NATIONAL BUREAU OF STANDARDS REPORT

3996

STATISTICAL INVESTIGATION OF FATIGUE LIFE OF BALL BEARINGS

(Cont'd)

by

Julius Lieblein and Marvin Zelen

APPENDICES

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

FOR THE USE OF THE ASA LOAD RATING COMMITTEE ONLY, NOT FOR PUBLICATION

U. S. DEPARTMENT OF COMMERCE
Sinclair Weeks, *Secretary*

NATIONAL BUREAU OF STANDARDS
A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

Electricity and Electronics. Resistance and Reactance. Electron Tubes. Electrical Instruments. Magnetic Measurements. Process Technology. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Engineering Metrology.

Heat and Power. Temperature Measurements. Thermodynamics. Cryogenic Physics. Engines and Lubrication. Engine Fuels.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Nuclear Physics. Radioactivity. X-rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. AEC Radiation Instruments.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Fluid Mechanics. Engineering Mechanics. Mass and Scale. Capacity, Density, and Fluid Meters. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Organic Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

Mineral Products. Porcelain and Pottery. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure.

Building Technology. Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. Components and Techniques. Digital Circuitry. Digital Systems. Analogue Systems.

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Radio Propagation Physics. Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services.

Radio Propagation Engineering. Frequency Utilization Research. Tropospheric Propagation Research.

Radio Standards. High Frequency Standards. Microwave Standards.

- Office of Basic Instrumentation
- Office of Weights and Measures

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
1103-40-5145	28 March 1955	3996

STATISTICAL INVESTIGATION OF FATIGUE LIFE OF BALL BEARINGS
(Cont'd)

by

Julius Lieblein and Marvin Zelen
Statistical Engineering Laboratory

APPENDICES

Final report to the
American Standards Association
Committee B3, Subcommittee 7

NBS

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

FOR THE USE OF THE ASA LOAD RATING COMMITTEE ONLY, NOT FOR PUBLICATION

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

APPENDIX A

SUMMARY OF THE ORIGINAL DATA EMPLOYED IN THIS STUDY

This appendix summarizes in tabular form the worksheets submitted by the working group of the ASA Subcommittee to the National Bureau of Standards for statistical analysis. Separate tables are presented for deep-groove data from SKF, New Departure, and Marlin-Rockwell Corporation (M.R.C.); a table is also included for the self-aligning data from SKF. These five tables (A-1 to A-5) are followed by Table A-6 which gives a synopsis of the number of test groups and the number of bearings for each company.

Tables A-1 to A-5 give the identifying material, size of test group, the values for the quantities P, Z, D_a, and the estimates* for L_{10}, L_{50} and the "Weibull slope" $\underline{e}$. All of these variables are directly observed or specified quantities except for the estimates L_{10}, L_{50}, and $\underline{e}$. These last three quantities are based on statistical calculations which made use of the results of individual endurance tests. These calculations are explained in APPENDIX B.

In addition to the above quantities, Tables A-1 to A-5 also show the number of rows, $\underline{i}$, and contact angle, $\underline{\alpha}$. Since these are the same for all deep-groove bearings, namely, $i = 1$, $\alpha = 0^{\circ}$, they are merely indicated in the title. However, for the self-aligning bearings (SKF), this is not the case; hence Table A-5 has two additional columns for $\underline{i}$ and $\underline{\alpha}$.

SKF has also furnished summary sheets that list the key items for each test group, (deep-groove and self-aligning bearings), and also show the computed ratio C/P, which appears in one form of the life formula (see equation (1) in the main text of this report). The value of this ratio was obtained from the formula

---

* The estimates for L_{10} and L_{50} are given in millions of revolutions for all companies except M.R.C. The life estimates for M. R. C. are shown in hours, the same units in which the original endurance data were given.

$$C^* = f_c Z^{0.7} D_a^{1.8} (i \cos\alpha)^{0.7} .$$

The asterisk denotes that the approximation 0.7 was used for the exponent of Z in place of the value 2/3 specified in A/P 1947 (page 32, equation 120). These values are also included in Tables A-1 and A-5 and were utilized in one aspect of the analysis covered in this report, (see Appendix C, section 7).

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 Fafnir, omits 4 tests performed on other manufacturers' bearings; Table A-4, for M.R.C., omits 3 tests.

The six tables described above are followed by a specimen worksheet. 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 had accompanied the worksheets submitted to the Statistical Engineering Laboratory. The specimen worksheet was taken from the New Departure data because these sheets, being original ribbon copies, were in most suitable form for reproduction. The worksheets for the other companies showed substantially similar information, differing only in matters of minor detail. 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 SKF (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e (i = 1, $\alpha = 0^{\circ}$)

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	C^*/P	L_{10}	L_{50}	e Weibull slope
1- 1	End. 150-Group 1	1936	6309	24	4240	8	11/16	2.308	19.2	84.5	1.27
1- 2	" 152- " 1	1937	6309	20	4240	8	11/16	2.308	26.2	74.2	1.81
1- 3	" 152 " 2	1937	6309	14	4240	8	11/16	2.308	11.1	68.1	1.04
1- 4	" 152 " 3	1937	6309	19	4240	8	11/16	2.308	11.8	66.8	1.09
1- 5	" 152 " 4	1937	6309	18	4240	8	11/16	2.308	13.5	79.4	1.06
1- 6	" 169 " 1	1938	6209	21	2530	9	1/2	2.404	5.80	25.7	1.27
1- 7	" 170 " 1	1938	6309	28	4240	8	11/16	2.308	18.3	44.7	2.10
1- 8	" 170 " 2	1938	6309	27	4240	8	11/16	2.308	5.62	73.2	.73
1- 9	" 171 " 1	1940	6309	20	4240	8	11/16	2.308	15.8	82.7	1.14
1-10	" 171 " 2	1940	6309	22	4240	8	11/16	2.308	8.70	41.6	1.20
1-11	" 171 " 4	1940	6309	19	4240	8	11/16	2.308	11.6	160	.72
1-12	" 181 " 2	1940	6207	15	1940	9	7/16	2.463	20.6	71.4	1.52
1-13	" 181 " 3	1940	6207	15	1940	9	7/16	2.463	14.5	88.2	1.04
1-14	" 186 " 5	1940	6209	15	2536	9	1/2	2.399	12.1	33.1	1.87
1-15	" 186 " 6	1940	6209	14	2536	9	1/2	2.399	15.1	46.4	1.67
1-16	" 186 " 7	1940	6209	15	2536	9	1/2	2.399	14.0	43.6	1.66
1-17	" 186 " 8	1940	6209	14	2536	9	1/2	2.399	19.3	51.8	1.91
1-18	" 189 " 2	1940	6309	26	4240	8	11/16	2.308	46.2	110	2.17
1-19	" 189 " 3	1940	6309	14	4240	8	11/16	2.308	30.0	88.2	1.74
1-20	" 194 " 1	1942	6309	20	4240	8	11/16	2.308	21.1	57.4	1.89
1-21	" 194 " 2	1942	6309	20	4240	8	11/16	2.308	17.3	45.7	1.94
1-22	" 196 " 2	1942	6309	37	4240	8	11/16	2.308	37.5	118	1.64
1-23	" 196 " 3	1942	6309	36	4240	8	11/16	2.308	20.3	77.1	1.41
1-24	" 196 " 6	1942	6309	32	4240	8	11/16	2.308	4.03	42.5	.80
1-25	" 198 " 3	1941	6307	28	2544	8	17/32	2.423	8.38	84.7	.81
1-26	" 199 " 1	1943	6308	23	3975	8	19/32	1.899	1.79	13.5	.93
1-27	" 200 " 1	1942	6212	30	4400	10	5/8	2.224	11.7	45.1	1.39
1-28	" 200 " 2	1942	6312	31	6920	8	7/8	2.168	4.15	15.8	1.41
1-29	" 211 " 2	1943	6205	30	990	9	5/16	2.637	7.23	41.0	1.09
1-30	" 211 " 3	1943	6305	30	1509	7	7/16	2.589	22.9	110	1.20
1-31	" 211 " 4	1943	6303	30	932	7	11/32	2.660	9.54	31.6	1.57
1-32	" 217 " 4	1944	6308	26	3180	8	19/32	2.374	6.28	23.0	1.45
1-33	" 217 " 5	1944	6308	29	3180	8	19/32	2.374	4.81	21.2	1.27
1-34	" 219 " 1	1944	6218	33	8640	10	7/8	2.073	4.17	12.8	1.68
1-35	" 219 " 2	1944	6318	26	14080	8	1-1/4	1.877	5.42	31.6	1.07
1-36	" 246 " 1	1951	6207	28	1940	9	7/16	2.463	7.47	49.5	1.00
1-37	" 246 " 2	1951	6207	34	2330	9	7/16	2.051	4.80	21.3	1.26
1-38	" 246 " 3	1951	6207	27	1550	9	7/16	3.083	14.8	78.4	1.13
1-39	" 246 " 4	1951	6207	29	1165	9	7/16	4.101	84.9	460	1.11
1-40	" 246 " 5	1951	6207	27	2910	9	7/16	1.642	3.40	16.5	1.19
1-41	" 246 " 6	1951	6207	27	3880	9	7/16	1.232	1.24	3.23	1.97
1-42	" 246 " 7	1951	6207	26	776	9	7/16	6.158	241	951	1.37
1-43	F 669 -	1951	6326	30	19750	8	1-3/4	2.150	3.01	12.6	1.31
1-44	E64C 1-Tb5-Ser 3-1		6309A	30	2112	8	11/16	4.633	89.1	486	1.11
1-45	E64C 1-" 6- " 4-2		6309A	30	4224	8	11/16	2.316	15.2	104	.98
1-46	E64C 1-" 7- " 5-3	Not given	6309A	30	8448	8	11/16	1.158	2.04	10.2	1.17
1-47	E64X -" 4- " 3-1		6309	30	2112	8	5/8	3.937	51.0	376	.94
1-48	E64X -" 5- " 4-2		6309	30	4224	8	5/8	1.969	5.26	58.8	.78
1-49	E64X -" 6- " 5-3		6309	30	8448	8	5/8	0.984	.883	4.94	1.09
1-50	E64El -	1944	6309A	30	4224	8	11/16	2.316	14.8	57.4	1.39

TABLE A-2

SUMMARY BALL BEARING DATA FOR NEW DEPARTURE (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e $(i=1, \alpha=0^{\circ})$

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
2- 1	L03-1	1940	3L03	19	570	10	3/16	6.68	13.4	2.72
2- 2	L04-2	1944	3L04	20	570	9	5/16	29.8	70.0	2.22
2- 3	L04-3	1946	3L04	23	580	9	1/4	16.3	55.1	1.55
2- 4	L04-3A	1946	3L04	23	580	9	1/4	28.5	69.2	2.13
2- 5	L04-3B	1947	3L04	23	580	9	1/4	16.4	49.3	1.71
2- 6	L04-4	1943	3L04	10	665	9	1/4	10.3	40.1	1.50
2- 7	L04-4A	1944	3L04	10	665	9	1/4	25.7	46.4	3.19
2- 8	L05-1	1942	3L05	19	580	10	1/4	9.55	39.6	1.32
2- 9	L05-2	1946	3L05	33	620	10	1/4	17.9	62.1	1.51
2-10	L05-2A	1947	3L05	15	620	10	1/4	19.9	73.2	1.45
2-11	L05-2B	1947	3L05	31	620	10	1/4	12.9	50.4	1.39
2-12	L05-3	1944	3L05	19	625	10	1/4	19.3	46.2	2.18
2-13	L05-4	1941	3L05	17	720	10	1/4	11.1	23.3	2.54
2-14	L06-1	1946	3L06	60	980	11	9/32	15.7	43.5	1.85
2-15	L06-1A	1947	3L06	32	980	11	9/32	11.2	38.1	1.54
2-16	L07-1	1950	3L07	49	600	11	5/16	417	809	2.85
2-17	L07-2	1949	3L07	60	600	11	5/16	216	709	1.58
2-18	L07-3	1943	3L07	20	900	11	5/16	35.6	100	1.82
2-19	L07-4	1946	3L07	67	1220	11	5/16	12.0	42.2	1.50
2-20	L07-4A	1947	3L07	34	1220	11	5/16	8.53	46.6	1.11
2-21	L07-5	1940	3L07	20	1370	11	5/16	6.77	18.9	1.85
2-22	L07-6	1950	3L07	60	1415	11	5/16	13.5	46.5	1.53
2-23	L07-7	1950	3L07	60	2243	11	5/16	2.32	8.06	1.51
2-24	L08-1	1942	3L08	20	720	12	5/16	36.7	141	1.40
2-25	L08-2	1946	3L08	55	1300	12	5/16	19.0	57.2	1.71
2-26	L08-2A	1947	3L08	30	1300	12	5/16	19.5	60.6	1.67
2-27	L10-1	1944	3L10	20	1650	14	11/32	17.0	74.4	1.37
2-28	L10-2	1946	3L10	59	1760	14	11/32	20.9	53.7	2.00
2-29	L10-2A	1947	3L10	34	1760	14	11/32	9.56	40.7	1.30
2-30	L11-1	1940	3L11	20	2010	13	13/32	5.49	33.3	1.05
2-31	L11-1A	1940	3L11	9	2010	13	13/32	1.39	44.0	.54
2-32	L11-2	1944	3L11	19	2140	13	13/32	9.80	82.7	.88
2-33	L13-1	1943	3L13	11	2630	15	13/32	5.19	54.9	.80
2-34	L18-1	1942	3L18	12	5900	14	19/32	6.36	17.5	1.86
2-35	L18-1A	1947	3L18	19	5900	14	19/32	3.68	22.1	1.05
2-36	L22-1	1947	3L22	20	8070	14	23/32	8.34	23.6	1.81
2-37	L22-2	1942	3L22	12	8075	14	23/32	6.78	36.4	1.12
2-38	202-1	1938	3202	10	565	9	.210	9.27	18.4	2.75
2-39	203-1	1940	3203	23	720	8	9/32	18.2	56.9	1.66
2-40	203-1A	1940	3203	24	720	8	9/32	22.8	56.2	2.09

Continued next page

TABLE A-2, continued

SUMMARY BALL BEARING DATA FOR NEW DEPARTURE (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e $(i = 1, \alpha = 0^{\circ})$

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
2-41	203-1B	1940	3203	25	720	8	9/32	3.99	15.6	1.38
2-42	203-1C	1940	3203	21	720	8	9/32	9.07	29.4	1.60
2-43	203-1D	1940	3203	25	720	8	9/32	7.14	28.5	1.36
2-44	203-1E	1940	3203	25	720	8	9/32	12.5	26.4	2.51
2-45	203-1F	1941	3203	25	720	8	9/32	18.8	48.7	1.98
2-46	203-1G	1941	3203	23	720	8	9/32	21.5	53.2	2.08
2-47	204-1A	1947	3204	33	860	8	5/16	17.1	59.0	1.52
2-48	204-1B	1948	3204	8	860	8	5/16	15.2	87.6	1.08
2-49	205-1	1943	3205	20	900	9	5/16	30.1	92.3	1.68
2-50	205-1A	1944	3205	18	900	9	5/16	15.0	47.6	1.63
2-51	205-2	1945	3205	27	940	9	5/16	17.5	52.8	1.71
2-52	205-2A	1947	3205	34	940	9	5/16	14.4	65.6	1.24
2-53	205-3	1938	3205	10	1180	9	5/16	8.76	22.1	2.04
2-54	206-1	1945	3206	30	1580	9	3/8	12.1	43.3	1.47
2-55	206-1A	1947	3206	33	1580	9	3/8	17.2	64.6	1.42
2-56	206-1B	1948	3206	8	1580	9	3/8	10.7	34.6	1.61
2-57	207-3	1945	3207	31	2160	9	7/16	10.9	37.6	1.52
2-58	207-3B	1947	3207	30	2160	9	7/16	12.7	53.7	1.30
2-59	207-4	1938	3207	9	2200	9	7/16	3.73	43.5	.77
2-60	207-5	1947	3207	30	2480	9	7/16	16.6	78.3	1.21
2-61	208-13	1950	3208	40	1340	9	15/32	180	275	4.44
2-62	208-16	1937	3208	19	1660	10	7/16	85.2	234	1.86
2-63	208-18	1941	3208	19	1700	9	15/32	57.1	230	1.35
2-64	208-20	1939	3208	24	2480	9	15/32	15.7	55.8	1.48
2-65	208-20A	1939	3208	25	2480	9	15/32	27.1	97.8	1.47
2-66	208-20B	1939	3208	23	2480	9	15/32	21.7	122	1.09
2-67	208-20C	1939	3208	28	2480	9	15/32	13.2	42.3	1.62
2-68	208-20D	1939	3208	28	2480	9	15/32	35.8	145	1.35
2-69	208-20E	1939	3208	20	2480	9	15/32	12.7	34.7	1.87
2-70	208-20F	1944	3208	20	2480	9	15/32	10.1	27.8	1.87
2-71	208-20G	1945	3208	20	2480	9	15/32	8.83	34.3	1.39
2-72	208-20H	1938	3208	10	2480	9	15/32	16.5	60.3	1.45
2-73	208-20I	1942	3208	11	2480	9	15/32	17.9	65.8	1.45
2-74	208-20J	1943	3208	10	2480	9	15/32	15.7	63.1	1.35
2-75	208-20K	1943	3208	20	2480	9	15/32	10.8	42.1	1.38
2-76	208-20L	1944	3208	18	2480	9	15/32	14.2	39.9	1.83
2-77	208-20M	1944	3208	18	2480	9	15/32	19.0	67.8	1.48
2-78	208-20N	1944	3208	18	2480	9	15/32	16.3	57.7	1.49
2-79	208-20O	1944	3208	20	2480	9	15/32	2.93	18.0	1.04
2-80	208-20P	1944	3208	20	2480	9	15/32	5.69	25.4	1.26

Continued next page

TABLE A-2, continued

SUMMARY BALL BEARING DATA FOR NEW DEPARTURE (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e (i=1, $\alpha = 0^o$)

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
2- 81	208-20Q	1944	3208	28	2480	9	15/32	9.54	39.9	1.32
2- 82	208-20S	1944	3208	22	2480	9	15/32	12.6	55.7	1.27
2- 83	208-20T	1944	3208	23	2480	9	15/32	5.10	37.5	.94
2- 84	208-20U	1944	3208	18	2480	9	15/32	16.0	53.7	1.56
2- 85	208-20V	1944	3208	20	2480	9	15/32	1.98	22.1	.78
2- 86	208-20W	1945	3208	20	2480	9	15/32	5.65	28.8	1.16
2- 87	208-20X	1945	3208	20	2480	9	15/32	12.8	43.6	1.58
2- 88	208-20Y	1945	3208	20	2480	9	15/32	9.84	32.3	1.59
2- 89	208-20Z	1945	3208	20	2480	9	15/32	12.1	43.0	1.48
2- 90	208-20AA	1945	3208	20	2480	9	15/32	5.48	40.8	.94
2- 91	208-20BB	1945	3208	20	2480	9	15/32	6.64	25.3	1.41
2- 92	208-20CC	1945	3208	32	2480	9	15/32	13.9	41.9	1.70
2- 93	208-20DD	1946	3208	35	2480	9	15/32	9.02	45.4	1.17
2- 94	208-20EE	1946	3208	34	2480	9	15/32	11.0	49.2	1.26
2- 95	208-20GG	1947	3208	31	2480	9	15/32	14.5	73.6	1.16
2- 96	208-20II	1944	3208	9	2480	9	15/32	5.91	37.2	1.02
2- 97	208-20JJ	1944	3208	10	2480	9	15/32	18.1	40.5	2.33
2- 98	208-20KK	1945	3208	10	2480	9	15/32	17.1	53.3	1.65
2- 99	208-20LL	1945	3208	10	2480	9	15/32	32.6	61.8	2.95
2-100	208-20NN	1945	3208	10	2480	9	15/32	24.1	66.2	1.87
2-101	208-2000	1945	3208	20	2480	9	15/32	36.1	71.6	2.75
2-102	208-20PP	1946	3208	20	2480	9	15/32	63.3	104	3.82
2-103	208-20QQ	1946	3208	12	2480	9	15/32	14.4	59.0	1.33
2-104	208-20RR	1946	3208	11	2480	9	15/32	15.1	92.9	1.04
2-105	208-20SS	1945	3208	10	2480	9	15/32	18.8	39.4	2.55
2-106	208-20TT	1950	3208	12	2480	9	15/32	5.63	34.7	1.04
2-107	208-20UU	1950	3208	12	2480	9	15/32	7.23	34.5	1.21
2-108	208-20VV	1951	3208	30	2480	9	15/32	16.7	71.8	1.29
2-109	208-20WW	1951	3208	63	2480	9	15/32	26.5	90.3	1.54
2-110	208-20XX	1950	3208	23	2480	9	15/32	8.35	49.1	1.06
2-111	208-21	1943	3208	19	3250	9	15/32	3.79	9.30	2.10
2-112	208-22	1937	3208	10	3470	10	7/16	9.05	36.6	1.35
2-113	208-23	1944	3208	20	4000	9	15/32	2.98	7.35	2.08
2-114	209-1	1943	3209	19	2300	10	15/32	22.5	73.4	1.59
2-115	209-2	1938	3209	10	2730	10	15/32	3.82	31.7	.89
2-116	209-3	1946	3209	22	2660	10	17/32	6.55	20.8	1.63
2-117	210-3	1944	3210	20	2250	11	15/32	17.5	64.3	1.45
2-118	210-4	1943	3210	16	2300	11	15/32	61.7	152	2.10
2-119	210-6	1945	3210	48	2840	11	15/32	18.6	42.7	2.27
2-120	210-6A	1947	3210	28	2840	11	15/32	21.6	66.3	1.68

Continued next page

TABLE A-2, continued

SUMMARY BALL BEARING DATA FOR NEW DEPARTURE (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e $(i = 1, \alpha = 0^{o})$

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
2-121	210-6B	1947	3210	8	2840	11	15/32	11.9	39.1	1.59
2-122	210-6C	1948	3210	8	2840	11	15/32	13.9	50.6	1.46
2-123	210-7	1943	3210	19	3200	11	15/32	7.80	33.1	1.30
2-124	210-9	1944	3210	28	4000	11	15/32	3.55	13.9	1.38
2-125	210-10	1943	3210	19	4000	11	15/32	9.40	23.4	2.06
2-126	216-1	1947	3216	23	6350	11	11/16	4.76	22.7	1.21
2-127	222-1	1944	3222	20	12000	11	1-1/16	3.23	9.86	1.69
2-128	222-1A	1944	3222	20	12000	11	1-1/16	2.62	9.52	1.46
2-129	222-2	1944	3222	9	12700	8	1-1/2	7.89	39.7	1.17
2-130	222-3	1949	3222	18	16500	11	1-1/16	4.93	20.4	1.33
2-131	222-3A	1950	3222	20	16500	11	1-1/16	6.26	16.2	1.98
2-132	302-1	1938	3302	8	565	7	5/16	37.3	103	1.85
2-133	302-2	1944	3302	20	900	7	5/16	14.0	38.6	1.86
2-134	305-3	1938	3305	10	1650	8	13/32	30.3	87.6	1.77
2-135	306-7	1944	3306	20	2250	8	15/32	25.7	71.2	1.85
2-136	306-8	1943	3306	20	2300	8	15/32	10.5	60.4	1.07
2-137	306-9	1944	3306	19	3200	8	15/32	10.3	24.1	2.21
2-138	306-10	1944	3306	19	4000	8	15/32	4.56	12.9	1.81
2-139	307-7	1937	3307	10	1710	8	17/32	25.1	274	.79
2-140	307-8	1938	3307	9	2360	8	17/32	48.8	264	1.12
2-141	307-9	1937	3307	10	2680	8	17/32	7.53	60.7	.90
2-142	307-10	1937	3307	11	3850	8	17/32	14.9	62.6	1.32
2-143	313-1	1947	3313	21	7760	8	29/32	4.57	43.4	.84
2-144	316-1	1943	3316	12	9550	8	1-1/16	3.90	40.7	.80
2-145	316-2	1947	3316	21	9750	8	1-1/16	15.5	79.4	1.16
2-146	318-1	1948	3318	16	11400	8	1-3/16	10.2	43.9	1.29
2-147	318-1A	1948	3318	20	11400	8	1-3/16	4.71	16.9	1.48
2-148	318-2	1947	3318	18	11420	8	1-3/16	10.1	34.2	1.55

TABLE A-3

SUMMARY BALL BEARING DATA FOR FAFNIR (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e

($i=1$, $\alpha=0^{\circ}$)

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
3-1	24-L-7	1942	307 K	94	1580	7	9/16	16.9	64.8	1.40
3-2	24-L-7-S-1-50%	1949	307 K	29	790	7	9/16	211	729	1.52
3-3	24-L-7-S-1	1949	307 K	35	1185	7	9/16	74.4	287	1.40
3-4	53-L-14	1940	208 K	29	1600	9	1/2	9.62	40.1	1.32
3-5	53-L-14-S-1	1945	208Con-rad	10	1600	8	15/32	11.9	66.3	1.10
3-6	147-L-27	1943	306 K	9	2275	7	17/32	13.8	58.0	1.31
3-7	195-L-34-S-2 A	1946	207 K	13	2540	8	15/32	2.38	11.3	1.21
3-8	195-L-34-S-2 B	1946	207 K	12	2540	8	15/32	2.38	11.5	1.19
3-9	242-L-42-A	1949	M307KCR	12	1580	7	9/16	8.75	62.2	0.96
3-10	242-L-42-B	1949	307 K	12	1580	7	9/16	25.7	113	1.27
3-11	285-L-46	1947	MM208K	24	1600	9	1/2	14.5	113	0.92
3-12	329-L-57	1949	204 K-Cl	12	610	8	5/16	26.8	65.6	2.10

TABLE A-4

SUMMARY BALL BEARING DATA FOR MARLIN-ROCKWELL CORPORATION (Deep Groove), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e

($i=1$, $\alpha=0^{\circ}$)

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	L_{10}	L_{50}	e Weibull slope
4-1	16	1946	207 S	19	1750	9	7/16	159	963	1.05
4-2	19	1951	207 S	34	1750	9	7/16	71.7	526	0.94
4-3	20	1951	207 S	56	1750	9	7/16	113	582	1.15

NOTE: In TABLE A-4 the life estimates L_{10} , L_{50} are in hours.

TABLE A-5

SUMMARY BALL BEARING DATA FOR SKF (Self-Aligning), WITH COMPUTED VALUES FOR L_{10}, L_{50}, AND WEIBULL SLOPE e

Record No.	Company Reference No.	Year of test	Bearing No.	Number in test group	Load lbs.	Z Number of balls	D_a Ball diam. in.	i	α	C^*/P	L_{10}	L_{50}	e Weibull slope
5- 1	End. 152-Group 5	1936	1309	27	4240	15	1/2	2	9°25'	1.690	4.90	14.7	1.72
5- 2	" 152 " 6	1936	1309	26	4240	15	1/2	2	9°25'	1.690	5.68	16.8	1.73
5- 3	" 170 " 3	1938	1309	37	4240	15	1/2	2	9°25'	1.690	8.43	28.8	1.53
5- 4	" 171 " 3	1939	1309	21	4240	15	1/2	2	9°25'	1.690	1.85	16.1	0.87
5- 5	" 172 " 1	1940	1309	35	4240	15	1/2	2	9°25'	1.690	1.39	12.3	0.86
5- 6	" 172 " 2	1940	1309	36	4240	15	1/2	2	9°25'	1.690	0.716	8.24	0.77
5- 7	" 172 " 3	1940	1309	35	4240	15	1/2	2	9°25'	1.690	0.942	9.98	0.80
5- 8	" 176 " 2	1940	1312	30	7040	16	5/8	2	8°33'	1.533	1.97	8.02	1.34
5- 9	" 177 " 1	1940	1315	30	10137	16	3/4	2	8°26'	1.468	2.27	9.51	1.31
5-10	" 186 " 9	1940	1309	17	4240	15	1/2	2	9°25'	1.690	2.19	7.30	1.56
5-11	" 188 " 3-4	1941	2318	19	9266	13	1 1/8	2	14° 6'	3.083	19.3	61.7	1.62
5-12	" 188 " -	1940	1218	30	8360	19	5/8	2	6°23'	1.293	2.10	5.64	1.91
5-13	" 189 " 1	1940	1309	29	4240	15	1/2	2	9°25'	1.690	9.10	23.6	1.98
5-14	" 192 " -	1941	2215	30	6315	20	17/32	2	9°32'	1.328	3.58	11.8	1.58
5-15	" 193 " 1	1940	1207	30	1760	16	5/16	2	8°44'	1.690	4.48	17.0	1.41
5-16	" 193 " 2	1940	1307	30	2604	14	13/32	2	9°18'	1.809	4.00	21.1	1.13
5-17	" 194 " 4	1942	1309	31	4240	15	1/2	2	9°25'	1.690	1.90	11.7	1.04
5-18	" 198 " 1	1941	2207	30	2216	14	3/8	2	13°55'	1.819	4.81	16.4	1.54
5-19	" 198 " 2	1941	2307	30	3496	11	17/32	2	17°7'	2.078	3.01	19.4	1.01
5-20	" 203 " 1	1942	2212	29	4440	18	15/32	2	10°34'	1.445	5.78	23.1	1.36
5-21	" 235 " -	1949	1309	59	4240	15	1/2	2	9°25'	1.690	3.01	9.64	1.62
5-22	" 239 " 1	1950	1309	30	4240	15	1/2	2	9°25'	1.690	2.25	10.8	1.20
5-23	" 246 " 11	1951	1207	26	1760	16	5/16	2	8°44'	1.693	4.53	12.9	1.80
5-24	" 246 " 12	1951	1207	27	1410	16	5/16	2	8°44'	2.109	6.97	26.9	1.40
5-25	" 246 " 13	1951	1207	31	2110	16	5/16	2	8°44'	1.409	1.56	6.73	1.29
5-26	" 246 " 14	1951	1207	25	1060	16	5/16	2	8°44'	2.805	7.79	34.3	1.27
5-27	" 246 " 15	1951	1207	28	1060	16	5/16	2	8°44'	2.805	12.1	41.3	1.53
5-28	" 246 " 16	1951	1207	29	795	16	5/16	2	8°44'	3.740	42.2	118	1.84
5-29	E64GL-Tb 4-Ser 4		1309	30	4224	15	1/2	2	9°25'	1.697	5.89	14.8	2.05
5-30	64G1 " 3 " 3		1309	30	2112	15	1/2	2	9°25'	3.394	31.2	122	1.38
5-31	64Q1 " 5 " 5		1309	30	8448	15	1/2	2	9°25'	0.848	0.625	1.56	2.06
5-32	64S1 " 1 " -		1309	30	1056	15	1/2	2	9°25'	6.789	166	477	1.78
5-33	64Q1 " 1 " -		1309	30	4224	15	1/2	2	9°25'	1.697	9.09	30.6	1.55
5-34	64F2 " 1 " -	Not given	1309	30	528	15	1/2	2	9°25'	13.575	286	2100	0.94
5-35	64F1 " 6		1309	29	528	15	1/2	2	9°25'	13.575	174	2720	0.69
5-36	64F1 " 7 " 2		1309	30	1056	15	1/2	2	9°25'	6.789	120	563	1.22
5-37	64F1 " 8 " 3		1309	30	2112	15	1/2	2	9°25'	3.394	24.4	130	1.12
5-38	64F1 " 9 " 4		1309	30	4224	15	1/2	2	9°25'	1.697	10.2	42.9	1.31
5-39	64F1 " 10 " 5		1309	30	8448	15	1/2	2	9°25'	0.848	0.459	1.60	1.51
5-40	64Q " 3	1939	1309	30	4224	15	1/2	2	9°25'	1.697	4.37	16.4	1.42

TABLE A-6

RECAPITULATION OF TEST GROUPS OF BALL BEARING DATA

Company and Bearing Type		Number of test groups	Total Number of bearings in test groups
DEEP GROOVE, Total		213	4948
SKF		50	1259
New Departure		148	3289
Extra light, Series 3L00	37		
Light, Series 3200	94		
Medium, Series 3300	17		
Fafnir*		12	291
Marlin-Rockwell Corporation*		3	109
SELF-ALIGNING, SKF Only		40	1196
Total, All Bearings		253	6144

* Bearings of other manufacture not included. See text.

SPECIMEN WORKSHEET

Referen No. L04-3A
Bearing Mfg. by New Departure
Bearing Tested by New Departure
Date of Test 8-26-46
Bearing No. 3L04
Load 580 R.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

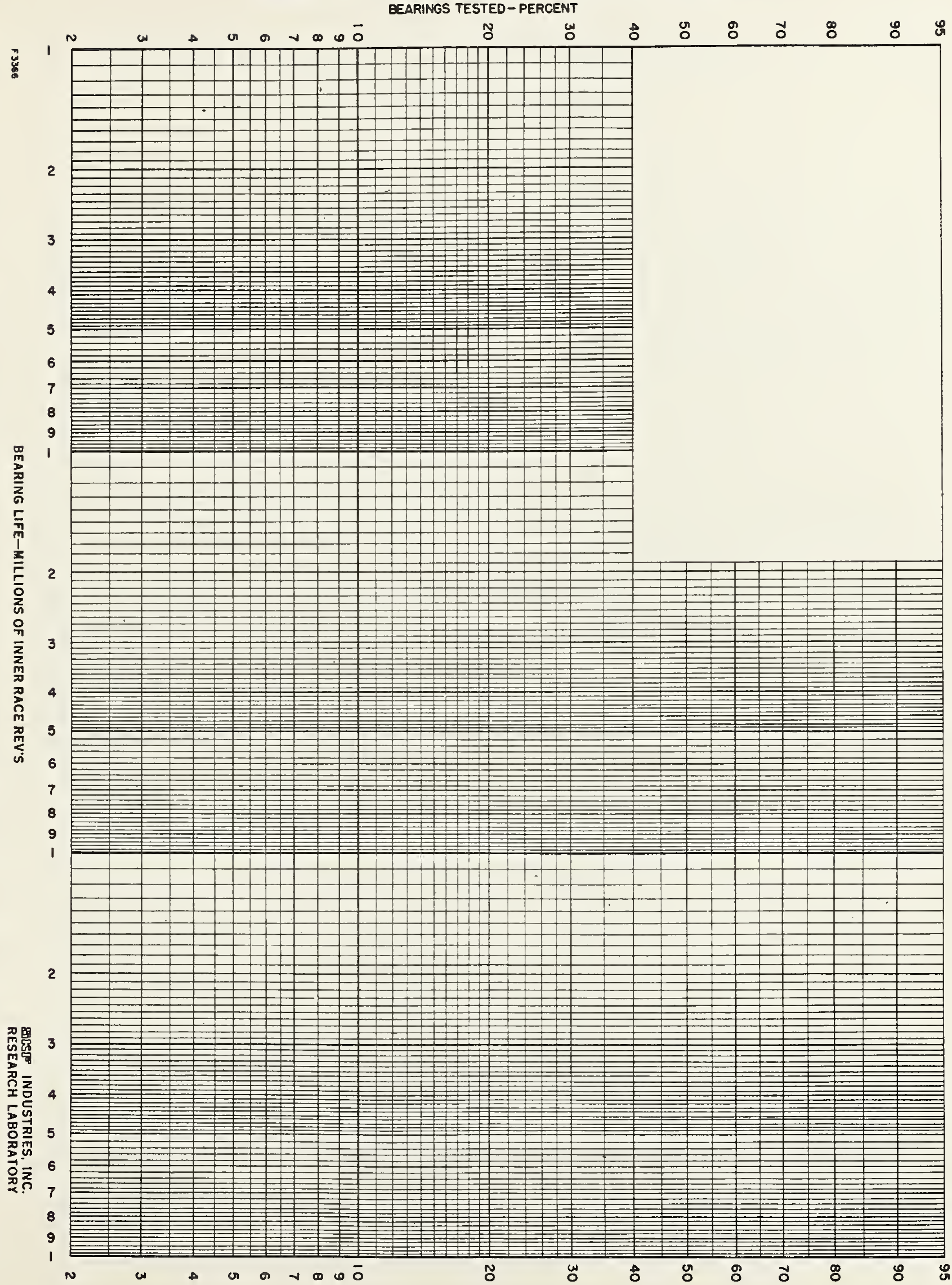
Ball Failure 13 52%
Inner Ring Failure 5 20%
Outer Ring Failure 1 4%

Test life in 10^6 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	
~~2~~	~~67.80~~	~~L.Bore~~	~~Omitted~~
~~3~~	~~67.80~~	~~L.Bore~~	~~Omitted~~
4	68.64	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



FOR THE USE OF THE ASA LOAD RATING COMMITTEE ONLY, NOT FOR PUBLICATION

APPENDIX B

EVALUATION OF L_{10}, L_{50}, AND WEIBULL SLOPE e BY USE OF ORDER STATISTICS

This is a technical appendix designed to present the mathematical and statistical bases for estimating, for each test group, the values of L_{10} and L_{50} for use in the regression analysis discussed in Appendix C. Estimation of the Weibull slope e is also considered.

1. The Weibull distribution

As noted in the main text of the report, the fundamental basis for estimation of L_{10}, L_{50}, and e for each individual test group was the assumption that the probability distribution of fatigue lives of individual bearings within a given test group could be represented by a "Weibull distribution".* This means that the observed fatigue lives 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 c.d.f.)

---

* So named for W. Weibull (cf. [15], pages 16 ff.), who is considered to be one of the first to study it extensively.

is*

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

where a and e are the two parameters to be fitted. They are related to L_{10} and L_{50} in a manner to be explained subsequently (see equation (B2a)). 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 of the types will be discussed in the following section.) This matter was first studied chiefly by Fisher and Tippett [3] and for this reason the type (B1) is sometimes referred to as Fisher-Tippett Type III. (A complete treatment of such limiting "extreme-value distributions" from the modern viewpoint may be found in Gnedenko [5].)

---

* Actually, an approximation is involved here, as the number of cycles, L, can take only integral values. However, in the application with which this report is concerned, L is generally of the order of millions. Therefore it is felt that use of a continuous instead of discrete probability distribution introduces no appreciable error.

There are two reasons for choosing the Weibull distribution (B1) as the underlying probability distribution for fatigue life: (a) theoretical, and (b) empirical.

(a) The theoretical basis for the choice lies in the assumption 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 which proceeds from this assumption is mentioned by a number of authors, and is given explicitly, for example, by Freudenthal and Gumbel in [4], pages 316 to 318. It leads precisely to the form (B1) (see equation (2.9) in [4]).

It is recognized that this statistical approach has not received universal acceptance. The present report is, however, not concerned with the relative merits of the various theories of fatigue, but merely with consequences of a reasonable choice from among them.

(b) The practical application of the Weibull distribution received extensive attention by W. Weibull in [16], where he showed that a (more general) distribution of the type (B1) represented certain fatigue-life data quite satisfactorily. Other treatments have discussed the suitability of this distribution. In addition, a report by L. G. Johnson [9], while also dealing with a more general form of the Weibull distribution, concludes (page 4) "from

several large lots of ball bearing [fatigue] failures" that "they support the theory of a Weibull distribution very well." Finally, there is in addition the concrete fact that inspection of the special "Weibull" plots accompanying each of the worksheets indicates that many can be fitted satisfactorily by a straight line representing a Weibull distribution, as explained below.

The manner in which these plots, called "Weibull plots", are constructed is described in detail by Johnson in [9]. A sample of Weibull-function coordinate paper used for this purpose is included in Appendix A above. The essence of the method is that equation (B1) may be converted, by taking logarithms twice, into *

(B2) $$e(\ln L) - (e \ln a) = \ln[\ln(1/S)] ,$$

where "ln" denotes "natural logarithm", i.e., with base $\epsilon = 2.71828\ldots$. From equation (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 equation (B2) give

(B2a)
$$e(\ln L_{10}) - e(\ln a) = \ln[\ln(1/.90)] = - 2.25037$$
$$e(\ln L_{50}) - e(\ln a) = \ln[\ln(1/.50)] = - 0.36651$$

---

* It will be recalled that the natural logarithm (base ϵ) is related to the ordinary logarithm (base 10) by $\ln a = 2.3025851 \log_{10} a$.

These are the relationships between the parameters a, e, L_{10}, and L_{50}. The right-hand numerical values will later be denoted by $y_{.90}$, $y_{.50}$, respectively. Equation (B2) may be written

$$\text{(B3)}\qquad \begin{gathered} ex - a' = y\ , \quad \text{where} \\ x = \ln L,\ a' = e \ln a,\ y = \ln[\ln(1/S)]\ . \end{gathered}$$

These 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 grid. The variable y is represented through the percentage surviving, S, or rather through the (vertical) scale for "Bearings tested - percent" = percent failed = 1 - S = F, 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 relation, 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.

For these reasons the Weibull distribution was adopted.

2. The extreme-value distribution

a. Relation to Weibull distribution

A by-product of the preceding discussion is that it appears that the logarithms of lives, rather than the lives themselves, are the natural units in terms of which to carry out the analysis. This idea has been adopted even by those who do not make use of the Weibull distribution either because they are unaware of its existence or because they do not feel it fits their data. Such authors advocate the use of "log life", but they maintain that this quantity is normally distributed. For example, a tentative draft (Ransom [14]) of statistical recommendations in connection with fatigue testing for the use of the A.S.T.M. Committee E-9 on Fatigue, adopts this logarithmic basis and presents a number of procedures suitable for the case of a normal distribution. Study of such sources suggests that one important reason for recommending the (log) normal is that there already exists a body of well-developed procedures for this distribution, and that this is less true for other distributions.

If the Weibull distribution is adopted for life, L, then the variate $x = \ln L$ cannot be taken as normal.* In fact, it has the c.d.f., obtained from (B1),**

---

* cf. Weibull [16], pages 29-35.

** cf. Freudenthal and Gumbel [4], equations (2.5), (2.6), (2.8), (2.9).

$$G(x) = \text{Prob}\{\ln(\text{life}) \geq x\} = \text{Prob}\{\text{life} \geq \epsilon^{x}\}$$

(B4)

$$= \mathcal{L}(\epsilon^{x}) = \exp(-a^{-e}\epsilon^{ex}) = \exp[-\epsilon^{(x-\ln a)/e^{-1}}] ,$$

$$-\infty < x < \infty .$$

This c.d.f. has the form

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

(B5)

$$y = (x-u)/\beta, \; -\infty < x < \infty ,$$

and

(B6) $$u = \ln a \;, \; \beta = 1/e$$

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

The form (B5) is another of the three asymptotic distributions of extreme (smallest) values, sometimes designated as Fisher-Tippett Type I. This distribution has been studied extensively chiefly by Professor E. J. Gumbel, a pioneer in the field of extreme values (e.g., [6, 7, 8]). Its full designation is "the asymptotic distribution of smallest values", and it is closely related to the one for largest values usually treated. Either of these is a distribution of extreme values. However, since only the one for smallest values will be applied in this

report, the "extreme-value distribution" will be taken to refer to the smallest-values case, unless otherwise specified.

From the above discussion it appears that procedures relevant to the extreme-value distribution (B5) rather than the normal distribution are called for. Fortunately, a mathematical basis for such procedures was recently developed by one of the authors of this report, and has been described in detail, Lieblein [11].

There were additional considerations which led to the use of methods based upon the extreme-value distribution rather than a method based on fitting a straight line equation of the type (B3) by a least squares method.

In fitting equation (B3) the variable $x = \ln L$ is obtained from the known data. However the variable y, measured through the percentage failing, F, presents difficulties. The problem of how to plot F is known as the problem of "plotting position". Solution of this problem is essential to the use of least squares since least squares requires pairs of data (x_i, y_i), $i=1,2,\ldots,n$, whereas the observations provide only the x_i directly. The values y_i needed can only come from the plotting positions adopted for F_i.

It is clear that values, F_i, of the plotted variable, F, must somehow be related to the rank order of the bearings as they fail. Thus, for the first bearing that fails

out of a test group of 10, we have F_1 = 10 percent failed. However, plotting the value .10 on the vertical scale on the Weibull coordinate paper, or, in general, plotting of the cumulative ratio f/n, where f is the rank order of failing in a test group of n, is not recommended. Plotting of this simple ratio has serious theoretical drawbacks. These are indicated by Johnson in [9], page 6, and more fully discussed by Gumbel in [8], page 14, where he advocates the plotting position of f/(n+1).* However, even this is open to objection. Dr. Bradford F. Kimball, another of the present-day experts in the field of extreme-value analysis, has maintained in private correspondence with one of the authors of this report that this plotting position is subject to bias for some purposes, and recommends still another, less simple, plotting position related to the means of order statistics tabulated in Lieblein and Salzer [12]. This, however, also raises certain difficulties. Numerous other plotting positions have been tried and discarded throughout the past several decades, and the search for a satisfactory solution is still continuing, a paper on the subject having appeared as recently as December 1954 (see Chernoff and Lieberman [1]). The

---

* This plotting position was also used by Weibull in [15] (cf. equation (73) and the vertical scale in Figures 3 and 4).

controversy over choice of plotting position thus by no means can be regarded as settled.

Another difficulty with the use of least squares is that as usually used it fails to take into complete account the number remaining intact in incompleted tests. Since many of the test groups contained intact bearings, it was desirable to have a method which could take the number intact into account. Finally, it is to be noted that the successive points are not independent, since they represent the observed lives in increasing order. A correct use of least squares procedures would have to take into account all the intercorrelations. This, however, is not done in the usual application of "the method of least squares".

It was therefore deemed preferable to adopt an approach that obviated the necessity of choice among a number of controversial procedures. For this purpose the order statistics approach mentioned above and outlined below was already at hand as an objective basis upon which to develop the analyses necessary for determining the life estimates for individual test groups.

The method adopted depends on the use of <u>order statistics</u>, defined as the set of observations in a sample when arranged in increasing or decreasing numerical order.

If the lives L_i for a test group of <u>n</u> bearings are in increasing size-order, so are their natural logarithms. Hence the test in general yields a set of order statistics $x_i = \ln L_i$; where

$$\text{(B7)} \quad x_1 \leq x_2 \leq \cdots \leq x_k \quad (\leq x_{k+1} \leq \cdots \leq x_n) \;, \quad k \leq n \;,$$

and k is the rank of the last bearing that fails before the test is stopped, the remaining (n-k) remaining intact. The potential lives $x_{k+1}, \ldots, x_n$, which would have been observed had the test been run to completion, are unavailable.

The k values $x_1, x_2, \ldots, x_k$ in (B7) constitute a <u>censored sample</u>. This is defined as one where the total number of observations (n) is known but information is not available concerning some of them. The statistical method used in this study for estimating rating life L_{10} or median life L_{50} was to find a linear function of the order statistics of the censored sample, of the form

$$T = \sum_{j=1}^{k} w_i x_i \;,$$

with weights w_i, which would in repeated samples (i) have the true population value as its long-run average, and (ii) have smallest sampling fluctuation about this population value. An estimating function with the first of these two properties is called <u>unbiased</u>; one having also the

second is said to have <u>minimum variance</u> (from among the class of such unbiased linear estimators). The search for such "optimum" estimators is an important part of the statistical determination of an underlying population.

The development in [11] yielded procedures for obtaining such estimators in the special case where all failures can be observed, i.e., $k = n$. What was needed in the present case was research directed to generalizing such results to the case $k < n$. An outline of this latter work was presented by one of the authors at the 17th Annual Meeting of the Institute of Mathematical Statistics in December 1954. The abstract was circulated to members of the American Standards Association working group on ball bearings and is included at the end of this Appendix for convenience.

b. Description of extreme-value distribution

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 report. It will be seen that the problem of estimating life is equivalent to that of estimating the parameters u and β.

The parameters of the extreme-value distribution (B5) are depicted in Figure 1. The quantity u is the position

of the mode or highest point of the (frequency) distribution. The quantity β is a scale parameter, analogous to the standard deviation σ in the case of the normal distribution. In fact, β is $\sqrt{6}/\pi$ (or, about 3/4) times the standard deviation of the extreme-value distribution.

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

(B8) $$t = u + \beta y$$

which are linear combinations of the parameters u and β 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 β simultaneously. Thus if t can be obtained as a + by with a, b known and y arbitrary, then we can read off at once the values u = a, β = b.

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 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_{10} (life exceeded by 90 percent of bearings) by the relation

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

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

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

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 β, may be written in the form (B8):

(B8') $$t_F = u + \beta y_F ,$$

where y is a quantity, depending only on the probability F, determined as follows. We have from (B8')

(B11) $$y_F = (t_F - u)/\beta ,$$

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) and (B11),

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

Thus, solving for y_F, we obtain

(B13) $$y_F = \ln(-\ln F) ;$$

this is the reduced variable corresponding to the probability F, and may be obtained by a simple change in sign from Table 2 of [13], which tabulates the function

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

where Φ_y, a probability, takes on values from 0 to 1. Thus,

(B14)
$$\text{for } F = .90, \quad y_F = -2.25037 ;$$
$$\text{for } F = .50, \quad y_F = -0.36651 .$$

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 (B10) 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 [11] were for problems, such as maximum gust-loads on airplanes, that required the (extreme-value) distribution of largest 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 considerable 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 will be lost by use of reduced variates and distributions throughout. Thus, in (B5), x will be replaced by the reduced variate y, and, for typing ease, the symbol $G(y)$ will be used instead of $\Phi(y)$:

(B14a) $$G(x) \equiv \Phi(y) = \exp(-\epsilon^{x}) \ .$$

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

(B15) $$\text{Prob}\{Y_S \geq y\} \equiv G_S(y) = \exp(-\epsilon^{y}) \ , \quad -\infty < y < \infty \ ,$$

where Y_S denotes the reduced smallest value. The corresponding distribution of largest values is (Gumbel [8], page 21, equation (I))

(B16) $$\text{Prob}\{Y_L \geq y\} \equiv H_L(y) = 1 - \text{Prob}\{Y_L < y\} = 1 - \exp(-\epsilon^{-y}) \ , \quad -\infty < y < \infty \ .$$

Comparison with (B15) shows that

(B17) $$G_S(y) = 1 - H_L(-y) \ .$$

The corresponding relation for the density functions is obtained by differentiation, with $g_S(y) = G_S'(y)$, and $h_L(y) = H_L'(y)$:

(B18) $$g_S(y) = h_L(-y) \ .$$

Hence the two distributions are merely mirror images of each other. The moments of the distributions then follow:

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

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

(B20) $$\nu_{1S} = -\gamma = -\nu_{1L} \ , \quad \sigma_S^2 = \frac{\pi^2}{6} = \sigma_L^2 \ .$$

These values are given, for example, in [8], page 23, equation (3.27).

Finally, there are needed the corresponding relationships between the order statistics for the two distributions. Since 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 will turn out to give simpler results. Thus we are interested in the i^{th} order statistic in the series

(B21) $$S: \quad y_1' \geq y_2' \geq \cdots \geq y_i' \geq \cdots \geq y_n' \ .$$

A prime on order statistics or on functions of them (such as u, v, h, in what follows) will be used as a reminder that the order is descending, not ascending. Thus in Tables B-1, B-2, B-3 the absence of primes indicates that the order statistics are in increasing order.

(B21) is the analogue of the series

$$\text{(B22)} \qquad (L): \quad y_1 \leq y_2 \leq \cdots \leq y_i \leq \cdots \leq y_n$$

of order statistics in the largest-value case. Whenever a distinction is necessary the subscripts S or L will be used with the y's.

The first task is to express the distribution of the i^{th} member of (S), with density function denoted by $u'_{S,i}(y)$, in terms of the density function, $v_{L,i}(y)$, of the corresponding i^{th} member of L. According to Wilks [17], page 16, equation (16), and in view of the definitions of symbols in (B15), (B16),

$$\text{(B23)} \qquad v_{L,i}(y) = \frac{[1-H_L(y)]^{i-1}[H_L(y)]^{n-i}}{B(i,n-i+1)} h_L(y) \ ,$$

where $1-H_L(y)$ is used in place of $F(x_{(\gamma)})$ of the reference and B denotes the (complete) Beta function.

The reference cited gives the density function for order statistics in ascending order. For the case of <u>descending</u> order, it is only necessary to replace <u>i</u> by n-i+1 in the cited equation, giving

$$\text{(B24)} \qquad u'_{S,i}(y) = \frac{[G_S(y)]^{i-1}[1-G_S(y)]^{n-i}}{B(n-i+1,i)} g_S(y) \ .$$

Substituting (B17) and (B18) there is obtained

$$u'_{S,i}(y) = \frac{[1-H_L(-y)]^{i-1}[H_L(-y)]^{n-i}}{B(n-i+1,i)} h_L(-y) = v_{L,i}(-y) , \tag{B25}$$

by (B23) and the fact that the Beta function is symmetric in its arguments. Thus, the density functions for the order statistics are related in precisely the same manner as those for the original populations, namely, by merely a change of sign in the variable, provided the sequence of the order statistics is reversed.

Equation (B25) yields immediately the following relations for the (pure) moments, covariances, and variances of the order statistics, analogously to (B19):

$$\begin{aligned} E'_S(y_i^m) &= (-1)^m E_L(y_i^m) \\ E'_S(y_i y_j) &= E_L(y_i y_j) \\ \sigma'^2_{S,i} \equiv \sigma'^2_S(y_i) &= \sigma^2_L(y_i) \equiv \sigma^2_{L,i} \\ \sigma'_{S,ij} \equiv \sigma'_S(y_i, y_j) &= \sigma_L(y_i, y_j) \equiv \sigma_{L,ij} \end{aligned} \tag{B26}$$

In words, these relationships are quite simple: the even moments remain the same; the odd moments change only in sign. This holds true for the order statistics as well as for the original extreme-value distributions.

It is to be remembered that the primed quantities, involving S are for order statistics in descending order;

those without primes, involving L, are for order statistics in ascending order.

The above development shows that the numerical results for moments of order statistics previously obtained in [11] for the largest-value case can be used here for smallest values without any substantive change.

3. Method of order statistics

a. For small samples

Returning to the order statistics method for censored samples, the problem may be stated as follows. An independent random sample of n items from the distribution of smallest values is taken, 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:

$$(x'_1 \geq x'_2 \geq \ldots \geq x'_{n-k} \geq)\; x'_{n-k+1} \geq \ldots \geq x'_n \,, \tag{B27}$$

where the parentheses denote the (n-k) (largest) unobservable values, and the remaining k values are known. Primes will 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

$$T'_{n,k} = w'_1 x'_{n-k+1} + w'_2 x'_{n-k+2} + \ldots + w'_k x'_n \,, \quad k \leq n \tag{B28}$$

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

(B29) $$t_F = u + \beta y_F$$

of the extreme-value population (B5), such that T' in (B28) is (i) unbiased and (ii) of minimum variance. Mathematically, this means respectively, that

(B30) $$E(T') = t_F ,$$

where E denotes mathematical expectation, and

(B31) $$\mathrm{Var}(T') = \text{a minimum}$$

subject to the above condition.

From (B5),

(B32) $$x = u + \beta y$$

where y is the reduced variable and x the observed variable. The corresponding relation for the order statistics x_i and y_i is

(B33) $$x'_j = u + \beta y'_j , \quad j = n-k+1, n-k+2, \ldots, n ,$$

where

(B34) $$x'_{n-k+1} \geq x'_{n-k+2} \geq \cdots \geq x'_n$$

and

(B35) $$y'_{n-k+1} \geq y'_{n-k+2} \geq \cdots \geq y'_n .$$

From (B33) it follows that

(B36) $$E_S(x'_j) = u + \beta E_S(y'_j) ,$$

since u and β, though unknown, are constants not subject to sampling variation when the operation of expectation is performed. The values $E_S(y'_j)$ may be obtained with the aid of the table in [12]. 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 (B36) are therefore obtained as

(B37) $$E_S(y'_s) = - E_L(y'_{n-s+1}) .$$

The reference gives the values of $E_L(y'_r)$ for r = 1(1)min(n,26), n = 1(1)10(5)60(10)100.

The relation (B30) gives, in view of (B28) and (B36),

(B38) $$E(T') = \sum_{j=1}^{k} w'_j[u+\beta E(y'_{n-k+j})] = t_F = u + \beta y_F .$$

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

(B39)
$$\sum_{j=1}^{k} w'_j = 1$$
$$\sum_{j=1}^{k} (Ey'_{n-k+j})w'_j = y_F ,$$

where the numerical values Ey'_{n-k+j} may be obtained from [12] as already indicated.

For the variance condition (B31), we have, in view of (B28),

$$\text{(B40)} \qquad \sigma_S^2(T') = \sum_{j=1}^{k} w_j'^2 \sigma^2(x'_{n-k+j}) + \sum_{j=1}^{k} \sum_{\substack{i=1 \\ i \neq j}}^{k} w_i' w_j' \sigma(x'_{n-k+i}, x'_{n-k+j}) \ .$$

From (B33) we have

$$\text{(B41)} \qquad \begin{aligned} \sigma^2(x_r') &= \beta^2 \sigma^2(y_r') = \beta^2 \sigma_r'^2 \ , \\ \sigma(x_r', x_s') &= \beta^2 \sigma(y_r', y_s') = \beta^2 \sigma'_{rs} \ . \end{aligned}$$

Hence the condition (B31) becomes

$$\text{(B42)} \qquad \begin{aligned} \mathrm{Var}(T') &= [\Sigma w_j'^2 \sigma'^2_{n-k+j} + \Sigma\Sigma' w_i' w_j' \sigma'_{n-k+i,n-k+j}]\beta^2 = V_k^{(n)'} \beta^2 \\ &= \text{minimum subject to (B30).} \end{aligned}$$

This is a constrained minimum problem for variation in the unknown w_j, and is equivalent to finding the (unconstrained) minimum of

$$\begin{aligned} M_1 = {} & (\sum_j \sigma'^2_{n-k+j} w_j'^2 + \sum_j \sum_i{}' \sigma'_{n-k+i,n-k+j} w_i' w_j')\beta^2 \\ & + \lambda_1 (\sum_j w_j' - 1) + \mu_1 [\sum_j (Ey'_{n-k+j}) w_j' - y_F] \ , \end{aligned}$$

where λ_1, μ_1 are Lagrange multipliers. Since $\beta^2 > 0$ is constant, though unknown, this is the same as minimizing

(B44) $$M = \frac{M_1}{\beta^2} = \Sigma\sigma'_{n-k+j} w'^2_j + \Sigma\Sigma'\sigma'_{n-k+i,n-k+j} w'_i w'_j + \lambda(\Sigma w'_j - 1) + \mu[\Sigma(Ey'_{n-k+j})w'_j - y_F] ,$$

where $\lambda = \lambda_1/\beta^2$, $\mu = \mu_1/\beta^2$. Setting the derivatives with respect to w'_j, j=1,2,...,k, equal to zero and dividing by 2 gives

(B45) $$\sigma'^2_{n-k+j} w'_j + \sum_{\substack{i=1 \\ i\neq k}}^{k} \sigma'_{n-k+i,n-k+j} w'_j + \lambda + \mu Ey'_{n-k+j} = 0$$

$$j = 1, 2, \ldots , k .$$

For each fixed value of $k \leq n$ these are k linear equations which, with the two in (B39), form a simultaneous system of (k+2) equations in the (k+2) unknowns $w'_1, w'_2, \ldots, w'_k$, λ, μ. The values of λ and μ are useful as a check, since, if (B45) is multiplied by w_j and summed for j=1,2,...,k, the result is, in view of (B39) and (B42),

(B46) $$V^{(n)'}_{k,min} + \lambda + \mu Y_F = 0 \quad ;$$

that is, we should have

(B47) $$V^{(n)'}_{k,min} = - \lambda - \mu Y_F \quad .$$

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

In general, there will be a set of (k+2) linear equations to solve for each k=2,...,n. To avoid cumbersome notation the case k = n will be discussed first and the case k < n then referred to in terms of it.

(i) Case k = n. For k = n, the matrix of coefficients and right-hand "constant terms" of (B45) and (B39) is the (n+2) x (n+3) matrix

$$\text{(B48)} \qquad \Lambda_n^o = \begin{bmatrix} \sigma'_{11} & \sigma'_{12} & \cdots & \sigma'_{1n} & 1 & Ey'_1 & 0 \\ \sigma'_{21} & \sigma'_{22} & \cdots & \sigma'_{2n} & 1 & Ey'_2 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ \sigma'_{n1} & \sigma'_{n2} & \cdots & \sigma'_{nn} & 1 & Ey'_n & 0 \\ 1 & 1 & \cdots & 1 & 0 & 0 & 1 \\ Ey'_1 & Ey'_2 & \cdots & Ey'_n & 0 & 0 & y_F \end{bmatrix} .$$

The ordinary (n+2) x (n+2) matrix of coefficients, without the constant terms, will be denoted by Λ_n. If Γ_n denotes the vector column of constant terms, then

$$\text{(B49)} \qquad \Lambda_n^o = [\Lambda_n \,\vdots\, \Gamma_n] ,$$

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

(B50) $$\Lambda_n W'_n = \Gamma_n ,$$

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

Before solving the set of equations (B50) its coefficients must be evaluated. The means Ey'_j have been tabulated, as already mentioned. The variances and covariances involve complicated integrals that have been evaluated in terms of simpler functions, Lieblein [10]. The numerical values of the coefficients are shown in Table B-1 for n = 2 to n = 6. While this table is primarily for the smallest-value case, it can also be used for the largest-value case in the manner indicated. In the smallest-value case the order-statistics are in descending order, and therefore primed. The means $E_S(y'_i)$ are for the smallest-value case, a fact denoted by the subscript S. In the largest-value case the order is the reverse, ascending, but the same column can be used for the means $E_L(y_i)$ provided all signs are reversed:

$$E_L(y_i) = - E_S(y'_i) .$$

The (n+2) solutions of (B50) are all expressible linearly in terms of the components of Γ_n. Thus the solutions all take the form

$$w'_j = a'_j + b'_j y_F , \qquad j = 1,2,\ldots,n$$

(B51) $$\lambda = c'_1 + d'_1 y_F$$

$$\mu = c'_2 + d'_2 y_F$$

Substituting these w'_j in (B42) gives an expression of the form

(B52) $$Q'_{n,n} = V^{(n)'}_{n,min} = A'_n \, y_F^2 + B'_n \, y_F + C'_n \quad .$$

The quantities a'_j, b'_j for the weights w'_j, for n = 2 to 6, may be found from Table B-2 in a manner to be discussed under the case k < n. 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, may likewise be found from Table B-3.

The solution of the system of equations became increasingly lengthy for increasing values of n, with correspondingly diminishing yield in the number of decimal places. On the other hand, the "efficiency" of the result (see below) increased as n increased, so that at n = 6 it had reached useful levels for practical purposes. For these reasons calculations were not pushed beyond n = 6 in the present study. Methods for larger values of n will be taken up subsequently.

Table B-3 shows that as sample size increases from $n = 2$ to 6 (in the case $k = n$), the variance diminishes as regards 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 theoretically specified 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 [11], pages 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 as regards 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 very good. As regards x_{50} ($F = .50$), 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 heavy computations necessary, calculations were not carried beyond n = 6 in [11].

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 not unreasonable, in view of its high efficiency in estimating the logarithms, x_{10}, x_{50}.

(ii) Case k < n. For the case k < n, the procedure is very similar. One starts with a (k+2) x (k+3) order matrix Λ^{o}_{k} derived from Λ^{o}_{n} in (B48) by striking out the first (n-k) rows and columns. Thus, when k = n - 1, the matrix of the linear system in (n+1) variables to be solved is

$$\text{(B53)} \quad \Lambda^{o}_{n-1} = \begin{bmatrix} \sigma'_{22} & \cdots & \sigma'_{2n} & 1 & Ey'_2 & 0 \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ \sigma'_{n2} & \cdots & \sigma'_{nn} & 1 & Ey'_n & 0 \\ 1 & \cdots & 1 & 0 & 0 & 1 \\ Ey'_2 & \cdots & Ey'_n & 0 & 0 & y_F \end{bmatrix}$$

etc. until when k = 2 it becomes

$$\text{(B54)} \qquad \Lambda^{o}_{2} = \begin{bmatrix} \sigma'_{n-1,n-1} & \sigma'_{n-1,n} & 1 & Ey'_{n-1} & 0 \\ \sigma'_{n,n-1} & \sigma'_{nn} & 1 & Ey'_{n} & 0 \\ 1 & 1 & 0 & 0 & 1 \\ Ey'_{n-1} & Ey'_{n} & 0 & 0 & y_F \end{bmatrix},$$

representing a set of 4 equations in 4 unknowns.

This procedure is convenient for high-speed automatic computation, since the master matrix Λ^{o}_{n} can be read in for each n, and the matrix for each value of k then obtained by the machine from the one for the preceding value by successively deleting rows and columns starting with the first row and column. It is thus not necessary to feed in the n-1 separate matrices (corresponding to k = 2,3,...,n) for each value of n. The resulting weights w'_j and variances $Q'_{n,k}$ were obtained in similar manner to those for k = n in (B51), (B52). These, it will be recalled, are primed quantities, associated with descending order of the order statistics. Since, however, the observations, x, for successive failures naturally arise 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 $w_i = a_i + b_i y_F$, and in Table B-3, giving the variances $Q_{n,k} = Ay_F^2 + By_F + C$ for the estimators $T_{n,k}$ formed with the above weights, and also these variances evaluated as regards the parameters $\bar{x}_{.90} = \ln L_{10}$, $\bar{x}_{.50} = \ln L_{50}$. The relationships of these unprimed quantities (smallest values--ascending order) to the primed ones (smallest values--descending order) 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 remain unchanged, by the following reasoning. Since the reversal of order changes x_i' into x_{n-i+1}, that is $x'_{n-k+1} \geq x'_{n-k+2} \geq \ldots \geq x'_n$ into $x_k \geq x_{k-1} \geq \ldots \geq x_2 \geq x_1$, which is the same as $x_1 \leq x_2 \leq \ldots \leq x_k$ in Tables B-2 and B-3, it follows that the sum of square terms in the variance of

$$T'_{n,k} = \sum_{j=1}^{k} w'_j \, x_{n-k+j}$$

in (B40) becomes

$$\sum_{j=1}^{k} w_j'^2 \sigma^2(x'_{n-k+j}) = \sum_{j=1}^{k} w_{k-j+1}^2 \sigma^2(x_{k-j+1}) \quad ,$$

i.e., the corresponding sum $\sum_{r=1}^{k} w_r^2 \sigma^2(x_r)$ in the variance of $T_{n,k} = \sum_{j=1}^{k} w_i x_j$, and similarly for the cross-product sum. Thus all variances Q remain unaffected.

b. <u>Extension to larger samples</u>

The key to handling samples of more than six items is to break them up into independent samples of 6 with a remainder subgroup, if necessary, of from 2 to 5 items. Since, however, 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 which were generated in the electronic computer (the SEAC) as needed.

Such randomization is not desirable when it can be avoided, as it may lose information embodied in the order of the bearings under test. It is therefore recommended that whenever a group of bearings are placed upon test machines for simultaneous testing, a record should be kept of their order of placement, or order of manufacture, or some other meaningful order that could be considered to be random. Then the fatigue lives will also represent a series in random order, and it is in this order that the endurance data should be available for analysis. If any need should occur for ordering the fatigue lives in increasing length, this can then be obtained, but the original order will be available if needed.

The original order will not be necessary, however, if calculations of the weights, etc. would be extended up to the number of bearings in a test group, such as n = 20. In that case it would not be necessary to use subgroups and the sample could be treated as a whole.

In the present investigation, however, subgroups were necessary, as indicated. Each subgroup was treated as a random sample by the methods already developed for size 6 or less. That is, a "sub-estimator" 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 overall sample, was obtained for each of 4 population quantities:

$$u, \beta, \; t_{.90} = x_{10} = \ln L_{10} = u + y_{.90}\beta \quad ,$$

(B55)

$$t_{.50} = x_{50} = \ln L_{50} = u + y_{.50}\beta \quad .$$

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

$$\hat{u} = \sum_{i=1}^{k} a_i x_i \quad , \quad \hat{\beta} = \sum_{i=1}^{k} b_i x_i \quad ,$$

(B56) $$\hat{x}_{10} = T_{n,k}(10) = \hat{u} + y_{.90}\hat{\beta}, \; \hat{x}_{50} = T_{n,k}(50) = \hat{u} + y_{.50}\hat{\beta}$$

$$y_{.90} = -2.25037 \quad , \; y_{.50} = -0.36651,$$

where $x_1 \leq x_2 \leq \ldots \leq x_k$, $2 \leq k \leq n \leq 6$, are the logarithms of the actual observed lives in a subgroup arranged in ascending order, the a_i and b_i are read directly from Table B-2, and $y_{.90}$ and $y_{.50}$ were given in (B14). For the overall sample estimator, the sub-estimators $T_{n,k}$ are merely averaged, as already noted.

For later use (APPENDIX C) the variance of the overall estimator T and its relation to sample size will be considered here. Consider first the case of a complete sample--no intact bearings are present because the test is run to completion. Let n be the sample size. There are two cases, according as (i) $n \leq 6$, or (ii) $n > 6$.

(i) $n \leq 6$. 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} are seen to lie on a straight line of slope unity. This shows that 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 unity. Hence the underlined statement is approximately applicable here too.

(ii) n > 6. Larger samples are broken up into subgroups and an overall estimator formed by taking a weighted average of the sub-estimators as already described. It will now be shown that the underlined rule holds here as well: If the variances of a set of independent quantities formed for different-size samples varies inversely with sample size, so also does the variance of a weighted average of the set, weighted by sample size.

Let T_{n_i} denote the given set of quantities formed for sample size n_i, $i = 1,2,\ldots,t$ and let the variances, by assumption, be

$$\sigma^2 (T_{n_i}) = \frac{A}{n_i}, \quad i = 1,2,\ldots,t,$$

where A is independent of sample size. Denote the weighted average by

$$\bar{T}_n = \sum_{i=1}^{t} \left(\frac{n_i}{n}\right) T_{n_i}, \quad \sum_{i=1}^{t} n_i = n.$$

Then, since the T_{n_i} are independent,

$$\sigma^2 (\bar{T}_n) = \Sigma \left(\frac{n_i}{n}\right)^2 \sigma^2 (T_{n_i}) = \Sigma \left(\frac{n_i}{n}\right)^2 \frac{A}{n_i}$$

$$= \frac{A}{n^2} \Sigma n_i = \frac{A}{n} \quad ;$$

i.e., the variance of $\bar{T}_n$ formed for a sample of n is also inversely proportional to sample size n.

Thus the desired relationship is demonstrated, to a reasonable approximation, for complete samples of any size.

When a sample contains intact bearings, so that we have k observed lives out of n, k is defined to be the "effective" sample size. Study of the variances in Table B-3, along the lines of that above for complete samples, indicates that for incomplete or censored samples, variance varies approximately as effective sample size, provided the number of intact bearings remains the

same. To the extent that the number intact varies in the different subgroups of an over-all sample, the relationship of variance to sample size would depart from that in the more simple situation. However, it is felt that in most of the cases the relationship, variance inversely proportional to effective sample size, holds sufficiently to be adopted as a working rule. In that case, the weights used in the regression analysis, which are proportional to the reciprocal of the variance, would merely be effective sample size. However, in the interests of simplicity, the total actual, rather than effective sample size was used in weighting. But since the actual weights used in APPENDIX C were obtained by reducing by a factor of five, it was felt that the difference between effective and actual sample size would not significantly affect the analysis.

Two complete runs were made on the SEAC for each of the 253 test groups of data and the two results were averaged for each group, giving values of the averages

(B56) $\quad \bar{u}, \ \bar{\beta}, \ \bar{t}_{.90} = \overline{\ln L_{10}}, \ \bar{t}_{.50} = \overline{\ln L_{50}}$.

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

In determining the number of runs necessary to minimize the variation due to randomization, the following considerations come into play. The effect of randomizing the individual results within a test group is to introduce an additional source of variability. Consider the average value $\overline{\ln L_{10}}$ (or $\overline{\ln L_{50}}$, etc.), an estimate which is made up of (say) $\underline{r}$ values of $\ln L_{10}$ (or $\ln L_{50}$, etc.), each value obtained from a different randomization of the same data within a test group. Then the variance, σ^2, of this combined estimate is made up of two additive components as follows:

$$\sigma^2 = \sigma_b^2 + \frac{\sigma_r^2}{r} .$$

The component σ_b^2 is the sampling variance, namely, the variance if randomization were not necessary (such as if the individual test results had already been given in some order considered random); σ_r^2 is the variance introduced by randomization. It is to be noted that as r increases (many randomizations) the component of variation due to randomization, σ_r^2/r, decreases, whereas σ_b^2 is not affected.

For these purposes, only rough approximations were necessary, and they were obtained by making a few trial runs on each of several sets of data, giving $\sigma_b^2 = 0.4$, $\sigma_r^2 = 0.5$. Thus

$$\sigma^2 = 0.4 + \frac{0.5}{r} ,$$

which gives the following approximate values of the standard deviation, σ, for 1, 2, and 3 randomizations or runs:

r	σ	Percentage improvement over single run
1	0.95	-----
2	.81	$\frac{.14}{.95}$ = 15 per cent
3	.75	$\frac{.20}{.95}$ = 21 per cent

This indicates that the greatest improvement in precision came from making a second run, and the additional improvement from a third run did not seem to be worth the effort.

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 selected for this purpose was that designated "Record

No. 1-1" in Table A-1, for SKF deep groove bearings. The test group consisted of 24 ball bearings, Bearing No. 6309, 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.* 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 $T_{n,k}$ written in column (4), 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).

* Since sample size 24 is an exact multiple of n = 6, it so happened that there was no "remainder subgroup". 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 4. Cross-products. The cross-products

$T_{n,k} = \sum_{i=1}^{k} a_i x_i, \quad \sum_{i=1}^{k} b_i x_i$ were then evaluated and placed as shown for each of the subgroups.

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 $\overline{T_1} = \hat{u}$, $\overline{T_2} = \hat{\beta}$, respectively. These are the order-statistics estimates of the two parameters u, β of the extreme-value distribution which represents the underlying Weibull distribution.

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

The logarithmic life estimates were given by following linear combinations of the estimates $\hat{u}$ and $\hat{\beta}$, using the given values of $y_{.90}$, $y_{.50}$:

$$\hat{x}_{10} = \ln L_{10} = \hat{u} - 2.25037\,\hat{\beta} = 2.982305 \qquad \text{(B)}$$

$$\hat{x}_{50} = \ln L_{50} = \hat{u} - 0.36651\,\hat{\beta} = 4.404120 \qquad \text{(C)}$$

The rating life L_{10} = antilog $(\hat{x}_{10})$ (base ϵ) = 19.2333.

The median life L_{50} = antilog $(\hat{x}_{50})$ (base ϵ) = 81.7872.

These three values (A), (B), (C) represent the outcome of the calculation. Convenient methods of finding confidence intervals for the population parameters L_{10}, L_{50} are discussed in APPENDIX D.

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 e shown here differ slightly from those recorded in Table A-1. The reason is that after all calculations for all companies were completed, an entirely independent

second calculation was run using different random numbers, and therefore giving another set of values. This not only served as a check on the first run, but also made it possible to eliminate much of the variation in results due to randomization by averaging the two determinations in the manner already mentioned. It may be of interest to compare the results of the two runs in this example:

	L_{10}	L_{50}	e
1st run	19.73	81.79	1.325
2nd run	18.62	87.31	1.219
recorded value (Table A-1)	19.2	84.5	1.27

It should be noted that the recorded value was not obtained as a simple average of the two runs shown. The average was taken of their logarithms or reciprocals. The recorded values therefore represent the geometric mean in the case of L_{10} and L_{50}, and the harmonic mean in the case of the Weibull slope e.

SYNOPSIS OF APPENDIX B

1. The Weibull distribution

This distribution is described and two types of reasons are discussed for choosing this as the underlying population for fatigue life of ball bearings.

2. The extreme-value distribution

a. Relation to Weibull distribution

This relation is the consequence of a simple logarithmic transformation. The reasonableness of the extreme-value distribution (for logarithm of life) is discussed in comparison with the use of the normal distribution by some authors. Reasons are also discussed for the use of this distribution as a basis for estimating life in preference to a "least squares" type of procedure commonly used.

b. Description of extreme-value distribution

Here are discussed several useful parameters (not all independent) of this distribution and their relationship to the concepts rating life, L_{10}, and median life, L_{50}.

c. Conversion from largest to smallest values

This discussion is necessary in order to be able to make immediate use of previous results in extreme-value analysis. By a careful choice of orientation of the order statistics, i.e., to rank them in descending order for smallest values and in ascending order for largest values, it becomes possible to take advantage of very simple relations between the distributions of largest and of smallest values and their related order statistics. This solves the problem for the case of complete or uncensored samples.

3. Method of order statistics

a. For small samples (i.e., $n \leq 6$)

SYNOPSIS (Cont'd)

For a censored sample ($k < n$), the minimization procedure is described that yields the "optimum" (this term is defined in the body of Appendix) estimators of the various quantities (parameters) of interest: u, β, $\ln L_{10}$, $\ln L_{50}$. For purposes of tabulation for convenience in applications, the necessary weights and variances for the estimators are rearranged so that they can be used with the observations in increasing order. The relevant tables, B-1 to B-4, are explained in detail.

b. Extension to larger samples ($n > 6$)

For larger samples than 6 the procedure is to act as though one had started with the appropriate number of separate samples each of size 6 or less, make the required calculations upon each, and average the results. Care should be taken that the separate samples should be independent. This is not the case, for example, when the over-all sample is already arranged in order, in which case randomization is necessary. This situation should be avoided, whenever possible.

It is also shown, for use in APPENDIX C, that with minor exceptions, variance varies inversely with sample size, and this relation is a good general working rule.

c. Worked example

A complete example is worked through in detail, taken from an actual test group in the data, and the results of the various steps shown in a concise table, together with the final outcomes.

4. References to APPENDIX B

[1] Chernoff, H. and Lieberman, G. J., "Use of Normal Probability Paper," Journal of the American Statistical Association, 49, pp. 778-785 (December 1954).

[2] Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

[3] Fisher, R. A. and Tippett, L. H. C., "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," Proceedings of the Cambridge Philosophical Society, 24, Pt. 2, pp. 180-190 (1928).

[4] Freudenthal, A. M. and Gumbel, E. J., "On the Statistical Interpretation of Fatigue Tests," Proceedings of the Royal Society, A, 216, pp. 309-332 (1953).

[5] Gnedenko, B., "Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire," Annals of Mathematics (2), 44, pp. 423-453 (1943).

[6] Gumbel, E. J., "Les Valeurs Extrêmes des Distributions Statistiques," Annals de l'Institut Henri Poincaré, 5, Pt. 2, pp. 110-113 (1935).

[7] Gumbel, E. J., "The Return Period of Flood Flows," Annals of Mathematical Statistics, 12, pp. 163-190 (June 1941).

[8] Gumbel, E. J., 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.)

[9] Johnson, L. G., Some Statistical Techniques in Analyzing Fatigue Data, Report No. ME-1-68, Research Laboratories Division, General Motors Corporation, Detroit, Michigan, January 20, 1950.

References (Cont'd)

[10] Lieblein, J., "On the Exact Evaluation of the Variances and Covariances of Order Statistics in Samples from the Extreme-Value Distribution," Annals of Mathematical Statistics, 24, pp. 282-287 (June 1953).

[11] Lieblein, J., A New Method of Analyzing Extreme-Value Data, National Advisory Committee for Aeronautics Technical Note 3053, January 1954.

[12] Lieblein, J. and Saltzer, H. E., "Table of the First Moment of Ranked Extremes," (To be submitted to a technical journal).

[13] National Bureau of Standards, 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.).

[14] Ransom, J. T., "A Guide to Statistical Methods for Use in Fatigue Testing," Tentative draft prepared for the use of A.S.T.M. Committee E-9 on Fatigue.

[15] Weibull, W., "The Phenomenon of Rupture in Solids," Ingeniörs Vetenskaps Akademiens Handlingar Nr. 153 (1949).

[16] Weibull, W., "A Statistical Representation of Fatigue Failure in Solids," Transactions of the Roayl Institute of Technology, Stockholm, Sweden, No. 27 (1949).

[17] Wilks, S. S., "Order Statistics," Bulletin of the American Mathematical Society, 54, pp. 6-50 (January 1948).

A B S T R A C T

"Estimation from 'Censored' Samples of Extreme Data," J. Lieblein
National Bureau of Standards

In a previous report (NACA Technical Note 3053, [11]), the author has obtained unbiased minimum-variance order statistics estimators of an arbitrary linear function of the two parameters of the extreme-value distribution with c.d.f. $P(x) = \exp[-\exp[-(x-u)/\beta]]$. These estimators are "optimum" within the class of all linear functions of the order statistics of a sample of given size n. The present paper extends these methods, together with the necessary tables, to the case of a "censored" sample, defined as one where the total number of observations is known but full information is not available with regard to some of them. For example, in fatigue testing, the test may be discontinued before all test specimens have failed, so that the endurance lives for the "intacts" are not available. The method is applied to an example of fatigue-life testing of ball bearings.

TABLE B-1

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

For distribution of largest values, $y_1 \le y_2 \le \cdots \le y_n$

For distribution of smallest values, $y'_1 \ge y'_2 \ge \cdots \ge y'_n$

n	i	MEANS* $E_S(y'_i)$	VARIANCES AND COVARIANCES*, $\sigma_{ij} = \sigma_{ji}$ j = 1	j = 2	j = 3	j = 4	j = 5	j = 6
2	1	0.11593 152	0.68402 804	0.48045 301				
	2	-1.27036 285		1.64493 407				
3	1	0.40361 359	0.44849 796	0.30137 144	0.24375 810			
	2	-0.45943 263		0.65852 235	0.54629 438			
	3	-1.67582 795			1.64493 407			
4	1	0.57351 263	0.34402 417	0.22455 344	0.17903 454	0.15388 918		
	2	-0.10608 352		0.41553 113	0.33720 966	0.29271 188		
	3	-0.81278 175			0.65180 236	0.57432 356		
	4	-1.96351 003				1.64493 407		
5	1	0.69016 715	0.28486 447	0.18202 536	0.14358 737	0.12257 865	0.10901 329	
	2	0.10689 454		0.30849 748	0.24676 731	0.21226 644	0.18967 383	
	3	-0.42555 061			0.40598 292	0.35267 072	0.31716 095	
	4	-1.07093 582				0.64907 319	0.58991 519	
	5	-2.18665 358					1.64493 407	
6	1	0.77729 368	0.24658 20	0.15496 74	0.12121 61	0.10291 64	0.09116 19	0.08285 42
	2	0.25453 448		0.24854 56	0.19670 62	0.16806 28	0.14945 32	0.13619 10
	3	-0.18838 534			0.29761 59	0.25616 60	0.22887 90	0.20925 46
	4	-0.66271 588				0.40185 52	0.36145 55	0.33204 51
	5	-1.27504 579					0.64769 96	0.59985 67
	6	-2.36897 513						1.64493 41

* The means are for smallest values; for largest values, change all signs. The σ_{ij} are the same for both.

TABLE B-2

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

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

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

n	k		x_1	x_2	x_3	x_4	x_5	x_6
2	2	a_i	.0836269	.9163731				
		b_i	- .7213475	.7213475				
3	2	a_i	- .3777001	1.3777001				
		b_i	- .8221012	.8221012				
	3	a_i	.0879664	.2557135	.6563201			
		b_i	- .3747251	- .2558160	.6305411			
4	2	a_i	- .7063194	1.7063194				
		b_i	- .8690149	.8690149				
	3	a_i	- .0801057	.0604316	1.0196741			
		b_i	- .4143997	- .3258576	.7402573			
	4	a_i	.0713800	.1536799	.2639426	.5109975		
		b_i	- .2487965	- .2239192	- .0859035	.5586192		
5	2	a_i	- .9598627	1.9598627				
		b_i	- .8962840	.8962840				
	3	a_i	- .2101141	- .0860231	1.2961372			
		b_i	- .4343419	- .3642463	.7985882			
	4	a_i	- .0153832	.0519642	.1520750	.8113440		
		b_i	- .2730342	- .2499429	- .1491094	.6720865		
	5	a_i	.0583502	.1088236	.1676091	.2462831	.4189341	
		b_i	- .1844826	- .1816564	- .1304534	- .0065354	.5031278	
6	2	a_i	-1.1655650	2.1655650				
		b_i	- .9141358	.9141358				
	3	a_i	- .3153968	- .2034315	1.5188283			
		b_i	- .4466018	- .3886492	.8352510			
	4	a_i	- .0865378	- .0280534	.0649390	1.0496521		
		b_i	- .2858647	- .2654739	- .1858756	.7372142		
	5	a_i	.0057311	.0465729	.1002523	.1722784	.6751653	
		b_i	- .2015431	- .1972753	- .1536040	- .0645894	.6170118	
	6	a_i	.0488669	.0835221	.1210527	.1656192	.2254909	.3554481
		b_i	- .1458072	- .1495332	- .1267277	- .0731937	.0359868	.4592751

TABLE B-3

Variance $Q_{n,k}\beta^2$ 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 $t_{.90} = x_{10} = \ln L_{10}$, $t_{.50} = x_{50} = \ln L_{50}$, respectively, for a censored sample of n=2 to 6. Variances in units of β^2.

$$Q_{n,k} = A\, y_F^2 + B\, y_F + C\ ^*$$

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

n	k	A	B	C	Q_{10}	Q_{50}
2	2	.6595467	.0643216	.7118574	3.975015	.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	.4331573	.1180273	.3922328	1.888278	.399329
	4	.2934587	- .0346903	.2252828	1.590460	.349150
5	2	1.7891720	1.0115594	.8950462	1.769068	1.167910
	3	.5293953	.2353740	.4168155	1.580861	.412852
	4	.2918142	.0385708	.2537913	1.403458	.297633
	5	.2313953	- .0339905	.1666472	1.228307	.278697
6	2	2.2440055	1.2082248	.9132926	1.431164	1.481035
	3	.6529409	.3332488	.4321160	1.341381	.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}, $y_F = -2.25037$; for Q_{50}, $y_F = -0.36651$.

TABLE B-4

Efficiency of order-statistics estimator of logarithmic life $x_{10} = \ln L_{10}$ and $x_{50} = \ln L_{50}$ 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_{50}, e from endurance data for a given test group.

Test Group No. 1-1, SKF (Deep Groove) Bearing No. 6309

Endurance, million revs.			Natural Logarithms x_i	Weights	
Observed Order	Randomized Order	Ascending Order Within Subgroups		a_i	b_i
(1)	(2)	(3)	(4)	(5)	(6)
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
			$T_{6,6}$: $\sum_1^k a_i x_i = 4.817310$,	$\sum_1^k b_i x_i = 0.400992$	
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	$x_5 = 4.61935$	.675165	.617012
89.0	6.0	(150.0)→	(x_6)		
			$T_{6,5}$: $\sum_1^k a_i x_i = 4.446363$,	$\sum_1^k b_i x_i = 1.213655$	
109.0	74.0	18.0	$x_1 = 2.89037$	Same as for first subgroup	
118.0	118.0	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	$x_6 = 4.96981$		
			$T_{6,6}$: $\sum_1^k a_i x_i = 4.628081$,	$\sum_1^k b_i x_i = 0.619440$	

Continued next page

TABLE B-5, continued

Example showing use of order statistics method of computing values of L_{10}, L_{50}, e from endurance data for a given test group.

Test Group No. 1-1, SKF (Deep Groove) Bearing No. 6309

Endurance, million revs.			Natural Logarithms x_i	Weights	
Observed Order	Randomized Order	Ascending Order Within Subgroups		a_i	b_i
(1)	(2)	(3)	(4)	(5)	(6)
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
		74.0	$x_3 = 4.30407$	1.518828	.835251
151.0	(153.0)→	(153.0)→	(x_4)		
(153.0)→	(153.0)→	(153.0)→	(x_5)		
		(153.0)→	(x_6)		
(153.0)→	50.5				
(153.0)→	74.0		$T_{6,3}$:	$\sum_{i}^{k} a_i x_i = 4.831197$,	$\sum_{i}^{k} b_i x_i = 0.784853$

$\bar{T}_1 = 4.680738$ $\bar{T}_2 = 0.754735$

SUMMARY

$\bar{T}_1 = \hat{u} = 4.680738$ $\bar{T}_2 = \hat{\beta} = 0.754735$ $e = 1/\hat{\beta} = 1.32497$

$y_{.90} = -2.25037$

$y_{.50} = -0.36651$

$\hat{x}_{10} = \ln L_{10} = \hat{u} + y_{.90}\hat{\beta} = 2.982305$, $L_{10} = 19.7333$

$\hat{x}_{50} = \ln L_{50} = \hat{u} + y_{.50}\hat{\beta} = 4.404120$, $L_{50} = 81.7872$

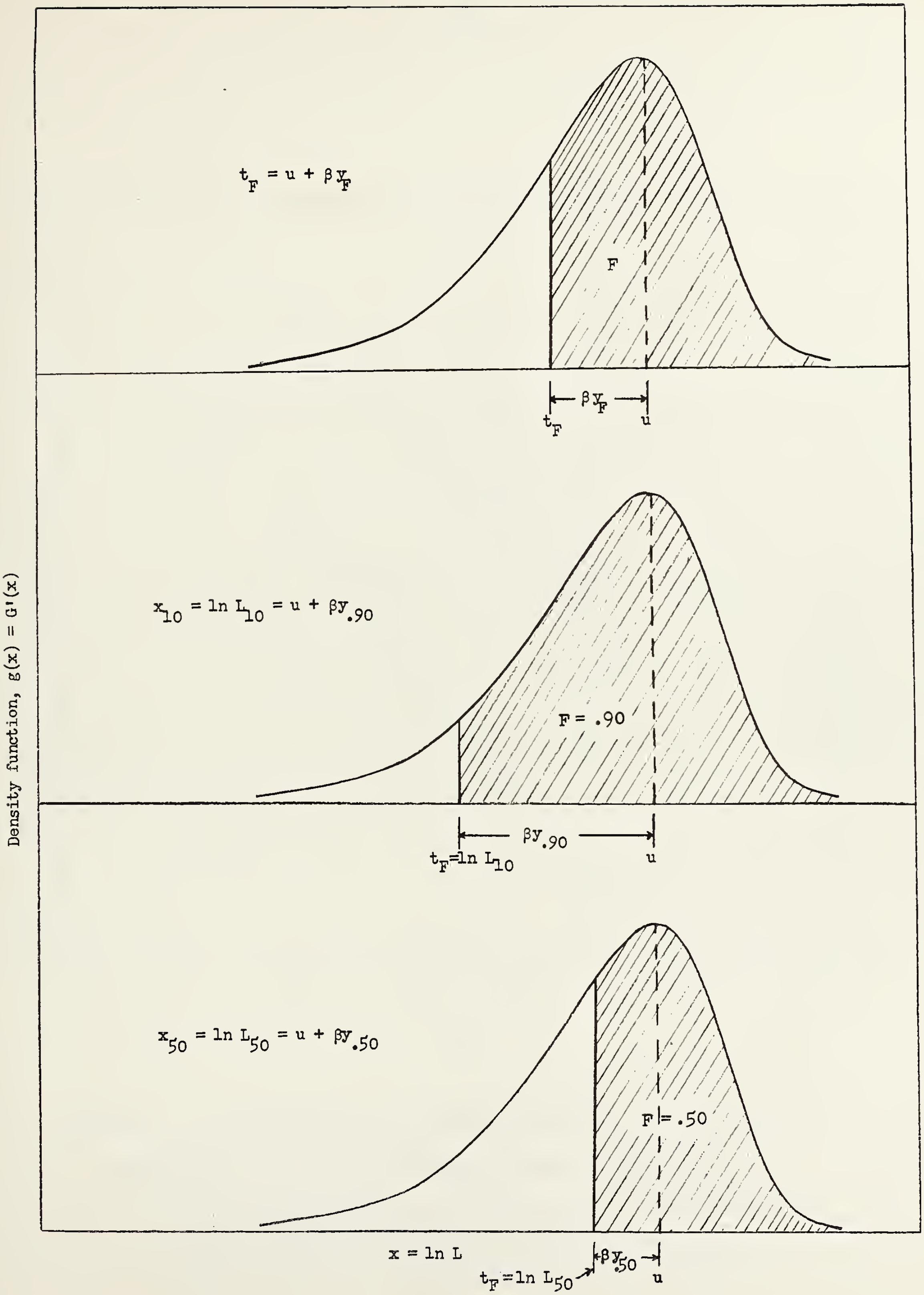


Figure 1. General form of extreme-value distribution (for smallest values) showing relationship of parameters t_F, $x_{10} = \ln L_{10}$, and $x_{50} = \ln L_{50}$, to u and β.

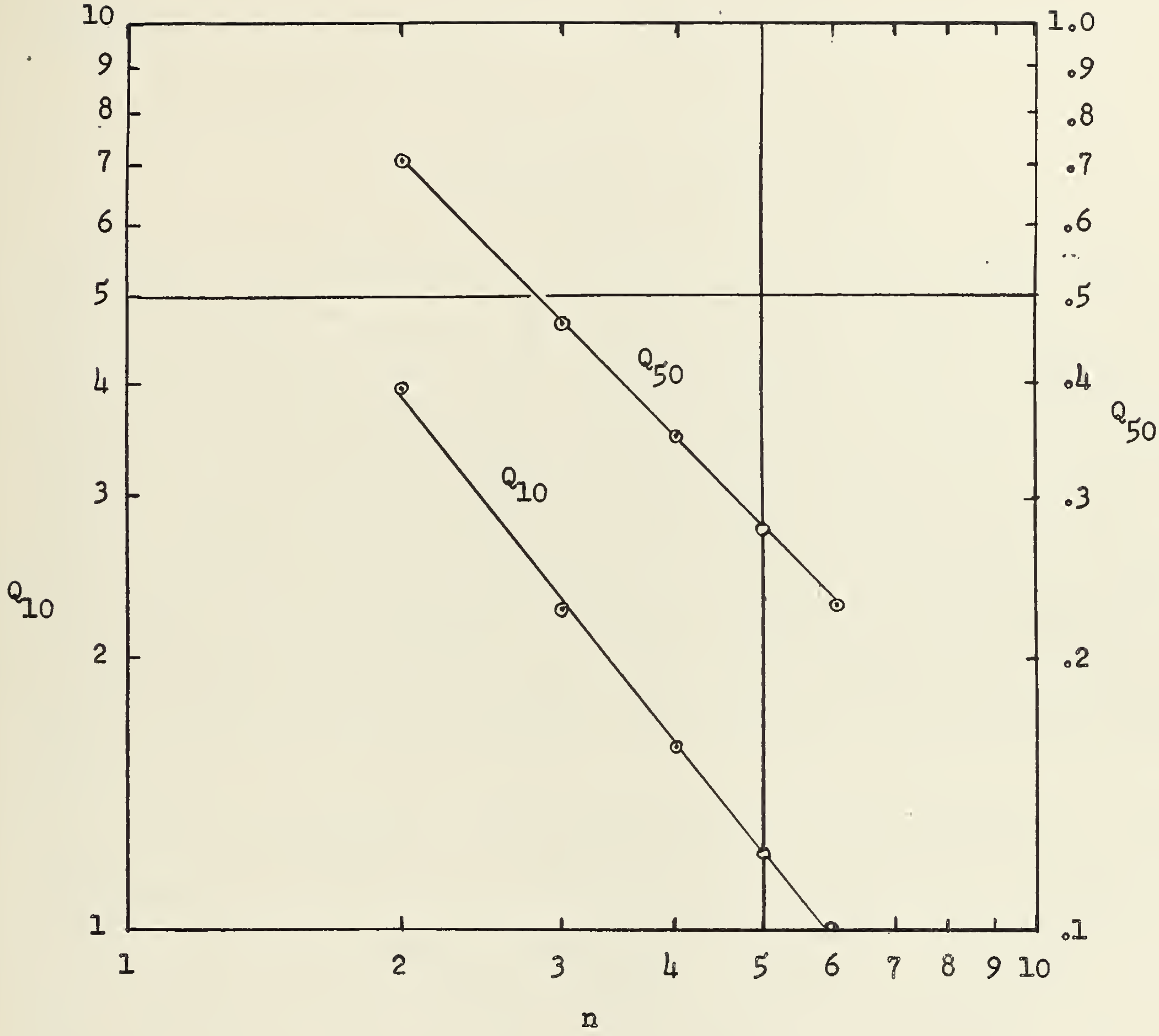


Figure 2. Relationship of variances Q_{10} and Q_{50} to sample size n for n = 2 to 6 (double-logarithmic scale).

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 β^2.)

APPENDIX C

EVALUATION AND ANALYSIS OF THE UNKNOWN PARAMETERS IN THE LIFE EQUATION WITH RESPECT TO COMPANIES AND BEARING TYPES

Summary

Equation (2) of the main text, namely

$$L = \left[\frac{f_c \; Z^{a_1} \; D_a^{a_2} \; (i \cos \alpha)^{a_3}}{P}\right]^p ,$$

expresses the dependence of fatigue-life L on the design characteristics of the bearing (Z, D_a, i, and α), 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, a_1, a_2, a_3, 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 to us by the Committee, 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 basic endurance tests data as described in APPENDIX B.

Section 1 outlines, and summarizes the application of the statistical methods used to determine best empirical values, and intervals of uncertainty, for the parameters f_c, a_1, a_2, and p in the case of deep-groove bearings, for which $i \cos \alpha = 1$, so that the parameter a_3 may be ignored. Section 6 gives the corresponding analysis for the case of self-aligning bearings, for which $i \cos \alpha$ varies, so that a_3 may not be ignored.

Sections 2-5 outline the statistical analyses employed to answer various questions about the values of f_c, a_1, a_2, and p for the deep-groove bearings of SKF, New Departure, and Fafnir. In particular, section 2 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 3 gives the analysis appropriate to determining whether the data are consistent with the supposition that the value of $\underline{p}$ 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 $\underline{p}$ is equal to 3.

Section 4 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 (Extra light, Light, and Medium) for which New Departure data 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 5 has to do (a) with the extent to which the L_{10} and L_{50} values for deep-groove bearings of SKF, New Departure, and Fafnir 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 A/P 1947; and (b), with the determination of more precise values for f_c in those cases in which the foregoing supposition is supported.

Finally, Section 7 presents an analysis of the consistency of the semi-empirical values for $C = f_c\ Z^{0.7}\ D_a^{1.8}\ (i \cos \alpha)^{.7}$ currently used by SKF, with the L_{10} values derived from the SKF data for deep-groove and self-aligning bearings by the method of APPENDIX B.

1. <u>Determination of "best" values for the parameters and their associated intervals of uncertainty</u>.

As shown in Section III of the main text, if (natural) logarithms are taken of both sides of the life equation (equation 2), the resulting equation expresses the (natural) logarithm of rating life L_{10} (or median life L_{50}) as a linear function of the (natural) logarithms of the characteristics of the bearing (Z, D_a, i, and α), and the bearing load P, with coefficients that are simple functions of the "workmanship factor" f_c, and the exponents a_1, a_2, a_3, and p. Furthermore, in the case of deep-groove bearings, $i = 1$ and $\alpha = 0^{\circ}$ throughout, so that the term in $\ln (i \cos \alpha)$ vanishes identically, irrespective of the value of a_3. Consequently for deep-groove bearings, the <u>logarithmic form of the life equation</u> is

(C-1) $$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

where

(C-2) $$\begin{cases} Y = \ln L \\ x_1 = \ln Z \\ x_2 = \ln D_a \\ x_3 = \ln P \end{cases}$$

and

(C-3) $$\begin{cases} b_0 = p \ln f_c = p\, a_0 \\ b_1 = p\, a_1 \\ b_2 = p\, a_2 \\ b_3 = -p \end{cases}$$

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 <u>variance</u>, 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, that is

$$\text{(C-4)} \qquad \begin{cases} \underline{\text{mean}} \text{ of } y = Y \\ \underline{\text{variance}} \text{ of } y = \dfrac{\sigma^2}{w}, \\ \text{i.e., standard deviation} = \sigma/\sqrt{w} \end{cases}$$

where σ^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 [5], Kempthorne [6], Mood [7], and Wilks [9] give exten-

* Numbers in square brackets refer to references at the end of this Appendix.

sive discussion of these techniques. For completeness some of the techniques and rationale of regression analysis bearing on the work enbodied 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 (α = 1, 2, ..., n) where y_α is the estimate of $\ln L_{10}$ or $\ln L_{50}$, $x_{1\alpha}$, $x_{2\alpha}$ are the logarithms of the bearing characteristics Z and D_a, $x_{3\alpha}$ is the logarithm of the load, and w_α is the number of individual bearings tested for α th test group; to estimate the values of the parameters b_0, b_1, b_2, and b_3, in equation (C-1) using some optimum method of estimation.

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

$$Q = \sum_{\alpha=1}^{n} w_\alpha (y_\alpha - b_0 - b_1 x_{1\alpha} - b_2 x_{2\alpha} - b_3 x_{3\alpha})^2 \tag{C-5}$$

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

---

* Estimates having these properties are usually termed minimum variance unbiased estimates, and the proof of the fact that the procedure used here leads to such estimates is sometimes referred to as the Gauss-Markoff theorem. Its relation to Gauss' development of the Method of Least Squares is given by Plackett [8].

The resulting equations defining the parameter estimates are termed the normal equations and can be written in the form*

(C-6)
$$\begin{cases} a_{00}\hat{b}_0 + a_{01}\hat{b}_1 + a_{02}\hat{b}_2 + a_{03}\hat{b} = g_0 \\ a_{10}\hat{b}_0 + a_{11}\hat{b}_1 + a_{12}\hat{b}_2 + a_{13}\hat{b}_3 = g_1 \\ a_{20}\hat{b}_0 + a_{21}\hat{b}_1 + a_{22}\hat{b}_2 + a_{23}\hat{b}_3 = g_2 \\ a_{30}\hat{b}_0 + a_{31}\hat{b}_1 + a_{32}\hat{b}_2 + a_{33}\hat{b}_3 = g_3 \end{cases}$$

where

(C-7) $$a_{ij} = \sum_{\alpha=1}^{n} w_\alpha \; x_{i\alpha} \; x_{j\alpha} \qquad i, j = 0, 1, 2, 3.$$

(C-8) $$g_i = \sum_{\alpha=1}^{n} w_\alpha \; x_{i\alpha} \; y_\alpha \qquad i = 0, 1, 2, 3.$$

and $x_{0\alpha} = 1$ for all α.

If the coefficient matrix $\|a_{ij}\|$ of the normal equations is not singular, then unique solutions will exist for the $\hat{b}_i$. These can be written as

(C-9) $$\hat{b}_i = \sum_{j=0}^{3} c_{ij} \, g_j \qquad \text{for} \quad i = 0, 1, 2, 3.$$

where c_{ij} (i, j = 0, 1, 2, 3) are the elements of the inverse matrix to $\|a_{ij}\|$.

---

*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 variances and covarances for the b's are estimated by

(C-10) $$\begin{cases} \text{variance } \hat{b}_i = c_{ii}\, s^2 & i = 0, 1, 2, 3. \\ \text{covariance } (\hat{b}_i, \hat{b}_j) = c_{ij}\, s^2 & i \neq j = 0, 1, 2, 3. \end{cases}$$

where the quantity s^2 is an estimate of the constant σ^2 in the second of equations C-4. 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

(C-11) $$S = \sum_{\alpha=1}^{n} w_\alpha y_\alpha^2 - \sum_{i=0}^{3} \hat{b}_i g_i ,$$

then the optimum estimate of σ^2 is

(C-12) $$s^2 = \frac{S}{n-4} .$$

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

The above method of estimation does not depend on y_α having a particular assumed probability distribution. All that is necessary to specify about the probability distribution of y_α is that it possess a finite mean and variance. The solution of the normal equations (C-6) has the property that the weighted sum of squares of the deviations about y_α 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 which 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 using either the ln L_{10} or ln L_{50} values for y_α. 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_α. However something more must be assumed about the distribution of y_α 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_α will not be of this form. However the estimate y_α (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_α will be approximated by a normal distribution when n is large (cf. Cramer [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 estimate y_α follow a normal distribution.

The intervals of uncertainty calculated for each parameter are 95 per cent 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 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. Finney [4], p. 27). The main result states that confidence intervals for the a_i are the roots of a quadratic equation, which for the problem considered here, can be written as

$$\text{(C-13)} \qquad \theta^2 \left[\hat{p}^2 - t^2 c_{33} s^2\right] + 2\,\theta \left[\hat{p}\,\hat{b}_i - t^2 c_{i3} s^2\right] + \left[\hat{b}_i^2 - t^2 c_{ii} s^2\right] = 0$$

where $\hat{p} = -\hat{b}_3$ and t^2 is the square of "Student's t" (tabulated in most statistical texts) for the degrees of freedom associated with s^2.

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 report. These are all based on a test statistic F, termed the variance ratio or F-ratio, which takes the form

$$F = \frac{Q_1/\nu_1}{Q_2/\nu_2}$$

where Q_1 and Q_2 are quadratic forms calculated from the data and ν_1, ν_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 which employ a variance ratio statistic also give the explicit expressions for the two quadratic forms.

The distribution of the variance ratio is tabulated in most statistical texts for the case where the hypothesis tested is true. This distribution depends only on the degrees of freedom of the numerator and denominator, i.e., ν_1 and ν_2. If the hypothesis being tested is true, then the calculated variance ratio will deviate from unity in accordance with the 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.

Since the calculated variance ratio is based upon the results of endurance tests which themselves are subject to considerable variability, the calculated variance ratio may exceed unity by chance even if the hypothesis in question is true. In order to objectively determine if a calculated variance ratio is significantly greater than unity, one selects a critical value of F from tables of the variance ratio distribution, 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 in all variance ratio tests here, has been selected such that there is only a probability of 0.05 of its being exceeded by a calculated variance ratio from purely chance causes. This critical value will be denoted by $F_{.05}(\nu_1, \nu_2)$. More extensive discussions of these procedures are found in Kempthorne [6], Chapter 5.

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

2. Analysis to determine whether companies have common values for all the parameters in the life equation (deep-groove).

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 (equation C-19), which is used to test this hypothesis, is obtained from the following procedures:

First, a single set of parameters b_i (i = 0, 1, 2, 3) is obtained by fitting all the data, irrespective of company, to the

logarithmic life equation and then calculating the resulting residual sums of squares S (equation C-17) 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 (equation C-16) 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., $\left\{S - (S_1 + S_2 + S_3)\right\}$, 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 $\left\{S - (S_1 + S_2 + S_3)\right\}$ will be large. To determine whether this difference is statistically significant the variance ratio (C-19) is employed.

Mathematical formulation

It will be convenient to adopt the following notation: let the superscript u = 1, 2, 3 refer to SKF, New Departure, and Fafnir respectively. Also, for each manufacturer, let $\hat{b}_i^{(u)}$ for i=0,1,2,3 refer to the estimates of the parameters in equation (C-1), and $a_{ij}^{(u)}$, $g_i^{(u)}$ denote the sums of cross products defined in (C-7) and (C-8).* Then the normal equations which give the parameter estimates for the uth company are

* Since it is only the ratio of the number of observations which is important for weighting, the weights w used in the calculation of the sums of cross products 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. This is the way the weights were taken in all analyses.

$$\text{(C-14)} \qquad \sum_{j=0}^{3} a_{ij}^{(u)} \hat{b}_j^{(u)} = g_i^{(u)} \qquad (i = 0, 1, 2, 3)$$

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 (C-3). These results are summarized in Tables Va, Vb and II respectively in the main text.

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

$$\text{(C-15)} \qquad b_i^{(1)} = b_i^{(2)} = b_i^{(3)} \qquad \text{for } i = 0, 1, 2, 3.$$

Define the residual sum of squares for the uth company by

$$\text{(C-16)} \qquad S_u = \sum_{\alpha=1}^{n_u} w_{u\alpha} y_{u\alpha}^2 - \sum_{i=0}^{3} \hat{b}_i^{(u)} g_i^{(u)} \qquad \text{for } u = 1, 2, 3,$$

and let

$$\text{(C-17)} \qquad S = \sum_{u=1}^{3} \sum_{\alpha=1}^{n_u} w_{u\alpha} y_{u\alpha}^2 - \sum_{i=0}^{3} \hat{b}_i g_i \ , \quad g_i = \sum_{u=1}^{3} g_i^{(u)} .$$

where $\hat{b}_i$ (i = 0, 1, 2, 3) are the estimates obtained from the solutions of the normal equations without regard to company differences, i.e.,

$$\text{(C-18)} \qquad \sum_{j=0}^{3} \left[\sum_{u=1}^{3} a_{ij}^{(u)} \right] \hat{b}_j = \sum_{u=1}^{3} g_i^{(u)} \qquad (i = 0, 1, 2, 3).$$

Then the variance ratio for testing the hypothesis given by (C-15) is

$$\text{(C-19)} \qquad F = \frac{(S - S_1 - S_2 - S_3)/8}{(S_1 + S_2 + S_3)/198}$$

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

Numerical calculations

The numerical values for $a_{ij}^{(u)}$ are shown in Tables C-1, C-2, and C-3 where the $a_{ij}^{(u)}$ are given as a triangular array, e. g.,

$$\text{(C-20)} \quad a^{(u)} = \begin{bmatrix} a_{00}^{(u)} & a_{01}^{(u)} & a_{02}^{(u)} & a_{03}^{(u)} \\ & a_{11}^{(u)} & a_{12}^{(u)} & a_{13}^{(u)} \\ & & a_{22}^{(u)} & a_{23}^{(u)} \\ & & & a_{33}^{(u)} \end{bmatrix}$$

The values for $g_i^{(u)}$ ($i = 0, 1, 2, 3$) and $\sum_{=1}^{n_u} w_u \ y_u^2$ ($u = 1, 2, 3$) are summarized in Tables C-4 and C-5 for L_{10} and L_{50} respectively. Table C-6 summarizes the values for S_u and s_u^2 ($u = 1, 2, 3$).

The values of the F statistic (equation C-19) calculated from the quantities given in Table C-6 are

$$\text{(C-21)} \quad \begin{cases} \text{for } L_{10}\text{:} & F = 3.32 \\ \text{for } L_{50}\text{:} & F = 3.98 \end{cases}$$

These calculated F values are both larger than the critical value, $F_{.05}$ (8, 198) = 1.98; actually the probability is less than 0.0005 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 the parameters in the life equation.

TABLE C-1. VALUES OF $a_{ij}^{(1)}$ FOR SKF

$$a^{(1)} = \begin{bmatrix} 252.0000000 & 533.2092214 & -125.9088049 & 2052.028218 \\ & 1129.678133 & -267.6498233 & 4340.657154 \\ & & 86.37042375 & -981.0460941 \\ & & & 16816.18212 \end{bmatrix}$$

TABLE C-2. VALUES OF $a_{ij}^{(2)}$ FOR NEW DEPARTURE

$$a^{(2)} = \begin{bmatrix} 668.0000000 & 1516.872777 & -598.2684700 & 5027.985042 \\ & 3459.745060 & -1362.350294 & 11420.85716 \\ & & 617.5688586 & -4337.475983 \\ & & & 38223.15468 \end{bmatrix}$$

TABLE C-3. VALUES OF $a_{ij}^{(3)}$ FOR FAFNIR

$$a^{(3)} = \begin{bmatrix} 58.00000000 & 116.8290299 & -37.23287485 & 422.3706781 \\ & 235.9119681 & -75.38686387 & 851.2083966 \\ & & 24.72718606 & -270.8359778 \\ & & & 3082.079643 \end{bmatrix}$$

TABLE C-4. SUMMARY OF $g_i^{(u)}$ AND $\sum_{\alpha=1}^{n_u} w_{u\alpha} \; y_{u\alpha}^2$ BY COMPANIES FOR L_{10}

	SKF	New Departure	Fafnir
g_0	596.2960159	1757.380696	174.9101606
g_1	1260.741544	3990.060572	349.2541794
g_2	-310.4200596	-1631.001443	-109.6707412
g_3	4763.812166	13016.17843	1255.435105
$\sum_{\alpha=1}^{n_u} w_{u\alpha} \; y_{u\alpha}^2$	1721.123970	5206.3983364	602.1702144

TABLE C-5. SUMMARY OF $g_i^{(u)}$ AND $\sum_{\alpha=1}^{n_u} w_{u\alpha} \; y_{u\alpha}^2$ BY COMPANIES FOR L_{50}

	SKF	New Departure	Fafnir
g_0	987.7087610	2605.650712	258.9632205
g_1	2087.560362	5916.331934	519.2863296
g_2	-506.1125459	-2367.253498	-163.5201590
g_3	7948.474829	19453.90754	1869.950560
$\sum_{\alpha=1}^{n_u} w_{u\alpha} \; y_{u\alpha}^2$	4181.969283	10613.239803	1220.597660

TABLE C-6. 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	s_u^2	S_u	s_u^2
Companies combined	206	446.684633		348.285324	
SKF	46	102.552212		89.734079	
New Dep.	144	285.121962		201.935797	
Fafnir	8	6.160858		8.357822	
sum	198	393.835032	1.9891	300.027698	1.5153
Difference	8	52.849601	6.6062	48.257626	6.0322
F (equation C-14)		$\frac{6.6062}{1.9891} = 3.32$		$\frac{6.0322}{1.5153} = 3.98$	

3. Analysis to determine whether companies have a common value for the exponent p (deep groove)

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 (C-22), 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'' (equation C-25) having 200 degrees of freedom, 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 2 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 (C-26) is employed.

Mathematical formulation

The logarithmic life equation, having a common value for the exponent p, can be written for the α^{th} test group in the u^{th} company as

$$(C\text{-}22) \qquad Y_{u\alpha} = b_o^{(u)} + b_1^{(u)} x_{1\alpha}^{(u)} + b_2^{(u)} x_{2\alpha}^{(u)} + b_3^{(u)} x_{3\alpha}^{(u)}$$

$$\text{for } \alpha = 1, 2, \ldots, n_u; \; u = 1, 2, 3.$$

Note that although each company has the same parameter b_3 in the above equation, the parameters $b_o^{(u)}$, $b_1^{(u)}$, 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

$$(C\text{-}23) \qquad \begin{cases} \sum_{j=0}^{2} a_{ij}^{(u)} \hat{b}_j^{(u)} + a_{i3}^{(u)} \hat{b}_3 = g_i^{(u)} & (i=0,1,2;\ u=1,2,3) \\ \sum_{u=1}^{3} \sum_{j=0}^{2} a_{3j}^{(u)} \hat{b}_j^{(u)} + a_{33} \hat{b}_3 = g_3 \end{cases}$$

where

$$a_{33} = \sum_{u=1}^{3} a_{33}^{(u)} , \qquad g_3 = \sum_{u=1}^{3} g_3^{(u)} .$$

Thus the set of equations given by (C-23) 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

$$(C\text{-}24) \qquad \hat{a}_i^{(u)} = -\frac{\hat{b}_i^{(u)}}{\hat{b}_3} \qquad \text{for } i = 0,1,2; \quad u = 1,2,3.$$

The residual sums of squares (denoted by S') associated with fitting the life formula (C-22) to the data is given by

$$(C\text{-}25) \qquad S' = \sum_{u=1}^{3} \sum_{\alpha=1}^{n_u} w_{u\alpha} y_{u\alpha}^2 - \sum_{u=1}^{3} \sum_{i=0}^{2} \hat{b}_i^{(u)} g_i^{(u)} - \hat{b}_3 g_3$$

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

$$(C\text{-}26) \qquad F = \frac{(S' - S_1 - S_2 - S_3)/2}{(S_1 + S_2 + S_3)/198}$$

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

Numerical calculations

All the quantities needed for the normal equations (C-23) are summarized in tables C-1, 2, 3, 4, 5. The values of S' for both L_{10} and L_{50} are

$$(C-27)\quad \begin{cases} L_{10}: & S' = 393.272847 & d.f. = 200 \\ L_{50}: & S' = 301.687871 & d.f. = 200 \end{cases}$$

Thus the calculated F values (using equation C-26) give

$$(C-28)\quad \begin{cases} L_{10}: & F = -\ 0.141^{*} \\ L_{50}: & F = 0.548 \end{cases}$$

Since 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

---

* From theory, the calculated value for the F-ratios can never be negative. The reason for the negative value of F for L_{10} is that the value for the numerator of (C-26) is only accurate numerically to one decimal place on account of round off errors arising from the solution of the normal equations (C-23). Thus if the hypothesis of a common p value is true, then the F-ratio will not be large and round off errors may effect the resulting calculation. Alternatively, if the null hypothesis was false, then the calculated F-ratio would be larger then 3.04 and the round off error should be of no consequence.

table I 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 (C-24). These results are summarized in tables IVa and IVb of the main text.

4. Analysis to determine whether three New Departure bearing types, i.e., Extra light, Light, and Medium, have common values for all the parameters in the life equation (deep-groove)

The analyses, discussed in previous sections, dealt with determining whether where are differences in the parameters of the life equation between companies. This section investigates (a) whether three different bearing types made by New Departure 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 2 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 New Departure data, irrespective of bearing type. The variance ratio for statistically testing (a) is given by equation (C-32).

The analyses 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 C-33.

Mathematical formulation

The 148 test groups from New Departure can be divided into three bearing types corresponding to 37 groups for Extra light, 94 groups for Light, and 17 groups for Medium bearings. Let the bearing type be denoted by $v = 1,2,3$ for Extra light, Light, and Medium type bearings, respectively. Also define

$$\text{(C-29)} \quad \begin{cases} A_{ij}^{(v)} = \sum_{\alpha=1}^{n_v} w_{v\alpha} \, x_{i\alpha}^{(v)} \, x_{j\alpha}^{(v)} & i,j = 0,1,2,3; \quad v=1,2,3. \\ G_{i}^{(v)} = \sum_{\alpha=1}^{n_v} w_{v\alpha} \, x_{i\alpha}^{(v)} \, y_{v\alpha}^{(v)} & i = 0,1,2,3; \quad v = 1,2,3. \end{cases}$$

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

$$(C-30) \qquad \sum_{j=0}^{3} A_{ij}^{(v)} \hat{b}_j^{(v)} = G_j^{(v)} \qquad i = 0,1,2,3; \quad v = 1,2,3.$$

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

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

$$\hat{p}^{(v)} = -b_3^{(v)}$$

Therefore the residual sums of squares for the v^{th} bearing type is

$$(C-31) \qquad S^{(v)} = \sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2 - \sum_{i=0}^{3} \hat{b}_i^{(v)} G_i^{(v)} \qquad v = 1,2,3.$$

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

$$(C-32) \qquad F = \frac{(S_2 - S^{(1)} - S^{(2)} - S^{(3)})/3}{(S^{(1)} + S^{(2)} + S^{(3)})/136}$$

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

Now since the analysis given in section 3 of this appendix reached the conclusion that all companies have a

common value for p, and since this value (cf. table I of main text) with its associated uncertainty includes the value p = 3 given in A/P 1947, it seems desirable to also test a second hypothesis that the values of p for each of the bearing types is consistent with p = 3. The F-ratio for this hypothesis is given by

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

where the $\hat{p}^{(v)}$ (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 $\|A_{ij}^{(v)}\|$ (v = 1,2,3). The critical value for the variance ratio (C-33) is $F_{.05}(3,136) = 2.67$.

Numerical results

The estimates of $a_i^{(v)}$ (i = 0,1,2) and $p^{(v)}$ obtained from the solutions of the normal equations (C-30) are summarized in tables VIa, VIb, and III in the main text, respectively. Tables C-7, 8 and 9 summarize the values for $A_{ij}^{(v)}$ in the form of a triangular array. Table C-10 and C-11 summarize the values for $G_i^{(v)}$ and $\sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2$. The

TABLE C-7. VALUES OF $A_{ij}^{(1)}$ FOR EXTRA LIGHT BEARINGS

$A^{(1)} =$

214.0000000	518.1383084	-248.9295851	1506.575333
	1258.277638	-597.8263023	3661.658351
		300.7637617	-1725.381891
			10693.46018

TABLE C-8. VALUES OF $A_{ij}^{(2)}$ FOR LIGHT BEARINGS

$A^{(2)} =$

401.0000000	889.3252551	-325.2355955	3083.034363
	1975.516399	-715.3345135	6852.948565
		294.2929935	-2433.705085
			23862.08756

TABLE C-9. VALUES OF $A_{ij}^{(3)}$ FOR MEDIUM BEARINGS

$A^{(3)} =$

53.00000000	109.4092134	-24.10328938	438.3753462
	225.9510228	-49.18947836	906.2502425
		22.51210338	-178.3890070
			3667.606940

TABLE C-10. SUMMARY OF $G_i^{(v)}$ AND $\sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2$ BY BEARING TYPE FOR L_{10}

	Extra Light	Light	Medium
G_0	602.0143335	1027.110445	128.2559173
G_1	1451.867253	2273.867843	264.3254755
G_2	-710.5400693	-854.1381299	-66.32324354
G_3	4148.761977	7826.050226	1041.366231
$\sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2$	1984.783747	2885.087835	336.5267544

TABLE C-11. SUMMARY OF $G_i^{(v)}$ AND $\sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2$ BY BEARING TYPE FOR L_{50}

	Extra Light	Light	Medium
G_0	862.7854631	1537.980372	204.8848770
G_1	2087.149852	3406.325505	422.8565766
G_2	-1006.992510	-1263.434071	-96.82691696
G_3	6004.772590	11767.86813	1681.266822
$\sum_{\alpha=1}^{n_v} w_{v\alpha} y_{v\alpha}^2$	3709.734322	6081.616419	821.8890620

calculations for the variance ratio (C-32) are summarized in table C-12.

TABLE C-12. SYNOPSIS OF CALCULATIONS TO DETERMINE WHETHER ALL BEARING TYPES HAVE COMMON PARAMETERS IN LIFE EQUATION (NEW DEPARTURE ONLY)

Type of Bearing	Degrees of Freedom	L_{10}		L_{50}	
		$s^{(v)}$	$s^{(v)^2}$	$s^{(v)}$	$s^{(v)^2}$
All types combined	144*	285.121962*		201.935797*	
Extra light	33	68.558298		33.016973	
Light	90	174.325939		131.142017	
Medium	13	11.537480		8.505638	
Sum	136	254.421716	1.8707	172.664628	1.2696
Difference	8	30.700246	3.8375	29.271169	3.6589
F (equation C-32)		$\frac{3.8375}{1.8707} = 2.05$		$\frac{3.6589}{1.2696} = 2.88$	

* From table C-6

Corresponding to the hypothesis that the three bearing types have the same parameters in the life equation, the calculated variance ratios (C-32) yield

$$\text{(C-34)} \quad \begin{cases} L_{10}: & F = 2.05 \\ L_{50}: & F = 2.88 \end{cases}$$

Since the critical F value is $F_{.05}(8,136) = 2.01$, both the L_{10} and L_{50} calculated variance ratios are statistically significant, although the L_{10} value is "just barely" significant. Therefore, one would conclude from the above F-ratios that the 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 equation (C-33) results in

$$\text{(C-35)} \quad \begin{cases} L_{10}: & F = 1.18 \\ L_{50}: & F = 2.32 \end{cases}$$

where

$$C_{33}^{(1)} = 0.53294$$

$$C_{33}^{(2)} = 0.107395$$

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

Thus since 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 $a_i^{(v)}$ $(i = 0,1,2; \quad v = 1,2,3)$ 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_a. 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_a. Thus the estimates for a_i based on all bearing types for New Departure (tables IVa, IVb, or Va, Vb) have substantially smaller confidence intervals as compared to the intervals based only on a single bearing type.

5. Determination of f_c assuming values $a_1 = 2/3$, $a_2 = 1.8$, $p = 3$; and the analysis, both by company and bearing type to determine whether these assumed parameter values are consistent with the given data (deep-groove).

The values for the parameters a_1, a_2, and p given in A/P 1947 are

$$a_1 = 2/3; \quad a_2 = 1.8, \quad p = 3.$$

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 A/P 1947 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 A/P 1947 values for the other parameters.

The procedure for determining whether the values $a_1 = 2/3$, $a_2 = 1.8$, $p = 3$ 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 (equation C-42), can be calculated having (n-1) degrees of freedom. Alternatively, the life equation can be fitted to the data such that <u>all</u> 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 A/P 1947 parameter values are not consistent with the given data, R will be appreciably larger than S. The variance ratio (C-43) is used to determine if the difference between these two residual sum 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 αth test group can be written as

(C-36) $$Y_\alpha = 3 \left\{ a_0 + x_\alpha \right\} \qquad (\alpha = 1, 2, \ldots, n),$$

where

(C-37) $$x_\alpha = \left[\frac{2}{3} x_{1\alpha} + 1.8\, x_{2\alpha} - x_{3\alpha}\right] .$$

The resulting normal equation for estimating a_0 is

(C-38) $$\hat{a}_0 = \frac{\sum_{\alpha=1}^{n} \frac{w_\alpha y_\alpha}{3} - \sum_{\alpha=1}^{n} w_\alpha x_\alpha}{\sum_{\alpha=1}^{n} w_\alpha}$$

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

(C-39) $$\hat{a}_0 = \frac{1}{a_{00}} \left\{ \frac{g_0}{3} - \left(\frac{2}{3} a_{01} + 1.8\, a_{02} - a_{03}\right) \right\} .$$

Hence the variance for this estimate of a_0 is

(C-40) $$\text{variance } \hat{a}_0 = \frac{s^2}{9a_{00}}$$

where

(C-41) $$s^2 = \frac{R}{n-1} = \frac{9\left[\sum_{\alpha=1}^{n} w_\alpha \left(\frac{y_\alpha}{3} - x_\alpha\right)^2 - a_{00}\, \hat{a}_0^2\right]}{n - 1}$$

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, e.g.,

(C-42) $$R = 9 \left\{ \frac{1}{9} \sum_{\alpha=1}^{n} w_\alpha y_\alpha^2 - \frac{4}{9} g_1 - 1.2 g_2 + \frac{2}{3} g_3 + \frac{4}{9} a_{11} + 3.24 a_{22} + a_{33} + 2.4 a_{12} - \frac{4}{3} a_{13} - 3.6 a_{23} - a_{00} \hat{a}_0^2 \right\}$$

In the analyses made in the preceeding 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 $a_1 = 2/3$, $a_2 = 1.8$, $p = 3$ as given in A/P 1947. The variance ratio used to test this hypothesis is

(C-43) $$F = \frac{(R-S)/3}{S/(n-4)} ,$$

where for testing within companies

$$S = \begin{cases} S_1 & \text{for SKF} \\ S_2 & \text{for New Dep.} \\ S_3 & \text{for Fafnir} \end{cases}$$ defined by equation (C-16) and given in Table C-6

and for testing within bearing types, $a_{ij}^{(u)}$ is replaced by $A_{ij}^{(v)}$,

$$S = \begin{cases} S^{(1)} & \text{for Extra light} \\ S^{(2)} & \text{for Light} \\ S^{(3)} & \text{for Medium} \end{cases}$$ defined by equation (C-31) and given in Table C-12

and R (equation C-42) refers to the calculated residual sum of squares within the particular classification.

Numerical calculations

Table C-13 summarizes with respect to SKF, New Departure, and Fafnir the values of $\hat{a}_0$, R, the calculated F-ratio (equation C-43), and the critical F value. Table C-14 summarizes the same quantities for the Extra light, Light, and Medium type bearings made by New Departure.

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 only consistent for the SKF data.

The results for the same analysis made on the three different bearing types, indicate that the A/P 1947 parameter values are only consistent with the Extra light and Light bearing with respect to median life (L_{50}).

Note that the analysis for New Departure (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 Medium type bearings. This apparent inconsistency stems from the fact that the analysis for New Departure, taken as a whole, is dominated by those bearing types having the larger number of test groups, i.e. Extra light, Light.

The estimates for f_c assuming $a_1 = 2/3$, $a_2 = 1.8$, and $p = 3$, are summarized in Table VII of the main text for rating life (L_{10}). This summary also includes the value of f_c for M.R.C. computed from only three test groups. (Because of the small number of test groups, it was not possible to verify whether the A/P 1947 parameter values are valid for these data.)

TABLE C-13. SUMMARY OF COMPUTATIONS (BY COMPANIES) TO TEST HYPOTHESIS THAT DATA ARE CONSISTENT WITH ASSUMED VALUES $a_1 = 2/3$, $a_2 = 1.8$, $a_3 = 3$.

Company	L_{10}			L_{50}			Critical F
	$a_0 = \ln f_c$	R	F	$a_0 = \ln f_c$	R	F	
SKF	8.4205	102.539286	-0.06*	8.9382	94.367916	0.79	2.80
New Dep.	8.5021	287.613198	0.42	8.9254	217.560672	3.71	2.65
Fafnir	8.1001	8.375562	0.96	8.5832	8.488584	9.92	3.59

TABLE C-14. SUMMARY OF COMPUTATIONS (BY BEARING TYPE) TO TEST HYPOTHESIS THAT DATA ARE CONSISTENT WITH ASSUMED VALUES $a_1 = 2/3$, $a_2 = 1.8$, $a_3 = 3$.

Type	L_{10}			L_{50}			Critical F
	$a_0 = \ln f_c$	R	F	$a_0 = \ln f_c$	R	F	
Extra light	8.4575	73.821978	0.84	8.8636	44.883126	3.95	2.86
Light	8.5236	181.128312	1.17	8.9482	141.731604	2.42	2.70
Medium	8.5203	27.014247	5.81	9.0022	18.905517	5.21	3.24

*Negative value due to rounding computations.

6. Determination of the parameter values for self-aligning bearings (SKF)*

This section deals with finding estimates of the parameters in the life equation for self-aligning bearings using data from 40 test groups of SKF self-aligning bearings. The only difference between this and the deep-groove case is that the self-aligning bearings require the additional term $(i \cos\alpha)^{a_3}$ in the life equation. Thus the logarithmic life equation can be written as

(C-44) $$Y_\alpha = b_0 + b_1 x_{1\alpha} + b_2 x_{2\alpha} + b_3 x_{3\alpha} + b_4 x_{4\alpha} \quad (\alpha = 1, 2, \ldots, n),$$

where b_i (i = 0, 1, 2, 3) and x_i (i = 1, 2, 3) have the same meanings as for deep-groove bearings and $b_4 = (pa_3)$, $x_4 = \ln(i \cos \alpha)$. The normal equations for estimating the b_i (i = 0, 1, 2, 3, 4) are

(C-45) $$\sum_{j=0}^{4} a_{ij} \hat{b}_j = g_i \qquad (i = 0, 1, 2, 3, 4)$$

where

$$a_{ij} = \sum_{\alpha=1}^{n} w_\alpha x_{i\alpha} x_{j\alpha}$$

$$g_i = \sum_{\alpha=1}^{n} w_\alpha x_{i\alpha} y_\alpha \qquad (n = 40)$$

Estimates for a_i (i = 0, 1, 2, 3) can be obtained from the $\hat{b}_i$ using the relations

(C-46) $$\begin{cases} \hat{a}_i = -\dfrac{\hat{b}_i}{\hat{b}_3} & \text{for } i = 0, 1, 2 \\ \hat{a}_3 = -\dfrac{\hat{b}_4}{\hat{b}_3}, \quad \hat{p} = -\hat{b}_3 & . \end{cases}$$

*All the notation used in this section will refer to the self-aligning bearings only, unless otherwise specified.

The appropriate intervals of uncertainty for the a_i can be calculated using Fieller's theorem (as given in equation C-13). Where s^2 is taken to be

$$s^2 = \frac{\sum_{\alpha=1}^{n} w_\alpha y_\alpha^2 - \sum_{i=0}^{4} \hat{b}_i g_i}{35} \qquad \text{(C-47)}$$

The estimates for the a_i together with their associated intervals of uncertainty are summarized in Table VIII of the main text. The values for a_{ij} $(i,j = 0, 1, 2, 3, 4)$ are summarized in Table C-15. Table C-16 gives the values for $\sum_{\alpha=1}^{n} w_\alpha y_\alpha^2$ and g_i $(i = 0, 1, 2, 3, 4)$ both for L_{10} and L_{50}.

TABLE C-15. VALUES OF a_{ij} FOR SELF-ALIGNING BEARINGS (SKF ONLY)

238.00000000	648.7845195	-172.0909813	1911.497205	161.5118702
	1770.460480	-472.2281104	5211.817531	440.3712025
		156.9382746	-1362.806927	-116.8392512
			15486.38403	1297.119642
				109.6145790

TABLE C-16. SUMMARY OF $\sum_{\alpha=1}^{n} w_\alpha y_\alpha^2$ AND g_i $(i=0,1,2,3,4)$

SELF-ALIGNING BEARINGS (SKF ONLY)

	L_{10}	L_{50}
g_0	418.8179296	765.4465183
g_1	1139.906247	2082.557309
g_2	-308.8485452	-554.0338357
g_3	3158.390004	5923.501002
g_4	284.2203906	519.3826284
$\sum_{\alpha=1}^{n} w_\alpha y_\alpha^2$	1285.621913	3058.794005

7. Analysis to determine whether values for dynamic capacity ($C = f_c\ Z^{0.7}\ D_a^{1.8}\ (i \cos \alpha)^{0.7}$) in current use by SKF are consistent with data.

The life equation can be written in the alternative form (cf. equation 1 of main text)

$$L = (C/P)^p \qquad \text{(C-48)}$$

where the constant C is the basic dynamic capacity. Taking logarithms of both sides of equation (C-48) results in an alternate form of the logarithmic life equation, i.e.

$$Y = p\,x \qquad \text{(C-49)}$$

where

$$Y = \ln L, \quad x = \ln (C/P) \quad .$$

This section deals with the problem of determining whether the "semi-empirical" values of C in current use by SKF contain a bias.* For this purpose, the SKF data included values for (C/P), both for deep-groove and self-aligning bearing, where $C = f_c\ Z^{0.7}\ D_a^{1.8}\ (i \cos \alpha)^{0.7}$.

The rationale guiding this analysis was that if the values of C, in current use by SKF, are free from a bias or systematic error, then the interval of uncertainty (95 per cent confidence limit) for p based on equation (C-49), would overlap the intervals of uncertainty for p obtained from Sections 2 and 6 of this Appendix where no *a priori* values for the parameters were assumed. If the confidence intervals do not overlap, then one would suspect that the values for (C/P), and hence C, currently used by SKF contain some type of bias.

---

* SKF was the only company which included values of (C/P) with their raw data.

The estimate of $\underline{p}$ using equation (C-49) is

$$\text{(C-50)} \qquad \hat{p} = \frac{\sum_{\alpha=1}^{n} w_\alpha y_\alpha x_\alpha}{\sum_{\alpha=1}^{n} w_\alpha x_\alpha^2}$$

and the variance of $\hat{p}$ is estimated by

$$\text{(C-51)} \qquad \text{variance } (\hat{p}) = \frac{s^2}{\left(\sum_{\alpha=1}^{n} w_\alpha x_\alpha^2\right)}$$

where

$$\text{(C-52)} \qquad s^2 = \frac{\sum_{\alpha=1}^{n} w_\alpha y_\alpha^2 - (\hat{p})^2 \left(\sum_{\alpha=1}^{n} w_\alpha x_\alpha^2\right)}{(n - 1)}$$

Confidence intervals for the parameter $\underline{p}$ are then given by

$$\text{(C-53)} \qquad \hat{p} \pm t \sqrt{(\text{variance } \hat{p})}$$

where $\underline{t}$ is the appropriate value taken from Student's table having (n-1) degrees of freedom.

Substituting the numerical values with respect to deep-groove bearings (L_{10} only) in equations (C-50) and (C-51) results in

$$\text{(C-54)} \qquad \hat{p} = \frac{573.7085232}{202.3046333} = 2.84$$

and

$$\text{(C-55)} \qquad \text{variance } (\hat{p}) = \frac{(94.164039)/49}{202.3046333} = 0.009488 \ .$$

Using the same equations but substituting the numerical values for self-aligning bearings (L_{10} only) results in

$$\text{(C-56)} \qquad \hat{p} = \frac{506.902149}{216.141147} = 2.35$$

and

$$\text{(C-57)} \qquad \text{variance } (\hat{p}) = \frac{(96.816238)/39}{216.141147} = 0.011485$$

Table C-17 compares the above results with those obtained in Sections 2 and 7 of the Appendix.

TABLE C-17. COMPARISON OF VALUES FOR $\hat{p}$

	$\hat{p}$	
	Using (C/P)	No a priori values for C
deep groove (L_{10})	2.84 ± 0.19	3.00 ± 0.64
self-aligning (L_{10})	2.34 ± 0.22	1.77 ± 0.46

It is clear from Table C-17 that the values of $\hat{p}$ using the variate (C/P) show good agreement with those values calculated without making assumptions as to the nature of C. Hence there is no indication of a bias in the use of the semi-empirical values for C which SKF is currently employing.

8. References to APPENDIX C.

[1] Anderson, R. L. and Bancroft, T. A., Statistical Theory in Research, McGraw-Hill Book Co., New York, 1952, Chapters 13, 14, 15.

[2] Cramér, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1946.

[3] Dixon, W. J. and Massey, F. J., Introduction to Statistical Analysis, McGraw-Hill Book Co., New York, 1951, Chapter 11.

[4] Finney, D. J., Statistical Method in Biological Assay, Charles Griffin and Co., London, 1952.

[5] Hald, A., Statistical Theory with Engineering Applications, John Wiley and Sons, New York, 1952, Chapters 18, 20.

[6] Kempthorne, O., The Design and Analysis of Experiments, John Wiley and Sons, New York, 1952, Chapters 4, 5.

[7] Mood, A. M., Introduction to the Theory of Statistics, McGraw-Hill Book Co., New York, 1950, Chapter 13.

[8] Plackett, R. L., "A Historical Note on the Method of Least Squares," Biometrika, 36 (1949), pp. 458-460.

[9] Wilks, S. S., Mathematical Statistics, Princeton University Press, Princeton, 1946, Chapter VIII.

APPENDIX D

NON-PARAMETRIC CONFIDENCE INTERVALS FOR L_{10}, L_{50}

The purpose of this Appendix is to indicate a method of obtaining confidence intervals for L_{10}, L_{50} that does not depend on the underlying functional form of the distribution of fatigue lives, is simple to employ, and is independent of any particular estimates of L_{10}, L_{50}. On the other hand, because it requires so little information, the method will generally yield rather wide limits, unless the sample size is very large. This method is termed non-parametric because it does not depend on a specific type of distribution (such as Weibull, log-normal, etc.).

Description of method

Let the n observed fatigue lives for a test group of bearings (including "intacts") be ranked in increasing order and denoted by $x_1 \leq x_2 \leq \ldots \leq x_i \leq \ldots \leq x_n$. The confidence intervals are of the form (x_r, x_s). Table D-1 shows, for given size n, the ranks r and s of the x's which will enclose the unknown population value, L_{10}, with probability approximately 95 percent. Similarly, table D-2 gives the ranks for determining confidence intervals for L_{50}. An example will make this clear.

The data in the specimen Worksheet, Appendix A, having n = 23 bearings are taken for illustration. This worksheet is summarized as Record No. 2-4 in Table A-2. The number of bearings in the test group is n = 23. To calculate the confidence intervals for rating life L_{10}, enter table D-1 with n = 23 and read r = 0, s = 5 and $P_{10} = 0.927$. This means that the .927 confidence interval for L_{10} is given by the interval between 0 and the 5th smallest observation, i.e., $x_5 = 42.12$. Hence with probability .927, the interval (0,42.12) will contain the population value of L_{10}.

To calculate a confidence interval for the median life L_{50}, enter table D-2 with n = 23 and read r = 7, s = 17, and $P_{50} = 0.965$. This means that the .965 confidence interval for L_{50} is given by (x_7, x_{17}) = (48.48, 93.12).

It is interesting to note that the estimates of L_{10} and L_{50}, 28.5 and 69.2 respectively, calculated on the assumption of an underlying Weibull distribution, both fall within the above non-parametric confidence limits.

Mathematical formulation

Let $x_1 \leq x_2 \leq \ldots \leq x_n$ be an ordered sample of n observations from some population having the probability density function f(x). Define $L_{(100p)}$ by

$$\text{(D-1)} \qquad p = \int_{L_{(100p)}}^{\infty} f(x)\, dx \qquad (0 < p < 1),$$

that is, if p = 1/10, then L_{10} is that value exceeded by 90 percent of the population; and if p = 1/2, then L_{50} is the value exceeded by 50 percent of the population. Then the interval determined by the r^{th} and s^{th} ordered observations define a confidence interval such that

$$\text{(D-2)} \qquad \text{Prob}\left\{x_r \leq L_{(100p)} \leq x_s\right\} = P_{(100p)}$$

where

$$\text{(D-3)} \qquad P_{(100p)} = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i}$$

Equation (D-2) is read as "the probability that the interval (x_r, x_s) includes the true value, $L_{(100p)}$, is $P_{(100p)}$". Tables D-1 and D-2 which refer to L_{10} and L_{50} respectively, list values of r and s for n = 6(1)50, and selected higher values of n corresponding to the larger size test groups included in the original data.

An outline of the derivation of equation (D-2) can be found in Mood [1], page 388 ff. and Wilks [3], p. 14 ff.

Discussion

Note that in order to calculate the non-parametric confidence intervals for L_{10} or L_{50}, it is not necessary for all the n bearings to be tested to fatigue failure. Referring to the Specimen Worksheet in Appendix A, the test could have ceased after the 5th bearing failed, if all that was desired was a confidence interval for L_{10}.

However, the problem of finding a suitable non-parametric estimate for L_{10} (or L_{50}) has not yet been satisfactorily solved. An obvious device is to say that x_i estimates the (i/n)-percent point, e.g., for L_{10} and L_{50}, i = n/10, n/2, respectively; some workers use x_i as an estimate for the $\left(\frac{i}{n+1}\right)$-percent point--a procedure

which is rigorously correct only for the rectangular distribution; still others use $\frac{(i-\frac{1}{2})}{n}$ in place of $\frac{i}{n+1}$. Which, if any, of these procedures may be correct for ball bearing fatigue data would be a subject for future research. Below are summarized the results of the above procedures applied to the aforementioned worksheet in Appendix A.

	Method of Appendix B	i/n	i/(n+1)	(i-½)/n	Non-parametric Confidence Intervals
L_{10}	28.5	30.14	30.55	32.18	(0,42.12)
L_{50}	69.2	68.22	67.80	67.80	(48.48,93.12)

To obtain these results, interpolation has been used as illustrated by the following example. For n = 23, if the "(i/n)" procedure is used to estimate the .10-percent point L_{10}, then i/23 = .10, i = 2.3, giving x_i = "$x_{2.3}$" as the estimate. The quantity in quotes is merely a shorthand expression for the interpolated value between x_2 and x_3:

$$x_{2.3} = .7\, x_2 + .3 x_3 = 30.14$$

References to Appendix D

[1] Mood, A. M., Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, 1950.

[2] Nair, K. R., "Table of Confidence Intervals for the Median in Samples from Any Continuous Population," Sankhyā, 4, pp. 551-558.

[3] Wilks, S. S., "Order Statistics,", Bulletin of the American Mathematical Society, 54, pp. 6-50 (January 1948).

TABLE D-1 NON-PARAMETRIC CONFIDENCE INTERVALS FOR L_{10}

Sample size n	r	s	P_{10}	Sample size n	r	s	P_{10}
6	0	3	.984	31	1	8	.952
7	0	3	.974	32	1	8	.954
8	0	3	.962	33	1	8	.955
9	0	3	.947	34	1	8	.955
10	0	3	.930	35	1	8	.955
11	0	4	.982	36	1	8	.954
12	0	4	.974	37	1	8	.952
13	0	4	.966	38	1	8	.950
14	0	4	.956	39	1	8	.947
15	0	4	.944	40	1	8	.943
16	0	4	.932	41	1	8	.939
17	0	5	.978	42	1	8	.934
18	0	5	.972	43	1	9	.965
19	0	5	.965	44	1	9	.962
20	0	5	.957	45	1	9	.959
21	0	5	.948	46	1	9	.956
22	0	5	.938	47	1	9	.952
23	0	5	.927	48	1	9	.947
24	0	6	.972	49	1	9	.942
25	0	6	.967	50	2	10	.942
26	0	6	.960	55	2	10	.934
27	0	6	.953	60	2	11	.954
28	0	6	.945	67	2	13	.95*
29	1	7	.931	94	3	15	.95*
30	1	8	.950				

*These values are based on using a normal approximation for large samples.

TABLE D-2. NON-PARAMETRIC CONFIDENCE INTERVALS FOR L_{50}

(ADAPTED FROM NAIR, [2])

Sample Size n	r	s	P_{50}	Sample Size n	r	s	P_{50}
6	1	6	.969	29	9	21	.976
7	1	7	.984	30	10	21	.957
8	1	8	.992	31	10	22	.971
9	2	8	.961	32	10	23	.980
10	2	9	.979	33	11	23	.965
11	2	10	.988	34	11	24	.976
12	3	10	.961	35	12	24	.959
13	3	11	.978	36	12	25	.971
14	3	12	.987	37	13	25	.953
15	4	12	.965	38	13	26	.966
16	4	13	.979	39	13	27	.976
17	5	13	.951	40	14	27	.962
18	5	14	.979	41	14	28	.972
19	5	15	.981	42	15	28	.956
20	6	15	.959	43	15	29	.968
21	6	16	.973	44	16	29	.951
22	6	17	.983	45	16	30	.964
23	7	17	.965	46	16	31	.974
24	7	18	.977	47	17	31	.960
25	8	18	.957	48	17	32	.971
26	8	19	.971	49	18	32	.956
27	8	20	.981	50	18	33	.967
28	9	20	.964	55	20	36	.970
				60	22	39	.973
				67	26	42	.950
				94	38	56	.95*

* This value is based on normal approximation for large samples.

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.25) and its Supplement ($0.75), available from the Superintendent of Documents, Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS