r}$  written in column (3), and they determined the weights  $a_i$  and  $b_i$  to be selected from table B-2. These weights are represented in columns (5) and (6).

*Step 4. Cross-products.* The cross-products

$$T_{n,r} = \sum_{i=1}^k a_i x_i, \quad \sum_{i=1}^k b_i x_i$$

were then evaluated and placed as shown for each of the subgroups.

<sup>14</sup> As sample size 24 is an exact multiple of  $n=6$ , it so happened that there was no "remainder subgroup" in this case. This will not usually be true, but the procedure is identical for other values of  $n$ , differing merely in the numerical weights to be used.

Step 5. Estimates. A simple arithmetic average of the four values was taken for each of the two columns (5) and (6), and denoted by  $\bar{T}_1 = \hat{u}$ ,  $\bar{T}_2 = \hat{\beta}$ , respectively. These are the order-statistics estimates of the two parameters  $u$  and  $\beta$  of the extreme-value distribution that represents the underlying Weibull distribution.

The reciprocal of  $\hat{\beta}$  yields the Weibull slope  $\epsilon = 1.32497$ . (A)

The following logarithmic life estimates were given by following linear combinations of the estimates  $\hat{u}$  and  $\hat{\beta}$ , using the given values of  $y_{.05}$ ,  $y_{.50}$ :

$$x_{10} = \ln L_{10} = \hat{u} - 2.25037 \hat{\beta} = 2.982305$$

$$x_{50} = \ln L_{50} = \hat{u} - 0.36651 \hat{\beta} = 4.404120$$

$$\text{rating life } L_{10} = \text{antilog}(\hat{x}_{10}) \text{ (base } \epsilon) = 19.2333 \text{ million rev} \quad (\text{B})$$

$$\text{median life } L_{50} = \text{antilog}(\hat{x}_{50}) \text{ (base } \epsilon) = 81.7872 \text{ million rev.} \quad (\text{C})$$

These three values (A), (B), and (C) represent the outcome of the calculation.

In the full-scale computing program, calculations were carried out by the SEAC to a larger number of places than is shown in the table for presentation purposes. In general, however, the number of places shown here should be adequate. The values  $L_{10}$ ,  $L_{50}$ , and  $\epsilon$  shown here differ slightly from those recorded in table A-1 because the latter represent averages of two separate runs.

TABLE B-1. Means, variances, and covariances of order statistics  $y_i$  in samples of  $n$  from the reduced extreme-value distribution  $G(y) = \exp(-e^{-y})$ ,  $n = 2$  to 6

For distribution of largest values,  $y_1 \leq y_2 \leq \dots \leq y_n$   
For distribution of smallest values,  $y_1' \geq y_2' \geq \dots \geq y_n'$

n	i	Means* $E_i(y_i)$ $= -E_i(y_i')$	Variances and covariances,* $\sigma_{ii}' = \sigma_{ii} = \sigma_{ii} = \sigma_{ii}$					
			j=1	j=2	j=3	j=4	j=5	j=6
2	1	0.11593 152	0.68402 804	0.48045 301				
	2	-1.27038 285		1.64493 407				
3	1	0.40361 359	.44849 796	0.30137 144	0.24375 810			
	2	-.45943 263		.65852 235	.54629 438			
	3	-1.67582 795			1.64493 407			
4	1	0.57351 263	.34402 417	.22455 344	0.17903 454	0.15388 918		
	2	-.10608 352		.41553 113	.33720 966	.29271 188		
	3	-.81278 175			.65180 236	.57432 356		
	4	-1.96351 003				1.64493 407		
5	1	0.69016 715	.28486 447	.18202 536	.14358 737	0.12257 865	0.10901 329	
	2	.10689 454		.30849 748	.24676 731	.21226 644	.18967 383	
	3	-.42555 061			.40598 292	.35267 072	.31716 095	
	4	-1.07093 582				.64907 319	.58991 519	
	5	-2.18665 358					1.64493 407	
6	1	0.77729 368	.24658 20	.15496 74	.12121 61	.10291 64	0.09116 19	0.08285 42
	2	.25453 448		.24854 56	.19670 62	.16806 28	.14945 32	.13619 10
	3	-.18938 534			.29761 59	.25616 60	.22887 90	.20925 46
	4	-.66271 588				.40185 52	.36145 55	.33204 51
	5	-1.27594 579					.64769 96	.59985 67
	6	-2.36897 513						1.64493 41

\*The means are for smallest values (denoted by subscript  $S$ ); for largest values, change all signs and reverse order of  $y$ 's. The  $\sigma_{ij}$  are the same for both.

TABLE B-2. Weights  $w_i$  for the order-statistics estimator  $T_{n,k}$  for the parameter  $\tau_T = u + \beta_T$  of the extreme-value distribution (smallest values) from a censored sample of  $n = 2$  to  $6$ , where only the  $k$  smallest values are known

$$T_{n,k} = w_1x_1 + w_2x_2 + \dots + w_kx_k, w_i = a_i + b_ix_p$$

$$x_1 \leq x_2 \leq \dots \leq x_k, k = 2 \text{ to } n$$

$n$	$k$		$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	
2	2	$\{a_i$	0.0836289	0.9163731					
		$\{b_i$	-.7213476	.7213475					
3	2	$\{a_i$	-.3777001	1.3777001					
		$\{b_i$	-.8221012	0.8221012					
	3	$\{a_i$	.0879664	.2557135	0.6563201				
		$\{b_i$	-.3747251	-.2558100	.6305411				
4	2	$\{a_i$	-.7063194	1.7063194					
		$\{b_i$	-.8690149	0.8690149					
	3	$\{a_i$	-.0801057	.0604316	1.0198741				
		$\{b_i$	-.4143997	-.3258576	0.7402573				
	4	$\{a_i$	.0713800	.1536799	.2639426	0.5109975			
		$\{b_i$	-.2487965	-.2239192	-.0859035	.5586192			
	5	2	$\{a_i$	-.9598627	1.9598627				
			$\{b_i$	-.8962840	0.8962840				
3		$\{a_i$	-.2101141	-.0860231	1.2961372				
		$\{b_i$	-.4343419	-.3642463	0.7986882				
4		$\{a_i$	-.0153832	.0519642	.1520750	.8113440			
		$\{b_i$	-.2730342	-.2499429	-.1491094	.6720865			
5		$\{a_i$	.0583502	.1088236	.1676091	.2462831	0.4189341		
		$\{b_i$	-.1844826	-.1816564	-.1304534	-.0063354	.5031278		
6		2	$\{a_i$	-1.1655650	2.1655650				
			$\{b_i$	-0.9141358	0.9141358				
	3	$\{a_i$	-.3153968	-.2034315	1.5188283				
		$\{b_i$	-.4466018	-.3886492	0.8352510				
	4	$\{a_i$	-.0865378	-.0280534	.0649390	1.0496521			
		$\{b_i$	-.2858647	-.2654739	-.1858756	0.7372142			
	5	$\{a_i$	.0057311	.0465729	.1002523	.1722784	.6751653		
		$\{b_i$	-.2015431	-.1973753	-.1536040	-.0846894	.6170118		
	6	$\{a_i$	.0488669	.0835221	.1210527	.1656192	.2234909	0.3554481	
		$\{b_i$	-.1458072	-.1495332	-.1267277	-.0731937	.0359868	.4592751	

TABLE B-3. Variances  $Q_{n,k}\beta^2$  of order-statistics estimator  $T_{n,k}$ , given in table B-2, and its numerical values  $Q_{n,k}(10) = Q_{10}$ ,  $Q_{n,k}(50) = Q_{50}$  for estimators of parameters  $t_{50} = x_{50} = \ln L_{50}$ ,  $t_{10} = x_{10} = \ln L_{10}$ , respectively, for a censored sample of  $n = k$  to 6

Variances in units of  $\beta^2$

$$Q_{n,k} = A + 2By_p + Cy_p^2$$

$$x_1 \leq x_2 \leq \dots \leq x_n, k = 2 \text{ to } n$$

$n$	$k$	A	B	C	$Q_{10}$	$Q_{50}$
2	2	0.6595467	0.0643216	0.7118574	3.975015	0.708021
3	2	.9160386	.4682465	.8183654	2.952920	.682735
	3	.4028637	-.0247719	.3447117	2.260033	.467327
4	2	1.3340189	.7720298	.8670220	2.250056	.884572
	3	0.4331573	.1180273	.3022328	1.888278	.399329
	4	.2934587	-.0346903	.2252828	1.590460	.349150
5	2	1.7891720	1.0115594	.8950462	1.769068	1.167910
	3	0.5298953	0.2353740	.4168155	1.580861	0.412852
	4	.2918142	.0385708	.2537913	1.403458	.297633
	5	.2313953	-.0339905	.1866472	1.228307	.278697
6	2	2.2440055	1.2082248	.9132926	1.431164	1.481035
	3	0.6529409	0.2332488	.4321160	1.341381	0.466709
	4	.3237185	.1020223	.2697162	1.230430	.285165
	5	.2236063	.0105329	.1861069	1.118677	.240885
	6	.1911738	-.0313731	.1319601	1.000644	.231897

\*For  $Q_{10}B_p = -2.75987$ ; for  $Q_{50}B_p = -0.38651$ .

TABLE B-4. Efficiency of order-statistics estimator of logarithmic life  $x_{50} = \ln L_{50}$  and  $x_{10} = \ln L_{10}$  for complete samples ( $k = n$ ) of size  $n = 2$  to 6

$n$	Efficiency (in percent) with respect to	
	$x_{10} = \ln L_{10}$	$x_{50} = \ln L_{50}$
2	67.2	97.4
3	78.8	98.3
4	84.0	98.7
5	87.0	98.9
6	89.0	99.1

TABLE B-5. Example showing use of order statistics method of computing values of  $L_{10}$ ,  $L_{30}$ , and  $e$  from endurance data

Test group 1-1, company A

1	2	3	4	5	6
Endurance (million revolutions)			Natural logarithms $x_i$	Weights	
Observed order	Randomized order	Ascending order within subgroups		$a_i$	$b_i$
6.0	119.0	27.5	$x_1=3.31419$	0.048867	-0.145807
8.6	138.0	69.0	$x_2=4.23411$	.083522	-.149533
17.8	146.0	119.0	$x_3=4.77912$	.121053	-.126728
18.0	151.0	138.0	$x_4=4.92725$	.165619	-.073194
27.5	27.5	146.0	$x_5=4.98361$	.225491	.035987
33.5	69.0	151.0	$x_6=5.01728$	.355448	.459275
$T_{6,6}$ -----				$\sum_1^6 a_i x_i = 4.817310$	$\sum_1^6 b_i x_i = 0.400992$
50.5	(150.0)→	6.0	$x_1=1.79176$	0.005731	-0.201543
51.5	8.6	8.6	$x_2=2.15176$	.046573	-.197275
69.0	51.5	51.5	$x_3=3.94158$	.100252	-.153604
74.0	89.0	89.0	$x_4=4.48864$	.172278	-.064589
74.0	109.0	109.0	$x_5=4.61835$	.2675165	.617012
89.0	6.0	(150.0)→	( $x_6$ )		
$T_{6,5}$ -----				$\sum_1^5 a_i x_i = 4.446362$	$\sum_1^5 b_i x_i = 1.213655$
109.0	74.0	18.0	$x_1=2.89037$		
118.0	181.0	33.5	$x_2=3.51155$		
119.0	141.0	74.0	$x_3=4.30407$		
138.0	18.0	118.0	$x_4=4.77068$		
141.0	33.5	141.0	$x_5=4.94876$		
144.0	144.0	144.0	$x_6=4.96981$		
$T_{6,4}$ -----				$\sum_1^4 a_i x_i = 4.628081$	$\sum_1^4 b_i x_i = 0.619440$
146.0	17.8	17.8	$x_1=2.87920$	-0.315397	-0.446602
(150.0)→	(153.0)→	50.5	$x_2=3.92197$	-.203432	-.388649
151.0	(153.0)→	74.0	$x_3=4.30407$	1.518828	.835251
(153.0)→	(153.0)→	(153.0)→	( $x_4$ )		
(153.0)→	50.5	(153.0)→	( $x_5$ )		
(153.0)→	74.0	(153.0)→	( $x_6$ )		
$T_{6,3}$ -----				$\sum_1^3 a_i x_i = 4.831197$	$\sum_1^3 b_i x_i = 0.784853$
Over-all estimator-----				$\bar{T}_1 = 4.680738$	$\bar{T}_2 = 0.754735$
SUMMARY					
$\bar{T}_1 = \hat{\alpha} = 4.680738$ $\bar{T}_2 = \hat{\beta} = 0.754735$ $e = 1/\hat{\beta} = 1.32497$					
$y_{.10} = -2.25037$ $y_{.30} = -0.36651$					
$\hat{\alpha}_{10} = \ln L_{10} = \hat{\alpha} + y_{.10} \hat{\beta} = 2.982305, \quad L_{10} = 19.7333$					
$\hat{\alpha}_{30} = \ln L_{30} = \hat{\alpha} + y_{.30} \hat{\beta} = 4.404120, \quad L_{30} = 81.7872$					

[Weights are same as for first subgroup]

TABLE B-6. Tabulation of estimated values of Weibull slopes,  $e$ , for the 213 test groups of companies A, B, C, and D

Weibull slope, $e$	Company				Total
	A	B	C	D	
0.50 to 0.74	2	1			3
.75 to .99	6	11	2	1	20
1.00 to 1.24	16	22	3	2	43
1.25 to 1.49	11	40	5		56
1.50 to 1.74	7	32	1		40
1.75 to 1.99	6	17			23
2.00 to 2.24	2	12	1		15
2.25 to 2.49		2			2
2.50 to 2.74		4			4
2.75 to 2.99		4			4
3.00 to 3.24		1			1
3.25 to 3.49					
3.50 to 3.94					
3.75 to 3.99		1			1
4.00 to 4.26					
4.25 to 4.49		1			1
Total	50	148	12	3	213
Mean	1.33	1.60	1.31	1.04	1.51
Median	1.27	1.49	1.30	1.06	1.43
50% interval*	1.07 to 1.59	1.27 to 1.82	1.08 to 1.45	0.94 to 1.16	1.17 to 1.74

\*First quartile to third quartile.

## 6. Appendix C. Evaluation and Analysis of the Unknown Parameters in the Life Equation with Respect to Companies and Bearing Types

### 6.1. Summary

Equation (2a) of the main text,

$$L = \left[ \frac{f_c Z^{\alpha_1} D_a^{\alpha_2}}{P} \right]^p,$$

expresses the dependence of fatigue-life  $L$  on the design characteristics of the deep-groove bearing ( $Z$ ,  $D_a$ ), the bearing load  $P$ , and the "workmanship factor"  $f_c$ . This appendix outlines the statistical methods that were used (a) to determine "best" empirical values for the parameters  $f_c$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $p$  of this life formula, (b) to derive the associated intervals of uncertainty, and (c) to answer various questions about the values of these parameters, from the basic endurance data furnished by the ASA Subcommittee, which are summarized in appendix A. These methods are applied separately in each case to the rating life  $L_{10}$  and median life  $L_{50}$  values derived from the endurance test data as described in appendix B.

Section 6.2 summarizes the application of the statistical methods used to determine best empirical values and intervals of uncertainty for the parameters  $f_c$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $p$  in the case of deep-groove bearings. Sections 6.3 to 6.6 outline the statistical analyses employed to answer various questions about the values of  $f_c$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $p$  for the ball bearings of companies A, B, and C. In particular, section 6.3 gives the analysis employed to determine whether values of these four parameters are the same for the bearings of the three companies. This analysis is carried out separately for rating life  $L_{10}$  and median life  $L_{50}$ , and the postulated "complete between-companies homogeneity" is not supported in either instance. Section 6.4 gives the analysis appropriate to determining whether the data are consistent with the supposition that the value of  $p$  is the same for the three companies (regardless how the values of the other



parameters may differ); this analysis is applied to the  $L_{10}$  and  $L_{50}$  data, with an affirmative conclusion in both instances. It is concluded further that the data are consistent with the supposition that this common value for  $p$  is equal to 3.

Section 6.5 presents the analysis employed to determine whether the values of the parameters  $f_c$ ,  $a_1$ ,  $a_2$ , and  $p$  are the same for the three types of deep-groove bearings (B-1, B-2, and B-3) for which the data from company B were available. A negative conclusion is reached. These data are then reanalyzed to determine whether they are consistent with the supposition that  $p=3$  for each of the three types, regardless of differences in the other parameters. An affirmative conclusion is reached in this case.

Section 6.6 concerns (a) the extent to which the  $L_{10}$  and  $L_{50}$  values are consistent with the supposition that the values of  $a_1$ ,  $a_2$ , and  $p$  are 2/3, 1.8, and 3, respectively, as given in [15]; and (b), the determination of more precise values for  $f_c$  in those cases in which the foregoing supposition is supported.

## 6.2. Determination of "Best" Values for the Parameters and Their Associated Intervals of Uncertainty

As shown in section 2.2 of the main text, if natural logarithms are taken of both sides of the life equation (eq (2a)), the resulting equation expresses the logarithm of rating life  $L_{10}$  (or median life  $L_{50}$ ) as a linear function of the logarithms of the characteristics of the bearing ( $Z, D_a$ ), and the bearing load  $P$ , with coefficients that are simple functions of the "workmanship factor"  $f_c$  and the exponents,  $a_1$ ,  $a_2$ , and  $p$ ; i. e.,

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \quad (C1)$$

where

$$\left. \begin{aligned} Y &= \ln L \\ x_1 &= \ln Z \\ x_2 &= \ln D_a \\ x_3 &= \ln P \end{aligned} \right\} \quad (C2)$$

and

$$\left. \begin{aligned} b_0 &= p \ln f_c = p a_0 \\ b_1 &= p a_1 \\ b_2 &= p a_2 \\ b_3 &= -p \end{aligned} \right\} \quad (C3)$$

are unknown constants to be estimated from the data.

The variables  $x_1$ ,  $x_2$ , and  $x_3$  are fixed variates. Their values are uniquely determined by the design of the bearing and the bearing loads that are used in the tests. The variable  $Y$ , on the other hand, denotes the mean values of  $\ln L_{10}$ , or  $\ln L_{50}$ , for the population of all bearings with characteristics  $x_1$  and  $x_2$ , tested at load  $x_3$ .

In the practical situation  $Y$  is never known, but must be determined from the results of endurance tests. The methods used for obtaining such estimates of  $Y$  from endurance-test data are given in appendix B. To distinguish  $Y$  from an empirical estimate of it, the estimate will be denoted by the lower-case letter  $y$ .

Generally speaking, an estimate  $y$  is a random variable, having a probability distribution that depends on  $x_1$ ,  $x_2$ , and  $x_3$ . We assume that the *mean* of this distribution is  $Y = Y(x_1, x_2, x_3)$ , and that its dispersion, or more precisely, its *variance*, is inversely proportional to the number of bearings  $w$  in the test group from which the estimate  $y = y(x_1, x_2, x_3; w)$  is derived (cf.

appendix B); that is,

$$\left. \begin{array}{l} \text{mean of } y=Y \\ \text{variance of } y=\frac{\sigma^2}{w} \end{array} \right\} \quad (C4)$$

where  $\sigma^2$  denotes some positive constant, and  $w$  is the number of bearings in the test group.

The statistical methods used to estimate the unknown parameters  $b_0$ ,  $b_1$ ,  $b_2$ , and  $b_3$  from the data are termed regression techniques. The books by Anderson and Bancroft [1], Dixon and Massey [3], Hald [10], Kempthorne [11], Mood [16], and Wilks [20] give extensive discussion of these techniques. For completeness, some of the techniques and rationale of regression analysis bearing on the work embodied in this report are summarized below. More detailed discussions can be found in the above references.

*Estimation.* The problem of estimating the unknown parameters in the life equation can be stated as follows: Given *independent* observations  $(y_\alpha; x_{1\alpha}, x_{2\alpha}, x_{3\alpha}; w_\alpha)$  from  $n$  test groups ( $\alpha=1, 2, \dots, n$ ) where  $y_\alpha$  is the estimate of  $\ln L_{10}$  or  $\ln L_{20}$ ,  $x_{1\alpha}$ ,  $x_{2\alpha}$  are the logarithms of the bearing characteristics  $Z$  and  $D_0$ ,  $x_{3\alpha}$  is the logarithm of the load, and  $w_\alpha$  is the number of individual bearings tested for the  $\alpha$ th test group; required to estimate the values of the parameters  $b_0$ ,  $b_1$ ,  $b_2$ , and  $b_3$ , in eq (C1) using some optimum method of estimation.

Estimates for the  $b_i$  ( $i=0, 1, 2, 3$ ) that are free of systematic error and have smaller variances than any other linear unbiased estimates are obtained by minimizing the quadratic form,

$$Q = \sum_{\alpha=1}^n w_\alpha (y_\alpha - b_0 - b_1 x_{1\alpha} - b_2 x_{2\alpha} - b_3 x_{3\alpha})^2, \quad (C5)$$

with respect to each of the  $b_i$  ( $i=0, 1, 2, 3$ ).

The resulting normal equations defining the parameter estimates can be written in the form,<sup>15</sup>

$$\left. \begin{array}{l} a_{00}\hat{b}_0 + a_{01}\hat{b}_1 + a_{02}\hat{b}_2 + a_{03}\hat{b}_3 = g_0 \\ a_{10}\hat{b}_0 + a_{11}\hat{b}_1 + a_{12}\hat{b}_2 + a_{13}\hat{b}_3 = g_1 \\ a_{20}\hat{b}_0 + a_{21}\hat{b}_1 + a_{22}\hat{b}_2 + a_{23}\hat{b}_3 = g_2 \\ a_{30}\hat{b}_0 + a_{31}\hat{b}_1 + a_{32}\hat{b}_2 + a_{33}\hat{b}_3 = g_3 \end{array} \right\} \quad (C6)$$

where

$$a_{ij} = \sum_{\alpha=1}^n w_\alpha x_{i\alpha} x_{j\alpha} \quad i, j = 0, 1, 2, 3. \quad (C7)$$

$$g_i = \sum_{\alpha=1}^n w_\alpha x_{i\alpha} y_\alpha \quad i = 0, 1, 2, 3. \quad (C8)$$

and  $x_{0\alpha} = 1$  for all  $\alpha$ .

If one defines the *residual sum of squares* by

$$S = \sum_{\alpha=1}^n w_\alpha \left[ y_\alpha - \sum_{i=0}^3 \hat{b}_i x_{i\alpha} \right]^2,$$

which also can be written in the alternative form

$$S = \sum_{\alpha=1}^n w_\alpha y_\alpha^2 - \sum_{i=0}^3 \hat{b}_i g_i, \quad (C9)$$

then the optimum estimate of  $\sigma^2$  is

$$s^2 = \frac{S}{n-4}. \quad (C10)$$

<sup>15</sup> A caret (^) is used here to distinguish the fact that the solutions of the normal equations are estimated values of the parameters and not the parameters themselves.

The quantity  $(n-4)$  is the rank of the quadratic form  $S$ , and is termed the *degrees of freedom* associated with  $S$ .

The above method of estimation does not depend on  $y_a$  having a particular assumed probability distribution. All that is necessary to specify about the probability distribution of  $y_a$  is that it possess a finite mean and variance. The solution of the normal equations (C3) has the property that the weighted sum of squares of the deviations about  $y_a$  will be a minimum. This is a so-called "least-square" property of the solutions, but it is only a consequence of the method and is not the justification for using this method of estimation. The justification for the method is that this is the only one that results in minimum variance unbiased estimates for the  $b_i$  ( $i=0, 1, 2, 3$ ).

Values for  $b_i$  ( $i=0, 1, 2, 3$ ), and hence for  $a_i$  and  $p$ , can be obtained by using either the  $\ln L_{10}$  or  $\ln L_{50}$  values for  $y_a$ . In all cases where the  $a_i$  and  $p$  have been obtained for rating life  $L_{10}$ , another set of parameters have also been calculated for median life  $L_{50}$ .

*Intervals of uncertainty and inferences.* The methods so far discussed for finding estimates of unknown parameters need no assumption as to the form of the underlying probability distribution of  $y_a$ . However, something more must be assumed about the distribution of  $y_a$  if (a) one wishes to place an interval about an estimate of a parameter that will include the "true" (or population) value of the parameter with given assurance, or (b) if one desires to make inferences about the parameters of the life equation for the population from which the bearings are a sample.

Although the endurance lives for individual bearings may follow a Weibull distribution, the distribution of  $y_a$  will not be of this form. However, the estimate  $y_a$  (cf. appendix B) is an *average* of several independent estimates, each based on linear functions of six or less order statistics. Hence, by the central limit theorem, the distribution of the estimate  $y_a$  will be approximated by a normal distribution when  $n$  is large (cf. Cramér [2, p. 213]). The statistical tests of significance used in this report are not greatly affected by moderate departures from normality. Therefore, for making all inferences, it will be further assumed that the estimates  $y_a$  follow a normal distribution.

The intervals of uncertainty calculated for each parameter are 95-percent confidence limits (equivalent to the usual "two-sigma" limits), which were referred to in the main text. Confidence limits for the parameter  $p = -b_3$  can be calculated by using conventional methods. However the confidence limits for  $a_0, a_1,$  and  $a_2$  are somewhat more complicated. The method used here, for this purpose, is sometimes referred to as Fieller's theorem (cf. Fieller [4]).

In order to make inferences about the parameters with respect to the different companies or bearing types, certain statistical tests of significance were used in this paper. These are all based on a test statistic  $F$ , termed the variance ratio or  $F$ -ratio, which takes the form

$$F = \frac{Q_1/v_1}{Q_2/v_2},$$

where  $Q_1$  and  $Q_2$  are quadratic forms calculated from the data and  $v_1, v_2$  are the respective ranks of the quadratic forms. The explicit expressions for  $Q_1$  and  $Q_2$  depend upon the particular hypothesis being tested. The subsequent sections that employ a variance-ratio statistic also give the explicit expressions for the two quadratic forms.

If the hypothesis being tested is true, then the calculated variance ratio will deviate from unity in accordance with its tabulated distribution. However, if the hypothesis is false, then the variance ratio will be substantially greater than unity, and the "more false" the stated hypothesis, the larger the value for the variance ratio. Thus when the hypothesis tested is false, this will be detected by an abnormally large  $F$ -ratio. In order to objectively determine whether a calculated variance ratio is *significantly* greater than unity, one selects from tables of the variance-ratio distribution a *critical value* of  $F$ , such that there is only a small probability of the calculated variance ratio exceeding the critical value from purely chance causes. The critical value for  $F$  used for all variance-ratio tests, in this paper, has been selected so that there

is only a probability of .05 of its being exceeded by a calculated variance ratio from purely chance causes. This critical value will be denoted by  $F_{.05}(v_1, v_2)$ . More extensive discussions of these procedures are given in Kempthorne [11, chap. 5].

All statistical tests of significance to test relevant hypotheses have been carried out both for rating life  $L_{10}$  and median life  $L_{50}$ .

### 6.3. Analysis to Determine Whether Companies Have Common Values for All the Parameters in the Life Equation

This section deals with the details of estimating the values of the parameters in the life equation for each company. Furthermore, a statistical analysis is made to determine whether the companies have common values for all the parameters in the life equation. The  $F$ -ratio (eq (C16)), which is used to test this hypothesis, is obtained from the following procedures: A single set of parameters,  $\hat{b}_i$  ( $i=0, 1, 2, 3$ ), is obtained by first fitting all the data, irrespective of company, to the logarithmic life equation, and then calculating the resulting residual sums of squares  $S$  (eq (C14)) having 206 degrees of freedom. If the hypothesis of common values for all the parameters is not true, then a better fit to the data can be made by fitting the life equation separately to each company. These calculations result in the individual residual sums of squares  $S_1$ ,  $S_2$ , and  $S_3$  (eq (C13)) having 46, 144, and 8 degrees of freedom, respectively. Thus the total residual sums of squares ( $S_1 + S_2 + S_3$ ) will have  $46 + 144 + 8 = 198$  degrees of freedom. Then the difference between  $S$  and ( $S_1 + S_2 + S_3$ ), i. e.,  $\{S - (S_1 + S_2 + S_3)\}$ , is also a quadratic form having  $206 - 198 = 8$  degrees of freedom. If a substantially better fit was obtained by fitting a separate life equation to the data for each company, as compared to a single life equation, the difference between the two residual sums of squares  $\{S - (S_1 + S_2 + S_3)\}$  will be large. To determine whether this difference is statistically significant the variance ratio (C16) is employed.

*Mathematical formulation.* It will be convenient to adopt the following notation: Let the superscript ( $u$ ) = 1, 2, 3 refer to companies A, B, and C, respectively. Also, for each company, let  $\hat{b}_i^{(u)}$  ( $i=0, 1, 2, 3$ ) refer to the estimates of the parameters in eq (C1), and  $a_{ij}^{(u)}$ ,  $g_i^{(u)}$  denote the sums of cross products defined in eq (C7) and (C8)<sup>14</sup>. Then the normal equations that give the parameter estimates for the  $u$ th company are

$$\sum_{j=0}^3 a_{ij}^{(u)} \hat{b}_j^{(u)} = g_i^{(u)} \quad i=0, 1, 2, 3, \quad (C11)$$

and the estimates for the parameters  $a_i$  ( $i=0, 1, 2$ ) and  $p$ , in the life equation, are obtained from the relations given by eq (C3). These results are summarized in tables 6 and 3, respectively, in the main text.

The hypothesis that all parameters in the life equation are the same for each company is equivalent to the hypothesis that

$$\hat{b}_i^{(1)} = \hat{b}_i^{(2)} = \hat{b}_i^{(3)} \quad i=0, 1, 2, 3. \quad (C12)$$

Define the residual sum of squares for the  $u$ th company by

$$S_u = \sum_{\alpha=1}^{n_u} w_{u\alpha} y_{u\alpha}^2 - \sum_{i=0}^3 \hat{b}_i^{(u)} g_i^{(u)} \quad u=1, 2, 3, \quad (C13)$$

and let

$$S = \sum_{u=1}^3 \sum_{\alpha=1}^{n_u} w_{u\alpha} y_{u\alpha}^2 - \sum_{i=0}^3 \hat{b}_i g_i, \quad g_i = \sum_{u=1}^3 g_i^{(u)}, \quad (C14)$$

where  $\hat{b}_i$  ( $i=0, 1, 2, 3$ ) are the estimates obtained from the solutions of the normal equations

<sup>14</sup>As it is only the ratio of the numbers of observations that is important for weighting, the weights  $w$  used in the calculation of the sums of cross products in all the analyses have been taken as integral multiples of 5; e. g., if the number of bearings in a test group is 26 (say), then  $w=5$ .

without regard to company differences, i. e.,

$$\sum_{j=0}^3 \left[ \sum_{u=1}^3 a_i^{(u)} \right] \hat{b}_j = \sum_{u=1}^3 g_i^{(u)} \quad i=0,1,2,3. \quad (C15)$$

Then the variance ratio for testing the hypothesis given by eq (C15) is

$$F = \frac{(S - S_1 - S_2 - S_3)/8}{(S_1 + S_2 + S_3)/198}, \quad (C16)$$

and the critical  $F$  value is  $F_{.05}(8, 198) = 1.98$ .

Table C-1 summarizes the values for  $S_u$  ( $u=1,2,3$ ) and the other quantities needed to evaluate eq (C16).

TABLE C-1. Synopsis of calculations to determine whether companies have common values for all parameters in life equation

	Degrees of freedom	$L_{10}$		$L_{50}$	
		$S_u$	Mean square	$S_u$	Mean square
Companies combined.....	206	446.684699	-----	248.285324	-----
A.....	46	102.552212	-----	89.734079	-----
B.....	144	235.121982	-----	201.935797	-----
C.....	8	6.160858	-----	8.357822	-----
Sum.....	198	393.835082	1.9801	300.027698	1.5153
Difference.....	8	52.849601	6.6062	48.257626	6.0322
From eq (C16).....		$F = \frac{6.6062}{1.9891} = 3.32$		$F = \frac{6.0322}{1.5153} = 3.98$	

The values of the  $F$ -ratio (eq (C16)) calculated from the quantities in table C-1 are

$$L_{10}: F=3.32$$

$$L_{50}: F=3.98$$

These calculated  $F$  values are both larger than the critical value,  $F_{.05}(8, 198) = 1.98$ ; actually the probability is less than .0005 of having an  $F$ -ratio as large as those above from purely chance causes. Thus from the above statistical tests of significance, the conclusion is reached that *the three companies do not have common values for all of the parameters in the life equation.*

#### 6.4. Analysis to Determine Whether Companies Have a Common Value for the Exponent $p$

The previous analysis resulted in the conclusion that the parameters in the life equation are different for each company. However, this does not exclude the possibility that all companies may have a common  $p$ , even though the  $a_i$  ( $i=0,1,2$ ) differ from company to company. This section discusses the analysis made to determine whether the companies have a common value for the exponent  $p$ . The analysis given here consisted of the following procedure. First, the logarithmic life equation (C17), having a common value of  $p$ , but allowing the  $a_i$  to vary for each company, was fitted to all the data, and the resulting residual sum of squares  $S'$  (eq (C20)), having 200 degrees of freedom, was calculated. The total residual sum of squares from fitting the life equation separately to each company (allowing  $p$  to vary in addition to the  $a_i$ ) is given by  $(S_1 + S_2 + S_3)$  having 198 degrees of freedom (cf. section 6.3 of this appendix).

Then the reduction in the residual sum of squares achieved by using a different exponent  $p$  for each company is  $\{S' - (S_1 + S_2 + S_3)\}$  having  $200 - 198 = 2$  degrees of freedom. To test whether this reduction in the residual sums of squares is statistically significant, the variance ratio (C21) is employed.

*Mathematical formulation.* The logarithmic life equation, having a common value for the exponent  $p$ , can be written for the  $\alpha$ th test group in the  $u$ th company as

$$Y_{\alpha u} = b_0^{(u)} + b_1^{(u)} x_{1\alpha}^{(u)} + b_2^{(u)} x_{2\alpha}^{(u)} + b_3 x_{3\alpha}^{(u)} \quad \alpha = 1, 2, \dots, n_u; u = 1, 2, 3. \quad (C17)$$

Note that although each company has the same parameter  $b_3$  in the above equation, the parameters  $b_0^{(u)}$ ,  $b_1^{(u)}$ , and  $b_2^{(u)}$  are different for each of the three companies. Thus there [are 10 different parameters, i. e.,  $b_i^{(u)}$  ( $i=0,1,2; u=1,2,3$ ) and  $b_3$ , to be estimated from the data.

The normal equations for estimating these parameters are

$$\left. \begin{aligned} \sum_{j=0}^2 a_{ij}^{(u)} \hat{b}_j^{(u)} + a_{i3}^{(u)} \hat{b}_3 &= g_i^{(u)} \quad i=0, 1, 2; u=1, 2, 3 \\ \sum_{u=1}^3 \sum_{j=0}^2 a_{3j}^{(u)} \hat{b}_j^{(u)} + a_{33} \hat{b}_3 &= g_3 \end{aligned} \right\} \quad (C18)$$

where

$$a_{zz} = \sum_{u=1}^3 a_{iz}^{(u)} \quad g_3 = \sum_{u=1}^3 g_3^{(u)}.$$

Thus the set of equations given by (C18) is a system of 10 linear equations in 10 unknowns. Once the solutions are obtained, the estimates for  $a_i$  ( $i=0,1,2$ ) in the life equation are calculated from the relationship

$$\hat{a}_i^{(u)} = -\frac{\hat{b}_i^{(u)}}{\hat{b}_3} \quad i=0, 1, 2; u=1, 2, 3. \quad (C19)$$

The residual sums of squares (denoted by  $S'$ ) associated with fitting the life formula (C17) to the data is given by

$$S' = \sum_{u=1}^3 \sum_{\alpha=1}^{n_u} w_{\alpha u} y_{\alpha u}^2 - \sum_{u=1}^3 \sum_{i=0}^2 \hat{b}_i^{(u)} g_i^{(u)} - \hat{b}_3 g_3. \quad (C20)$$

Then, to test the hypothesis that the companies have a common value of  $p$ , regardless of the values for the other parameters in the life equation, the variance ratio

$$F = \frac{(S' - S_1 - S_2 - S_3)/2}{(S_1 + S_2 + S_3)/198}, \quad (C21)$$

having 2 and 198 degrees of freedom is used. The critical  $F$  value is  $F_{.05}(2, 198) = 3.04$ .

The values of  $S'$  for both  $L_{10}$  and  $L_{30}$  are

$$\left. \begin{aligned} L_{10}: \quad S' &= 393.272847 \quad \text{d. f.} = 200 \\ L_{30}: \quad S' &= 301.687871 \quad \text{d. f.} = 200, \end{aligned} \right\} \quad (C22)$$

and the calculated <sup>17</sup>  $F$  values (using eq (C21)) give

$$\left. \begin{aligned} L_{10}: \quad F &= -0.141 \\ L_{30}: \quad F &= 0.548. \end{aligned} \right\} \quad (C23)$$

<sup>17</sup> From theory, the calculated value for the  $F$ -ratio can never be negative. The reason for the negative value of  $F$  for  $L_{10}$  is that the value for the numerator of eq (C21) is only accurate numerically to one decimal place on account of round-off errors arising from the solution of the normal equations (C18). Thus, if the hypothesis of a common  $p$  value is true, then the  $F$ -ratio will not be large and round-off errors may affect the resulting calculation. Alternatively, if the null hypothesis is false, then the calculated  $F$ -ratio will be larger than 3.04 and the round-off error should be of no consequence.

Because both variance ratios are smaller than the critical value  $F_{.05}(2, 198)=3.04$ , the conclusion drawn from this statistical analysis is that the data support the hypothesis of a common value of  $p$  for the three companies. This holds both for rating life ( $L_{10}$ ) and median life ( $L_{50}$ ). The values for the common  $p$  are given in table 2 of the main text. The values for the remaining constants in the life equation  $a_i (i=0,1,2)$  are found from the relations (C19). These results are summarized in table 5 of the main text.

### 6.5. Analysis to Determine Whether the Three Bearing Types From Company B Have Consistent Parameter Values

The analyses, discussed in previous sections, dealt with determining whether there are differences in the parameters of the life equation between companies. This section investigates (a) whether three different bearing types made by company B have common values for all the parameters in the life equation, and (b) whether the exponents  $p$  calculated for each bearing type are consistent with the value of  $p=3$ .

The analysis for (a) is similar to the analysis made in section 6.3 of this appendix; i. e., separate life equations were fitted to each bearing type and the resulting residual sum of squares was compared with the residual sum of squares arising from fitting a single equation to all data from company B, irrespective of bearing type. The variance ratio for statistically testing (a) is given by eq (C27).

The analysis for (b) was governed by the following considerations. If the true (or population) value of the exponent  $p$  is  $p=3$ , regardless of bearing type, then the estimates for  $p$  obtained by fitting a separate life equation to each bearing type should not differ from  $p=3$  by more than the dispersion inherent in the endurance lives of the bearings. The agreement of the values of  $p$  estimated for each bearing type with  $p=3$  is tested for statistical significance by the variance ratio (C28).

*Mathematical formulation.* The 148 test groups from company B can be divided into three bearing types corresponding to 37 groups for B-1, 94 groups for B-2, and 17 groups for B-3-type bearings. Let these types be denoted by  $v=1, 2, 3$ , respectively. Also define

$$\left. \begin{aligned} A_i^{(v)} &= \sum_{a=1}^{n_v} w_{ra} x_{ia}^{(v)} x_{ja}^{(v)} & i, j=0, 1, 2, 3; v=1, 2, 3 \\ G_i^{(v)} &= \sum_{a=1}^{n_v} w_{ra} x_{ia}^{(v)} y_{ia}^{(v)} & i=0, 1, 2, 3; v=1, 2, 3 \end{aligned} \right\} \quad (C24)$$

where  $n_v$  is the number of test groups for bearing type  $v$ . Then the normal equations, which determine the estimates for the parameters in the logarithmic life equation, are

$$\sum_{j=0}^3 A_i^{(v)} \hat{b}_j^{(v)} = G_i^{(v)} \quad i=0, 1, 2, 3; v=1, 2, 3. \quad (C25)$$

Thus for each bearing type, the values for the parameters  $a_i (i=0,1,2)$  and  $p$  in the life equation can be estimated from the relations

$$\hat{a}_i^{(v)} = -\frac{\hat{b}_i^{(v)}}{\hat{b}_3^{(v)}}$$

$$\hat{p}^{(v)} = -\hat{b}_1^{(v)}$$

Therefore the residual sum of squares for the  $v$ th bearing type is

$$S^{(v)} = \sum_{a=1}^{n_v} w_{v,a} y_{v,a}^2 - \sum_{i=0}^3 \hat{\delta}_i^{(v)} G_i^{(v)} \quad v=1, 2, 3, \quad (C26)$$

having  $(n_v - 4)$  degrees of freedom. Then the hypothesis of common parameters for the three bearing types can be tested by the variance ratio,

$$F = \frac{(S_1 - S^{(1)} - S^{(2)} - S^{(3)})/8}{(S^{(1)} + S^{(2)} + S^{(3)})/136}, \quad (C27)$$

having a critical value of  $F_{.05}(8, 136) = 2.01$ .

Since the analysis given in section 6.4 of this appendix reached the conclusion that all companies have a common value for  $p$ , and since this value (cf. table 2 of main text) with its associated uncertainty includes the value  $p=3$  given in [15], it seems also desirable to test a second hypothesis that the value of  $p$  for each bearing type is consistent with  $p=3$ . The  $F$ -ratio for this hypothesis is given by

$$F = \frac{\frac{1}{3} \left[ \frac{(\hat{p}^{(1)} - 3)^2}{C_{33}^{(1)}} + \frac{(\hat{p}^{(2)} - 3)^2}{C_{33}^{(2)}} + \frac{(\hat{p}^{(3)} - 3)^2}{C_{33}^{(3)}} \right]}{(S^{(1)} + S^{(2)} + S^{(3)})/136}, \quad (C28)$$

where the  $p^{(v)}$  ( $v=1,2,3$ ) refer to the estimates of  $p$  obtained for each bearing type, and  $C_{33}^{(v)}$  ( $v=1,2,3$ ) is the element occurring in the last row, last column of the inverse matrix to  $\|A_i^{(v)}\|$  ( $v=1,2,3$ ). The critical value for the variance ratio (C28) is  $F_{.05}(3, 136) = 2.67$ .

The estimates of  $a_i^{(v)}$  ( $i=0, 1, 2$ ) and  $p^{(v)}$  obtained from the solutions of the normal equations (C25) are summarized in tables 7 and 4, respectively, in the main text. The calculations for the variance ratio (C27) are summarized in table C-2.

Corresponding to the hypothesis that the three bearing types have the same parameters in the life equation, the calculated variance ratios (C27) yield

$$\left. \begin{array}{l} L_{10}: F=2.05 \\ L_{20}: F=2.88. \end{array} \right\} \quad (C29)$$

TABLE C-2.—Synopsis of calculations to determine whether all bearing types have common parameters in life equation (company B only)

Type of bearing	Degrees of freedom	$L_{10}$		$L_{20}$	
		$S^{(v)}$	Mean square	$S^{(v)}$	Mean square
All types combined.....	*144	*285. 121962	.....	*201. 935797	.....
B-1.....	33	68. 558298	.....	33. 016973	.....
B-2.....	90	174. 325939	.....	131. 142017	.....
B-3.....	12	11. 537480	.....	8. 505638	.....
Sum.....	136	254. 421716	1. 8707	172. 664628	1. 2696
Difference.....	8	30. 700246	3. 8375	29. 271169	3. 6589
From eq (C27).....			$F = \frac{3.8375}{1.8707} = 2.05$		$F = \frac{3.6589}{1.2696} = 2.88$

\*From table C-1.



Because the critical  $F$  value is  $F_{.05}(8, 136)=2.01$ , both the  $L_{10}$  and  $L_p$  calculated variance ratios are statistically significant. Therefore, one could conclude from the above  $F$ -ratios that the data support the hypothesis that parameters do differ between bearing types. However, this does not exclude the possibility that the values for the exponent  $p$  are consistent with the value  $p=3$ . Substituting the appropriate quantities in eq (C28) results in

$$\left. \begin{array}{l} L_{10}: F=1.18 \\ L_{50}: F=2.32 \end{array} \right\}, \quad (C30)$$

where

$$C_{11}^{(3)}=0.53294$$

$$C_{22}^{(3)}=0.107395$$

$$C_{33}^{(3)}=0.396744.$$

Thus, as both of the above calculated  $F$ -ratios are less than the critical value,  $F_{.05}(3, 136)=2.67$ , the conclusion can be made that the values of  $p$  are consistent with  $p=3$  for different bearing types, although possibly some (or all) of the other parameters in the life formula (i. e.,  $a_i (i=0, 1, 2)$ ) may differ among the different bearing types.

The values for  $\hat{a}_i^{(v)}$  ( $i=0, 1, 2$ ;  $v=1, 2, 3$ ) arising from the analysis by ball-bearing types have very large confidence limits (intervals of uncertainty). This is mainly due to the fact that an analysis restricted to one bearing type is essentially an analysis on bearings having almost the same values for  $Z$  and  $D_n$ . In order to estimate the  $a_i^{(v)}$  with good precision, it is necessary to have results for bearings having wide variations with respect to  $Z$  and  $D_n$ . Thus the estimates for  $a_i$  based on all bearing types for company B (table 5 or table 6) have substantially smaller confidence intervals as compared to the intervals based only on a single bearing type.

#### 6.6. Determination and Analysis of $f_c$ Based on the Parameter Values

$$a_1=2/3, a_2=1.8, \text{ and } p=3$$

The values for the parameters  $a_1$ ,  $a_2$ , and  $p$  given in [15] are  $a_1=2/3$ ,  $a_2=1.8$ , and  $p=3$ . If these parameter values are valid for the data at hand, then more precise estimates for the "workmanship" parameter  $a_0$  (or  $f_c$ ) can be made for each company or bearing type. These will generally have better precision compared to the estimates of  $a_0$  made when the other parameters in the life equation are simultaneously estimated along with  $a_0$ . This section considers the problem of verifying whether the parameter values, given above, are valid for the given data, and for those cases where this is true, estimates of  $a_0$  (or  $f_c$ ) are obtained assuming these values for the other parameters.

The procedure for determining whether the values  $a_1=2/3$ ,  $a_2=1.8$ , and  $p=3$  are valid for a given classification of the data (with respect to a company or bearing type) is to fit the data to the life equation, using the assumed values for  $a_1$ ,  $a_2$ , and  $p$ . Thus there is only one unknown parameter,  $a_0$ , in the life equation to be estimated. Then the resultant residual sum of squares, denoted by  $R$  (eq (C37)), can be calculated having  $(n-1)$  degrees of freedom. Alternatively, the life equation can be fitted to the data such that all the unknown parameters are simultaneously estimated. The residual sum of squares from this latter fit,  $S$ , will have  $(n-4)$  degrees of freedom. Then, if the above parameter values are not consistent with the given data,  $R$  will be appreciably larger than  $S$ . The variance ratio (C38) is used to determine if the difference between these two residual sums of squares, i. e.,  $(R-S)$  having  $[(n-1)-(n-4)]=3$  degrees of freedom, is statistically significant.

*Mathematical formulation.* Let  $n$  be the number of test groups within a particular classification (either by company or bearing type). Then, assuming the values  $a_1=2/3$ ,  $a_2=1.8$ , and

$p=3$ , the logarithmic life equation for the  $\alpha$ th test group can be written as

$$Y_\alpha = 3 \{ a_0 + x_\alpha \} \quad \alpha = 1, 2, \dots, n, \quad (C31)$$

where

$$x_\alpha = \left[ \frac{2}{3} x_{1\alpha} + 1.8 x_{2\alpha} - x_{3\alpha} \right], \quad (C32)$$

The resulting normal equation for estimating  $a_0$  is

$$\hat{a}_0 = \frac{\sum_{\alpha=1}^n \frac{w_\alpha y_\alpha}{3} - \sum_{\alpha=1}^n w_\alpha x_\alpha}{\sum_{\alpha=1}^n w_\alpha}, \quad (C33)$$

which also can be written as a function of the sums of cross products  $g_i, a_{ij}$ , e. g.,

$$\hat{a}_0 = \frac{1}{a_{00}} \left\{ \frac{g_0}{3} - \left( \frac{2}{3} a_{01} + 1.8 a_{02} - a_{03} \right) \right\}. \quad (C34)$$

Hence the estimated variance of  $\hat{a}_0$  is

$$\text{variance } (\hat{a}_0) = \frac{s^2}{9a_{00}}, \quad (C35)$$

where

$$s^2 = \frac{R}{n-1} = \frac{9 \left[ \sum_{\alpha=1}^n w_\alpha \left( \frac{y_\alpha}{3} - x_\alpha \right)^2 - a_{00} \hat{a}_0^2 \right]}{n-1}. \quad (C36)$$

The residual sum of squares  $R$ , having  $(n-1)$  degrees of freedom, can also be written as a function of the sums of cross products,

$$R = 9 \left\{ \frac{1}{9} \sum_{\alpha=1}^n w_\alpha y_\alpha^2 - \frac{4}{9} g_1 - 1.2 g_2 + \frac{2}{3} g_3 + \frac{4}{9} a_{11} + 3.24 a_{22} + a_{33} + 2.4 a_{12} - \frac{4}{3} a_{13} - 3.6 a_{23} - a_{00} \hat{a}_0^2 \right\}. \quad (C37)$$

In the analyses made in the preceding sections, the ball-bearing data have been analyzed with respect to individual companies or bearing types. It thus seems desirable to determine whether the data within these classifications support the hypothesis that  $a_1=2/3$ ,  $a_2=1.8$ , and  $p=3$ . The variance ratio used to test this hypothesis is

$$F = \frac{(R-S)/3}{S/(n-4)}, \quad (C38)$$

where, for testing within companies,

$$S = \left\{ \begin{array}{l} S_1 \text{ for company A} \\ S_2 \text{ for company B} \\ S_3 \text{ for company C} \end{array} \right\} \text{ defined by eq (C13) and given in table C-1,}$$

and for testing within bearing types,  $a_i^{(j)}$  is replaced by  $A_i^{(j)}$ ,

$$S = \left\{ \begin{array}{l} S^{(1)} \text{ for B-1} \\ S^{(2)} \text{ for B-2} \\ S^{(3)} \text{ for B-3} \end{array} \right\} \text{ defined by eq (C26) and given in table C-2,}$$

and  $R$  (eq (C37)) refers to the calculated residual sum of squares within the particular classification.

Table C-3 summarizes with respect to companies A, B, and C the values of  $\hat{a}_0$ ,  $R$ , the calculated  $F$ -ratio (eq C38), and the critical  $F$  value. Table C-4 summarizes the same quantities for the B-1, B-2, and B-3-type bearings made by company B.

The results of this analysis show that the values  $a_1=2/3$ ,  $a_2=1.8$ , and  $p=3$  are consistent for rating life  $L_{10}$  data with respect to each of the three companies. However, for median life  $L_{50}$ , these assumed parameter values are consistent only for the data from company A.

The results for the same analysis made on the three different bearing types, indicate that the assumed parameter values are consistent only for the B-1 and B-2-type bearings with respect to median life  $L_{50}$ .

It is interesting to note that the analysis for company B (ignoring bearing types) showed that the assumed values for the parameters are consistent with the  $L_{10}$  data. However, a finer analysis by bearing type revealed that these values are not valid for the B-3-type bearings. This apparent inconsistency stems from the fact that the analysis for company B, taken as a whole, is dominated by those bearing types having the larger number of test groups, i. e., B-1 and B-2, and for these types the parameter values were found to be consistent with the  $L_{10}$  data.

The estimates for  $f_c$ , assuming  $a_1=2/3$ ,  $a_2=1.8$ , and  $p=3$ , are summarized in table 8 of the main text for rating life  $L_{10}$ . This summary also includes the value of  $f_c$  for company D computed from only three test groups. (Because of the small number of test groups, it was not possible to verify whether the assumed parameter values are valid for these data.)

TABLE C-3. Summary of computations (by companies) to test hypothesis that data are consistent with assumed values  $a_1=2/3$ ,  $a_2=1.8$ ,  $p=3$

Company	$L_{10}$			$L_{50}$			Critical $F$
	$\hat{a}_0 = \ln f_c$	$R$	$F$	$\hat{a}_0 = \ln f_c$	$R$	$F$	
A	8.4205	102.539286	*-0.06	8.9382	94.367916	0.79	2.80
B	8.5021	287.613198	0.42	8.9254	217.600672	3.71	2.65
C	8.1001	8.375582	0.96	8.5832	8.488584	9.92	3.59

\*Negative value due to rounding in calculations.

TABLE C-4. Summary of computations (by bearing type) for company B to test hypothesis that data are consistent with assumed values  $a_1=2/3$ ,  $a_2=1.8$ ,  $p=3$

Type	$L_{10}$			$L_{50}$			Critical $F$
	$\hat{a}_0 = \ln f_c$	$R$	$F$	$\hat{a}_0 = \ln f_c$	$R$	$F$	
B-1	8.4575	73.821978	0.84	8.8636	44.883126	3.95	2.86
B-2	8.5236	181.128312	1.17	8.9482	141.731604	2.42	2.70
B-3	8.6203	27.014247	5.81	9.0022	18.905517	5.21	3.24

## 7. References

- [1] R. L. Anderson and T. A. Bancroft, *Statistical theory in research*, ch. 13, 14, 15 (McGraw-Hill Book Co., Inc., New York, N. Y., 1952).
- [2] H. Cramér, *Mathematical methods of statistics* (Princeton University Press, Princeton, N. J., 1946).
- [3] W. J. Dixon and F. J. Massey, *Introduction to statistical analysis*, ch. 11 (McGraw-Hill Book Co., Inc., New York, N. Y., 1951).
- [4] M. A. Fieller, *Some problems in interval estimation*, *J. Roy. Stat. Soc. [B]* **XVI**, 175 (1954).
- [5] R. A. Fisher and L. H. C. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, *Proc. Cambridge Phil. Soc.* **24**, pt. 2, 180 (1928).
- [6] A. M. Freudenthal and E. J. Gumbel, *On the statistical interpretation of fatigue tests*, *Proc. Royal Soc. [A]* **216**, 309 (1953).
- [7] E. J. Gumbel, *Les valeurs extrêmes des distributions statistiques*, *Ann. inst. Henri Poincaré* **3**, pt. 2, 110 (1935).
- [8] E. J. Gumbel, *The return period of flood flows*, *Ann. Math. Stat.* **12**, 163 (June 1941).
- [9] E. J. Gumbel, *Statistical theory of extreme values and some practical applications*, National Bureau of Standards Applied Mathematics Series 33 (1954) (U. S. Government Printing Office, Washington 25, D. C.).
- [10] A. Hald, *Statistical theory with engineering applications*, ch. 18, 20 (John Wiley & Sons, New York, N. Y., 1952).
- [11] O. Kempthorne, *The design and analysis of experiments*, ch. 4, 5 (John Wiley & Sons, New York, N. Y., 1952).
- [12] J. Lieblein, *On the exact evaluation of the variances and covariances of order statistics in samples from the extreme-value distribution*, *Ann. Math. Stat.* **24**, 282 (June 1953).
- [13] J. Lieblein, *A new method of analyzing extreme-value data*, *Nat. Advisory Comm. Aeronau. Tech. Note* 3053 (January 1954).
- [14] J. Lieblein and H. E. Salzer, *Table of the first moment of ranked extremes (to be submitted to a technical journal)*.
- [15] G. Lundberg and A. Palmgren, *Dynamic capacity of rolling bearings*, *Acta Polytech.* **7**, Mech. Eng. Series, **1**, No. 3 (1947).
- [16] A. M. Mood, *Introduction to the theory of statistics*, ch. 13 (McGraw-Hill Book Co., Inc., New York, N. Y., 1950).
- [17] *Probability tables for the analysis of extreme-value data*, National Bureau of Standards Applied Mathematics Series 22 (1953) (U. S. Government Printing Office, Washington 25, D. C.).
- [18] W. Weibull, *The phenomenon of rupture in solids*, *Ing. Vetenskaps Akad., Handl. No.* 153 (1939).
- [19] W. Weibull, *A statistical representation of fatigue failure in solids*, *Trans. Roy. Inst. Technol., Stockholm.* No. 27 (1949).
- [20] S. S. Wilks, *Mathematical statistics*, ch. VIII (Princeton University Press, Princeton, N. J., 1946).
- [21] S. S. Wilks, *Order statistics*, *Bul. Am. Math. Soc.* **54**, 6 (January 1948).

WASHINGTON, March 22, 1956.