STATISTICAL ANALYSIS OF FATIGUE DATA

By Robert Plunkett[1]

It is commonly assumed in analysing fatigue data that there is a definite functional relationship between life in number of cycles and stress level. However, as has been pointed out several times (**1**, **2**),[2] an examination of the data shows considerable scatter. Even with carefully prepared smooth specimens, all from the same heat of steel, treated in the same manner and tested in the same laboratory, a range of 2 to 1 in number of cycles for failure at the same stress level is normal (**1**) and a range of 10 to 1 is not unusual (**2**). If the specimens are tested by different laboratories, slightly varying techniques will introduce further scatter (**3**).

For full-size factory-run items the scatter becomes considerably worse (**3**, p. 75; **4**). In addition to the reasons for the scatter in laboratory specimens, there is the usual production tolerance in metal strength and the variable surface conditions due to production finish. This variation in life for a given stress makes a statistical analysis of the data advisable so that full benefit may be derived from expensive data.

Statistical Approach

Previous work on statistical analysis may be divided into two approaches. The first is an attempt to predict the data from an analysis of the distribution of flaws in the material (**5**). The second is an analysis of the experimental data itself (**1**, **2**). This report falls in the second category.

The approach is based on that of Weibull. If we have a large number of similar specimens and test them in fatigue at the same stress level, S_1, we will find a range of values for the number of cycles to failure. If P is the fraction of the total number of specimens that fail at or below a certain number of cycles, N, then N is a monotonically increasing function of P, $N_1(P)$. If the same set of specimens had been tested at a different stress level, S_2, there would be another function of P, $N_2(P)$. Experience shows that for a given value of P, N is a monotonically decreasing function of S. If $N(P)$ were a constant, that is, all specimens failed at the same number of cycles for a given S, then by definition N must be a monotonically decreasing function of S; actually this carries over to the variable case. Thus, N is a continuous function of S and P, $N\ (S,P)$, with continuous derivatives. Since the derivatives are continuous, P must be a continuous function of N and S, $P\ (N,S)$.

It is the purpose of this report to give a method for plotting the curves of constant P in the N-S plane, that is, the curves $P\ (N,S)$ = constant. It is possible to interpret these curves as the S-N curves for a given specimen under the impossible theoretical condition that one specimen could be tested to destruction at different stress levels. This would be true if the S-N curves for the different specimens did not cross; if the scatter

[1] Electro-Mechanical Division, General Engineering Laboratory, General Electric Co., Schenectady, N. Y., formerly with Hughes Tool Co., Houston, Tex.

[2] The boldface numbers in parentheses refer to the list of references appended to this paper, see p. 53.

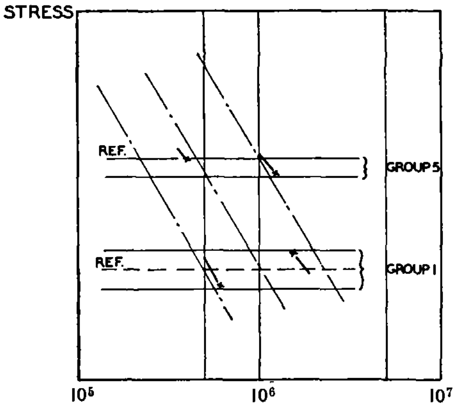

FIG. 1.—Adjustment of Points to Nominal Stresses.

is caused only by different strengths of the metal due to heat treatment and stress raisers, this is undoubtedly true, if it is due to different chemical compositions it may not be. In either case it does not affect the problem as previously stated.

As Weibull has pointed out, if there are available sets of values of N at different constant S values, the analysis is relatively simple. However, for practical reasons, this is not always possible. The data considered here is for $4\frac{1}{2}$ in., 16.6 lb per ft, grade D drill pipe which was loaded in a rotating cantilever-beam machine. The specimen did not break at the point of maximum stress, so that it was impossible to determine in advance the stress level. It will be assumed that it is possible to estimate roughly by eye the general trend of the points and thus project them along lines of constant P to a small number of different values of S. In common with usual practice, it will be found easier to plot the experimental points either as S *versus* log N or log S *versus* log N. The procedure then is to group the points in bands of constant width in the S or log S direction, project them along ap-

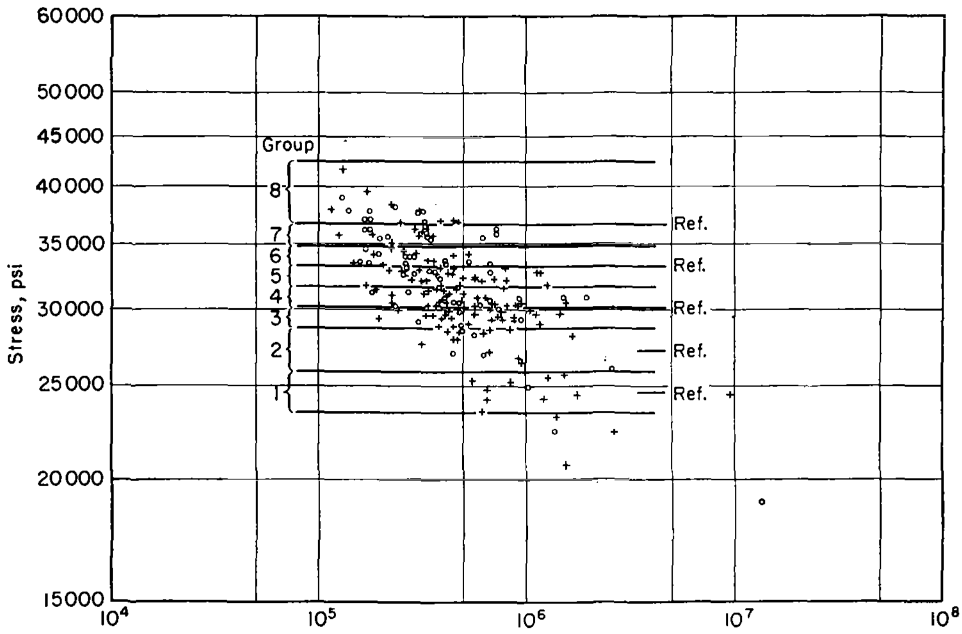

FIG. 2.—Cycles to Failure of Drill Pipe.

proximate lines of constant P to values of S which are the mid values or boundary values of the bands, and record these adjusted values on N. On these plots, the S log N or log S-log N curves will not deviate appreciably from straight lines for small changes in S. If the width of the band is small enough, quite appreciable errors in the slopes of the curves will have little effect on the adjusted values of N. Figure 1 is a typical example showing the adjustment of four such points used in the subsequent analysis.

the analysis were where the short diagonals cross the values of the selected stress and will be called the adjusted values of N. The 10, 50, and 90 per cent lines have been plotted as later determined from Fig. 3 to indicate the amount of error in the assumed slopes. If the error is too great, it may be necessary to repeat the analysis a second time using the more accurate P curves. While the log S - log N curves at constant P seemed to be straight lines in this case, this is not essential to the method and any smooth curve will do.

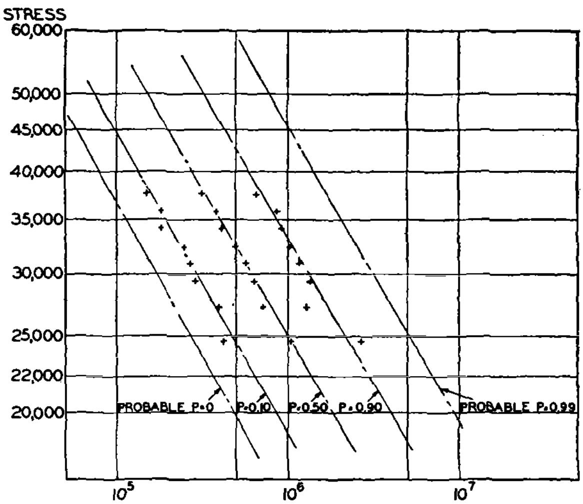

FIG. 3.—Constant Probability S-N Curves for Drill Pipe.

The four crosses are points giving the stress and number of cycles at failure for four specimens selected from Fig. 2. The slope of the P curves was estimated roughly by counting off half the specimens in various bands of stress range in Fig. 2, and calling these the median values. It was possible in this case to fair in a reasonably good straight line through these points. The experimental points were then adjusted to previously selected stress levels by projecting them as indicated by the short diagonal lines; the values of N used in

EFFECT OF FINITE SET OF POINTS

Weibull (1) shows that if a set of n cards is drawn at random from a set of m cards which are numbered from 1 to m, and then the n cards are arranged in ascending order of their numbers, the mathematical expectation of the kth card of the set is $\frac{k}{1 + n}\, m$. This means that if we test n specimens at a given stress level, and arrange the values of N so determined in ascending order, the mathematical expectation of the value of P for the kth value of N is

$\frac{k}{1 + n}$. There will, of course, be a certain amount of scatter in these values, with the scatter decreasing as n increases.

What was done in this case was to plot, on probability paper, the adjusted values of log N *versus* $\frac{k}{n+1}$ for each band of log S. It was not expected nor necessary that these would be normally distributed and thus show a straight line on probability paper, but since the scatter is due to a large number of different causes, the central limit theorem (6) indicates that the curve will be approximately a straight line for values of P which are not near either 0 or 1.

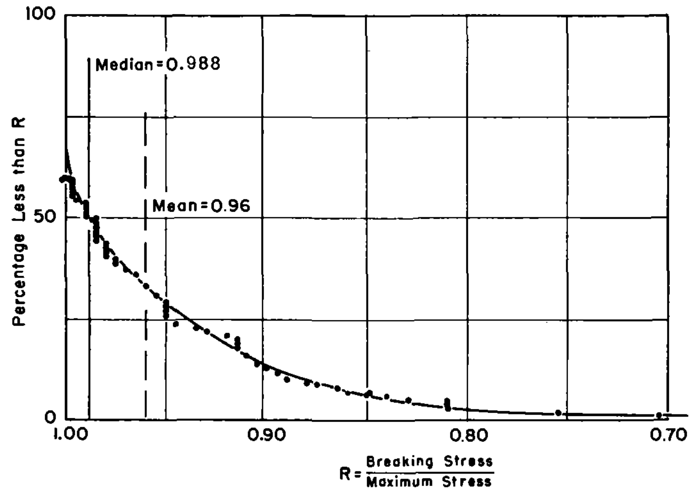

FIG. 4.—Deviation of Breaking Stress from Maximum Stress (cantilever beam testing).

INTERRUPTED TESTS

These particular data are further complicated by the fact that some of the tests were interrupted before failure took place. This was caused by the fact that both the pipe and the tool joint were being tested, and when either one failed the test was perforce halted. If these values were ignored, it is clear that the data would be biased in favor of the lower strength values. The assumption was made that the distribution of P and N for the interrupted tests in a given stress band was the same as that for those which failed at higher values of N. This is based on the doctrine of equal probabilities, which is admittedly dangerous, but used in default of anything better. If N_1 is the value of N at which a test was discontinued, this would be equivalent to weighting the remaining points in the ratio of the total number of points remaining above N_1 plus the dropped one, to the number of points remaining. Thus, if there were 20 points originally arranged in ascending order of N, and the sixth one were discontinued, each of the first 5 would have

TABLE I.—SPECIMENS TESTED.

Group	ln 10^{-4} S	Reference ln 10^{-4} S	Total Points	Incomplete	Failed
1........	0.85 to 0.95	0.90	11	1	10
2........	0.95 to 1.05	1.00	17	5	12
3........	1.05 to 1.10	1.10	27	4	23
4........	1.10 to 1.15	1.10	38	10	28
5........	1.15 to 1.20	1.20	33	8	25
6........	1.20 to 1.25	1.20	23	13	10
7........	1.25 to 1.30	1.30	19	11	8
8........	1.30 to 1.45	1.30	18	9	9
			186	61	125
					61
					186

weight 1 and each of the last 14 would have weight $\frac{15}{14}$; it should be noted that the sum of the weights is still 20. It may be assumed that the value of P to be used is the average value for the weighted points, since any other reasonable assumption would make little difference in P; that is, if w_i is the weight of the ith point,

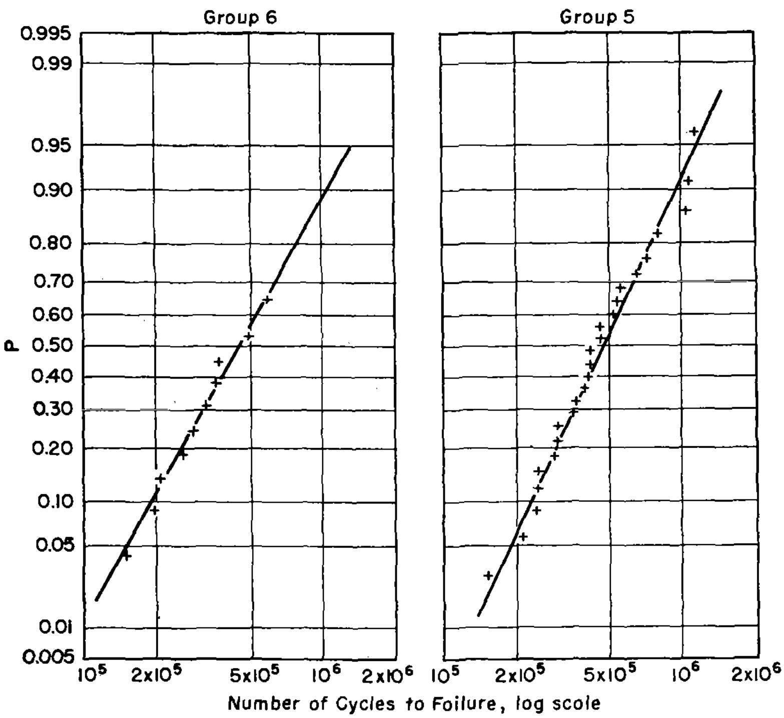

FIG. 5.—Sample P-ln N Plot.

$$E(P_k) = \frac{\sum_{1}^{k-1} w_i + \frac{w_k + 1}{2}}{\sum_{1}^{m} w_i + 1}$$

$$\sum_{1}^{m} w_i = n$$

where:

$E\ (P_k)$ = mathematical expectation of the kth point,

m = total number of completed tests, and

n = total number of tests both completed and incomplete.

This is complicated still further by the fact that failure does not occur at the maximum stress in the pipe, whereas the maximum stress is the only one available for the incomplete test. Figure 4 is a plot of the ratio of the actual breaking stress to the maximum stress for those specimens which did break; it shows the percentage of those specimens which had less than a given ratio and indicates that small error is introduced by using the maximum stress. The error in N is that which is caused by decreasing the stress by the given percentage and can be seen from Fig. 1 to be small; this is further helped by the fact that the P values of the points are little affected by tests which were abandoned at values of N just less than those of the given specimen.

METHOD OF PLOTTING DATA

First all the points were plotted on a log S - log N graph, Fig. 2; the crosses are the completed tests, and the circles are the values of N at which tests were abandoned. Log S was used, as a preliminary study indicated that this would give almost straight lines for P = constant for the range of interest. The set of data was broken up into eight groups of points, those falling in the bands of log S shown in Table I. It would probably have been better to use a middle value of S for reference in all groups; then the errors for those greater than the reference value would balance the errors for those less, see Fig. 1. The values for $E(P_k)$ were calculated as indicated previously (see Appendix). Then the values of log N_k *versus* $E(P_k)$ were plotted on probability paper (Fig. 5). It can be seen that regardless of the shape of the curve, the $P = 0.50$ points may be determined very accurately. Because of the large number of discontinued tests, many of the $P = 0.90$ points are not reached, but the $P = 0.10$ points may be found with a little more uncertainty than the median points. For reasons that will be shown later, it is felt that straight lines give the best approximation to the curves between $P = 0.10$ and $P = 0.80$. Thus, an extrapolation to $P = 0.90$ will not be far in error. The values of the 0.10, 0.50, and 0.90 points were plotted in Fig. 3, after being adjusted to the mid points of the group for each band, with the exception of group 1 where they were plotted on the ln $10^{-4}S = 1.325$ line. It can be seen that straight lines can be drawn through each set of points and that they are almost parallel; one such set of data is insufficient to say that they do not meet at a point at S equal to the ultimate load, as has been suggested by Almen (3, p. 66), but it is not readily apparent that they must.

A further check on the shape of the curve was made by plotting all the points on one consolidated graph. Since the P lines seem to be parallel, it is unnecessary to adjust the width of the curves

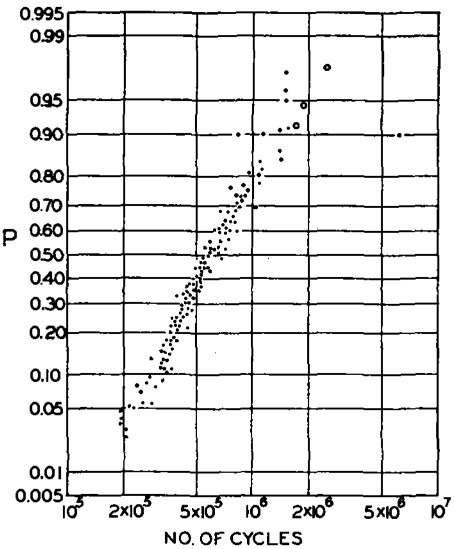

FIG. 6.—Consolidated P-ln N Plot, Eight Groups. All groups displaced sideways so that $P = 0.50$ points coincide.

for log N. The eight curves for the different bands were shifted sideways enough to make all the $P = 0.50$ points fall on the same ordinate and then all the points were picked off on Fig. 6. It can be seen that the points fit a straight line very nicely. The amount of scatter is misleading, since if the points were formed into one sequence there can be only one value of N for each value of P and the points must lie on a fairly smooth curve, at least in the middle of the range. The only purpose of Fig. 6 then, is to show that a straight line of a given slope will fit all groups fairly well.

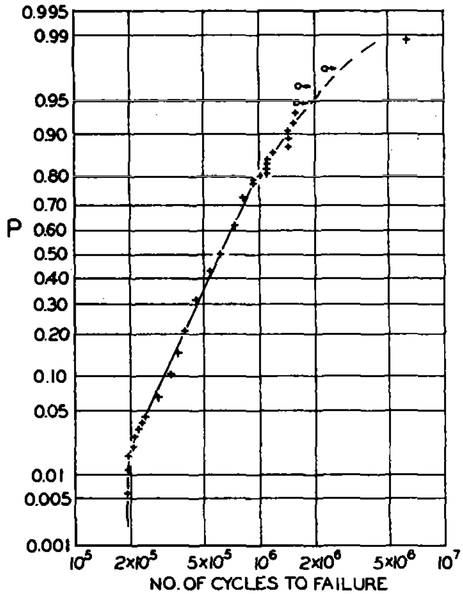

FIG. 7.—Consolidated P-ln N Plot, Replotted from Fig. 6 in One Sequence.

the last 15 points were plotted on Fig. 7, with about every tenth point plotted in the middle of the range. The last three discarded points are also indicated. There seems to be a possibility that the lower end of the scale cuts off, that is, there is a definite curve for $P = 0$, no specimen will fail below this number of cycles for each stress level. The upper end of the curve shows more scatter and is less accurate because of the weighting process, but there seems to be a definite indication of curving over to the right as indicated by the dash line. There is undoubtedly a definite $P = 1.00$ curve, but it may well coincide with the S-N curve for small polished specimens which is well out of the range shown. The small amount of scatter in the middle range of Fig. 7, is also of interest.

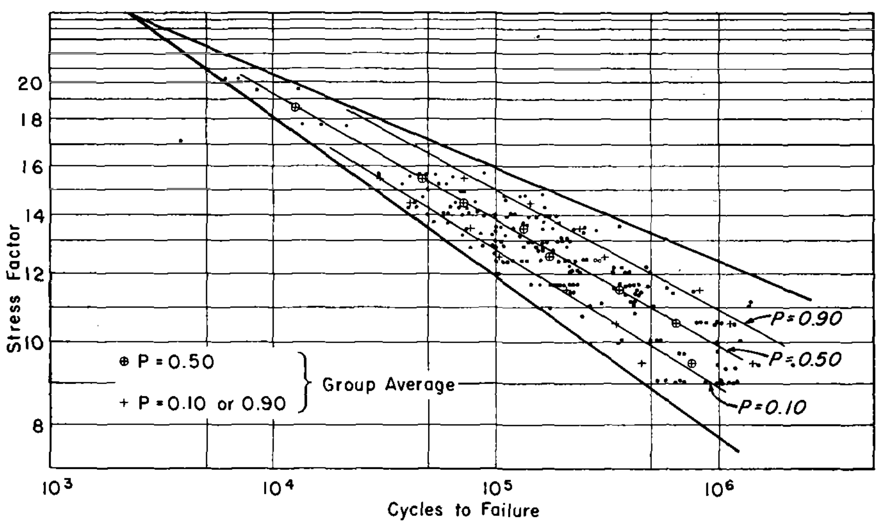

FIG. 8.—Constant Probability S-N Curves for Rear Axle Pinions (after Almen).

The points in Fig. 6 were arranged in one sequence to include the discarded points and the percentages for the whole group figured as before. The first 8 and

TEST ON ALMEN'S DATA

To test the method further, it was used on the data from rear axle pinions shown in Fig. 28 of Almen's paper (3).

The points were divided into eight groups centered at 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, 15.5 and 18.5 on his stress level scale. These were plotted on probability paper and the 10, 50 and 90 per cent points replotted on the original chart, Fig. 8. It will be seen that straight lines, slightly converging, fit all points fairly well except at the 9.5 level. This last indicated that some 15 points were missing from the original plot.

It is interesting to note that if the P lines do converge, it is at a point considerably above the ultimate strength of the steel.

CONCLUSIONS

As a result of this analysis several conclusions may be reached:

1. Figure 3 gives, with considerable accuracy, the number of cycles at which 10, 50, and 90 per cent of the mill run of $4\frac{1}{2}$ in., 16.6 lb per ft, grade D drill pipe may be expected to fail. Very few, if any, will fail below the line marked $P = 0$.

2. A very small fraction of the tests may be expected to show extremely large values of N at failure for a given stress—beyond the line marked $P = 0.99$.

3. Practically no pipe will last indefinitely at stresses above 22,000 psi. If there is an endurance limit it is below this value, but the data are insufficient to show it.

4. The stress range ratio at 5×10^5, cycles for 90 per cent of the specimen, from $P = 0$ to $P = 0.9$, is about $2\frac{1}{4}$, the range from $P = 0.5$ to $P = 0.9$ is a little less than half of this. The median stress at 10^6 cycles is about 25,000 psi, the 90 per cent stress at 10^6 cycles is about 33,000 psi.

5. If it is desired to design a tool joint such that 99 per cent of the failures will occur in the body of the pipe, a safe method would be to make sure that the $P = 0$ line for the tool joint approximated the $P = 0.9$ curve for the drill pipe, then the $P = 0.1$ curve for the tool joint would be far enough out that less than 1 per cent of the failures would occur in the tool joint.

The exact equation is

$$F = \int_{Q=0}^{P=1} (1 - P(\log N)) \frac{dQ}{d \log N} d(\log N)$$

where:

$P(\log N)$ = P as a function of $\log N$ for a given stress,

Q = similar function for the tool joint, and

F = fraction of failures in the tool joint.

On integrating by parts it is easily seen that, also:

$$F = \int_{Q=0}^{P=1} Q \frac{dP}{d \log N} d(\log N)$$

An upper limit is easily found since

$$F = \int_{Q=0}^{P=1} Q \frac{dP}{d \log N} d(\log N) \leq Q(N_2) \int dP$$

$$= Q(N_2)(P(N_2) - P(N_1)) = Q(N_2)(1 - P(N_1))$$

where:

$Q(N_2)$ = the value of Q where $P = 1$ and

$P(N_1)$ = the value of P where $Q = 0$.

For example, if $Q(N_2) = 0.1$, $P(N_1) = 0.9$ then:

$$F \leq 0.1\ (1 - 0.9) = 0.01$$

Less than 1 per cent will fail in the tool joint.

This inequality is equivalent to saying that if 0.9 of the pipe specimens have broken before any of the joints have broken (N_1 cycles), and 0.1 of the joints would have broken by the number of cycles (N_2) at which all of the pipe would have broken, each being tested independently, then the worst possible situation is where the additional 0.1 of the pipe specimens do not fail until N_2 cycles. In the combination of the two, only 0.1 of the total number is left by N_1 cycles due to failure in the pipe. Of this 0.1, 0.1 will fail in the joint by N_2 cycles at which time all the remaining specimens will fail in the pipe. Thus under this worst condition 0.1×0.1 will fail in the tool joint or 0.01 all told. Thus 0.90 will fail in the pipe by N_1 cycles, 0.01 in the joint between N_1 and N_2 and 0.09 in the pipe at N_2, a total of 1.00 in all. The actual number of failures in the joint must be less than this since the pipe specimens continue to fail between N_1 and N_2 cycles.

Acknowledgment:

I should like to thank F. C. Scott, Head of Research Engineering, Hughes Tool Co., for permission to publish this paper and to use the data accumulated by his laboratory over a period of some twenty years. The stimulating discussion of this problem with H. B. Woods of the same department has been most helpful.

REFERENCES

(1) W. Weibull, "Statistical Representation of Fatigue Failures in Solids," *Kungl Tekn. Hogsk. Hand.*, No. 27, Stockholm (1949).

(2) R. E. Peterson, "Approximate Statistical Method for Fatigue Data," ASTM BULLETIN, No. 156, p. 50, January, 1949.

(3) J. O. Almen, "Fatigue of Metals as Influenced by Design and Internal Stresses," Symposium on Surface Stressing of Metals, Am. Soc. Metals, p. 68, February, 1946.

(4) W. S. Bachman, "Fatigue Testing and Development of Drill Pipe-to-Tool Joint Connections," *World Oil*, Vol. 132, p. 104, January, 1951.

(5) A. M. Freudenthal, "Statistical Aspect of the Fatigue of Materials," *Proceedings*, Royal Soc. (A), Vol. 187, pp. 416–429 (1946).

(6) H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, p. 213 (1946).

APPENDIX

SAMPLE CALCULATIONS

$$P(N_k) = \frac{\sum_{1}^{k-1} w_i + \frac{w_k + 1}{2}}{n + 1}$$

$\sum w_i = n$ (sum of all weighted values)

$$w_i = \prod_{k=1}^{i} \frac{j_k + m_k}{j_k}$$

j_k = total number left
m_k = number dropped between $k - 1$ and k

EXAMPLE (*Group 3*):

$i = 17$ $j_i = 8$ (total left)
$m_i = 1$ (dropped between $i = 16$ and $i = 17$)

$$\frac{j_i + m_i}{j_i} = \frac{8 + 1}{8} = 1.125$$

$$w_i = 1.125 \times 1.111 = 1.250$$

$$\sum w_i = 17.002 + 1.250 = 18.252$$

$$P(N) = \frac{\sum_{1}^{16} w_i + \frac{w_{17} + 1}{2}}{n + 1}$$

$$= \frac{17.002 + \frac{1.250 + 1}{2}}{27 + 1} = 64.8$$

GROUP 3, $n = 27$

i	Number Dropped	$\frac{j + m_i}{j_i}$	w_i	Σw_i	P_i
1		1	1.000	1.000	3.6
2		1	1.000	2.000	7.1
	1				
3		1.042	1.042	3.042	10.8
4		1	1.042	4.084	14.5
5		1	1.042	5.126	18.2
6		1	1.042	6.168	21.9
7		1	1.042	7.210	25.6
8		1	1.042	8.252	29.4
9		1	1.042	9.294	33.1
10		1	1.042	10.336	36.8
	1				
11		1.067	1.111	11.447	40.6
12		1	1.111	12.558	44.6
13		1	1.111	13.669	48.6
14		1	1.111	14.780	52.6
15		1	1.111	15.891	56.5
16		1	1.111	17.002	60.5
	1				
17		1.125	1.250	18.252	64.8
18		1	1.250	19.502	69.2
19		1	1.250	20.752	73.7
	1				
20		1.250	1.563	22.315	78.6
21		1	1.563	23.878	84.3
22		1	1.563	25.441	89.9
23		1	1.563	27.004	95.5
23 +	4 = 27				

DISCUSSION

Mr. J. O. Almen.[1]—Mr. Plunkett has shown an S-N diagram of a large number of automobile spiral bevel gears tested to failure in complete rear axle assemblies. It is seen that the stress scale in this chart is given as a "load factor" which is a necessary deviation from conventional fatigue diagrams because the stresses in conventional units are not known.

In early plots of the same data[2] the stress scale was arbitrarily shown as pounds per square inch for the purpose of simplifying the job of selling the resulting empirical gear strength formula to potential users. After the gear strength formula had been universally accepted the diagrams were altered to express stress as proportional to the applied load but not in terms of pounds per square inch.

There has been considerable misunderstanding about these tests which, in view of their present use by Mr. Plunkett, seems to call for clarification.

The laboratory tests were devised to conform to the failure of gears in actual owner service. The laboratory test loads were selected to produce fractures of gear teeth that duplicated the fractures that occurred in owner service in regard to character and location of fractures. This required testing the gears by applying to each propeller shaft a torque equal to the maximum torque delivered by its engine divided by the low gear ratio of a conventional three speed transmission. No reduction was made to allow for friction losses in the transmission.

Since the great majority of service fractures occurred in automobiles that had been subjected to abusive driving on steep hills, in deep mud, and with "grabbing" clutches, etc., it will be appreciated that the plotted data were only remotely related to normal fatigue expectations of automobiles used by normal drivers of the period (15 to 20 yr ago) when the tests were made. This relationship is even more remote today because better roads have replaced steep hills and mud and the old style friction clutches have been largely replaced by hydraulic clutches. By these improvements the abusive driver is protected against himself.

Mr. Plunkett has made proper use of the spiral bevel gear fatigue data in so far as they relate to fatigue tests under controlled loads, but it is not possible to establish permissible service loads for use in gear design from statistical analysis of these data.

After 20 yr experience in many millions of automobiles with gears designed by the empirical formula that was developed from the data shown in Mr. Plunkett's paper we can only say that our original estimate of adequate automobile rear axle gear strength has been fully justified. That estimate is represented by a line approximately bisecting the scatter band shown in Mr. Plunkett's paper.

Mr. W. Weibull[3] (*by letter*).—The

[1] Research Consultant, Research Laboratories Div., General Motors Corp., Detroit, Mich.

[2] J. O. Almen and A. L. Boegehold, "Rear Axle Gears: Factors Which Influence Their Life," *Proceedings*, Am. Soc. Testing Mats., Vol. 35, Part II, p. 99 (1935).

[3] Scientific Director, Bofors Co., Bofors Sweden.

adjustment process shown in this article seems sound and successful. In principle, the weighting process is a proper method for dealing with the interrupted tests. For reasons stated below, I am proposing some small modifications. In addition to this approach, it would be possible, I think, to analyze the combined distribution of the pipes and the tool joints into separate distributions if the tool joints all belonged to the same family.[4] If the

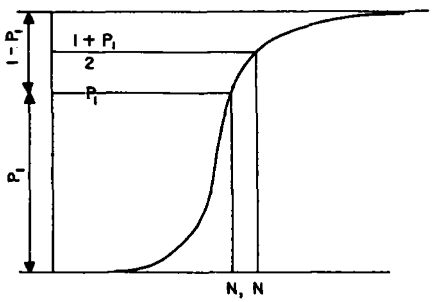

FIG. 9.—Most Probable Life of Incomplete Test.

distribution function of the first one is P_1 and of the second one is P_2, the combined distribution function is found from:

$$(1 - P) = (1 - P_1)(1 - P_2)$$

In some cases it is possible to separate P_1 and P_2.[5]

METHODS OF DEALING WITH INTERRUPTED FATIGUE TESTS

Method 1:

Let us first suppose that we have only one incomplete test, interrupted after N_1 stress cycles and $(n - 1)$ complete tests. If the cumulative distribution function (*cdf*) $P = F(N)$ is known, the probability P_1 corresponding to the life N_1 is easily determined (see Fig. 9).

Putting

$$P = P_1 + \frac{1 - P_1}{2} = \frac{1 + P_1}{2}$$

[4] Does not obtain in this case.—AUTHOR.
[5] W. Weibull, "The Phenomenon of Rupture in Solids," Ing. Vetenskaps Akad., Hand. No. 153, pp. 32 and 48, Stockholm (1939).

The corresponding value $N(P)$ may be said to be the most probable life of the incompletely tested specimen, as there will be 50 chances in 100 of its being greater, and just as many of its being smaller than N.

Accordingly, if we arrange the n points, including N_1, in ascending order of N, and N_1 is the k-th point of this array, the most probable order number of this specimen will not be k, but

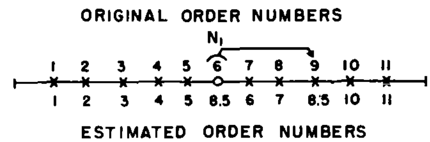

FIG. 10.—Estimated Order Numbers Method.

TABLE II.—COMPARISON OF THREE METHODS.

Order Number of Complete Tests	Corrected Order Number		
	Method 1	Method 2	Method 3
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6.2	6.17
7	7	7.4	7.33
8	8.5	8.6	8.50
9	10	9.8	9.67
10	11	11.0	10.83

$$K + \frac{n - k}{2} = \frac{n + k}{2}$$

Figure 10 shows a sample calculation using Group 1, where $n = 11$ and $k = 6$. The order number 6 has to be raised to $6 + (11 - 6) = 8.5$ and the intermediate order numbers have to be reduced by one point.

The probable life of the interrupted test is thus equal to the orignal N_9 (see Table II).

If there is more than one interrupted test, the above-mentioned rule may still be applied. The incomplete N values have to be counted after their positions have been altered. For this reason, we have to start with the highest incomplete value, then take the next highest, etc.

The estimated life of an interrupted test is evidently rather uncertain. Thus, it seems better to drop these values when plotting the data. They have, nevertheless, been useful, as they have improved the estimates of the other P-values. There is, of course, no necessity to plot all the points of the array. In the case of grouped data, for instance, it is usual to plot the class limits only, leaving most of the observations unplotted.

Method 2 (Plunkett):

The preceding method may be modified as follows. Returning to Group 1, it may be said that it is very probable that the incompletely tested specimen will have a life greater than N_7 and smaller than N_{11} (original order number), but it has just the same chance of being greater as of being smaller than N_9. Thus, the uncertainty is greater in the middle of the array above N_1 and smaller at its boundaries. This uncertainty may be equalized by the weighting process of Plunkett. Instead of giving the incomplete test the order number 8.5, we may distribute the unity over the total number left, increasing each one by 0.2 and thus having the values of Table II.

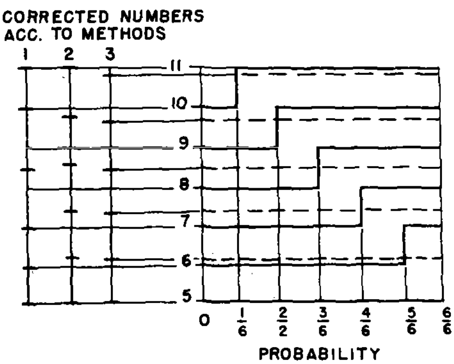

FIG. 11.—Comparison of Three Methods.

The difference between Method 1 and 2 may be illustrated by Fig. 11, where the step curves give the probabilities of the original order numbers. For instance, as there is 1 chance in 6 that N_1 falls below N_6, there is also 1 chance in 6 that the original order number 6 in reality may be 7, and 5 chances in 6 that it may be 6. Method 1 gives accordingly (on the average) the correct value in 5 cases in 6, whereas it gives an error of 1, in 1 case out of 6. Method 2 gives an error of $+0.2$ in 5 cases in 6 and an error of -0.8 in 1 case in 6 For the original order number 7, Method 1 gives 4 chances in 6 of this number being the correct one and 2 chances in 6 that the error is -1, whereas Method 2 gives the same probabilities for the errors $+0.4$ -0.6, respectively, etc.

Method 3:

This method is a slight modification of Method 2. It seems better to divide the weight into 6 parts, giving each of the complete tests an extra weight of 1/6 only (instead of 1/5). The difference between Methods 2 and 3 may be practically insignificant, but the latter method presents a more symmetrical solution. The fact that the last order number deviates from the total number of tests, n, is quite all right, as it reflects the possibility of N_1 in some cases (1/6) being greater than the greatest value of the complete tests.

The values of w_i may be regarded as the corrected order numbers. For this reason it seems better to calculate the P-values according to the formula

$$E(P) = W_i/(n + 1)$$

There is no reason, as far as I can see, to introduce the average $\frac{(w_k + 1)}{2}$. The advantage of the above-mentioned expression for $E(P)$ lies in the fact that it holds good even if $m_k = 0$.

MR. ROBERT PLUNKETT (*author's closure*).—Mr. Weibull's method for calculating the P values is the theoretically correct one according to the assumptions of the paper. The only reason for using $(W_k + 1)/2$ was to make the revised order numbers give better P values and is incorrect for his method. Using this, however, there is negligible difference between his P values and mine, even for the last complete test (99.1 *versus* 98.5). This procedure was also communicated to the author by L. G. Johnson, Research Laboratories Division, General Motors Corp.

Mr. Almen's clarifying remarks on the origin of his data should prove interesting.